

**A GUIDE TO SELECTED ALGORITHMS, DISTRIBUTIONS, AND
DATABASES USED IN EXPOSURE MODELS DEVELOPED
BY THE OFFICE OF AIR QUALITY PLANNING AND STANDARDS**

by

Ted Johnson

TRJ Environmental, Inc.
713 Shadylawn Road
Chapel Hill, NC 27514

REVISED DRAFT

Purchase Order No. P0047620

EPA Grant No. CR827033

Dr. Panos Georgopoulos, Project Manager

Environmental and Occupational Health Sciences Institute
170 Frelinhuysen Road, Piscataway, New Jersey 08855

Prepared for

U.S. Environmental Protection Agency
Office of Research and Development
Research Triangle Park, North Carolina 27711

May 22, 2002

DISCLAIMER

This draft report is being furnished to the U.S. Environmental Protection Agency by TRJ Environmental, Inc. in partial fulfillment of Grant No. CR827033. The opinions, findings, and conclusions expressed are those of the author and are not necessarily those of the U. S. Environmental Protection Agency. Comments should be addressed to Ted Johnson, TRJ Environmental, Inc., 713 Shadylawn Road, Chapel Hill, North Carolina 27514.

MAIN SECTION TITLES

<u>Section</u>		<u>Page</u>
1.	Introduction	1-1
2.	A Modeling Example	2-1
3.	Spatial and Temporal Coverage of Exposure Analysis	3-1
4.	Representing Demographic Variability	4-1
5.	Exposure Event Sequence	5-1
6.	Estimation of Microenvironmental Concentrations	6-1
7.	Processing Ambient Air Quality Data	7-1
8.	Simulation of Indoor Emissions	8-1
9.	Estimation of Human Ventilation Rate	9-1
10.	Estimation of Carboxyhemoglobin Levels	10-1
11.	Useful Recent Research	11-1

CONTENTS

Disclaimer	ii
Main Section Titles	iii
Figures	xii
Tables	xiii
Acknowledgments	xxi

<u>Section</u>	<u>Page</u>
1. Introduction	1-1
1.1 Purpose of This Guide	1-1
1.2 Exposure Models Developed by OAQPS	1-1
1.2.1 NEM-Type Models	1-2
1.2.1.1 The Deterministic Versions of NEM	1-2
1.2.1.2 The Probabilistic Versions of NEM	1-2
1.2.1.3 APEX	1-5
1.2.2 HAPEM-Type Models	1-5
1.3 Organization of This Report	1-7
1.4 References for Section 1	1-12
2. A Modeling Example	2-1
2.1 Define Study Area, Populations-of-Interest, Appropriate Subdivisions of Study Area, and Exposure Period	2-1
2.2 Divide Population-of-Interest into Representative Units	2-4
2.3 Develop Exposure Event Sequence for Each Population Unit ..	2-6
2.4 Estimate Exposure and Resulting Physiological Effects	2-14
2.4.1 Pollutant Concentration	2-14
2.4.2 Alveolar Ventilation Rate	2-18
2.4.3 Carboxyhemoglobin Level	2-20
2.4.4 The Physiological Profile Generator	2-20
2.4.5 Hourly Average Exposure Estimates	2-21
2.5 Extrapolate Unit Exposures to Population-of-Interest	2-21
2.5.1 General Population	2-21
2.5.2 Persons with Ischemic Heart Disease	2-22

CONTENTS (continued)

<u>Section</u>		<u>Page</u>
2.6	Advice to Modelers	2-25
2.7	References for Section 2	2-26
3.	Spatial and Temporal Coverage of Exposure Analysis	3-1
3.1	Census Units	3-1
3.2	Combining Census Units into Exposure Districts	3-6
3.2.1	Exposure Districts Based on Proximity to Monitor Locations	3-6
3.2.2	Exposure Districts Based on Neighborhood Types	3-6
3.2.3	Exposure Districts Based on Receptor Locations	3-7
3.3	Population-of-Interest	3-10
3.4	Exposure Period	3-13
3.5	References for Section 3	3-14
4.	Representing Demographic Variability	4-1
4.1	Defining a Comprehensive Set of Cohorts	4-1
4.2	Randomly Generate a Representative Set of Individuals	4-2
4.2.1	APEX	4-2
4.2.2	SimExpo-HAP	4-4
4.3	Choosing a Method	4-4
4.4	References for Section 4	4-6
5.	Exposure Event Sequence	5-1
5.1	24-Hour Time/Activity Data	5-1
5.1.1	The Consolidated Human Activity Database	5-2
5.1.2	Assembling Multi-Day Sequences	5-4
5.1.3	Effect of Daily Maximum Temperature on Activity Patterns	5-5
5.1.4	Factors Affecting Time/Activity Patterns Reported by Ten Studies	5-7
5.1.5	Use of Random Walk Techniques to Construct 24-Hour Activity Patterns	5-11

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
5.2 Residential Occupancy Periods	5-15
5.2.1 Introduction	5-15
5.2.2 Overview of the Methodology	5-17
5.2.3 Application of Methodology to the General U.S. Population	5-18
5.2.4 The ROPSIM Program	5-21
5.2.5 Assumptions and Possible Biases of the Methodology .	5-24
5.3 Commuting Patterns	5-25
5.3.1 Home-to-Work	5-25
5.3.1.1 The Census Commuting Database (1990)	5-25
5.3.1.2 Denver Exposure Districts	5-27
5.3.1.3 Origin-Destination Table for Denver	5-27
5.3.1.4 Denver Data Quality	5-31
5.3.2 Home-to-School	5-33
5.3.2.1 Background	5-33
5.3.2.2 The Gravity Model: Structure and Calibration .	5-38
5.3.2.3 Revisions to the HAPEM4 Model Code	5-42
5.3.2.4 Sample Application	5-43
5.4 References for Section 5	5-48
6. Estimation of Microenvironmental Concentrations	6-1
6.1 Selection of Microenvironments	6-1
6.2 General Version of the Mass Balance Model	6-3
6.3 Mass Balance Model Used in Recent Applications of pNEM/O ₃	6-7
6.3.1 Theoretical Basis and Assumptions	6-8
6.3.2 Simulation of Microenvironmental Ozone Concentrations	6-10
6.3.3 Air Exchange Rate Distributions	6-13
6.3.4 Window Status Algorithm	6-15

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
6.4	Mass Balance Model Used in Version 2.1 of pNEM/CO 6-16
6.4.1	Overview of the Model 6-16
6.4.2	Application of the Mass Balance Model to Specific Microenvironments 6-20
6.4.2.1	Indoors - Residence 6-20
6.4.2.2	Restaurants and Bars 6-22
6.4.2.3	Other Indoor Microenvironments 6-22
6.4.2.4	Automobiles and Trucks 6-23
6.4.2.5	Mass Transit Vehicles 6-29
6.4.2.6	Estimation of Mass-Balance Parameters 6-29
6.4.3	Estimation of Air Exchange Rate 6-34
6.4.3.1	The Air Exchange Algorithms 6-34
6.4.3.2	Air Exchange Rate Distributions 6-36
6.4.4	Indoor CO Emission Rates 6-39
6.4.5	Enclosed Volumes 6-40
6.4.5.1	Residences 6-40
6.4.5.2	Passenger Vehicles 6-42
6.5	The Linear Relationship Model 6-42
6.6	Parameter Estimates for the Linear Relationship Used in HAPEM4 6-46
6.6.1	Ambient Pollutant Concentration [AMB(c,t)] 6-46
6.6.2	PROX Factor 6-48
6.6.3	PEN Factor 6-49
6.6.3.1	Literature Survey 6-49
6.6.3.2	Grouping HAPs and Microenvironments 6-51
6.6.4	Estimated Values 6-54
6.6.5	Discussion 6-57
6.7	References for Section 7 6-58

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
7. Processing Ambient Air Quality Data	7-1
7.1 Selection of Data	7-1
7.2 Identification of Data Anomalies	7-3
7.2.1 The Gap Test	7-3
7.2.2 The Pattern Test	7-3
7.2.3 Visual Inspection	7-8
7.3 Estimation of Missing Values	7-8
7.3.1 The Time Series Model	7-9
7.3.2 Initial Treatment of Missing Values	7-11
7.3.3 Procedure for Simulating Missing Values	7-11
7.3.4 Examples of Filled-In Data Sets	7-12
7.4 Adjustment to Simulate Regulatory Scenarios	7-19
7.4.1 Specifications of AQI	7-20
7.4.2 Proportional Rollback — No Background	7-21
7.4.3 Proportional Rollback with Constant Background Concentration	7-22
7.4.4 Quadratic Rollback	7-23
7.4.5 Weibull Rollback	7-24
7.4.6 Evaluation of Alternative Air Quality Adjustment Procedures	7-29
7.5 Estimation of Background Concentrations and Sample Results for Proportional Rollback	7-38
7.6 Simulating Geographic Variability	7-41
7.6.1 Adding Random Components to Monitoring Data	7-44
7.6.2 Spatial Interpolation of Monitoring Data	7-49
7.7 References for Section 7	7-53

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
8. Simulation of Indoor Emissions	8-1
8.1 Simulation of Gas Stove Operation	8-1
8.1.1 Probability of Stove Use	8-1
8.1.2 Gas Stove Emission Rates	8-5
8.1.2.1 Denver Estimates	8-6
8.1.2.2 Los Angeles Estimates	8-7
8.1.3 Gas Stove Prevalence Rates	8-9
8.1.4 Probability of Gas Stove Having Electronic Ignition	8-9
8.2 Simulation of Passive Smoking	8-11
8.2.1 Estimation of CO Emission Rate from Passive Smoking	8-11
8.2.2 Residential Locations	8-11
8.2.3 Restaurants and Bars	8-14
8.2.4 Passenger Vehicles	8-16
8.3 References for Section 8	8-17
9. Estimation of Human Ventilation Rate	9-1
9.1 Estimating Oxygen Uptake Rate	9-1
9.2 Estimating Ventilation Rate from Oxygen Uptake Rate	9-2
9.3 Tests to Identify Unrealistic Values of Oxygen Uptake Rate	9-7
9.4 A Probabilistic Algorithm for Estimating V_E	9-9
9.5 Estimation of Alveolar Ventilation Rate (V_A)	9-17
9.6 Limitations of the Algorithm	9-17
9.7 References for Section 9	9-21
10. Estimation of Carboxyhemoglobin Levels	10-1
10.1 The Coburn-Forster-Kane Model	10-1
10.2 Mathematical Structure of the CFK Model Used in pNEM/CO	10-2
10.3 Implementation of the CFH Model in pNEM/CO	10-6
10.3.1 The COHb Module	10-6
10.3.2 Overview of Input Data Requirements	10-6

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
10.4	Computation of Input Data for the COHb Module 10-7
10.4.1	Input Data Supplied at Start of the pNEM/CO Computation 10-13
10.4.2	Input Data Supplied with Each Exposure Event 10-16
10.5	References for Section 10 10-21
11.	Useful Recent Research 11-1
11.1	Advanced Temperature-Related Comfort Indices 11-1
11.1.1	Heat Index 11-2
11.1.1.1	The Steadman Model 11-2
11.1.1.2	NWS Heat Index 11-6
11.1.1.3	Approximate Calculation of Heat Index 11-7
11.1.2	Wind Chill 11-7
11.1.2.1	Development of the WCT Index 11-8
11.1.2.2	The WCT Index Model 11-8
11.1.2.3	WCT Index Algorithms 11-9
11.1.2.4	Ongoing Work on Temperature Indices 11-9
11.2	Factors Affecting Window Position 11-11
11.2.1	Residences (Durham, NC) 11-11
11.2.2	Motor Vehicles (Houston, TX) 11-11
11.2.2.1	Study Methodology 11-11
11.2.2.2	Results 11-14
11.2.2.3	Discussion and Recommendations 11-26
11.2.3	Motor Vehicles (Chapel Hill, NC) 11-28
11.3	Air Exchange Rates in Residences 11-28

CONTENTS (continued)

<u>Section</u>	<u>Page</u>
11.4	Variation in Ozone Concentrations with Height Above Ground 11-28
11.4.1	Introduction 11-28
11.4.2	The Cincinnati Tower Study 11-29
11.4.2.1	Test Procedures 11-29
11.4.2.2	Test Results 11-30
11.4.2.3	Supplemental Analyses of Tower Data 11-33
11.4.2.4	Sample Application of Tower Study Results to pNEM/O3 Analyses 11-34
11.4.3	The 1995 Cincinnati Roadside Ozone Study 11-34
11.4.3.1	Test Procedures and Results 11-34
11.4.3.2	Sample Application of Roadside Results to pNEM/O3 Analyses 11-44
11.4.4	Development of the Final Adjustment Algorithm 11-44
11.5	References for Section 11 11-46
 <u>Appendices</u>	
A	Distribution of Energy Expenditure Rates Associated with CHAD Activity Descriptors A-1
B	Microenvironmental Factors Developed for HAPEM4 B-1
C	Distributions and Equations Used in the Ventilation Rate Algorithm C-1

FIGURES

<u>Figure</u>	<u>Page</u>
3-1 Selected census tracts in Durham, North Carolina	3-2
5-1 Fixed-site CO monitoring sites used to define Denver exposure districts . . .	5-28
5-2 Overview of prototype school commute module	5-37
5-3 Map of Harris County showing boundaries of school districts and census tracts	5-44
7-1 Computer plot of 1977 hourly average nitrogen dioxide data (ppm) reported by monitoring site 050230001I01	7-13
7-2 Computer plot of 1977 hourly average nitrogen dioxide data (ppm) reported by monitoring site 050230001I01 Data plot for fixed-site monitor after final simulation of missing values	7-14
7-3 Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 1H1EX formulation (see Table 7-4)	7-35
7-4 Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 8H1EX formulation (see Table 7-4)	7-36
7-5 Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 8H5EX formulation (see Table 7-4)	7-37
7-6 Location of eight PM monitoring stations providing data for development of interpolation method. Meteorological data provided by NE Airport station	7-50
11-1 Map of Houston study area showing videotaping locations	11-9

TABLES

<u>Table</u>	<u>Page</u>
1-1 Abbreviations and acronyms appearing in this report	1-8
2-1 Descriptive statistics for hourly average values in 1995 data sets selected to represent Denver exposure districts after estimation of missing values	2-3
2-2 Descriptive statistics for hourly average values in 1997 data sets selected to represent Los Angeles exposure districts after estimation of missing values	2-3
2-3 Demographic groups defined for the pNEM/CO analyses and number of associated cohorts by study area	2-5
2-4 Assignment of CHAD location codes to pNEM/CO microenvironments	2-8
2-5 Methodology used to estimate carbon monoxide concentrations in each microenvironment defined for the Denver and Los Angeles pNEM/CO analyses	2-13
2-6 Estimated values of parameters in Equation 2-2	2-17
2-7 Percentage of persons with ischemic heart disease (IHD) by demographic group	2-23
2-8 Estimates of population residing in each Denver exposure district	2-24
2-9 Estimates of population residing in each Los Angeles exposure district	2-24
3-1 Description of 1990 Summary Tape File 1A (STF 1A)	3-3
3-2 Description of 1990 Summary Tape File 3a (STF 3A)	3-4
3-3 Description of 2000 Summary Tape File 1 (STF 1)	3-5
3-4 Radial grid used for interpolation example presented in text	3-8
3-5 Information on the 2000 decennial census	3-12
4-1 Characteristics for defining demographic profiles in APEX3 as applied to CO	4-3
5-1 Brief descriptions of studies providing data to CHAD (McCurdy et al., 2000)	5-3

TABLES (continued)

<u>Table</u>	<u>Page</u>
5-2 Results of applying five statistical tests to grouped data on the number of minutes spent per day outdoors as the critical temperature used to partition the data is varied (Johnson, 1988)	5-6
5-3 Hypothetical event probabilities for transitions from Exposure State A between 11:-- a.m. and 11:59 a.m.	5-13
5-4 Descriptive statistics for residential occupancy periods (males and females combined by current age)	5-16
5-5 Sample simulation for a male aged 14 years which yields a residential occupancy period of 11 years	5-22
5-6 Descriptive statistics for residential occupancy period and age of death	5-23
5-7 Format for commuting database obtained from the Bureau of Census	5-26
5-8 Exposure districts for Denver pNEM/CO analysis	5-29
5-9 Number and fraction of Denver commuters associated with each combination of home and work district	5-30
5-10 Percentage of Denver commuters (WR) assigned by the Bureau of Census to each home-work combination based on supplemental data	5-32
5-11 Estimated values of coefficient “a” in Equation 5-7, based on calibration of school enrollment data from San Diego, CA	5-41
5-12 Distribution of U.S. enrolled students aged 14 by grade level, October 1999	5-41
5-13 Harris County, TX school-related characteristics	5-48
6-1 Microenvironments defined for HAPEM4 as used in NATA	6-4
6-2 Means, standard deviations, and confidence intervals for estimates of $k_d(A/V)$ provided by Weschler, Shields, and Nike (1992)	6-10
6-3 Distributions of air exchange rate values used in the pNEM/O3 mass balance model	6-13

TABLES (continued)

<u>Table</u>	<u>Page</u>
6-4 Probability of window status for day by air conditioning system and temperature range	6-17
6-5 Probability of windows being open by clock hour, temperature range, and window status of preceding hour (PH) for residences with central air conditioning	6-17
6-6 Probability of windows being open by clock hour, temperature range, and window status of preceding hour (PH) for residences with window air conditioning units	6-18
6-7 Probability of windows being open by clock hour, temperature range, and window status of preceding hour (PH) for residences with no air conditioning system	6-18
6-8 Selection frequencies for probabilistic parameters used in mass-balance model	6-21
6-9 Algorithm used to model the trip-related exposures associated with each cohort for the year-long exposure period	6-24
6-10 Algorithm for estimating enclosed volumes of cars and trucks	6-25
6-11 Algorithms for determining trip-related values of parameters used in the mass-balance model applied to cars and trucks	6-26
6-12 Vehicle speed distribution	6-28
6-13 Distributions of parameter values used in the application of the pNEM/CO mass-balance model to Denver	6-30
6-14 Distributions of parameter values used in the application of the pNEM/CO mass-balance model to Los Angeles	6-31
6-15 Distributions for air exchange rate (v) for enclosed, nonresidential microenvironments	6-32
6-16 Percentage of person-days with indicated window ratio by air conditioning system and temperature range	6-35

TABLES (continued)

<u>Table</u>	<u>Page</u>
6-17 Air exchange rate measured by Rodes et al. (1998) under varying conditions	6-40
6-18 Statistics on square footage of occupied units in DE nver and Los Angeles (Bureau of the Census, 1995)	6-41
6-19 1998 automobile sales by segment and enclosed volume by vehicle type . .	6-43
6-20 Allocation of market segments to vehicle types	6-44
6-21 Interior volumes for selected pickup trucks classified by curb weight	6-45
6-22 List of HAPs for development of ME factors	6-47
6-23 Microenvironments assigned PROX factors greater than 1.0 for acetaldehyde, acrolein, 1,3-butadiene, and formaldehyde for the “onroad” mobile-source category only	6-49
6-24 Description of ME groups	6-52
6-25 HAP groups for use in HAPEM4	6-55
7-1 Techniques for validating data	7-4
7-2 Descriptive statistics for one-hour carbon monoxide concentrations reported by Denver monitors before and after estimation of missing values	7-15
7-3 Descriptive statistics for one-hour carbon monoxide concentrations reported by Los Angeles monitors before and after estimation of missing values	7-17
7-4 Formulations of the National Ambient Air Quality Standard (NAAQS) for ozone used in evaluating three procedures for adjusting air quality data to simulate attainment of an ozone NAAQS	7-24
7-5 Design values for baseline and “attainment” conditions	7-25
7-6 Quadratic air quality adjustment procedure for 1H1EX formulation	7-26
7-7 Quadratic air quality adjustment procedure for 8H1EX formulation	7-27
7-8 Quadratic air quality adjustment procedure for 8H5EX formulation	7-28

TABLES (continued)

<u>Table</u>	<u>Page</u>
7-9 Weibull air quality adjustment procedure for 1H1EX formulation	7-30
7-10 Weibull air quality adjustment procedure for 8H1EX formulation	7-31
7-11 Weibull air quality adjustment procedure for 8H5EX formulation	7-32
7-12 Selected descriptive statistics for Los Angeles monitoring sites based 1997 carbon monoxide data	7-40
7-13 Parameter values specified for application of proportional rollback procedure to carbon monoxide data reported by fixed-site monitors in Los Angeles . . .	7-40
7-14 Descriptive statistics for carbon monoxide concentrations reported by 10 Los Angeles monitors before and after adjustment to simulate attainment of eight-hour CO NAAQS	7-42
8-1 Statistics on gas stove use obtained from a survey by Koontz, Mehegan, and Nagda (1992)	8-1
8-2 Probability of gas stove use by clock hour and assumed burner operation period	8-3
8-3 Proportion of PEM values in gas stove residences with stove in operation by clock hour and work status	8-4
8-4 Gas stove data provided by Demand Analysis Office of the California Energy Commission	8-8
8-5 Fuel use statistics from the American Housing Survey (1996)	8-10
8-6 Algorithm for estimating CO emission rate from passive smoking in the residential microenvironment	8-14
8-7 Algorithm for estimating CO emission rate from passive smoking in restaurants and bars (Method B in text)	8-17
9-1 Regression-based estimates for parameters of Equation 9-6 obtained from statistical analysis of data provided by Adams (1998)	9-3
9-2 Estimates for parameters of Equation 9-6 applicable to males aged 0 to 17	9-4

TABLES (continued)

<u>Table</u>	<u>Page</u>
9-3 Estimates for parameters of Equation 9-6 applicable to females aged 0 to 17	9-5
9-4 Values of the upper limit of $PCTVO_{2max}$ for specified averaging times	9-8
9-5 Probabilistic method for generating estimates of ventilation rates as a function of energy expenditure rate	9-10
9-6 Probabilistic algorithms used to determine physiological parameters associated with person k	9-11
9-7 Principal parameters used in probabilistic algorithm for estimating ventilation rates (Tables 9-5 and 9-6)	9-12
9-8 Parameters for lognormal distributions fitted to body mass (BM) data in kilograms organized by age and gender	9-13
9-9 Descriptive statistics for VO_2 and VER measured at maximal exertion by various researchers	9-14
9-10 Estimates of the energy conversion factor (ECF) based on data in ESmal, Bhambhani, and Brintnell (1995)	9-16
9-11 Regression equations for predicting basal metabolic rate provided by Schofield (1985) as compiled by McCurdy (1998)	9-18
9-12 Probabilistic method for generating estimates of ventilation rates as a function of energy expenditure rate	9-19
9-13 Parameters appearing in algorithms used to convert activity codes to estimates of ventilation rate	9-20
10-1 Definitions of CFK model variables	10-3
10-2 Parameters included in physiological profile for adults in Version 2.1 of pNEM100	10-3
10-3 Literature data used to derive geometric mean and standard deviation for lognormal distribution of endogenous CO production rate	10-17
11-1 Assumed parameter values used in Steadman Apparent Temperature Model	11-3

TABLES (continued)

<u>Table</u>	<u>Page</u>
11-2 Derived quantities calculated in Steadman Apparent Temperature Model . . .	11-4
11-3 Primary heat transfer variables in Steadman Apparent Temperature Model .	11-5
11-4 Heat index table from National Weather Service showing HI in degrees F . .	11-6
11-5 2001/2002 WCT Index table in English units (National Weather Service 2002)	11-10
11-6 2001/2002 WCT Index table in metric units (Nelson et al., 2002)	11-10
11-7 Characteristics of videotaping sessions	11-15
11-8 Frequency distributions for selected variables relating to monitoring conditions	11-16
11-9 Frequency distributions for selected variables relating to vehicle characteristics	11-18
11-10 Frequency distributions for specific combinations of vehicle type and window configuration	11-20
11-11 Frequency distributions for specific combinations of heat index and vehicle speed	11-21
11-12 Frequency distributions for specific combinations of location, date, and time of day	11-22
11-13 Variables created for stepwise linear regression analyses	11-23
11-14 Results of stepwise linear regression analyses by vehicle type when dependent variable = OPEN using candidate predictor variables in Table 11-11 as appropriate	11-24
11-15 Results of stepwise linear regression analyses by location when dependent variable = OPEN using candidate predictor variables in Table 11-11 as appropriate	11-25
11-16 Three-minute average ozone concentrations measured at specified heights above ground by the Cincinnati Tower Study	11-31
11-17 Median values of DIFF and RATIO by height above ground	11-32

TABLES (continued)

<u>Table</u>	<u>Page</u>
11-18 Results of regression analyses in which the dependent variable was the ozone concentration at 2 meters and the predictor variable was ozone concentration at indicated height	11-33
11-19 Characteristics of Chicago ozone monitoring sites and 1991 maximum ozone concentrations	11-36
11-20 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 11 on August 10, 1995 in Cincinnati, Ohio	11-37
11-21 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 12 on August 10, 1995 in Cincinnati, Ohio	11-37
11-22 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 21 on August 11, 1995 in Cincinnati, Ohio	11-38
11-23 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 2 on August 11, 1995 in Cincinnati, Ohio	11-38
11-24 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 31 on August 12, 1995 in Cincinnati, Ohio	11-39
11-25 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 41 on August 14, 1995 in Cincinnati, Ohio	11-39
11-26 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 42 on August 14, 1995 in Cincinnati, Ohio	11-40
11-27 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 51 on August 17, 1995 in Cincinnati, Ohio	11-40
11-28 Mean ozone and nitric oxide (NO) concentrations measured during Test No. 52 on August 17, 1995 in Cincinnati, Ohio	11-41
11-29 Ratios of ozone concentrations measured at 2 and 10 meters by test — summary of values listed in Tables 11-18 through 11-26	11-42
11-30 Characteristics of Chicago ozone monitoring sites used in pNEM/O3 and proposed values of RATIO(h) adjustment factor	11-45
11-31 Characteristics of Los Angeles ozone monitoring sites used in pNEM/O3 and proposed values of RATIO(h) adjustment factor	11-47

ACKNOWLEDGMENTS

This report was prepared and edited by Ted Johnson of TRJ Environmental, Inc. He was assisted by Tom Long. The work was funded by Purchase Order No. P0047620 of EPA Grant No. CR827033 awarded to the Environmental and Occupational Health Sciences Institute (EOHSI). Dr. Panos Georgopoulos of EOHSI supervised the project.

Hundreds of individuals have contributed to the development of the exposure models, algorithms, distributions, and databases described in this report. Although it is not possible to list all of these contributors here, we would like to acknowledge the special contributions of the following people.

William C. Adams	William F. Hunt	Will Ollison
Gerry Akland	Ted Johnson	Ted Palma
Kim Balassiano	Tad Kleindienst	Roy Paul
William F. Biller	Neil Klepeis	Anne Pope
David Burmaster	Jackie LaPointe	Michael Ragan
Dianne Byrne	Jim Laurenson	Harvey Richmond
Jim Capel	Mark Lee	Mike Riley
Jonathan Cohen	William Linn	Arlene Rosenbaum
Katherine Craig	Tom Long	Luther Smith
Tom Curran	Michael McCoy	Pat Stiefer
Marcus Duff	Tom McCurdy	Lance Wallace
George Duggan	Dave McKee	Mark Weaver
Tom Feagans	Greg Meiners	Charles J Weschler
Alicia Ferdo	Robert Mersch	Louis Wijnberg
Kris Fletcher	Gary Mihlan	Tony Wisbith
Graham Glen	Jill (Warnasch) Mozier	
Stan Hayes	William Nelson	
Michael Huang	Eduardo Olaguer	

SECTION 1

INTRODUCTION

1.1 Purpose of This Guide

Within the U.S. Environmental Protection Agency (EPA), the Office of Air Quality Planning and Standards (OAQPS) has the responsibility of developing air quality standards that are protective of human health and welfare. These standards include the National Ambient Air Quality Standards (NAAQS), National Emission Standards for Hazardous Air Pollutants (NESHAPS), and other standards regulating hazardous pollutants. In evaluating potential standards, OAQPS frequently uses one or more computer-based models to estimate the number of people who will be exposed to the air pollution levels that are expected to occur under various air quality scenarios. These models have been under continuous development since 1979 and have been described by a series of contractor reports prepared under the direction of OAQPS staff.

Each exposure model consists of a set of components that may be conveniently classified as (1) input databases, (2) algorithms that operate on these databases, and (3) values for various parameters within the algorithms. Although each exposure model was developed to estimate exposures with respect to a specific pollutant and exposed population, many of the components used in a particular existing model are likely to be useful in future models and models currently under development. To facilitate the use of these components by other researchers, the author has surveyed the components of selected models developed by OAQPS (pNEM, APEX, and HAPEM) and identified the components that are most likely to be of use to other modelers. This report provides a detailed description of each of these components together with an explanation of its role within the source exposure model. It is hoped that model builders will make use of this information whenever appropriate and thus avoid "reinventing the wheel."

1.2 Exposure Models Developed by OAQPS

The exposure models included in this report fall into two general categories: NEM-type models and HAPEM-type models. Both model types treat human activity patterns as sequences of exposure events in which each event is defined by a geographic location and a microenvironment. NEM-type models provide an exposure estimate (defined by concentration and breathing rate) for each individual exposure event and then average these estimates over periods ranging from one hour to one year. HAPEM-type models typically determine annual average exposure estimates (concentration only) based on the quantity of time spent per year in each combination of geographic location and microenvironment. The NEM-type models tend to incorporate estimation procedures (algorithms) that are more sophisticated than those included in the HAPEM models. Recent probabilistic versions of NEM (labeled pNEM) attempt to simulate the variability in the principal factors affecting exposure; earlier versions treated most exposure factors as constants. The following two subsections provide brief histories of the NEM and HAPEM model types.

1.2.1 NEM-Type Models

In evaluating alternative NAAQS proposed for a particular pollutant, OAQPS assesses the risks to human health of air quality meeting each of the standards under consideration (Richmond and McCurdy, 1988). This assessment of risk requires estimates of the number of persons exposed at various pollutant concentrations for specified periods of time. The estimates may be specific to an urbanized area such as Los Angeles or apply to the entire nation.

Several researchers (Ott, 1982; Duan, 1982) have recommended that such estimates be obtained by simulating the movements of people through zones of varying air quality so as to approximate the actual exposure patterns of people living within a defined area. OAQPS has implemented this approach through an evolving methodology referred to as the NAAQS Exposure Model (NEM). An early overview of the NEM methodology is provided by a paper by Biller et al. (1981). From 1979 to 1988, IT Air Quality Services (ITAQS) assisted OAQPS in developing and applying pollutant-specific versions of NEM to nitrogen dioxide (Johnson and Paul, 1981a; Paul and Johnson, 1984), particulate matter (Johnson and Paul, 1981b), CO (Johnson and Paul, 1983), and ozone (Paul and McCurdy, 1986). These versions of NEM are referred to as “deterministic” versions in that no attempt was made to model random processes within the exposure simulation. To better address the observed variability in the factors affecting exposure, researchers incorporated probabilistic components in all versions of NEM created after 1988. These versions are jointly referred to as a “probabilistic” NEM (pNEM) models. Subsections 1.2.1.1 and 1.2.1.2 provide brief descriptions of the deterministic and probabilistic versions of NEM.

The various versions of NEM and pNEM are similar in that each estimates the exposures of homogeneous population groups (called cohorts) rather than the exposures of individuals. Demographic data are used to extrapolate the cohort exposure estimates to the population-of-interest within each study area. This approach was largely dictated by the computational limitations of the main-frame computers used to run the NEM and pNEM models. With the advent of fast personal computers, EPA began development of the Air Pollution Exposure Model (APEX), an enhanced version of pNEM that uses Monte Carlo methods to estimate the exposures of individuals whose characteristics are realistically sampled from the population of each study area. Subsection 1.2.1.3 provides a description of the APEX model currently under development.

1.2.1.1 The Deterministic Versions of NEM

The deterministic versions of NEM were similar in that each was capable of simulating the movements of selected segments of an urban population through a set of environmental settings. Each environmental setting was defined by a geographic area and a microenvironment. Analysts employed either an “exposure district” or “neighborhood-type” approach to determine the borders of each geographic area, consistent with the ambient characteristics of the pollutant. Ambient (outdoor) pollutant levels in each geographic area were estimated from either fixed-site monitoring data or dispersion model estimates. To better utilize fixed-site monitoring data, researchers developed special time series techniques to fill in missing values and special “roll-back”

techniques to adjust the monitoring data to simulate conditions under attainment of a particular NAAQS.

The population-of-interest in each study area was divided into an exhaustive set of cohorts, and an activity pattern was developed for each cohort. The activity pattern assigned the cohort to a geographic location and a microenvironment for each time interval of a defined exposure period. In some models, the activity pattern also assigned an exercise level or respiration rate to each time interval. In early NEM analyses, the time interval was 1 hour; in later NEM analyses the time interval was reduced to 10 minutes. Exposure periods varied from three months to one year. The activity patterns were based on reviews of the time use surveys. Researchers estimated the number of persons represented by each cohort by accessing census and commuting data at the census-tract level.

The pollutant concentration in a particular microenvironment was estimated by a linear model that included terms relating to the ambient pollutant level and to emission sources within the microenvironment. Researchers developed both point estimates and distributions for the parameter values of these linear models by conducting comprehensive reviews of the scientific literature associated with each pollutant.

The earliest versions of NEM were developed for nitrogen dioxide (Johnson and Paul, 1981a; Paul and Johnson, 1984) and particulate matter (Johnson and Paul, 1981b). The activity patterns used in these models assigned each cohort to a geographic location and a microenvironment for each time interval but did not provide an estimate of associated energy expenditure or respiration rate. Consequently, the exposure estimates obtained from these models were limited to pollutant concentrations. The activity patterns used in the versions of NEM applicable to CO (NEM-CO) and ozone (NEM-O₃) were enhanced to provide a qualitative estimate of exercise level (low, medium, or high) for each activity. In NEM-CO, each exercise level was subsequently converted to a point estimate of ventilation rate for use in an algorithm that estimated the blood level of carboxyhemoglobin (COHb) at the end of each hour of exposure (Johnson and Paul, 1983). NEM-CO was the only deterministic NEM-series model to characterize exposures by both pollutant concentration and ventilation rate. Although NEM-O₃ specified an exercise level for each activity, these levels were not converted into estimates of ventilation rate for use in other parts of the model (Paul, Johnson, Pope, et al., 1986).

Additional details concerning the evolution of the deterministic version of NEM are provided by Paul, Johnson, and McCurdy (1988). Critiques of deterministic NEM are included in surveys of exposure models by Pandian (1987) and Ryan (1991). Two staff papers (McKee, Richmond, Johnson, and McCurdy, 1984; McKee, Johnson, McCurdy, and Richmond, 1989) prepared by EPA discuss the use of NEM in evaluating proposed NAAQS for CO and ozone.

1.2.1.2 The Probabilistic Versions of NEM

In 1988, OAQPS began to incorporate probabilistic elements into the NEM methodology and to apply the resulting model (pNEM) to the criteria pollutants. The initial result of this work was an early version of pNEM applicable to ozone (pNEM/O₃). In this model, the activity pattern for each cohort consisted of a sequence of exposure events with durations varying from one to 60 minutes. Each exposure event assigned

the cohort to a particular exposure district, microenvironment, and breathing rate. The activity patterns were based on real-time activity data collected during the Cincinnati Activity Diary Study (Johnson, 1986). Four breathing rate categories were defined according to codes appearing in the activity diary data base: slow - sleeping, slow - awake, medium, and fast. A probabilistic algorithm converted the breathing rate category assigned to each exposure event to an estimate of "equivalent ventilation rate" (EVR), defined as ventilation rate divided by body surface area.

This early version of pNEM/O₃ also used a probabilistic, regression-based relationship to estimate indoor ozone concentrations from outdoor concentrations. In addition, a probabilistic algorithm was used to simulate the opening and closing of residential windows as a function of air conditioning system and temperature. A report by Johnson, Paul, Capel, and McCurdy (1990) describes this model and its application to Houston, Texas. A later report by Johnson, Capel, Olaguer, and Wijnberg (1992) describes the application of this model to a 20-state region in the Northeastern United States. Researchers also created a special version of the model applicable to children attending summer camp (pNEM/CAMP). A technical letter by Johnson (1994) describes the application of this model to a summer camp in California.

A probabilistic version of NEM applicable to carbon monoxide (pNEM/CO) was developed in 1991. This model marked the first time in the evolution of NEM that a mass balance model was used to estimate indoor pollutant concentrations. The model also included a probabilistic algorithm that estimated an EVR value for each exposure event. These EVR values were converted to corresponding ventilation rates for use in a probabilistic COHb algorithm. The application of pNEM/CO to Denver, Colorado, has been described in a report by Johnson, Capel, Paul, and Wijnberg (1992). This report also includes the results of initial efforts to validate the model using personal monitoring data collected in Denver by Johnson (1984). A report by Johnson, Capel, McCoy, and Warnasch (1994b) describes the application of pNEM/CO to Toronto, Ontario. In a related study, researchers used a modified version of the model to estimate benzene and 1,3-butadiene exposures in the Toronto study area (Johnson, Capel, McCoy, and Warnasch, 1994a).

A new version of pNEM/O₃ was developed in early 1992. Unlike the earlier version of pNEM/O₃, the new model used a mass balance model to estimate indoor ozone concentrations. Consistent with 1991 version of pNEM/CO, the new version of pNEM/O₃ included a probabilistic algorithm for estimating EVR. A report by Johnson, Capel, Olaguer, and Wijnberg (1993) describes this version of pNEM/O₃ and summarizes the results of an initial application of the model to 10 cities. Subsequent to this report, ITAQS made a number of significant enhancements to pNEM/O₃ and its input databases, including revisions to the methods used to estimate equivalent ventilation rates, to determine commuting patterns, and to adjust ambient ozone levels to simulate attainment of proposed NAAQS. A report by Johnson, Capel, and McCoy (1993) describes the enhanced model and summarizes the results of applying the model to nine of the ten cities included in the previous pNEM/O₃ exposure assessment. (Tacoma, Washington, was excluded from the analysis because of insufficient monitoring data.) The report also includes the results of initial efforts to validate pNEM/O₃ using personal monitoring data collected in Houston, Texas. A report by Johnson, Capel, McCoy, and Warnasch (1994c) describes the application of pNEM/O₃ to three Canadian cities.

In early 1994, EPA directed ITAQS to develop a special version of pNEM/O₃ applicable to outdoor workers and to use it to estimate the ozone exposures of outdoor workers residing in each of the nine areas. The activity pattern database employed by the model was expanded to include data from seven activity diary studies. A summary of this work can be found in a report by Johnson, Capel, McCoy, and Mozier (1996).

In a follow-up work effort for EPA, ITAQS developed a second special version of pNEM/O₃ applicable to children who tend to be active outdoors (referred to as “outdoor children”). This version of pNEM/O₃ estimated EVR using a sophisticated Monte Carlo algorithm based on ventilation rate data obtained from students in Los Angeles. A report by Johnson, Capel, Mozier, and McCoy (1996) summarizes the results of applying the model to outdoor children residing in the nine study areas. No subsequent versions of pNEM/O₃ have been developed by OAQPS.

During 1999 and 2000, researchers made a series of refinements to pNEM/CO that culminated in Version 2.1. This version contained enhanced algorithms for simulating gas stove use, estimating alveolar ventilation rate (a measure of human respiration), and modeling home-to-work commuting patterns. A report by Johnson, Mihlan, LaPointe, et al. (2000) describes Version 2.1 and presents results of applying the model to Denver and Los Angeles.

1.2.1.3 APEX

EPA is currently developing an advanced version of pNEM referred to as the Air Pollution Exposure Model (APEX). APEX differs from earlier pNEM models in that the probabilistic features of the model are incorporated into an explicit Monte Carlo framework. Instead of dividing the population-of-interest into a set of cohorts, APEX generates individuals as if they were being randomly sampled from the population. APEX provides each generated individual with a demographic profile that specifies values for all parameters required by the model. The values are selected from distributions and databases that are specific to age, gender, and other specifications stated in the demographic profile. Although the model is applicable to a wide range of pollutants, the current version (APEX3) is being used primarily to estimate carbon monoxide exposures and associated COHb levels (Glen, 2002). EPA plans to develop future versions of APEX that are optimized for application to ozone and other criteria pollutants.

1.2.2 HAPEM-Type Models

The initial version of HAPEM was developed for Emissions Standards Division of EPA by Johnson, Capel, Paul, and Olaguer (1991). HAPEM was created specifically to provide more accurate exposure estimates than EPA’s existing Human Exposure Model (HEM). In its original form, HAPEM was intended to be applied to factories, refineries, and other point emission sources. For convenience, this point-source version of HAPEM is referred to as HAPEM-PS. An article by Johnson, Capel, and Byrne (1991) describes the commuting model incorporated into HAPEM-PS.

Following the development of HAPEM-PS, EPA’s Office of Mobile Sources funded the development of HAPEM-MS1, a special version of HAPEM applicable to toxic pollutants emitted by mobile sources (Johnson, Paul, and Capel, 1992). The

Office of Research and Development (ORD) subsequently funded the development of an enhanced version of HAPEM-MS designated HAPEM-MS2 (Johnson, McCoy, and Capel, 1993; Johnson, Warnasch, McCoy, et al., 1996). This version of HAPEM provided a larger number of microenvironments (37) and greater time resolution than either HAPEM-PS or HAPEM-MS1. Guides to the program code and run procedures of HAPEM-MS2 were prepared by Glen (1994a, 1994b).

In 1996, ORD funded a number of enhancements to the model, including (1) the capability of defining custom demographic groups, (2) the use of 1990 census data, and (3) an improved modeling algorithm for estimating ambient impacts in residences with attached garages. A report by Palma, Riley, and Capel (1996) describes this model, designated HAPEM-MS3, and provides results of applying the model to San Francisco.

The principal difference between HAPEM-PS and the various versions of HAPEM-MS concerns the structure of the input air quality data bases required by each model. HAPEM-PS requires an air quality indicator (e.g., annual mean concentration) for each point in a receptor grid surrounding the point source under evaluation. Receptor air quality values are typically determined through the use of emissions data and a dispersion model. Although the HAPEM-MS series of models can theoretically use air quality data organized by receptor points, the models have historically been applied to air quality data obtained from a relatively small number of fixed-site monitors.

Further enhancements were made to HAPEM-MS3 that culminated in the development of HAPEM4, the exposure model currently being used in the National-Scale Air Toxics Assessment (USEPA, 2001). This assessment is a nationwide study of potential inhalation exposures and health risks associated with 32 hazardous air pollutants and diesel particulate matter based on 1996 emission estimates. It is one component of the National Air Toxics Assessment (NATA), the technical support component of EPA's National Air Toxics Program.

All versions of HAPEM provide exposure estimates for user-defined cohorts. In the most recent version (HAPEM4), each cohort is defined by age, gender, race, residential location, work location (adults), and school location (children). Activity diary data obtained from CHAD are analyzed to determine the fraction of time members of each cohort are likely to spend in various combinations of geographic location and microenvironments by time of day, day type, and season. HAPEM4 employs a linear relationship that accounts for ambient (outdoor) concentration, proximity to sources, and penetration efficiency to determine exposures within each microenvironment. The microenvironmental exposures are weighted by time spent in each microenvironment to determine an annual average exposure for each cohort. Census and commuting data are used to extrapolate the cohort exposure estimates to the study area population.

As part of a NATA exposure analysis conducted during 2001, HAPEM4 was applied to the residents of all census tracts within the contiguous United States, Puerto Rico, and the Virgin Islands. Researchers used the 1996 National Toxics Inventory (NTI) and the ASPEN (Assessment System for Population Exposure Nationwide) dispersion model to develop estimates of ambient pollutant concentrations for each pollutant by census tract. Microenvironmental concentrations specific to each census tract were derived from the ambient concentration estimate for the census tract (obtained from ASPEN) and a set of three factors: PEN, PROX, and ADD. The penetration factor (PEN) was an estimate of the ratio of microenvironment concentration to the concurrent outdoor concentration in the immediate vicinity of the

microenvironment. These pollutant-specific estimates were derived from concentration values reported by various measurement studies. The proximity factor (PROX) was an estimate of the ratio of the outdoor pollutant concentration in the immediate vicinity of the microenvironment to the outdoor concentration represented by the air concentration data. ADD was an additive factor that accounted for emission sources within or near a particular microenvironment, i.e., indoor emission sources. A report by Rosenbaum (2002) provides a description of the version of HAPEM4 employed in the 2001 NATA exposure analysis.

HAPEM4 was designed to predict inhalation exposures for population groups, not individuals within these groups. However, HAPEM4 contains a stochastic feature that attempts to capture some of the variability in activity patterns within these groups. The stochastic feature predicts the annual exposure concentrations for a randomly selected set of 30 activity patterns representing each demographic group.

All versions of HAPEM have accounted for home-to-work commuting. HAPEM4 is the first model in the series to also account for home-to-school commuting. The model estimates the probability that students residing in a particular census tract will attend a particular school as a function of student age, school type, school capacity, and home-to-school distance. A report by Laurenson, et al. (2001) describes the development of the school commuting algorithm and its implementation within HAPEM4. This approach may be incorporated into future versions of pNEM and APEX.

1.3 Organization of This Report

This report is intended to provide an inventory of the most useful components (input databases, algorithms, distributions, etc.) to be found in the NEM-type and HAPEM-type models developed by OAQPS. Section 2 demonstrates how the various components work together within the structure of a typical exposure model through an overview of the most recently-developed NEM-type model (Version 2.1 of pNEM/CO). Sections 3 through 10 provide detailed descriptions of the components organized by function as follows.

- Section 3: Spatial and Temporal Coverage of Exposure Analysis
- Section 4: Representing Demographic Variability
- Section 5: Exposure Event Sequence
- Section 6: Estimation of Microenvironmental Concentrations
- Section 7: Processing Ambient Air Quality Data
- Section 8: Simulation of Indoor Emissions
- Section 9: Estimation of Human Ventilation Rate
- Section 10: Estimation of Carboxyhemoglobin Levels.

Section 11 provides a summary of recent research that may be useful in developing components for future model applications. A list of the abbreviations and acronyms appearing in this report can be found in Table 1-1.

Table 1-1. Abbreviations and Acronyms Appearing in This Report.

Abbreviation or Acronym	Section(s) containing definition or first appearance	Explanation
AC	6.4.2.1	Air conditioning
ADD	1.2.2	Additive factor used in HAPEM4
AER	6.3.2	Air exchange rate
AHS	8.1.3	American Housing Survey
AM	6.4.2.6	Arithmetic mean
APEX	1.2.1	Air Pollution Exposure Model
APEX3	1.2.1.3	Current version of APEX
API	6.4.3.2	American Petroleum Institute
AQAP	7.4	Air quality adjustment procedure
AQI	7.4	Air quality indicator
AR(2)	7.3.1	Autoregressive process of order 2
ASD	6.4.2.6	Arithmetic standard deviation
ASHRAE	8.2.3	American Society of Heating, Refrigeration, and Air Conditioning Engineers
ASPEN	6.6.1	Assessment System for Population Exposure Nationwide
AUB	6.4.2.1	Burner annual fuel use
AUP	6.4.2.1	Pilot light annual fuel use
BG	7.5	Background concentration
BEADS	6.6.5	Benzene Exposure and Absorbed Dose Simulation
BM	9.1	Body mass
BMR	9.4	Basal metabolic rate
BNL	6.4.3.2	Brookhaven National Laboratory
BOC	--	<u>B</u> ureau of <u>C</u> ensus
BOP	6.4.2.1, 8.1.1	Burner operation period

Abbreviation or Acronym	Section(s) containing definition or first appearance	Explanation
CADS	5.1.3	Cincinnati Activity Diary Study
CARB		California Air Resources Board
CEC	6.4.3.2	California Energy Commission
CFK	10.1	Coburn Forster Kane
CHAD	2.3, 5.1.1	Consolidated Human Activity Database
CLV	7.4.1	Characteristic largest value
CO	1.0	Carbon monoxide
COHb	2.43	Carboxyhemoglobin
ECC	7.3.1	Essential cyclical component
EAD	7.4.6	Estimated attainment data
ECF	2.4.2	Energy conversion factor
EE	2.4.2	Energy expenditure rate
EES	2.3	Exposure event sequence
EFBURN	6.4.2.1	Burner emission factor
EFPILOT	6.4.2.1	Pilot light emission factor
EPA	1.1	U.S. <u>E</u> nvironmental <u>P</u> rotection <u>A</u> gency
EPOC	9.6	Excess post-energy oxygen consumption
ETS	2.4.1	Environmental tobacco smoke
EVR	1.2.1.2	<u>E</u> quivalent <u>V</u> entilation <u>R</u> ate
GM	2.4.1, 6.4.2.6	Geometric mean
GSD	2.4.1, 6.4.2.6	Geometric standard deviation
HAP	6.6.1	Hazardous air pollutant
HAPEM	1.1	Hazardous Air Pollutant Exposure Model
HAPEM4	1.2.2	Version of HAPEM used in NATA assessment
HAPEM-MS1	1.2.2	First version of HAPEM applicable to mobile sources

Abbreviation or Acronym	Section(s) containing definition or first appearance	Explanation
HAPEM-MS2	1.2.2	Second version of HAPEM applicable to mobile sources
HAPEM-PS	1.2.2	Version of HAPEM applicable to point sources
HI	11.1.1	Heat index
HST	5.3.2.1	Home-to-school trips
HU	8.1.4	Housing unit
ICF	6.6.1	ICF Consulting
IHD	2.5.2	Ischemic heart disease
I & M	11.2.2.1	Inspection and Maintenance
I/O		Indoor/outdoor
IT		International Technology
ITAQS		IT Air Quality Services
JAG/TI	11.1.2.1	Joint Action Group for Temperature Indicators
ME		Microenvironment
MET	2.3	Metabolic equivalent
MTBE		Methy tertiary butyl ether
NAAQS	1.1	National Ambient Air Quality Standard(s)
NATA	1.2.2	National Air Toxics Assessment
NEM	1.2.1	NAAQS Exposure Model
NEM-CO	1.2.1.1	NEM applicable to carbon monoxide
NEM-O3	1.2.1.1	NEM applicable to ozone
NERL	5.1.1	National Exposure Research Laboratory
NHAPS	5.1.1	National Human Activity Pattern Survey
NIGAS	8.1.2.1	Northern Illinois Gas Company
NO	11.4.3.1	Nitric oxide
NSA	5.3.2.1	National Scale Assessment

Abbreviation or Acronym	Section(s) containing definition or first appearance	Explanation
NVO _{2max}	9.3	Maximum oxygen uptake rate per kg of body mass
NVR	6.4.3.2	Normalized ventilation rate
NMS	11.1.1	National Weather Service
O2Hb	10.1	Oxyhemoglobin
OAD	7.4.6	Observed attainment data
OAQPS	1.1	Office of Air Quality Planning and Standards
OCT	5.1.3	Optimum critical temperature
PAQM	6.3.3	
PCTVO _{2max}	9.3	Percentage of VO _{2max}
PD	5.2.3	Probability that person will die during next year
PEM	3.2.2	Personal exposure monitor
PEN	1.2.2	Penetration factor used in HAPEM4
pNEM	1.2.1	Probabilistic NEM
pNEM/CO	1.2.1.2	pNEM applicable to carbon monoxide
pNEM/O3	1.2.1.2	pNEM applicable to ozone
POP	5.2.3	Population
PROX	1.2.2	Proximity factor used in HAPEM4
PSR	5.2.3	Probability that person was in same residence one year earlier
RMR	2.4.2	Resting metabolic rate
RN	6.4.3.1	Random number
ROP	5.2	<u>R</u> esidential <u>O</u> ccupancy <u>P</u> eriod
ROPSIM	5.2	Program developed to simulate residential occupancy periods
SHEDS-PM	7.6.2	Exposure model developed by NERL applicable to particulate matter

Abbreviation or Acronym	Section(s) containing definition or first appearance	Explanation
SimExpo- HAP	4.2.2	Exposure model applicable to hazardous air pollutants
TRJ	--	TRJ Environmental, Inc.
UBC	7.6.2	University of British Columbia
USEPA	--	U.S. Environmental Protection Agency
UTCI	11.2.4	Universal Thermal Climate Index
V_A	2.4.1	Alveolar ventilation rate
V_E	9	Ventilation rate
VER	9.2	Ventilatory equivalence ratio
VER_{max}	9.6	Maximum ventilatory equivalence ratio
VO_2	2.4.1	Oxygen uptake rate
VO_{2max}	9.2	Maximum oxygen uptake rate
WCT	11.1.2	Wind Chill Temperature

1.4 References for Section 1

Biller, W. F., T. B. Feagans, T. R. Johnson, G. R. Duggan, R. A. Paul, T. McCurdy, and H. C. Thomas. 1981. "A General Model for Estimating Exposure Associated With Alternative NAAQS." Paper No. 81-18.4. Presented at the 74th Annual Meeting of the Air Pollution Control Association, Dallas, Texas, June.

Duan, N. 1982. "Models for Human Exposure to Air Pollution." Environment International, Vol. 8, pp. 305 - 309.

Glen, G. 1994a. Details of the HAPEM-MS2 Exposure Model. Report prepared by ManTech Environmental Technology, Inc., for the Atmospheric Research and Exposure Assessment Laboratory, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina, March.

Glen, G. 1994b. Running the HAPEM-MS2 Exposure Model. Report prepared by ManTech Environmental Technology, Inc., for the Atmospheric Research and Exposure Assessment Laboratory, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina, March.

Glen, G. 2002. Draft User's Guide for the APEX3 Model. Report prepared by ManTech Environmental Technology, Inc., for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. March 15.

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. 1986. A Study of Human Activity Patterns in Cincinnati, Ohio, prepared for Environmental Assessment Department of the Electric Power Research Institute under Contract RP940-06, June.

Johnson, T. 1994. "Enhancements to the pNEM Summer Camp Methodology." Letter to Harvey Richmond, Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. March 21 (revised).

Johnson, T. J. Capel, and D. Byrne. 1991. "The Estimation of Commuting Patterns in Applications of the Hazardous Air Pollutant Exposure Model (HAPEM)," Paper No. 91-172.6, 84th Annual Meeting of the Air and Waste Management Association, Vancouver, June.

Johnson, T., J. Capel, and M. McCoy. 1993. Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM and 1990 Population Data. Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., J. Capel, M. McCoy, and J. Mozier. 1996. Estimation of Ozone Exposures Experienced by Outdoor Workers in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Johnson, T., J. Capel, M. McCoy, and J. Warnasch. 1994a. Estimation of Benzene and 1,3-Butadiene Exposures Experienced by Residents of Toronto, Ontario Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for Health Canada, Ottawa, March.

Johnson, T., J. Capel, M. McCoy, and J. Warnasch. 1994b. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels Experienced by Residents of Toronto, Ontario Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for Health Canada, Ottawa, March.

Johnson, T., J. Capel, M. McCoy, and J. Warnasch. 1994c. Estimation of Ozone Exposures Experienced by Residents of Three Canadian Cities Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for Health Canada, Ottawa, July.

Johnson, T., J. Capel, J. Mozier, and M. McCoy. 1996. Estimation of Ozone Exposures Experienced by Outdoor Children in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Johnson, T., J. Capel, E. Olaguer, and L. Wijnberg. 1992. Estimation of Ozone Exposures Experienced by Residents of ROMNET Domain Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., J. Capel, E. Olaguer, and L. Wijnberg. 1993. Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, February.

Johnson, T., J. Capel, R. Paul, and E. Olaguer. 1991. A Comparison of the Hazardous Air Pollutant Exposure Model (HAPEM) and the Human Exposure Model (HEM). Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM. Report prepared by International Technology Air Quality Services under EPA Contract No. 68-D0-0062. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., M. McCoy, and J. Capel. 1993. Enhancements to the Hazardous Air Pollutant Exposure Model (HAPEM) as Applied to Mobile Source Pollutants. Report prepared by IT Air Quality Services for the Atmospheric Research and Exposure Assessment Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1), prepared for the Risk and Exposure Assessment Division of U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

Johnson, T. and R. Paul. 1981a. The NAAQS Exposure Model (NEM) and Its Application to Nitrogen Dioxide. Report prepared by PEDCo Environmental, Inc. for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. R. and R. A. Paul. 1981b. The NAAQS Exposure Model (NEM) Applied to

Particulate Matter. Report prepared by PEDCo Environmental, Inc. for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. R. and R.A. Paul. 1983. The NAAQS Exposure Model (NEM) Applied to Carbon Monoxide. EPA Report No. 450/5-83-003. Prepared by PEDCo Environmental, Inc. for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., R. Paul, and J. Capel. 1992. Application of the Hazardous Air Pollutant Exposure Model (HAPEM) to Mobile Source Pollutants. Report prepared by IT Air Quality Services for the Office of Mobile Sources, U. S. Environmental Protection Agency, Ann Arbor, Michigan, August.

Johnson, T. R., R. A. Paul, J. E. Capel, and T. McCurdy. 1990. "Estimation of Ozone Exposure in Houston Using a Probabilistic Version of NEM." Paper No. 90-150.1, presented at the 83rd Annual Meeting of the Air and Waste Management Association, Pittsburgh, Pennsylvania.

Johnson, T., J. Warnasch, M. McCoy, J. Capel, and M. Riley. 1996. Developmental Research for the Hazardous Air Pollutant Exposure Model (HAPEM) as Applied to Mobile Source Pollutants. Report prepared by IT Air Quality Services for the Human Exposure and Field Research Division, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina, February.

Laurenson, J., A. S. Rosenbaum, M. Huang, and J. P. Cohen. 2001. "Preliminary Results of Prototype School Commute Module for HAPEM4 (Contract 68-D6-0064, Work Assignment 4-11)." Memorandum to Ted Palma, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. September 30.

McKee, D., P. Johnson, T. McCurdy, and H. Richmond. 1989. Review of the National Ambient Air Quality Standard for Ozone: Assessment of Scientific and Technical Information. EPA Report No. EPA-450/2-92-001. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

McKee, D., H. Richmond, P. Johnson, and T. McCurdy. 1984. Review of the NAAQS for Carbon Monoxide: Reassessment of Scientific and Technical Information. EPA Report No. EPA-450/5-84-004. U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina.

Ott, W. R. 1982. "Concepts of Human Exposure to Air Pollution." Environment International, Vol. 7, pp. 179 - 196.

Palma, T., M. Riley, and J. Capel. 1996. Development and Evaluation of Enhancements to the Hazardous Air Pollutant Exposure Model (HAPEM-MS3). Report prepared by International Technology Corporation for the Office of Air Quality Planning

and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Pandian, M. D. 1987. Evaluation of Existing Total Human Exposure Models. EPA Report No. EPA-600/4-87-004. U. S. Environmental Protection Agency, Las Vegas, Nevada.

Paul, R. A., and T. Johnson. 1984. The NAAQS Exposure Model (NEM) Applied to Nitrogen Dioxide. Report prepared by PEDCo Environmental, Inc. for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Paul, R. A., T. Johnson, and T. McCurdy. 1988. "Advancements in Estimating Urban Population Exposure." Paper No. 88-127.1. Presented at the 81st Annual Meeting of the Air Pollution Control Association, Dallas, Texas, June.

Paul, R. A., T. Johnson, A. Pope, A. Ferdo, and W. F. Biller. 1986. National Estimates of Exposure to Ozone Under Alternative National Standards. Report prepared by PEI Associates, Inc. for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina

Paul, R. A. and T. McCurdy. 1986. "Estimation of Population Exposure to Ozone." Paper No. 86-66.2. Presented at the 79th Annual Meeting of the Air Pollution Control Association, Dallas, Texas, June.

Richmond, H. M. and T. McCurdy. 1988. "Use of Exposure Analysis and Risk Assessment in the Ozone NAAQS Review Process." Paper No. 88-121.3. Presented at the 81st Annual Meeting of APCA, Dallas, Texas.

Rosenbaum, A. 2002. The HAPEM4 User's Guide, Draft Version (Hazardous Air Pollutant Model - Version 4). Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. UPDATE WHEN REPORT IS FINALIZED.

Ryan, P. B. 1991. "An Overview of Human Exposure Modeling." Toxicology and Industrial Health, Vol. 1, No. 4, pp. 453-474.

USEPA. 2001. National-Scale Air Toxics Assessment for 1996. Report No. EPA-453/R-01-033. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. January.

SECTION 2

A MODELING EXAMPLE

This report is intended to provide an inventory of the most useful components (input databases, algorithms, distributions, etc.) to be found in the NEM-type and HAPEM-type models developed by OAQPS. Later sections of this report provide a detailed description of each of these components together with an explanation of its role within the source exposure model. To better understand how the various parts work together, it may be useful to consider the structure of the most recently-developed NEM-type model, Version 2.1 of pNEM/CO. This section provides an overview of the model organized according to the five steps of the NEM modeling methodology:

1. Define a study area, one or more populations-of-interest, appropriate subdivisions of the study area, and an exposure period.
2. Divide the population-of interest into an exhaustive set of cohorts.
3. Develop an exposure event sequence for each cohort for the exposure period.
4. Estimate the pollutant concentration, ventilation rate, and physiological indicator (if applicable) associated with each exposure event.
5. Extrapolate the cohort exposures to each population-of-interest.

The remainder of this section describes how Version 2.1 of pNEM/CO implements each step of the pNEM methodology. Pertinent information concerning the application of the methodology to the Denver and Los Angeles study areas is included as appropriate. Additional details can be found in the project report by Johnson, Mihlan, LaPointe, et al. (2000).

2.1 Define a Study Area, Populations-of-Interest, Appropriate Subdivisions of the Study Area, and an Exposure Period

The pNEM/CO methodology provides estimates of the distribution of CO exposures and associated COHb levels within a defined population (the population-of-interest) for a specified exposure period. The exposure period is usually a recent calendar year for which good data are available with respect to ambient CO levels. The population-of interest is typically defined as people with specific demographic characteristics (e.g., adults with ischemic heart disease) who live and work within a defined set of exposure districts. Each exposure district is a contiguous set of census units surrounding one or more fixed-site CO monitors selected as representative of the district.

Analysts selected seven fixed-site monitors as the basis for developing the Denver exposure districts. Subsection 3.1.1 of Johnson, Mihlan, LaPointe, et al. (2000)

describes in detail the process used to select these monitors and to define the district boundaries. Briefly, analysts identified seven monitors which (1) were located within 50 km of the center of Denver, (2) were located in areas of appropriate urban land use, and (3) reported sufficient air quality data for 1995 through 1997. Five of the seven sites were identical to sites used in an analysis of Denver in 1992; the remaining two sites were located in downtown Boulder (28th Street and Marine Street). The locations of five sites used in the 1992 analysis were considered appropriate for defining five separate exposure districts with 10 km radii. However, the Boulder sites were considered too close together to support separate exposure districts. Consequently, analysts defined six exposure districts -- one for each of the 1992 Denver sites and a "composite" Boulder site. For purposes of constructing the associated exposure district, the composite site was assigned a location midway between the two Boulder sites.

Analysts evaluated the quality and completeness of the data available for the seven monitors for the years 1995 through 1997. Based on this evaluation, EPA selected 1995 as the year for the pNEM/CO analysis of Denver. Each of the selected monitors provided an hourly average data set that was at least 96 percent complete for this year. Section 3.2 of Johnson, Mihlan, LaPointe, et al. (2000) describes the method used to estimate the missing values in each data set. Table 2-1 provides descriptive statistics for the data sets after missing values were estimated.

A similar approach was used in selecting fixed-site monitors for the application of pNEM/CO to Los Angeles. As discussed in Subsection 3.1.2 of Johnson, Mihlan, LaPointe, et al. (2000), analysts designated all sites within Los Angeles, Orange, Riverside, San Bernardino, and Ventura Counties which reported CO data between 1995 and 1997 as potential sites for the pNEM/CO analysis.

Of the 30 CO sites which met these criteria, 24 satisfied a further requirement that the site reported data that were at least 75 percent complete for each of the three years. Analysts omitted seven of these monitors which were located in outlying areas and reported relatively low CO levels (three-year averages for the second largest 8-hour maximum CO concentration less than 4.5 ppm). EPA evaluated the siting characteristics and locations of the remaining 17 monitors and selected the 10 monitors listed in Table 2-2. These monitors all reported relatively high CO levels, and they appeared to provide good coverage of the highly urbanized areas within greater Los Angeles.

EPA selected 1997, the most recent of the three years evaluated, as the year for the pNEM/CO analysis of Los Angeles. All 10 sites had adequate data completeness for 1997. Table 2-2 provides descriptive statistics for each data set after estimation of missing values.

After application of the fill-in procedure described in Subsection 3.2 of Johnson, Mihlan, LaPointe, et al. (2000), each district in the Denver and Los Angeles study areas was represented by a complete, year-long sequence of 1-hour outdoor concentrations. These sequences were assumed to represent existing ("as is") outdoor air quality conditions in each study area. To represent outdoor air quality that just meets the current 8-hour NAAQS for CO, the concentration values in each Los Angeles sequence were adjusted according to the procedure described in Subsection 7.4.3 of this report. The Denver data were not adjusted, as the "as is" air quality was judged to be roughly equivalent to conditions expected to occur when the current 8-hour NAAQS is attained.

Table 2-1. Descriptive Statistics for Hourly Average Values in 1995 Data Sets Selected to Represent Denver Exposure Districts After Estimation of Missing Values.

Location	Monitor ID	Descriptive statistics for hourly-average CO concentrations, ppm					
		50th	90th	95th	99th	99.5th	Maximum
Littleton	005-0002	0.3	0.7	1.0	1.8	2.2	3.6
Broadway	031-0002	1.2	2.7	3.4	6.1	7.7	24.5
Albion	031-0013	0.9	2.5	3.4	5.5	6.4	14.6
Julian	031-0014	0.7	2.3	3.2	5.3	6.5	10.4
Arvada	059-0002	0.6	2.0	2.7	4.8	5.8	11.9
Boulder 28 th St.	013-0010	0.8	2.1	2.8	4.8	5.5	10.6
Boulder Marine St.	013-1001	0.4	0.9	1.3	2.3	2.9	8.3
Composite Boulder monitor	---	0.7	1.5	2.0	3.3	3.8	9.5

Table 2-2. Descriptive Statistics for Hourly Average Values in 1997 Data Sets Selected to Represent Los Angeles Exposure Districts After Estimation of Missing Values.

Location	Monitor ID	Descriptive statistics for hourly-average CO concentrations, ppm					
		50th	90th	95th	99th	99.5th	Maximum
West Los Angeles	60370113	0.6	2.0	2.6	3.6	4.1	7.3
Burbank	60371002	1.4	3.5	4.4	6.0	6.6	8.8
Los Angeles	60371103	1.0	3.0	3.8	5.4	5.8	8.9
Lynwood	60371301	1.7	4.9	6.7	11.2	13.5	19.2
Pico Rivera	60371601	1.2	3.0	3.6	5.0	5.6	9.2
Pasadena	60372005	0.9	2.1	2.8	4.2	4.7	8.1
Long Beach	60374002	0.7	2.7	3.6	5.2	5.9	9.0
Hawthorne	60375001	0.5	3.7	5.1	7.3	8.2	12.4
Anaheim	60590001	0.8	2.3	2.9	4.6	5.5	8.4
La Habra	60595001	1.0	2.8	3.6	6.1	7.1	11.9

Two populations-of-interest were defined for the exposure assessment: (1) the

general population and (2) adults (18 and older) with ischemic heart disease who lived and worked within the exposure districts in each urban study area. [EPA's Criteria Document and OAQPS Staff Paper used several terms (i.e., coronary heart disease, cardiovascular disease, coronary heart disease, and ischemic heart disease) which all referred to the same population group.]

2.2 Divide Population-of-Interest Into Representative Units

In a pNEM analysis, the population-of-interest is divided into a set of cohorts such that each person is assigned to one and only one cohort. Each cohort is assumed to contain persons with identical exposures during the specified exposure period. Cohort exposure is typically assumed to be a function of demographic group, location of residence, and location of work place. Specifying the home and work district of each cohort provides a means of linking cohort exposure to ambient CO concentrations. Specifying the demographic group provides a means of linking cohort exposure to activity patterns which vary with age, work status, and other demographic variables. In some analyses, cohorts are further distinguished according to factors relating to proximity to emission sources or time spent in particular microenvironments.

In the application of pNEM/CO (Version 2.1) to Denver and Los Angeles, each cohort was identified as a distinct combination of (1) home district, (2) demographic group, (3) work district (if applicable), (4) residential cooking fuel, and (5) replicate number. The home district and work district of each cohort were identified according to the districts defined above. Table 2-3 lists 10 adult demographic groups defined for the pNEM/CO analyses. Four of the demographic groups were identified as workers. Each cohort associated with one of these groups was identified by both home and work district. The remaining cohorts were identified only by home district. Note that although children have been included within the demographic groups defined for previous pNEM/CO analyses, the exposure analyses summarized here were limited to adult demographic groups.

The residential cooking fuel of each cohort was identified as either "natural gas" or "other." This cohort index was used because a personal monitoring study (Johnson, 1984) conducted in Denver suggested that proximity to operating natural gas stoves contributed significantly to CO exposure. A review of the scientific literature concerning five other sources (kerosene space heaters, gas space heaters, wood stoves, fireplaces, and attached garages) indicated that (1) fireplaces and stoves did not contribute significantly to indoor CO levels, (2) kerosene and gas space heaters were used in less than 1 percent of the residences in each study area, and (3) attached garages could not be adequately characterized by available data (Fletcher and LaPointe, 1998). Section 3.5 of the Criteria Document (CD) for CO (USEPA, 2000) also provides a review of the available information on indoor and in-vehicle sources of CO and observed CO concentrations.

Earlier versions of pNEM/CO defined cohorts solely according to home district, demographic group, work district (if applicable), and residential cooking fuel. A new feature was installed in pNEM/CO (Version 2.1) which permitted the user to specify a "replication" value (n) such that the model will produce n cohorts for each combination of these four indices. Because pNEM/CO uses a Monte Carlo process to construct an

Table 2-3. Demographic Groups Defined for the pNEM/CO Analyses and Number of

Associated Cohorts by Study Area.

Demographic group	Includes commuting cohorts?	Number of cohorts associated with demographic group	
		Los Angeles	Denver
Males, 18 to 44, working	yes	600	360
Males, 18 to 44, nonworking	no	60	60
Males, 45 to 64, working	yes	600	360
Males, 45 to 64, nonworking	no	60	60
Males, 65+	no	60	60
Females, 18 to 44, working	yes	600	360
Females, 18 to 44, nonworking	no	60	60
Females, 45 to 64, working	yes	600	360
Females, 45 to 64, nonworking	no	60	60
Females, 65+	no	60	60
Total		2,760	1,800

activity pattern for each cohort, each of the n cohorts associated with a particular combination of home district, demographic group, work district, and residential cooking fuel is associated with a distinct exposure sequence. The replication feature permits the analyst to divide the population-of-interest into a larger number of smaller cohorts -- a process that decreases the “lumpiness” of the exposure simulation. Replication values of n = 5 and n = 3, respectively, were specified for the Denver and Los Angeles exposure analyses described in this report. Consequently, the pNEM/CO model analyzed n times the number of cohorts it would have considered for each city had the cohorts been defined solely by home district, demographic group, work district, and residential cooking fuel.

Table 2-3 lists the number of cohorts associated with each demographic group by study area. Each of the six nonworking demographic groups defined for Los Angeles is associated with 60 cohorts, one for each combination of home district, residential cooking fuel, and replicate number ($10 \times 2 \times 3 = 60$). Each of the four working demographic groups is associated with 600 cohorts, one for each combination of home district, work district, residential cooking fuel, and replicate number ($10 \times 10 \times 2 \times 3 = 600$). The total number of Los Angeles cohorts is thus $(6 \times 60) + (4 \times 600)$ or 2,760. As indicated in Table 2-3, a similar process using a replicate number of 5 produced 1,800 cohorts for the Denver study area.

2.3 Develop Exposure Event Sequence for Each Population Unit

In the pNEM/CO methodology, the exposure of each cohort is determined by an exposure event sequence (EES) specific to the cohort. Each EES consists of a series of events with durations from 1 to 60 minutes. To permit the analyst to determine average exposures for specific clock hours, the exposure events are defined such that no event falls within more than one clock hour. Each exposure event assigns the cohort to a particular combination of geographic area and microenvironment. In addition, each event specifies whether or not the cohort is in the presence of smokers. Each event also provides an indication of respiration rate. In the original (1992) version of pNEM/CO, this indicator was a classification of breathing rate as sleeping, slow, medium, or fast. In Version 2.1 of pNEM/CO, this indicator is a specific activity descriptor such as “raking” or “playing baseball.”

In typical pNEM applications, the EESs are determined by assembling activity diary records relating to individual 24-hour periods into a year-long series of records. Because each subject of a typical activity diary study provides data for only a few days, the construction of a year-long EES requires either the repetition of data from one subject or the use of data from multiple subjects. The latter approach is used in pNEM analyses to better represent the variability of exposure that is expected to occur among the persons included in the cohort.

In the pNEM/CO (Version 2.1) analysis, activity diary data were obtained from the Consolidated Human Activity Database (CHAD). At the time of this analysis (1999), CHAD was comprised of approximately 17,000 person-days of 24-hour time/activity data developed from eight surveys (Glen et al., 1997). The surveys included probability-based recall studies conducted by EPA and the California Air Resources Board, as well as real-time diary studies conducted in individual U.S. metropolitan studies using both probability-based and volunteer subject panels. All ages of both genders were represented in CHAD. The data for each subject consisted of one or more days of sequential activities, in which each activity was defined by start time, duration, activity type (140 categories), and microenvironment classification (110 categories). Activities varied from one minute to one hour in duration, with longer activities being subdivided into clock-hour durations to facilitate exposure modeling. A distribution of values for the ratio of oxygen uptake rate to body mass (referred to as metabolic equivalents or “METs”) was provided for each activity type listed in CHAD. The forms and parameters of these distributions were determined through an extensive review of the exercise and nutrition literature. The primary source of distributional data was Ainsworth et al. (1993), a compendium developed specifically to “facilitate the coding of physical activities and to promote comparability across studies.”

The CHAD database was processed to create a special database appropriate for input in pNEM/CO. This database consisted of diary records organized by study subject and calendar day. The diary records for one subject for one calendar day were designated a “person-day.” The CHAD-derived database contained 14,048 usable person-days, each of which was indexed by the following factors:

1. Demographic group
2. Season: "summer" (June through August) or "winter" (all other months)
3. Temperature classification: cool or warm
4. Day type: weekday or weekend.

The demographic group index was determined by the demographic group to which the subject filling out the diary belonged. The season and day indices were based on the date of the calendar day. The temperature classification was based on the daily maximum temperature (in °F) of the associated geographic location on that date. The cool range was defined as temperatures below 55° in winter and temperatures below 84° in summer.

The EES for each cohort was determined by a computerized sampling algorithm. The algorithm was provided with the sequence of daily maximum temperatures reported by the city for the year of the analysis and with a list of cohorts. The temperature data were used to assign each calendar day to one of the temperature ranges used in classifying the activity diary data. To construct the EES for a particular cohort, the algorithm selected a person-day from the CHAD-derived database for each calendar day according to the demographic group of the cohort and the season, day type, and temperature classification associated with the time period.

Each exposure event within an EES was defined by (1) district, (2) CHAD location descriptor, (3) microenvironment, (4) CHAD activity descriptor, and (5) passive smoking status. The district was either the home or work district associated with the cohort. The home/work determination was based on a decision rule which was applied to the sequence of exposure events associated with each person-day. Starting with the midnight event, each event was assigned to the home district unless the activity code associated with the event explicitly indicated the subject was at work. Whenever an explicit work code was encountered in the sequence, each subsequent event was assigned to the work district until an explicit home event was encountered. Each subsequent event was then assigned to the home district until an explicit work event was encountered. This assignment procedure was continued until the end of the person-day was reached.

The CHAD location descriptor refers to the location code associated with the event in the CHAD database. The location descriptor was used to assign the event to a microenvironment and to determine the contribution of passive smoking (if applicable) to the CO concentration experienced during the event. Table 2-4 lists the 120 codes used to define the location descriptors of exposure events.

Table 2-5 lists the 15 microenvironments used for event assignments. Each microenvironment is identified as to a general location (e.g., outdoors) and a specific location (e.g., near road). The list includes two indoor microenvironments related to residences, seven indoor microenvironments related to nonresidential buildings, three outdoor microenvironments, and three vehicle microenvironments. The majority of these microenvironments are aggregates of two or more of the CHAD location descriptors. Only location descriptions associated with similar average CO exposures were combined in defining the aggregate microenvironments. Researchers determined these similarities through an analysis of personal CO monitoring data obtained from the Denver activity diary study (Johnson, 1984). Table 2-4 shows the assignment of CHAD location descriptors to microenvironments.

Table 2-4. Assignment of CHAD Location Codes to pNEM/CO Microenvironments

CHAD Location Code	pNEM/CO Microenvironment Code ^a
<30> Home	
30000: residence, general	1
30010: your residence	1
30020: other's residence	1
30100: residence, indoor	1
30120: your residence, indoor	1
30121: kitchen	1
30122: living room/ family room	1
30123: dining room	1
30124: bathroom	1
30125: bedroom	1
30126: study/ office	1
30127: basement	1
30128: utility room/ laundry room	1
30129: other indoor	1
30130: other's residence, indoor	1
30131: kitchen	1
30132: living room/ family room	1
30133: dining room	1
30134: bathroom	1
30135: bedroom	1
30136: study/ office	1
30137: basement	1
30138: utility room/ laundry room	1
30139: other indoor	1

CHAD Location Code	pNEM/CO Microenvironment Code ^a
30200: residence, outdoor	11
30210: your residence, outdoor	11
30211: pool, spa	11
30219: other outdoor	11
30220: other's residence, outdoor	11
30221: pool, spa	11
30229: other outdoor	11
30300: garage	9
30310: indoor garage	9
30320: outdoor garage	11
30330: your garage	9
30331: indoor garage	9
30332: outdoor garage	11
30340: other's garage	9
30341: indoor garage	9
30342: outdoor garage	11
30400: other, residence	1
<31> Travel	
31000: travel, general	12
31100: motorized travel	12
31110: car	12
31120: truck	13
31121: truck (pick-up or van)	13
31122: truck (other than pick-up or van)	13
31130: motorcycle/ moped/ motorized scooter	10
31140: bus	14

CHAD Location Code	pNEM/CO Microenvironment Code ^a
31150: train/ subway/ rapid transit	14
31160: airplane	CO = 0
31170: boat	11
31171: motorized boat	11
31172: unmotorized boat	11
31200: non-motorized travel	11
31210: walk	11
31220: bicycle/ skateboard/ roller-skates	10
31230: in a stroller or carried by an adult	11
31300: waiting	11
31310: wait for bus, train, ride (at stop)	10
31320: wait for travel, indoors	7
31900: other travel	14
31910: other vehicle	14
<32-34> Other Indoor	
32000: other indoor, general	7
32100: office building/ bank/ post office	7
32200: industrial plant/ factory/ warehouse	8
32300: grocery store/ convenience store	7
32400: shopping mall/ non-grocery store	3
32500: bar/ night club/ bowling alley	5
32510: bar/ night club	5
32520: bowling alley	6
32600: repair shop	
32610: auto repair shop/ gas station	2
32620: other repair shop	3

CHAD Location Code	pNEM/CO Microenvironment Code ^a
32700: indoor gym/ sports or health club	8
32800: childcare facility	8
32810: childcare facility, house	1
32820: childcare facility, commercial	8
32900: public building/ library/ museum/ theater	7
32910: auditorium, sport's arena, concert hall	6
32920: library, courtroom, museum, theater	7
33100: laundromat	7
33200: hospital/ health care facility/ doctor's office	8
33300: beauty parlor/ barber shop/ hair dresser's	7
33400: at work: no specific location, moving among locations	7
33500: school	8
33600: restaurant	4
33700: church	8
33800: hotel/ motel	7
33900: dry cleaners	8
34100: parking garage	15
34200: laboratory	7
34300: other, indoor (specify)	7
<35-36> Other Outdoor	
35000: other outdoor, general	11
35100: sidewalk/ street/ neighborhood	10
35110: within 10 yards of street	10
35200: public garage/ parking lot	15
35210: public garage	15
35220: parking lot	15

CHAD Location Code	pNEM/CO Microenvironment Code ^a
35300: service station/ gas station	15
35400: construction site	11
35500: amusement park	11
35600: school grounds/ playground	11
35610: school grounds	11
35620: playground	11
35700: sports stadium and amphitheater	11
35800: park/ golf course	11
35810: park	11
35820: golf course	11
35900: pool, river, lake	11
36100: restaurant, picnic	11
36200: farm	11
36300: other outdoor (specify)	11

^aSee Table 2-5 for explanation of microenvironment codes.

Activity descriptors were defined according to activity classifications appearing in CHAD. CHAD provides a distribution of energy expenditure rate for each activity classification which was used in a later step to estimate a ventilation rate for each activity (see Section 5 of Johnson, Mihlan, LaPointe, et al., 2000). Appendix A lists the CHAD descriptors and associated distributions for energy expenditure rate.

The effects of active smoking on CO exposure were not addressed in the exposure analysis described here. Because of the coding conventions used in the CHAD diary studies, passive smoking patterns could be determined for nonsmoking subjects only. Consequently, the activity diaries sampled in constructing EESs were limited to those of nonsmokers (a total of 8,077 adult person-days of data). The diary record associated with each exposure event provided information on whether or not the subject was in the presence of smokers. This information was used to assign a passive smoking status to each event.

Table 2-5. Methodology Used to Estimate Carbon Monoxide Concentrations in Each Microenvironment Defined for the Denver and Los Angeles pNEM/CO Analyses

Microenvironment			Activity diary locations included in microenvironment	CO sources treated by hour mass-balance model	CO sources treated by minute mass-balance model
Code	General location	Specific location			
1	Indoors	Residence	Indoors - residence	Outdoor CO Gas stoves	ETS
2	Indoors	Nonresidence A	Service station Auto repair	Outdoor CO	
3	Indoors	Nonresidence B	Other repair shop Shopping mall	Outdoor CO	
4	Indoors	Nonresidence C	Restaurant	Outdoor CO ETS	
5	Indoors	Nonresidence D	Bar	Outdoor CO ETS	
6	Indoors	Nonresidence E	Other indoor location Auditorium	Outdoor CO	
7	Indoors	Nonresidence F	Store Office Other public building	Outdoor CO	
8	Indoors	Nonresidence G	Health care facility School Church Manufacturing facility	Outdoor CO	
9	Indoors	Residential garage	Residential garage	Outdoor CO	
10	Outdoors	Near road	Near road Bicycle Motorcycle	Not applicable	
11	Outdoors	Other locations	Outdoor residential garage Construction site Residential grounds School grounds Sports arena Park or golf course Other outdoor location	Not applicable	
12	Vehicle	Automobile	Automobile		Outdoor CO ETS
13	Vehicle	Truck	Truck		Outdoor CO ETS
14	Vehicle	Mass transit vehicles	Bus Train/subway Other vehicle	Outdoor CO	
15	Outdoor	Public parking or fueling facility	Indoor parking garage Outdoor parking garage Outdoor parking lot Outdoor service station	Not applicable	

2.4 Estimate Exposure and Resulting Physiological Effects

In the pNEM/CO analyses described in this example, each cohort was represented by a sequence of exposure events spanning a calendar year. Probabilistic algorithms within pNEM/CO provided estimates of the CO concentration and alveolar ventilation rate associated with each exposure event. A biokinetics model within pNEM/CO then processed these estimates together with physiological data specific to the cohort to develop an estimate of the COHb level at the end of each hour.

2.4.1 Pollutant Concentration

In the pNEM/CO analysis, each exposure event within a particular EES was indexed according to district *d*, microenvironment *m*, person-day *p*, clock hour *h*, start time *t*, and duration *u*. The exposure associated with a particular event (CEXP(*d,m,p,h,t,u*)) was estimated by the expression

$$\text{CEXP}(d,m,p,h,t,u) = \text{CHR}(d,m,p,h) + \text{CMIN}(d,m,t,u). \quad (2-1)$$

CHR(*d,m,p,h*) is the hourly-average CO concentration determined for microenvironment *m* in district *d* for person-day *p* and hour *h*. Values of CHR(*d,m,p,h*) for enclosed microenvironments (i.e., buildings and enclosed vehicles) were obtained from a mass-balance model with an averaging time of one hour. CMIN(*d,m,t,u*) is the average of *u* one-minute CO concentrations spanning the exposure event. A mass-balance model with an averaging time of one minute was used to estimate values of CMIN(*d,m,t,u*). As discussed in Section 4 of Johnson, Mihlan, LaPointe, et al. (2000), each of the two mass-balance models was capable of accounting for the effects of outdoor CO concentration, air exchange rate, indoor sources, and enclosure volume.

The method used in applying the mass-balance models varied according to the microenvironment where the event occurred. Table 2-5 lists the 15 microenvironments defined for Version 2.1 of pNEM/CO and indicates the CO sources considered in modeling each microenvironment.

The one-hour mass-balance model was used to estimate the contribution of outdoor CO levels to indoor CO levels in Microenvironments Nos. 2, 3, 6, 7, 8, and 9. These indoor microenvironments were assumed to have no significant indoor CO sources. The one-hour model was also used to estimate the combined contribution of outdoor CO and indoor gas stove emissions to indoor CO levels for Microenvironment No. 1 (indoors - residence). In addition, this model was used to estimate the combined contribution of outdoor CO and indoor environmental tobacco smoke (ETS) to CO levels in Microenvironment Nos. 4 (restaurants) and 5 (bars), when smoking was permitted by local regulations (i.e., Denver only).

The one-minute mass-balance model was used to estimate the minute-by-minute contribution of ETS to CO levels in Microenvironment No. 1 (indoors - residence). The one-minute model was also used to estimate the combined minute-by-minute contribution of outdoor CO and inside ETS to CO levels in Microenvironments Nos. 12 (automobiles) and 13 (trucks).

ETS was assumed to occur in Microenvironments Nos. 1 (indoors - residence), 12 (automobiles), and 13 (trucks) whenever the presence of one or more smokers was

indicated by the diary data used to define the exposure event. In applications of pNEM/CO to Denver, ETS was assumed to occur constantly in Microenvironment Nos. 4 (restaurants) and 5 (bars), as Denver does not restrict smoking in these microenvironments. As California currently bans smoking in bars and restaurants, analysts assumed that Microenvironments Nos. 4 and 5 were always free of ETS when applying pNEM/CO to Los Angeles.

ETS was not considered to be a significant source of CO in the remaining microenvironments. This assumption may underestimate CO levels in some of these microenvironments. However, analysts were unable to find sufficient data to develop realistic estimates for the contribution of passive smoking to these microenvironments.

As indicated previously, a review of the scientific literature concerning five other sources (kerosene space heaters, gas space heaters, wood stoves, fireplaces, and attached garages) indicated that (1) fireplaces and stoves did not contribute significantly to indoor CO levels, (2) kerosene and gas space heaters were used in less than 1 percent of the residences in each study area, and (3) attached garages could not be adequately characterized by available data (Fletcher and LaPointe, 1998). EPA and the model-development team recognized that Version 2.1 of pNEM/CO did not characterize the elevated CO exposures that may occur in certain circumstances related to these additional indoor sources (e.g., extended use of unvented kerosene space-heaters with inadequate room ventilation).

In general, the CHR(d,m,p,h) term was used to represent the component of exposure contributed by ambient (outdoor) CO concentrations, by the operation of residential gas stoves, and by ETS in smokey environments such as bars and restaurants. An array of CHR(d,m,p,h) values was created for each cohort. Each array consisted of a set of year-long sequences of hourly-average CHR(d,m,p,h) values, one sequence for each combination of microenvironment and district. The district was either the home or work district specified for the cohort. When an exposure event occurring during hour h assigned a cohort to a particular combination of microenvironment and district, the cohort was assigned the CHR(d,m,p,h) value specified for hour h in the designated microenvironment/district sequence.

Each year-long sequence of hourly average CHR(d,m,p,h) values was generated by the hour mass-balance algorithm described in Section 4 of Johnson, Mihlan, LaPointe, et al. (2000). Briefly, this algorithm estimated the hourly average indoor CO concentrations during hour h as a function of the indoor CO concentration during the preceding hour (i.e., hour h - 1), the CO concentration outdoors during hour h, the air exchange rate during hour h, and the indoor emissions of CO from gas stoves (Microenvironment No. 1) or ETS (Microenvironment Nos. 4 and 5) during hour h. Values for the air exchange rate, gas stove emission rate, and ETS emission rate were sampled from appropriate distributions on a yearly, seasonal, or daily basis. During each clock hour, gas stoves were probabilistically determined as "on" for 30 minutes, "on" for 60 minutes, or "off" for the entire hour. The probability of being on varied with time of day according to use patterns observed during the Denver activity diary study (Akland, Hartwell, Johnson, and Whitmore, 1985; Johnson, 1984).

The CMIN(d,m,t,u) term was used for two purposes: (1) representing total CO exposure for exposure events occurring in automobiles (Microenvironment No. 12) and trucks (Microenvironment No. 13) and (2) representing the contribution of ETS to exposures occurring in the indoors - residence microenvironment. In both cases, the

minute mass-balance model was used to generate one-minute CO concentrations on an “as needed” basis. These values were then averaged over the duration of the event.

Whenever a cohort was assigned to an automobile or truck for a trip consisting of one or more sequential exposure events, the minute mass-balance model was used to estimate one-minute CO concentrations for the total duration of the trip. These calculations accounted for outdoor CO levels, air exchange rate, vehicle volume, and CO emissions from ETS (if any). The resulting one-minute CO concentrations were averaged over each exposure event of the trip.

Whenever a cohort was assigned to the indoors - residence microenvironment and the exposure event indicated one or more smokers were present, the minute mass-balance model was used to generate a series of one-minute CO concentrations spanning the event. These calculations accounted for air exchange rate, building volume, and CO emissions from ETS. The one-minute CO concentrations were averaged over the exposure event to determine a value for CMIN(d,m,t,u). Note that this value represented the contribution of ETS only. The contribution of outdoor CO and gas stoves emissions were included in the CHR(d,m,p,h) term associated with the exposure event.

The outdoor CO concentration required by each mass-balance algorithm was determined for each hour through a Monte Carlo process based on the equation

$$CO_{out}(c,m,d,h) = M(m) \times L(c, m, d) \times T(c,m,d,h) \times [CO_{mon}(d,h)]^A \quad (2-2)$$

in which

- $CO_{out}(c,m,d,h)$ = outdoor CO concentration for cohort c with respect to microenvironment m in district d during hour h,
- $M(m)$ = multiplier (> 0) specific to microenvironment m,
- $L(c,m,d)$ = location multiplier (> 0) specific to cohort c, microenvironment m, and district d (held constant for all hours),
- $T(c,m,d,h)$ = time-of-day multiplier (> 0) specific to cohort c, microenvironment m, district d, and hour h,
- $CO_{mon}(d,h)$ = monitor-derived CO concentration for hour h in district d, and
- A = exponent ($A > 0$).

This equation was used to generate a year-long sequence of outdoor one-hour CO concentrations for each combination of cohort (c), microenvironment (m), and district (d). The exponent A was set equal to 0.621 and held constant for all sequences. The value of M(m) varied only with microenvironment as indicated in Table 2-6.

A value of the location factor L(c, m, d) was specified for each individual sequence and held constant for all hours in the sequence. The value was randomly selected from a lognormal distribution with geometric mean (GM_L) equal to 1.0 and geometric standard deviation (GSD_L) equal to 1.5232. The natural logarithms of this

Table 2-6. Estimated Values of Parameters in Equation 2-2.

Microenvironment ^a			Activity diary locations included in microenvironment	Parameter Estimates for Equation 2-2			
Code	General location	Specific location		A	σ_L	σ_T	M(m)
1	Indoors	Residence	Indoors - residence	0.621	0.4208	0.4879	1.034
2	Indoors	Nonresidence A	Service station Auto repair	0.621	0.4208	0.4879	2.970
3	Indoors	Nonresidence B	Other repair shop Shopping mall	0.621	0.4208	0.4879	1.213
4	Indoors	Nonresidence C	Restaurant	0.621	0.4208	0.4879	1.213
5	Indoors	Nonresidence D	Bar	0.621	0.4208	0.4879	1.213
6	Indoors	Nonresidence E	Other indoor location Auditorium	0.621	0.4208	0.4879	1.213
7	Indoors	Nonresidence F	Store Office Other public building	0.621	0.4208	0.4879	1.213
8	Indoors	Nonresidence G	Health care facility School Church Manufacturing facility	0.621	0.4208	0.4879	0.989
9	Indoors	Residential garage	Residential garage	0.621	0.4208	0.4879	1.034
10	Outdoors	Near road	Near road Bicycle Motorcycle	0.621	0.4208	0.4879	1.607
11	Outdoors	Other locations	Outdoor res. garage Construction site Residential grounds School grounds Sports arena Park or golf course Other outdoor location	0.621	0.4208	0.4879	1.436
12	Vehicle	Automobile	Automobile	0.621	0.4208	0.4879	3.020
13	Vehicle	Truck	Truck	0.621	0.4208	0.4879	3.020
14	Vehicle	Mass transit vehicles	Bus Train/subway Other vehicle	0.621	0.4208	0.4879	3.020
15	Outdoor	Public parking or fueling facility	Indoor parking garage Outdoor parking garage Outdoor parking lot Outdoor service station	0.621	0.4208	0.4879	2.970

^aAggregate microenvironments defined for statistical analysis of Denver PEM data: residence (1 and 9), service/parking (2 and 15), commercial (3 through 7), and vehicle (12 through 14).

distribution can be characterized by a normal distribution with an arithmetic mean (μ_L) equal to zero and an arithmetic standard deviation (σ_L) equal to 0.4208.

A value of the time-of-day factor $T(c, m, d, h)$ was randomly selected for each hour within a sequence from a lognormal distribution with geometric mean (GM_T) equal to 1.0 and geometric standard deviation (GSD_T) equal to 1.6289. The natural logarithms of this distribution follow a normal distribution with an arithmetic mean (μ_T) equal to zero and an arithmetic standard deviation (σ_T) equal to 0.4879.

The $CO_{out}(c, m, d, h)$ term is interpreted as the outdoor CO concentration in the immediate vicinity of microenvironment m in district d during hour h . $CO_{mon}(d, h)$ is the CO concentration reported for hour h by a nearby fixed-site monitor selected to represent district d .

Equation 2-2 is based on the results of data analyses that suggest that the relationship between $CO_{out}(c, m, d, h)$ and $CO_{mon}(d, h)$ should account for the identity of the microenvironment, the geographic location of the microenvironment, and the time of day. Numerous statistical models could be developed. In specifying the Equation 2-2 model, analysts attempted to balance the need for simplicity and parsimony with the need to model the most important patterns in the available data. Most of the model development was based on a comparison of hourly averages of 10-minute concentrations measured outside residences in southern California (Wilson, Colome, and Tian, 1995) with hourly averages measured at the nearest fixed site monitor. The derivation of Equation 2-2 and the associated parameter estimates is described by Johnson, Mihlan, LaPointe, et al. (2000). A summary can be found in Subsection 7.6.1 of this report.

Equation 2-2 requires a complete (gapless) year of hourly average fixed-site monitoring values for each district. Section 3.2 of Johnson, Mihlan, LaPointe, et al. (2000) describes the method used to fill in missing hourly-average values. The resulting filled-in data sets were assumed to represent existing conditions at each monitor.

As discussed previously, each exposure event assigned the cohort to a district and a microenvironment. For exposure events occurring in non-vehicle microenvironments, it seemed reasonable to assign the event to a single district represented by a single fixed-site monitor. Consequently, the values of $CO_{mon}(d,h)$ for a non-vehicle microenvironment in district d were obtained from the fixed-site monitoring station assigned to district d . The locations of exposure events occurring in vehicle microenvironments were more difficult to characterize, as some trips were likely to have crossed two or more districts. Researchers assumed that an average of the fixed-site monitoring values from all districts would be more appropriate in these situations. Consequently, the value of $CO_{mon}(d,h)$ for a vehicle microenvironment during hour h was obtained by averaging the $CO_{mon}(d,h)$ values for all districts for hour h . Note that this approach is likely to underestimate the occurrence of high ambient CO levels during periods when people are assigned to the vehicle microenvironments.

2.4.2 Alveolar Ventilation Rate

In addition to CO concentration, a value for alveolar ventilation rate (V_A) value was estimated for each exposure event. V_A is expressed as liters of air respired per minute (liters min^{-1}). The algorithm used to estimate V_A was developed for Version 2.1 of pNEM/CO and had not been used previously in pNEM analyses. Section 5 of

Johnson, Mihlan, LaPointe, et al. (2000) provides a detailed description of the algorithm.

Briefly, the CHAD database provided an activity indicator for each exposure event. Each activity type was assigned a distribution of values for the metabolic equivalent of work (MET). MET is a dimensionless quantity defined by the ratio

$$\text{MET} = \text{EE}/\text{RMR}, \quad (2-3)$$

where EE is the rate of energy expenditure during a particular activity (expressed in kcal/min), and RMR is a person's typical resting metabolic rate (also expressed in kcal/min). For example, activity no. 11300 -- "outdoor chores" -- was represented by a normal distribution of MET values with mean equal to 5 and a standard deviation equal to 1. Appendix A lists the distribution assigned to each of the CHAD activity codes.

A probabilistic procedure was used to assign a RMR value to each cohort for the entire 365-day exposure period. An EE value was calculated for each exposure event by the equation

$$\text{EE}_a(i,j,k) = [\text{MET}(i,j,k)][\text{RMR}(k)], \quad (2-4)$$

in which $\text{EE}_a(i,j,k)$ was the average energy expenditure rate (kcal min^{-1}) for cohort k during exposure event i on day j; $\text{MET}(i,j,k)$ was a value for MET randomly selected from the distribution associated with activity type a; a was the activity type associated with exposure event e; and $\text{RMR}(c,d)$ was the RMR value randomly generated for cohort k. Section 5.5 of Johnson, Mihlan, LaPointe, et al. (2000) describes the methods used to randomly select or generate the required parameter values.

Energy expenditure requires oxygen which is supplied by ventilation (respiration). Let ECF(k) indicate an energy conversion factor defined as the volume of oxygen required to produce one kilocalorie of energy in person k. The oxygen uptake rate (VO_2) associated with a particular activity can be expressed as

$$\text{VO}_2(i,j,k) = [\text{ECF}(k)][\text{EE}_a(i,j,k)], \quad (2-5)$$

in which $\text{VO}_2(i,j,k)$ has units of liters oxygen min^{-1} , ECF(k) has units of liters oxygen kcal^{-1} , and $\text{EE}(i,j,k)$ has units of kcal min^{-1} . In pNEM/CO, the value of $\text{VO}_2(i,j,k)$ is determined from $\text{MET}(i,j,k)$ by substituting Equation 2-4 into Equation 2-5 to produce the relationship

$$\text{VO}_2(i,j,k) = [\text{ECF}(k)][\text{MET}(i,j,k)][\text{RMR}(k)]. \quad (2-6)$$

Subsection 5-5 of Johnson, Mihlan, LaPointe, et al. (2000) describes the probabilistic methods used to estimate values of ECF(k) and RMR(k) for person k.

V_A represents the portion of the minute ventilation that is involved in gaseous exchange with the blood. VO_2 is the oxygen uptake that occurs during this exchange. The absolute value of V_A is known to be affected by total lung volume, lung dead space, and respiration frequency -- parameters which vary according to person and/or exercise rate. However, it is reasonable to assume that the ratio of V_A to VO_2 is relatively constant regardless of a person's physiological characteristics or energy

expenditure rate. Consistent with this assumption, Version 2.1 of pNEM/CO converts each estimate of $VO_2(i,j,k)$ to an estimate of $V_A(i,j,k)$ by the proportional relationship

$$V_A(i,j,k) = (19.63)[VO_2(i,j,k)] \quad (2-7)$$

in which both V_A and VO_2 are expressed in units of liters min^{-1} . This relationship was obtained from an article by Joumard, Chiron, Vidon, et al. (1981), who based it on research by Galetti (1959). Equation 2-7 was applied to all cohorts under all energy expenditure rates.

The V_A algorithm included a method for identifying “impossible” values which were occasionally generated by the estimation process. This method determined a maximum VO_2 value for each exposure event which accounted for the duration of the activity and for the age, weight, and gender of the person. No estimate of VO_2 (and the corresponding estimate of V_A) was permitted to exceed this limit. Subsection 5.4 of Johnson, Mihlan, LaPointe, et al. (2000) provides a more detailed description of this procedure.

2.4.3 Carboxyhemoglobin Level

An algorithm developed by Biller and Richmond [included as an appendix in Johnson, Capel, Paul, and Wijnberg (1992)] was used in the 1992 version of pNEM/CO to estimate the COHb level at the end of each exposure event. The algorithm is based on a differential equation proposed by Coburn, Forster, and Kane (1963). Inputs to the algorithm include

- Percent COHb at the start of the event
- Average CO exposure concentration during the event, ppm
- Time duration of the event, min
- Alveolar ventilation rate, ml/min
- Haldane Constant
- Atmospheric pressure at sea level, torr
- Altitude above sea level, feet
- Blood volume, ml
- Total hemoglobin content of blood, gm/100 ml
- Pulmonary CO diffusion rate, ml/min per torr
- Endogenous CO production rate, ml/min

Version 2.1 of pNEM/CO employed a slightly revised version of the algorithm to estimate end-of-event COHb levels. Section 10 provides a detailed description of the COHb algorithm, the various physiological parameters that are inputs to the COHb algorithm, and a list of related references.

2.4.4 The Physiological Profile Generator

As discussed in Subsections 2.4.2 and 2.4.3, the algorithms used to estimate V_A and COHb required values for various physiological parameters such as body mass, blood volume, and RMR. Section 10 provides a complete list of these parameters. A

special algorithm within pNEM probabilistically generated a value for each parameter on the list (collectively referred to as a “physiological profile”) for each cohort processed by pNEM/CO. Each of the generated physiological profiles was internally consistent, in that the functional relationships among the various parameters were maintained. For example, blood volume was determined as a function of weight and height, where height was estimated as a function of weight. Weight was in turn selected from a distribution specific to gender and age. Section 10 describes the method used to estimate values for each parameter in the application of pNEM/CO to Denver and Los Angeles.

2.4.5 Hourly Average Exposure Estimates

Algorithms within pNEM/CO provided four estimates for each exposure event: average CO concentration, average V_A , the product of average CO concentration and V_A (represented as “CO x V_A ”), and the COHb level at the end of the event. These estimates were processed to produce time-weighted estimates of CO concentration, V_A , and CO x V_A for each clock hour, as well as end-of-hour estimates of COHb. The result was a year-long sequence of hourly values for CO, V_A , CO x V_A , and COHb for each cohort. These sequences were statistically analyzed to determine the value of various multihour exposure indicators of interest, including the largest eight-hour daily maximum CO concentration occurring each year and the number of times the end-of-hour COHb level exceeded a specified percentage value.

2.5 Extrapolate Unit Exposures and Effects to Population-of-Interest

2.5.1 General Population

The cohort-specific exposure estimates developed in Step 4 of the pNEM methodology (Section 2.4) were extrapolated to the general study area population by estimating the population size of each cohort. Cohort populations were estimated in three steps. The 1990 population of each demographic group within a particular home district [$POP_{90}(d,h)$] was first estimated from 1990 census data specific to that district. Each of these groups was subdivided into a group residing in homes with gas stoves and a group residing in homes with other cooking fuels. The population of each of these groups in the target year specified for the pNEM/CO analysis (Denver - 1995, Los Angeles -1997) was determined by the expression

$$POP_{9x}(d,h,f) = F(h,f) \times AF_{9x} \times POP_{90}(d,h) \quad (2-8)$$

where $POP_{9x}(d,h,f)$ is the target-year population of a group associated with demographic group d , home district h , and cooking fuel f . $F(h,f)$ is the fraction of homes in Home district h that use cooking fuel f .

Analysts estimated that $F(h,f) = 19.6$ percent for the Denver study area and $F(h,f) = 79$ percent for the Los Angeles area. Subsection 4.4.3 of Johnson, Mihlan, LaPointe, et al. (2000) describes the methodology used in developing these estimates.

AF_{9x} is a city-specific factor which adjusts 1990 census data to provide 1997 population estimates for Los Angeles and 1995 population estimates for Denver. In

developing the Los Angeles AF_{9x} value (1.051), analysts first determined the average annual growth rate of Los Angeles and Orange Counties from 1990 to 1998 (0.72 percent/year). Based on this assumed annual growth rate, analysts estimated the Los Angeles study area population increased by 5.1 percent from 1990 to 1997. The Denver AF_{9x} value (1.087) was developed in a similar manner based on the population growth of the four counties containing parts of the Denver study area (Arapahoe, Boulder, Denver, and Jefferson). Analysts determined that the annual rate of population increase was 1.69 percent. Using this value, they estimated that the population of the population of the Denver study area increased by 8.7 percent from 1990 to 1995.

The $POP_{9x}(d,h,f)$ values provided an estimate of the target-year population of each non-commuting cohort residing within home district h . The target-year populations of the commuting cohorts (assumed to include all working cohorts) were determined by the expression

$$COM_{9x}(d,h,f,w) = POP_{9x}(d,h,f) \times COM(h,w)/WORK(h). \quad (2-9)$$

$COM_{9x}(d,h,f,w)$ is the number of persons in the commuting cohort associated with demographic group d , home district h , cooking fuel f , and work district w ; $COM(h,w)$ is the number of workers in all demographic groups that commute from home district h to work district w ; and $WORK(h)$ is the total number of workers in home district h . Estimates of $WORK(h)$ were developed from census data specific to each district. Section 6 of Johnson, Mihlan, LaPointe, et al. (2000) describes the method used to estimate $COM(h,w)$ from origin-destination data provided by the BOC.

2.5.2 Persons with Ischemic Heart Disease

The cohort-specific exposure estimates developed in Step 4 of the pNEM methodology were also extrapolated to the sensitive population defined as persons with diagnosed and undiagnosed ischemic heart disease (IHD). The extrapolation was performed using the procedure described in Subsection 2.5.1 with a single variation: the following equation was substituted for Equation 2-8.

$$POP_{9x}(d,h,f) = IHD(d) \times F(h,f) \times AF_{9x} \times POP_{90}(d,h). \quad (2-10)$$

The term $IHD(d)$ is the fraction of persons in demographic group d with IHD.

Estimates of the prevalence of IHD by demographic group were provided by H. Richmond (memorandum, August 25, 1998). Table 2-7 lists these estimates as percentages. In each case, a total prevalence rate is provided which is the sum of a prevalence rate for diagnosed IHD and a prevalence rate for undiagnosed IHD. Estimates of diagnosed IHD were obtained from the National Health Interview Survey (Adams and Marano, 1995), in which U.S. prevalence rates were disaggregated by age and gender. The estimated prevalence of diagnosed IHD for children (age 0 to 17) is 0.01 percent. According to the National Health Interview Survey, approximately 8.0 million individuals are estimated to have diagnosed IHD in the civilian, non-institutionalized population. These estimates do not include individuals in the military or

Table 2-7. Percentage of Persons with Ischemic Heart Disease (IHD) by Demographic Group

Demographic group	Percentage of persons with IHD		
	Diagnosed ^a	Undiagnosed ^b	Total
1. Children, 0 to 17	0.01	0.004	0.014
2. Males, 18 to 44, working	0.38	0.17	0.55
3. Males, 18 to 44, nonworking	0.38	0.17	0.55
4. Males, 45 to 64, working	8.19	3.60	11.8
5. Males, 45 to 64, nonworking	8.19	3.60	11.8
6. Males, 65+	19.2	8.45	27.7
7. Females, 18 to 44, working	0.13	0.06	0.19
8. Females, 18 to 44, nonworking	0.13	0.06	0.19
9. Females, 45 to 64, working	3.25	1.43	4.68
10. Females, 45 to 64, nonworking	3.25	1.43	4.68
11. Females, 65+	12.3	5.41	17.7

^aSource: Richmond (1998) compilation based on Adams and Marano (1995).

^bSource: Richmond (1998) based on American Heart Association (1990) estimate and assumption that persons with undiagnosed IHD are distributed within the population in the same proportions as persons with diagnosed IHD.

individuals in nursing homes or other institutions.

The estimates of undiagnosed IHD in Table 2-7 were based on two assumptions: (1) there are 3.5 million persons in the U.S. with undiagnosed IHD and (2) persons with undiagnosed IHD are distributed within the population in the same proportions as persons with diagnosed IHD. The 3.5 million statistic was based on an estimate by the American Heart Association (1990) that there are between three and four million persons with undiagnosed IHD.

Table 2-8 lists the resulting Denver population estimates by exposure district for (1) all adults and (2) adults with IHD. The total number of adults with IHD in the six-district study area is approximately 48,400. District No. 3 has the largest number of adults with IHD (about 14,700), accounting for 30 percent of the total. On average, 5.7 percent of the adults are estimated to have IHD.

Table 2-9 provides similar estimates for Los Angeles. As expected, the Los Angeles study area has a larger number of adults with IHD (approximately 258,000). Accounting for 16 percent of the total number, District No. 3 has the largest number of adults with IHD (about 41,800). The Los Angeles study area is estimated to have a
 Table 2-8. Estimates of Population Residing in Each Denver Exposure District.

Denver exposure district	All adults		Adults with ischemic heart disease	
	Number	Percent of total	Number	Percent of total
1	119,085	14.0	6,430 (5.4) ^a	13.3
2	83,805	9.9	4,740 (5.7)	9.8
3	237,061	28.0	14,703 (6.2)	30.3
4	161,963	19.1	10,665 (6.6)	22.0
5	154,395	18.2	8,369 (5.4)	17.3
6	91,584	10.8	3,550 (3.9)	7.3
All	847,892	100.0	48,457 (5.7)	100.0

^a Number in parentheses is percentage of adults with ischemic heart disease.

Table 2-9. Estimates of Population Residing in Each Los Angeles Exposure District.

Los Angeles exposure district	All adults		Adults with ischemic heart disease	
	Number	Percent of total	Number	Percent of total
1	514,488	10.6	31,701 (6.2) ^a	12.3
2	437,960	9.0	25,631 (5.8)	9.9
3	888,622	18.3	41,813 (4.7)	16.2
4	571,207	11.8	25,388 (4.4)	9.8
5	298,199	6.1	15,701 (5.3)	6.1
6	443,409	9.1	26,678 (6.0)	10.3
7	453,220	9.3	25,336 (5.6)	9.8
8	506,428	10.4	26,010 (5.1)	10.1
9	484,451	10.0	25,268 (5.2)	9.8
10	257,760	5.3	14,653 (5.7)	5.7
All	4,855,744	100.0	258,180 (5.3)	100.0

^a Number in parentheses is percentage of adults with ischemic heart disease.

slightly lower prevalence rate of adults with IHD (5.3 percent).

In extrapolating the cohort-specific exposure estimates developed in Step 4 to persons with IHD, analysts assumed the activity patterns of IHD were similar to those of the general population. Subsection 5.6 of Johnson, Mihlan, LaPointe, et al. (2000) presents the results of a statistical analysis performed to evaluate the reasonableness of this assumption.

2.6 Advice to Modelers

The pNEM/CO methodology was developed specifically to meet the requirements of OAQPS for a computer-based model capable of simulating the CO exposures and resulting COHb levels of specific population groups under alternative NAAQS. In addition to meeting these needs, the designers of pNEM/CO attempted to create a model which is flexible in application and easy to upgrade. The model was deliberately constructed as a collection of stand-alone algorithms organized within a modular framework. For this reason, analysts can revise individual algorithms without the need to make major changes to other parts of the model.

It is important to understand that the structure of each algorithm in Version 2.1 of pNEM/CO was largely determined by the characteristics of the available input data. For example, the algorithm used to construct a year-long exposure event sequence for each cohort was constrained by the fact that none of the available time/activity studies provided more than three days of diary data for any one subject. To make maximum use of the available diary data, the pNEM/CO sequencing algorithm constructed each exposure event sequence by sampling data from more than one subject. The other pNEM/CO algorithms were similarly designed to make best use of available data bases.

The series of pNEM models applicable to ozone (pNEM/O3) are similar to the pNEM/CO models in that each model incorporates the general features of the five-step methodology summarized at the beginning of Section 2. However, the pNEM/O3 and pNEM/CO models differ greatly with respect to the details of implementing each step. These differences in methodology are the result of a number of fundamental differences in the properties of the two pollutants, for example:

Chemical reactivity: Ozone is a highly reactive; CO is non-reactive.

Emission sources and formation mechanisms: Ozone is typically produced by photochemical reactions of various precursors. CO is emitted directly into the atmosphere

Ambient concentration gradients: CO tends to exhibit steeper spatial concentration gradients than ozone

Indoor sources: Indoor sources of CO are much more prevalent than indoor sources of ozone.

Sensitive population groups: EPA has identified outdoor workers and children who tend to be active outdoors as populations of particular concern with respect to ozone exposure. In recent pNEM/CO applications, EPA has defined the sensitive population as adults (18 and older) with ischemic heart disease.

In addition, ozone and CO differ with respect to the databases available for implementing each step, particularly the databases required to characterize pollutant concentrations in microenvironments.

Sections 3 through 10 of this report describe selected modeling components (input databases, algorithms, parameter distributions, etc.) that have been used in various NEM-type and HAPEM-type models. Researchers who plan to incorporate any of these components in other models should (1) carefully review the history of each candidate component, (2) note any features of the component that are pollutant specific, and (3) consider how data constraints determined the final form of the component.

2.7 References for Section 2

Adams, P. F., and M. A. Marano. 1995. Current Estimates from the National Health Interview Survey, 1994. National Center for Health Statistics, Vital Health Statistics 10 (193).

Ainsworth, et al. 1993. "Compendium of Physical Activities: Classification of Energy Costs of Human Physical Activities," Med. Sci. Sports Exer. Volume 25, pp. 71 - 80.

Akland, G., T. D. Hartwell, Ted R. Johnson, and Roy W. Whitmore. 1985. "Measuring Human Exposure to Carbon Monoxide in Washington, D.C., and Denver, Colorado, during the Winter of 1982-1983." Environmental Science and Technology, Volume 19, No. 10, pp. 911-918 (October 1985).

American Heart Association. 1990. 1990 Heart and Stroke Facts. American Heart Association, Dallas, Texas. p. 13.

Coburn, R., R. Forster, and R. Kane. 1963. "Considerations of the Physiology and Variables that Determine Blood Carboxyhemoglobin Concentration in Man." Journal of Clinical Investigations, Vol. 44, pp. 1899-1910.

Fletcher, K. and J. LaPointe. November 22, 1998. "Review of Documents on Indoor Sources of Carbon Monoxide. (Attachment)." Johnson, Ted. December 2, 1998. "Memorandum No. 2: Evaluation of Five Potential Indoors Sources of Carbon Monoxide." Report prepared by TRJ Environmental, Inc., under EPA Contract No. 68-D6-0064.

Galetti, P. M. "Respiratory Exchanges During Muscular Effort." Helv. Physiol. Acta. Vol. 17, pp. 34 - 61 (1959).

Glen, J. A., Y. Lakkadi, J. A. Trippett, and M. del Valles-Torres. 1997. Development of NERL/CHAD: The National Exposure Research Laboratory Consolidated Human Activity Database. Report prepared by ManTech Environmental Technology, Inc., under EPA Contract No. 68-D5-0049.

Johnson, T. R. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014. U. S. Environmental Protection Agency,

Research Triangle Park, North Carolina.

Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM. Report prepared by International Technology Air Quality Services under EPA Contract No. 68-D0-0062, U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina 27711.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared for the Risk and Exposure Assessment Division of U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

Joumard, R, M. Chiron, R. Vidon, M. Maurin, and J-M Rouzioux. 1981. "Mathematical Models of the Uptake of Carbon Monoxide on Hemoglobin at Low Carbon Monoxide Levels." Environmental Health Perspectives. Vol. 41, pp. 277 - 289 (October 1981).

Richmond, H. 1998. "Memorandum: Sensitive Population Estimates for Use in Carbon Monoxide Exposure Analysis." U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina 27711. August 25.

USEPA. 2000. Air Quality Criteria for Carbon Monoxide. Report No. EPA/600/P-99/001. Office of Research and Development, U.S. Environmental Protection Agency. , Research Triangel Park, North Carolina.

Wilson, A.L., S.D. Colome, and Y. Tian. 1993. California Residential Indoor Air Quality Study, Volume 1, Methodology and Descriptive Statistics, Appendix. Integrated Environmental Services, Irvine, California.

SECTION 3

SPATIAL AND TEMPORAL COVERAGE OF EXPOSURE ANALYSIS

In past applications of the NEM and HAPEM-type exposure models, analysts have referred to the geographical region where exposures can occur as the “study area”. The study area is typically defined as the unification of two or more exposure districts. To facilitate population estimates, each exposure district is defined as a collection of one or more census units as defined by the Bureau of Census (BOC). Depending on the application, the census units are either blocks, block groups, or census tracts. Demographic data for these census units are acquired from publically-available data files provided by the BOC.

This section provides guidance for selecting appropriate census units to be used in an exposure analysis, for combining these units into exposure districts, and for determining the air quality that should be assigned to each district. The section concludes with guidance for selecting the time period of the exposure analysis.

3.1 Census Units

Census data are available for blocks, block groups, census tracts, and block numbering areas (BNA's). A block, the smallest geographic unit defined by the BOC, is defined as a geographical area surrounded on all sides by roads (or similar boundaries such as rivers and railroads) but not intersected by a road. Note that land area and population can vary widely from block to block. A block group is a collection of two or more blocks. A census tract is a collection of up to nine block groups with a total population of 1,500 to 8,000 people (600 to 3,000 housing units) with an average of about 4,000 people (1,500 housing units). In sparsely-populated rural areas, the BOC substitutes “block numbering areas” (BNA's) for census tracts. Figure 3-1 shows selected census tracts in Durham, North Carolina.

Each census tract has a basic census tract number composed of no more than four digits, and may have an optional two-digit decimal suffix. Blocks are identified within census tracts by a three digit format in which the first digit indicates the block group ID and the last two digits indicates the block ID.

The BOC provides a variety of files on tape and compact disc listing 1990 and 2000 demographic data for hierarchies of census units. Two of the most useful databases for 1990 are Summary Tapes 1A and 3A. Overviews of these databases can be found in Tables 3-1 and 3-2, respectively. Note that both databases provide data at the census tract and block group level, but not at the block level. Data for 1990 at the block level by state can be obtained from Summary Tape 1B, although these data are not as detailed as the data provided by Summary Tapes 1A and 3A.

Summary Tape 1 for the 2000 census replaces Summary Tapes 1A and 1B for 1990, providing data for census tracts, block groups, and blocks. Table 3-3 provides a brief description of this database. The 2000 version of Summary Tape 3 will not be available until 2002.

Top of Figure 3-1

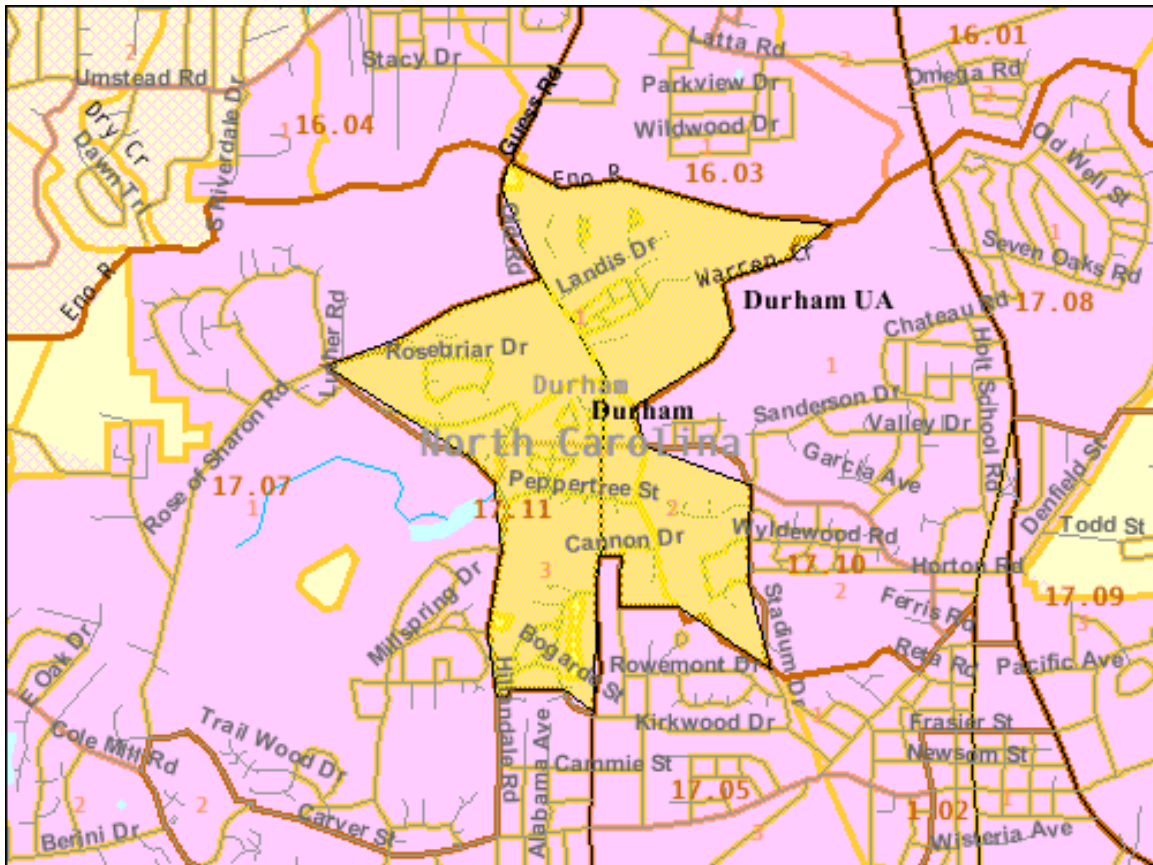


Figure 3-1. Selected census tracts in Durham, North Carolina.

Table 3-1. Description of 1990 Summary Tape File 1A (STF 1A)^a

Cost: \$100 per disc; \$900 for set of 17 discs ordered at same time. Add \$25 shipping fee for international orders (including Canada and Mexico). Technical documentation: \$10 per copy (FREE with product). See Bureau of Census web site for ordering information (http://www.census.gov/geo/www/cen_tract.html).

Subject content: 100 percent data. Population items include age, race, sex, marital status, Hispanic origin, household type, and household relationship. Population items are cross-tabulated by age, race, Hispanic origin, or sex. Housing items include occupancy/vacancy status, tenure, units in structure, contract rent, meals included in rent, value, and number of rooms in housing unit. Housing data are cross-tabulated by race or Hispanic origin of householder or by tenure. Selected aggregates and medians also are provided.

Geography: States and their subareas in hierarchical sequence down to the block group level; also summaries for the State portion of American Indian and Alaska Native areas, whole places, whole tracts/BNA's, whole county subdivisions in selected States and whole block groups.

Note: Also available in its entirety on computer tape, and partially in printed reports

File format & Software: dBase III+; comes with GO display/retrieval software.

Year 2000 Compliant: The software on the Summary Tape File 1A (STF 1A) CD-ROM is Year 2000 compliant. It does not process date information and can be used during and after calendar year 2000 without any interruption or error.

Census contact: Customer Services, 301-763-INFO(4636).

^aSource: http://www.census.gov/geo/www/cen_tract.html.

Table 3-2. Description of 1990 Summary Tape File 3A (STF 3A)^a

Cost: \$100 per disc; \$4,340 for entire set of 62 discs. Add \$25 shipping fee for international orders including Canada and Mexico). Technical documentation: \$15 if purchased separately, FREE with product purchase. See Bureau of Census web site for ordering information (http://www.census.gov/geo/www/cen_tract.html).

Subject content: Contains sample data weighted to represent the total population. In addition, the file contains 100-percent counts and unweighted sample counts for total persons and total housing units.

Population items: Age, Mobility limitation status, Ancestry, Occupation, Citizenship, Place of birth, Class of worker, Place of work, Educational attainment, Poverty status, Employment status, Private vehicle occupancy, Family type, Race, Farm and nonfarm population, Residence in 1985, Foreign-born status, School enrollment, Group quarters, Self-care limitation status, Hispanic origin, Sex, Household type and relations, Travel time to work, Income in 1989, Urban and rural population, Industry, Veteran/military status, Language spoken at home, Work disability status, Marital status, Work status in 1989, Means of transportation to work, Workers in family in 1989.

Housing Items: Age of householder, Rooms, Bedrooms, Selected monthly owner costs, Condominium status, Sewage disposal, Farm and nonfarm housing, Telephone availability, Heating fuel, Tenure, Hispanic origin of householder, Units in structure, Housing units, Urban and rural housing, Kitchen facilities, Utilities in rent, Meals included in rent, Value of housing unit, Mortgage status, Vehicles available, Occupancy status, Water source, Plumbing facilities, Year householder moved into unit, Race of householder, Year structure built, Rent.

Geography: States and their subareas in hierarchical sequence down to the State, county, place (or place part), census tract/block numbering area, block group (or part), State portion of American Indian and Alaska Native area (with trust lands and with no trust lands), Alaska Native Regional Corporation, State portion of MSA, CMSA, and PMSA, State portion of urbanized area. Also, inventory (complete) summaries for: Census tract/block numbering area, block group, place, consolidated city.

File format & software: dBASE III+; files for each State are labeled STF300ss.DBF through STF334ss.DBF where ss is the two-character State abbreviation. Disc comes with GO display/retrieval software.

Year 2000 Compliant: The software on the Summary Tape File 3A (STF 3A) CD-ROM is Year 2000 compliant. It does not process date information and can be used during and after calendar year 2000 without any interruption or error.

Census contact: Customer Services (orders), 301-763-INFO(4636); Population Division's Statistical Information Staff (content) 301-457-2422.

^aSource: http://www.census.gov/geo/www/cen_tract.html.

Table 3-3. Description of 2000 Summary Tape File 1 (STF 1)^a

Census 2000 Summary File 1 (ASCII) on DVD (two disks cover the entire U.S., \$70 per disk). Summary File 1 contains population and housing unit data based on Census 2000 questions asked of all people and about every housing unit. A total of 286 data tables are included in this file. Population items include: sex, age, race, Hispanic or Latino origin, household relationship, and household and family characteristics. Housing items include occupancy status and tenure (whether the unit is owner- or renter-occupied). For most subjects, data for census block groups and census blocks are also shown.

Technical documentation for Summary File 1 provides information on subject matter, a list of table matrices, summary level sequence charts as well as a data dictionary. Technical documentation is included with the purchase of the CD-ROM but may also be purchased as an individual item. Also available on the Internet in PDF format.

See <https://catalog.mso.census.gov/esales4boc> for information on ordering files.

^aSource: <https://catalog.mso.census.gov/esales4boc>

3.2 Combining Census Units Into Exposure Districts

3.2.1 Exposure Districts Based on Proximity to Monitor Locations

In applications of the NEM-and HAPEM-type models, analysts typically assume that ambient (outdoor) air quality is the same at all locations within a particular exposure district during a particular time period. If the air quality value is derived from data reported by fixed-site monitors, then it may be appropriate to use the locations of these monitors as centers for the exposure districts. In past applications of pNEM to ozone and CO, analysts have defined a radius for each monitor such that all census tracts that fell within that radius were assigned to the exposure district centered on that monitor. In some applications (see Subsection 5.3.1.2), the same radius value was used for all monitors. In other applications, the value for the radius depended on the criteria used in citing the monitor. Monitors characterized by EPA as “regional” were assigned larger radius values than monitors characterized as “neighborhood”.

3.2.2 Exposure Districts Based on Neighborhood Types

In an early application of NEM to CO (Johnson and Paul, 1983), analysts employed a method based on “neighborhood types” to allocate census tracts to exposure districts. Each census tract in the study area was categorized as to its predominant land use (residential, commercial, industrial, etc.). An exposure district (referred to as a neighborhood type) was defined for each land use such that the district contained only the census tracts assigned to that category. The air quality data for the exposure district was obtained from the fixed-site monitor within the study area judged to best represent the specified land use. In implementing this approach, analysts believed that land use would be a better predictor of ambient CO levels than geographic location.

In later applications of pNEM to CO, analysts examined more sophisticated approaches for classifying neighborhood types. In a technical letter to EPA, Johnson (1991) compared data measured by personal exposure monitors (PEM's) in Denver, Colorado (Johnson, 1984; Akland, Hartwell, Johnson, and Whitmore, 1985) with concurrent data measured by fixed-site monitors in the same area to determine which fixed-site monitors were the best predictors of PEM values. He investigated schemes for matching fixed-site monitors to PEM locations (defined by census tracts) based on four neighborhood-type indicators (population density, housing density, distance from center of city, and traffic density) and on distance. He also evaluated matching the same monitor to all PEM locations, using a “composite monitor” defined by averaging the data from all fixed-site monitors, and matching the nearest monitor to each PEM location. Matching PEM and fixed-site locations by any of the neighborhood-type indicators was found to be inferior to the other approaches in producing a good correlation between PEM and fixed-site values; matching PEM locations with the nearest fixed-site monitor generally produced the best results. Based on these findings, EPA elected to use the monitor-proximity method described in Subsection 3.2.1 to define exposure districts for the pNEM/CO exposure analyses conducted after 1991 (e.g., Johnson, Mihlan, LaPointe, et al., 2000). However, analysts should continue to investigate the neighborhood type approach as a viable option whenever ambient

pollutant levels are expected to be affected by land use.

3.2.3 Exposure Districts Based on Receptor Locations

In exposure analyses of toxic pollutants, a dispersion model is often used to estimate ambient (outdoor) air quality at each point within a defined receptor array or grid. As the BOC provides locations for the geographic centroids of blocks and census tracts, it is convenient to use these centroids as the basis for the receptor array. The choice of unit (block vs. census tract) depends on the level of spatial resolution required by the analysis. Smaller units (i.e., blocks) are usually required in areas where there are steep gradients with respect to any of the exposure factors of concern (pollutant concentration, emission density, population demographics, etc.). Larger census units (census tracts or aggregates of census tracts) may be adequate for areas where gradients are more gradual.

In many applications the receptor grid is predefined as a rectangular or polar array of points which do not line up with the centroids of census units. In these cases, the analyst can use an interpolation scheme to map the concentration values onto the centroids of appropriate census units. The following example demonstrates the use of an interpolation scheme appropriate when the dispersion model array is a polar grid.

Consider a radial grid defined by 12 concentric rings and 16 equally-spaced radial spokes as specified in Table 3-4. Each point in the polar grid can be identified by a ring index (r) and a spoke index (s). The coordinates of each point are (rd, az) in which rd is the distance from the center of the grid to ring r and az is the azimuth (angle) of spoke s measured clockwise from north (0 degrees). The concentration for period h at each point of the grid is denoted by $C_{r,s}$. The following approach can be used to estimate the concentration for period h at the centroid of a particular census tract located within the polar grid.

1. Identify the four grid points that surround the centroid. These points are defined by two consecutive rings ($r = a$ and b) and two adjacent spokes ($s = c$ and d) and can be identified as (a,c), (a,d), (b,c), and (b,d). The coordinates of these four points are (rd_a, az_c), (rd_a, az_d), (rd_b, az_c), and (rd_b, az_d), respectively. The pollutant concentrations for time period h associated with these four points are denoted as $C_{a,c}$, $C_{a,d}$, $C_{b,c}$, and $C_{b,d}$, respectively.

The coordinates of the centroid are denoted (rd, az) where rd is the radial distance from the grid center to the centroid and az is the azimuth of the centroid.

Example: A centroid is located at (0.65 km, 110 degrees) with respect to the radial grid presented in Table 3-4. The four nearest grid points are thus located at (0.5 km, 90 degrees), (0.5 km, 112.5 degrees), (1.0 km, 90 degrees), and (1.0 km, 112.5 degrees). The dispersion model has estimated the pollutant concentrations associated with these four points as 10.0, 16.0, 5.0, and 8.0 $\mu\text{g}/\text{m}^3$, respectively.

Table 3-4. Radial Grid Used for Interpolation Example Presented in Text.

Index type	Index	Value
Ring	1	0.1 km
	2	0.5 km
	3	1.0 km
	4	2.0 km
	5	5.0 km
	6	10.0 km
	7	15.0 km
	8	20.0 km
	9	25.0 km
	10	30.0 km
	11	40.0 km
	12	50.0 km
Spoke	1	0 degrees
	2	22.5 degrees
	3	45.0 degrees
	4	67.5 degrees
	5	90.0 degrees
	6	112.5 degrees
	7	135.0 degrees
	8	157.5 degrees
	9	180.0 degrees
	10	202.5 degrees
	11	225.0 degrees
	12	247.5 degrees
	13	270.0 degrees
	14	292.5 degrees
15	315.0 degrees	
16	337.5 degrees	

2. Calculate the natural logarithm of concentration for each of the four grid points surrounding the centroid.

$$\text{LNC}_{a,c} = \ln(C_{a,c})$$

$$\text{LNC}_{a,d} = \ln(C_{a,d})$$

$$\text{LNC}_{b,c} = \ln(C_{b,c})$$

$$\text{LNC}_{b,d} = \ln(C_{b,d})$$

Example: Taking natural logarithms of the four concentration values, we obtain $\text{LNC}_{a,c} = \ln(10.0) = 2.303$, $\text{LNC}_{a,d} = \ln(16.0) = 2.773$, $\text{LNC}_{b,c} = \ln(5.0) = 1.609$, and $\text{LNC}_{b,d} = \ln(8.0) = 2.079$.

3. Calculate the radial interpolation ratio

$$\text{ratio} = [\ln(rd) - \ln(rd_a)] / [\ln(rd_b) - \ln(rd_a)]$$

Example: Using the specified input values ($rd = 0.65$, $rd_a = 0.5$, and $rd_b = 1.0$), we calculate

$$\begin{aligned} \text{ratio} &= [\ln(0.65) - \ln(0.5)] / [\ln(1.0) - \ln(0.5)] \\ &= (-0.431 + 0.693) / (0 + 0.693) = (0.262) / (0.693) = 0.378. \end{aligned}$$

4. Calculate the azimuthal interpolation ratio

$$\text{aratio} = (az - az_c) / (az_d - az_c)$$

All azimuths are expressed as angles.

Example: Using the specified input values ($az = 110$, $az_c = 90$, and $az_d = 112.5$), we calculate

$$\text{aratio} = (110 - 90) / (112.5 - 90) = 20 / 22.5 = 0.889.$$

5. Calculate interpolated logarithmic concentrations on spokes c and d (denoted LNC_c and LNC_d , respectively)

$$\text{LNC}_c = \text{LNC}_{a,c} + (\text{LNC}_{b,c} - \text{LNC}_{a,c}) \times \text{ratio}$$

$$\text{LNC}_d = \text{LNC}_{a,d} + (\text{LNC}_{b,d} - \text{LNC}_{a,d}) \times \text{ratio}$$

Example: Using the values calculated in Steps 3 and 5, we calculate

$$\text{LNC}_c = 2.303 + (1.609 - 2.303) \times 0.378 = 2.041$$

$$\text{LNC}_d = 2.773 + (2.079 - 2.773) \times 0.378 = 2.511.$$

6. Exponentiate the logarithmic concentration values calculated in Step 5.

$$C_c = \exp(\text{LNC}_c)$$

$$C_d = \exp(\text{LNC}_d).$$

Example: Using the values calculated in Step 5, we calculate

$$C_c = \exp(2.041) = 7.7 \mu\text{g}/\text{m}^3$$

$$C_d = \exp(2.511) = 12.3 \mu\text{g}/\text{m}^3.$$

7. Linearly interpolate in the azimuthal direction to yield C, an estimate of the pollutant concentration at the centroid during time period h.

$$C = C_c + (C_d - C_c) \times \text{aratio}.$$

Example: Using the values calculated in Steps 4 and 6, we calculate

$$C = 7.7 + (12.3 - 7.7) \times 0.889 = 11.8 \mu\text{g}/\text{m}^3.$$

In our example, we estimate the pollutant concentration at the centroid during time period h to be $11.8 \mu\text{g}/\text{m}^3$.

After the analyst has selected a method for assigning air quality values to census units, he or she has two basic options with respect to defining exposure districts:

- (1) treat each census unit as a separate exposure district with its own distinct air quality or
- (2) create aggregate exposure districts by combining census units and averaging air quality values.

The first alternative obviously provides a greater degree of spatial resolution. However, the run time of an exposure model is usually proportional to the number of exposure districts (or the square of this number). Consequently, the second alternative (aggregating census units into a relatively small number of exposure districts) may be preferable when short run times are desired and pollutant gradients are not too steep.

3.3 Population-of-Interest

The population-of-interest in an exposure assessment may be the entire population of a defined study area or a subpopulation with specified demographic characteristics. Subpopulations are typically identified as having either unusual sensitivity to the pollutant under analysis or greater-than-average exposure to the

pollutant. For example, EPA specified “adults (18 and older) with ischemic heart disease” as the sensitive group for the pNEM/CO application described in Section 2. In the pNEM/O3 analysis described by Johnson, Capel, McCoy, and Mozier (1996), the population-of-interest was defined according to exposure potential as “outdoor workers”.

To the extent possible, the analyst should define a subpopulation in terms of the classifications appearing in the BOC databases for the census units used to define the exposure districts (blocks, block groups, or census tracts). These classifications include various breakdowns of the population by age, gender, work status, education, housing characteristics, and many other factors.

If the definition of the subpopulation includes characteristics not found in the BOC database (e.g., presence of a specific health condition), then the analyst must apply appropriate adjustment factors. This process is illustrated by the example presented in Subsections 2.5.1 and 2.5.2 in which adjustment factors for cooking fuel use and ischemic heart disease were determined from non-BOC sources and then applied to BOC-derived population estimates that accounted for gender, age, and work status.

The census data used for estimating the population-of-interest will typically be obtained from a recent decennial census. The BOC “American Factfinder” web site provides the following brief description of the decennial census:

The decennial census occurs every 10 years, in the years ending in "0," to count the population and housing units for the entire United States. Its primary purpose is to provide the population counts that determine how seats in the U.S. House of Representatives are apportioned. Census figures also are required to draw congressional and state legislative district boundaries, to allocate federal and state funds, to formulate public policy, and to assist with planning and decision-making in the private sector.

The decennial census uses both short- and long-form questionnaires to gather information. The short form asks a limited number of basic questions. These questions are asked of all people and housing units, and are often referred to as 100-percent questions because they are asked of the entire population. The long form asks more detailed information from approximately a 1-in-6 sample, and includes the 100-percent questions as well as questions on education, employment, income, ancestry, homeowner costs, units in a structure, number of rooms, plumbing facilities, etc.

Source: <http://factfinder.census.gov/home/en/decennialdata.html>

Table 3-5 provides a brief description of the 2000 census and lists the types of data items appearing on the long-form and short-form questionnaires.

As the data obtained from a particular census represents population counts at the time of the census, it may be necessary to adjust the census data for population growth or decline since the census. Perhaps the simplest method for making this adjustment is to apply a single factor to all population values where the factor represents the percent change in total population between the most recent census year

Table 3-5. Information on the 2000 Decennial Census^a.

Census 2000

The Census Bureau conducted censuses in the United States, Puerto Rico, American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, and the United States Virgin Islands. Statistical data from all the censuses are available through American FactFinder. The reference date for Census 2000 is April 1, 2000 (Census Day).

Census 2000 Short-Form Questionnaire

The data from the Census 2000 short-form questionnaire are also referred to as the “100-percent characteristics”.

- | | |
|---------------------------|-------------------------|
| Age | Sex |
| Hispanic or Latino origin | Tenure |
| Household relationship | Vacancy characteristics |
| Race | |

Census 2000 Long-Form Questionnaire

The data from the Census 2000 long-form questionnaire are also referred to as the “sample characteristics”.

- | <u>Population</u> | <u>Housing</u> |
|--|---|
| Ancestry | Heating fuel |
| Disability | Number of rooms, number of bedrooms |
| Farm residence | Occupation, industry, class of worker |
| Grandparents as caregivers | Plumbing and kitchen facilities |
| Income in 1999 | Telephone service |
| Labor force status | Units in structure |
| Language spoken at home | Utilities, mortgage, taxes, insurance, fuel |
| Marital status | Value of home or monthly rent paid |
| Migration (residence since 1995) | Vehicles available |
| Place of birth, citizenship, year of entry | Work status in 1999 |
| Place of work, journey to work | Year moved into residence |
| School enrollment and educational attainment | Year structure built |
| Veteran status | |

^aSource: <http://factfinder.census.gov/home/en/decennialdata.html>

and the year of the exposure assessment. For example, Johnson, Mihlan, LaPointe, et al. (2000) multiplied 1990 population counts for Denver, Colorado, by 1.087 to obtain population estimates for 1995. The 1.087 value was determined by examining the growth of the four counties that contained parts of the Denver study area (Arapahoe, Boulder, Denver, and Jefferson). Analysts determined that the average annual rate of population increase was 1.69 percent. Using this value, they estimated that the population of the Denver study area increased 8.7 percent from 1990 to 1995.

More sophisticated adjustment methods may be required when the analyst believes that population change varies significantly among population groups. For example, there may be evidence for a desirable retirement community that the percentage increase in population from in-migration is larger for retirees than young adults. In this case, the population adjustment method should account for varying rates of migration by age and work status.

Information on population projection methods used by the BOC can be found at the following BOC website:

<http://www.census.gov/ftp/pub/population/www/projections/popproj.html>.

This site also provides access to the Population Projections Program, a series of files that list projections by year (1995 through 2025) of the resident population for each of the 50 states and the District of Columbia by age, sex, race, and Hispanic origin, consistent with the national population projections. The state projections are based on state-specific data on births, deaths, international migration, and domestic migration. Although several alternative series are produced, the preferred or “middle series” (based on the assumption that past and current trends will continue) is most commonly used.

3.4 Exposure Period

Researchers have used exposure models to assess exposure periods ranging from a single event measured in minutes to an entire lifetime. The NEM and HAPEM series of models have typically been applied to exposure periods ranging from three months to a entire calendar year (i.e., January 1 to December 31). In these cases, the model provides estimates for relatively short time periods that are aggregated and tabulated by the model over the defined exposure period to provide exposure estimates for various averaging times of interest. For example, pNEM/CO constructs an exposure event sequence for each cohort that spans an entire calendar year. The model then provides an exposure estimate for each individual exposure event in the sequence where the events have durations of 1 to 60 minutes. The event-specific exposure estimates can be aggregated as required to calculate exposure estimates with averaging times of one hour, 8 hours, 24 hours, one year, or any other averaging time specified as an integer number of hours.

Researchers must frequently balance two or more competing goals in selecting the exposure period. For example, analysts attempted to satisfy the following goals in selecting a year for the application of pNEM/CO to Denver, Colorado:

- the year should fall between 1993 and 1997,
- the air quality data for the selected year should be at least 75 percent complete at a minimum of six monitoring sites, and
- the air quality data for the selected year should represent conditions at, or near to, attainment of the current NAAQS for CO.

Subsection 3.1.1 of Johnson, Mihlan, LaPointe, et al. (2000) describes the evaluation process used in selecting 1995 as the exposure period for Denver.

3.5 References for Section 3

Akland, G., T. Hartwell, T. Johnson, and R. Whitmore. 1985. "Measuring Human Exposure to Carbon Monoxide in Washington, D.C., and Denver, Colorado, during the Winter of 1982-1983." Environmental Science and Technology, Volume 19, No. 10, pp. 911-918 (October 1985).

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. 1991. "Evaluation of the Neighborhood Concept as a Means of Estimating Population Exposure to Carbon Monoxide." Letter to Tom McCurdy, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina, May 1.

Ted Johnson, J. Capel, M. McCoy, and J. Mozier. 1996. Estimation of Ozone Exposures Experienced by Outdoor Workers in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared by IT Air Quality Services for the Air Quality Management Division, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, April.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared for the Risk and Exposure Assessment Division, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

Johnson, T., and R. Paul. 1983. The NAAQS Exposure Model (NEM) Applied to Carbon Monoxide. Report No. EPA-450/5-83-003. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

SECTION 4

REPRESENTING DEMOGRAPHIC VARIABILITY

This section describes two methods that have been used in NEM-type and HAPEM-type exposure models to represent the demographic variability of people included in the population-of-interest:

- (1) defining a comprehensive set of cohorts and
- (2) using Monte Carlo techniques to generate a representative set of individuals.

These methods are described in Sections 4.1 and 4.2, respectively. Section 4.3 provides guidance in choosing between the two methods.

4.1 Define a Comprehensive Set of Cohorts

Until recently, all NEM-type and HAPEM-type exposure models have divided the population-of-interest into a set of cohorts such that each person is assigned to one and only one cohort. Each cohort is assumed to contain persons who experience identical exposures during the specified exposure period. Cohort exposure is typically assumed to be a function of demographic group (age, gender, work status, etc.), location of residence, and location of work place. Specifying the home and work district of each cohort provides a means of linking cohort exposure to ambient pollutant concentrations. Specifying the demographic group provides a means of linking cohort exposure to activity patterns which vary with age, work status, and other demographic variables. In some analyses, cohorts are further distinguished according to factors relating to proximity to emission sources or time spent in particular microenvironments. Analysts use census and other demographic data to estimate the number of people represented by each cohort.

In the application of Version 2.1 of pNEM/CO to Denver and Los Angeles (Johnson, Mihlan, LaPointe, et al., 2000), each cohort was identified as a distinct combination of (1) home district, (2) demographic group, (3) work district (if applicable), (4) residential cooking fuel, and (5) replicate number. The home district and work district of each cohort were identified according to a list of potential districts defined for each study area. Table 2-3 lists 10 adult demographic groups defined for the pNEM/CO analyses. Four of the demographic groups were identified as workers. Each cohort associated with one of these groups was identified by both home and work district. The remaining cohorts were identified only by home district. The residential cooking fuel of each cohort was identified as either “natural gas” or “other.”

Earlier versions of pNEM/CO defined cohorts solely according to home district, demographic group, work district (if applicable), and residential cooking fuel. A new feature was installed in Version 2.1 of pNEM/CO which permitted the user to specify a

“replication” value (n) such that the model will produce n cohorts for each combination of these four indices. Because pNEM/CO uses a probabilistic process to construct an activity pattern for each cohort, each of the n cohorts associated with a particular combination of home district, demographic group, work district, and residential cooking fuel is associated with a distinct exposure sequence. The replication feature permits the analyst to divide the population-of-interest into a larger number of smaller cohorts -- a process that decreases the “lumpiness” of the exposure simulation. Replication values of n = 5 and n = 3, respectively, were specified for the Denver and Los Angeles exposure analyses described in this report. Consequently, the pNEM/CO model analyzed n times the number of cohorts it would have considered for each city had the cohorts been defined solely by home district, demographic group, work district, and residential cooking fuel.

Table 2-3 lists the number of cohorts associated with each demographic group by study area. Each of the six nonworking demographic groups defined for Los Angeles is associated with 60 cohorts, one for each combination of home district, residential cooking fuel, and replicate number ($10 \times 2 \times 3 = 60$). Each of the four working demographic groups is associated with 600 cohorts, one for each combination of home district, work district, residential cooking fuel, and replicate number ($10 \times 10 \times 2 \times 3 = 600$). The total number of Los Angeles cohorts is thus $(6 \times 60) + (4 \times 600)$ or 2,760. As indicated in Table 2-3, a similar process using a replicate number of 5 produced 1,800 cohorts for the Denver study area.

4.2 Randomly Generate a Representative Set of Individuals

The newly developed APEX3 and SimExpo-HAP models employ an alternative method for representing variability within the population-of-interest. Each of these models employs a Monte Carlo process that randomly generates n individuals, where n is determined by the user. Each individual is assigned a demographic profile that is used by the model to determine exposure. In general, a demographic profile defines a set of characteristics for the individual that could include age, gender, residential location, work location, weight, occupation, or any other attribute that may be useful in estimating exposure. Ideally, the probability that the model will assign a particular demographic profile to an individual is equal to the probability that the profile will occur in the population-of-interest. Extrapolation of the results to the population-of-interest is simplified because the exposure determined for each individual receives equal weighting. As the number of individuals generated by the model increases, the resulting distribution of individual exposures should converge on the distribution of exposures within the population-of-interest.

4.2.1 APEX3

The current version of APEX (APEX3) is described in a user’s manual prepared by Glen (2002). Briefly, APEX3 employs Monte Carlo processes to estimate year-long patterns of CO exposures and associated COHb levels for a set of randomly-generated individuals. APEX3 provides 22 characteristics (Table 4-1) that can be used to define the demographic profiles of the generated individuals. The characteristics include home location, work location, age, gender, race, employment status, and a collection of

Table 4-1. Characteristics for Defining Demographic Profiles in APEX3 as Applied to CO.

Code for demographic characteristic	Explanation
HSEct	Sector in which the person lives (home)
WSEct	Sector in which the person works (= HSEct for non-workers)
HDis	Air quality (exposure) district for HSEct
WDis	Air quality (exposure) district for WSEct
Zone	Temperature zone for HSEct
DGRP	Demographic group number (1 - 11) as defined in pNEM/CO
Age	Age in years
Gender	Male or female
Race	One of five categories as defined by BOC
Employed	Indicates employment outside home
Stove	Indicates presence of gas stove in home
Pilot	Indicates the presence of a gas pilot light
ACHom	Indicates the presence of air conditioning at home
ACCar	Indicates the presence of air conditioning in the car
Height	Height of the person in inches
Weight	Body mas of the person in pounds
Hemoglob	Amout of hemoglobin in the blood
DiffDay	Lung diffusivity parameter used in COHb calculation
BloodVol	Volume of blood in the body
HemFac	Accounts for change in hemoglobin level with altitude
Endgn1	Endogenous CO production rate no. 1
Endgn2	Endogenous CO production rate no. 2

physiological parameters that are used in estimating COHb levels. To facilitate the comparison of estimates obtained from APEX3 and pNEM/CO, one of the APEX3

profile characteristics (DRGP) indicates the pNEM/CO demographic group that most resembles the APEX3 profile with respect to age and work status.

4.2.2 SimExpo-HAP

The current version of SimExpo-HAP is described in a technical memorandum by Johnson (2001). The exposure model employs a Monte Carlo procedure that generates a user-specified number of demographic profiles. Each profile specifies five characteristics of the person to be modeled: age, gender, work status (in home/away from home), home district, and work district (work district = home district for in-home workers). Districts are defined by individual census tracts. The model also specifies a day of the year (Julian date) which is converted to month and weekday/weekend attributes for the profile.

During each iteration, the model randomly determines these seven profile attributes based on probabilities obtained from appropriate data bases (census data, commuting origin-destination tables, etc.). The model next identifies all activity patterns in the time/activity database (derived from CHAD) that match the age, gender, work status, month, and weekday/weekend attributes of the profile. It then randomly selects one activity pattern for use in determining exposure. The selected activity pattern consists of an unbroken sequence of exposure events that spans a 24-hour midnight-to-midnight period. Each exposure event assigns the person to a particular microenvironment (m) in either the home or work district (d) for a specified time period (t). The exposure during each event [exposure (t, m, d)] is calculated by the same formula used in HAPEM4; i.e.,

$$\text{exposure}(t, m, d) = \text{ADD}(m) + [\text{PEN}(m)][\text{PROX}(m)][\text{AMBCONC}(d, t)] \quad (4-1)$$

in which ADD(m), PEN(m), and PROX(m) are specific to the microenvironment and AMBCONC(d, t) is the hourly average ambient concentration specific to home or work district d and time period t. The event-specific exposures are averaged over the 24-hour period to produce a 24-hour average exposure for the model iteration. The output of the model consists of a file listing each estimated 24-hour average exposure and the corresponding profile information. (With a minor change to the program, the user can output one-hour exposure estimates.)

Currently, the model is set up to step through the 365 days of a specified calendar year and produce n 24-hour exposure estimates per day, where n is specified by the user. The demographic characteristics of the n people associated with each day are determined by a Monte Carlo process such that the distribution of sampled demographic characteristics should approach those of the underlying population as n increases. The memorandum by Johnson (2001) describes the use of SimExpo-HAP to estimate benzene exposures in a study area consisting of 483 census tracts in and around Philadelphia, PA.

4.3 Choosing a Method

In choosing between the two methods described above (predefined cohorts vs. randomly-generated individuals), the analyst should consider the relative merits of each

method with respect to (1) demographic data requirements, (2) model run time, (3) ease of modification, and (4) granularity.

Demographic Data Requirements. Both methods require that the user provide demographic data. The cohort method uses the demographic data to determine the number of people represented by each cohort. The generated-individual method employs demographic data as the basis for tables of probabilities used in assigning a demographic profile to each individual. However, an exposure model based on the generated-individual method will typically require demographic data of greater detail and complexity than one based on the predefined-cohort method. For example, a generated individual will typically be defined by a one-year age interval, whereas a cohort may be defined by multi-year interval (e.g., 21 to 35 years). In this case, the generated-individual method will also require one-year breakdowns in the data provided for any other defining demographic characteristic correlated with age (gender, weight, etc.). In past pNEM and HAPEM analyses, cohorts have often been defined using broad BOC classifications that facilitate data acquisition.

Model Run Time. When running an exposure model incorporating predefined cohorts, the analyst must run the model on the complete set of cohorts to obtain a distribution of exposure estimates for the entire population-of-interest. If the analyst desires to reduce model run times, he or she must reduce the number of cohorts by combining or redefining cohorts. As the cohort definitions are usually “hard-wired” into the model, the analyst may be required to make significant changes to the computer program and/or its databases.

The run time of models that incorporate the Monte Carlo method tends to be proportional to n -- the number of individuals generated. As the user determines n , the user can reduce run time by simply reducing the specified value of n . The user may decide to increase n to provide a larger sample size of exposure estimates to represent the population-of-interest. In either case, the user is not required to modify the structure of the program.

Ease of Modification. In general, exposure models based on predefined cohorts are more difficult to modify than models that employ the Monte Carlo method to generate individuals. Cohort definitions are usually programmed into the model in such a way that changing a cohort definition in one part of the model requires modifications to other parts of the model.

Monte Carlo models such as APEX3 and SimExpo-HAP minimize this problem by generating the demographic profiles of individuals according to detailed probability tables. These tables can be modified by the user without changing other parts of the model.

Granularity. Each cohort represents a fraction of the population-of-interest which is determined by the cohort definition. This fraction may vary considerably among the cohorts such that some cohorts are significantly larger than other cohorts. If cohorts with large population fractions are frequently associated with high exposures, then the exposure distribution may exhibit excessive graininess in the upper tail of the distribution. Under such conditions, it may not be possible to determine precisely the concentration associated with a particular high percentile of interest (e.g., the 99th percentile).

Each individual generated by a Monte Carlo model represents an equal fraction of the population-of-interest. Consequently, all portions of the exposure distribution will

exhibit the same degree of detail. The number of individuals generated (n) determines the maximum exposure percentile.

4.4 References for Section 4

Glen, G. 2001. Draft User's Manual for the APEX3 Model. ManTech Environmental Technology, Inc. Research Triangle Park, NC. March 15.

Johnson, T. 2001. "Brief Description of SimExpo-HAP and Results of Initial Model Runs." Technical memorandum prepared by TRJ Environmental, Inc., for Ted Palma, Risk and Exposure Assessment Division, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. March 11.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared for the Risk and Exposure Assessment Division, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

SECTION 5

EXPOSURE EVENT SEQUENCE

Each of the exposure models discussed in this report provides exposure estimates for a group of defined population units. Depending on the model, this set of population units is composed of either

- (1) a set of individuals generated by a Monte Carlo process to represent the distribution of demographic characteristics within a study area or
- (2) a set of cohorts in which each cohort represents a defined portion of the study area population.

In either case, the exposure model provides a time-ordered exposure event sequence (EES) for each population unit that serves as the basis for developing a corresponding time-ordered sequence of exposures. Some exposure models go further and convert the exposure estimates to metrics relating to dose and/or resulting physiological conditions. For example, Version 2.1 of pNEM/CO provides a year-long EES for each cohort which is used as the basis for estimating a sequence of CO exposures (each defined by CO concentration and alveolar ventilation rate). A module within pNEM/CO converts the sequence of CO exposures into a sequence of dose estimates that are subsequently used to estimate a sequence of COHb levels.

It should be noted that the EES is not the sequence of estimated exposures but rather a sequence of input data that the model uses to develop the exposure estimates. Each event in an EES is typically defined by a start time, a duration, a geographic location, and a microenvironment. The event may also be defined by the presence or absence of specific indoor sources (e.g., an operating gas stove), an activity type (e.g., meal preparation), an exertion level (e.g., low), and other information that may affect exposure and resulting dose. As discussed in Subsection 5.1, the majority of information included in the EES is typically derived from 24-hour time/activity data collected through the use of diaries. Additional information may be obtained from analyses of tenancy and commuting data (see Subsections 5.2 and 5.3, respectively).

5.1 24-HOUR TIME/ACTIVITY DATA

The exposure models described in this report typically employ time/activity data that are organized in 24-hour (midnight-to-midnight) activity patterns. These patterns were obtained from diary studies in which subjects either (1) made diary entries as the events occurred ("real-time" diaries) or (2) completed diaries from memory (retrospective diaries). The majority of the diaries available for use span a single 24-hour period (typically midnight-to-midnight); the remainder include two or three 24-hour periods. Depending on the application, analysts have placed the time/activity data from one or more diary studies into a time/activity database which is accessed by the exposure model. In modeling applications prior to 1994, these databases typically contained data from diary studies conducted in Denver, Colorado; Washington, D.C.;

and Cincinnati, Ohio. As new diary studies have been completed, the database has been expanded to include data from four personal monitoring studies of special populations in Los Angeles, California; a telephone survey of a random sample of California residents; and a personal monitoring study in Valdez, Alaska. A report by Johnson (1995) provides an analysis of an aggregate database that contained time/activity data from 10 diary studies. Currently, the primary source of time/activity data for OAQPS exposure models is the Consolidated Human Activity Database (CHAD). At the time of this report, CHAD contained time/activity data from 12 studies. Subsection 5.1.1 provides an overview of CHAD and lists the diary studies incorporated into the database.

APEX, pNEM/CO, SimExpo-HAP, and other exposure models recently developed by OAQPS use special probabilistic algorithms to select appropriate 24-hour activity patterns from a CHAD-derived database and assemble these patterns into a multi-day EES for each population unit. To illustrate the process, Subsection 5.1.2 describes the method used by pNEM/CO (Version 2.1) to construct an EES for each cohort. This method is based on the assumption that activity patterns are primarily affected by four factors: demographic group, season, daily maximum temperature, and day type (weekday or weekend). Subsection 5.1.3 summarizes the research that led to the decision to include daily maximum temperature as one of these factors. Subsection 5.1.4 identifies other factors that affect activity patterns based on the analysis of 10 time/activity studies conducted by Johnson (1995).

Historically, each 24-hour activity pattern used in the OAQPS exposure models has been based directly (event-by-event) on a single 24-hour activity pattern obtained from a diary study. However, researchers have also evaluated an alternative approach in which 24-hour activity patterns are constructed using “random walk” techniques. Briefly, diary data are statistically analyzed to determine a set of transition probabilities (probabilities of moving from one exposure state to another exposure state at a given time). A 24-hour activity patterns is generated by using these probabilities to simulate the state transitions of a person over a 24-hour period. Subsection 5.1.4 describes an evaluation of this approach conducted in 1987. At that time, analysts found that the method failed to adequately represent correlated activities that were separated by several hours (e.g., morning and evening commutes). Because such correlated activities are important in estimating exposures, the random walk technique has not been used to construct EES’s for the OAQPS exposure models.

5.1.1 The Consolidated Human Activity Database

Currently, the primary source of time/activity data for OAQPS exposure models is the Consolidated Human Activity Database (CHAD). Various versions of this database have been developed by ManTech Environmental Technology, Inc. under contract to the National Exposure Research Laboratory (NERL). In 1999, CHAD was comprised of approximately 17,000 person-days of 24-hour time/activity data developed from eight surveys (Glen et al., 1997). This version was used as the source of time/activity data for Version 2.1 of pNEM/CO. The later June 2000 version (the most recent version at the time of this report) contained approximately 23,000 person-days of 24-hour data obtained from 12 surveys (Table 5-1) drawing subjects from five cities (Denver, CO; Washington, DC; Cincinnati, OH; Valdez, AK; and Los Angeles, CA), the

Table 5-1. Brief descriptions of studies providing data to CHAD (McCurdy et al., 2000).

Study name	Calendar time period of study	Characteristics of subjects ^a	Number of subject-days of data ^b	Type of diary instrument used ^c	Time period of diary ^d	Activity-specific breathing rate data?	Documentation or reference	Notes
Baltimore	January - February 1997; July - August, 1998	Ages 65+	391	Diary, 15 min blocks	24 h standard	No	Williams et al. (2000)	Multiple days, varying from 5 to 15; part of a PM _{2.5} PEM ^e study
CARB: adolescents and adults	October 1987 - September 1998	Ages 12 - 94	1762	Retrospective	24 h standard	No	Robinson et al. (1989), Wiley et al. (1991b)	
CARB: children	April 1989 - February 1990	Ages 0 - 11	1200	Retrospective	24 h standard	No	Wiley et al. (1991a)	
Cincinnati (EPRI)	March - April and August 1985	Ages 0 - 86	2614	Diary	24 h; nominal: 7 pm - 7 am	Yes	Johnson (1989)	3 consecutive days; 186 PD removed
Denver (EPA)	November 1982 - February 1983	Ages 18 - 70	805	Diary	24 h; nominal: 7 pm - 7 pm	No	Johnson (1984), Akland et al. (1985)	Part of CO PEM ^e study; 55 PD removed
Los Angeles: elementary school children	October 1989	Ages 10 - 12	51	Diary	24 h standard	Yes	Spier et al. (1992)	7 PD removed ^f
Los Angeles: high school students	September - October 1990	Ages 13 - 17	43	Diary	24 h standard	Yes	Spier et al. (1992)	23 PD removed ^f
NHAPS - A	September 1992 - October 1994	Ages 0 - 93	4723	Retrospective	24 h standard	No ^g	Klepeis et al. (1995), Tsang and Klepeis (1996)	A national random probability survey
NHAPS - B	As above	Ages 0 - 93	4663	Retrospective	24 h standard	No ^g	As above	As above
University of Michigan: children	February - December 1997	Ages 0 - 13	5616	Retrospective	24 h standard	No	http://www.isr.umich.edu/frc/childevelopment/home.html	2 days of data; one is s weekend
Valdez, Alaska	November 1990 - October 1991	Ages 11 - 71	401	Retrospective	Varying 24 h period	No	Goldstein et al. (1992)	4 PD removed ^f
Washington, DC (EPA)	November 1982 - February 1983	Ages 18 - 98	699	Diary	24 h; nominal: 7 pm - 7pm	No	Hartwell et al. (1984), Akland et al. (1985)	Part of a CO PEM ^e study; 6 PD removed

^aAll studies included both genders. The age range depicted is for the subjects actually included; in most cases, there was not an upper limit for the adult studies. Ages are inclusive. Age 0 = infants < 1 year old.

^bThe actual number of person-days of data in CHAD after the "flagging" and removal of questionable data. See McCurdy et al. (2000) for a discussion of these procedures.

^cRetrospective: a "what did you do yesterday" type of survey; also known as an *ex post* survey. Diary: a "real time" paper diary that a subject carried as he or she went through the day.

^dStandard = midnight-to-midnight.

^ePEM = a personal monitoring study. In addition to the diary, a subject carried a small CO or PM_{2.5} monitor throughout the sampling period.

^fPD removed = the number of person-days of activity pattern data removed from CHAD due to missing activity and formation: completeness criteria are listed in McCurdy et al. (2000).

^gA question was asked regarding which activities (within each 6 h time block in the day) involved "heavy breathing," lifting heavy objects, and running hard.

^hNHAPS = National Human Activity Pattern Study; A = the air version; B = the water version. The activity data obtained on the two versions are identical in format.

states of California and Michigan, and the nation as a whole [the National Human Activity Pattern Survey (NHAPS)]. The surveys include both real-time studies in which subjects enter diary data as the events occur and retrospective studies in which subjects complete diaries after the fact from memory.

CHAD consists of 24-hour activity patterns indexed by age, gender, day of the week, season, health status, temperature, and other useful factors. Each 24-hour activity pattern (a “person-day”) is a sequence of activities in which each activity is defined by start time, duration, activity type (140 categories), and microenvironment classification (110 categories). Activities vary from one minute to one hour in duration, with longer activities being subdivided into clock-hour durations to facilitate exposure modeling.

A distribution of values for the ratio of oxygen uptake rate to body mass (referred to as metabolic equivalents or “METs”) is provided for each activity type listed in CHAD. The forms and parameters of these distributions were determined through an extensive review of the exercise and nutrition literature. The primary source of distributional data was Ainsworth, Haskell, Leon, et al. (1993), a compendium developed specifically to “facilitate the coding of physical activities and to promote comparability across studies.” Section 9 presents a method by which MET values can be assigned to individual events and then used to estimate oxygen uptake rate and ventilation rate.

It should be noted that CHAD is being continually updated by EPA and that the June 2000 version of CHAD discussed here is likely to be superceded by later versions that include additional time/activity data. Subsection 5.1.2 describes the method by which pNEM/CO selects activity data from CHAD and assembles the data into an for each cohort.

5.1.2 Assembling Multi-Day Sequences

In a typical pNEM application, each EES is developed by assembling activity diary records relating to individual 24-hour periods into a year-long series of records. Because each subject of a typical activity diary study provides data for only a few days, the construction of a year-long EES requires either the repetition of data from one subject or the use of data from multiple subjects. The latter approach (using data from multiple subjects) has been employed in the majority of pNEM analyses to better represent the variability of exposure that is expected to occur among the persons included in the cohort. The principal deficiency of this approach is that it may not adequately account for the day-to-day repetition of activities common to individuals. Using activities from different subjects may underestimate multiple occurrences of high exposure situations (e.g., long commutes) for segments of the population who engage in highly repetitive activities.

The multiple-subject approach has been recently used in constructing EES for pNEM/CO. In this application, the CHAD database was processed to create a special database appropriate for input to pNEM/CO. This database consisted of diary records organized by study subject and calendar day. The diary records for one subject for one calendar day were designated a “person-day.” The CHAD-derived database contained 14,048 usable person-days, each of which was indexed by the following factors:

1. Demographic group
2. Season: "summer" (June through August) or "winter" (all other months)
3. Temperature classification: cool or warm
4. Day type: weekday or weekend.

The demographic group index was determined by the demographic group to which the subject filling out the diary belonged. The season and day indices were based on the date of the calendar day. The temperature classification was based on the daily maximum temperature (in °F) of the associated geographic location on that date. The cool range was defined as temperatures below 55° in winter and temperatures below 84° in summer.

The EES for each cohort was determined by a computerized sampling algorithm. The algorithm was provided with the sequence of daily maximum temperatures reported by the city for the year of the analysis and with a list of cohorts. The temperature data were used to assign each calendar day to one of the temperature ranges used in classifying the activity diary data. To construct the EES for a particular cohort, the algorithm selected a person-day from the CHAD-derived database for each calendar day according to the demographic group of the cohort and the season, day type, and temperature classification associated with the time period.

5.1.3 Effect of Daily Maximum Temperature on Time/Activity Patterns

Subsection 5.1.2 describes a method for sequencing EES's in which two critical temperatures (55° for winter and 84° for summer) are used to classify time/activity data. This classification scheme, which has been used in the majority of pNEM analyses since 1988, is based on the results of an analysis of time spent outdoors by subjects of the Cincinnati Activity Diary Study (CADS).

CADS provided 2614 person-days of time/activity data collected during March and August of 1985. In a special study described by Johnson (1988), researchers created a separate database for each month listing the number of minutes spent outdoors during each person-day (as reported in the subject's activity diary) together with the maximum temperature reported for the day. The person-days in each database were partitioned into two groups: the person-days with daily maximum temperatures below a specified "critical temperature" and the remaining person-days associated with daily maximum temperatures that equaled or exceeded the critical temperature. Researchers conducted a series of statistical tests (Table 5-2) to determine the optimum critical temperature (OCT) for each monthly database (March and August). The OCT was defined as the temperature that partitioned the person-days into two groups that exhibited the greatest difference with respect to the distribution of minutes spent outdoors.

To determine the OCT, the critical temperature was varied and a "figure of merit" was determined for each trial temperature. The figure of merit was the p value that resulted from performing a particular two-sample statistical test on the null hypothesis that the two subgroups were from populations that had the same distribution or had the same measure of central tendency (e.g., mean). The trial temperature yielding the smallest p value for a particular test was designated the OCT. Five statistical tests were employed: analysis of variance (ANOVA), the rank sum test, the median test, the

Table 5-2. Results of Applying Five Statistical Tests to Grouped Data on the Number of Minutes Spent Per Day Outdoors as the Critical Temperature Used to Partition the Data is Varied (Johnson, 1988).

Month	Critical temperature, degrees F	p value of indicated statistical test				
		ANOVA	Rank sum	Median	van der Waerden	Savage
March	49	0.0075	0.0066	0.0061	0.0039	0.0037
	51	0.0044	0.0088	0.0035	0.0035	0.0035
	53	0.0021	0.0001	0.0001	0.0001	0.0003
	55	0.0007 ^a	0.0001 ^a	0.0001 ^a	0.0001 ^a	0.0001 ^a
	57	0.0060	0.0009	0.0051	0.0008	0.0016
	59	0.0150	0.0035	0.0136	0.0038	0.0051
August	79	0.0108	0.0011	0.0001	0.0027	0.0070
	81	0.0011	0.0001	0.0001	0.0001	0.0005
	83	0.0001	0.0001	0.0001	0.0001	0.0001
	84	0.0001 ^a	0.0001 ^a	0.0001 ^a	0.0001 ^a	0.0001 ^a
	85	0.0001	0.0001	0.0001	0.0001	0.0001
	87	0.0008	0.0146	0.0555	0.0090	0.0013
	89	0.0970	0.1936	0.2003	0.1789	0.1125

^aMinimum p value (determined by z value if p is less than or equal to 0.0001).

van der Waerden (VDW) test, and the Savage test. ANOVA assumes the data are normally distributed and tests that the means are equal. Researchers noted that this test was probably not appropriate as the data tested were highly skewed. The remaining four techniques test the null hypothesis that the distributions of rank scores are similar. Although these tests are nonparametric, each is most powerful against the distribution indicated below:

<u>Nonparametric test</u>	<u>Distribution</u>
rank-sum	logistic
median	double exponential
VDW	normal
Savage	exponential

Table 5-2 presents the results of applying the five tests to each of the two monthly databases. In tests of the March data, the critical temperature was varied from 49

degrees to 59 degrees in 2 degree increments. In tests of the August data, the critical temperature was varied from 79 degrees to 89 degrees in 2 degree increments with an additional set of tests at 84 degrees. For each test, a minimum p value (or maximum z value if p was less than or equal to 0.0001) was found within the range searched. The temperature associated with the minimum p value is the OCT for that particular test. Note that the tests were unanimous in selecting 55 degrees as the OCT for the March data and in selecting 84 degrees as the OCT for the August data.

In this analysis, the activity indicator being tested was “minutes spent per day outdoors.” The indicated OCT values were thus specific to this activity indicator. Johnson (1988) cautions that the OCT values may not be optimum for another activity indicator such as “minutes per day spent in vehicle.” Consequently it is important when using this methodology to select an activity indicator that exhibits a strong relationship with the exposure indicator of interest.

5.1.4 Factors Affecting Time/Activity Patterns Reported by Ten Studies

A number of researchers have analyzed time/activity data from individual diary studies to identify factors that significantly affect time spent in various activities and microenvironments. Few researchers have attempted to determine whether the time/activity data obtained from multiple studies present similar patterns for specific demographic groups. A report by Johnson, Capel, and McCoy (1995) describes an analysis of data from 10 time/activity studies that addressed both of these goals. In addition, the report provides useful statistics concerning the time subjects spent in microenvironments during periods when high exposures to CO or ozone were likely to have occurred.

The research effort, conducted during 1994, began with a survey to identify and evaluate all time/activity databases developed since 1980. Researchers identified the following 10 databases as promising candidates for an in-depth statistical analysis.

California - 11 and under	Los Angeles - construction workers
California - 12 and over	Los Angeles - elementary school
Cincinnati, Ohio	Los Angeles - high school
Denver, Colorado	Los Angeles - outdoor workers
Washington, DC	Valdez, Alaska

Eight of the databases relate to five individual urban areas; the remaining two databases relate to the entire state of California. All ten databases were obtained from studies that are summarized in Table 5-1.

A complete time/activity database was obtained for each of the 10 studies. These databases differed greatly with respect to format, documented time periods, data items, coding conventions, and other factors. To facilitate the statistical comparison of the 10 databases, researchers developed a common data format and coding scheme that could be applied to all 10 databases. The result was a set of ten simplified “person-hour” databases.

Each record in one of the person-hour databases pertained to a single person-hour of time/activity data. Each person-hour began on a clock hour (midnight, 1 a.m., 2 a.m., etc.) and ended 60 minutes later. To permit sorting by demographic

characteristics, each person-hour record included entries assigning the subject to a demographic group and to categories relating to gender, race, and family income. Each person-hour was also classified as to season (summer or winter), day type (weekday or weekend), and maximum temperature of the calendar day associated with the person-hour.

The person-hour databases associated with Cincinnati and Los Angeles were obtained from studies that used an activity diary originally developed for the Cincinnati study. The Cincinnati diary provides data on both the microenvironment and the breathing rate associated with each activity documented by a subject. Researchers found it useful to classify the Cincinnati-style data according to the following general categories for factors.

<u>Microenvironment</u>	<u>Breathing Rate Category</u>
Indoors - residence	Sleeping
Indoors - other	Slow
Outdoors - near road	Moderate
Outdoors - other	Fast
In vehicle	Any

Based on these categories, analysts defined a set of 25 distinct combinations of microenvironment and breathing rate. Each person-hour record associated with a study that used the Cincinnati diary included 25 values for MIN(m,b) and 25 values for ACT(m,b), where MIN(m,b) was defined as the number of minutes associated with microenvironment m and breathing rate b and ACT(m,b) equaled the number of distinct activities associated with the m,b combination.

Breathing rate data were not available for the remaining studies. Consequently, the person-hour databases developed for these studies included entries for MIN(m) and ACT(m) where each variable was classified as to microenvironment but not to breathing rate. The five microenvironments listed above were employed in developing each of the Denver, Washington, and Los Angeles person-hour databases. As the diary location codes used in the two California studies and in the Valdez study did not permit analysts to differentiate outdoor activities conducted near roads from other outdoor activities, these two categories were combined into a general outdoor category in these databases.

A set of 28 binary or “dummy” variables were defined to facilitate statistical analyses of the person-hour databases. The variables provided information on season, day of week, time of day, temperature, demographic characteristics (including gender, race, and income), work status, and other factors identified as likely predictors of time spent in microenvironments and breathing rate categories. The value of each variable was set equal to 1 when the indicated condition(s) were met and equal to zero in all other cases.

According to the mean MIN statistics, all databases indicated that the majority of time was spent in the two indoor microenvironments (indoors - residence and indoors - other). The total time spent in the two outdoor microenvironments was higher for the Los Angeles databases than for the other databases. This pattern was probably the result of (1) limiting the subjects of the Los Angeles studies to young people and outdoor workers and (2) omitting nighttime hours from the sampled periods in the Los

Angeles - construction worker study. The two studies with the most similar data collection protocols and study characteristics (Denver and Washington) produced almost identical results with respect to the mean values of MIN.

Consistent with expectations, the apportionment of time among microenvironments and breathing rates varied with time of day. The general pattern observed in the Cincinnati subjects was typical. Time spent in the indoors - other microenvironment peaked in the early afternoon. Time spent in the outdoor and in-vehicle microenvironments was highest between the hours of 3:00 p.m. and 8:00 p.m. With respect to breathing rate categories, subjects were most likely to report time at the moderate and fast levels during afternoon hours.

Analysts calculated "activity rates" as the average number of activities (all microenvironments) occurring per person-hour within eight specified time periods. According to this indicator, most of the databases showed less activity during nighttime hours, particularly the hours between midnight and 5 a.m, and more activity during daylight hours. The Cincinnati database exhibited a typical pattern for activity rate. The activity rate averaged 1.1 activities per person-hour for the midnight-to-5 a.m. time period, increased to 2.1 or 2.2 for the early morning and mid-morning time periods, rose to a value between 2.3 and 2.5 for the late morning and afternoon time periods, fell back to 2.1 for early evening, and then decreased to 1.6 for the late evening hours before midnight.

Comparison of the 10 databases with respect to activity rates suggested that the data collection protocol had a significant effect on the level of detail in the collected data. The combined use of observed and real-time Cincinnati diary in the Los Angeles construction study appears to have increased the level of detail (as quantified by the activity rate statistic) by a least a factor of four. The use of the real-time Cincinnati diary appears to have contributed to the relatively high levels of detail observed in the other Los Angeles databases and in the Cincinnati database. The use of a retrospective diary in the Valdez study may have contributed to the relatively low level of detail observed in that database.

A series of stepwise linear regression analyses were performed on the person-hour databases to identify factors affecting the time spent in microenvironments and breathing rate categories. In each analysis, the dependent variable was either MIN(m,b) or MIN(m). The set of independent (or explanatory) variables in each regression analysis included all or most of the 28 dummy variables.

The stepwise linear regression analyses were performed on seven person-hour databases representing individual studies (California - 11 and under, California - 12 and over, Cincinnati, Denver, Washington, Valdez, and Los Angeles - elementary school) and one aggregate database (Los Angeles - aggregate). The aggregate Los Angeles database was formed by combining the high school, outdoor worker, and construction worker databases.

The results of these regression analyses indicated that factors related to the major divisions of the data (particularly time periods associated with sleeping, work, and school), student status, gender, season, and day of the week were relatively good predictors of time/activity patterns. Factors related to temperature, school commuting time periods, general work status, and family income were relatively poor predictors. It should be noted, however, that these results were dependent on the set of candidate variables available for selection into each regression model. The set of 28 variables

defined for this analysis may not have adequately represented some of the proposed explanatory factors such as temperature. Note that the analysis of temperature described in Subsection 5.1.3 found that daily maximum temperature appeared to affect time spent outdoors.

Researchers performed a series of two-sample t tests to evaluate the effects of gender, race, and income on time spent in various combinations of microenvironment and breathing rate category. The results supported the following conclusions:

- Females, non-whites, and persons with incomes below \$10,000 tended to spend more time in the indoors - residence microenvironment.
- Whites and persons with incomes above \$10,000 tended to spend more time in the indoors - other microenvironment.
- Males tended to spend more time in the outdoors - other microenvironment and in the outdoors - combined microenvironment.
- Males tended to spend more time at fast breathing rates in the outdoor - other microenvironment.
- Males, whites, and persons with incomes above \$10,000 tended to spend more time in the in-vehicle microenvironment.

Note that these statements describe general patterns observed in the analyzed data and are not intended to indicate cause-and-effect relationships. Gender, race, and income are correlated with other explanatory factors such as employment status and car ownership which may better represent the actual causes of the patterns identified above.

The report by Johnson, Capel, and McCoy (1995) also included results of special analyses relevant to EPA's on-going assessment of exposure to CO and ozone. In CO exposure assessments, high exposures are more likely to occur during periods when ambient (outdoor) CO concentrations are high and a person is located in a motor vehicle or near a roadway. In a similar manner, high ozone exposures are more likely to occur during periods when ambient ozone concentrations are high and a person is located outdoors away from roadways. [Johnson (1995) and others have demonstrated that outdoor ozone concentrations near roadways and in vehicles tend to be lower than other outdoor locations because nitric oxides emitted by motor vehicles scavenge ozone.] As previously indicated, researchers used the designation "outdoors - other" for outdoor locations away from roadways.

Researchers selected Denver, Colorado, to represent an urban area with relatively high CO levels. CO levels tended to be higher in Denver during the cold-season months (October through March). Researchers analyzed Denver monitoring data for the cold season and found that (1) weekday CO concentrations were high during morning and afternoon commute periods and (2) weekend CO concentrations were high during the early evening hours. The report provides MIN(m) statistics by database for these time periods for two microenvironments: vehicles and outdoors - near road. These statistics can be used to evaluate the effects of using different

time/activity databases to represent the activities of Denver residents in CO exposure assessments such as those performed with Version 2.1 of pNEM/CO.

Los Angeles was selected to represent an urban area with relatively high ozone concentrations. As in most urban areas, ozone concentrations in Los Angeles tended to be higher during the warm season of April through September. Researchers analyzed Los Angeles monitoring data for the warm season and found that ozone concentrations were high during afternoon hours on both weekdays and weekends. The report provides MIN(m) statistics for these time periods for two microenvironments: outdoors - other and outdoors - combined. Analysts can use these statistics to evaluate the effects of using different time/activity databases to represent the activities of Los Angeles residents in ozone exposure assessments performed using pNEM/O3 and similar models.

5.1.5 Use of Random Walk Techniques to Construct 24-Hour Activity Patterns

During the development of the NEM model for ozone in 1987, researchers considered “random walk” techniques as a possible means of constructing the 24-hour activity patterns required by the model. This investigation culminated in a report by Johnson, Wijnberg, and Mersch (1987) describing the Activity Pattern Simulator (APS). This subsection provides a brief summary of this report.

The APS was designed to use data obtained from field studies in which subjects documented their activities over 24-hour periods in specially-prepared diaries. The activity diary entries for a particular person-day could be characterized as a series of transition from exposure state s_i to a different exposure state, s_j . Let the time of the k th transition be denoted $t(k)$. The time period between the k and $k+1$ transitions will be referred to as the occupancy period (OP). The duration in minutes of the OP starting with the k th transition will be denoted as $d(k)$. Thus

$$d(k) = t(k+1) - t(k). \tag{5-1}$$

At the k th transition, a person enters the $s(k)$ exposure state and departs from the $s(k - 1)$ exposure state. If three exposure states (A, B, and C) are defined, there are six possible transitions:

<u>s(k-1)</u>	<u>s(k)</u>
A	B
A	C
B	A
B	C
C	A
C	B

There are two possible $s(k)$ outcomes for each of three $s(k-1)$ possibilities. We can define $P[s(k) = j | s(k-1) = i]$ as the probability that state j will be entered given a person is departing from state i . We can then add time dependency and define

$$P[s(k) = j | s(k-1) = i \text{ and } t_a \leq t(k) \leq t_b]$$

as the probability that state j will be entered given departure from state i in the period between t_a and t_b (e.g., noon and 1:00 p.m.). We can go one step further and define

$$P[s(k) = j | s(k-1) = i \text{ and } d(k) = h | s(k-1) \text{ and } t_a \leq t(k) \leq t_b]$$

as the probability that state j will be entered and that the OP will last h minutes given $s(k-1) = i$ and $t_a \leq t(k) \leq t_b$. We will refer to this as an event probability.

Consider a subject currently in exposure state A who is making a transition at 11:15 a.m. We wish to predict the state being entered and its duration. Assume that activity diary data have been used to determine the event probabilities listed in Table 5-3 for the 11:00 a.m. to 11:59 a.m. time period, given departure from state A. A random number generator (uniform distribution from 0 to 1) can be used to select one of 20 possible events based on the associated cumulative probabilities. Assume the number 0.33 is generated. Since $0.296 \leq 0.33 \leq 0.371$, event 6 [$s(k) = B$ and $d(k) = 15$] is selected.

A table of event probabilities similar to Table 5-3 can be determined for each hour of the day for each possible $s(k-1)$ state and assembled into a matrix. For more sophisticated applications, the matrix could provide event probabilities indexed by day type (weekday, Saturday, Sunday), season, and demographic group as well as hour of day. The analyst would also need to provide a table listing the probabilities associated with initial conditions (i.e., the exposure state at midnight). Figure 1 of the report by Johnson, Wijnberg, and Mersch (1987) presents the algorithm employed by APS, a pattern sequencer program capable of generating 24-hour activity patterns based on input data of this kind. Briefly, the algorithm begins by randomly selecting the initial state and duration using the probabilities listed in the initial probability table. If the duration of the event does not exceed 1440 minutes (24 hours), the identity of the next event is determined using the event probability table listings for the hour that brackets the start time of the event. If the combined durations of the first and second events do not exceed 1440 minutes, a third event is determined and so on. When the combined durations of all selected events equal or exceed 1440 minutes, no further events are selected and the activity pattern is terminated at minute 1440.

The APS concept was tested by using it to construct simplified activity patterns appropriate for the estimating ozone exposure. Initially researchers defined nine exposure states based on the possible combinations of generalized microenvironment (vehicle-affected, other indoors, or other outdoors) and breathing rate (slow, medium, or fast). The nine states were subsequently aggregated into the following five exposure states.

<u>Code</u>	<u>Exposure state</u>
1	Vehicle-affected — low or medium breathing rate
2	Vehicle-affected — fast breathing rate
3	Other indoors — breathing rate not specified
4	Other outdoors — low or medium breathing rate
5	Other outdoors — fast breathing rate

The event probabilities were obtained from an analysis of a time/activity database provided by the Cincinnati Activity Diary Study (Johnson, 1987). This database was

Table 5-3. Hypothetical Event Probabilities for Transitions from Exposure State A between 11:00 a.m. and 11:59 a.m.

Event	State entered	Duration, min	Event probability	Cumulative probability
1	B	2	0.042	0.042
2	B	5	0.093	0.135
3	B	7	0.053	0.188
4	B	10	0.087	0.275
5	B	13	0.021	0.296
6	B	15	0.075	0.371
7	B	25	0.056	0.427
8	C	6	0.033	0.460
9	C	9	0.036	0.496
10	C	20	0.083	0.579
11	C	33	0.041	0.620
12	C	42	0.045	0.665
13	C	64	0.043	0.708
14	C	90	0.097	0.805
15	C	118	0.020	0.825
16	C	172	0.011	0.836
17	C	243	0.028	0.864
18	C	291	0.037	0.901
19	C	382	0.044	0.945
20	C	ROD ^a	0.055	1.000

^aRemainder of day (i.e., duration is number of minutes remaining in day).

selected because it was the largest database available in 1987 that contained information on both the subject's microenvironmental location and his or her breathing rate (characterized as low, medium, or fast). Subsections 4.3 and 4.4 of the report by Johnson, Wijnberg, and Mersch (1987) describe the method used to construct an exposure probability matrix based on the Cincinnati time/activity data. The event probabilities within this matrix were indexed by hour of day, day type, season, and demographic group (19 classifications).

If the assumptions incorporated into APS are valid, the distribution of a particular characteristic in the activity patterns generated by APS for a particular population group should be similar to the distribution of that characteristic in the activity diaries obtained from members of that group, i.e., the diaries which served as input data to APS. Researchers selected two characteristics for testing purposes: the number of events in an activity pattern and the partitioning of minutes in the activity pattern among exposure states. To perform the tests, researchers used APS to generate 100 weekday, 100 Saturday, and 100 Sunday patterns based on 68 weekday, 15 Saturday, and 14 Sunday patterns obtained from Cincinnati subject during March 1985. The subjects were all members of "Demographic Group No. 7" (age 18 years or older, employed male, indoor work morning shift, commute 0 to 29 minutes). Population weights were not used in the simulation, so that the resulting APS patterns could be compared directly to the diary patterns.

The results of various statistical tests suggested that the distribution of number of events in APS-generated patterns compared favorably with the corresponding diary-based distribution for weekdays and Saturdays. The Sunday distributions compared favorably in the middle and upper percentiles but diverged in the lower percentiles. [The likely cause of this particular discrepancy was related to inadequate sample size; it is discussed in detail by Johnson, Wijnberg, and Mersch (1987).]

Each activity pattern, whether simulated or diary-based, contained 1440 minutes which were partitioned among the five exposure states listed above. The results of the statistical tests suggested that there was generally good agreement between simulated and diary-based patterns with respect to this partitioning of time.

The favorable results of these tests do not ensure that simulated patterns will be similar to activity patterns in all important respects. Researchers noted that the simulated patterns frequently exhibited a "disconnect" between morning and afternoon events. For example, morning and afternoon commute times often differed significantly, whereas one would expect that morning and afternoon commute times for a particular individual would be roughly equal. This phenomenon was expected, as the transition probabilities relevant to changing states were specific to the current state rather than to states that occurred hours earlier. To minimize the effect, the algorithm would need to be modified so that transition probabilities were affected by both current and previous states. Researchers elected not to pursue this alternative, choosing instead to construct EES using "observed" 24-hour activity patterns as discussed in Subsection 5.1.2.

5.2 RESIDENTIAL OCCUPANCY PERIODS

5.2.1 Introduction

EPA regulates the emissions of benzene and other toxic pollutants emitted by mobile sources because of their potential for causing cancer in exposed populations. When performing risk assessments for a known or suspected carcinogen, analysts typically consider each person's cumulative lifetime exposure to the pollutant. A reasonable approach to estimating lifetime exposure is to divide each person's life into a series of residential occupancy periods (ROPs), where each ROP is the period between the date that the person moves into a new residence and the date the person dies or moves out of the residence. The analyst estimates an exposure for each ROP separately and then sums the exposures to produce the lifetime exposure estimate.

In many exposure assessments, ROP is assumed to be 70 years, a period roughly equal to the average lifetime. Strictly speaking, this value should be considered an upper bound for ROP, as several studies have indicated that the average ROP of the general population is less than 20 years (U.S. EPA, 1989).

A value of 30 years has been suggested previously as the 90th percentile of ROP. This value is based on a survey of households conducted by the Bureau of Census (BOC) in 1983. In this survey, a household member was asked to provide the year which the household moved into their present residence (BOC, 1983). An analysis of the responses for owner-occupied housing yielded a 50th percentile of 9.37 years and a 90th percentile of 29.84 years (U.S. EPA, 1989). These estimates are not entirely appropriate as indicators of the distribution of ROP in the general population. In particular, these estimates (1) refer to elapsed time in the residence rather than the total occupancy period, (2) omit persons living in rental housing, (3) apply to entire households rather than individuals, and (4) are based on data which may be out of date.

To provide more representative estimates of residential duration, researchers developed and implemented a methodology capable of estimating the distribution of ROP in the national population. The resulting methodology employed a Monte Carlo approach to simulate a distribution of ROP values using readily available data on mobility and mortality. A report by Johnson and Capel (1992) provides a description of the methodology, the results of applying the methodology to the national population through the use of a computer-based algorithm (ROPSIM), and a discussion of various assumptions employed in the methodology and how they may bias the simulation.

The ROPSIM program was run for 500,000 iterations using values for population, mobility, and mortality appropriate for the general U.S. population. The program produced tables listing detailed descriptive statistics for ROP by gender and current age that can be found in the report by Johnson and Capel (1992). Table 5-4 provides selected descriptive statistics taken from this run for males and females combined. Results are provided by age in six-year intervals. The last row of the table provides results for all ages combined. The estimated median ROP for the entire U.S. population is 9 years; the 90th percentile ROP is 26 years.

Price, Sample, and Strieter (1992) have used a similar Monte Carlo approach to determine the fraction of one's life spent in the current residence. The Price methodology has been incorporated into a computer model for estimating lifetime

Table 5-4. Descriptive Statistics for Residential Occupancy Period (Males and Females Combined by Current Age).

Current age, years	Cumulative percentage of simulated persons	Residential occupancy period, years							Percentage of group who die in residence
		Mean	Percentile						
			25	50	75	90	95	99	
6	10.47	8.0	4	7	10	15	18	22	0.29
12	18.79	9.3	5	9	13	16	18	23	0.18
18	27.46	8.2	4	7	11	16	19	23	0.31
24	37.23	5.2	2	4	6	11	15	25	0.52
30	48.11	7.3	3	6	9	14	19	32	1.26
36	58.21	10.4	5	8	13	21	28	47	3.98
42	66.73	13.5	6	11	18	27	35	49	8.20
48	73.27	16.6	8	14	22	32	39	52	14.80
54	78.76	18.3	9	16	25	34	40	50	24.56
60	84.25	19.7	11	18	27	35	40	51	35.53
66	89.52	20.7	12	20	28	36	41	50	48.12
72	93.87	21.6	13	20	29	37	43	53	60.57
78	97.00	21.4	12	19	29	38	44	53	65.96
84	98.86	20.3	11	19	28	37	44	56	76.41
90	99.69	18.9	8	15	27	40	47	56	83.11

All ages	100.00	11.7	4	9	16	26	33	4
----------	--------	------	---	---	----	----	----	---

exposures to hazardous pollutants and associated cancer risks (Price, Young, and Chaisson, 2001).

5.2.2 Overview of the Methodology

The general methodology developed by Johnson and Capel (1992) can be described as a six-step process. Steps 1 through 5 consist of tasks in which the analyst defines the population being modeled and develops appropriate data bases to represent this population. These data bases are used to generate a distribution of ROP values in Step 6.

In Step 1, the analyst defines a specific population of interest. Examples include the general U.S. population, all blacks living in the Southeast, all women of child-bearing age living in urban areas, and all persons living within 20 kilometers of a particular point source.

In Step 2, the analyst develops a comprehensive set of demographic groups such that each member of the specified population can be assigned to one and only one demographic group. The demographic groups should be defined to make optimum use of the data on population, mobility, and mortality available for the specific population of interest. The demographic group definitions should always include current age as one of the descriptive factors. Other descriptive factors which may be appropriate in a particular analysis include gender, race, location (urban vs. rural), and income.

In Step 3, the analyst estimates the fraction of the specified population that falls into each demographic group. If the demographic groups are defined according to descriptors used by the BOC (e.g., age and gender), then these estimates can be developed directly from census data.

In Step 4, the analyst develops a mobility table for the population of interest. The mobility table provides listings by age of the probability that a person with specific demographic characteristics did not move during the previous year. Data for this table can be obtained from the P-20 series of reports on geographic mobility which have been published by the BOC since 1981. These reports provide mobility data organized by age, gender, race, region, level of urbanization, status within household (head of household, other family member, etc.), presence of children, and a variety of other factors.

In Step 5, a mortality table is developed for the population. The mortality table provides listings by age of the probability that a person with specified demographic characteristics will die during the upcoming year. Data for this table can be obtained from various reports on vital statistics published by the National Center for Health Statistics.

In Step 6, a computer-based algorithm is used to apply a Monte Carlo process to a series of persons selected at random from the population being analyzed. The process begins by randomly assigning each person to one of the demographic groups defined in Step 2 according to the population fractions developed in Step 3. The process then simulates a residential history for this person using data on mobility and mortality developed in Steps 4 and 5 which are consistent with the person's demographic group.

The residential history specifies when the person moved into his or her current

residence and when the person will die or move out of the current residence. Thus the residential history determines the ROP of the person with respect to his or her current residence.

The process in Step 6 is repeated for n persons, where n is determined by the analyst. As n becomes large, the distribution of simulated ROP values is assumed to approach the actual distribution of ROP in the population of interest.

The next section describes how this six-step process was used to develop a distribution of ROP for the general U.S. population. It also provides a more detailed description of the algorithm employed in Step 6.

5.2.3 Application of Methodology to the General U.S. Population

Definition of Population-of-Interest. The population-of-interest was defined as the U.S. population as of July, 1987. This year was selected because it was the most recent year for which data on population, mobility, and mortality were simultaneously available. (The example analysis was conducted in 1992).

Definition of Demographic Groups. A review of available data suggested that gender (male/female) and age (in years) were good choices for defining demographic groups with respect to the general U.S. population. As census data (BOC, 1990) indicate that less 0.03 percent of the U.S. population exceeds 100 years of age, the age categories were limited to 0 through 100. A total of 202 demographic groups were defined according to this approach.

Estimation of Demographic Group Populations. Estimates of the number of persons within each group were developed from data provided in a report by the BOC (1990). Population values at one-year intervals for ages 0 to 64 were taken directly from a table in the report listing population estimates for July 1987. As only five-year intervals were provided in this table for ages above 64, values for one-year intervals above 64 were estimated through an interpolation procedure. These estimates were developed by (1) converting the population values for each five-year interval into cumulative population values, (2) fitting a cubic spline to the five-year cumulative values, (3) using the spline to estimate the cumulative values at one-year intervals, and converting the one-year cumulative values to one-year interval values. The report by Johnson and Capel (1992) provides the estimated demographic group populations for ages 0 through 100.

Development of Mobility Table. Let $PSR(i,j)$ indicate the probability that a person of gender i and age j was in the same residence one year earlier; that is, the person did not move during the preceding year. The value of $PSR(i,j)$ can be estimated by the equation

$$PSR(i,j) = POPSR(i,j)/POPTOT(i,j) \quad (5-2)$$

where $POPSR(i,j)$ is the number of persons of gender i at age j who indicated that they occupied the same residence one year earlier and $POPTOT(i,j)$ is the total number of persons of gender i at age j.

At the time of this example analysis, data on $POPSR$ and $POPTOT$ were available from the BOC for the years 1981 through 1987. The BOC conducted surveys

in these years in which each respondent was asked where he or she resided one year earlier. Permissible answers included:

- Same residence
- Different residence - same county
- Different residence - other county - same state
- Different residence - other state - contiguous U.S.
- Different residence - other state - noncontiguous U.S.
- Different residence - other country.

The BOC extrapolated the survey responses to the general U.S. population.

The resulting mobility statistics for 1987 are presented in BOC (1989) as a series of tables organized by various demographic categories. Under the categories “male” and “female”, data are tabulated by age in single year intervals from 1 to 74. Totals for the age range “75 years and over” are also provided.

The ratio of males reporting “same residence” in a particular age interval to the total number of males in the age interval was used as an estimate of PSR(i,j) for the interval. PSR(i,j) values for females are estimated in a similar manner. For each gender, a single value of PSR(i,j) is used for ages 75+. The report by Johnson and Capel (1992) lists the PSR(i,j) values used in the analysis. Values increase from birth to the late teens and then fall to a minimum in the middle twenties. The values then increase again to a plateau which starts in the middle 60's.

Development of Mortality Table. Let PD(i,j) indicate the probability that a person of gender i and age j will die during the next year. The value of PD(i,j) can be estimated from the equation

$$PD(i,j) = [POPA(i,j) - POPS(i,j)]/POPA(i,j) \quad (5-3)$$

where POPA(i,j) is the number of persons of gender i and age j alive at the beginning of the current year and POPS(i,j) is the number of persons who survive to the beginning of the next year. Values of POPA and POPS can be obtained from “current life” tables. These tables, prepared by the National Center for Health Statistics (1990), consider a hypothetical cohort containing 100,000 persons which is subject throughout its lifetime to the death rates prevailing in a specified year.

To determine PD(i,j), the value in the life table for age j was substituted for POPA(i,j), and the value in the table for j + 1 was substituted for POPS(i,j). This procedure produced gender-specific estimates of PD(i,j) for ages 0 through 84.

Another procedure was used to determine PD(i,j) for ages 85 and above. In this case, PD(i,j) was estimated by the equation

$$PD(i,j) = POPD(i,j)/POP(i,j) \quad (5-4)$$

where POP(i,j) is the number of persons of gender i at age j and POPD(i,j) is the number of people of gender i at age j who die within the next year. Estimates of POP(i,j) for ages 85 to 100 were developed by Johnson and Capel (1992) using the cubic spline procedure described above. Estimates of POPD(i,j) for 1987 were taken from a general mortality table prepared by the National Center for Health Statistics

(1990). For simplicity, a 100 percent death rate was assumed for age 99. As previously indicated, less than 0.03 percent of the population lives to be older than 100. The report by Johnson and Capel (1992) lists the values of $PD(i,j)$ estimated by the two procedures for males and females. In general, the death rate for males at a given age is higher than the corresponding death rate for females.

Implementation of the Monte Carlo Algorithm. The Monte Carlo algorithm consists of the following four-step process. The process is repeated for each person selected from the population-of-interest.

In Step 1, the person's gender is randomly determined with the probability proportional to the male/female breakdown of the defined population. In 1987, 48.8 percent of the general U.S. population was male (BOC, 1990).

In Step 2, the person's current age in years is randomly determined according to the distribution of ages in the defined population for the gender determined in Step 1.

Let E equal the elapsed time already spent by a person in his or her current residence, and let F equal the future time that the person will spend in the current residence. A value for E is generated in Step 3 through the use of the following iterative process. A randomly-generated number between zero and one is first compared with $PSR(i,j)$, where $PSR(i,j)$ is determined by the person's gender and current age. If the random number is less than $PSR(i,j)$, the person is assumed to have lived in current residence for the entire preceding year. The process is then repeated using a $PSR(i,j)$ values consistent with the person's age one year earlier. The process ends when the person is found to have moved during the preceding year or reaches the age of zero. The elapsed time is E years, where E is the number of years going backward in which the person did not move during the preceding year.

The value of F , the future time spent in the residence, is determined in Step 4 by another iterative process. A randomly-generated number between zero and one is first compared with $PD(i,j)$, where $PD(i,j)$ is the probability that a person of gender i at age j will die during the next year. If the random number is less than $PD(i,j)$, the person is assumed to die during the next year. The process stops and F is set equal to zero.

If the random number is larger than $PD(i,j)$, the person is assumed to survive for one more year. The age of this person at the end of this year is $j + 1$. Another random number is generated and compared with $PSR(i,j+1)$, the probability that someone of gender i and age $j + 1$ will have moved during the preceding year. If the random number is less than $PSR(i,j+1)$, then the person is assumed to have lived in his current residence for this entire year. The process is then completed using values of $PD(i,j)$ and $PSR(i,j)$ consistent with the person's age during each future year. The process ends when the person is found to have died or moved during a particular year. The parameter F is set equal to the number of future years spent in the same residence.

The value of ROP for the simulated person is calculated by the equation

$$\text{Residential occupancy period} = E + F + 1. \quad (5-5)$$

The extra year added to the sum of E and F compensates for the fact that E tends to underestimate the actual elapsed time spent in the residence and F tends to underestimate the actual future time in the residence. For example, a value of E equal to 6 years indicates that the actual elapsed time falls between 6 and 7 years. Similarly, a value of 4 for F indicates that the actual future time falls between 4 and 5 years. In

this case, the sum of the actual elapsed time and the actual future time falls somewhere between 10 and 12 years. A reasonable “best” estimate would be 11 years, the value of $E + F + 1$.

Table 5-5 presents a sample simulation for a male, aged 14, which yields a value of E of 6 years and a value of F of 4 years. Based on Equation 5-5, the estimated ROP for this person is $6 + 4 + 1 = 11$ years.

5.2.4 The ROPSIM Program

A computer program (ROPSIM) was developed which implements the Monte Carlo algorithm described above. The ROPSIM user provides three input data sets:

- (1) the number of persons in the population of interest by gender and age,
- (2) the probability of moving during the preceding year by gender and age, and
- (3) the probability of dying during the coming year by gender and age.

The user also specifies the number of persons to be considered in the simulation.

Based on these inputs, the ROPSIM program produces a series of tables indicating the distribution of ROP for various population groups. Each table lists ROP values in one year intervals from zero to 100 years and indicates the number of persons assigned each value in the simulation. Mean values and cumulative percentages are also provided.

In addition to the ROP statistics, the program provides the distribution of current age (i.e., the age used to start each simulation). The program also provides the distribution of age at death for all simulated persons who die while residing at the current residence. Results are provided for the following population groups:

- Both genders, all ages (one group)
- Males, all ages (one group)
- Females, all ages (one group)
- Both genders, current age = 3, 6, 9, ..., 90 (30 groups)
- Males, current age = 3, 6, 9, ..., 90 (30 groups)
- Females, current age = 3, 6, 9, ..., 90 (30 groups).

The program can be easily modified to tabulate results for other groups defined by gender and current age.

The ROPSIM program was run for 500,000 persons using the values for population, mobility, and mortality developed by Johnson and Capel (1992) for the general population. Table 5-6 presents descriptive statistics concerning the distribution of ROP for various population groups obtained from this run. More detailed statistics can be found in the report by Johnson and Capel (1992).

Table 5-5. Sample Simulation for a Male Aged 14 years Which Yields a Residential Occupancy period of 11 Years

Elapsed time simulation:				
Calendar year	Age	Random number	PSR	Moved during preceding year?
1987	14	0.738	0.831	No
1986	13	0.478	0.853	No
1985	12	0.139	0.834	No
1984	11	0.417	0.830	No
1983	10	0.588	0.823	No
1982	9	0.783	0.834	No
1981	8	0.849	0.828	Yes
Elapsed time = 6 years				

Future time simulation:						
Calendar year	Age	Random number	PD	Survive another year?	PSR	Moved during preceding year?
1987	14	0.995	0.00062	Yes		
1988	15	0.462 0.367	0.00083	Yes	0.844	No
1989	16	0.340 0.840	0.00105	Yes	0.860	No
1990	17	0.245 0.137	0.00123	Yes	0.865	No
1991	18	0.507 0.220	0.00137	Yes	0.826	No
1992	19	0.899			0.761	Yes
Future time = 4 years						

Residential occupancy period = elapsed time + future time + 1 = 11 years

Table 5-6. Descriptive Statistics for Residential Occupancy Period and Age of Death

Statistic	Value of statistic		
	Both genders n = 500,00	Males only n = 244,274	Females only n = 255,726
Residential occupancy period, years			
Mean	11.7	11.1	12.3
5 th percentile	2	2	2
10 th percentile	2	2	2
25 th percentile	4	4	5
50 th percentile	9	8	9
75 th percentile	16	15	17
90 th percentile	26	24	28
95 th percentile	33	31	35
98 th percentile	41	39	43
99 th percentile	47	44	49
99.5 th percentile	51	48	53
99.8 th percentile	55	53	58
99.9 th percentile	59	56	61
Second largest value	75	73	75
Largest value	87	73	87
Age at death, years ^a			
Mean	77.2	74.9	79.7
25 th percentile	71	69	73
50 th percentile	80	77	82
75 th percentile	86	84	89
90 th percentile	92	90	93
95 th percentile	95	92	96
Percentage of simulated persons who die in current residence	13.5	13.3	13.4

^aIncludes only deaths which occur while occupying current residence.

5.2.5 Assumptions and Possible Biases of the Methodology

The ROP-estimation methodology developed by Johnson and Capel (1992) was designed to make use of the most detailed mobility and mortality data available at the time of the analysis (1992). As described above, these data were used to estimate the one-year probabilities of dying (PD) and of not moving (PSR) as a function of gender and age. These probabilities were in turn used in a Monte Carlo process which simulated the occurrence of these events over multiyear periods.

It is important to note that PD and PSR are functions only of gender and age. In particular, the Monte Carlo process assumes that these probabilities are independent of (1) the calendar year to which they are applied and (2) the past history of the person being simulated. Neither of these assumptions is entirely correct.

The first assumption concerns the applicability of a single set of PD and PSR values to all calendar years included in the simulation. The values of PSR and PD used in the sample application were determined from 1987 data. While they are probably representative of mobility and mortality rates of the 1980's, they are unlikely to be representative of rates in effect during earlier decades. Moreover, it is difficult to predict the applicability of these rates to future decades. These uncertainties may produce a bias in the distribution of simulated occupancy periods with respect to the unknown "true" distribution of occupancy periods associated with the general U.S. population of 1987. This bias would be most evident in the upper tail of the distribution, where occupancy periods include calendar years far removed from 1987. The direction and magnitude of this bias is not known.

The second assumption concerns the applicability of a single set of PD and PSR values to persons with varying histories of mobility. In the Monte Carlo procedure, a man aged 55 who has lived in his current residence for 30 years is given the same probability of not moving as a man of the same age who has lived in his current residence for only 5 years. This probability is based on the average behavior of 55-year-old males as determined by a survey questionnaire administered by the BOC to a sample of persons in this age group.

Intuitively, one would expect that a person who has lived in his current residence for an extended period of time is more tied to his home than another person who has lived in his current residence for a relatively brief period of time. The former person is less likely to move in the coming year; the latter person is more likely to move. If this pattern holds for the general population, then the Monte Carlo process described above will tend to underestimate the occurrence of very large occupancy periods. The degree of this bias is difficult to characterize, as there are no known data bases available which relate a person's future plans to move to his or her past moving history.

The reader should also note that the ROP statistics presented in Table 5-6 are based on a single run of the ROPSIM program. Because ROPSIM employs a Monte Carlo approach to estimate occupancy period, each run of the program may produce results that differ slightly. These differences will be most evident in the extreme upper tails of the simulated distributions.

Johnson and Capel (1992) made five runs of ROPSIM ($n = 500,000$) for the population group labeled "both genders, all ages." The same 99.9th percentile ROP value (59 years) was estimated by each of the five runs. The values determined for the largest ROP varied from 79 to 92 years.

5.3 COMMUTING PATTERNS

Historically, the NEM/HAPEM series of exposure models have accounted for the commuting trips of various population groups between home and work locations. Recently, efforts have begun to enhance the models to account for commuting trips of students between home and school. In both cases, the commuting patterns are represented by means of an origin-destination table that shows the number of people associated with each possible combination of origin location (e.g., home census tract) and destination location (e.g., work place census tract). Subsection 5.3.1 presents methods that have been used to develop origin-destination tables for home-work trips. Subsection 5.3.2 presents “tentative” methods that have been proposed for developing origin-destination tables for home-school trips. (As of December 2001, none of the NEM/HAPEM series of models had been enhanced to account for home-school trips.)

5.3.1 Home-to-Work

A report by Johnson, Mihlan, LaPointe, et al. (2000) describes the method researchers used in developing home-to-work trip tables for the application of Version 2.1 of pNEM/CO to Denver and Los Angeles. This method, which makes direct use of a special commuting database developed by the Bureau of Census (1994), supercedes an earlier indirect estimation method developed by Johnson, Capel, and Byrne (1991). Subsection 5.3.1.1 describes the special commuting database. Subsection 5.3.1.2 presents the method used to aggregate census tracts into six “home” districts and seven “work” districts. Subsection 5.3.1.3 describes the method used to develop an origin-destination table for this set of home and work districts. Data quality issues related to the use of BOC data in developing origin-destination tables are discussed in Subsection 5.3.1.4.

5.3.1.1 The Census Commuting Database (1990)

Table 5-7 presents the format of the special commuting database, referred to as “STP154: Census Tract of Work by Census Tract of Residence” by the BOC (1994). The following two data items were used in the analyses which follow.

- WF: weighted count of workers in specified flow.
- WA: weighted count of workers allocated to this tract of work.

WF is the total number of workers estimated by BOC to commute (“flow”) from the indicated residence census tract to the indicated work census tract. Each value of WF is a weighted estimate based on responses to the 1990 census “long form”. The estimate includes both (1) people who explicitly indicated they lived and worked in the indicated census tracts and (2) other people who the BOC allocated to the flow based on supplemental information. WA indicates only the number of people who fall into the second category. As WA increases as a percentage of WF, confidence in the value of WF is reduced.

To develop an origin-destination table for a particular study area, analysts first

Table 5-7. Format of Commuting Database Obtained from the Bureau of Census.

Bytes	Data item	Explanation of codes
1 - 2	FIPS state code of <u>place of work</u>	01 - 56 = Alabama - Wyoming 99 = did not work in U.S.
3 - 5	FIPS county code of <u>place of work</u>	000 = did not work in U.S. NNN = county code
6 - 11	Census tract/BNA code of <u>place of work</u>	000000 = did not work in U.S. 000100...999999 = legal tract code range
12 - 14	Census MCD code of <u>place of work</u>	000 = did not work in New England state NNN = MCD code in New England (FIPS State = 09, 23, 25, 33, 44, 50)
15 - 18	Census place code of <u>place of work</u>	0000 = did not work in U.S. 0001...9998 = census place code 9999 = did not work in a place
19	Blank	
20 - 21	FIPS state code of <u>residence</u>	01 - 56 = Alabama - Wyoming 99 = did not work in U.S.
22 - 24	FIPS county code of <u>residence</u>	NNN = county code
25 - 30	Census tract/BNA code of <u>residence</u>	000100...999999 = legal tract code range
31 - 33	Census MCD code of place of <u>residence</u>	000 = did not live in New England state NNN = MCD code in New England (FIPS State = 09, 23, 25, 33, 44, 50)
34 - 37	Census place code of <u>residence</u>	0001...9998 = census place code 9999 = did not live in a place
38	Blank	
39 - 44	Weighted count of workers in this flow, including all place of work allocation (everyone forced into a tract of work)	
45 - 50	Weighted count of workers allocated to this tract of work	
51 - 56	Weighted count of workers allocated to this place of work	
57 - 62	Weighted count of workers allocated to this county or MCD (in 6 New England states) of work	

identify the census tracts included within each exposure district defined within the study area. Next, the flows among the individual census tracts in each district are combined to determine flows among the districts. Each district-to-district flow originating in a particular district is then converted to a fraction of the total workers commuting from the district. These fractions are organized by home and work districts to create the required origin-destination table. The following subsections describe the application of this approach to the exposure districts defined for the application of Version 2.1 of pNEM/CO to Denver. The report by Johnson, Mihlan, LaPointe, et al. (2000) provides a similar description for the application of pNEM/CO to Los Angeles.

5.3.1.2 Denver Exposure Districts

As described in Subsection 2.1, analysts defined six exposure districts for application of the Version 2.1 of pNEM/CO to the Denver study area. Five of the districts are centered on individual fixed-site CO monitors in the Denver metropolitan area. The sixth district is centered on a point midway between two monitors located in Boulder, CO. Figure 5-1 shows the locations of these seven fixed-site monitors. Table 5-8 identifies the monitor(s) associated with each district.

Each district was defined as a collection of census tracts as delineated by the BOC for the 1990 decennial census. The census tracts were drawn from a comprehensive listing of the census tracts located in the following nine counties:

Adams	Clear Creek	Gilpin
Arapahoe	Denver	Jefferson
Boulder	Douglas	Weld.

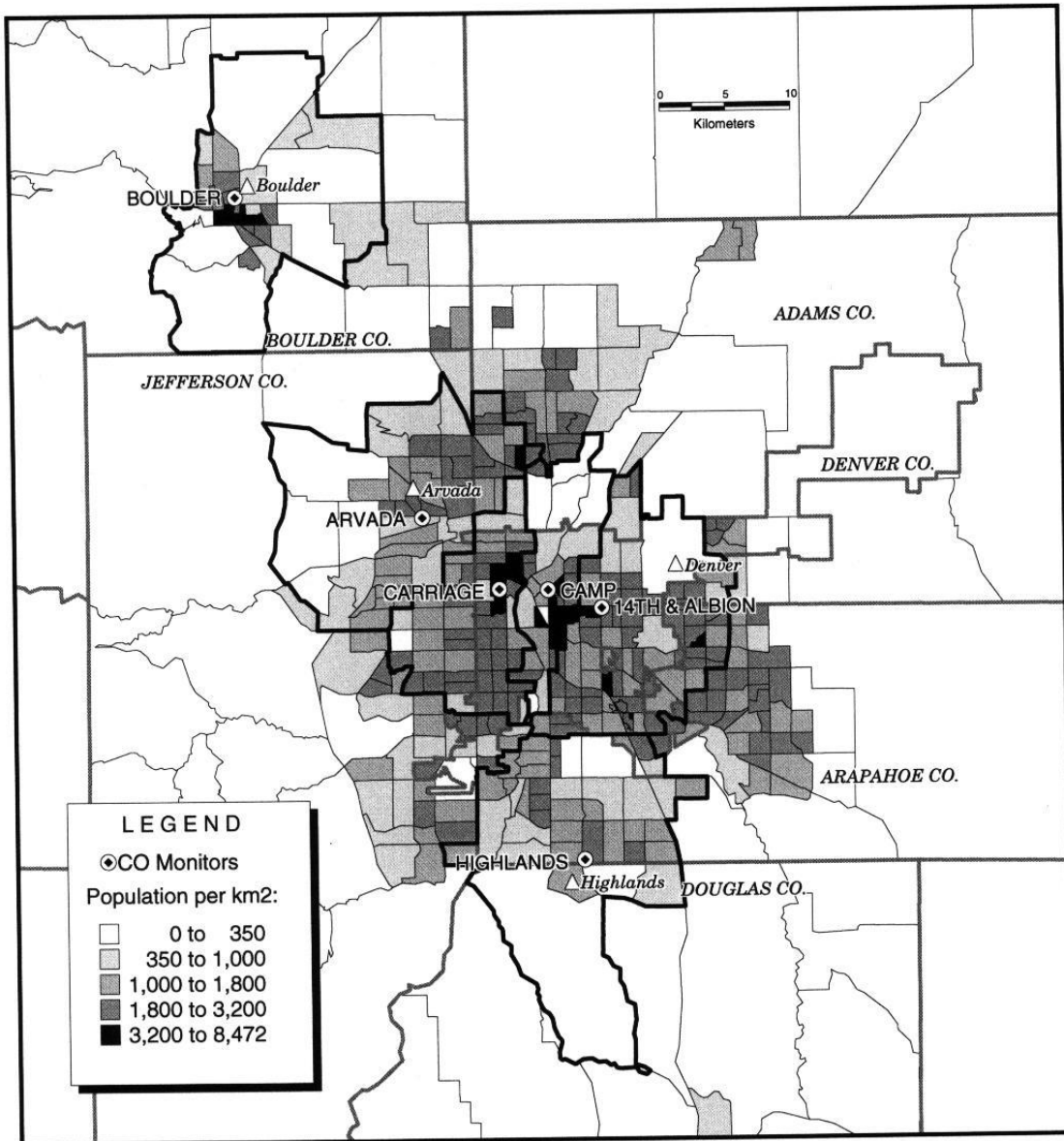
In developing the districts, analysts assigned each census tract located within 10 km of one or more district centers to the nearest district. People residing in the remaining census tracts were excluded from the pNEM/CO analysis, based on the assumption that pNEM/CO could not accurately estimate the exposures of people who lived more than 10 km from a monitoring station.

Table 5-8 lists the number of census tracts assigned to each of the six exposure districts and the number of remaining census tracts (designated “District 19”). Among the home districts (1 through 6), the number of census tracts contained within a district ranges from 28 (No. 6: Boulder) to 116 (No. 3: 14th and Albion). The six home districts account for 351 (59 percent) of the 593 census tracts in the nine-county area.

5.3.1.3 Origin-Destination Table for Denver

Analysts defined six “home” exposure districts and seven “work” exposure districts. The six home districts were the six monitor-derived exposure districts listed in Table 5-8. The seven “work” districts included the six districts in Table 5-8 and a seventh district (District 19) containing all areas not included in the six home districts. Each flow value in the BOC database was assigned to the appropriate combination of home and work districts according to the residence and work census tracts listed for the flow. Table 5-9 lists the sums of the flows assigned to each combination of home and work district. It also presents each sum as a fraction of the total flow originating at the

Top of Figure 5-1



**CO Monitors and Population Density
in the Denver Study Area**

Figure 5-1. Fixed-Site CO Monitoring Sites Used to Define Denver Exposure Districts.

Table 5-8. Exposure Districts Defined for Denver pNEM/CO Analysis.

District no.	Monitor(s) included in district			Number of census tracts assigned to district
	AIRS ID	City	Address	
1	005-0002	Littleton	8100 So. University Blvd. (Highlands)	47
2	031-0002	Denver	2105 Broadway (CAMP)	44
3	031-0013	Denver	14th and Albion St. (NJHE)	116
4	031-0014	Denver	23 rd and Julian (Carriage)	61
5	059-0002	Arvada	W. 57 th Ave. and Garrison	55
6	013-0010 013-1001	Boulder	2150 28 th Street 2320 Marine Street	28
19	Remaining census tracts in nine county region			242

Table 5-9. Number and Fraction of Denver Commuters Associated with Each Combination of Home and Work District.

Home district	Statistic	Work District							Totals by Home District
		Districts containing CO monitors						District 19 ^a	
		1	2	3	4	5	6		
1	Commuters	35,045	15,334	14,929	3,974	1,499	470	18,288	89,539
	Fraction ^b	0.391	0.171	0.167	0.044	0.017	0.005	0.204	1.000
2	Commuters	2,184	27,440	8,649	3,457	1,789	669	4,511	48,699
	Fraction	0.045	0.563	0.178	0.071	0.037	0.014	0.093	1.000
3	Commuters	12,804	42,602	64,969	7,844	2,651	1,434	22,028	154,332
	Fraction	0.083	0.276	0.421	0.051	0.017	0.009	0.143	1.000
4	Commuters	5,207	26,728	13,773	28,053	9,283	898	14,563	98,505
	Fraction	0.053	0.271	0.140	0.285	0.094	0.009	0.148	1.000
5	Commuters	2,811	22,282	11,666	14,765	39,110	3,428	19,675	113,737
	Fraction	0.025	0.196	0.103	0.130	0.344	0.030	0.173	1.000
6	Commuters	814	4,149	1,771	766	1,305	44,630	6,839	60,274
	Fraction	0.014	0.069	0.029	0.013	0.022	0.740	0.113	1.000
Totals by Work District		58,865	138,535	115,757	58,859	55,637	51,529	85,904	565,086

^a District 19 includes all census tracts in the nine-county area that are not located in Districts 1 through 6.

^b Fraction = $(Com_{ij}) / (Com_i)$

Com_{ij} = number of workers commuting from home district i to work district j

Com_i = total number of workers commuting from home district i to all work districts (including District 19)

indicated home district. For example, the flow from Home District No. 1 to Work District No. 2 is 15,334. Listed under this value is the fraction 0.171, calculated by dividing 15,334 by the total flow (89,539) from Home District No. 1 to all districts.

Home District No. 3, the district containing the largest number of census tracts, is associated with the largest number of commuters (154,332). Home District No. 2 has the smallest number of commuters (48,699). Work District No. 2 is associated with the largest number of commuters (138,535); Work District No. 6 (Boulder) has the smallest number of commuters (51,529). Note that these values include only those commuters which move among the home and work districts listed in Table 5-8.

Table 5-9 accounts for the commuting patterns of 565,086 workers. Of these, 85,904 workers (15 percent) reside in one of the six home districts but work in District 19 (i.e., at a location more than 10 km from the nearest fixed-site monitor). People working in District 19 were excluded from the pNEM/CO analysis of Denver, based on the assumption that pNEM/CO could not accurately estimate the exposures of people who worked more than 10 km from a monitoring station. Consequently, the working cohorts are limited to people with residential and work locations within Districts 1 through 6.

5.3.1.4 Denver Data Quality

It should be noted that the BOC Commuting Database is not a perfect representation of the commuting patterns of Denver residents. The data were acquired from the census “long form” (see Table 3-5) which the BOC administered to about one-sixth of the Denver residents. The BOC extrapolated the data from this subset of the population to the remainder of the population using various assumptions and supplemental information.

Analysts defined WR as the ratio of WA (number of workers allocated to a particular home-work combination by indirect methods) to WF (total number of workers associated with the home-work combination); i.e.,

$$WR = (WA)/(WF) \quad (5-6)$$

WR provided a crude indication of the uncertainty associated with each WF value in the database.

Table 5-10 presents the population-weighted mean value of WR (expressed as a percentage) for each of the 42 combinations of home and work district. The 42 WR values range from 10.28 percent (Home District No. 6, Work District No. 3) to 43.90 percent (Home District No. 1, Work District No. 5). The largest values (indicating the highest degree of uncertainty) are associated with Work Districts Nos. 5 and 19.

Table 5-10 also provides an aggregate WR value for all flows originating in each home district. The six values range from 22.14 percent (Home District No. 1) to 28.59 percent (Home District No. 2). Aggregate WR values for all flows ending in each work district can also be found in Table 5-10. The values range from 18.46 percent (Work District No. 1) to 35.27 percent (Work District No. 19).

The overall aggregate value of WR is 23.98 percent. In general, this analysis indicates that less than 25 percent of the commute trips were estimated indirectly using supplemental data.

Table 5-10. Percentage of Denver Commuters (WR) Assigned by Bureau of Census to Each Home-Work Combination Based on Supplemental Data.

Home District	Work District							All work districts
	Districts containing CO monitor(s)						19	
	1	2	3	4	5	6		
1	15.21	18.47	23.93	23.70	43.90 ^a	21.70	34.93 ^a	22.14
2	23.99	30.09 ^a	22.68	22.36	29.29 ^a	22.87	38.40 ^a	28.59 ^a
3	19.31	22.87	16.21	32.66 ^a	35.16 ^a	24.48	35.28 ^a	22.27
4	30.44 ^a	22.96	35.05 ^a	14.95	33.14 ^a	35.63 ^a	36.14 ^a	25.79 ^a
5	24.55	17.72	25.12 ^a	21.38	25.73 ^a	33.07 ^a	24.18	24.18
6	32.56 ^a	21.43	10.28	11.49	31.34 ^a	22.47	38.60 ^a	24.06
All home districts	18.46	22.96	20.74	19.91	26.93 ^a	22.97	35.27 ^a	23.98

^a Value exceeds 25 percent.

5.3.2 Home-to-School

As discussed in Subsection 5.3.1, the NEM/HAPEM series of exposure models have historically accounted for commuting trips between home and work locations. Prior to 2001, none of the exposure models developed by OAQPS included an algorithm for simulating the commuting trips of students between home and school. As part of the development of HAPEM4 during 2001, researchers with ICF Consulting and TRJ Environmental, Inc., conducted an in-depth review of the transportation and planning literature to identify methods that could be used to allocate students residing within a specific study area among various private and public schools in and around the area (Johnson, 2000). They selected a subset of the candidate methods for further evaluation and obtained several school enrollment data sets appropriate for testing and calibrating each method. The work concluded with the development and testing of a gravity model that accounts for residential patterns, school enrollments, and home-to-school distances.

Results of this effort are summarized in the following four subsections as follows:

- Subsection 5.3.2.1 provides background on EPA's decisions to (1) develop a special version of HAPEM4 that would account for school commuting patterns and (2) use a gravity model to develop the required HST files;
- Section 5.3.2.2 describes the selected gravity model in mathematical terms, presents a method for calibrating the model, and describes a method for converting students' ages (the classification scheme appearing in census data) to grade level (the classification scheme required for allocating students to schools);
- Section 5.3.2.3 identifies revisions to the HAPEM4 program that were made to implement a new school commuting module incorporating the proposed gravity model;
- Section 5.3.2.4 describes a test application in which the school commute module was used to estimate the school commuting patterns of public school students in Harris County, TX.

Additional information concerning the results of the test application can be found in a technical memorandum prepared by Laurenson, Rosenbaum, Huang, and Cohen (2001).

5.3.2.1 Background

As part of the National Air Toxics Assessments (NATA) activities, OAQPS used HAPEM4 to conduct an initial national-scale assessment (NSA) during 2000. This initial NSA had two principal goals: (1) to demonstrate a proposed approach for characterizing air toxics risks nationwide and (2) to characterize health risks to sensitive population subgroups, such as children, associated with inhalation exposures to 33

hazardous air pollutants (HAPs) identified as priority pollutants by the Integrated Urban Air Toxics program. A report prepared by the U.S. EPA (2001) describes the 2000 version of HAPEM4 used in the initial NSA. A supplemental report by ICF Consulting and TRJ Environmental, Inc. (2001) provides a more detailed account of the methods used by HAPEM4 to estimate microenvironmental concentrations. Additional information concerning HAPEM4 can be found in a user's guide prepared by Rosenbaum (2002).

The 2000 version of HAPEM4 accounted for home-to-work commuting but not for home-to-school commuting. Because a child's exposure is expected to vary between home and school locations, EPA directed a team of researchers to evaluate the feasibility of simulating home-to-school trips in future versions of HAPEM4 as well as in the Air Pollutant Exposure Model (APEX). Researchers assumed that these models would require home-to-school trip (HST) tables that listed the number of children associated with each possible combination of age, home census tract, and school census tract. The home census tracts would include all residential census tracts within a defined study area. Ideally, the school census tracts would include the locations of all schools (public and private) attended by students residing in the study area, including schools that fall outside the study area.

Researchers were unable to identify a national HST database that met the specified requirements. Consequently, they investigated two approaches to constructing the required HST database: (1) aggregating local sources of HST data into a single national database and (2) employing one or more "school commute models" that would estimate school attendance patterns according to available data on residential patterns and school enrollments. The data aggregation approach was subsequently judged to be infeasible because data sets from local sources were found to differ significantly with respect to coverage, content, and format. In addition, usable HST data were not available for many areas.

The use of school commute models was judged to be the more promising approach. Researchers identified a set of modeling approaches based on a "gravity" model that could be combined into a general methodology that would permit the use of school enrollment data with varying characteristics. A technical memorandum by Lee, Laurenson, Balassiano, and Johnson (January 22, 2001) provides preliminary recommendations for using these methods to allocate students to public and private schools in a HAPEM4-type exposure analysis. The memorandum also describes and evaluates potential databases available for implementing each method.

A follow-up memorandum prepared by Lee, Balassiano, Ragan, and Laurenson (March 30, 2001) provides a preliminary method for converting data obtained from five available databases into the input data appropriate for a generic gravity model. The databases are listed below.

1998-99 Common Core of Data (CCD) Public School Database

2000 Delorme Street Database

1999 ESRI Maps and Data CD

1997-98 National Center for Education Statistics (NCES) Private School

Database

1990 U.S. Census Database

The cited memorandum discusses various limitations of these databases and provides recommendations for addressing these deficiencies. The authors also noted that the expected substitution of 2000 census data for 1990 data would require updating and/or reformatting the other databases. In addition, the authors proposed the development of a table that could be used to map student age onto grade level, given that the public and private school databases provide data specific to grade level and the U.S. Census provides age-specific data.

Researchers subsequently revised the list of candidate methods to include a “nearest school with available capacity” model, a “gravity” model, and an “enrollment-driven distribution” model (a special version of the gravity model). These methods are briefly described below.

The **nearest school with available capacity model** was designed to assign students to the nearest school with available space. School proximity is determined by the straight-line distance between a student’s residence (approximated by the census tract centroid containing the residence) and each available school. The model uses a step-wise process to assign students to the nearest school in small increments, verifying that the school has available capacity between steps. If the nearest school becomes filled, students are assigned to the next nearest school. This process continues until all students are assigned to a school or all schools are filled to capacity. Two alternatives were proposed for the case where all of the schools were filled to capacity before all students were assigned. According to the first alternative, the excess students would be “forced” into nearby schools under the assumption that the specified enrollment values were lower than the true enrollments of these schools. Under the second alternative, analysts would adjust enrollments of all schools proportionally to match the total number of students to be allocated.

The **gravity model** is a more complex model that assumes the probability of a student attending a particular school increases as the home-to-school distance decreases and the school’s enrollment capacity increases. The probabilities (fractions) are estimated by optimizing a mathematical function through an iterative procedure that incorporates information on home-to-school distances and enrollments for all combinations of residential locations and schools. The procedure requires that the user provide a value for a particular parameter in the gravity model (referred to as coefficient “a”) that is likely to vary according to factors relating to public/private classification, enrollment policy (open vs. closed), school type (elementary, middle, or high school), and land use (urban vs. rural). Appropriate values of coefficient “a” would be estimated by fitting the gravity model to “calibration” data sets consisting of actual HST data for particular geographical areas considered representative of a specific set of conditions (e.g., public high school students in an urban school district). Researchers evaluated two methods for fitting the gravity model to calibration

data sets (maximum likelihood and minimum chi-squared estimation) and determined the latter to be the more feasible method given projected resource constraints.

The **enrollment-driven distribution model** is a special version of the gravity model that considers only school enrollment capacities in assigning students within specified boundaries (e.g., school district, radius). The probability of a student attending a particular school increases with the school's enrollment capacity but is unaffected by home-to-school distance within the boundary. Consequently, a student is more likely to attend a distant school with a large enrollment than a nearby school with a small enrollment. The enrollment-driven distribution model is applied in the same manner as the gravity model, except that the user is not required to provide a value for coefficient "a" (this coefficient is effectively set to zero).

Appendix A of a technical memorandum prepared by Johnson, Lee, Laurenson, Cohen, and Rosenbaum (May 31, 2001) provides more detailed descriptions of these three methods together with a discussion of the appropriate applications for each method.

The second method (the gravity model) was subsequently selected for further development and testing as part of a "school commute module" for use with HAPEM4. Figure 5-2 is a flow diagram indicating how the school commute module and its associated input and output databases fit within the larger HAPEM4 modeling framework. Essentially, the school commute module creates input data for HAPEM4 by applying a calibrated gravity model to data on school district boundaries, student residential populations and locations, school locations, and school enrollments. Subsection 5.3.2.2 presents the mathematical structure of the gravity model and describes the method developed by researchers for its calibration.

A potential advantage of the gravity model is that the "a" coefficient of the gravity model can be varied to represent different categories of schools as characterized by degree of open enrollment and other factors. Given this flexibility, researchers determined that the gravity model was particularly applicable to public school students who have a variety of enrollment options, including magnet schools, charter schools, and other schools with "open" enrollment policies. Subsection 5.3.2.4 describes an initial test of the gravity model in which it was applied to public school students in Harris County, Texas. Additional information concerning the applicability of the gravity model to public school students can be found in a technical memorandum by Laurenson, Rosenbaum, Huang, and Cohen (September 30, 2001).

To date, there has been no comparable application of any model to private school students. However, the memorandum by Laurenson, Rosenbaum, Huang, and Cohen (September 30, 2001) includes proposals for applying the gravity model and enrollment-driven distribution model to private schools. In applications of the gravity model, researchers would use estimates of the "a" coefficient previously developed for magnet and/or charter public schools as the basis for estimating the value of "a" for private schools. This approach is considered reasonable because the commuting patterns of students attending private schools are expected to be similar to those attending magnet and charter schools; that is, students are not necessarily bound to the nearest school, but are likely to show a preference for closer schools. Researchers

Top of Figure 5-2

Figure 1. Overview of prototype school commute module

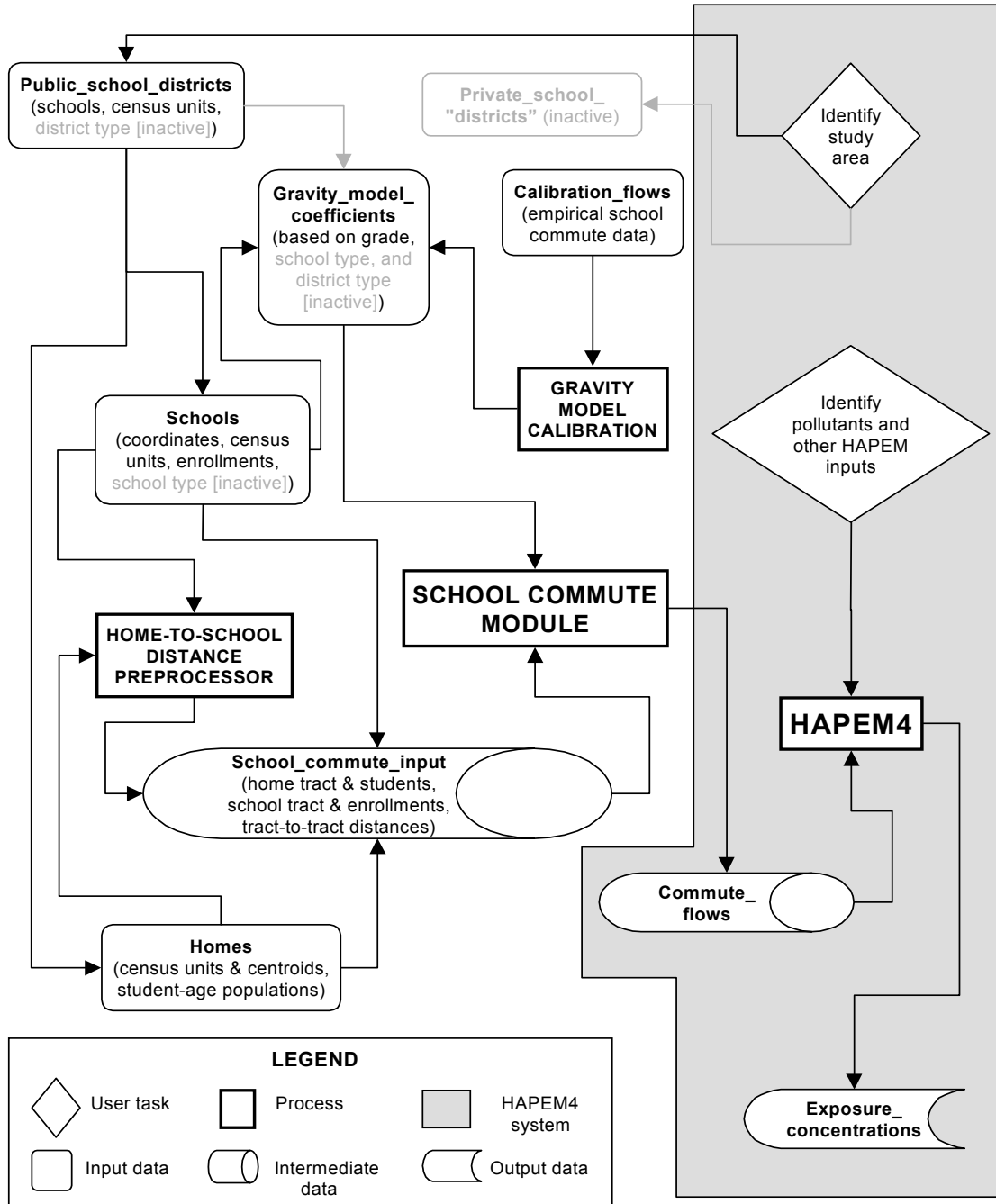


Figure 5-2. Overview of Prototype School Commute Module.

judged the enrollment-driven distribution model to be more appropriate in applications in which students are equally likely to attend any private school within a given area (such as a county).

Regardless of the model(s) selected for public and private schools, analysts must convert students' ages (from US Census data) to grade level before allocating students to schools. Subsection 5.3.2.2 summarizes a conversion method proposed by Johnson, Lee, Cohen, and Rosenbaum (May 31, 2001).

5.3.2.2 The Gravity Model: Structure and Calibration

The gravity model proposed by Laurenson, Rosenbaum, Huang, and Cohen (September 30, 2001) assumes that the fraction of students residing in a given census tract that attend a particular school will increase as the home-to-school distance decreases and the school's enrollment capacity increases. That is, given

a set of m census tracts with residential student populations, $R(i)$,

a set of n schools with student enrollments $E(j)$, and

a set of distances (in miles) from the home census tract centroids to the schools, $D(i,j)$,

the allocation of students to schools is estimated based on the following "gravity model" relationship:

$$P(i,j) = \frac{[G(j)]}{\left\{ [D(i,j)]^a [M'(i)] \right\}} \quad (5-7)$$

where $P(i,j)$ is the probability that a student in census tract i will attend school j , the exponent "a" will be determined by calibration, and

$$M'(i) = \sum_{j=1}^n \frac{[G(j)]}{[D(i,j)]^a} . \quad (5-8)$$

$G(j)$ is an "attractor" function for school j and is monotonic in the enrollment, $E(j)$. For a given value of a , the values of the $G(j)$'s are derived in an iterative manner to be consistent with the $R(i)$'s and $E(j)$'s.

To apply the gravity model, the user first defines the census tract expected attendance (or allocation) as

$$A(i,j) = P(i,j) \times R(i) . \quad (5-9)$$

When Equation 5-9 has been applied to all combinations of census tracts and schools, each school will be allocated a total of $A(j)$ students, where

$$A(j) = \sum_{i=1}^m A(i, j) \quad (5-10)$$

Thus

$$A(j) = \sum_{i=1}^m \left(\frac{[G(j)]}{\{[D(i, j)]^a [M'(i)]\}} \times R(i) \right) \quad (5-11)$$

As shown below, one can solve this equation for $G(j)$ to make $A(j) = E(j)$:

$$E(j) = \sum_{i=1}^m \left(\frac{[G(j)]}{\{[D(i, j)]^a [M'(i)]\}} \times R(i) \right) \quad (5-12)$$

Note that the enrollment $E(j)$ increases when $G(j)$ increases, as can be shown by partially differentiating Equation 5-12 with respect to $G(j)$. Thus $G(j)$ increases when $E(j)$ increases, as desired, so the attraction of school j is an increasing function of the enrollment, although it also depends upon the set of school-tract distances.

To solve the nonlinear Equation 5-12 for $G(j)$, the following iterative scheme is applied.

1. Start with $G(j) = 1/n$ for all j .
2. Compute $M'(i)$ for each i from Equation 5-8.
3. Solve the n linear Equations 5-12 for the $G(j)$ treating $M'(i)$ as constants.
4. Divide each $G(j)$ by the sum of the $G(j)$ so that $\sum G(j) = 1$.
5. Recompute $M'(i)$ for each i from Equation 5-8 applied to the new $G(j)$'s.
6. Solve the n linear Equations 5-12 for the $G(j)$ treating the new $M'(i)$ as constants.
7. Divide each $G(j)$ by the sum of the $G(j)$ so that $\sum G(j) = 1$.
8. Repeat until convergence. Note that the $G(j)$'s are defined only up to a multiplicative factor, since the same probabilities $P(i, j)$ are obtained if all the $G(j)$ are multiplied by the same number, so the algorithm keeps the sum of $G(j)$ fixed at 1.

Appendix A in the technical memorandum by Johnson, Lee, Laurenson, and Cohen (May 31, 2001) provides a brief history of the derivation of this approach.

The gravity model requires an estimate for the “a” coefficient that appears in Equation 5-7. An estimate applicable to a particular type of school system and geographic region can be obtained through the use of a representative “calibration” data

set; i.e., a data set consisting of known values, $f(i,j)$, for the number of students residing in each of m census tracts that attend each of n schools. Using the following method, one can estimate a value for coefficient “a” that minimizes a mean square error (MSE) statistic.

For each candidate value of coefficient “a”, compute $M'(i)$, $G(j)$, and hence $P(i, j)$ as described above (the results will depend on a , $R(j)$, and the $E(j)$, but not on the $f(i,j)$). The expected attendance $A(i, j)$, which depends upon the value of coefficient “a”, is given in Equation 5-9.

The mean square error (MSE) statistic measures the discrepancy between the expected attendance $A(i, j)$ and the observed attendance $f(i,j)$:

$$MSE = \sum \frac{\{f(i, j) - A(i, j)\}^2}{n} \quad (5-13)$$

The sum is calculated across all i, j pairs, and n is the number of i, j pairs. The estimated value of coefficient “a” is found by choosing a value for coefficient “a” that minimizes the MSE.

The value of coefficient “a” can be expected to vary according to a variety of factors, including school classification (regular public vs public charter/magnet vs private), enrollment type (open enrollment vs closed enrollment), school type (elementary vs middle vs high schools, or primary vs secondary), land use classification (urban vs suburban vs rural), and general location of school (coastal vs inland, plain vs mountainous, East Coast vs West Coast, North vs South). Ideally, analysts would employ a wide range of calibration data sets to obtain estimates of coefficient “a” that fully account for all of these factors. However, initial testing of the prototype was based on the simplifying assumption that the value of coefficient “a” was primarily affected by school classification and school type. Grade-specific estimates of coefficient “a” were obtained for grade levels K - 12 within a large public school system (private schools were not included in this analysis).

Analysts initially identified two sets of HST data appropriate for the calibration task: one for the Blue Valley, KS school district (approximately 17,000 students) and the other for the San Diego, CA school district (approximately 140,000 students). Because of the larger size of the CA data set and its greater similarity to Harris County, Texas (the initial test area), analysts selected the CA data as the basis for developing parameter values by grade. The results of this calibration are presented in Table 5-11. The 13 values for coefficient “a” range from 2.39 (kindergarten) to 3.36 (9th grade) with 11 of the values falling between 2.40 and 2.81. The average value of coefficient “a” for grades K through 6 is 2.51; the average for grades 7 through 12 is 2.75. In general, higher grades tend to have slightly larger estimated values of coefficient “a”, although this pattern may not be statistically significant.

Note that the gravity model is intended for allocating the students in each combination of residential census tract and grade range to the appropriate schools in a defined school district. Ideally, the user would acquire the required data on students by

Table 5-11. Estimated Values of Coefficient “a” in Equation 5-7, Based on Calibration of School Enrollment Data from San Diego, CA.

Grade Level	Estimated Value of Coefficient “a” in Equation 5-7	Grade Level	Estimated Value of Coefficient “a” in Equation 5-7
K	2.39	7	2.40
1	2.49	8	2.60
2	2.56	9	3.36
3	2.58	10	2.81
4	2.44	11	2.63
5	2.60	12	2.68
6	2.51		

Table 5-12. Distribution of U.S. Enrolled Students Aged 14 by Grade Level, October 1999.

Category	Students, thousands (percent of total in parentheses)							
	Total	Enrolled	6	7	8	9	10	11
All	3971 (100.0)	3903 (98.3)	15 (0.4)	86 (2.2)	919 (23.1)	2680 (67.5)	199 (5.0)	5 (0.1)
Male	2035 (100.0)	1984 (97.5)	8 (0.4)	48 (2.4)	524 (25.7)	1311 (64.4)	87 (4.3)	5 (0.2)
Female	1936 (100.0)	1919 (99.1)	6 (0.3)	38 (2.0)	395 (20.4)	1368 (70.7)	112 (5.8)	0 (0.0)

census tract and grade range directly from the Bureau of Census (BOC). Unfortunately, the BOC provides a breakdown of the students in each census tract by age rather than grade level. Consequently, the user must employ a procedure for distributing the students in each age group among the appropriate grade levels prior to implementing a gravity model or equivalent student allocation procedure. The following procedure was proposed by Johnson, Lee, Laurenson, Cohen, and Rosenbaum (May 31, 2001).

Let $K(i,x)$ indicate the number of students in census tract i that are age x . Let $N(i,x,y)$ indicate the number of students in census tract i that are age x and belong in grade y . One can estimate $N(i,x,y)$ by the expression

$$N(i,x,y) = C(x,y) \times K(i,x) \quad (5-13)$$

where $C(x,y)$ is the fraction of students aged x that are in grade y . Values of $C(x,y)$ can

be derived from enrollment data classified by race and gender found at the following web site:

<http://www.census.gov/population/www/socdemo/school/p20-533.html>.

Table 5-12 presents an excerpt from this tabulation showing the data for 14 year old students by gender only. Lacking data specific to each census tract, one can assume that the percentage values in Table 5-12 can be applied to each of the census tracts within a study area. If there are 100 males aged 14 years in a particular census tract, Equation 5-13 would provide the following grade distribution for these children:

<u>Status</u>	<u>Percentage</u>	<u>Number (rounded)</u>
Not enrolled	2.5	3
6 th grade	0.4	0
7 th grade	2.4	2
8 th grade	25.7	26
9 th grade	64.4	64
10 th grade	4.3	4
11 th grade	0.2	0

The percentage values indicate that the ninth grade is the most probable grade level for males aged 14 years (64.4 percent), followed by the eighth grade (25.7 percent) and tenth grade (4.3 percent). Note that 2.5 percent of males aged 14 years are not enrolled in school. These children can be either (1) omitted from the analysis entirely or (2) distributed among grades 6 through 11 in the same proportions as the children with known grade assignments.

The above example shows how males aged 14 would be distributed using the 1999 national data. The same procedure would be repeated for each age and gender combination within the census tract until all children were distributed among the grade levels. The resulting totals by grade level and census tract would be the inputs to the gravity model.

5.3.2.3 Revisions to the HAPEM4 Model Code

The original version of HAPEM4 accounted for home-to-work commuting but not home-to-school commuting. During 2001, researchers developed a special version of HAPEM4 that addressed school commuting for four demographic groups (males and females in the 5-11 age group and in the 12-17 age group) using a general approach similar to that used previously for home-to-work commuting. When an activity pattern was selected that indicated a child was in the school microenvironment, HAPEM4 selected a school tract from the set of possible school tracts specified for that resident

tract and age group. The probability of selection from among the possible school tracts was determined by the estimated proportion of students in that resident tract/age group combination that attended school in the tract. The HAPEM4 User's Guide (Rosenbaum, 2002) provides a detailed explanation of the worker commuting algorithm.

5.3.2.4 Sample Application

Figure 5-2 shows how the school commute module is implemented with the HAPEM4 modeling framework. The procedure requires four data sets that must be provided by the user:

<i>Calibration_flows:</i>	empirical school commute data appropriate for estimation of coefficient "a" in Equation 5-7
<i>Public_school_districts:</i>	a listing of the census units included in each school district
<i>Schools:</i>	school locations and enrollments
<i>Homes:</i>	home locations and student-age populations.

These databases are processed by various programs to create two intermediate databases (*Gravity_model_coefficients* and *School_commute_input*). The school commute module operates on these databases to produce *Commute_flows*, an output database that contains school commute data that are (1) specific to the study area and (2) correctly formatted for input into the special version of HAPEM4 described in Subsection 5.3.2.3.

Analysts tested the procedure by using it to develop a *Commute_flows* database appropriate for applying HAPEM4 to the public school students of Harris County, Texas (Figure 5-3); private school students were not included in the test. Analysts selected Harris County (which includes Houston) because its large industrial base was expected to produce significant spatial variations in the ambient concentrations of pollutants associated with emissions from large stationary sources. These spatial variations were expected to affect exposure estimates obtained from HAPEM4 runs performed with and without the school commute module activated, enabling analysts to evaluate the sensitivity of HAPEM4 exposure estimates to school commuting.

In the sample application, the *Gravity_model_coefficients* database provided estimates for coefficient "a" for public schools by grade level only (future applications may also classify values by school type and other factors). Researchers assumed that the values obtained from the analysis of San Diego data (see Table 5-11) were acceptable for the Harris County test. Subsection 5.3.2.2 describes the method used to develop these estimates.

The remainder of this subsection briefly summarizes the steps followed in developing a *School_commute_input* database for Harris County applicable to public school students. In reviewing this material, the reader should note that analysts first created national versions of the *Schools* and *Homes* input databases and then extracted data specific to the census units in Harris County for inclusion in the

top of Figure 5-3

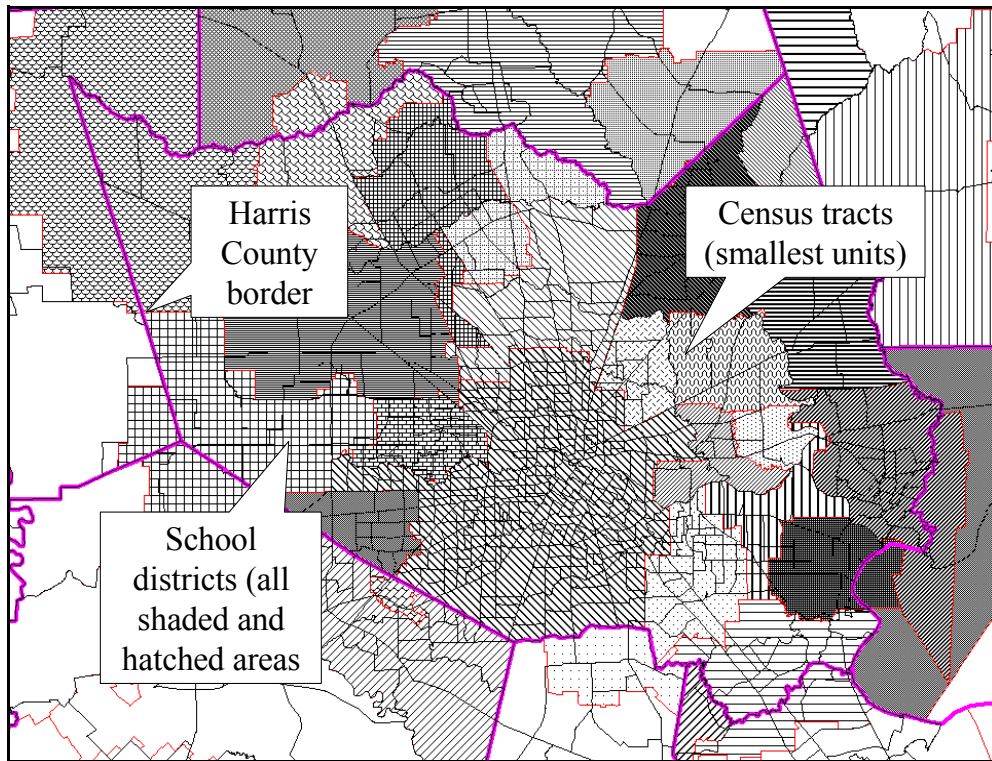


Figure 5-3. Map of Harris County showing boundaries of school districts and census tracts.

School_commute_input database.

Public_school_districts (school district boundaries). The primary source of school district data for the school module was the Proximity One school district database, containing boundary coordinates for approximately 14,000 public school districts. Analysts used this data set and a “point-in-polygon” procedure to identify the schools and census tracts located within each school district in the U.S. When a census tract fell within more than one school district, each portion was designated a “census unit.” Consequently, each census unit could be identified by an unique code specifying census tract and school district.

An automated procedure within the MapInfo GIS tool was used to estimate the area of each census unit. The population of each census unit was estimated by assuming the population of the census tract was proportionately divided among the constituent census units according to the areas of the census units.

Schools (school locations and enrollments). The primary source of public school data for the test run was the National Center for Education Statistics (NCES) Common Core of Data (CCD) database. The CCD contains detailed data on school enrollment by grade, school location, and other relevant school characteristics for every public school in the county. Using the CCD school location address information (i.e., address, city, state, and zip code) and geocoding (address-matching) software, analysts determined the coordinates (in Albers meters) of each public school in the U.S. The CCD database also provided enrollment counts by grade level for each school. The resulting *Schools* database contained location and enrollment data for approximately 94,000 public schools.

The reader should note that the CCD database was not complete or totally accurate. Address defects occurring frequently in the database included (1) blank address fields, (2) school addresses placing the school in the wrong county, and (3) school addresses placing the school in the wrong zip code. In such cases, analysts could frequently use a map-overlay procedure to estimate the school’s approximate location based on one of the available address data items (street address, street block, zip code, etc.) More than 70% of the school locations were determined by street address - the most accurate geocoding method available; the remainder relied on the less accurate methods. (Proprietary databases are currently available that provide more accurate school coordinates than the CCD.)

Homes (home locations and student-age populations). In the next phase of the process, analysts applied the following four-step process to 1990 U.S. Census data to determine the number of children, by age and grade, attending public school that resided in each census unit.

1. Estimate population in each census unit according to relative area. The populations of children within census units (i.e., census tracts divided into sections by school district boundaries) were determined by first assuming that the population was evenly distributed over the census tract. Using the census unit and associated census tract area estimates described above, the population in each census unit was estimated based on the area of the census unit relative to the area of its associated census tract (e.g., if the census unit comprised 50 percent of the area of the census tract, it would be assumed that half the population of the census tract resided in the

given census unit).

2. Estimate number of children in public elementary schools and public high schools by census tract. Analysts calculated the fraction of school age children in each census tract attending elementary public schools as the ratio of total public elementary school enrollment to the total number of children in the 5 to 17 age range. In a similar manner, the fraction of school age children in each census tract attending high schools was estimated as the ratio of total public high school enrollment to the total number of children in the 5 to 17 age range. These census unit-specific fractions were then multiplied by the total number of students in each age range in each census tract to produce estimates of the number of public school students by age range and census unit. (It should be noted that this approach was not specific to grade and thus did not capture the likely variation in public/private school attendance by grade. As there were no readily available sources of public versus private school enrollment data that were both grade and census unit-specific at the time of this analysis, this method was deemed acceptable given EPA's needs and the available resources.)
3. Convert Ages to Grades. The procedure described at the end of Subsection 5.3.2.2 was used to allocate the public school students in each combination of residential census unit and age range among the appropriate grade levels. Because the data provided by the US Census Bureau provided population estimates for age ranges rather than individual years of age, analysts made an additional assumption in implementing the procedure: the population for a given range of ages was evenly distributed among the individual years included in the range. If, for example, the census data provided an estimate of that 100 people ages 7 and 8 resided in a given census tract, analysts assumed that 50 were 7 years of age and 50 were 8 years of age. After completing this process, the census data consisted of the number of public school students in each grade in each census unit for the entire United States.
4. Match Census Data to Enrollments. The final step in processing the census data was to adjust the number of students in each grade in each school district obtained from the census data so as to match the grade-level enrollments in each school district obtained from the school data base. This step was necessary because the gravity model required that the total school enrollment for all schools by grade within a school district equal the total census data by grade for all census units in the district. These values did not match up initially because the census data were obtained from the 1990 US Census and the school data were specific to the 1997-1998 school year. Analysts summed the grade-level-specific student data over each school district for all census units and schools in the United States. The results were total number of students by grade for each school district in the United States for both the census and school

data sets. Next, they calculated the ratio of the total students from the school data set to the total students from the census data set for each grade and school district. The final step was to multiply these ratios by the data developed for public schools under Step 3, being careful to apply to ratios to the appropriate grade levels and census units (which was determined based on the associated school district). The results of this process were revised estimates of number of public school students in each grade in each census unit in the United States.

This approach yielded data on centroid locations and school-age populations by grade for approximately 150,000 census units within the U.S.

School_commute_input (combined school and home data). Data specific to Harris County were extracted from the *Homes* and *Schools* databases and processed to create 13 grade-specific (K-12) school commute input files collectively referred to as the *School_commute_input* database. As indicated in Figure 5-2, the *Home-to-School Distance Processor* was used to determine the straight-line distance between the centroid of each home census unit and the location of each school for all home-school combinations within the same school district. The *School_commute_input* database also includes the FIPS census tract codes for the home and school census tracts, the number of grade-specific students in each home tract, and the grade-specific enrollments for each school.

School Commute Module. As indicated in Figure 5-2, the inputs to the *school commute module* consist of home-to-school distances, student populations, and student enrollments from the grade-specific files in the *school_commute_input* database and estimates of coefficient “a” from the *gravity_model_coefficients* database. In the Harris County test application, the module first used these data to assign student populations from each home census tract to appropriate schools within the same school district. These assignments were then organized into a grade-specific HST file that reflected the HST file used to calibrate the gravity model (in this case San Diego). Next, the school commute module consolidated the 13 grade-specific (i.e., K-12) input files by first combining the duplicate tract-to-tract flows in each grade file (e.g., for two or more schools with that grade in the same tract) by summing the student flows and averaging the distances. Next, the grade-specific files were combined into a single file (*commute_flows*) by summing the student flows for each tract-to-tract combination across grades. Finally, the grade-specific flows were converted back to age-specific flows and then aggregated into total flows for the HAPEM4 age groups associated with elementary school (ages 5 to 11) and high school (ages 12 to 17) students.

Table 5-13 presents selected commute-related statistics for Harris County based on the input files used by the school commute module and the commute flow estimates produced by the module. Note that several statistics are given in two forms: including zero flows and excluding zero flows. Of the 67,000 possible home-school tract combinations considered by the school commute module, only 22,805 (34 percent) were assigned non-zero flows.

Table 5-13. Harris County, TX school-related characteristics.

Characteristic	Value
Public school districts	20
Public schools	381
Census tracts	587
Public school students: total	587,399
Public school students: age 5-11	328,703
Public school students: age 12-17	258,696
Within-district home tract-to-school tract combinations: total, <u>including</u> zero flows	67,000
Within-district home tract-to-school tract combinations: total, excluding zero flows	22,805
Within-district home tract-to-school tract combinations: age 5-11, excluding zero flows	18,649
Within-district home tract-to-school tract combinations: age 12-17, excluding zero flows	11,115
Average home tract-to-school tract distance (miles): both age groups, <u>including</u> zero flows	8.3
Average home tract-to-school tract distance (miles): age 5-11, excluding zero flows	2.3
Average home tract-to-school tract distance (miles): age 12-17, excluding zero flows	2.7

5.4 REFERENCES FOR SECTION 5

Ainsworth, B., W. Haskell, A. Leon, et al. 1993. "Compendium of Physical Activities: Classification of Energy Costs of Human Physical Activities," Med. Sci. Sports Exer. Volume 25, pp. 71 - 80.

Akland, G. G., T. D. Hartwell, T. R. Johnson, and R. W. Whitmore. 1985. "Measuring Human Exposure to carbon Monoxide in Washington, DC, and Denver, Colorado During the Winter of 1982 - 83." Environ Sci Technol. Volume 19, pp. 911 - 918.

Bureau of Census. 1983. Current Housing Reports, Series H-150-83, "Housing Characteristics of Recent Movers for the United States and Regions: 1983. Annual Housing Survey: 1983, Part D," Washington, D.C.

Bureau of Census. 1989. Current Population Reports, Series P-20, Number 430, "Geographical Mobility: March 1986 to March 1987," U.S. Department of Commerce.

Bureau of Census. 1990. Current Population Reports, Series P-25, Number 1057, "U.S. Population Estimates by Age, Sex, Race, and Hispanic Origin: 1989," U.S.

Department of Commerce.

Bureau of the Census. 1994. "STP154: Census Tract of Work by Census Tract of Residence." Journey-to-Work and Migration Statistics Branch, Population Division, U. S. Bureau of the Census.

Glen, J., et al. 1997. Development of NERL/CHAD: The National Exposure Research Laboratory Consolidated Human Activity Database. Report prepared by ManTech Environmental Technology, Inc., under EPA Contract No. 68-D5-0049.

Goldstein, B.R., G Tardiff, G. Hoffnagle, and R. Kester. 1992. Valdez Health Study: Summary Report. Alyeska Pipeline Service Company, Anchorage, AK.

Hartwell, T. D., C. A. Clayton, R. M. Mitchie, et al. 1984. A Study of Carbon Monoxide Exposure of Residents of Washington, DC, and Denver, Colorado. Report No. EPA-600/S4-84-031. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. 600/S4-84-014. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Johnson, T. R. 1987. A Study of Human Activity Patterns in Cincinnati, Ohio. Report prepared by PEI Associates, Inc. for Electric Power Research Institute, Palo Alto. Available from Ted Johnson, TRJ Environmental, Inc., 713 Shadylawn Road, Chapel Hill, North Carolina, 27514.

Johnson, T. 1988. "Partitioning Activity Diary Data According to Temperature." Technical letter to Tom McCurdy, U.S. Environmental Protection Agency, Research Triangle Park, N.C. March 25.

Johnson, T. 1989. A Study of Human Activity Patterns in Cincinnati, Ohio. Electric Power Research Institute. Palo Alto, CA.

Johnson, T. R. 1995. "Recent advances in the estimation of population exposure to mobile source pollutants." J. Expos. Anal. Environ. Epidemiol. 1995: 5: 551-571.

Johnson, T. 2000. "Methods for Estimating School Commuting Patterns for Children." Memorandum to Jim Laurenson, ICF Consulting, Fairfax, Virginia. September 28.

Johnson, T., and J. Capel. 1992. A Monte Carlo Approach to Simulating Residential Occupancy Periods and Its Application to the General U.S. Population. Report No. EPA-450/3-92-011. U.S. Environmental Protection Agency, Research Triangle Park, N.C. August.

Johnson, T., J. Capel, and D. Byrne. 1991. "The Estimation of Commuting Patterns in Applications of the Hazardous Air Pollutant Model (HAPEM)." Paper No. 91-172.6. Presented at the 84th Annual meeting of the Air and Waste Management Association,

Vancouver, British Columbia, June 16 - 21.

Johnson, T., J. Capel, and M. McCoy. 1995. An Analysis of Ten Time/Activity Databases: The Effects of Selected Factors on the Apportionment of Time Among Microenvironments and Breathing Rate Categories, American Petroleum Institute, May.

Johnson, L., M. Lee, J. Laurenson, J. Cohen, and A. Rosenbaum. 2001. "Options for Prototype Algorithms for Estimating Home-to-Work Commuting Patterns." Memorandum to Ted Palma, U.S. Environmental Protection Agency, Research Triangle Park, N.C. May 31.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared for the Risk and Exposure Assessment Division of U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

Johnson, T., L. Wijnberg, and R. Merch. 1987. A Probabilistic Model for Simulating Human Activity Patterns. Report prepared for Strategies and Air Standards Division of the U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. February.

Klepeis, N., A. Tsang, and J. V. Behar. 1995. Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment. National Exposure Research Laboratory. U. S. Environmental Protection Agency, Las Vegas, NV.

Laurenson, J., A. Rosenbaum, M. Huang, and J. Cohen. 2001. "Preliminary Results of Prototype School Commute Module for HAPEM4." Memorandum to Ted Palma, U.S. Environmental Protection Agency, Research Triangle Park, N.C. September 30.

Lee, M., K. Balassiano, M. Ragan, and J. Laurenson. 2001. "Description of Draft Data Bases for Estimating Children's Exposure to Criteria and Hazardous Air Pollutants While Attending School." Memorandum to Ted Palma, U.S. Environmental Protection Agency, Research Triangle Park, N.C. March 30.

Lee, M., J. Laurenson, K. Balassiano, K. Craig, and T. Johnson. 2001. "Draft Outline of Methods and Data to Track Children's Exposure to Criteria and Hazardous Air Pollutants While Attending School." Memorandum to Ted Palma, U.S. Environmental Protection Agency, Research Triangle Park, N.C. January 22.

National Center for Health Statistics. 1990. Vital Statistics of the United States: 1987, Volume II - Mortality. Part A. U.S. Department of Health and Human Services, Hyattsville, Maryland.

Price, P., J. Sample, and R. Strieter. 1992. "Determination of Less-than-Lifetime Exposures to Point Source Emissions." Risk Analysis, Volume 12, pp. 367 - 382.

Price, P.S., Young, J.S., and Chaisson, C.F. (2001). "Assessing aggregate and cumulative pesticide risks using a probabilistic model." Ann. Occup. Hyg. 45: S131-S142.

Robinson, J. P., J. A. Wiley, T. Piazza, K. Garrett, and K. Cirksena. 1989. Activity Patterns of California Residents and Their Implications for Potential Exposure to Pollution. Report No. CARB-A6-177-33. California Air Resources Board, Sacramento, CA.

Rosenbaum, A. 2002. The HAPEM4 User's Guide, Draft Version (Hazardous Air Pollutant Model - Version 4). Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. UPDATE WHEN REPORT IS FINALIZED.

Spier, C. E., D. E. Little, S. C. Trim, T. R. Johnson, W. S. Linn, and J. Hackney. 1992. "Activity Patterns in Elementary and High School Students Exposed to Oxidant Pollution." J Expos Anal Environ Epidemiol, Volume 2, pp. 277 - 293.

Tsang, A. M., and N. E. Klepeis. 1996. Descriptive Statistics from a Detailed Analysis of the National Human Activity Pattern Survey (NHAPS) Data. Report No. 600/R-96/148. U. S. Environmental Protection Agency, Las Vegas, NV.

U.S. Environmental Protection Agency. 1989. Exposure Factors Handbook. Report No. EPA-600/8-89-43. U.S. Environmental Protection Agency, Washington, D.C., July.

U.S. Environmental Protection Agency. 2001. National-Scale Air Toxics Assessment for 1996. Report No. EPA-453/R-01-003. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, N.C. Draft for EPA Science Advisory Board review, January 18.

Wiley, J. A., J. P. Robinson, Y.-T. Cheng, T. Piazza, L. Stork, and K. Pladsen. 1991a. Study of Children's Activity Patterns. Survey Research Center, University of California, Berkeley, CA.

Wiley, J. A., J. P. Robinson, T. Piazza, K. Garrett, K. Cirksena, Y.-T. Cheng, and G. Martin. 1991b. Activity Patterns of California Residents. Survey Research Center, University of California, Berkeley, CA.

Williams, R., J Suggs, J. Creason, C. Rodes, P. Lawless, R. Kwok, R. Zweidinger, and L. Sheldon. 2000. "The 1998 Baltimore Particulate Matter Epidemiology-Exposure Study: Part 2. Personal Exposure Associated with Elderly Study Population." J Expos Anal Environ Epidemiol, Volume 6, No. 6, pp. 533 - 543.

SECTION 6

ESTIMATION OF MICROENVIRONMENTAL CONCENTRATIONS

The NEM-type and HAPEM-type exposure models discussed in this report simulate the movements of people through zones defined by geographic location and microenvironment. Potential geographic locations are typically defined in terms of a set of exposure districts constructed by aggregating census tracts or other units defined by the Bureau of Census (see Subsection 3.1). Potential microenvironments are typically defined as aggregations of the location descriptor codes used in various time/activity studies. The exposure model estimates the exposure of a person in a particular zone as a function of the ambient (outdoor) concentration in the district; the air exchange, deposition, and decay properties of the microenvironment; and selected emission sources associated with the microenvironment.

In recent NEM-type models, a probabilistic mass balance model is used to estimate the pollutant concentration in indoor and in-vehicle microenvironments. Concentrations in source-oriented outdoor microenvironments are estimated by applying a linear relationship to ambient concentrations. HAPEM-type models have typically used linear relationships to estimate all microenvironmental concentrations.

Analysts have defined a set of microenvironments for each exposure model that can account for all time spent by each modeled individual or cohort. Ideally, each of these microenvironments (1) represents a relatively homogenous set of exposure conditions, (2) is not duplicated by any other microenvironment, and (3) can be adequately modeled using available input data.

Section 6.1 describes the general methods used in defining microenvironments in the NEM-type and HAPEM-type exposure models. Section 6.2 provides a description of the general mass balance model. Sections 6.3 and 6.4 describe special versions of this mass balance model that have been incorporated into pNEM/O₃ and pNEM/CO, respectively. Section 6.5 describes an alternative approach (the linear relationship method) used by HAPEM4 to estimate microenvironmental concentrations. Section 6.6 provides estimates for the parameters of the linear relationship used in HAPEM4. References are listed in Section 6.7

6.1 Selection of Microenvironments

As indicated above, a microenvironment is typically defined by one or more of the location descriptor codes appearing in the time/activity database. For example, the analyst might define an "in vehicle" microenvironment as including the codes for car, truck, bus, van, train, airplane, and unspecified vehicle. Alternatively, the analyst may define a separate microenvironment for each of these seven codes or combine the four road vehicles into one microenvironment. Ultimately, each of the individual codes in the time/activity database must be assigned to one and only one microenvironment.

Historically, the developers of the NEM and HAPEM-series models have attempted to achieve the following objectives in developing the set of microenvironments to be included in a particular exposure assessment:

- (1) the set provides a good representation of overall exposure variability,

- (2) the set includes all microenvironments of special interest to the researcher (frequently microenvironments associated with high exposures, particular sources, or special populations),
- (3) each microenvironment can be defined unequivocally by one or more location descriptor codes in the proposed time/activity database,
- (4) each microenvironment can be adequately modeled with existing data concerning air exchange rates, decay rates, sources, and other factors affecting pollutant concentration in the microenvironment,
- (5) the resulting model run times are not excessive (increasing the number of microenvironments tends to increase run time).

Early versions of NEM and pNEM tended to have a small number of microenvironments. For example, the following seven microenvironments were defined in 1993 for pNEM/O3:

1. Indoors - residence - central air conditioning system
2. Indoors - residence - window air conditioning units
3. Indoors - residence - no air conditioning system
4. Indoors - nonresidential locations
5. Outdoors - near road
6. Outdoors - other
7. In vehicle.

Each microenvironment is identified as to a general location (e.g., outdoors) and a specific location (e.g., near road). The indoor - residence classification is further subdivided according to residential air conditioning system. In defining this set of microenvironments, researchers focused on two factors assumed to significantly affect ozone concentrations: air exchange rate and proximity to mobile source emissions. The presence or absence of an air conditioner in the residence was found to significantly affect the probability that residents open windows during periods of high ambient ozone; open windows increase the indoor-outdoor air exchange rate which in turn increases indoor ozone levels. Researchers distinguished between outdoors locations near roads and other outdoor locations because ozone concentrations near roads tend to be reduced through reaction with nitric oxide emitted by motor vehicles.

In 1999, researchers defined 15 microenvironments for Version 2.1 of pNEM/CO (Johnson, Mihlan, LaPointe, et al., 2000). Table 2-5 lists the activity diary locations included in each of the 15 microenvironments together with the methodology used to estimate carbon monoxide concentrations within the microenvironment. The list includes two indoor microenvironments related to residences, seven indoor microenvironments related to nonresidential buildings, three outdoor microenvironments, and three vehicle microenvironments. The majority of these microenvironments are aggregates of two or more of the CHAD location descriptors. Only location descriptors associated with similar average CO exposures were combined in defining each aggregate microenvironment. Researchers determined these similarities through

an analysis of personal CO monitoring data obtained from the Denver activity diary study (Johnson, 1984). Table 2-4 shows the assignment of CHAD location descriptors to microenvironments.

The number of microenvironments used by the HAPEM-series of exposure models has also increased with time. The original point-source version (HAPEM-PS) described by Johnson, Capel, Paul, and Olaguer (1991) included only two microenvironments (indoors and outdoors). The first mobile-source version (HAPEM-MS1) -- developed by Johnson, Paul, and Capel (1992) -- included five microenvironments (indoors - residence, indoors - other locations, outdoors - near road, outdoors - other locations, and inside motor vehicle). The next model in the series (HAPEM-MS2) included 37 microenvironments corresponding to the 37 location descriptors appearing in the time/activity database developed for HAPEM-MS2 (Johnson, McCoy, and Capel, 1993). Subsequent models in the HAPEM series have employed this same set of microenvironments with minor modifications. Table 6-1 lists the 37 microenvironments defined for HAPEM4 as used in the recent NATA (ICF Consulting and TRJ Environmental, 2000).

Version 2.1 of pNEM/CO employs calculation-intensive methods for estimating microenvironmental concentrations with a time resolution of one minute or one hour, depending on the microenvironment. Consequently, the use of 37 microenvironments in pNEM/CO would produce unacceptably long run times for most applications. HAPEM4 can handle the large number of microenvironments because it employs simple linear relationships to estimate microenvironmental concentrations with relatively low time resolution.

6.2 General Version of the Mass Balance Model

The 1992 application of pNEM/CO to Denver marked a milestone in the evolution of the NEM methodology in that it represented the first time that a mass-balance model had been incorporated directly into the NEM methodology (Johnson, Capel, Paul, and Wijnberg, 1992). Researchers updated the mass-balance model for use in Version 2.1 of pNEM/CO (Johnson, Mihlan, LaPointe, et al., 2000). A variation of this mass balance model has been used in various pNEM/O3 applications. In all cases, the structure of the mass balance model was based on the generalized mass-balance model presented by Nagda, Rector, and Koontz (1987). As originally proposed, this model assumed that pollutant concentration decays indoors at a constant rate. For use in pNEM, the Nagda model was revised to incorporate an alternative assumption that the indoor decay rate is proportional to the indoor concentration. The resulting model can be expressed by the differential equation

$$dC_{in}/dt = (1 - F_B)(v)(C_{out}) + S/(cV) - (m)(v)(C_{in}) - (F_d)(C_{in}) - (q)(F)(C_{in})/(cV) \quad (6-1)$$

in which

Table 6-1 Microenvironments Defined for HAPEM4 as Used in NATA.

Identification No.	Microenvironment	
	Specific	General
1	Car	In vehicle
2	Bus	In vehicle
3	Truck	In vehicle
4	Other	In vehicle
5	Public garage	Indoors
6	Parking lot/garage	Outdoors
7	Near road	Outdoors
8	Motorcycle	Outdoors
9	Service station	Indoors
10	Service station	Outdoors
11	Residential garage	Indoors
12	Other repair shop	Indoors
13	Residence - no gas stove	Indoors
14	Residence - gas stove	Indoors
15	Residence - attached garage	Indoors
16	Residential - stove and garage	Indoors
17	Office	Indoors
18	Store	Indoors
19	Restaurant	Indoors
20	Manufacturing facility	Indoors
21	School	Indoors
22	Church	Indoors
23	Shopping mall	Indoors

Identification No.	Microenvironment	
	Specific	General
24	Auditorium	Indoors
25	Health care facility	Indoors
26	Other public building	Indoors
27	Other location	Indoors
28	Not specified	Indoors
29	Construction site	Outdoors
30	Residential grounds	Outdoors
31	School grounds	Outdoors
32	Sports arena	Outdoors
33	Park/golf course	Outdoors
34	Other location	Outdoors
35	Not specified	Outdoors
36	Train/subway	In vehicle
37	Airplane	In vehicle

C_{in}	=	Indoor concentration (units: mass/volume)
F_B	=	Fraction of outdoor concentration intercepted by the enclosure (dimensionless fraction)
F_d	=	Pollutant decay coefficient (1/time)
v	=	Air exchange rate (1/time)
C_{out}	=	Outdoor concentration (mass/volume)
S	=	Indoor generation rate (mass/time)
cV	=	Effective indoor volume where c is a dimensionless fraction (volume)
m	=	Mixing factor (dimensionless fraction)
q	=	Flow rate through air-cleaning device (volume/time)
F	=	Efficiency of the air-cleaning device (dimensionless fraction).

To simplify the notation, let $F_p = 1 - F_B$ in which F_p is the “penetration factor.” If one assumes that c and m are each equal to 1 (indicating perfect mixing) and that $F = 0$ (indicating the absence of an air cleaning device), the resulting differential equation is

$$dC_{in}/dt = (F_p)(v)(C_{out}) + S/V - (v')(C_{in}) \quad (6-2)$$

in which $v' = v + F_d$. It can be shown that this equation has the following exact solution:

$$C_{in}(t) = k_1 C_{in}(t - \Delta t) + k_2 C_{out}(t - \Delta t) + k_3 \quad (6-3)$$

where

$$k_1 = e^{-v'\Delta t}, \quad (6-4)$$

$$k_2 = (F_p)(v/v')(1 - k_1), \quad (6-5)$$

$$k_3 = (S)(1 - k_1)/(v'V), \quad (6-6)$$

Δt is a fixed time interval, and C_{out} is assumed to be constant over the interval t to $t + \Delta t$.

In typical pNEM applications, Δt is assumed to equal one hour. When $\Delta t = 1$ hour, the average indoor pollutant concentration of hour h [$CAVG_{in}(h)$] can be calculated by the expression

$$CAVG_{in}(h) = (a_1)[C_{in}(h - 1)] + (a_2)[CAVG_{out}(h)] + a_3 \quad (6-7)$$

in which $C_{in}(h - 1)$ is the indoor concentration at the end of the preceding hour and $CAVG_{out}(h)$ is the average outdoor concentration during hour h. The other variables appearing in Equation 6-7 are defined by the following equations:

$$a_1 = z(h) \quad (6-8)$$

$$a_2 = (F_p)(v/v')[1 - z(h)] \quad (6-9)$$

$$a_3 = (S)[1 - z(h)]/(v'V) \quad (6-10)$$

$$z(h) = (1 - e^{-v'})/v' \quad (6-11)$$

The instantaneous indoor concentration at the end of a particular hour h [i.e., $C_{in}(h)$] is calculated by the equation

$$C_{in}(h) = k_1 C_{in}(h - 1) + k_2 CAVG_{out}(h) + k_3 \quad (6-12)$$

in which $C_{in}(h - 1)$ is the instantaneous indoor concentration at the end of hour h - 1, $CAVG_{out}(h)$ is the average outdoor concentration during hour h, and

$$k_1 = e^{-v'} \quad (6-13)$$

$$k_2 = (F_p)(v/v')(1 - k_1) \quad (6-14)$$

$$k_3 = (S)(1 - k_1)/(v'V). \quad (6-15)$$

The same set of equations can be used for $\Delta t = 1$ minute, with each hourly index (h) replaced by a corresponding minute index (m).

6.3 Mass Balance Model Used in Recent Applications of pNEM/O3

During the mid-1990's, EPA applied a version of pNEM/O3 to three different population groups in selected cities:

- (1) the general population of urban residents (Johnson, Capel, and McCoy, 1996),
- (2) outdoor workers (Johnson, Capel, McCoy, and Mozier, 1996), and
- (3) children who tend to spend more time outdoors than the average child (Johnson, Capel, Mozier, and McCoy, 1996).

This version of pNEM/O3 used a probabilistic mass balance model to determine ozone concentrations over one-hour periods in indoor and in-vehicle microenvironments. The model was designed to operate under the following conditions:

1. There are no indoor sources of ozone.

2. The indoor ozone concentration at the end of the preceding hour is specified.
3. The outdoor ozone concentration during the clock hour is constant at a specified value.
4. The air exchange rate during the clock hour is constant at a specified value.
5. Ozone decays at a rate that is proportional to the indoor ozone concentration. The proportionality factor is constant at a specified value.

The mass balance model employed in these calculations was based on the generalized mass balance model described in Section 6.2, hereafter referred to as the Nagda model. As originally proposed, this model assumed that pollutant concentration decays indoors at a constant rate. For use in pNEM/O₃, the Nagda model was revised to incorporate the alternative assumption that the indoor decay rate is proportional to the indoor concentration. The Nagda model was further revised to incorporate ozone-specific assumptions concerning various parameter values suggested by Weschler, Shields, and Naik (1992) and others.

Subsection 6.3.1 presents the theoretical basis for the pNEM/O₃ mass balance model and the principal model assumptions. Subsection 6.3.2 describes the algorithms which were used to generate hourly values of ozone for the indoor and in-vehicle microenvironments. Subsection 6.3.3 presents the procedure used to determine air exchange rate for the mass balance model. An algorithm for simulating the opening and closing of windows is described in Subsection 6.3.4.

The reader should note that many of the algorithms and distributions discussed in Section 6.3 have been improved or replaced in later pNEM-type models, particularly in Version 2.1 of pNEM/CO.

6.3.1 Theoretical Basis and Assumptions

The general mass balance model used in the pNEM series of models is expressed by Equation 6-1. If one assumes that c and m are each equal to 1 (indicating perfect mixing) and that $F = 0$ (indicating the absence of an air cleaning device), one obtains the model expressed by Equation 6-2:

$$dC_{in}/dt = (F_p)(v)(C_{out}) + S/V - (v')(C_{in})$$

A steady-state version of this model can be developed by solving Equation 6-2 under the conditions that

$$dC_{in}/dt = 0 \tag{6-16}$$

and C_{out} is constant. In this case, the mass balance equation is

$$0 = (F_p)(v)(C_{out}) + S/V - (v')(C_{in}) \tag{6-17}$$

which can be rearranged as

$$C_{in} = (F_p)(v/v')(C_{out}) + S/(v'V). \quad (6-18)$$

The ratio of indoor concentration to outdoor concentration is

$$C_{in}/C_{out} = (F_p)(v/v') + S/(v'VC_{out}). \quad (6-19)$$

Weschler, Shields, and Naik (1992) developed a steady-state equation for the indoor/outdoor ratio that is expressed in their notation as

$$I/O = E_x/[E_x + k_d(A/V)], \quad (6-20)$$

in which I = indoor concentration, O = outdoor concentration, E_x = air exchange rate, k_d = deposition velocity, A = surface area, and V = volume. (This model is hereafter referred to as “Weschler’s model”.) With respect to the general mass balance model described in Section 6.2, Weschler’s model implies that there are no indoor sources ($S = \text{zero}$), no air cleaning devices ($F = \text{zero}$), the penetration factor is unity ($F_p = 1$), $c = 1$, and $m = 1$. Under these conditions, the general mass balance model can be expressed as

$$dC_{in}/dt = (v)(C_{out}) - (v + F_d)(C_{in}) \quad (6-21)$$

and Equation 6-19 becomes

$$C_{in}/C_{out} = v/(v + F_d). \quad (6-22)$$

Weschler’s model (Equation 6-20) and Equation 6-22 are equivalent if the following substitutions are made:

$$C_{in} = I \quad (6-23)$$

$$C_{out} = O \quad (6-24)$$

$$v = E_x \quad (6-25)$$

$$F_d = k_d(A/V). \quad (6-26)$$

Equation 6-26 is a particularly useful relationship, as Weschler has identified a number of studies that suggest that $k_d(A/V)$ is relatively constant from building to building. He suggests that $1.0 \times 10^{-3} \text{ sec}^{-1}$ is a good general estimate of this quantity.

Weschler, Shields, and Naik (1992) present 14 estimates of $k_d(A/V)$ based on data obtained from specific studies. Nine of these values are based on the observed first-order decay of ozone in isolated rooms. The remaining five values are based on reported I/O values and air exchange rates. Table 6-2 presents means and standard

deviations for the first nine estimates, for the last five estimates, and for all 14 estimates. Two-sided 95 percent confidence intervals for the means are also provided.

Table 6-2. Means, Standard Deviations, And Confidence Intervals For Estimates of $k_d(A/V)$ Provided by Weschler, Shields, and Naik (1992)

Parameter	Source of $k_d(A/V)$ estimate		
	Observed first-order decay	Reported I/O values	All
Sample size	9	5	14
Mean, sec^{-1}	1.133×10^{-3}	1.098×10^{-3}	1.121×10^{-3}
Standard deviation, sec^{-1}	0.447×10^{-3}	0.143×10^{-3}	0.374×10^{-3}
Two-sided 95% confidence interval, sec^{-1}	$(0.789, 1.477) \times 10^{-3}$	$(0.920, 1.276) \times 10^{-3}$	$(0.906, 1.335) \times 10^{-3}$

The values in Table 6-2 can be converted to units of h^{-1} by multiplying each value by 3600. Expressed in these units, the mean and standard deviation for the 14 estimates are 4.04 h^{-1} and 1.35 h^{-1} , respectively. A normal distribution with these parameters was assumed to represent the distribution of F_d values for the non-vehicle indoor microenvironments. The value of F_d was not permitted to be less than 1.44 h^{-1} or more than 8.09 h^{-1} . The lower bound was based on the smallest value cited by Weschler, Shields, and Naik (1992) which was measured in a stainless steel room. The upper bound corresponds to the 99.87 percentile (i.e., $z = 3$) of a normal distribution with mean equal to 4.04 and standard deviation equal to 1.35. The largest value cited by Weschler, Shields, and Naik (1992) was 7.2 h^{-1} .

The mass balance model was also used to simulate ozone concentrations for the in-vehicle microenvironment. Ideally, the in-vehicle microenvironment would have been represented by a distribution of F_d values based on ozone decay rates measured in a representative sample of motor vehicles. Because of the scarcity of research concerning ozone decay rates in motor vehicles, researchers were not able to develop such a distribution. Instead, a point estimate of 72.0 h^{-1} was assumed for the F_d of the in-vehicle microenvironment. This value was derived by Hayes (1989) from an analysis of data for one vehicle presented by Petersen and Sabersky (1975). Hayes has used this value in applications of the PAQM exposure model (Hayes, 1989).

The use of a point estimate based on a single motor vehicle is likely to produce a bias in the ozone concentrations estimated for the in-vehicle microenvironment. The direction of this bias is uncertain.

6.3.2 Simulation of Microenvironmental Ozone Concentrations

Consistent with the theoretical considerations discussed in Subsection 6.3.1, the following equation was used to estimate the hourly ozone concentration in a particular indoor or in-vehicle microenvironment during hour h :

$$\text{CAVG}_{\text{in}}(h) = (a_1)[C_{\text{in}}(h - 1)] + (a_2)[\text{CAVG}_{\text{out}}(h)] \quad (6-27)$$

in which $\text{CAVG}_{\text{in}}(h)$ is the average indoor ozone concentration during hour h , $C_{\text{in}}(h - 1)$ is the indoor concentration at the end of the preceding hour, and $\text{CAVG}_{\text{out}}(h)$ is the average outdoor concentration during hour h . The other variables appearing in Equation 6-27 are defined by the following equations:

$$a_1 = z(h) \quad (6-28)$$

$$a_2 = (v/v')[1 - z(h)] \quad (6-29)$$

$$z(h) = (1 - e^{-v'})/v' \quad (6-30)$$

$$v' = v + F_d \quad (6-31)$$

The instantaneous indoor concentration at the end of a particular hour h [i.e., $C_{\text{in}}(h)$] is calculated by the equation

$$C_{\text{in}}(h) = k_1 C_{\text{in}}(h - 1) + k_2 \text{CAVG}_{\text{out}}(h) \quad (6-32)$$

in which $C_{\text{in}}(h - 1)$ is the instantaneous indoor concentration at the end of hour $h - 1$, $\text{CAVG}_{\text{out}}(h)$ is the average outdoor concentration during hour h , and

$$k_1 = e^{-v'} \quad (6-33)$$

$$k_2 = (v/v')(1 - k_1) \quad (6-34)$$

and v' is determined by Equation 6-31.

The following algorithm was used to generate a sequence of hourly-average ozone concentrations for each combination of microenvironment and district.

1. Go to first/next day.
2. Select value of air exchange rate for day from appropriate distribution or use point estimate. If microenvironment is residential, select one air exchange value for hours when windows are open and one for hours when windows are closed. If microenvironment is a nonresidential building or vehicle, then one air exchange rate is used for all hours of the day.
3. Select value of decay rate (F_d) for day from appropriate distribution or use point estimate. If microenvironment is non-vehicular enclosure, select value of F_d from normal distribution with mean = 4.04 h^{-1} and standard deviation = 1.35 h^{-1} . Value is not permitted to be less than 1.44 h^{-1} or more than 8.09 h^{-1} . If microenvironment is "in vehicle", use point estimate of 72.0 h^{-1} .
4. Go to first/next clock hour.

5. If microenvironment is residential, use supplementary window algorithm to determine window status for current hour (open or closed). Window status determines which air exchange rate determined in Step 2 applies to current hour.
6. Use Equation 6-27 to determine ozone concentration for current hour based on air exchange rate specified for hour, outdoor ozone concentration during hour, and ozone concentration at end of preceding hour.
7. Use Equation 6-32 to determine instantaneous ozone concentration at end of current hour based on air exchange rate specified for hour, outdoor ozone concentration during hour, and instantaneous ozone concentration at end of preceding hour. This value is saved for input into Equation 6-27 during the next hour.
8. If end of day, go to Step 1. Otherwise, go to Step 4.

Step 2 requires the random selection of an air exchange rate from a specified distribution. Four enclosure categories were established for this purpose.

- Residential buildings - windows open
- Residential buildings - windows closed
- Nonresidential buildings
- Vehicles.

A survey of the scientific literature determined that there were sufficient data available to define distributions for only two of the four enclosure categories: "residential building - windows closed" or "nonresidential building". In each case, a two-parameter lognormal distribution was found to provide a good fit to the data. Point (single-valued) estimates were developed for the remaining two enclosure categories.

Each of the two lognormal distributions was defined by the expression

$$\text{AER} = \text{GM} \times \text{GSD}^Z \quad (6-35)$$

where AER is the air exchange rate, GM is the geometric mean, and GSD is the geometric standard deviation. The values for GM and GSD were determined by fitting lognormal distributions to representative data sets (Subsection 6.3.3). A value of AER was selected at random from a particular lognormal distribution by randomly selecting a value of Z from the unit normal distribution [N(0,1)] and substituting it into Equation 6-35. Table 6-3 lists the values of GM and GSD for the two lognormal distributions and the values of the point estimates.

The distributions used to determine AER are discussed in more detail in Subsection 6.3.3. Subsection 6.3.4 provides a description of the algorithm used to determine window status in the residential microenvironments (Step 5).

Table 6-3. Distributions of Air Exchange Rate Values Used in the pNEM/O3 Mass

Balance Model

Enclosure category	Air exchange rate distribution
Residential building-windows closed	Lognormal distribution <ul style="list-style-type: none"> ◦ Geometric mean = 0.53 ◦ Geometric standard deviation = 1.704 ◦ Lower bound = 0.063 ◦ Upper bound = 4.47
Residential building-windows open	Point estimate: 6.4
Nonresidential building	Lognormal distribution <ul style="list-style-type: none"> ◦ Geometric mean = 1.285 ◦ Geometric standard deviation = 1.891 ◦ Lower bound = 0.19 ◦ Upper bound = 8.69
Vehicle	Point estimate: 36

6.3.3 Air Exchange Rate Distributions

A review of the scientific literature relating to air exchange rates identified 31 relevant references (list available on request). Of these, only a few were found to contain sufficient data to construct a distribution of air exchange rates relating to a particular building type such as residence or office. The two most useful studies were conducted by Grimsrud, Sherman, and Sondregger (1982) and by Turk, Grimsrud, Brown, et al. (1989).

Residential Locations

Grimsrud, Sherman, and Sondregger (1982) measured AER's in 312 residences. Reported AER values ranged from 0.08 to 3.24. Researchers analyzed these data to determine which of two distributions (normal versus lognormal) better characterized the data. The lognormal distribution was found to yield a better fit, as the data were highly skewed. The fitted lognormal parameters were geometric mean = 0.53 and geometric standard deviation = 1.704. This distribution was used in pNEM/03 to represent the distribution of AER's in residences with windows closed. Upper and lower limits of 4.47 and 0.063 air changes per hour were established to prevent the selection of unusually extreme values of AER. These limits corresponded to the substitution of $Z = 4$ and $Z = -4$ in Equation 6-35 when $GM = 0.53$ and $GSD = 1.704$. The upper bound was 38 percent larger than the largest reported AER (3.24). The lower bound was 21 percent smaller than the smallest reported AER (0.08).

No comparable data bases were identified which were considered representative of residences where windows are open. Hayes (1989) has used 6.4 h^{-1} as the AER value for open windows in applications of the PAQM model. This value was based on an analysis by Hayes and Lundberg (1985) of a hypothetical building plan with an assumed "orifice coefficient." Orifice coefficient is defined as the ratio of the equivalent

area of all openings in a building to the building's volume. In support of this approach, Hayes cites a report by Moschandreas, Zabransky, and Peltan (1981) which suggests that infiltration is proportional to a building's orifice coefficient.

At the time of the pNEM/O₃ analysis described here, researchers considered Hayes's estimate to be the best available estimate of AER for residences with windows open. Consequently, the AER for residences with windows open was treated as a point estimate (6.4 h⁻¹). Note that the use of an AER estimate representing a single set of conditions is likely to produce a bias in the ozone concentrations estimated for this microenvironment. The direction of this bias is uncertain. The reader is referred to Subsection 6.4.3.2 for a discussion of the AER distributions used for residences in Version 2.1 of pNEM/CO.

Nonresidential Locations

Turk, Grimsrud, Brown, et al. (1989) measured AER's in 40 public buildings identified as schools (n = 7), offices (n = 25), libraries (n = 3), and multipurpose buildings (n = 5). The minimum reported AER was 0.3; the maximum was 4.1. Researchers fit normal and lognormal distributions to the data for all 40 buildings and found that the lognormal distribution produced a slightly better fit, although it had a tendency to over-predict high values. The fitted lognormal parameters were geometric mean = 1.285 and geometric standard deviation = 1.891.

The buildings can be grouped as offices (n = 25) and non-offices (n = 15). Lognormal fits to these data sets yielded geometric means and standard deviations of 1.30 and 1.93 for offices and 1.27 and 1.87 for non-offices. Analysts performed a two-sample t test on the two data sets and found no significant difference in the means or standard deviations of the data. Consequently, a single lognormal distribution (geometric mean = 1.285, geometric standard deviation = 1.891) was used in pNEM/O₃ for all nonresidential buildings. To prevent the over-prediction of high AER values, an upper bound of 8.69 was established. This value results when Z = 3 is substituted into Equation 6-35 with GM = 1.285 and GSD = 1.891. This value is over twice the largest AER value (4.1) reported for the 40 buildings and corresponds to the 99.87 percentile of the specified lognormal distribution. A lower bound of 0.19 was also established. This value corresponds to a Z value of -3 and represents the 0.13 percentile of the lognormal distribution.

Analysts considered the AER data obtained from Turk, Grimsrud, Brown, et al. (1989) to be generally representative of buildings with closed windows. Consequently, the lognormal AER distribution derived from these data may not be applicable to non-residential buildings which are ventilated by open windows. As comparable data were not available for non-residential buildings with open windows at the time of this analysis, researchers applied the lognormal AER distribution for closed windows to all non-residential buildings. This approach is likely to under-estimate the ozone exposures of people who frequently occupy buildings with open windows.

In Vehicle Locations

A point estimate of 36 air changes per hour was used for in-vehicle locations. This value was obtained from Hayes (1991) based on his analysis of data for a single vehicle presented by Peterson and Sabersky (1975). Hayes notes that the greater AER observed in vehicles, even with the windows closed, is due to wind effects on the moving vehicle and the "leakiness" of typical automobiles.

At the time of the pNEM/O₃ analysis described here, researchers considered Hayes's estimate to be the best available estimate of AER for the in-vehicle microenvironment. Consequently, in-vehicle AER was treated as a point estimate (36 h⁻¹). It should be noted that the use of an AER estimate representing a single set of conditions is likely to produce a bias in the ozone concentrations estimated for this microenvironment. The direction of this bias is uncertain. A more advanced method for estimating AER in vehicles can be found in Subsection 6.4.3.

6.3.4 Window Status Algorithm

The opening and closing of windows in the three residential microenvironments defined for pNEM/O₃ were simulated by an algorithm which specified a window status (open or closed) for each clock hour. The algorithm consisted of the following eight-step procedure.

1. Identify air conditioning system associated with cohort (central, window units, none).
2. Go to first/next day.
3. Determine average temperature for day from supplementary file. Identify temperature range which contains this value (below 32, 32 to below 63, 63 to 75, above 75).
4. Select random number between zero and 1. Compare random number with probabilities listed in Table 6-4 for specified air conditioning system and temperature range. Determine window status for day. If day status is "windows open all day" or "windows closed all day", set window status for all clock hours of day as indicated and go to Step 2. If day status is "windows open part of day", go to Step 5.
5. Go to first/next clock hour.
6. Determine window status of preceding clock hour.
7. Select random number between zero and 1. Compare random number with probabilities listed in Table 6-5, 6-6, or 6-7 for specified air conditioning system, clock hour, temperature range, and window status for preceding hour. If the random number is less than the specified probability, the window will be open during the clock hour. Otherwise, the window will be closed.

8. If end of day, go to Step 2. Otherwise, go to Step 5.

This algorithm assigns each day to one of three categories: (1) windows closed all day, (2) windows open all day, and (3) windows open part of day. These assignments are made according to the air conditioning system associated with the cohort and the average temperature of the day. If the day assignment is "windows open part of day", the algorithm assigns window status on an hourly basis for each of the 24 clock hours in the day. These hourly assignments are made according to the (1) cohort's air conditioning system, (2) clock hour, (3) average temperature for the day, and (4) window status of the preceding hour. Both the daily and hourly assignments are made probabilistically by comparing random numbers to the probabilities that the specified window status will occur under the stated conditions.

The window status probabilities listed in Tables 6-4, 6-5, 6-6, and 6-7 were developed through a statistical analysis of data on window openings obtained from the CADS (Johnson, 1987). This analysis indicated that air conditioning system, temperature, clock hour, and window status of preceding hour were statistically significant factors affecting window status.

6.4 Mass Balance Model Used in Version 2.1 of pNEM/CO

The 1992 application of pNEM/CO to Denver represented the first time that a mass-balance model had been incorporated directly into the NEM methodology (Johnson, Capel, Paul, and Wijnberg, 1992). In 1999, researchers updated the mass-balance model for use in Version 2.1 of pNEM/CO (Johnson, Mihlan, LaPointe, et al., 2000). This section provides an overview of the pNEM/CO mass-balance model together with descriptions of the algorithms used in the model to estimate air exchange, emissions from gas stoves, and emissions from passive smoking. It also describes the data used for the input parameters to the mass-balance model.

6.4.1 Overview of the Model

The pNEM/CO methodology includes a mass-balance model which is used to estimate CO concentrations when a cohort is assigned to an indoor or motor vehicle microenvironment. The model used in Version 2.1 of pNEM/CO was based on the generalized mass-balance model presented in Equation 6-1 in Section 6.2. In developing the CO model, researchers recognized that CO is a nonreactive pollutant and assumed that (1) that the enclosure does not intercept any of the CO as it moves indoors, (2) that the CO does not decay once it enters the enclosure, and (3) that no CO is removed by air-filtration devices. Under these assumptions, the parameters F_B , F_d , and F in Equation 6-1 would be set equal to zero. If the additional assumptions are made that c and m are each equal to 1, the resulting differential equation is

$$dC_{in}/dt = (v)(C_{out}) + S/V - (v)(C_{in}) \quad (6-36)$$

It can be shown that this equation has the following exact solution:

$$C_{in}(t) = k_1 C_{in}(t - \Delta t) + k_2 C_{out}(t - \Delta t) + k_3 \quad (6-37)$$

TABLE 6-4. PROBABILITY OF WINDOW STATUS FOR DAY BY AIR CONDITIONING SYSTEM AND TEMPERATURE RANGE

Air conditioning system	Temperature range, °F	Probability of window status for day		
		Closed all day	Open all day	Open part of day
Central	Below 32	1.000	0	0
	32 to 62	0.851	0.009	0.140
	63 to 75	0.358	0.343	0.299
	Above 75	0.633	0.167	0.200
Room units	Below 32	1.000	0	0
	32 to 62	0.734	0.028	0.238
	63 to 75	0.114	0.505	0.381
	Above 75	0.160	0.380	0.460
None	Below 32	1.000	0	0
	32 to 62	0.812	0.011	0.177
	63 to 75	0.095	0.672	0.233
	Above 75	0.016	0.823	0.161

TABLE 6-5. PROBABILITY OF WINDOWS BEING OPEN BY CLOCK HOUR, TEMPERATURE RANGE, AND WINDOW STATUS OF PRECEDING HOUR (PH) FOR RESIDENCES WITH CENTRAL AIR CONDITIONING

Clock hour	Probability of windows being open					
	32°F to 62°F		63°F to 75°F		Above 75°F	
	PH=open	PH=closed	PH=open	PH=closed	PH=open	PH=closed
1-3	1.000	0.000	0.978	0.011	0.986	0.020
4-6	1.000	0.005	0.989	0.000	1.000	0.017
7-9	0.837	0.038	0.932	0.074	0.961	0.094
10-12	0.679	0.126	0.865	0.235	0.860	0.174
13-15	0.857	0.149	0.912	0.240	0.923	0.263
16-18	0.932	0.131	0.935	0.161	0.912	0.000
19-21	0.646	0.043	0.892	0.136	0.893	0.047
22-24	0.811	0.036	0.913	0.101	0.909	0.066

TABLE 6-6. PROBABILITY OF WINDOWS BEING OPEN BY CLOCK HOUR, TEMPERATURE RANGE, AND WINDOW STATUS OF PRECEDING HOUR (PH) FOR RESIDENCES WITH WINDOW AIR CONDITIONING UNITS

Clock hour	Probability of windows being open					
	32°F to 62°F		63°F to 75°F		Above 75°F	
	PH=open	PH=closed	PH=open	PH=closed	PH=open	PH=closed
1-3	0.970	0.006	0.947	0.007	0.974	0.010
4-6	0.975	0.000	0.994	0.016	0.989	0.017
7-9	0.864	0.040	0.934	0.101	0.989	0.092
10-12	0.929	0.121	0.917	0.303	0.849	0.351
13-15	0.860	0.244	0.969	0.400	0.819	0.152
16-18	0.859	0.103	0.956	0.125	0.930	0.043
19-21	0.684	0.063	0.925	0.176	0.902	0.056
22-24	0.919	0.042	0.851	0.064	0.865	0.121

TABLE 6-7. PROBABILITY OF WINDOWS BEING OPEN BY CLOCK HOUR, TEMPERATURE RANGE, AND WINDOW STATUS OF PRECEDING HOUR (PH) FOR RESIDENCES WITH NO AIR CONDITIONING SYSTEM

Clock hour	Probability of windows being open					
	32°F to 62°F		63°F to 75°F		Above 75°F	
	PH=open	PH=closed	PH=open	PH=closed	PH=open	PH=closed
1-3	1.000	0.015	0.974	0.031	1.000	0.000
4-6	1.000	0.000	1.000	0.000	1.000	0.000
7-9	0.950	0.000	0.868	0.057	1.000	0.000
10-12	0.889	0.200	0.933	0.400	0.875	0.500
13-15	0.923	0.130	1.000	0.286	0.917	0.000
16-18	0.848	0.200	0.964	0.000	0.818	0.667
19-21	0.609	0.067	0.909	0.500	1.000	0.200
22-24	0.684	0.043	0.800	0.167	0.769	0.500

where

$$k_1 = e^{-v\Delta t} \quad (6-38)$$

$$k_2 = 1 - e^{-v\Delta t} \quad (6-39)$$

$$k_3 = (S)(1 - e^{-v\Delta t})/(vV) \quad (6-40)$$

and Δt is a fixed time interval. In Version 2.1 of pNEM/CO, Δt is either one hour or one minute, depending on the time resolution required by a particular modeling algorithm.

When $\Delta t = 1$ hour, the average indoor pollutant concentration of hour h [$CAVG_{in}(h)$] can be calculated by the expression

$$CAVG_{in}(h) = (a_1)[C_{in}(h - 1)] + (a_2)[CAVG_{out}(h)] + a_3 \quad (6-41)$$

where $C_{in}(h - 1)$ is the indoor concentration at the end of the preceding hour and $CAVG_{out}(h)$ is the average outdoor concentration during hour h . The other variables appearing in Equation 6-41 are defined by the following equations:

$$a_1 = z(h) \quad (6-42)$$

$$a_2 = 1 - z(h) \quad (6-43)$$

$$a_3 = (S)[1 - z(h)]/(vV) \quad (6-44)$$

$$z(h) = (1 - e^{-v})/v \quad (6-45)$$

The instantaneous indoor concentration at the end of a particular hour h [i.e., $C_{in}(h)$] is calculated by the equation

$$C_{in}(h) = k_1 C_{in}(h - 1) + k_2 CAVG_{out}(h) + k_3 \quad (6-46)$$

in which $C_{in}(h - 1)$ is the instantaneous indoor concentration at the end of hour $h - 1$, $CAVG_{out}(h)$ is the average outdoor concentration during hour h , and

$$k_1 = e^{-v} \quad (6-47)$$

$$k_2 = 1 - e^{-v} \quad (6-48)$$

$$k_3 = (S)(1 - e^{-v})/(vV). \quad (6-49)$$

The same set of equations can be used for $\Delta t = 1$ minute, with each hourly index (h) replaced by a corresponding minute index (m).

To achieve reasonable run times, the hour version of the mass-balance model was used to estimate hour-by-hour CO concentrations in indoor microenvironments from sources other than environmental tobacco smoke (ETS). In addition, a special version of the hour mass-balance model was used to estimate hour-by-hour CO

concentrations in restaurants and bars in areas which permit smoking. The minute version of the mass-balance model was used to estimate the minute-by-minute contribution of ETS in the indoors - residence microenvironment. It was also used to estimate total exposure in the automobile microenvironment. Subsection 6.4.2 describes these applications in more detail.

The majority of parameters included in each version of the mass-balance model were treated as probabilistic variables, in that the values for each parameter were randomly selected from appropriate distributions as they were required by the model. Table 6-8 indicates the selection frequency applied to each parameter (annual - i.e., one value per cohort; seasonal; daily; hourly; or by trip when in a vehicle). The parameters will be defined as they appear throughout this section.

6.4.2 Application of the Mass-balance Model to Specific Microenvironments

Table 2-6 lists the 15 microenvironments defined for Version 2.1 of pNEM/CO. As discussed above and in Section 2.4, the mass-balance model was used to estimate CO levels (partial or total) in 12 of these microenvironments. Subsections 6.4.2.1 through 6.4.2.6 describe the specific methodology used for each microenvironment.

6.4.2.1 Indoors - Residence

In simulating CO concentrations in this microenvironment, the hour version of the mass-balance model was used to estimate the hourly average contribution of CO from outdoor sources and from indoor gas stoves. The minute version of the mass-balance model was used to estimate a minute-by-minute contribution from ETS which was averaged over each exposure event in which passive smoking occurred. If more than one passive smoking event occurred in sequence, the minute mass-balance model was run for the entire time period during which passive smoking occurred. The CO contribution from passive smoking was assumed to be zero during all non-smoking periods, and the minute mass-balance model was not run during these periods. This approach, which significantly reduced model run-time, may produce estimates that are biased low, as it does not account for CO from smoking that lingers indoors after smoking stops. Analysts assumed that the effect was small enough to disregard.

In running the hour version of the mass-balance model, the value of $CAVG_{out}$ for a particular hour was set equal to the value for outdoor concentration determined for that hour by the algorithm described in Subsection 2.4.1. A value for air exchange rate (v) was selected for each season from the lognormal distributions specified in Subsection 6.4.3. This same air exchange rate was also used in the minute mass-balance model when estimating the contribution of passive smoking, with the value expressed in air changes per minute rather than air changes per hour.

In applying the hour mass-balance model to residences, the S parameter was assumed to represent CO emissions from a single gas stove in the residence. In applying the minute mass-balance model, the S parameter represented CO emissions from passive smoking. In both versions, the V parameter was assumed to represent the total volume of the residence.

Cohorts with gas stoves were randomly identified as having stoves with either (1) continuously operating pilot lights or (2) electronic ignitions (i.e., no emissions from pilot

Table 6-8. Selection Frequencies for Probabilistic Parameters Used in Mass-balance Model.

Version of mass-balance model	Microenvironment	Probabilistic parameter	Frequency of Selection
Hour	Indoors - residence	Open window air exchange rate (AER)	Daily (function of window status)
		Closed window AER	Seasonal
		AC type (central, window, none)	Annual
		Window status (open/closed)	Daily
		Enclosed volume (V)	Annual
		Burner status (on/off)	Hourly
		Burner emission factor (EFBURN)	Annual
		Burner annual fuel use (AUB)	Annual
		Pilot light emission factor (EFPILOT)	Annual
		Pilot light annual fuel use (AUP)	Annual
		Pilot light status (yes/no)	Annual
	Restaurants and bars	Air exchange rate (v)	Annual
		Cigarette emission rate ($CO_{cigarette}$)	Hourly
		Normalized ventilation rate (NVR)	Annual
	Other indoor ME's	Air exchange rate (v)	Annual
	Mass transit vehicles	Air exchange rate (v)	Daily
Minute	Indoor - residence (for ETS increment)	Cigarette emission rate ($CO_{cigarette}$)	Annual
		Air exchange rate (v)	^a
		Volume (V)	Annual
	Cars and trucks	Volume (V)	Annual
		$CO_{cigarette}$	Annual
		Air conditioning availability (Step 3 in Table 6-9)	Annual
		Vent status (Step 2 in Table 6-11)	Trip
		Speed (Step 3 in Table 6-11)	Trip
		Z value for AER (Step 4 in Table 6-11)	Trip

^aUse same value as selected for hour mass-balance model.

lights), based on the estimated proportions of gas stove homes with and without electronic ignition in the Denver and Los Angeles study areas. Subsection 8.1.2 describes the methods used to simulate the CO contribution from continuously operating pilot lights.

The probabilistic algorithm described in Section 8.1.1 was used to simulate the operation of gas stove burners. Briefly, burner operation was assumed to occur in discrete "burner operation periods" (BOPs) of 60 minutes duration during normal dinner hours and of 30 minutes duration at other times. No more than one BOP was permitted to occur within a given clock hour, and each BOP began and ended with the same clock hour. A Monte Carlo process was used to randomly assign BOPs to clock hours throughout the year based on a table listing the probability of a BOP occurring within each hour of a typical day. This table was developed from an analysis of gas stove use patterns observed during the Denver Personal Monitoring Study (Johnson, 1984).

Other probabilistic algorithms were used to determine values of annual fuel usage, pilot light emission rate (if required), and burner emission rate for each cohort with gas stoves. Section 8.1.2 describes these algorithms. The simulated burner and pilot light emissions were summed for each clock hour and presented to the mass-balance model as an hourly average value for S . The residential volume (V) receiving the CO emissions was determined for each cohort by selecting values from a distribution representing the housing stock of Denver or Los Angeles, as appropriate.

The probabilistic algorithm described in Subsection 8.2.1 was used to estimate emission rate per cigarette during passive smoking events. The algorithm assumed that one smoker was present and that two cigarettes were smoked per hour.

6.4.2.2 Restaurants and Bars

Equation 6-41 was used to estimate hourly average values of CO for restaurants and bars in both study areas. ETS was considered to be the only potential indoor source of CO in these microenvironments. Consequently, the a_3 term in Equation 6-41 was used solely to account for the effect of passive smoking. Because smoking is prohibited in Los Angeles bars and restaurants, a_3 was set to zero when pNEM/CO was applied to Los Angeles. Passive smoking was assumed to occur continuously in Denver bars and restaurants, as local regulations permit smoking in these locations. In Denver applications, pNEM/CO calculated a_3 using an alternative to Equation 6-44 which better utilized existing databases. The alternative equation and applicable parameter distributions are presented in Subsection 8.2.3.

6.4.2.3 Other Indoor Microenvironments

The hour mass-balance model was used to estimate hourly average CO concentrations for each of these indoor microenvironments. In each case, analysts assumed that either (1) local regulations did not permit smoking in the microenvironments or (2) the CO contribution from passive smoking was insignificant. Analysts further assumed that these microenvironments contained no other significant indoor CO sources. Consequently, the indoor CO emission rate (S) was set at zero for each application. With $S = 0$, no value was required for enclosure volume (V). An air exchange value was selected for each combination of cohort and microenvironment

from an appropriate distribution (see Subsection 6.4.3.2).

6.4.2.4 Automobiles and Trucks

Microenvironment No. 12 was defined as including automobiles and other non-truck passenger vehicles (vans, sport utility vehicles, etc.). Trucks were included in Microenvironment No. 13. In modeling these two microenvironments, the minute mass-balance model was used to estimate CO concentrations as a function of outside concentration, air exchange rate, vehicle volume, and ETS. The resulting one-minute CO values were averaged over the duration of each exposure event occurring in the microenvironment to determine the CO concentration to be applied to the event. The 14-step procedure presented in Table 6-9 was used to model the trip-related exposures associated with each cohort for the year-long exposure period. Tables 6-10, 6-11, and 6-12 describe the algorithms used to estimate the values of particular parameters required by the procedure.

According to this procedure, pNEM/CO determined vehicle volume, cigarette emission rate, and air conditioning availability on an annual basis. Vent status, speed, and air exchange rate were selected on a trip basis. Smoking status was determined on an event-by-event basis during each trip.

A memorandum by Cohen, Johnson, and Rosenbaum (1999) describes the derivation of the 14-step algorithm in detail. Briefly, analysts assumed that the outside concentration during each exposure event in a vehicle would be equal to the outdoor concentration associated with the motor vehicle for the clock hour containing the event. The passive smoking status would be determined by the diary entry for the event. A value for the emission rate of the cigarette would be randomly selected from a lognormal distribution with geometric mean = 71,400 μg and geometric standard deviation = 1.3 (see Section 8.2.1). The volume of the vehicle would be determined by probabilistically assigning a size classification (e.g., subcompact automobile) to the vehicle and then assuming that the volume of the vehicle was equal to the average enclosed volume of the passenger compartment for vehicles in that classification (e.g., 2.32 m^3). Subsection 6.4.5.2 describes the methods which analysts used to estimate the distribution of size classifications and the average passenger volume assigned to each classification.

Analysts evaluated the sensitivity of the pNEM/CO mass-balance model to variations in air exchange rate and found that the model was not very sensitive to the exact rate value when the value exceeded about 10 hr^{-1} . A review of the literature identified a recent study (Rodes, Sheldon, Whitaker, et al., 1998) funded by the California Air Resources Board (CARB) as the best existing source of air exchange data for passenger vehicles, although the data were limited to a few vehicles tested under a small number of driving scenarios (defined by speed, window status, and vent status). All air exchange rates measured while the vehicle vents were open (and the windows were either open or closed) exceeded 20 hr^{-1} . At very low speeds, air exchange rates as low as 2 hr^{-1} were measured when the vehicle vents were closed. Consistent with these findings, analysts developed a probabilistic algorithm in which air exchange rates for a given speed and vent status (open or closed) were randomly selected from specified log-normal distributions based on the CARB data (see Subsection 6.4.3.2). Vehicle speed was probabilistically determined by an algorithm

Table 6-9. Algorithm Used to Model the Trip-related Exposures Associated with Each Cohort for the Year-long Exposure Period.

1. The volume of the vehicle is determined by the algorithm presented in Table 6-10. This value is held constant for all trips associated with the cohort.
2. The CO emission rate for cigarettes ($CO_{cigarette}$) smoked in the vehicle is determined by randomly selecting a value from a lognormal distribution with geometric mean = 71,400 μg and a geometric standard deviation of 1.3. The total emission rate from cigarettes is determined by the equation

$$S_{\min} = (n_{\text{smokers}})(n_{\text{cigs/smoker/hr}})(CO_{cigarette})/60$$
 in which n_{smoker} is the number of smokers and $n_{\text{cigs/smoker/hr}}$ is the number of cigarettes smoked by each smoker per hour. Assume $n_{\text{smokers}} = 1$ and $n_{\text{cigs/smoker/hr}} = 2$. The resulting value of S_{\min} (expressed as $\mu\text{g CO min}^{-1}$) is held constant for all trips with smokers associated with the cohort.
3. Air Conditioner Availability. Select a random number (RN1) between zero and 1. If RN1 is 0.85 or below, then assume an air conditioner is available. Otherwise assume an air conditioner is unavailable (i.e., the vehicle does not have an air conditioner or the air conditioner is not functional). Apply this result to all trips associated with the cohort.
4. Parameters specific to the trip (vent status, speed, and air exchange rate) are determined by the algorithm presented in Table 6-11.
5. The inside CO concentration at the beginning of the first event of the trip is set equal to the outdoor concentration associated with the motor vehicle for the hour.
6. The outside CO concentration for the duration of the event is set equal to the outdoor concentration associated with the motor vehicle for the clock hour containing the event. (As exposure events do not cross clock hours, there is only one outside concentration associated with each event).
7. The smoking status of the event is determined by the entry for “smokers present (yes/no)” included in the CHAD database.
8. If smoking occurs during the event, the CO emission rate for ETS is set equal to the value determined in Step 2 above.
9. The air exchange rate for the event is set equal to the value determined by Step 4 above.

Continued

Table 6-9 (continued)

10. The one-minute version of the mass-balance model is used to determine the average CO concentration for each minute and the instantaneous CO concentration for the end of each minute.
11. The average CO concentration for the event is determined by averaging the minute-average CO concentrations for the event.
12. The inside CO concentration at the beginning of the next event is set equal to the instantaneous CO concentration calculated in Step 10 for the end of the last minute of the preceding event.
13. Repeat Steps 6 through 12 for each subsequent event in the trip sequence.
14. Repeat Steps 4 through 13 for each subsequent trip.

Table 6-10. Algorithm for Estimating Enclosed Volumes of Cars and Trucks.

1. Randomly select value between zero and 1.
2. Compare selected value to following ranges^a.

Microenvironment No. 12: Automobiles

0.000 to 0.034: mini-compact (1.93 m³)
0.035 to 0.068: sub-compact (2.32 m³)
0.069 to 0.275: compact (2.58 m³)
0.276 to 0.862: mid-size (2.78 m³)
0.863 to 0.988: large (3.09 m³)
0.989 to 0.992: small wagon (3.48 m³)
0.993 to 0.996: mid-size wagon (3.82 m³)
0.997 to 1.000: large wagon (4.81 m³)

Microenvironment No. 13: Trucks

0.00 to 0.31: curb weight < 3,500 lbs (1.52 m³)
0.32 to 0.62: 3,500 lbs ≤ curb weight ≤ 4,000 lbs (1.81 m³)
0.63 to 1.00: curb weight > 4,000 lbs (2.25 m³)
3. Use indicated value in parentheses for vehicle volume.

^aSee Subsection 6.4.5 for derivation of ranges and associated volumes.

Table 6-11. Algorithms for Determining Trip-Specific Values of Parameters Used in the Mass-balance Model Applied to Cars and Trucks.

1. Other algorithms have determined the availability of air conditioning in the cohort's vehicle and the daily average temperature (DAT) for each day of the year. Use these values as necessary in determining the following values to be applied to each vehicle trip taken by the cohort. (Trips by cars and trucks are treated separately).
2. Vent Status. Apply the residential window status algorithm as described in Subsection 6.4.3.1. This algorithm determines window status based on AC system and the daily average temperature according to the probabilities listed in Table 6-16. For the current purpose, vehicles with functional air conditioners are equated to residences with central air conditioning systems, and vehicles with vents open are equated to residences with windows open:
 - a) For each day, determine daily average temperature from step 1 and air conditioning (AC) system availability from step 2. Select RN2 between zero and 1.
 - b) Assume step (a) specified 65 degrees and functioning AC. RN2 will be evaluated against percentage values listed in Table 6-16 for functional AC - medium temperature range (i.e., 35.6, 29.4, and 34.6).
 - c) If $RN2 \leq 0.356$, vents are always closed. AER value is determined by "vent closed" equation in Step 4a below.
 - d) If $0.356 < RN2 \leq 0.650$, vents are always open. AER value is determined by "vent open" equation in Step 4a below.
 - e) If $0.650 < RN2$, vents are open for 58.2 percent of the time (see last column of Table 6-16) and therefore are closed 41.8 percent of the time. The AER is calculated as $0.582 \times (\text{AER for open windows}) + 0.418 \times (\text{AER for closed windows})$ in which the AER values are determined by Step 4a below.

Continued on next page

3. Speed. Select a random number (RN4) between zero and 1. Use this random number to select a vehicle speed using the distribution given in Table 6-12.
 - a) If $RN4 \leq 0.0462$, speed = 0 mph.
 - b) If $0.0462 < RN4 \leq 0.1124$, speed = 5 mph.
 - c) If $0.1124 < RN4 \leq 0.2400$, speed = 10 mph.
 - d) Etc. Final case is $0.9988 < RN4$, speed = 60 mph.
4. Air Exchange Rate. Simulate AER from the appropriate log-normal distribution as indicated below. Vent status and speed were simulated in Steps 2 and 3.
 - a) Compute mean (of the logarithms) using the formulae
If vent is closed: $\mu = 3.37311 - 2.46213 + 0.03696 \times \text{speed}$
If vent is open: $\mu = 3.37311 + 0.01798 \times \text{speed}$.
 - b) Variance (of the logarithms) = $\sigma^2 = 0.27323$.
 - c) Randomly select Z from a standard normal distribution. Values of Z are not permitted to fall below -1.645 or above 1.645.
 - d) Hour AER = $\exp(\mu + \sigma Z)$.
 - e) Divide AER by 60 to determine minute AER.

Table 6-12. Vehicle Speed Distribution.

Speed (mph)	Frequency (percent)	Cumulative percentage
0	4.62	4.62
5	6.62	11.24
10	12.76	24.00
15	21.75	45.75
20	21.52	67.27
25	16.12	83.39
30	7.57	90.96
35	4.10	95.06
40	2.18	97.24
45	1.46	98.70
50	0.91	99.61
55	0.27	99.88
60	0.12	100.00

^aReference: Table 4 of Cohen et al. (1999). Estimates were derived from data presented in Carlson and Austin (1997).

based on data obtained from the Spokane-Baltimore-Atlanta instrumented vehicle study (Carlson and Austin, 1997). Vent status was determined probabilistically by the same pNEM/CO algorithm that simulates the opening and closing of windows in residences as a function of daily average temperature and availability of air conditioning. To complete the simulation, the probability of having a functioning air conditioner was estimated using the results of an EPA Office of Mobile Sources study (Koupal, 1998).

Passive smoking was assumed to occur whenever the activity diary data indicated the presence of smokers in the vehicle. The probabilistic algorithm described in Subsection 8.2.1 was used to estimate emission rate per cigarette during passive smoking events. The algorithm specified that one smoker was present during smoking events and that two cigarettes were smoked per hour. These assumptions were consistent with estimates of smoking rates provided by Repace, Jinot, Bayard, et al. (1998).

6.4.2.5 Mass Transit Vehicles

Microenvironment No. 14 included buses, trains, subway trains, and other mass transit vehicles not included in Microenvironment Nos. 12 and 13. Analysts assumed that passive smoking did not occur in these vehicles. The hour mass-balance model was used to estimate hourly average CO concentrations inside each vehicle as a function of outside CO concentration and air exchange rate. Air exchange rates were selected from a uniform distribution with minimum equal to 1.8 hr^{-1} and maximum = 5.6 hr^{-1} as discussed in Subsection 6.4.3.2.

6.4.2.6 Estimation of Mass-balance Parameters

Subsection 6.4.3 and Section 8 provide descriptions of the algorithms and data bases used to determine the air exchange rates, burner operation probabilities, burner emission rates, pilot light emission rates, cigarette emission rates, and residential volumes used in the mass-balance model. Many of these algorithms require that values be selected at random from normal or lognormal distributions. This selection was performed by first defining the distribution of interest by one of the following expressions:

$$\text{Normal: } X = AM + (ASD)(z) \quad (6-50)$$

$$\text{Lognormal: } X = (GM)(GSD)^Z \quad (6-51)$$

In these expressions, AM is the arithmetic mean, ASD is the arithmetic standard deviation, GM is the geometric mean, and GSD is the geometric standard deviation. The distribution type (normal vs. lognormal) and the corresponding values for the mean and standard deviation were determined by fitting distributions to representative data sets. A value for X was selected from a particular distribution by randomly selecting a value for Z from the unit normal distribution $[N(0, 1)]$ and substituting it into the appropriate equation. Tables 6-13, 6-14, and 6-15 list the distribution types and parameter values for the majority of random variables used in the mass-balance model.

Table 6-13. Distributions of Parameter Values Used in the Application of the pNEM/CO Mass-Balance Model to Denver.

Parameter	Distribution of parameter	Reference
Air exchange rate, exchanges/h: residence - windows closed	Lognormal distributions by season Season 1 <ul style="list-style-type: none"> ○ Geometric mean = 0.450 ○ Geometric standard deviation = 1.960 ○ Lower bound = 0.120 ○ Upper bound = 1.683 Season 2 <ul style="list-style-type: none"> ○ Geometric mean = 0.308 ○ Geometric standard deviation = 2.241 ○ Lower bound = 0.063 ○ Upper bound = 1.498 Season 3 <ul style="list-style-type: none"> ○ Geometric mean = 0.653 ○ Geometric standard deviation = 2.010 ○ Lower bound = 0.166 ○ Upper bound = 2.566 Season 4 <ul style="list-style-type: none"> ○ Geometric mean = 0.309 ○ Geometric standard deviation = 1.716 ○ Lower bound = 0.107 ○ Upper bound = 0.890 	Johnson, Memorandum No. 1, 1998 Murray and Burmaster, 1995
Air exchange rate, exchanges/h: residence - windows open	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 1.34 ○ Geometric standard deviation = 1.55 ○ Lower bound = 0.57 ○ Upper bound = 3.16 	Johnson, Memorandum No. 1, 1998 Johnson, Weaver, Mozier, et al., 1998
Air exchange rate, exchanges/h: nonresidential, enclosed microenvironments, including motor vehicles	See Table 6-15	See Table 6-15
Annual gas usage by burners, kilojoules	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 2.11×10^6 ○ Geometric standard deviation = 1.48 ○ Lower bound = 0.98×10^6 ○ Upper bound = 4.55×10^6 	Menkedick, Niemuth, Hartford, and Landstrom, 1993
Annual gas usage by pilot lights, kilojoules	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 3.37×10^6 ○ Geometric standard deviation = 1.84 ○ Lower bound = 1.02×10^6 ○ Upper bound = 11.13×10^6 	Menkedick, Niemuth, Hartford, and Landstrom, 1993
Burner emission factor, mg/kilojoule	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 0.0294 ○ Geometric standard deviation = 2.77 ○ Lower bound = 0 ○ Upper bound = 0.400 	Davidson, Borrazzo, and Hendrickson, 1987
Residential volume, cubic meters	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 436 ○ Geometric standard deviation = 1.62 ○ Lower bound = 169 ○ Upper bound = 1122 	Bureau of Census, 1995

Table 6-14. Distributions of Parameter Values Used in Application of the pNEM/CO Mass-Balance Model to Los Angeles.

Parameter	Distribution of parameter	Reference
Air exchange rate, exchanges/h: residence - windows closed	Lognormal distributions by season Season 1 <ul style="list-style-type: none"> ○ Geometric mean = 0.507 ○ Geometric standard deviation = 1.910 ○ Lower bound = 0.143 ○ Upper bound = 1.802 Season 2 <ul style="list-style-type: none"> ○ Geometric mean = 0.619 ○ Geometric standard deviation = 1.950 ○ Lower bound = 0.167 ○ Upper bound = 2.292 Season 3 <ul style="list-style-type: none"> ○ Geometric mean = 1.054 ○ Geometric standard deviation = 2.489 ○ Lower bound = 0.176 ○ Upper bound = 6.296 Season 4 <ul style="list-style-type: none"> ○ Geometric mean = 0.607 ○ Geometric standard deviation = 2.034 ○ Lower bound = 0.151 ○ Upper bound = 2.441 	Johnson, Memorandum No. 2, 1999 (see Appendix I) Murray and Burmaster, 1995
Air exchange rate, exchanges/h: residence - windows open	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 1.34 ○ Geometric standard deviation = 1.55 ○ Lower bound = 0.57 ○ Upper bound = 3.16 	Johnson, Memorandum No. 1, 1998 Johnson, Weaver, Mozier, et al., 1998
Air exchange rate, exchanges/h: nonresidential, enclosed microenvironments, including motor vehicles	See Table 6-15	See Table 6-15
Annual gas usage by burners, kilojoules	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 1.73×10^6 ○ Geometric standard deviation = 1.48 ○ Lower bound = 0.80×10^6 ○ Upper bound = 3.73×10^6 	Menkedick, Niemuth, Hartford, and Landstrom, 1993
Annual gas usage by pilot lights, kilojoules	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 2.76×10^6 ○ Geometric standard deviation = 1.84 ○ Lower bound = 0.84×10^6 ○ Upper bound = 9.12×10^6 	Menkedick, Niemuth, Hartford, and Landstrom, 1993
Burner emission factor, mg/kilojoule	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 0.0294 ○ Geometric standard deviation = 2.77 ○ Lower bound = 0 ○ Upper bound = 0.400 	Davidson, Borrazzo, and Hendrickson, 1987
Residential volume, cubic meters	Lognormal distribution <ul style="list-style-type: none"> ○ Geometric mean = 363 ○ Geometric standard deviation = 1.64 ○ Lower bound = 138 ○ Upper bound = 957 	Bureau of Census, 1995

Table 6-15. Distributions for Air Exchange Rate (ν) for Enclosed, Nonresidential Microenvironments

Microenvironment			Activity diary locations included in microenvironment	Distribution of Air Exchange Rate (ν)					
				Distribution type	Lognormal Parameters		Bounds		Source of data
Code	General location	Specific location	GM		GSD	Lower	Upper		
2	Indoors	Nonresidence A	Service station or auto repair	Lognormal	1.24	1.93	0.34	4.50	a
3	Indoors	Nonresidence B	Other repair shop Shopping mall	Lognormal	1.24	1.93	0.34	4.50	a
4	Indoors	Nonresidence C	Restaurant	See Table 8-7	--	--	--	--	--
5	Indoors	Nonresidence D	Bar	See Table 8-7	--	--	--	--	--
6	Indoors	Nonresidence E	Other indoor location Auditorium	Lognormal	1.24	1.93	0.34	4.50	a
7	Indoors	Nonresidence F	Store Office Other public building	Lognormal	1.24	1.93	0.34	4.50	a
8	Indoors	Nonresidence G	Health care facility School Church Manufacturing facility	Lognormal	1.36	1.91	0.38	4.83	b
9	Indoors	Residential garage	Residential garage	Lognormal	1.24	1.93	0.34	4.50	a

Microenvironment			Activity diary locations included in microenvironment	Distribution of Air Exchange Rate (v)					
				Distribution type	Lognormal Parameters		Bounds		Source of data
Code	General location	Specific location			GM	GSD	Lower	Upper	
12	Vehicle	Automobile	Automobile	See Table 6-11	--	--	--	--	c
13	Vehicle	Truck	Truck	See Table 6-11	--	--	--	--	c
14	Vehicle	Mass transit vehicle	Bus Train/subway Other vehicle	Uniform	--	--	1.8	5.6	c

^aData set containing all non-school AER values provided by Turk, Grimsrud, Brown, et al. (1989) and CEC (Lagus Applied Technology, 1995).

^bData set containing all AER values provided by Turk et al. (1989) and CEC (Lagus Applied Technology, Inc., 1995).

^cRodes, Sheldon, Whitaker, et al. (1998).

6.4.3 Estimation of Air Exchange Rate

6.4.3.1 The Air Exchange Algorithms

In Version 2.1 of pNEM/CO, a probabilistic algorithm was used to estimate an air exchange rate (AER) value for each enclosed microenvironment (i.e., buildings and vehicles). In most cases, the estimation procedure consisted of randomly selecting an AER value for the cohort from a distribution specific to the microenvironment. These distributions are presented in Tables 6-11, 6-15, and 8-7.

A more sophisticated methodology was used for the indoor - residence microenvironment which distinguished between air exchange when windows were open and when windows were closed. The window status was conditioned on the air conditioning (AC) system assigned to the cohort's residence and the outdoor temperature. A probabilistic algorithm assigned one of three potential residential AC systems to each cohort (central, window units, or none). A window status algorithm was then used to probabilistically determine window status (closed or open). Based on this determination, a value of AER was selected from either the closed window distribution or the open window distribution. The closed window distribution varied with season; a single distribution was used to represent open window conditions.

The AC algorithm required that the user specify the proportion of residences in the study area that had central AC, window units, and no AC. According to the 1995 American Housing Survey for Denver (Bureau of the Census, 1995), the breakdown for Denver was 25.3 percent central, 14.3 percent window, and 60.4 percent none. Analysts estimated the corresponding statistics for the Los Angeles study area to be 32 percent central, 21 percent window, and 47 percent none, based on data for the Los Angeles - Long Beach Standard Metropolitan Statistical Area provided by the American Housing Survey (Bureau of the Census, 1997).

The application of the AC algorithm to Denver can be described as follows:

1. For each day, select a random number (RN) between zero and 1.
2. If $RN \leq 0.253$, the AC system is "central."
3. If $0.253 < RN \leq 0.396$, the AC system is "window units."
4. If $0.396 < RN$, the AC system is "none."

The same procedure with appropriate parameter substitutions was applied to Los Angeles.

The window status algorithm was originally developed for applications of pNEM/O3 (see Subsection 6.3.4). The pNEM/O3 algorithm determines window status based on AC system and the daily average temperature according to the probabilities listed in Tables 6-4 through 6-7. Version 2.1 of pNEM/CO combines a somewhat simplified version of the window status algorithm and an AER algorithm as follows.

1. The AC algorithm determines the AC system for the cohort.
2. Go to the first/next day. The average temperature for the day is obtained from a supplementary temperature file. Select RN between zero and 1.

3. Find the row in Table 6-16 that corresponds to the specified AC system and daily average temperature. Evaluate RN against the percentage values listed in this row consistent with the following example.
 - a. Assume Steps 1 and 2 specified AC system = central AC and average daily temperature = 65 degrees. RN will be evaluated against the percentage values listed in Table 6-16 for central AC - medium temperature range (i.e., 35.6, 29.4, and 34.6).
 - b. If $RN \leq 0.356$, windows are closed all day. AER value is selected from the "windows closed" AER distribution.
 - c. If $0.356 < RN \leq 0.650$, windows are open all day. AER value is selected from the "windows open" AER distribution.
 - d. If $0.650 < RN$, windows are open for 58.2 percent of the day (see last column). AER is determined by the expression

$$AER = (0.582)(\text{open AER}) + (0.418)(\text{closed AER}) \quad (6-52)$$

where open AER is selected from the open window AER distribution and closed AER is selected from the closed window AER distribution.

4. If last day, end. Otherwise, go to Step 2.

A special version of the mass-balance model was applied to restaurants and bars. This model characterized air exchange as normalized ventilation rate (NVR), expressed as volume of air changed per person per hour. Subsection 8.2.3 describes the model and provides a method for estimating NVR.

Table 6-16. Percentage of Person-Days With Indicated Window Ratio by Air Conditioning System and Temperature Range

Air conditioning system	Temperature range ^a	Percentage of person-days with indicated window ratio ^b			Mean of ratios not equal to 0 or 1
		Ratio = 0	Ratio = 1	0 < Ratio = <1	
Central	Low	86.0	0.8	13.2	0.260
	Medium	35.6	29.4	34.6	0.582
	High	62.1	12.9	25.0	0.503
Room units	Low	73.2	2.0	24.7	0.316
	Medium	12.0	44.2	43.8	0.618
	High	17.1	34.3	48.6	0.521
No air conditioning	Low	80.0	1.0	19.0	0.276
	Medium	4.7	59.1	36.2	0.716
	High	1.4	70.8	27.8	0.774

^a Low: 31° to 62°F. Medium: 63° to 75°F. High: 76°F and above.

^b Ratio = (minutes windows open)/(minutes spent in residence).

6.4.3.2 Air Exchange Rate Distributions

A review of scientific literature was conducted to identify references relating to air exchange rates (AERs). Of the references identified, only a few were found to contain sufficient data to construct a distribution of AERs relating to a particular building type such as residence or office. The three most useful studies were conducted by Murray and Burmaster (1995); Turk, Grimsrud, Brown, et al. (1989); and Lagus Applied Technology (1995).

Indoors - Residence

An article by Murray and Burmaster (1995) described their analysis of residential air exchange rate data compiled by the Brookhaven National Laboratory (BNL). The BNL data included AERs for 2,844 residences in the United States, classified according to four geographic regions and the four seasons. The data for Denver were included in Region 2. The BNL data for Region 2 includes a large number of AER values for winter and spring, but small sample sizes for summer and autumn ($n = 2$ and 23 , respectively). Statistical methods were used to estimate the geometric mean and standard deviation for the seasons with limited data (Johnson, 1998). The resulting seasonal AER distributions for Region 2 (which includes Denver) are included in Table 6-13. The lower and upper bounds of the distributions are based on the 2.5th and 97.5th percentile of the distributions.

A similar approach was used to develop seasonal distributions of air exchange rate for Los Angeles when windows were open. In this case, the BNL data for Region 4 were assumed to represent Los Angeles. A memorandum by Johnson (1999) describes how analysts applied statistical methods to these data to develop the seasonal distributions listed for Los Angeles in Table 6-14.

It should be noted that the estimates for Regions 2 and 4 presented by Murray and Burmaster were based solely on data derived from the BNL database. Pandian, Behar, Ott, et al. (1998) identified errors in a version of the BNL database previously used by Pandian, Ott, and Behar (1993) and provided corrected estimates of AER for various geographic regions. In evaluating other researcher's use of the BNL database, Pandian, Behar, Ott, et al. (1998) concluded that the errors they identified did not affect the AER statistics presented by Murray and Burmaster (1995). This conclusion is supported by the corrected statistics presented by Pandian, Behar, Ott, et al. (1998) for a region containing Denver which are consistent with the statistics presented by Murray and Burmaster (1995) for Region 2.

For residences with windows open, the AER distribution in Version 2.1 of pNEM/CO is based on a study of a single residence. The American Petroleum Institute (API) conducted a study of a typical suburban house over a 24-hour time period (Johnson, Weaver, Mozier, et al., 1998). In that study, researchers altered the ventilation characteristics of the house each hour according to a prepared script, and measured the resulting hourly average AER. Analysts determined that the data for hours when windows were open could be characterized by a lognormal distribution with a geometric mean of 1.34 air changes per hour, and a geometric standard deviation of 1.55 (Johnson, 1998). The upper and lower bounds of the distribution have been set at 0.57 and 3.16, which correspond to the 2.5th and 97.5th percentiles, respectively. The

distribution was applied to both Denver and Los Angeles.

Although a wealth of air exchange data currently exist for commercial and residential buildings with closed windows, there is a shortage of data representing buildings with open windows. Perhaps the best existing source of open-window data prior to 1998 is a study reported by Wallace and Ott (1996) in which researchers measured air exchange rates in a detached house in Redwood City, California. A continuous monitor (Brüel & Kjaer Model 1302) was used to track the decay of sulfur hexafluoride (SF_6) under a variety of conditions over a period of 16 months. The majority of measurements (88 of 101) were made with all external doors and windows closed. Nine measurements were made under maximum air exchange conditions in which all windows and one or two doors were open. The most useful data were obtained from 27 measurements in which one or two windows were opened to varying widths in a single room (the den). The measured air exchange rates varied from 0.35 to 5.6 h^{-1} , with half the rates between 0.59 and 2.75 h^{-1} . Linear regression analyses indicated that the air exchange rate increased by about 0.12 h^{-1} for every inch that the window was opened. Analysts judged the results of this study to be generally consistent with those of the Johnson, Weaver, Mozier, et al. (1998) study cited above.

Wallace and Ott (1996) also provide a useful survey of other studies which have measured air exchange rates. In most cases, the available data represent buildings with closed windows or buildings for which the window status is unknown.

Microenvironment Nos. 2 through 8

This group of microenvironments includes all nonresidential, indoor microenvironments. Two AER distributions are used in pNEM/CO to represent buildings in these microenvironments. Microenvironment Nos. 2 through 7 are represented by a lognormal distribution with geometric mean = 1.24 and geometric standard deviation = 1.93. Microenvironment No. 8 is represented by a lognormal distribution with geometric mean = 1.36 and geometric standard deviation = 1.91. These distributions were developed from statistical analyses of AER data provided by two studies. The first study, conducted by Turk, Grimsrud, Brown, et al. (1989), measured AERs in 40 public buildings identified as schools (n = 7), offices (n = 25), libraries (n = 3), and multipurpose buildings (n = 5). The second study was conducted by the California Energy Commission (Lagus Applied Technology, Inc., 1995), and included 49 public buildings identified as schools (15), offices (22), and retail stores (13).

Microenvironment Nos. 2 through 7 are similar in that each includes various types of public buildings but omits schools. To determine a representative distribution of AER for these microenvironments, analysts combined all non-school data from the Turk and CEC studies into a single data set containing 68 values. These values could be well-fit by a lognormal distribution with geometric mean of 1.24 air changes per hour and a geometric standard deviation of 1.93. As indicated in Table 6-15, AER values for Microenvironments Nos. 2, 3, 6, and 7 were randomly selected from this distribution. Values were not permitted to fall below 0.34 or exceed 4.50, corresponding to the 2.5th and 97.5th percentiles of the distribution.

Microenvironment No. 8 differs from the microenvironments discussed above in that it includes school and non-school buildings. Consequently, analysts used the

complete set of AER values from the Turk and CEC studies to represent this microenvironment. The resulting data set could be well fit by a lognormal distribution with geometric mean = 1.36 air changes per hour and geometric standard deviation = 1.91. AER values for Microenvironment No. 8 were randomly selected from this distribution. Values were not permitted to fall below 0.38 or exceed 4.83, corresponding to the 2.5th and 97.5th percentiles of the distribution.

An alternative to air exchange rate was used in calculating the a_3 term when Equation 6-7 was applied to Microenvironments No. 4 (restaurants) and No. 5 (bars). This parameter -- normalized ventilation rate (NVR) -- is defined as the volume of air exchanged per hour per person and is expressed as m³/hr/person. Section 8.2.3 provides distributions for the parameter and discusses its use in estimating the contribution of passive smoking to CO concentrations in restaurants and bars.

Passenger and Mass Transit Vehicles

Table 6-9 presents the 14-step method used to estimate the CO concentration associated with each exposure event in Microenvironment Nos. 12 (automobiles) and 13 (trucks). In Step 4, the algorithm presented in Table 6-11 is used to determine the air exchange rate for the exposure event. The algorithm determines air conditioning status with an 85 percent probability of having AC; vent status according to AC status and temperature; and speed. Air exchange rate is then determined probabilistically as a function of these three parameters.

In developing the algorithm summarized in Table 6-11, analysts considered three sources of data for estimating the distribution of air exchange rates in vehicles: Hayes (1991); Ott, Switzer, and Willis (1994); and Rodes, Sheldon, Whitaker, et al. (1998). Hayes (1991) provided a point estimate of 36 air changes per hour based on his analysis of data presented by Peterson and Sabersky (1975). This estimate was used for all vehicle-related microenvironments in the 1992 version of pNEM/CO. In a study reported by Ott, Switzer, and Willis (1994), researchers measured an AER value of 13.1 air changes per hour in a car moving at 20 mph with windows closed. AER values of 67 to 120 air changes per hour were measured in the car at the same speed with windows open.

During a study funded by the California Air Resources Board (Rodes, Sheldon, Whitaker, et al., 1998), researchers measured 11 AER values under test conditions which varied the ventilation setting, vehicle speed, and vehicle type (Table 6-17). (The draft version of the CARB report shows that multiple measurements were made for some vehicle/vent combinations - the analysis presented here used the average values given in the final CARB report). Three of the 11 values were obtained from a 91 Caprice which was tested under all three ventilation conditions. Unfortunately, this vehicle was tested at only one speed (55 mph). The 97 Taurus was tested at two ventilation settings and one speed (55 mph). The 97 Explorer was tested at the same two ventilation settings and at all three speeds. Note that the none of Taurus and Explorer values represent conditions with windows open. There is only one value for windows open -- the 91 Caprice driven at 55 mph. For the statistical analysis, analysts grouped this special case of vent open and windows (partially) open with the other cases of vent open and windows closed to produce data classified by only two ventilation conditions (open or closed).

The CARB data better represent AER under varying speeds and vent conditions than the point estimate provided by Hayes (1991) or the data provided by Ott, Switzer, and Willis (1994). Consequently, analysts used the CARB data as the basis for constructing the algorithm presented in Table 6-11. A memorandum by Cohen, Johnson, and Rosenbaum (1999) describes the methodology in more detail and provides justifications for the parameter values presented in Table 6-11.

Briefly, researchers began the process of constructing the algorithm by performing a sensitivity analysis of the mass-balance model used in pNEM/CO to determine whether it was sensitive to air exchange rates in the range associated with motor vehicles. The analysis showed that the mass-balance model was not very sensitive to the exact value of the AER when AER exceeded 10 hr^{-1} . The data obtained from CARB (Table 6-17) indicated that air exchange rate tended to exceed 20 hr^{-1} when vehicle vents were open (and the windows were open or closed). Furthermore, AER could be as low as 2 hr^{-1} when the vehicle vents were closed (at very low speeds). These results suggested that the AER algorithm should realistically simulate the opening and closing of vents.

Cohen, Johnson, and Rosenbaum (1999) used the CARB data to develop an algorithm that selected AER values from a lognormal distribution whose parameters varied according speed, vent status (open or closed), and air conditioning status (present or absent). The speed value was simulated using speed distributions (Table 6-12) developed from the Spokane-Baltimore-Atlanta instrumented vehicle study (Carlson and Austin, 1997). The algorithm determined window status using the same probabilistic procedure used elsewhere in pNEM/CO to determine window status in residences (see Subsection 6.4.3.1). The probabilities used in this procedure are a function of daily average temperature and the availability of an air conditioner. Researchers acknowledged that behavior patterns for opening windows in residences may not be the same as those for opening vents in vehicles, but were unable to find good data which were directly applicable for vehicles. Air conditioning status was estimated using the results of an EPA Office of Mobile Sources study (Koupal, 1998).

Researchers were unable to obtain specific measured data on air exchange rates for mass transit vehicles. A reasonable approximation was obtained from the 1997 Ford Explorer data provided by Rodes, Sheldon, Whitaker, et al (1998) and tabulated in Table 6-17. Preference was given to the first set of data listed in the table (vent closed, low fan speed) which was considered to be more representative of mass transit vehicles. Under these conditions, the measured air exchange rates were 1.8 per hour at 0 mph, 5.6 per hour at 35 mph, and 13.5 per hour at 55 mph. As mass transit vehicles typically travel at relatively low speeds, researchers selected the air exchange rates measured at 0 and 35 mph and assumed a uniform distribution over that range. Consequently, air exchange rates for mass transit vehicles were selected from a uniform distribution with minimum value equal to 1.8 hr^{-1} and maximum value equal to 5.6 hr^{-1} .

6.4.4 Indoor CO Emission Rates

Section 8 provides a detailed discussion of the methods used in Version 2.1 of pNEM/CO to simulate indoor emissions of CO from gas stoves and passive smoking.

Table 6-17. Air Exchange Rates Measured by Rodes, Sheldon, Whitaker, et al. (1998) Under Varying Conditions.

Test conditions		Air exchange rate (hr ⁻¹)		
Ventilation settings	Vehicle speed, mph ^a	1991 Caprice	1997 Taurus	1997 Explorer
Vent closed, low fan speed	55	39	14	13.5
	35			5.6
	0			1.8
Vent open, low fan speed	55	98	76	55.5
	35			35.7
	0			20.7
Vent open, low fan speed, front windows 1/3 open	55	160		

^aThe vehicle speed was constantly maintained throughout the AER measurement.

6.4.5 Enclosed Volumes

The mass balance model used in Version 2.1 of pNEM/CO requires distributions for the volume of residences and motor vehicles. Subsection 6.4.5.1 describes the method used to develop distributions of residential volume for Denver and Los Angeles based on the distribution of square footage reported for each city. Subsection 6.4.5.2 presents the method used to develop distributions of cabin volume for automobiles and trucks.

6.4.5.1 Residences

Table 6-18 presents data obtained from the American Housing Survey (Bureau of Census, 1995, 1997) representing the distribution for square footage of occupied units in Denver and Los Angeles. Plots of these statistics indicate that the data can be closely fit by lognormal distributions with the following values for geometric mean and geometric standard deviation.

		<u>Geometric mean</u>	<u>Geometric std. dev.</u>	<u>Median</u>
Square footage:	Denver	1926	1.62	2020
	Los Angeles	1604	1.64	1651

The values listed under “median” were provided in the American Housing Survey listings and agree closely with the estimated geometric means. (The geometric mean

Table 6-18. Statistics on Square Footage of Occupied Units in Denver and Los Angeles (Bureau of the Census, 1995, 1997).

Range of square footage for occupied units	Denver		Los Angeles	
	Number in thousands	Cumulative percent	Number in thousands	Cumulative percent
less than 500	1.0	0.2	20.7	1.7
500 to 749	9.3	2.5	37.0	4.7
750 to 999	34.1	10.6	104.9	13.2
1,000 to 1,499	75.6	28.6	354.8	42.0
1,500 to 1,999	86.0	49.2	326.1	68.5
2,000 to 2,499	86.7	69.8	199.6	84.7
2,500 to 2,999	49.8	81.7	66.8	90.1
3,000 to 3,999	53.4	94.5	77.1	96.3
4,000+	23.2	100.00	45.1	100.0
Total ^a	419.1	--	1232.1	--

^a Omits units which did not report square footage values.

of a “perfect” lognormal distribution is equal to its median).

Assuming an eight-foot ceiling and using 1 cubic meter = 35.315 cubic feet, the residential volumes can be modeled by lognormal distributions with the following parameters.

		<u>Geometric mean</u>	<u>Geometric std. dev.</u>
Volume, m ³ :	Denver	436	1.62
	Los Angeles	363	1.64

Version 2.1 of pNEM/CO selects residential volumes from these distributions. To prevent the occurrence of unrealistic values, volumes are not permitted to fall outside the 2.5th and 97.5th percentiles of each lognormal distribution. These values are listed below.

		<u>Lower bound</u>	<u>Upper bound</u>
Volume, m ³ :	Denver	169	1,122
	Los Angeles	138	957

6.4.5.2 Passenger Vehicles

The volume of the passenger compartment is determined by the algorithm presented in Table 6-10. For the specified microenvironment (automobiles or trucks), this algorithm randomly determines the vehicle type according to the estimated distribution of registered passenger vehicles and then assigns the vehicle the average volume for that vehicle type.

The distribution for automobiles by vehicle type was based on an analysis of the data listed in Table 6-19. The left-hand section of the table provides a breakdown of 1998 automobile sales by segment. The right-hand section provides values for average enclosed volume by vehicle type. Because data on enclosed volume were not available for the specified market segments (or vice versa), analysts made a series of assumptions as to how market segments could be allocated to vehicle types. Table 6-20 documents these assumptions and presents the resulting allocation of market share to each vehicle type.

Table 6-21 lists interior volume data for pickup trucks acquired from various industry sources. The trucks are classified according to the three weight classes for which 1995 sales data are available: less than 3,500 lbs (31 percent), 3,500 to 4,000 lbs (31 percent), and greater than 4,000 lbs (38 percent).

Analysts made the assumption that the average interior volume for each weight classification in Table 6-21 was equal to the average of interior volumes of the trucks listed for the classification. These average volumes are listed below together with the corresponding ranges of individual values (in parentheses).

<u>Weight classification</u>	<u>Percentage of sales</u>	<u>Average interior volume (m³)</u>
less than 3,500 lbs	31	1.52 (1.49 - 1.59)
3,500 to 4,000 lbs	31	1.81 (1.49 - 2.35)
greater than 4,000 lbs	38	2.25 (1.49 - 3.15)

These results were the basis for the algorithm used to estimate truck volume (see Table 6-10).

6.5 The Linear Relationship Method

Sections 6.2 through 6.4 provide descriptions of various mass balance models that have been incorporated into the NEM-type models. In general, analysts prefer mass balance models over alternative methods when microenvironmental pollutant levels are expected to change rapidly with time in response to changes in outdoor pollutant level, indoor emission rate, and/or indoor-outdoor air exchange rate. Because of their computational complexity, mass balance models may significantly increase program run time compared to other modeling methods. Analysts may also have difficulty obtaining good estimates for the various parameters of a mass balance model, particularly the indoor emission and decay rates.

In many applications, it is reasonable to substitute a relatively simple linear relationship for the more complex mass balance model. For example, HAPEM4 estimates concentrations of hazardous pollutants (HAP's) within each of 37 microenvironments using the linear relationship:

Table 6-19. 1998 Automobile Sales by Segment and Enclosed Volume by Vehicle Type.

1998 Automobile Sales by Segment ^a		Enclosed volume by vehicle type ^b	
Segment	Percent of market	Vehicle type	Average volume, m ³
Budget	3.0	Mini-compact	1.93
Small	18.7	Sub-compact	2.32
Lower mid-size	17.5	Compact	2.58
Mid-size	29.5	Mid-size	2.78
Upper mid-size	11.5	Large	3.09
Near luxury	5.4	Small wagon	3.48 ^c
Luxury	7.1	Mid-size wagon	3.82 ^c
Sporty	5.6	Large wagon	4.81 ^c
Specialty	1.3	--	--

^aSource: "1998 U.S. Car Sales by Segment," Automotive News Market Data Book Supplement, p. 33, May 1999.

^bSource: Average volumes by EPA class as listed in new car buying pages on Yahoo!

^cIncludes luggage space.

Table 6-20. Allocation of Market Segments to Vehicle Types.

Vehicle type (average passenger compartment volume, m ³)	Assigned market segment	Percent of market in segment (see Table 4-17)	Fraction of segment allocated to vehicle type	Resulting percent of market in vehicle type
Mini-compact (1.93)	Sporty	5.6	0.33	3.4 ^a
	Budget	3.0	0.50	
Sub-compact (2.32)	Sporty	5.6	0.33	3.4
	Budget	3.0	0.50	
Compact (2.58)	Sporty	5.6	0.33	20.7
	Small	18.7	1.00	
Mid-size (2.78)	Lower mid-size	17.5	1.00	58.7
	Mid-size	29.5	1.00	
	Upper mid-size	11.5	1.00	
Large (3.09)	Near luxury	5.4	1.00	12.6
	Luxury	7.1	1.00	
Small wagon (3.48)	Specialty	1.3	0.33	0.4
Mid-size wagon (3.82)	Specialty	1.3	0.33	0.4
Large wagon (4.81)	Specialty	1.3	0.33	0.4
Total				100.0

^aSample calculation: $(5.6\% \times 0.33) + (3.0\% \times 0.50) = 3.4\%$.

Table 6-21. Interior Volumes for Selected Pickup Trucks Classified by Curb Weight.

Curb weight category, lbs	Percentage of 1995 pickup sales ^a	Model	Curb weight, lbs	Interior volume		Source of volume data
				ft ³	m ³	
Less than 3,500	31	Dodge Dakota reg. cab	3378 - 3581	56.3	1.59	b
		Mazda 4x2 reg. cab	3356 - 3431	52.5	1.49	c
		Ford Ranger reg. cab	3429 - 3518	52.5	1.49	c
3,500 to 4,000	31	Dodge Dakota ext. cab	3611+	82.9	2.35	b
		Mazda supercab 2dr	3580	65.2	1.85	c
		Ford Ranger supercab 2dr	3576 - 3618	65.2	1.85	c
		Ford Ranger 4x4 reg. cab	3784	52.5	1.49	c
		Mazda 4x4 reg. cab	3862	52.5	1.49	c
Greater than 4,000	38	Dodge RAM reg. cab	4150+	62.8	1.78	b
		Dodge RAM ext. cab	4704+	111.2	3.15	b
		Ford F150	4247+	61.8	1.75	c
		Ford Ranger Elec	4427+	52.5	1.49	c
		Ford F150 supercab	4621+	106.9	3.03	c
		Ford F250 reg. cab	4957+	61.8	1.75	c
		Ford F250 supercab	5130+	106.9	3.03	c
		Maxda 4x4 supercab	4055	65.2	1.85	c
		Ford Ranger 4x4 supercab	4017 - 4057	65.2	1.85	c
		Ford F150 4x4 reg. cab	4615 - 4747	61.8	1.75	c
		Ford F150 4x4 supercab	5222 - 5364	106.9	3.03	c
		Ford F250 4x4 reg. cab	5426 - 5468	61.8	1.75	c
		Ford F250 4x4 supercab	5627 - 5642	106.9	3.03	c

^aInsurance Institute of Highway Safety, EPM Communications, Copyright 1998. In: "Utility Vehicles Overtake Passenger Vehicles," Research Alert, Vol. 16, No. 7, p. 7 (April 3, 1998).

^bPersonal communication from Mr. Chuck Paterka (248-576-5465), Daimler-Chrysler.

^cSpread sheet provided by Mr. Neil Whitbeck (313-322-9329), Fuel Economy and Quality, Environmental and Safety Engineering, Ford Motor Company.

^dSteve Cadle of General Motors stated that the cab volumes of GM pickups typically range from 60 to 70 ft³ (1.70 to 1.98 m³). Personal communication (November 4, 1999).

$$ME(m,c,t) = ADD(m) + [PROX(m)][PEN(m)][AMB(c,t)] \quad (6-52)$$

where:

ME(m,c,t): concentration in microenvironment *m* located in census tract *c* at time *t*,

ADD(m): additive factor for microenvironment *m*,

PROX(m): proximity factor for microenvironment *m*,

PEN(m): penetration factor for microenvironment *m*, and

AMB(c,t): ambient concentration for census tract *c* at time *t*.

The additive factor (*ADD*) accounts for emission sources within or near to a microenvironment. Its primary function is to account for emissions of a pollutant that originate indoors. Since this term is itself a concentration, the value of *ADD* can be set equal to zero, indicating no contribution from indoor sources.

The *PEN* factor accounts for the relationship between the concentration in the microenvironment and the concentration in the outdoor air in the immediate vicinity of the microenvironment, when the microenvironment contains no sources. For example, the *PEN* factor for the indoors, residence microenvironment would represent the ratio of the indoor concentration to the concentration in the air surrounding the residence when sources are absent in the residence.

The *PROX* factor accounts for the relationship between the outdoor concentration in the vicinity of the microenvironment and the concentration at the location represented by *AMB(c,t)*. In some applications of the HAPEM4, the ambient concentration is determined by measurements made by a fixed-site monitor. If this fixed-site monitor is located at a greater distance from the predominant emission sources (e.g., roadways) than the microenvironment is, then the monitor would tend to record lower values than the outdoor concentration at the microenvironment. In this case, the *PROX* factor should be greater than unity to compensate for the relatively low reading at the fixed-site monitor.

6.6 Parameter Estimates for the Linear Relationship Used in HAPEM4

The report by ICF Consulting and TRJ Environmental (2000) provides estimates for the parameters of the HAPEM4 linear model presented in Equation 6-52 for each combination of the 37 microenvironments (Table 6-1) and 35 HAP's (Table 6-22) considered in the NATA assessment. This section provides a summary of the methods used in developing these estimates.

6.6.1 Ambient Pollutant Concentration [AMB(c,t)]

Estimates of *AMB(c,t)* may be obtained from fixed-site monitors, dispersion models, or other sources. In the recent NATA application of HAPEM4 (EPA. 2001),

Table 6-22. List of HAPs for Development of ME Factors^a

ID No.	Name	ID No.	Name	ID No.	Name
1	acetaldehyde	13	1,2-dichloroethane (ethylene dichloride)	25	nickel compounds
2	acrolein	14	1,3-dichloropropene	26	7-PAH
3	acrylonitrile	15	ethylene dibromide (1,2-dibromethane)	27	polychlorinated biphenyls
4	arsenic	16	ethylene oxide	28	propylene dichloride (1,2-dichloropropane)
5	benzene	17	formaldehyde	29	quinoline
6	beryllium compounds	18	hexachlorobenzene	30	styrene ^b
7	1,3-butadiene	19	hydrazine	31	2,3,7,8-TCDD
8	cadmium compounds	20	lead compounds - organic	32	1,1,2,2-tetrachloroethane
9	carbon tetrachloride	21	lead compounds - inorganic	33	tetrachloroethylene (perchloroethylene)
10	chloroform	22	manganese compounds	34	trichloroethylene
11	chromium compounds	23	mercury compounds	35	vinyl chloride
12	coke oven emissions	24	methylene chloride (dichloromethane)		

^a In addition to this list, the following pollutants were also included for future development of ME factors: ethyl benzene, hexane, methyl-tertiary-butyl-ether (MTBE), propionaldehyde, toluene, and xylenes.

^b Styrene is the only HAP in this table that is not on EPA's urban HAPs list.

values for $AMB(c,t)$ for the entire nation were provided by the ASPEN (Assessment System for Population Exposure Nationwide) dispersion model. In this application, analysts began by defining a receptor grid for each emission source which extended 50 km in all directions from the source. This receptor grid, and hence the concentration estimates, varied according to emission source type (i.e., point sources, mobile sources, or area sources). Next, ASPEN used emissions and meteorological data specific to the census tract to estimate pollutant concentrations at each receptor point. The model then interpolated the receptor concentrations to the census tracts which fell within the 50 km study area. Finally, ASPEN weighted the estimated pollutant concentrations by population to produce an average concentration value for each census tract. Additional details concerning the use of ASPEN to estimate $AMB(c,t)$ can be found in the report "Development of Microenvironmental Factors for the HAPEM4 in Support of the National Air Toxics Assessment (NATA)" by ICF Consulting and TRJ Environmental (2000) and in the ASPEN User's Guide (SAI, 1999).

6.6.2 PROX Factor

Ideally, the concentration values provided by ASPEN are unbiased estimates of the outdoor concentration at locations within the census tract where people are likely to reside and work. The values can be used without adjustment to represent outdoor (ambient) concentrations for microenvironments that do not tend to be in close proximity to the predominant sources of the pollutant. Consequently, the appropriate *PROX* factor for such microenvironments is unity.

However, the ASPEN model estimates ambient HAP concentrations by source category (i.e., major, area, "on-road" vehicles, and "non-road" vehicles)¹, hence the estimates of outdoor concentrations are biased low for microenvironments that are usually located near predominant sources of a pollutant. The most prevalent example of this is for HAPs emitted by on-road motor vehicles that impact microenvironments in close proximity to roadways. For on-road mobile source pollutants such as benzene, these "close proximity" microenvironments include outdoor locations near roads (HAPEM4 microenvironment No. 7), vehicles (Nos. 1, 2, 3, 4, 8, and 36), service stations (No. 10), and parking facilities (No. 6). In these cases, the appropriate proximity factor should be greater than unity.

Table 6-23 lists the microenvironments whose *PROX* factors (from the literature) for acetaldehyde, acrolein, benzene, 1,3-butadiene, and formaldehyde for the on-road mobile source category are greater than unity. All other combinations of pollutants, microenvironments, and source categories have been assigned a default *PROX* factor equal to 1.0 [see Appendix C of the report by ICF Consulting and TRJ Environmental (2000)].

¹ The definitions for major and area sources are in Title I of the Clean Air Act Amendments (CAAA) of 1990, while the "mobile source" terms are defined in Title II of the CAAA.

Table 6-23. Microenvironments Assigned *PROX* Factors Greater Than 1.0 for Acetaldehyde, Acrolein, Benzene, 1,3-Butadiene, and Formaldehyde For the “Onroad” Mobile-Source Category Only

Microenvironment No.	Microenvironment	
	Specific	General
1	Car	In vehicle
2	Bus	In vehicle
3	Truck	In vehicle
4	Other	In vehicle
6	Parking lot/garage	Outdoors
7	Near road	Outdoors
8	Motorcycle	Outdoors
10	Service station	Outdoors
36	Train/subway	In vehicle

6.6.3 PEN Factor

6.6.3.1 Literature Survey

A comprehensive literature search was conducted of all available sources containing measurement data for use in developing the *PEN* factors for HAPEM4. The search was accomplished with the aid of on-line library search engines that captured relevant literature from the United States and other countries. These library search engines included, but were not limited to, the National Technical Information Service (NTIS), Enviroline, Transportation Research Information Services (TRIS), Wilson Applied Science & Technology Abstracts, Energy Science and Technology, PASCAL, MEDLINE, TOXLINE, Occupational Safety and Health (NIOSH[®]), and the Abstracts in New Technologies and Engineering (ANTE). The search was based on a large set of key words to ensure that it was comprehensive. The microenvironments and pollutants in Tables 6-1 and 6-22, respectively, were included as key words in the on-line search. Additional key words were used to focus the search on studies that had the potential for reporting results that had both indoor and outdoor measurements.

The search was conducted using the procedure described below.

Step 1

The on-line search was conducted using key words as described above. The search included all literature dating back to 1994. Literature published prior to 1994 was included in an earlier on-line search conducted by TRJ Environmental (see Step 5). Including years prior to 1994 in this search would have yielded more documents than could have been reviewed given the resources available for this project. The search

initially yielded titles and key word descriptors for 5,424 documents encompassing all of the HAPs in Table 6-22 (including the six HAPs listed at the end of the table). All of the titles and key word descriptors were reviewed and ranked for the potential of containing data useful to the development of a ME factor. Three categories of rankings were used: documents that had a high probability of containing relevant information, documents that had a moderate probability of containing relevant information, and documents that were unlikely to contain, or had a low probability of containing, relevant information.

Step 2

The next step was to retrieve the abstracts for each citation deemed to have either a high- or moderate probability of containing both indoor and outdoor measurements for the various HAPs. Of the original set of 5,424 titles obtained in Step 1, 251 abstracts were obtained because, based on the title review, they were thought to have the potential for providing the needed information for determining *PEN* factors, (i.e., high or moderate probability). The abstracts were obtained, reviewed, and ranked in a fashion similar to that in Step 1 to determine which articles or documents were most likely to provide the needed information for calculating *PEN* factors.

Step 3

The third step in the literature search was to obtain the complete paper or report for each abstract from Step 2 that was categorized as having either a high or moderate probability of containing relevant information for determining *PEN* factors. Of the 251 abstracts reviewed in Step 2, 100 were categorized with either a high or moderate potential for containing the necessary information for determining a *PEN* factor. The complete documents for these 100 references were obtained and reviewed for information needed to determine *PEN* factors.

Step 4

Because of the relative paucity of information obtained from the on-line literature search described in the three steps above, researchers obtained additional documents from studies that preceded those of the on-line search (i.e., pre-1994). These studies were thought to contain useful information on *PEN* factors and were identified during the review of papers and reports from the original on-line search. As a result of this step, researchers obtained and reviewed an additional 39 papers and reports for useful information on *PEN* factors.

Step 5

Concurrent with the activities described in Steps 1 through 4, researchers re-examined literature that had been obtained for an earlier review of scientific studies related to the measurement of *PEN* factors [refer to Fletcher, LaPointe, Long, and Johnson (1999) for a description of this work]. This research included a total of 68 documents that were also collected and reviewed. Researchers were careful not to

duplicate the documents collected in this step with those collected in Steps 1 through 3. Most of the documents collected in this step preceded the five-year period (i.e., 1994-1999) for which the studies in Steps 1 through 3 were collected.

6.6.3.2 Grouping HAPs and Microenvironments

Because of the paucity of published I/O measurement data for most of the 1,295 HAP-ME combinations (i.e., 35 HAPs × 37 microenvironments), researchers decided to group similar HAPs together. Microenvironments were also grouped based on similarity to each other. The purpose of this grouping was to allow assignment of ME factors to the groups, based on the best data available, so that more HAP-ME combinations could be covered (and to minimize the assignment of generic default values).

Microenvironmental Groups

Based on the characteristics of the 37 HAPEM4 microenvironments, each was assigned to one of five groups. Each group consists of microenvironments expected to have similar *PEN* factors, thus allowing ME factors developed for one microenvironment to be applied to other microenvironments in the same group. Table 6-24 shows the five ME groups used in this project, as well as the mapping between the original 37 HAPEM4 microenvironments and the five ME groups.

The five ME groups were established based on the general ME type (indoor, outdoor, or in-vehicle), expected location relative to roadways (near road, away from road), and building type (residence, other). This five-group classification shown in Table 6-24 is similar to the ME categories used in previous versions of the HAPEM. The current groupings include information on location for outdoor microenvironments and building type for indoor microenvironments. *PROX* factors are likely to be influenced by ME location relative to roadway sources, while *PEN* factors are affected by the structural design and usage patterns of indoor microenvironments.

In general, researchers used the actual HAPEM4 ME designations to assign a microenvironment to one of the five groups. However, for certain microenvironments, researchers used CO concentration data from a personal exposure monitoring study in Denver to determine whether the microenvironment should be assigned to a “near road” or “away from road” group (Johnson, 1984). Researchers assigned two indoor microenvironments with broad descriptors (“other indoor location” and “indoors, not specified”) to non-residence groups on the assumption that these microenvironments are not used to indicate residential locations.

Although these ME groups can be useful in situations where no data are available in the literature for a particular HAP-ME combination, there are limitations to this method of approximation. Different microenvironments, even those in the same group, have unique characteristics which may cause them to have different ME factors for some or all pollutants. However, the use of ME groups (essentially, extrapolating available data to closely related situations) is considered preferable to the use of more generic default assumptions.

Table 6-24. Description of ME Groups

ID No.	HAPEM4 Microenvironments		Location ^a	Bldg. Type ^b	ME Group ^c
	ME Designation	General ME Type			
1	Car	In vehicle	–	–	5
2	Bus	In vehicle	–	–	5
3	Truck	In vehicle	–	–	5
4	Other	In vehicle	–	–	5
5	Public garage	Indoors	–	O	2
6	Parking lot/garage	Outdoors	N	–	3
7	Near road	Outdoors	N	–	3
8	Motorcycle	Outdoors	N	–	3
9	Service station	Indoors	–	O	2
10	Service station	Outdoors	N	–	3
11	Residential garage	Indoors	–	R	1
12	Other repair shop	Indoors	–	O	2
13	Residence - no gas stove	Indoors	–	R	1
14	Residence - gas stove	Indoors	–	R	1
15	Residence - attached garage	Indoors	–	R	1
16	Residential - stove and garage	Indoors	–	R	1
17	Office	Indoors	–	O	2
18	Store	Indoors	–	O	2
19	Restaurant	Indoors	–	O	2
20	Manufacturing facility	Indoors	–	O	2
21	School	Indoors	–	O	2
22	Church	Indoors	–	O	2
23	Shopping mall	Indoors	–	O	2
24	Auditorium	Indoors	–	O	2
25	Health care facility	Indoors	–	O	2
26	Other public building	Indoors	–	O	2

ID No.	HAPEM4 Microenvironments		Location ^a	Bldg. Type ^b	ME Group ^c
	ME Designation	General ME Type			
27	Other location	Indoors	–	O	2
28	Not specified	Indoors	–	O	2
29	Construction site	Outdoors	A	–	4
30	Residential grounds	Outdoors	A	–	4
31	School grounds	Outdoors	A	–	4
32	Sports arena	Outdoors	A	–	4
33	Park/golf course	Outdoors	A	–	4
34	Other location	Outdoors	A/N	–	3/4 ^d
35	Not specified	Outdoors	A/N	–	3/4 ^d
36	Train/subway	In vehicle	–	–	5
37	Airplane	In vehicle	–	–	5

^a N = near road; A = away from road

^b R = residence; O = other building

^c 1 = indoors - residence; 2 = indoors - other building; 3 = outdoors, near road; 4 = outdoors, away from road; 5 = in vehicle

^d Use average of factors determined for Groups 3 and 4.

Pollutant Lifetime Groups

For the purpose of estimating *PEN* factors for pollutants without I/O ratios available in the literature, researchers used the lifetime grouping method developed by TRJ Environmental [see Long and Johnson, 1999; included in the report by ICF Consulting and TRJ Environmental (2000) as Appendix A]. This method classifies HAPs into five groups based on expected atmospheric residence lifetime. The HAPs belonging to each group are expected to have similar decay or removal rates in typical microenvironments. However, in the case of certain HAPs, insufficient data were available to definitively assign them exclusively into either the short, medium, or long lifetime categories. Therefore, these HAPs [identified as lifetime Groups 2 (medium short lifetime category) and 4 (medium long lifetime category) in Table 6-24] were assigned to Group 3 (medium lifetime category) for this project. Furthermore, Groups 2 and 4 contain only three HAPs for which no literature data were available for most microenvironments. In the future, as additional data on atmospheric lifetimes of HAPs become available, it may be possible to further subdivide these categories into additional groupings having narrower time increments and then reassign the HAPs among the larger number of groups. Because the atmospheric lifetime of a pollutant should be a major factor in determining its I/O ratio for a given microenvironment (in the absence of indoor sources), pollutants in each lifetime group should exhibit similar *PEN*

factors.

Researchers considered both atmospheric transformation reactions for gaseous HAPs and removal by deposition for particle-phase HAPs. Although the groups are not directly derived from structure or composition of each HAP, some broad associations can be made between structural class (e.g., aromatics, halogenated hydrocarbons) and lifetime group. Table 6-25 lists the lifetime group for each of the HAPs considered in this project, and the paper in Appendix A of the report by ICF Consulting and TRJ Environmental (2000) indicates the lifetime classification for each of the 188 HAPs.

There are a number of limitations to the pollutant lifetime grouping method. One limitation is that the method does not explicitly consider the effect of multiple phases of a single pollutant on HAP lifetimes. This approach treats semi-volatile organics and other pollutants that may be present in both gas and particle phases as gaseous pollutants. Another limitation of the method is that it does not account for the production of HAPs through atmospheric transformations. Also, the method does not consider effects of HAP sorption by materials in indoor microenvironments.

6.6.4 Estimated Values

The detailed results of the literature review are found in Appendix B of the report by ICF Consulting and TRJ Environmental (2000). An electronic version of this information is available at the following EPA web address:

www.epa.gov/ttn/atw/sab/appendix-b.pdf.

The majority of surveyed studies reported indoor concentrations of HAPs that were substantially greater than outdoor concentrations, that is, the I/O ratio exceeded 1. An I/O ratio greater than 1 suggests that indoor sources were responsible for a portion of the indoor concentration. The ideal data for estimating PEN values would have been measured in the absence of indoor sources and would exhibit I/O ratios less than or equal to 1. To make use of the less than ideal data, researchers developed the following guidelines for evaluating the results of each study.

- Case 1: If a study reports I/O measurement data whose median and/or mean is less than or equal to unity, then this value is used with preference given to the median. If more than one study reports mean and/or median I/O values less than or equal to unity, then the average of the values from these studies are used. Note, however, that median and mean values are not combined to form an average.
- Case 2: For situations not covered by Case 1, and where only the range of I/O values were reported, and the range extended below unity, the *PEN* (I/O) factor was calculated by splitting the difference between the minimum value in the range and 1.0.
- Case 3: For HAP/ME combinations where there was no indication of the I/O ratio being less than unity, 1.0 was used as a default value.

Table 6-25. HAP Groups for Use in HAPEM4

HAP ID No.	Pollutant	Lifetime Group ^b	Source Distribution Group ^c	Predominant Ambient Sources
1	acetaldehyde	1	LN	motor vehicles (rxn product)
2	acrolein	1	LN	motor vehicles (rxn product)
3	acrylonitrile	4 ^d	PS	chemical manufacturing facilities (vehicle exhaust)
4	arsenic compounds	5	PS	smelters; semiconductor manufacturing
5	benzene ^a	5	LN	motor vehicles; chemical manufacturing
6	beryllium compounds	5	PS	smelters; metal manufacturing
7	1,3-butadiene	1	LN	motor vehicles; chemical manufacturing
8	cadmium compounds	5	PS	smelters; power generation
9	carbon tetrachloride	5	PD	various manufacturing facilities; agricultural chemicals
10	chloroform	5	PD	dry cleaning
11	chromium compounds	5	PS	smelters; metal manufacturing facilities; power generation
12	coke oven emissions	--	PS	blast furnaces and steel mills
13	1,2-dichloroethane (ethylene dichloride)	5	PD	metal degreasing; chemical preparations
14	1,3-dichloropropene	1	AR	agricultural soil fumigation
15	ethylene dibromide (dibromoethane)	5	PS	chemical manufacturing
16	ethylene oxide	5	AR	agricultural fumigation
17	formaldehyde	3	LN	motor vehicles (rxn product)
18	hexachlorobenzene	5	AR	hydraulic cement; agricultural use as pesticide
19	hydrazine	1	PS	chemical manufacturing
20	lead compounds - organic	--	--	sources unknown (formerly motor vehicles; lead additives)
21	lead compounds - inorganic	5	PS	smelters; manufacturing facilities; power generation
22	manganese compounds	5	PS	various manufacturing

HAP ID No.	Pollutant	Lifetime Group ^b	Source Distribution Group ^c	Predominant Ambient Sources
23	mercury compounds	5	PS	power generation; various manufacturing facilities
24	methylene chloride (dichloromethane)	5	PS	chemical and semiconductor manufacturing; laboratories
25	nickel compounds	5	PS	smelters; power generation; various manufacturing
26	7-PAH (2 categories: lower and upper bound)	2 ^d	AR	residential burning (fireplaces, grills); diesel exhaust
27	polychlorinated biphenyls (aroclers)	5	PD	waste incineration and disposal; atmospheric redistribution
28	propylene dichloride (1,2-dichloropropane)	5	PD	dry cleaning; various manufacturing facilities
29	quinoline	3	PS	various manufacturing facilities
30	styrene	1	PS	polymer manufacturing facilities
31	2,3,7,8-TCDD (2 categories: lower and upper bound)	3	AR	residential burning; waste incineration; manufacturing; atmospheric redistribution
32	1,1,1,2-tetrachloroethane	5	PD	metal degreasing
33	tetrachloroethylene (perchloroethylene)	5	PD	dry cleaning; metal degreasing
34	trichloroethylene	5	PD	chemical manufacturing
35	vinyl chloride	2 ^d	PS	chemical manufacturing

^a Pollutants that have a long atmospheric lifetime and a line source distribution. These pollutants are expected to have ME factors similar to those developed for CO.

^b Estimated atmospheric lifetimes. 1 = short (<1 day); 2 = medium short; 3 = medium (1-5 days); 4 = medium long; 5 = long (>5 days).

^c Expected distribution of ambient sources. AR = area sources (e.g., residential fireplaces); LN = line sources (e.g., roadways); PD = point sources, densely distributed (e.g., dry cleaning establishments); PS = point sources, sparsely distributed (e.g., smelters, manufacturing facilities).

^d Were placed in Group 3 for estimation of ME factors.

Appendix B of this report presents the resulting estimates of *PEN* together with estimates of *PROX* and *MULT* (defined by the expression $MULT = PEN \times PROX$). Each page of the appendix provides estimates by microenvironment for one pollutant. Although a column has been provided for the additive factor (*ADD*), no estimates of this factor are currently presented. Note also that *PEN* factors for coke oven emissions (HAP #12) and MTBE could not be estimated since there was insufficient information to categorize them using the lifetime grouping methodology. The *Data Code* column indicates how the *PEN* and *PROX* factors were derived for each HAP/ME combination:

- 1 – indicates that measurement data for that particular combination were available and used,
- 2 – indicates that useful measurement data for the HAP/microenvironment pair were not available, and the *PEN* or *PROX* factor was obtained through the grouping scheme described in Subsection 6.6.3.2,
- 3 – indicates that a default value of 1.0 was used. Note that for the *PEN* factors, the use of this default value only applies to those outdoor microenvironments that were in Groups 3 and 4 (see Subsection 6.6.3.2).

Finally, the codes in the *Ref. Sources* column are directly matched to the *Sources* codes in Appendix B of the report by ICF Consulting and TRJ Environmental (2000).

6.6.5 Discussion

Equation 6-52 is acknowledged to be an oversimplification of the physical processes that affect ME concentrations. In addition, Equation 6-52 does not include probabilistic components to account for the inherent variability of these processes. Ideally, ME concentrations would be estimated through the use of a probabilistic mass balance algorithm similar to those incorporated in recent versions of pNEM/O₃ (Johnson, Capel, McCoy, and Mozier, 1996) and pNEM/CO (Johnson, Mihlan, LaPointe, et al., 2000). These exposure models provide short-term (i.e., less than 1 hour) estimates of ME concentrations which account for time-varying changes in outdoor (ambient) concentration, air exchange rate, filtration, decay and other removal processes, and indoor emission rates. Values for the various parameters are selected from appropriate distributions that are specific to pollutant and microenvironment.

A mass-balance modeling algorithm was not implemented in HAPEM4 because (1) the refinement would have greatly increased computer program run times and (2) available data were judged inadequate for developing distributions for many of the algorithm parameters. In addition, the initial NATA assessment was focused on annual exposures which were assumed to be relatively insensitive to the short-term variations in ME concentrations provided by mass balance models. This assumption was based on the observation that the relationship between ME concentration and ambient concentration tends to become more linear as averaging time increases. For example, Dockery and Spengler (1981) analyzed a mass balance model applicable to particulate matter and found that the long-term relationship between indoor and outdoor concentration could be approximated by a simple linear function. The coefficients of this function were determined by performing linear regression analyses on empirical data.

The assumption of linearity has been used in other exposure models applied to HAPs. For example, the Benzene Exposure and Absorbed Dose Simulation (BEADS) model developed by MacIntosh, Xue, Ozkaynak, et al. (1995) estimated the benzene concentration inside vehicles by the expression:

$$\text{benzene inside vehicle} = \alpha + (\beta)(\text{ambient benzene})$$

where the coefficients α and β were randomly selected from normal distributions. The means of these distributions were based on the results of a regression analysis performed by Chang, Ozkaynak, Spengler, and Sheldon (1991) on empirical data collected during a field study in Raleigh, North Carolina.

To summarize, mass balance models are preferred when short-term exposure estimates are required and when analysts are able to identify appropriate data for estimating model parameters. Linear models such as Equation 6-52 are considered more appropriate for screening level analyses such as the initial NATA that attempts to estimate long-term (e.g., annual) exposures. Compared to mass balance models, linear models typically require less data for parameter estimation and run faster.

6.7 References for Section 6

Automotive News Market Data Book Supplement. May 1999. "1998 U.S. Car Sales by Segment." p. 33.

Bureau of Census. 1995. "American Housing Survey for the Denver Metropolitan Area in 1995." Current Housing Reports. Report No. H170/95-46. U. S. Department of Commerce and U.S. Department of Housing and Urban Development.

Bureau of Census. 1997. "American Housing Survey for Los Angeles-Long Beach Metropolitan Area in 1995." Current Housing Reports. Report No. H170/95-7. U.S. Department of Commerce and U.S. Department of Housing and Urban Development.

Carlson, T.R., and T.C. Austin. 1997. Development of Speed Correction Cycles. Sierra Research, Inc. Report No. SR-97-04-01.

Chang, C.C., H. Ozkaynak, J. D. Spengler, and L. Sheldon. 1991. "Driver exposure to volatile organic compounds, CO, ozone, and NO₂ under different driving conditions." Environ. Sci. Technol., Volume 25, 964-972.

Cohen J., T. Johnson, and A. Rosenbaum. 1999. "Algorithm for Air Exchange Rates Inside Motor Vehicles in pNEM/CO." Memorandum to Harvey Richmond, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. November 17.

Davidson, C. L., J. E. Borrazzo, and C. T. Hendrickson. 1987. Pollutant Emission Factors for Gas Stoves. Report No. EPA/600/9-87-005. U. S. Environmental Protection Agency. Research Triangle Park, North Carolina.

Dockery, D.W. and J. D. Spengler. 1981. "Indoor-outdoor relationships of respirable sulfates and particles." Atmos. Environ., Volume 15, pp. 335-343.

EPA. 2001. National-Scale Air Toxics Assessment for 1996. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Fletcher, K., J. LaPointe, T. Long, and T. Johnson. 1999. "Memo No. 1: Review of scientific literature relating to microenvironmental factors." TRJ Environmental memorandum to EPA. EPA purchase order #8D-2091-NALX, April 5, 1999.

Grimsrud, D. T., M. H. Sherman, and R. C. Sondregger. 1982. "Calculating Infiltration: Implications for a Construction Quality Standard," Proceedings of the ASHRAE/DDE Conference: Thermal Performance of the Exterior Envelopes of Buildings II, Las Vegas, Nevada.

Hayes, S. R. 1991. "Use of an Indoor Air Quality Model (IAQM) to Estimate Indoor Ozone Levels." Journal of the Air and Waste Management Association, Vol. 41, pp.161-170.

Hayes, S.R. 1989. "Estimating the Effect of Being Indoors on Total Personal Exposure to Outdoor Air Pollution." Journal of the Air Pollution Control Association, Vol. 39, No. 11, pp. 1453-1461.

Hayes, S. R. and G. W. Lundberg. 1985. Further Improvement and Sensitivity Analysis of an Ozone Population Exposure Model. Report No. SYSAPP-85/061, Systems Applications, Inc., San Rafael, California.

ICF Consulting and TRJ Environmental. 2000. Development of Microenvironmental Factors for the HAPEM4 in Support of the National Air Toxics Assessment (NATA). Report prepared for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, May 8.

Insurance Institute of Highway Safety. 1998. "Utility Vehicles Overtake Passenger Vehicles." Research Alert, Volume 16, No. 7, p. 7 (April 3, 1998).

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014. U.S. Environmental Protection Agency, Research Triangle Park, N.C.

Johnson, T. R. 1987. A Study of Human Activity Patterns in Cincinnati, Ohio. Report prepared by PEI Associates, Inc. for Electric Power Research Institute, Palo Alto. Available from Ted Johnson, TRJ Environmental, Inc., 713 Shadylawn Road, Chapel Hill, North Carolina, 27514.

Johnson, T. 1998. "Analyses Supporting Proposed Modifications to the Mass-balance Algorithm and Gas Stove Algorithm in pNEM/CO." Memorandum dated September 23,

1998, prepared by TRJ Environmental, Inc., under EPA Contract No. 68-D6-0064.

Johnson, T. 1999. "Distributions for Use with the Mass-balance and Gas Stove Algorithms in Applying pNEM/CO to Los Angeles." Memorandum dated April 21, 1999, prepared by TRJ Environmental, Inc., under EPA Contract No. 68-D6-0064.

Johnson, T., J. Capel, and M. McCoy. 1996. Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM and 1990 Population Data. Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., J. Capel, M. McCoy, and J. Mozier. 1996. Estimation of Ozone Exposures Experienced by Outdoor Workers in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Johnson, T., J. Capel, J. Mozier, and M. McCoy. 1996. Estimation of Ozone Exposures Experienced by Outdoor Children in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Johnson, T., J. Capel, R. Paul, and E. Olaguer. 1991. A Comparison of the Hazardous Air Pollutant Exposure Model (HAPEM) and the Human Exposure Model (HEM). Report prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM. Report prepared by International Technology Air Quality Services under EPA Contract No. 68-D0-0062. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., M. McCoy, and J. Capel. 1993. Enhancements to the Hazardous Air Pollutant Exposure Model (HAPEM) as Applied to Mobile Source Pollutants. Report prepared by IT Air Quality Services for the Atmospheric Research and Exposure Assessment Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., G. Mihlan, J. Lapointe, K. Fletcher, J. Capel, A. Rosenbaum, J. Cohen, and P. Stiefer. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared under Contract No. 68-D6-0064 for the U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC.

Johnson, T., R. Paul, and J. Capel. 1992. Application of the Hazardous Air Pollutant Exposure Model (HAPEM) to Mobile Source Pollutants. Report prepared by IT Air Quality Services for the Office of Mobile Sources, U. S. Environmental Protection Agency, Ann Arbor, Michigan, August.

Johnson, T., M. Weaver, J. Mozier, et al. 1998. "Variation of Residential Air Exchange Rates Under Scripted Ventilation Conditions." Presented at the International Symposium on the Measurements of Toxic and Related Pollutants, Air and Waste management Association, September 1 - 3, 1998, Cary, North Carolina.

Koupal, J. 1998. Air Conditioning Activity Effects in MOBILE6. Report No. M6.AE.001. Office of Mobile Sources, U.S. Environmental Protection Agency.

Lagus Applied Technology, Inc. 1995. Air Exchange Rates in Non-Residential Buildings in California. Document No. P400-91-034. Prepared for the California Energy Commission under Contract No. 400-91-034.

Long, T. and T. Johnson. 1999. "Memo No. 2: Classification of hazardous air pollutants into groups likely to exhibit similar microenvironmental factors." TRJ Environmental memorandum to EPA. EPA purchase order #8D-2091-NALX, April 6, 1999.

MacIntosh, D.L., J. Xue, H. Ozkaynak, J. D. Spengler, and P. B. Ryan. 1995. "A population-based exposure model for benzene." J. of Exposure Anal. and Environ. Epidem. Volume 5, pp. 375-402.

Menkedick, J., N. Niemuth, P. Hartford, and D. K. Landstrom. 1993. Metered Ranges, Cooktops, and Ovens in the Northern Illinois Gas Residential Load Study Data Base. GRI Report No. 93/0204. Prepared by Battelle for the Gas Research Institute, 8600 West Bryn Mawr Avenue, Chicago, IL 60631.

Moschandreas, D. J., J. Zabransky, and D. J. Peltan. 1981. Comparison of Indoor and Outdoor Air Quality. Report No. EA-1733, Electric Power Research Institute.

Murray, D. M., and D. E. Burmaster. 1995. "Residential Air Exchange Rates in the United States: Empirical and Estimated Parametric Distributions by Season and Climatic Region." **Risk Analysis**, Vol. 15, No. 4, 459-465.

Nagda, N. L., H. E. Rector, and M. D. Koontz. 1987. Guidelines for Monitoring Indoor Air Quality. Hemisphere Publishing Corporation, Washington, D. C.

Ott, W., P. Switzer, and N. Willis. 1994. "Carbon Monoxide Exposures Inside an Automobile Traveling on an Urban Arterial Highway." Journal of the Air and Waste Management Association, Volume 44, pp. 1010 - 1018.

Pandian, D.P., W. R. Ott, and J. V. Behar. 1993. "Residential Air Exchange Rates for Use in Indoor Air and Exposure Modeling Studies." Journal of Exposure Analysis and Environmental Epidemiology, Vol. 3, No. 4, pp. 407-416.

Pandian, D. P., J. V. Behar, W. R. Ott, L. A. Wallace, A. L. Wilson, S. D. Colome, and M. Koontz. 1998. "Correcting Errors in the Nationwide Data Base of Residential Air Exchange Rates." Journal of Exposure Analysis and Environmental Epidemiology, Vol. 8, No. 4, pp. 577-586.

Peterson, G.A. and R.H. Sabersky. 1975. "Measurements of Pollutants Inside an Automobile." Journal of the Air Pollution Control Association, Vol. 25, No. 10, pp. 1028-1032, October.

Repace, J.L., J. Jinot, S. Bayard, K. Emmons, and S. K. Hammond. 1998. "Air Nicotine and Saliva Cotinine as Indicators of Workplace Passive Smoking Exposure and Risk." Risk Analysis, Volume 18, pp. 71 - 83.

Rodes, C, L. Sheldon, D. Whitaker, A. Clayton, K. Fitzgerald, J. Flanagan, F. DiGenova, S. Hering, and C. Frazer. 1998. "Measuring Concentrations of Selected Air Pollutants Inside California Vehicles." Final Report for Contract No. 95-339. California Air Resources Board, Sacramento, California.

SAI (Systems Applications International). 1999. User's guide: Assessment System for Population Exposure Nationwide (ASPEN, Version 1.1) – Volumes I and II. Report No. SYSAPP-98/25r2. Prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina.

Turk, B. H., D. T. Grimsrud, J. T. Brown, K. L. Geisling-Sobotka, J. Harrison, and R. J. Prill. 1989. "Commercial Building Ventilation Rates and Particle Concentrations." ASHRAE Transactions, Vol. 95, Part 1.

Wallace, L., and W. Ott. 1996. "Air Exchange Rate Experiments in a Detached House Using a Continuous Monitor." Presented at the Annual Meeting of the SRA/ISEA, New Orleans.

Weschler, C. J., H. C. Shields, and D. V. Naik. 1992. "Indoor Ozone: Recent Findings," Tropospheric Ozone in the Environment II, editor, R. Burglund, Air and Waste Management Association, Pittsburgh, PA, pp. 681 - 700.

SECTION 7

PROCESSING AMBIENT AIR QUALITY DATA

The exposure models discussed in this report typically require data representing ambient air pollutant levels with an averaging time of one hour. Ambient air quality data for the NEM-series models are usually obtained from fixed-site monitoring stations located in and around the study area defined for the exposure assessment. The HAPEM-series models tend to use data obtained from dispersion models, although fixed-site monitoring data have been used in some applications. This section focuses primarily on methods used in selecting, validating, and adjusting air quality data obtained from fixed-site monitors, as these data typically require more processing than data acquired from dispersion models. The reader should note, however, that many of these methods discussed in this section can be adapted to air quality data obtained from dispersion models.

7.1 Selection of Data

Subsection 2.1 provides an overview of the process used to select ambient CO data for the application of pNEM/CO to Denver and Los Angeles. The process is illustrative of the general approach employed in selecting air quality data sets for pNEM analyses.

For the Denver application, analysts identified seven monitors which (1) were located within 50 km of the center of Denver, (2) were located in areas of appropriate urban land use, and (3) reported sufficient air quality data for 1995 through 1997. Five of the seven sites were identical to sites used in an analysis of Denver in 1992; the remaining two sites were located in downtown Boulder. The locations of five sites used in the 1992 analysis were considered appropriate for defining five separate exposure districts with 10 km radii. However, the Boulder sites were considered too close together to support separate exposure districts. Consequently, analysts defined six exposure districts -- one for each of the 1992 Denver sites and a "composite" Boulder site. For purposes of constructing the associated exposure district, the composite site was assigned a location midway between the two Boulder sites.

Analysts evaluated the quality and completeness of the data available for the seven monitors for the years 1995 through 1997. Based on this evaluation, EPA selected 1995 as the year for the pNEM/CO analysis of Denver. Each of the selected monitors provided an hourly average data set that was at least 96 percent complete for this year.

A similar approach was used in selecting fixed-site monitors for the application of pNEM/CO to Los Angeles. Analysts began by designating all sites within Los Angeles, Orange, Riverside, San Bernardino, and Ventura Counties which reported CO data between 1995 and 1997 as potential sites for the pNEM/CO analysis. Of the 30 CO sites which met these criteria, 24 satisfied a further requirement that the site reported data that were at least 75 percent complete for each of the three years. Analysts omitted seven of these monitors which were located in outlying areas and reported

relatively low CO levels. EPA evaluated the siting characteristics and locations of the remaining 17 monitors and selected the 10 monitors listed in Table 2-2. These monitors all reported relatively high CO levels, and they appeared to provide good coverage of the highly urbanized areas within greater Los Angeles.

EPA selected 1997, the most recent of the three years evaluated, as the year for the pNEM/CO analysis of Los Angeles. All 10 sites had adequate data completeness for 1997.

In following this approach, researchers were attempting to identify air quality data that met the following requirements:

- (1) The spatial scale of each monitor is well characterized. The data reported by a particular monitor is typically assumed to apply to an “exposure district” centered on the monitor with a radius of X. Ideally, the researcher will have sufficient data on the site characteristics of the monitor (nearby land use, traffic patterns, etc.) to make a reasonable estimate of X.
- (2) The exposure districts determined for individual monitors span a sufficient portion of the area of interest. Ideally, each census tract in the area of interest (e.g., the Denver metropolitan area) will fall within an exposure district associated with one of the selected monitoring sites. In cases where some tracts fall outside the exposure districts, the analyst has the options of foregoing the assessment, expanding one or more of the exposure districts to include the orphan tracts, or defining the study area as only those tracts that fall within the exposure districts as initially defined. (The third option has been implemented in recent versions of pNEM/O3 and pNEM/CO.)
- (3) The data for each monitoring site represent an appropriate time period. Historically, researchers have given preference to monitoring data that are both recent and “typical” when selecting data sets for pNEM analyses. Such data are desirable because analysts frequently adjust air quality data to simulate conditions that may occur in the near future. These adjustment procedures are frequently based on the assumption that the future emission and meteorological patterns within a particular study area will be similar to those of the historical time period during which the monitoring data were collected.
- (4) The data for each monitoring period are relatively complete. The majority of exposure models discussed in this report require that each input site-year of monitoring data be complete (gapless). Because most data sets obtained from fixed-site monitors include gaps, researchers have developed methods such as the time series model described in Subsection 7.3 to estimate the missing values prior to model input. These methods perform better when the absolute number of missing values is small and the duration of each gap is small. Data sets missing more than 25 percent of hourly values are seldom used in pNEM exposure assessments.

All other factors being equal, researchers tend to give less preference to monitors sited

in outlying areas, particularly when these areas are characterized by relatively low pollutant concentrations.

7.2 Identification of Data Anomalies

Air quality data sets containing erroneous values can bias the results of exposure studies, especially if the errors occur as extreme values. A report by Nelson, Armentrout, and Johnson (1980) provides a number of techniques for validating data prior to use. Table 7-1 provides a summary of the techniques applicable to the types of data likely to be used in an exposure analysis.

In past NEM analyses, three techniques were used successfully to screen data for anomalous values: (1) the Gap Test, (2) the Pattern Test, and (3) visual inspection. The Gap Test and the Pattern Test were implemented using a standardized data review program developed by EPA's Monitoring and Reports Branch (MRB) referred to as the MRB Validation Report.

7.2.1 The Gap Test

The Gap Test is a statistical analysis of data over a one-month period. This analysis assumes that the data can be modeled reasonably well by a smooth probability distribution curve. Two exponential curves are fit to the data using the 50th and 95th percentiles of the data for one fit and the 50th and 99.9th percentiles for the other fit. Both fits emphasize the upper tail of the distribution. All data values are arranged in order of magnitude and the program examines "gaps" in the data, i.e., large differences between succeeding ordered values. Using the fitted distribution functions, the program calculates the probability that a gap of magnitude "x" could be obtained by chance. A gap that is greater than would be expected from the underlying assumptions is flagged as a data anomaly.

The strength of the Gap Test is that the criteria for identifying anomalies are based on an analysis of the data set itself. Its weakness is that the assumed distribution is not always appropriate; consequently some false failures may occur.

7.2.2 The Pattern Test

The Pattern Test is composed of five subtests performed on each 24-hour period in the data base.

1. High-Value Test: This test flags an hourly value that exceeds a predetermined limit. For CO, the criterion depends on whether or not the measurement was taken during rush hours (7 to 11 am or 4 to 9 pm). During rush hours the limit was 66 ppm; at other times the limit was 44 ppm.
2. Adjacent-Hour Difference Test: This test assumes that the data take the form of an auto-correlated time series, i.e., a large jump or drop in the values within 1 hour is not expected. If there is a jump or drop greater than 22 ppm, the suspect values is flagged.

Table 7-1. Techniques for Validating Data.

Technique (subsection of Nelson, Armen- trout, and Johnson, 1980)	Brief description	Reference
Data identification checks (3.1.1)	Check that data identification fields are correct. For example, critical identification codes associated with air quality data from Aerometric Information Retrieval System (AIRS) include (1) time and date, (2) location, (3) sampling/analytical method code, (4) pollutant method interval code, (5) parameter, and (6) decimal.	
Unusual event review (3.1.2)	Ask data source whether a log was maintained to record extrinsic events (construction activity, dust storms, forest fires, unusual traffic volume, etc.). Review log to determine whether log entries can explain unusual data.	
Deterministic relationship checks (3.1.3)	Data sets that contain two or more physically or chemically-related parameters should be routinely checked to ensure that the measured values on an individual parameter do not exceed the measured values of an aggregate parameter that includes the individual parameter. For example, NO ₂ values should not exceed NO _x values recorded at the same time and location.	
Data processing procedures (3.1.4)	<p>The following techniques can be used to improve data processing:</p> <ol style="list-style-type: none"> 1. Context and staged edits (e.g., a field edit for checking the data values against the field specification for length, character set, and value range). 2. Addition of quality flags to items in a data base to condition processing to avoid a mismatch between the data quality and its use. 3. Redundancy in batches, files, and inputs to improve reliability 4. Checks on data sequence (e.g., input data are checked for correct time sequence) 5. Editing by classification of category, class limits, normal limits, and trend limits. For example, the behavior of an individual data item can be compared to its previous behavior or the aggregate of individuals in its group. Procedures of this kind are included in Subsections 3.2 and 3.3 of Nelson, Armentrout, and Johnson (1980). 6. Parallel check calculations, useful when the same results can be obtained by two independent calculation procedures. 7. Built-in test data, verification tests, and diagnostics to provide a test environment without the risk of allowing an unchecked program access to real files. 8. On-line testing to exercise fault detection logic. 9. Clearly defined organizational responsibilities to ensure that the correct data validation procedures are continuously employed. 	U. S. Department of Commerce (1978)

Technique (subsection of Nelson, Armen- trout, and Johnson, 1980)	Brief description	Reference
Data flow diagram (3.1.4)	Analysts should construct a data flow diagram that indicates the steps in data handling at which an error could occur. In general these steps can be classified as (1) data input, (2) data transmission, (3) data processing, and (4) data output.	
Data input checks (3.1.4)	Check for data entry errors and mislabeled computer files. Include data input with computer output to facilitate review.	
Data transmission checks (3.1.4)	The principal result of transmission error is the loss or alteration of data. A simple check is to transmit the data a second time and compare the data streams.	
Data processing and data output checks (3.1.4)	Errors of this kind are typically caused by the computer programs that manipulate the data files, perform mathematical calculations, and format the output results. A standard method of checking for processing errors is to make up a small, typical data set, perform the appropriate data manipulations and calculations by hand, and compare with the results from the data processing system.	
Tests for Internal Consistency (3.2)	Internal consistency tests include data plots, the Dixon ratio test, the Grubbs test, and the gap test. These tests check for values in a data set which appear atypical when compared to values of the whole data set. Common anomalies of this type include unusually high or low values (outliers) and large differences in adjacent values. These tests will not detect errors which alter all values of the data set by either an additive or multiplicative factor (e.g., an error in the use of the scale of a meter or recorder).	
Data plots (3.2.1)	Data plotting (including strip chart records) is one of the most effective means of identifying possible data anomalies.	
Dixon Ratio Tests (3.2.2)	The Dixon ratio tests are the simplest of the statistical tests recommended for evaluating the internal consistency of data. Nelson et al. describes tests of (1) the largest value in a data set and (2) the largest pair of values in a data set. Both procedures test the assumption H_0 that the largest value(s) in a sample are consistent with the spread of the data. Other Dixon ratio tests which may be of interest to the data analyst are described in Barnett and Lewis (1978) and USEPA (1978).	Barnett and Lewis (1978), USEPA (1978)
Grubbs Test (3.2.3)	Like the Dixon ratio test, the Grubbs test can be used to determine whether the largest observation in a sample from a normal distribution is too large with respect to the internal consistency of the data set. The Grubbs test differs from the Dixon ratio test in that the test statistic is calculated using all of the values of the data set.	Barnett and Lewis (1978), Grubbs and Beck (1972)

Technique (subsection of Nelson, Armen- trout, and Johnson, 1980)	Brief description	Reference
Gap test (3.24)	The gap test identifies spurious outliers by examining the frequency distribution for large gaps. The occurrence of a gap length larger than a predetermined critical value indicates a possible data anomaly.	USEPA (1978), Curran, Hunt, and Faoro (1977)
Johnson "p" test (3.2.5)	In this test, the analyst fits a distribution to a data set and then uses the parameters of the distribution to determine an acceptance range for the largest value. If the largest value falls outside this range, it is flagged for further evaluation.	
Tests for Historical Consistency (3.3)	Tests for historical consistency check the data set with respect to similar data recorded in the past.	
Gross Limit Checks (3.3.1)	Gross limit checks are useful in detecting data values that are either highly unlikely or generally considered impossible. Upper and lower limits are developed by examining historical data for a site (or other sites in the area). Whenever possible, the limits should be specific to each monitoring site and should consider both the parameter and instrument/method characteristics. Although this technique can be easily adapted to computer application, it is particularly appropriate for technicians who reduce data manually or who scan strip charts to detect unusual events.	
Pattern and Successive Value tests (3.3.2)	The pattern and successive value tests check the data for pollutant behavior which has never or very rarely occurred. Like the gross limit checks, they require that a set of boundary values or limits be determined empirically from pre-screened historical data. Values representing pollutant behavior outside of these predetermined limits are then flagged for further investigation. EPA has recommended the use of pattern tests which place upper limits on (1) the individual concentration value (maximum hour test), (2) the difference in adjacent concentration values (adjacent hour test), (3) the difference or percentage difference between a value and both of its adjacent values (spike test), and (4) the average of four or more consecutive values (consecutive value test).	USEPA (1978)
Parameter relationship test (3.3.3)	Parameter relationship tests can be divided into two main categories: deterministic tests involving the theoretical relationships between parameters (e.g., $NO < NO_x$) and empirical tests which check whether or not a parameter is behaving normally in relation to the observed behavior of one or more other parameters.	
Shewhart Control Chart (3.3.4)	The Shewhart control chart is a valuable supplement to gross limit checks and patterns tests in that it identifies data sets which have a mean or range values that are inconsistent with past data sets.	Grant and Leavenworth (1996)

Technique (subsection of Nelson, Armen- trout, and Johnson, 1980)	Brief description	Reference
Tests for Consistency of Parallel Data Sets (3.4)	These tests are used to identify a systematic bias by comparing the data set under evaluation with other data sets which presumably have been sampled from the same population. Examples include (1) the sign test, (2) the Wilcoxon signed-rank test, (3) the Wilcoxon sum test, and (4) the intersite correlation test.	Siegel (1956)
The Sign test (3.4.1)	The sign test evaluates the assumption H_0 that two related (paired) samples, such as data sets from adjacent monitoring instruments, have the same median.	Siegel (1956)
Wilcoxon Signed- Rank Test (3.4.2)	Like the sign test, the signed-rank can be used to test the assumption H_0 that two samples come from populations having the same median. The Wilcoxon test is generally more powerful than the sign test since it considers both the sign and the magnitude of the difference between paired data.	Siegel (1956)
Rank sum test (3.4.3)	The rank sum procedure is useful in testing the assumption H_0 that two samples represent populations with the same distribution. Unlike the sign test and ranked sign test, the rank sum test is applicable to independent (unrelated) samples.	Remington and Schork (1970)
Intersite Correlation test (3.4.4)	The intersite correlation test is suggested as a means of comparing two correlated parameters being measured at the same site or at neighboring sites. The analyst plots parameter B vs. parameter A as a scatter plot and then uses the calculated correlation between the parameters to determine a probability ellipse that should contain 95 percent of the data points. If a point falls outside the ellipse, both values (A and B) should be flagged for checking.	Hald (1952)

3. Dixon-Ratio Test: This is a statistical analysis of the highest and lowest value found during the day. If A is the difference between the two highest values and B is the range of values for the day, then the Dixon Ratio is A/B. If this ratio is statistically significant, the suspect values are flagged.
4. Spike Test: The difference between a suspect hourly value and the preceding and following value is measured. If either of these two differences is greater than 20 ppm, or if the suspect value is 500 percent greater than either of the adjacent values, the suspect value is flagged.
5. Consecutive High Values: It is unusual for a long series of hourly measurements to remain at a high level. If four consecutive hourly values are greater than 40 ppm, the data are flagged.

Analysts evaluated the results of applying MRB Validation Program to SO₂ data in four cities (Paul, Faoro, and Hunt, 1981). This analysis led to the development of a revised validation program. This new program (denoted MRB-2) contained improvements in the patterns tests and an option for graphical output of flagged data. The Dixon Ratio Test was enhanced for all pollutants according to the recommendations of the data validation manual prepared by Nelson, Armentrout, and Johnson (1980). In the new version, the formulas for testing the high and low hourly values from each 24-hour period vary according to the number of hourly measurements recorded during the period. The spike test was enhanced so that an individual hourly value is compared to the two (rather than one) preceding values and the two succeeding values. In addition, MRB-2 incorporates a preliminary screening test, so that if all values in a 24-hour period are low, the pattern tests are skipped. By passing over data that do not require testing, the computer program runs more quickly.

The strength of the pattern test program is that it uses more than one test to identify possible anomalies. Its weakness is that the criteria for identifying anomalies are predetermined for all tests except the Dixon Ratio Test. The pattern test program could be improved by adjusting test criteria according to the data reported by each individual monitoring site.

7.2.3 Visual Inspection

Analysts developed a computer program that plotted all hourly values in a year of data on a single graph. Such data plots are useful in identifying unusually high values and long strings of identical values.

7.3 Estimation of Missing Values

The pNEM/CO model is typical of pNEM-type models in that it requires that each input site-year of monitoring data be complete (gapless). Historically, a time series model developed by Johnson and Wijnberg (1981) has been used to generate estimates for the missing values in each input data set used in a pNEM exposure analysis. The time series model is based on the assumption that hourly-average air quality values can be represented by a combination of cyclical, autoregressive, and

random processes. The parameter values of these processes are determined by a statistical analysis of the reported data.

Subsection 7.3.1 provides a theoretical description of the time series model. The model is applied in two stages. In the first stage, the analysts makes an initial estimate of the each missing value using one or more interpolation procedures. This process yields the initial augmented data set. In the second stage, the time series model is fit to the initial augmented data set and the resulting model parameter estimates are used to generate an alternative set of estimates for the missing values. These alternative estimates are inserted into the data set as appropriate to yield the final augmented data set. Subsection 7.3.2 discusses procedures used in past pNEM analyses to obtain initial (first-stage) estimates of missing values. Subsection 7.3.3 provides a nine-step procedure for conducting the entire two-stage process. Examples of data before and after implementation of the procedure are presented in Subsection 7.3.4.

7.3.1 The Time Series Model

A complete year of hourly average data takes the form of a time series

$$x_1, x_2, \dots, x_t, \dots, x_n$$

where $n = 8760$. We can fit this time series exactly by the model

$$x_t = \bar{x} + \sum_{j=1}^{4380} R_j \cos(\omega_j t + \theta_j) \quad (7-1)$$

where \bar{x} is the arithmetic mean of the series, R_j and θ_j are amplitude and phase angle values determined by Fourier analysis, and $\omega_j = 2\pi j/8760$. Omission of one or more of the 4380 Fourier cosine terms will yield an approximate fit. Because Fourier cosine functions are orthogonal and because the contribution of each cosine function to the representation of the original time series is proportional to its amplitude R_j , we can provide a least-squares fit to the original time series with m cosine terms by using the cosine terms with the m largest amplitudes. We denote each term in this estimated time series as \hat{x}_t where

$$\hat{x}_t = \bar{x} + \sum_{i=1}^m R_i \cos(\omega_i t + \theta_i) \quad (7-2)$$

and R_i , ω_i , and θ_i are the parameters of the Fourier term having the i th largest amplitude. For convenience, we will refer to the m Fourier terms in Equation 7-2 as the essential cyclical component (ECC).

The differences between the x_t series and the \hat{x}_t series comprise the d_t series, i.e.,

$$d_t = x_t - \hat{x}_t \quad (7-3)$$

We can define how well the \hat{x}_t series represents the x_t series by the goodness-of-fit

statistic

$$r^2 = 1 - \left[\sum_{t=1}^{8760} d_t^2 \right] / \left[\sum_{t=1}^{8760} (x_t - \bar{x})^2 \right] \equiv \left[\sum_{i=1}^m R_i^2 \right] / \left[\sum_{j=1}^{4380} R_j^2 \right]. \quad (7-4)$$

As m increases, r^2 increases and the goodness of fit improves. Note that $r^2 = 1$ when $m = 4380$.

If the x_t series exhibits autocorrelation, the d_t series is likely to exhibit autocorrelation. One means of characterizing a series that exhibits autocorrelation is to use an autoregressive process of order p [denoted AR(p)]. In this case, each d_t term can be expressed as

$$d_t = a_t + \phi_1 d_{t-1} + \phi_2 d_{t-2} + \dots + \phi_p d_{t-p} \quad (7-5)$$

where a_t is a normally-distributed random variate with mean 0 and variance σ_a^2 .

Estimates of $\phi_1, \phi_2, \dots, \phi_p$ can be obtained by first estimating each autocorrelation ρ_k , using the relationship $\hat{\rho}_k = r_k$ where

$$r_k = (c_k)/(c_0) \quad k = 1, 2, \dots, p \quad (7-6)$$

and

$$c_k = (1/8760) \sum_{k+1}^{8760} (d_t - \bar{d})(d_{t-k} - \bar{d}). \quad (7-7)$$

From these estimates, the Yule-Walker estimates of the autoregressive parameters can be obtained (Box and Jenkins, 1976).

Autocorrelation of the d_t series will decrease as m increases since an increasing portion of the x_t series autocorrelation is explained by the cosine functions. We assumed that most of the autocorrelation in the data corresponding to $k \geq 3$ would be contained in the ECC we selected and that an AR(2) process would suffice to characterize the d_t series. In this case,

$$\phi_1 = [(r_1)(1 - r_2)]/(1 - r_1^2), \quad (7-8)$$

$$\phi_2 = (r_2 - r_1^2)/(1 - r_1^2), \quad (7-9)$$

and

$$\hat{\sigma}_a^2 = c_0(1 - \phi_1 r_1 - \phi_2 r_2). \quad (7-10)$$

This AR(2) process represents a stationary time series if it meets certain conditions described by Box and Jenkins (1976).

A theoretical AR(2) process will have non-zero values of ρ_k for $k > 2$ that decrease gradually according to the relationship

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \quad k > 0 \quad (7-11)$$

until a point is reached where the distribution of r_k is approximately normal with mean zero and standard error

$$\sigma(r_k) \doteq [(1/n)(1 + 2\rho_1^2 + 2\rho_2^2)]^{0.5}. \quad (7-12)$$

The values of ρ_1 and ρ_2 are estimated by r_1 and r_2 . No more than 5 percent of the values of r_k for large values of k should deviate from zero by more than two standard errors (Box and Jenkins, 1976).

If we select an ECC such that the autocorrelations in the d_t series corresponding to $k > 2$ are consistent with Equations 7-11 and 7-12, then an AR(2) process should suffice to characterize the d_t series. To select this ECC, we can start by determining the d_t series that corresponds to $m = 1$. We then calculate r_k for values of k that are likely to be significant. These include $k = 3, 4, 6, 8, 12, 24,$ and 168 for typical air quality data. If the r_k values are not consistent with Equations 7-11 and 7-12, we determine the d_t for $m = 2$ and repeat the analysis. We continue increasing m until the r_k values for $k > 2$ meet our criteria. At this stage, we should have a combination of ECC and AR(2) process that will adequately characterize the data.

7.3.2 Initial Treatment of Missing Values

Fourier analysis cannot be applied to a time series if one or more values are missing. If air quality data to be analyzed are incomplete, some method of estimating missing values must be used prior to analysis. Bloomfield (1976) recommends replacing each missing observation by a linear combination of its neighbors in situations where most of the missing values tend to occur in small, isolated groups. If a gap containing $b - 1$ missing values occurs between values x_a and x_{a+b} , each missing value x_t can be estimated by linear interpolation as

$$\hat{x}_t = x_a + (1/b)(t - a)(x_{a+b} - x_a). \quad (7-13)$$

However, linear interpolation may not yield reasonable estimates of missing one-hour values for large gaps, especially if they are bounded by extreme values. In these cases, the arithmetic mean (\bar{x}) may be a better initial estimate of each missing value. Experience with the data sets used in past pNEM/CO analyses suggests that the arithmetic mean should be used to fill in gaps whenever gap length exceeds 72 hours and/or one of the boundary values exceeds the arithmetic mean by more than two standard deviations. In other cases, linear interpolation tends to produce reasonable results.

7.3.3 Procedure for Simulating Missing Values

The time model described above is the basis for the following procedure for simulating missing values.

1. Calculate the mean and standard deviation of the data set.

2. Identify gaps with lengths exceeding 72 hours and/or with boundary values exceeding the arithmetic mean by more than two standard deviations. Fill these gaps with the arithmetic mean.
3. Use linear interpolation to fill in the remaining gaps.
4. Apply Fourier analysis to the augmented time series created in Steps 1 and 2.
5. Construct for an ECC that contains the smallest number of cosine terms required to produce a d_t series consistent with Equations 7-11 and 7-12.
6. Use Equations 7-8, 7-9, and 7-10 to determine the values of ϕ_1 , ϕ_2 , and $\hat{\sigma}_a$ for the AR(2) process that will generate the d_t series.
7. Divide each term in a $N(0,1)$ random series by $\hat{\sigma}_a$ to generate the a_t series.
8. Simulate missing d_t values using the relationship

$$\hat{d}_t = \phi_1 \hat{d}_{t-1} + \phi_2 \hat{d}_{t-2} + a_t \quad (7-14)$$

9. Fill in missing x_t values using the model

$$\hat{x}_t = \bar{x} + \sum_{j=1}^m R_j \cos(\omega_j t + \theta_j) + \hat{d}_t \quad (7-15)$$

to create the final augmented data set.

Note that the final simulation (Equation 7-15) uses the information concerning the cyclical, autoregressive, and stochastic (random) properties of the time series that are omitted in the initial estimates made in Steps 2 and 3.

7.3.4 Examples of Filled-In Data Sets

The procedure described above was applied to hourly nitrogen dioxide data reported by Los Angeles site 050230001101. Figure 7-1 is a computer drawn graph of the data reported by this site for 1977. The dashes below the horizontal axis indicate hours for which no data were reported. The 2508 missing values were initially filled in using linear interpolation for small gaps and the arithmetic mean (57 ppb) for large gaps. Analysis of this augmented data set suggested an ECC with 93 terms was appropriate for characterizing the periodic behavior of the data. Estimates of parameters of the corresponding AR(2) process were $\phi = 0.884$, $\phi_2 = -0.147$, and $\hat{\sigma}_a = 0.0116$. Figure 7-2 shows the data set with missing values filled in using this model.

Before-and-after comparisons of the filled-in data sets used in applying Version 2.1 of pNEM/CO to Denver and Los Angeles (Johnson, Mihlan, LaPointe, et al., 2000) indicate that the distributional characteristics of each data set did not change significantly when missing values were filled in using the procedure proposed above. Table 7-2 provides descriptive statistics by monitoring site for the 1-hour CO concentrations in each Denver data set before and after estimation of the missing values, based on the 1995 data sets selected for the pNEM/CO analysis. The statistics indicate that the addition of missing-value estimates did not significantly affect the distribution of any data set. Each table also provides descriptive statistics for

Top of Figure 7-1

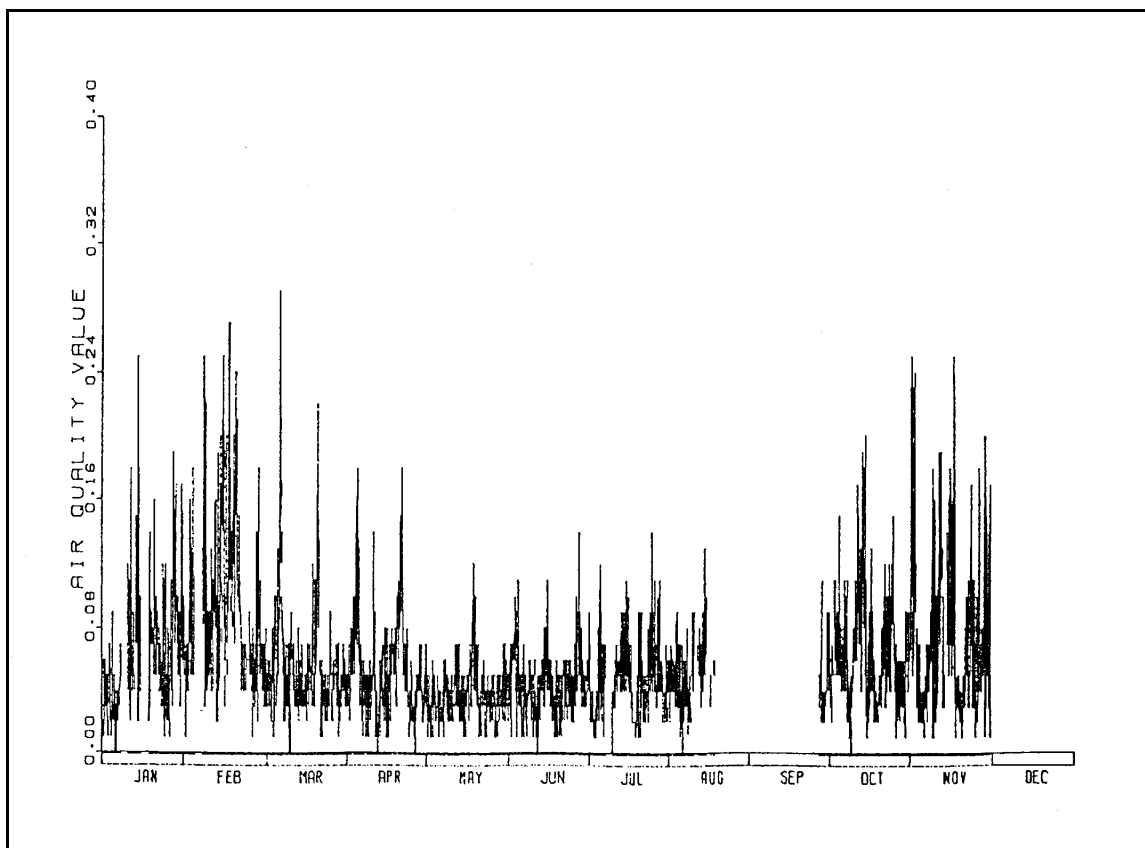


Figure 7-1. Computer plot of 1977 hourly average nitrogen dioxide data (ppm) reported by monitoring site 050230001I01.

Top of Figure 7-2

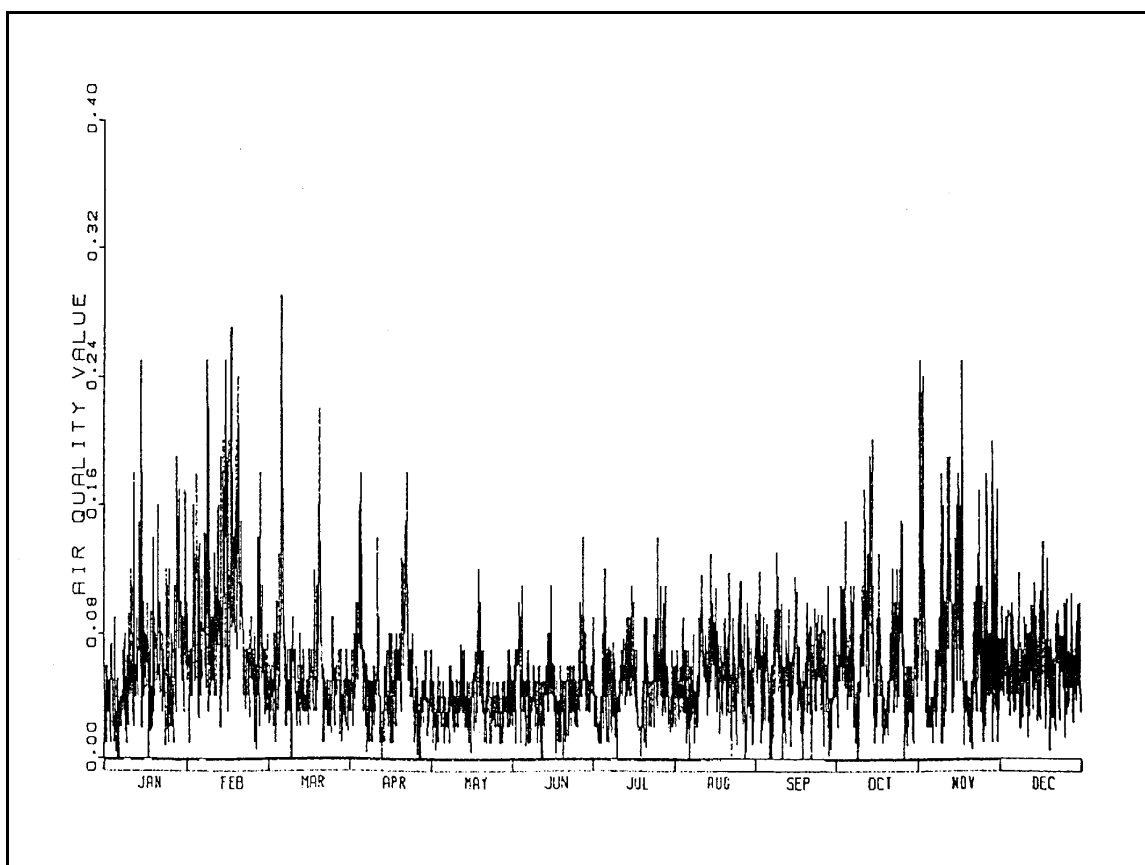


Figure 7-2. Computer plot of 1977 hourly average nitrogen dioxide data (ppm) for monitoring site 050230001101 after final simulation of missing values.

Table 7-2. Descriptive Statistics for One-Hour Carbon Monoxide Concentrations Reported by Denver Monitors Before and After Estimation of Missing Values.

Site	Data set ^a	No. of obs.	Descriptive statistics for 1-hour CO concentrations, ppm						
			50th	90th	95th	99th	99.5th	second largest	largest value
005-002 (Highlands, M)	1 h (o)	8670	0.3	0.7	1.0	1.8	2.2	3.3	3.6
	1 h (s)	8760	0.3	0.7	1.0	1.8	2.2	3.3	3.6
	8 h (s)	8760	0.3	0.7	0.8	1.4	1.7	2.5	2.6
031-0002 (Broadway, A)	1 h (o)	8697	1.2	2.7	3.4	6.1	7.7	16.4	24.5
	1 h (s)	8760	1.2	2.7	3.4	6.1	7.7	16.4	24.5
	8 h (s)	8760	1.3	2.4	3.0	4.7	5.8	10.8	11.0
031-0013 (Albion, C)	1 h (o)	8647	0.9	2.5	3.4	5.5	6.4	13.6	14.6
	1 h (s)	8760	0.9	2.5	3.4	5.5	6.4	13.6	14.6
	8 h (s)	8760	1.1	2.2	2.7	3.7	4.3	8.5	8.5
031-0014 (Julian, B)	1 h (o)	8701	0.7	2.3	3.2	5.3	6.4	9.9	10.4
	1 h (s)	8760	0.7	2.3	3.2	5.3	6.5	9.9	10.4
	8 h (s)	8760	0.8	2.1	2.7	4.1	4.8	7.2	7.3
059-0002 (Arvada, L)	1 h (o)	8680	0.6	2.0	2.7	4.8	5.8	8.9	11.9
	1 h (s)	8760	0.6	2.0	2.7	4.8	5.8	8.9	11.9
	8 h (s)	8760	0.8	1.8	2.3	3.1	3.5	5.0	5.1

Site	Data set ^a	No. of obs.	Descriptive statistics for 1-hour CO concentrations, ppm						
			50th	90th	95th	99th	99.5th	second largest	largest value
013-0010 (Boulder, 28 th Street)	1 h (o)	8608	0.8	2.1	2.8	4.8	5.5	10.3	10.6
	1 h (s)	8760	0.8	2.1	2.8	4.8	5.5	10.3	10.6
	8 h (s)	8760	0.9	1.8	2.2	3.1	3.6	5.2	5.3
013-1001 (Boulder, Marine Street)	1 h (o)	8651	0.4	0.9	1.3	2.3	2.9	8.2	8.3
	1 h (s)	8760	0.4	0.9	1.3	2.3	2.9	8.2	8.3
	8 h (s)	8760	0.5	0.9	1.1	1.8	2.1	3.8	3.9

^a 1 h (o): original 1-hour data set as down-loaded from AIRS.

1 h (s): supplemented 1-hour data set (includes estimates of missing values)

8 h (s): supplemented 8-hour running average data set [based on 1 h (s) data].

Table 7-3. Descriptive Statistics for One-Hour Carbon Monoxide Concentrations Reported by Los Angeles Monitors Before and After Estimation of Missing Values.

Site	Data set ^a	No. of obs.	Descriptive statistics for 1-hour CO concentration, ppm						
			50th	90th	95th	99th	99.5th	second largest	largest value
60370113 (West Los Angeles)	1 h (o)	8360	0.6	2.0	2.6	3.7	4.2	6.4	7.3
	1 h (s)	8760	0.6	2.0	2.6	3.6	4.1	6.4	7.3
	8 h (s)	8760	0.6	1.8	2.2	3.0	3.3	4.1	4.2
60371002 (Burbank)	1 h (o)	8025	1.4	3.5	4.5	6.1	6.6	8.6	8.8
	1 h (s)	8760	1.4	3.5	4.4	6.0	6.6	8.6	8.8
	8 h (s)	8760	1.5	3.3	4.1	5.3	5.7	7.2	7.3
60371103 (Los Angeles)	1 h (o)	8292	0.9	3.1	3.9	5.4	5.9	8.7	8.9
	1 h (s)	8760	1.0	3.0	3.8	5.4	5.8	8.7	8.9
	8 h (s)	8760	1.0	2.8	3.4	4.5	4.8	7.7	7.8
60371301 (Lynwood)	1 h (o)	8302	1.7	4.9	6.8	11.2	13.5	18.8	19.2
	1 h (s)	8760	1.7	4.9	6.7	11.2	13.5	18.8	19.2
	8 h (s)	8760	1.7	4.8	6.1	8.8	10.3	16.8	17.0
60371601 (Pico Rivera)	1 h (o)	7881	1.1	3.0	3.6	5.1	5.6	7.9	9.2
	1 h (s)	8760	1.2	3.0	3.6	5.0	5.6	7.9	9.2
	8 h (s)	8760	1.2	2.7	3.3	4.3	4.6	6.0	6.0

Site	Data set ^a	No. of obs.	Descriptive statistics for 1-hour CO concentration, ppm						
			50th	90th	95th	99th	99.5th	second largest	largest value
60372005 (Pasadena)	1 h (o)	8250	0.9	2.1	2.8	4.2	4.7	7.7	8.1
	1 h (s)	8760	0.9	2.1	2.8	4.2	4.7	7.7	8.1
	8 h (s)	8760	1.0	2.0	2.4	3.4	3.7	5.8	6.0
60374002 (Long Beach)	1 h (o)	8347	0.7	2.7	3.6	5.2	5.9	8.6	9.0
	1 h (s)	8760	0.7	2.7	3.6	5.2	5.9	8.6	9.0
	8 h (s)	8760	0.7	2.5	3.2	4.5	4.9	6.4	6.4
60375001 (Hawthorne)	1 h (o)	8125	0.5	3.7	5.1	7.4	8.3	12.3	12.4
	1 h (s)	8760	0.5	3.7	5.1	7.3	8.2	12.3	12.4
	8 h (s)	8760	0.7	3.4	4.5	6.1	6.9	10.1	10.3
60590001 (Anaheim)	1 h (o)	8354	0.8	2.3	2.9	4.6	5.5	8.2	8.4
	1 h (s)	8760	0.8	2.3	2.9	4.6	5.5	8.2	8.4
	8 h (s)	8760	0.9	2.1	2.6	3.5	3.8	5.7	5.7
60595001 (La Habra)	1 h (o)	8230	1.0	2.8	3.7	6.2	7.2	11.0	11.9
	1 h (s)	8760	1.0	2.8	3.6	6.1	7.1	11.0	11.9
	8 h (s)	8760	1.1	2.8	3.3	4.2	4.5	5.6	5.7

^a 1 h (o): original 1-hour data set as down-loaded from AIRS.
1 h (s): supplemented 1-hour data set (includes estimates of missing values)
8 h (s): supplemented 8-hour running average data set [based on 1 h (s) data].

running-average 8-hour concentrations after estimation of missing values.

Table 7-3 provides descriptive statistics by monitoring site for the 1-hour CO concentrations in each 1997 Los Angeles data set before and after estimation of the missing values. Again, the statistics indicate that the addition of missing-value estimates did not significantly affect the distribution of any data set. Each table also provides descriptive statistics for running-average 8-hour concentrations after estimation of missing values.

7.4 Adjustments to Simulate Regulatory Scenarios

In past applications of various exposure models to particular study areas, OAQPS has typically developed input air quality data sets for the study area that represent (1) baseline conditions and (2) conditions in which the area just attains a specific NAAQS. Baseline conditions are usually represented by unadjusted air quality data reported by fixed-site monitors in the area during a recent calendar year. Attainment conditions are simulated by applying an air quality adjustment procedure (AQAP) to the baseline data.

Since the first NEM-type model was implemented in 1981, OAQPS has used a variety of AQAP's to simulate attainment conditions. As many of the early methods have been superseded, this section will discuss only the AQAP's used in recent applications of pNEM/CO and pNEM/O₃ that are likely to be considered in future pNEM assessments. These methods can be conveniently identified as follows:

- Proportional rollback
- Proportional rollback with constant background concentration
- Quadratic rollback
- Adjustment of a fitted Weibull distribution

Each of these AQAP's incorporates the following four steps:

1. Specify an air quality indicator (AQI) to be used in evaluating the status of a monitoring site with respect to the NAAQS (or air quality scenario) of interest.
2. Determine the value of the AQI for each site within the study area under baseline conditions.
3. Determine the value of the AQI under conditions in which the air pollution levels within the study area have been reduced until the site with the highest pollution levels just attains a specified NAAQS.
4. Adjust the one-hour values of the baseline data set associated with each site to yield the AQI value determined in Step 3. The adjusted data set should retain the temporal profile of the baseline data set.

Subsection 7.4.1 discusses methods for specifying an appropriate AQI (Step 1) and the determination of baseline AQI values (Step 2). Subsections 7.4.2 through 7.4.5 present descriptions of the four adjustment methods listed above. Subsection 7.4.6

demonstrates a method for evaluating the performance of two or more candidate AQAP's.

The following discussion assumes that the air quality data selected to represent a study area have been obtained from fixed-site monitors located within the study area. When dispersion model estimates are used in place of fixed-site monitoring data, the analyst usually has the prerogative of simulating attainment conditions directly by varying the emissions rates used as input to the dispersion model. If this option is not available, the analyst may wish to consider the AQAP methods described here.

7.4.1 Specification of AQI

The AQI is a value derived from the data reported by a monitoring site that can be used to evaluate the status of the monitoring site with respect to the NAAQS (or air quality scenario) of interest. The AQI may be determined empirically or statistically. For example, analysts considered the following two AQI's for evaluating ozone monitoring data with respect to the one-hour ozone NAAQS.

Deterministic:	the observed second largest daily-maximum one hour ozone concentration
Statistical:	the characteristic largest daily-maximum one hour ozone concentration

The deterministic AQI has the advantage of being easily calculated by simply ranking the one-hour daily maximum values in the data reported by a monitor. The principal disadvantage of the deterministic AQI is that it is very sensitive to anomalous high values that may represent unusual meteorological conditions or data processing errors. The statistical AQI, which is determined by fitting a distribution to the data, is less affected by data anomalies.

As EPA chose to use the statistical AQI in the recent pNEM/O3 assessments described by Johnson, Capel, and McCoy (1996), it is instructive to consider how the statistic is calculated. The characteristic largest value (CLV) of a distribution is the value expected to be exceeded once in n observation. If F(x) is the cumulative distribution of x, then

$$F(x) = 1 - 1/n \quad (7-16)$$

when x is the CLV.

Selection of an appropriate cumulative distribution to fit data is important in determining a reasonable CLV. The Weibull and lognormal distributions have often been found to provide close fits to ambient air quality data. The Weibull distribution is defined as

$$F(x) = 1 - \exp[-(x/\delta)^k] \quad (7-17)$$

where δ is the scale parameter and k is the shape parameter. The lognormal distribution is defined as

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w \exp(-t^2/2) dt \quad (7-18)$$

where

$$w = [\ln(x) - \mu]/\sigma \quad (7-19)$$

and $\ln(x)$ is distributed normally with mean μ and variance σ^2 .

Consider the hourly-average ozone concentrations reported by a monitoring site for a specified ozone season. If the data set is complete (no missing values), it will take the form of a time series x_t ($t = 1, 2, 3, \dots, n$) that contains $n = (24)(N)$ values, where N is the number of days in the time period. From this time series a second time series of daily maximum 1-hour values can be constructed.

Assume that a Weibull distribution with parameters δ and k provides a good fit to the empirical distribution of hourly average values. If one disregards autocorrelation, the value expected to be exceeded once in $n = (24)(N)$ hours can be estimated as

$$CLVOH = \delta [\ln(24)(N)]^{1/k} \quad (7-20)$$

This is the characteristic largest one-hour value (CLVOH). If we again disregard autocorrelation, the daily maximum 1-hour value expected to be exceeded once in N days can be estimated as

$$CLVOHDM = \delta \left\{ -\ln \left[1 - \left(\frac{N-1}{N} \right)^{1/24} \right] \right\}^{1/k} \quad (7-21)$$

This is the characteristic largest daily maximum one-hour value. For 7-month and 12-month ozone seasons, N is equal to 214 and 365, respectively. For these values of N , CLOH and CLVOHDM are virtually indistinguishable in value over the range in k values typically found in ozone data ($0.6 < k < 2.5$). Consequently, the expression

$$CLVOHDM \doteq \delta [\ln(24)(N)]^{1/k} \quad (7-22)$$

can be used as an alternative to Equation 7-21 for calculating CLVOHDM.

7.4.2 Proportional Rollback -- No Background

This AQAP is based on the assumption that the air quality data reported by each monitoring site under attainment conditions will be proportional (relative to each time period) to the data reported under baseline conditions. In the case of one hour data, the adjustment equation is simply

$$CMON(m,h,s) = \rho(s) \times CMON(m,h,b) \quad (7-23)$$

in which $CMON(m,h,b)$ is the baseline 1-hour concentration for monitor m at hour h ,

CMON(m,h,s) is the adjusted 1-hour concentration for monitor m at hour h under air quality scenario s, and $\rho(s)$ is an adjustment factor specific to scenario s. The $\rho(s)$ term is calculated by the expression

$$\rho(s) = \text{CMAX}(s)/\text{CMAX}(b) \quad (7-24)$$

in which CMAX(s) is the largest concentration permitted under scenario s for a specified AQI and CMAX(b) is the value of this AQI based on the monitoring data selected to represent baseline conditions.

A letter report by Johnson (1997c) provides an evaluation of a “proportional” AQAP used in various applications of pNEM/O3 (Johnson, Mozier, and Capel, 1997; Johnson, 1997d). This AQAP applied a site-specific adjustment factor to all ozone concentrations reported by a fixed-site monitor. Consequently, the adjusted data exhibited the same degree of reduction at the high, middle, and low percentiles of the distribution.

7.4.3 Proportional Rollback with Constant Background Concentration

This AQAP applies the proportionality assumption to that portion of the concentration that lies above a specified background level. For one-hour data, the adjustment equation is now

$$\text{CMON}(m,h,s) = \text{BG} + \rho(s) \times \text{CDIF}(m,h,b) \quad (7-25)$$

in which CMON(m,h,b) is the baseline 1-hour concentration for monitor m at hour h, BG is the assumed background concentration, and $\rho(s)$ is the adjustment factor specific to scenario s. The CDIF(m,h,s) term is calculated as

$$\text{CDIF}(m,h,b) = \text{CMON}(m,h,b) - \text{BG} \quad (7-26)$$

in which CMON(m,h,b) is the 1-hour concentration associated with monitor m at hour h under existing conditions. The value of $\rho(s)$ is calculated by the expression

$$\rho(s) = [\text{CMAX}(s) - \text{BG}]/[\text{CMAX}(b) - \text{BG}] \quad (7-27)$$

when CDIF(m,h,b) > 0 and by the expression

$$\rho(s) = 1 \quad (7-28)$$

when CDIF(m,h,b) ≤ 0. As before, CMAX(s) is the largest concentration permitted under scenario s for a specified AQI and CMAX(b) is the value of this AQI based on the monitoring data selected to represent baseline conditions.

Section 3.3 of the report by Johnson, Mihlan, LaPointe, et al. (2000) describes the application of this AQAP to Los Angeles air quality data for use with Version 2.1 of pNEM/CO. In this application, analysts estimated the following values for the parameters of Equation 7-27: CMAX(e) = 15.0 ppm, CMAX(s) = 9.4 ppm, and BG = 0.2 ppm. The resulting estimate of $\rho(s)$ for CDIF(m,h,b) > 0 was 0.622. Results of the

adjustment procedure can be found in Subsection 7.5.

7.4.4 Quadratic Rollback

An article by Duff, Horst, and Johnson (1998) describes the theoretical basis of the quadratic AQAP and discusses its limitations. In its simplest form, the quadratic AQAP can be expressed by the relationship

$$\text{CMON}(m,h,s) = r(m,h) \times \text{CMON}(m,h,b) \quad (7-29)$$

in which

$$r(m,h) = V - B \times \text{CMON}(m,h,b). \quad (7-30)$$

As before, $\text{CMON}(m,h,b)$ is the baseline 1-hour concentration for monitor m at hour h and $\text{CMON}(m,h,s)$ is the adjusted 1-hour concentration for monitor m at hour h under air quality scenario s .

The degree of reduction applied to each $\text{CMON}(m,h,b)$ is determined by the magnitude of $\text{CMON}(m,h,b)$ such that large values are reduced by a larger degree than small values. V and B are positive constants; the method used to calculate these coefficients is relatively complicated and dependent on the specific application. To illustrate a typical application, this subsection describes how the quadratic AQAP was used to adjust 1-hour ozone data to simulate attainment conditions with respect to three ozone standard formulations. The formulations and associated AQI's are listed in Table 7-4.

Letter reports by Johnson (1997a, 1997b) describe how the quadratic AQAP was applied to ozone data obtained from fixed-site monitors located in each of the six cities listed in Table 7-5. Baseline conditions in each city were represented by ozone season data reported for a calendar year with relatively high ozone concentrations. Attainment conditions were defined by the ozone season data reported for a year with relatively low ozone levels. Table 7-5 lists the baseline and attainment "design values" developed for each city. To evaluate the performance of the quadratic AQAP, analysts applied the procedure to the baseline data for each monitor and compared the resulting adjusted data with the corresponding data actually reported by the monitor for the attainment year.

The AQAP procedure varied slightly according to the formulation of the standard. Tables 7-6, 7-7, and 7-8 present the implementation steps associated with the 1H1EX, 8H1EX, and 8H5EX formulations, respectively. Steps 1 through 3 in each table describe the method used to determine the baseline and attainment values of AQI for each monitor in a particular city. As indicated above, the baseline value for the AQI was the CLV1 for 1H1EX, the CLV8 for 8H1EX, and the EH6LDM for 8H5EX. (The CLV1 and CLV8 values were determined by fitting Weibull distributions to the data as discussed in Subsection 7.4.5).

The remaining steps were performed to adjust the one-hour values at each monitor so that the specified AQI of the resulting data set equaled the attainment (target) concentration determined for the monitor in Step 3. Steps 4 through 7 Table 7-4. Formulations of the National Ambient Air Quality Standard (NAAQS) for

ozone used in evaluating three procedures for adjusting air quality data to simulate attainment of an ozone NAAQS.

Formulation acronym	Attainment conditions	Air Quality Indicator (AQI) used to assess compliance
1H1EX	The expected number of daily maximum one-hour ozone concentrations exceeding a specific value shall not exceed one.	CLV1: the characteristic largest daily maximum one-hour ozone concentration ^a
8H1EX	The expected number of daily maximum eight-hour ozone concentrations exceeding a specific value shall not exceed one.	CLV8: the characteristic largest daily maximum eight-hour ozone concentration
8H5EX	The expected number of daily maximum eight-hour ozone concentrations exceeding a specific value shall not exceed five.	EH6LDM: the observed sixth largest daily maximum eight-hour ozone concentration

^aIn these analyses, the characteristic largest value (CLV) is defined as the value expected to be exceeded once in n observations, in which n is the number of days in the ozone season specified for a particular city. The CLV of each data set was estimated by fitting a Weibull distribution to the daily maximum values for the ozone season and then identifying the value located on the fitted distribution at fractile = 1 - 1/n.

produced a value for the parameter V for each monitor. In Step 8, the V value was used to determine which of two quadratic formulae would be used to adjust the one-hour ozone values associated with the monitor. Application of the selected formula yielded an adjusted data set representing attainment conditions. This data set was considered the final adjusted data set for the 1H1EX formulation. Two additional steps were used to “fine-tune” the data for the 8H1EX and 8H5EX formulations.

7.4.5 Weibull Rollback

To simulate attainment conditions, the Weibull AQAP adjusts each value in the baseline data set for a particular monitor by the equation

$$CMON(m,h,s) = a \times CMON(m,h,b)^b \quad (7-31)$$

in which $CMON(m,h,b)$ is the baseline 1-hour concentration for monitor m at hour h and $CMON(m,h,s)$ is the adjusted 1-hour concentration for monitor m at hour h under air quality scenario s. The a and b coefficients are functions of the parameters of a Weibull distribution fit to the baseline data for monitor m and the value of the characteristic largest value (CLV) predicted to occur under attainment conditions.

A letter report by Johnson (1997b) describes how the Weibull AQAP was applied

Table 7-5. Design Values for Baseline And "Attainment" Conditions.

Study area	Number of monitors ^a	Baseline conditions				"Attainment" conditions ^b			
		Calendar year	Design value, ppb			Calendar year	Design value, ppb		
			1H1EX	8H1EX	8H5EX		1H1EX	8H1EX	8H5EX
Chicago	11	1991	134	114	91	1993	96	84	76
Washington	10	1991	169	143	106	1992	118	101	90
Houston	10	1990	241	179	124	1994	172	119	111
Los Angeles	14	1991	321	215	180	1994	240	180	149
New York	10	1991	183	143	113	1994	155	116	93
Philadelphia	9	1991	162	138	115	1992	140	103	92

^aNumber used in evaluation of air quality adjustment procedure. Typically omits one or two monitors used in pNEM/O3 analyses because of incomplete data in attainment year.

^bThe so-called "attainment" conditions year represents a year with improved air quality and not necessarily attainment of the current or proposed air quality standards.

Table 7-6. Quadratic Air Quality Adjustment Procedure for 1H1EX Formulation.

-
1. Table 7-4 lists 1H1EX design values for the baseline and "attainment" years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.
 2. Determine the following adjustment factor:

$$\text{CAF} = [\text{CAQI}(\text{attainment})]/[\text{CAQI}(\text{baseline})].$$
 3. Let AQI(s, baseline) indicate the value of CLV1 for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$\text{TAQI}(s, \text{attainment}) = \text{CAF} \times \text{AQI}(s, \text{baseline})$$
 4. Operations on baseline data for one monitor: Let S =TAQI(s, attainment) from Step 3. Let J = maximum one-hour value. For each one-hour period ending at hour i, calculate

$$I_i = \text{mean of the one-hour ozone concentrations}$$

$$Q_i = \text{mean of the squares of the one-hour ozone concentrations}$$

$$X_i = (2)(J)(I_i) - Q_i$$

$$Z_i = (I_i - S)/Q_i.$$
 5. Define B as the second largest daily maximum value of Z_i determined in Step 4.
 6. Define X as the second largest daily maximum value of X_i. Let m equal the value of index i for the one-hour period associated with X. Save I_i and Q_i values for this time period as I_m and Q_m.
 7. Calculate the value

$$V = (2)(J)(S)/X.$$
 using S and J values from Step 4 and the X value from Step 6.
 8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s.

If $V \geq 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = x_t - (B)(x_t)^2.$$

If $V < 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = (V)(x_t) - \{[(V)(I_m) - S]/(Q_m)\}(x_t)^2.$$
 using the value of S from Step 4 and the values of I_m and Q_m from Step 6.
 9. The one-hour data set produced by Step 8 is considered the final one-hour data set representing attainment conditions at monitor s.
-

Table 7-7. Quadratic Air Quality Adjustment Procedure for 8H1EX Formulation.

-
1. Table 7-4 lists 8H1EX design values for the baseline and "attainment" years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.

2. Determine the following adjustment factor:

$$\text{CAF} = [\text{CAQI}(\text{attainment})]/[\text{CAQI}(\text{baseline})].$$

3. Let AQI(s, baseline) indicate the value of CLV8 for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$\text{TAQI}(s, \text{attainment}) = \text{CAF} \times \text{AQI}(s, \text{baseline})$$

4. Operations on baseline data for one monitor: Let S =TAQI(s, attainment) from Step 3. Let J = maximum one-hour value. For each eight-hour period ending at hour i, calculate

$$I_i = \text{mean of the one-hour ozone concentrations}$$

$$Q_i = \text{mean of the squares of the one-hour ozone concentrations}$$

$$X_i = (2)(J)(I_i) - Q_i$$

$$Z_i = (I_i - S)/Q_i.$$

5. Define B as the second largest daily maximum value of Z_i determined in Step 4.

6. Define X as the second largest daily maximum value of X_i . Let m equal the value of index i for the eight-hour period associated with X. Save I_i and Q_i values for this time period as I_m and Q_m .

7. Calculate the value

$$V = (2)(J)(S)/X.$$

using S and J values from Step 4 and the X value from Step 6.

8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s.

If $V \geq 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = x_t - (B)(x_t)^2.$$

If $V < 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = (V)(x_t) - \{[(V)(I_m) - S]/(Q_m)\}(x_t)^2.$$

using the value of S from Step 4 and the values of I_m and Q_m from Step 6.

9. Calculate the value of the second largest eight-hour daily maximum concentration using the one-hour values obtained in Step 8. Indicate this preliminary value as PAQI(s).

10. Multiply each y_t value by the ADJ(s) value determined below to obtain the final one-hour data set representing attainment conditions at monitor s.

$$\text{ADJ}(s) = \text{TAQI}(s)/\text{PAQI}(s).$$

Table 7-8. Quadratic Air Quality Adjustment Procedure for 8H5EX Formulation.

-
1. Table 7-4 lists 8H5EX design values for the baseline and "attainment" years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.

2. Determine the following adjustment factor:

$$\text{CAF} = [\text{CAQI}(\text{attainment})]/[\text{CAQI}(\text{baseline})].$$

3. Let AQI(s, baseline) indicate the value of EH6LDM for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$\text{TAQI}(s, \text{attainment}) = \text{CAF} \times \text{AQI}(s, \text{baseline})$$

4. Operations on baseline data for one monitor: Let S =TAQI(s, attainment) from Step 3. Let J = maximum one-hour value. For each eight-hour period ending at hour i, calculate

$$I_i = \text{mean of the one-hour ozone concentrations}$$

$$Q_i = \text{mean of the squares of the one-hour ozone concentrations}$$

$$X_i = (2)(J)(I_i) - Q_i$$

$$Z_i = (I_i - S)/Q_i.$$

5. Define B as the sixth largest daily maximum value of Z_i determined in Step 4.

6. Define X as the sixth largest daily maximum value of X_i . Let m equal the value of index i for the eight-hour period associated with X. Save I_i and Q_i values for this time period as I_m and Q_m .

7. Calculate the value

$$V = (2)(J)(S)/X.$$

using S and J values from Step 4 and the X value from Step 6.

8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s.

If $V \geq 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = x_t - (B)(x_t)^2.$$

If $V < 1$, then apply the following equation to each x_t value in the one-hour data set to produce the initial adjusted value at time t (y_t):

$$y_t = (V)(x_t) - \{[(V)(I_m) - S]/(Q_m)\}(x_t)^2.$$

using the value of S from Step 4 and the values of I_m and Q_m from Step 6.

9. Calculate the value of the specified sixth largest daily maximum concentration using the one-hour values obtained in Step 8. Indicate this preliminary value as PAQI(s).

10. Multiply each y_t value by the ADJ(s) value determined below to obtain the final one-hour data set representing attainment conditions at monitor s.

$$\text{ADJ}(s) = \text{TAQI}(s)/\text{PAQI}(s).$$

to the same ozone data sets used in the application of the quadratic AQAP described in Subsection 7.4.4. Consistent with the quadratic evaluation, analysts applied the Weibull procedure to the baseline data for each monitor and compared the resulting adjusted data with the corresponding data actually reported by the monitor for the attainment year. The methods used in this evaluation provide a good demonstration of the Weibull AQAP.

Table 7-5 lists the baseline and attainment “design values” developed for each of six cities. The applied AQAP procedure varied slightly according to the formulation of the standard. Tables 7-9, 7-10, and 7-11 present the implementation steps associated with the 1H1EX, 8H1EX, and 8H5EX formulations, respectively. Steps 1 through 3 in each table describe the method used to determine the baseline and attainment values of AQI for each monitor in a particular city. Consistent with the approach used for the quadratic evaluation, the baseline value for the AQI was the CLV1 for 1H1EX, the CLV8 for 8H1EX, and the EH6LDM for 8H5EX. (The CLV1 and CLV8 values were determined from earlier Weibull fits to the data.)

The remaining steps were performed to adjust the one-hour values at each monitor so that the specified AQI of the resulting data set equaled the attainment (target) concentration determined for the monitor in Step 3. In Step 4 of the procedure, the target concentration was multiplied by an “attainment” ratio value to obtain an estimate of the characteristic largest one-hour value expected under attainment conditions. The attainment ratio value was estimated using the corresponding ratio of characteristic largest one-hour value to AQI observed in the baseline one-hour data. Steps 5, 6, and 7 provide estimates of the shape and scale parameters (k and δ) which characterize the Weibull distribution expected under attainment conditions. In Step 8, the baseline and attainment Weibull parameters were inserted into an adjustment equation which was then applied to the one-hour baseline data. Application of the selected formula yielded an adjusted data set representing attainment conditions. This data set was considered the final adjusted data set for the 1H1EX formulation, and no further adjustments were made. For the 8H1EX and 8H5EX formulations, a small proportional adjustment factor was applied to the data to ensure that the AQI exactly equaled the target AQI value (Steps 9 and 10).

7.4.6 Evaluation of Alternative Air Quality Adjustment Procedures

In 1997, researchers conducted a comparative evaluation of three AQAP methods (proportional, quadratic, and Weibull) using the ozone data summarized in Table 7-5 (Johnson, 1997b; Duff, Horst, and Johnson, 1998). The methods used in this evaluation may be useful to other researchers in assessing the pros and cons of alternative AQAP's.

The evaluations of the proportional, quadratic, and Weibull AQAP's were performed using a common ten-step procedure:

1. Six cities were selected for evaluating the AQAP's: Chicago, Washington, Houston, Los Angeles, New York, and Philadelphia. These cities were selected because they showed a significant variation in ozone levels within a recent 10-year period (1985-94).

Table 7-9. Weibull Air Quality Adjustment Procedure for 1H1EX Formulation.

-
1. Table 7-4 lists 1H1EX design values for the baseline and “attainment” years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.
 2. Determine the following adjustment factor:

$$\text{CAF} = [\text{CAQI}(\text{attainment})]/[\text{CAQI}(\text{baseline})].$$
 3. Let AQI(s, baseline) indicate the value of CLV1 for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$\text{TAQI}(s, \text{attainment}) = \text{CAF} \times \text{AQI}(s, \text{baseline})$$
 4. Estimate attainment ACLV1(s) value corresponding to TAQI(s, NAAQS) using the following identity:

$$\text{ACLV1}(s, \text{NAAQS}) = \text{TAQI}(s, \text{NAAQS}).$$
 5. Fit a two-parameter distribution to the baseline one-hour data for monitor s to obtain values for the Weibull scale parameter δ and shape parameter k.
 6. Use the following equation to estimate the Weibull shape parameter for monitor i under attainment conditions (k').

$$1/k' = -0.2389 + (0.003367)[\text{ACLV1}(s, \text{NAAQS})] + (0.4726)(1/k)$$

where k is the Weibull shape parameter determined in Step 5.
 7. Use the following equation to estimate the Weibull scale parameter for monitor s under attainment conditions (δ').

$$\delta' = [\text{ACLV1}(s, \text{NAAQS})] / [\ln(n)]^{1/k'}$$

where n is the number of one-hour values in the exposure period (i.e., the ozone season specified for the city).
 8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s. Initially, the adjusted one-hour at time t will be

$$y_t = (\delta')(x_t/\delta)^{k/k'}$$

where δ , δ' , k, and k' have been determined in Steps 5, 6, and 7.
 9. The one-hour data set produced by Step 8 is considered the final one-hour data set representing attainment conditions at monitor s.
-

Table 7-10. Weibull Air Quality Adjustment Procedure for 8H1EX Formulation.

-
1. Table 7-4 lists 8H1EX design values for the baseline and "attainment" years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.

2. Determine the following adjustment factor:

$$CAF = [CAQI(attainment)]/[CAQI(baseline)].$$

3. Let AQI(s, baseline) indicate the value of CLV8 for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$TAQI(s, attainment) = CAF \times AQI(s, baseline)$$

4. Estimate attainment ACLV1(s) value corresponding to TAQI(s, NAAQS) using the following equation:

$$ACLV1(s, NAAQS) = TAQI(s, NAAQS) * \text{RATIO}(s, NAAQS)$$

where RATIO(s, NAAQS) is the ratio of CLV1(s) to CLV8(s) in the baseline data set.

5. Fit a two-parameter distribution to the baseline one-hour data for monitor s to obtain values for the Weibull scale parameter δ and shape parameter k.

6. Use the following equation to estimate the Weibull shape parameter for monitor i under attainment conditions (k').

$$1/k' = -0.2389 + (0.003367)[ACLV1(s, NAAQS)] + (0.4726)(1/k)$$

where k is the Weibull shape parameter determined in Step 5.

7. Use the following equation to estimate the Weibull scale parameter for monitor s under attainment conditions (δ').

$$\delta' = [ACLV1(s, NAAQS)] / [\ln(n)]^{1/k'}$$

where n is the number of one-hour values in the exposure period (i.e., the ozone season specified for the city).

8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s. Initially, the adjusted one-hour at time t will be

$$y_t = (\delta')(x_t/\delta)^{k/k'}$$

where δ , δ' , k, and k' have been determined in Steps 5, 6, and 7.

9. Calculate the value of the second largest eight-hour daily maximum concentration using the one-hour values obtained in Step 8. Indicate this preliminary value as PAQI(s).

10. Multiply each y_t value determined in Step 8 by the ADJ(s) value determined below to obtain the final one-hour data set representing attainment conditions at monitor s.

$$ADJ(s) = TAQI(i)/PAQI(s).$$

Table 7-11. Weibull Air Quality Adjustment Procedure for 8H5EX Formulation.

-
1. Table 7-4 lists 8H5EX design values for the baseline and "attainment" years selected for each city. Let CAQI(baseline) and CAQI(attainment) equal the design values for the baseline and attainment years, respectively.

2. Determine the following adjustment factor:

$$\text{CAF} = [\text{CAQI}(\text{attainment})]/[\text{CAQI}(\text{baseline})].$$

3. Let AQI(s, baseline) indicate the value of EH6LDM for monitor s for the year selected to represent baseline conditions. Determine a target value for the AQI of monitor s under attainment conditions by the equation

$$\text{TAQI}(s, \text{attainment}) = \text{CAF} \times \text{AQI}(s, \text{baseline})$$

4. Estimate attainment ACLV1(s) value corresponding to TAQI(s, NAAQS) using the following equation:

$$\text{ACLV1}(s, \text{NAAQS}) = \text{TAQI}(s, \text{NAAQS}) * \text{RATIO}(s, \text{NAAQS})$$

where RATIO(s, NAAQS) is the ratio of CLV1(s) to EH6LDM(s) in the baseline data set.

5. Fit a two-parameter distribution to the baseline one-hour data for monitor s to obtain values for the Weibull scale parameter δ and shape parameter k.

6. Use the following equation to estimate the Weibull shape parameter for monitor i under attainment conditions (k').

$$1/k' = -0.2389 + (0.003367)[\text{ACLV1}(s, \text{NAAQS})] + (0.4726)(1/k)$$

where k is the Weibull shape parameter determined in Step 5.

7. Use the following equation to estimate the Weibull scale parameter for monitor s under attainment conditions (δ').

$$\delta' = [\text{ACLV1}(s, \text{NAAQS})] / [\ln(n)]^{1/k'}$$

where n is the number of one-hour values in the exposure period (i.e., the ozone season specified for the city).

8. Let x_t indicate the one-hour ozone concentration at time t in the unadjusted data set for monitor s. Initially, the adjusted one-hour at time t will be

$$y_t = (\delta')(x_t/\delta)^{k/k'}$$

where δ , δ' , k, and k' have been determined in Steps 5, 6, and 7.

9. Calculate the value of the sixth largest eight-hour daily maximum concentration using the one-hour values obtained in Step 8. Indicate this preliminary value as PAQI(s).

10. Multiply each y_t value determined in Step 8 by the ADJ(s) value determined below to obtain the final one-hour data set representing attainment conditions at monitor s.

$$\text{ADJ}(s) = \text{TAQI}(i)/\text{PAQI}(s).$$

2. The baseline year used previously in the pNEM/O₃ analyses for each city (Table 7-5) was selected to represent the baseline ozone conditions of the city.
3. Analysts reviewed ozone data for the years 1985 to 1994 and selected a "low ozone" year for each city to represent "attainment" conditions. These years are also listed in Table 7-5.
4. Analysts filled in the missing values in each low-ozone data set using the same time series technique used in pNEM/O₃ assessments. The resulting data sets were labeled "observed attainment data" or OAD.
5. The OAD for all monitoring sites within a city were analyzed to identify the site with the highest value for each of the following AQI's: A = second-highest daily maximum one-hour ozone concentration, B = second-highest daily maximum eight-hour ozone concentration, and C = sixth-highest daily maximum eight-hour concentration. These sites were labeled Sites A, B, and C, respectively.
6. Site A was used to define "attainment" conditions with respect to a 1H1EX NAAQS formulation. In a similar manner, Sites B and C were used to define "attainment" conditions with respect to the 8H1EX and 8H5EX NAAQS formulations.
7. The one-hour version of the AQAP under evaluation was used to adjust the baseline data for each study area to meet the "attainment" conditions for the 1H1EX formulation specified in Step 6. Similarly, the eight-hour version of the AQAP was used to adjust the baseline data to meet the "attainment" conditions specified for the 8H1EX and 8H5EX formulations. For each NAAQS formulation, the adjusted data set for each monitor was considered to be an estimate of the data set that would be reported by the monitor under the specified "attainment" conditions. These data were labeled "estimated attainment data" or EAD.
8. At this stage, there were two data sets available for each combination of monitor and NAAQS formulation: the OAD and the EAD. Using the OAD and EAD, analysts calculated nine percentile statistics (50, 60, 70, 80, 90, 95, 99, 99.5, and 99.75) and three extreme-value statistics (sixth largest value, second largest value, and largest value). These 12 statistics were calculated for each of four averaging times: one-hour values, one-hour daily maximum values, eight-hour values, and eight-hour daily maximum values.
9. Analysts defined the term

$$\text{RATIO} = \text{EST}(X)/\text{OBS}(X) \quad (7-32)$$

where X is one of the 12 statistics listed in Step 8, EST(X) is the estimated

value of X as determined by the EAD, and OBS(X) is the observed value of X as determined by the OAD. Note that $RATIO = 1$ when the estimated value equals the observed value. Analysts calculated RATIO values for all combinations of statistics, NAAQS formulations, and averaging times.

10. The performance of each AQAP was evaluated by examining the distribution of RATIO values.

Briefly, analysts selected two years to represent baseline and "attainment" conditions in each of six cities. The baseline year was the same year used in the pNEM/O3 analyses to represent baseline or "as is" conditions. The attainment year was another recent year during which ozone levels were significantly lower than the baseline year. Analysts reviewed the attainment-year ozone data for all monitors in each city to determine the "design value" monitor with respect to each NAAQS formulation being evaluated. The appropriate AQAP was applied to the baseline data for the city to adjust the data to simulate attainment conditions as represented by the design value monitor.

The resulting monitor-specific adjusted data sets (referred to as "estimated attainment data" or EAD) were then compared with the corresponding observed data sets for the data sets (referred to as observed attainment data or OAD). The comparisons were performed by first calculating nine percentage statistics (50, 60, 70, 80, 90, 95, 99, 99.5, 99.75) and three extreme-value statistics (sixth largest value, second largest value, and largest value) for the EAD and OAD. Analysts then calculated the quotient

$$RATIO = EST(X)/OBS(X)$$

where X is one of the 12 statistics, EST(X) is the estimated value of X as determined by the EAD, and OBS(X) is the observed value of X as determined by the OAD. The distribution of RATIO was examined according to each combination of city, NAAQS formulation (e.g., 8H5EX), averaging time (e.g., 8-hour daily maximum), and statistic (e.g., 99th percentile). Values of RATIO near 1.0 were considered to indicate good performance of the AQAP with respect to the associated statistic.

To facilitate comparisons of RATIO values for the three AQAP's, Figures 7-3 through 7-5 present three graphs which plot average RATIO value (all sites - all cities, n = 64) against distribution statistic (50th percentile, 70th percentile, ..., second largest value, largest value). Each graph provides results of applying the proportional, quadratic, and Weibull AQAP's to one-hour ozone data for a single formulation (1H1EX, 8H1EX, or 8H5EX). Detailed tables of descriptive statistics for RATIO values associated with applications of the QR, W2, and PR procedures to the 1H1EX, 8H1EX, and 8H5EX formulations are presented in a technical letter by Johnson (1997b).

1H1EX. An AQAP is considered to perform well if it produces RATIO values near 1.0. The plots on the 1H1EX graph indicate that the quadratic yields RATIO values nearest to 1.0 throughout the distribution (from the 50th percentile to the largest value), except for the region around the 99th percentile where the Weibull AQAP produces RATIO values closer to 1.0. Below the 95th percentile, the Weibull AQAP produces RATIO values larger than those of the quadratic and proportional AQAP's.

top of figure 7-3

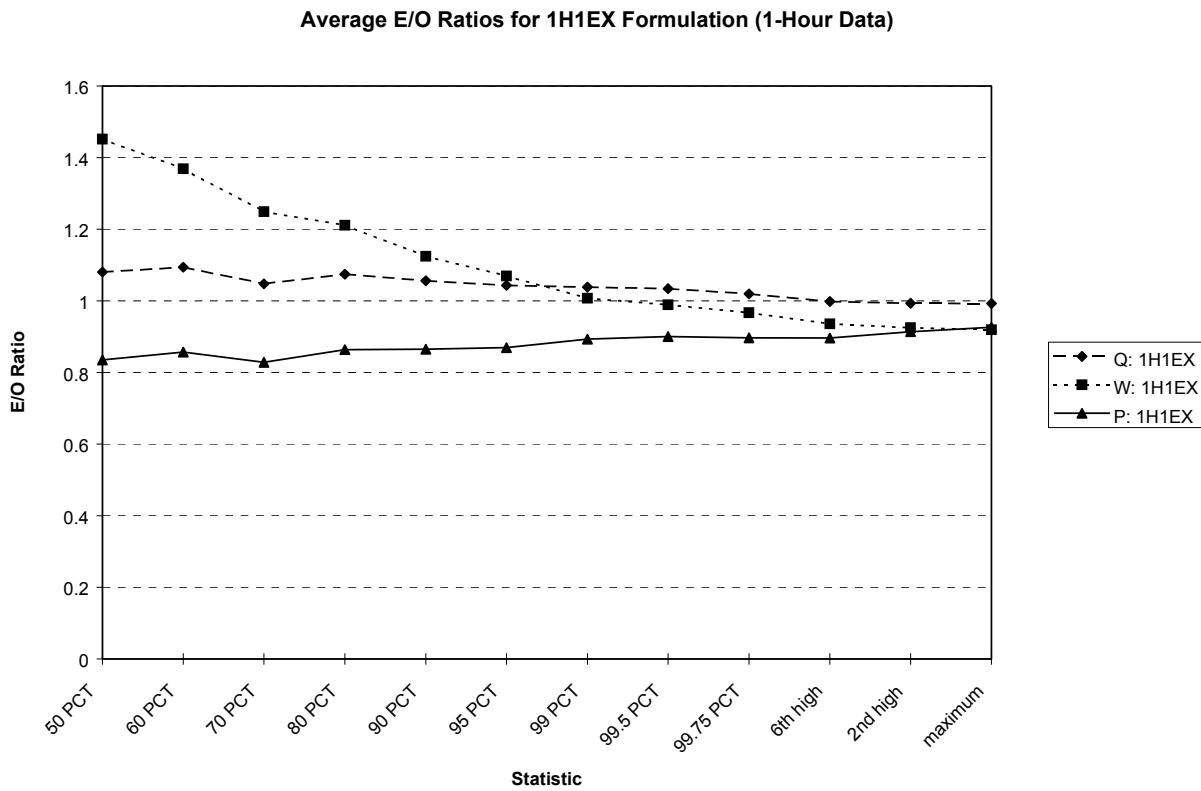


Figure 7-3. Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 1H1EX formulation (see Table 7-4).

top of figure 7-4

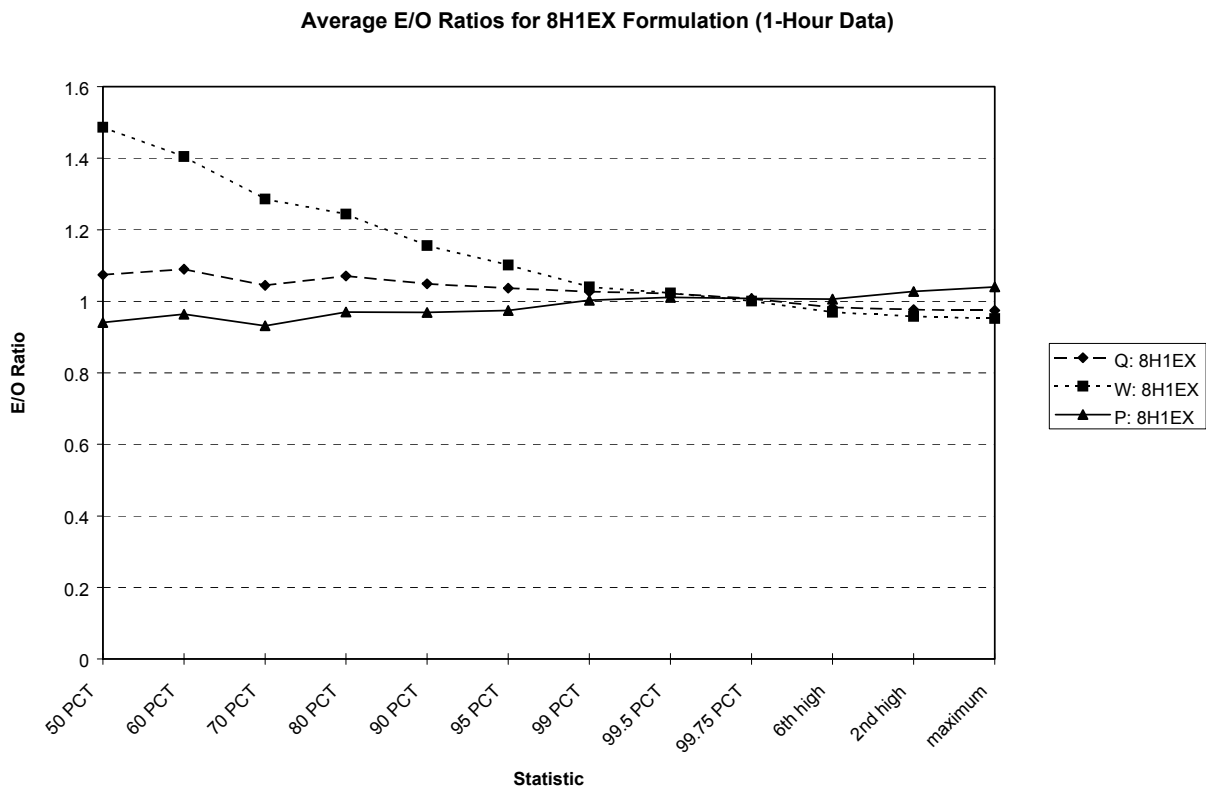


Figure 7-4. Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 8H1EX formulation (see Table 7-4).

top of figure 7-5

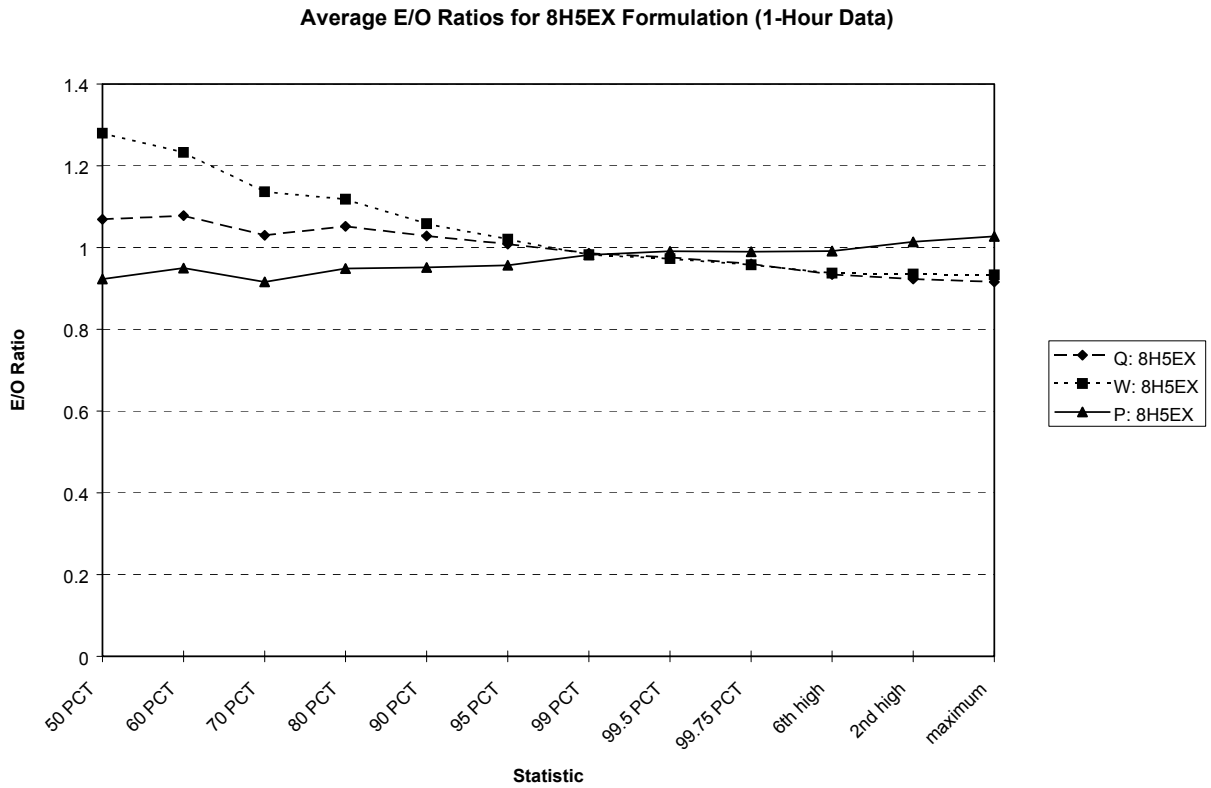


Figure 7-5. Plot of average RATIO value (all sites - all cities, n = 64) against selected distribution statistics for indicated air quality adjustment procedures (Q = quadratic, W = Weibull, and P = proportional) when applied to 8H5EX formulation (see Table 7-4).

The Weibull AQAP yields RATIO values farthest from 1.0 at the 50th percentile, where the RATIO value exceeds 1.4. The proportional AQAP produces RATIO values less than 1.0 throughout the distribution.

8H1EX. The quadratic and proportional AQAP's both produce RATIO values nearer to 1.0 than the Weibull AQAP on the 8H1EX graph. The quadratic produces RATIO values above 1.0 from the 50th percentile to the 99.75th percentile and RATIO values less than 1.0 at the higher percentiles. The proportional AQAP produces the reverse of this pattern. As in the 1H1EX graph, the Weibull yields RATIO values which differ significantly from 1.0 below the 95th percentile. Above the 99th percentile, the Weibull and quadratic produce similar RATIO values, tending to be less than 1.0.

8H5EX. The plots on the 8H5EX graph indicate that the quadratic AQAP produces RATIO values nearest to 1.0 from the 50th to the 99th percentile. Above the 99th percentile, the proportional AQAP yields values nearest to 1.0, with the quadratic and Weibull producing almost identical RATIO values that tend to be less than 1.0.

Summary. Based on this limited analysis, the quadratic AQAP appears to produce RATIO values that tend closest to 1.0 overall and the Weibull AQAP tends to produce RATIO values farthest from 1.0 overall. However, all three AQAP's perform adequately above the 99th percentile.

7.5 Estimation of Background Concentrations and Sample Results for Proportional Rollback

Many of the AQAP's discussed earlier in this section require an estimate of background concentration (BG). In an exposure assessment, BG is the portion of the ambient concentration that is not expected to change when a particular regulatory scenario (e.g., attainment of a specific NAAQS) is implemented. Typically, BG is considered to represent the contribution of upwind and natural sources that are unlikely to be significantly affected by the regulatory scenario, as indicated by the following definition of background PM taken from the current draft of the PM staff paper (USEPA, 2001).

For the purposes of this document, background PM is defined as the distribution of PM concentrations that would be observed in the U.S. in the absence of anthropogenic, or man-made emissions of primary PM and precursor emissions of VOC, NO_x, SO₂, and NH₃ in North America. Thus background includes PM from natural sources and transport of PM from outside of North America.

To simplify calculations, OAQPS has historically assumed that BG is constant during the exposure period, although this assumption is obviously not completely valid for pollutants like CO and ozone that exhibit diurnal and seasonal patterns. As BG is usually a relatively small quantity, the omission of diurnal and/or seasonal patterns does not tend to significantly affect the results of the analysis.

In recent pNEM assessments, researchers have developed estimates of BG by examining monitoring data associated with the study area and time period of interest together with data reported by remote sites considered to be relatively unaffected by anthropogenic emissions. Section 3.3 of the report by Johnson, Mihlan, LaPointe, et al. (2000) provides an example of the process as implemented in Version 2.1 of pNEM/CO.

Briefly, the proportional rollback AQAP described above (Equations 7-25 through 7-28 in Subsection 7.4.3) was applied to ambient CO data reported by monitoring sites in Los Angeles. At the time of the analysis, EPA was evaluating a CO standard expressed in terms of the second highest eight-hour non-overlapping average. Although the adjustment procedure is applied to one-hour data, the values of CMAX(s) and CMAX(e) required by Equation 7-27 are determined according to the relevant averaging time of the CO standard under evaluation. Consequently, the values for CMAX(s) and CMAX(e) inserted in Equation 7-27 were based on this averaging time.

Analysts selected 1997 monitoring data to represent existing conditions in Los Angeles. The AQAP was applied to these data sets with the goal of simulating attainment of the then current eight-hour NAAQS for CO, which stated that the second highest non-overlapping eight-hour average shall not exceed 9 ppm. Thus, the AQI of interest was the largest “second highest non-overlapping eight-hour average” reported by the monitoring sites of Los Angeles in the baseline year (1997). In implementing the eight-hour NAAQS, EPA used a rounding convention which specified that AQI values above 9.4 ppm were to be treated as nonattainment. Consequently, analysts specified that the largest “second highest non-overlapping eight-hour average” of the adjusted monitoring data should equal 9.4 ppm [i.e., CMAX(s) = 9.4 ppm in Equation 7-27].

To complete the adjustment procedure, analysts required values for CMAX(e) and BG specific to Los Angeles. Table 7-12 lists the value of second highest non-overlapping eight-hour average associated with each of the monitors previously selected to represent CO conditions in Los Angeles. The Los Angeles values range from 3.9 to 15.0 ppm. Based on the maximum value in this range, analysts specified that CMAX(e) would equal 15.0 ppm for Los Angeles.

In an earlier (1992) exposure analyses using Version 1.0 of pNEM/CO (Johnson, Capel, Paul, and Wijnberg, 1992), BG was defined as the smallest annual average CO concentration reported by a monitoring site within the defined study area. Analysts evaluated the results of using this same approach for the later pNEM/CO analysis. In implementing the approach, they considered only those monitors previously selected to represent CO conditions in Los Angeles. Table 7-12 lists the annual average for each monitor. The annual averages range from 0.84 to 2.33 for Los Angeles. Referring to the smallest value in the range, the value of BG would be 0.84 ppm for Los Angeles.

At the time of the analysis, the current criteria document (CD) for CO (EPA, 2000) supported a lower estimate for BG. In a discussion of CO measured at remote sites, the CD provided the following information concerning the relationship between global background CO and latitude.

Carbon monoxide concentrations range from a minimum of about 30 ppb during the summer in the Southern Hemisphere to about 200 ppb at high latitudes in the Northern Hemisphere during winter. Thus, CO concentration in remote areas of the Northern Hemisphere are only a small fraction (~ 1 to 2%) of those of concern to human health (as given by the National Ambient Air Quality Standards [NAAQS] for CO of 9 ppm for the second highest, nonoverlapping 8-h average concentration). [page 3-3]

For Version 2.1 of pNEM/CO, EPA determined that a reasonable estimate of BG for Los Angeles would be approximately 200 ppb (or 0.20 ppm). This value was based on

Table 7-12. Selected Descriptive Statistics for Los Angeles Monitoring Sites Based on 1997 Carbon Monoxide Data.

Monitoring site		Carbon monoxide concentration, ppm	
AIRS ID	Name or location	Second highest non-overlapping eight-hour average	Annual average
6-037-0113	West LA	3.9	0.84
6-037-1002	Burbank	7.2	1.76
6-037-1103	Los Angeles	5.8	1.36
6-037-1301	Lynwood	15.0	2.33
6-037-1601	Pico Rivera	5.9	1.49
6-037-2005	Pasadena	5.4	1.10
6-037-4002	Long Beach	6.2	1.11
6-037-5001	Hawthorne	7.9	1.28
6-059-0001	Anaheim	5.4	1.11
6-0595001	La Habra	5.3	1.36
Max/min values		Maximum = 15.0	Minimum = 0.84

Table 7-13. Parameter Values Specified for Application of Proportional Rollback Procedure to Carbon Monoxide Data Reported by Fixed-Site Monitors in Los Angeles.

Parameter	Los Angeles value
C _{MAX} (e), ppm	15.0
C _{MAX} (s), ppm	9.4
B _G , ppm	0.2
ρ(s), dimensionless	0.622

the winter maximum for remote sites observed at northern latitudes cited in the revised CD.

Table 7-13 lists the parameters used in Equation 7-27 when adjusting Los Angeles monitoring data. Table 7-14 lists descriptive statistics for the Los Angeles 1-hour data sets before and after application of the AQAP. The table also lists descriptive statistics for 8-hour running-average concentrations based on the adjusted data sets.

The Los Angeles results exhibit a significant difference between the unadjusted and adjusted one-hour data sets, with high values showing a slightly greater proportional reduction than low values. The adjustment procedure reduced the maximum one-hour values listed in Table 7-14 by approximately 37 percent. Depending on the site, the 50th percentile of the one-hour values was reduced by 30 to 35 percent. After adjustment, the second highest non-overlapping eight-hour average at the controlling Los Angeles site (Lynwood) was exactly equal to 9.4 ppm. [Note that this is the second highest non-overlapping eight-hour average at Linwood; the second highest eight-hour average based on all Linwood values (including overlapping averages) is 10.5 ppm, as indicated by Table 7-14.]

7.6 Simulating Geographic Variability

Each of the exposure models discussed in this report requires that the user provide estimates of ambient (outdoor) pollutant concentrations for a set of geographic units that together comprise the study area. For convenience, we will refer to these geographic units as “exposure districts” throughout this subsection. There are a variety of methods for determining ambient pollutant concentrations for a particular exposure district for a specified time period, the majority of which fall under one of the following general categories:

Direct use of monitoring data: data obtained from one or more fixed-site monitors located within the exposure district;

Interpolation of monitoring data: application of spatial interpolation techniques to data obtained from fixed-site monitors in neighboring districts;

Dispersion models: the dispersion model processes data on source characteristics, emission patterns, and meteorological conditions to determine ambient concentrations at specified receptor locations within each exposure district.

Ideally the method selected will adequately represent the spatial variability of ambient concentration throughout the study area. Regardless of method selected, small districts are generally to be preferred over large districts, as the former can better capture the spatial gradients of pollutant concentration. However, the lower limit on district size is likely to be constrained by a number of factors that are outside the control of the modeler, including the historical placement of fixed-site monitors and the spatial resolution of the input data available for dispersion modeling. In developing the various OAQPS exposure models discussed in this report, researchers were forced to make a number of compromises in the definition of exposure districts to address such

Table 7-14. Descriptive Statistics for Carbon Monoxide Concentrations Reported by 10 Los Angeles Monitors Before and After Adjustment to Simulate Attainment of Eight-Hour CO NAAQS.

Site	Data set ^a	Descriptive statistics for CO concentration, ppm						
		50th	90th	95th	99th	99.5th	second maximum	maximum
60370113 (West Los Angeles)	1 h (u)	0.6	2.0	2.6	3.6	4.1	6.4	7.3
	1 h (a)	0.4	1.3	1.7	2.3	2.6	4.1	4.6
	8 h (a)	0.5	1.2	1.4	2.0	2.1	2.6	2.7
60371002 (Burbank)	1 h (u)	1.4	3.5	4.4	6.0	6.6	8.6	8.8
	1 h (a)	0.9	2.3	2.8	3.8	4.2	5.4	5.5
	8 h (a)	1.0	2.1	2.6	3.4	3.6	4.5	4.6
60371103 (Los Angeles)	1 h (u)	1.0	3.0	3.8	5.4	5.8	8.7	8.9
	1 h (a)	0.7	1.9	2.4	3.4	3.7	5.5	5.6
	8 h (a)	0.7	1.8	2.2	2.8	3.1	4.8	4.9
60371301 (Lynwood)	1 h (u)	1.7	4.9	6.7	11.2	13.5	18.8	19.2
	1 h (a)	1.1	3.1	4.2	7.0	8.5	11.8	12.0
	8 h (a)	1.1	3.0	3.9	5.5	6.5	10.5	10.6
60371601 (Pico Rivera)	1 h (u)	1.2	3.0	3.6	5.0	5.6	7.9	9.2
	1 h (a)	0.8	1.9	2.3	3.2	3.6	5.0	5.8
	8 h (a)	0.8	1.8	2.1	2.7	2.9	3.8	3.8

Site	Data set ^a	Descriptive statistics for CO concentration, ppm						
		50th	90th	95th	99th	99.5th	second maximum	maximum
60372005 (Pasadena)	1 h (u)	0.9	2.1	2.8	4.2	4.7	7.7	8.1
	1 h (a)	0.6	1.4	1.8	2.7	3.0	4.9	5.1
	8 h (a)	0.7	1.3	1.6	2.2	2.4	3.7	3.8
60374002 (Long Beach)	1 h (u)	0.7	2.7	3.6	5.2	5.9	8.6	9.0
	1 h (a)	0.5	1.8	2.3	3.3	3.7	5.4	5.7
	8 h (a)	0.5	1.6	2.1	2.9	3.1	4.0	4.1
60375001 (Hawthorne)	1 h (u)	0.5	3.7	5.1	7.3	8.2	12.3	12.4
	1 h (a)	0.4	2.4	3.2	4.6	5.2	7.7	7.8
	8 h (a)	0.5	2.2	2.9	3.9	4.3	6.3	6.5
60590001 (Anaheim)	1 h (u)	0.8	2.3	2.9	4.6	5.5	8.2	8.4
	1 h (a)	0.6	1.5	1.9	2.9	3.5	5.2	5.3
	8 h (a)	0.6	1.4	1.7	2.3	2.4	3.6	3.6
60595001 (La Habra)	1 h (u)	1.0	2.8	3.6	6.1	7.1	11.0	11.9
	1 h (a)	0.7	1.8	2.3	3.9	4.5	6.9	7.5
	8 h (a)	0.8	1.8	2.1	2.7	2.9	3.6	3.6

^a 1 h (u): unadjusted 1-hour data (with missing values filled in).

1 h (a): adjusted 1-hour data set.

8 h (a): adjusted 8-hour running average data set [based on 1 h (a) data].

constraints. This subsection provides two examples of methods used by researchers to improve the representation of spatial variability in exposure assessments with limited monitoring data.

7.6.1 Adding Random Components to Monitoring Data

Section 2 describes the application of Version 2.1 of pNEM/CO to Denver and Los Angeles. As discussed in Subsection 2.1, analysts defined six exposure districts for Denver and 10 exposure districts for Los Angeles based on the availability of fixed-site monitoring data. The monitoring data were used to develop a sequence of one-hour CO concentrations spanning the exposure period for each of exposure district. Researchers could have used each sequence “as is” to represent the ambient (outdoor) CO concentrations throughout the applicable exposure district. However, this approach would have significantly under-represented the degree of spatial variability known to occur in ambient CO concentrations. Consequently, the researchers chose an alternative approach in which they created sequences of outdoor CO concentrations by adding random components of spatial variability to the monitoring data for each exposure district. These random components were based on statistical analyses of two supplemental data sets that contained outdoor CO data collected at hundreds of outdoor locations in Denver and in Southern California. The remainder of this subsection provides a detailed description of the development and implementation of this approach.

Subsection 2.4.1 briefly describes the method used to estimate the CO exposures in Version 2.1 of pNEM/CO. Note that the method requires estimates of outdoor CO concentration specific to cohort, microenvironment, district, and hour. These estimates are provided by the equation

$$CO_{out}(c,m,d,h) = M(m) \times L(c, m, d) \times T(c,m,d,h) \times [CO_{mon}(d,h)]^A \quad (7-33)$$

in which

$CO_{out}(c,m,d,h) =$	outdoor CO concentration for cohort c with respect to microenvironment m in district d during hour h,
$M(m) =$	multiplier (> 0) specific to microenvironment m,
$L(c,m,d) =$	location multiplier (> 0) specific to cohort c, microenvironment m, and district d (held constant for all hours),
$T(c,m,d,h) =$	time-of-day multiplier (> 0) specific to cohort c, microenvironment m, district d, and hour h,
$CO_{mon}(d,h) =$	monitor-derived CO concentration for hour h in district d, and
$A =$	exponent ($A > 0$).

This equation was used to generate a year-long sequence of outdoor one-hour CO

concentrations for each combination of cohort (c), microenvironment (m), and district (d). The exponent A was set equal to 0.621 and held constant for all sequences. The value of M(m) varied only with microenvironment as indicated by Table 2-6 in Section 2.

A value of the location factor L(c, m, d) was specified for each individual sequence and held constant for all hours in the sequence. The value was randomly selected from a lognormal distribution with geometric mean (GM_L) equal to 1.0 and geometric standard deviation (GSD_L) equal to 1.5232. The natural logarithms of this distribution can be characterized by a normal distribution with an arithmetic mean (μ_L) equal to zero and an arithmetic standard deviation (σ_L) equal to 0.4208.

A value of the time-of-day factor T(c, m, d, h) was randomly selected for each hour within a sequence from a lognormal distribution with geometric mean (GM_T) equal to 1.0 and geometric standard deviation (GSD_T) equal to 1.6289. The natural logarithms of this distribution follow a normal distribution with an arithmetic mean (μ_T) equal to zero and an arithmetic standard deviation (σ_T) equal to 0.4879.

The $CO_{out}(c, m, d, h)$ term is interpreted as the outdoor CO concentration in the immediate vicinity of microenvironment m in district d during hour h. $CO_{mon}(d, h)$ is the CO concentration reported for hour h by a nearby fixed-site monitor selected to represent district d.

Equation 7-33 is based on the results of data analyses that suggest that the relationship between $CO_{out}(c, m, d, h)$ and $CO_{mon}(d, h)$ should account for the identity of the microenvironment, the geographic location of the microenvironment, and the time of day. Numerous statistical models could be developed. In specifying the Equation 7-33 model, analysts attempted to balance the need for simplicity and parsimony with the need to model the most important patterns in the available data. Most of the model development was based on a comparison of hourly averages of 10-minute concentrations measured outside residences in southern California (Wilson, Colome, and Tian, 1995) with hourly averages measured at the nearest fixed site monitor. In this analysis, m represented the residence microenvironment in district d. The district d was initially taken to be the entire study region (i.e., San Diego and Los Angeles areas).

Analysts began model development by considering a simple linear regression model of the form

$$CO_{out}(c,m,d,h) = a(m,d) + A \times [CO_{mon}(d,h)] + e(c,m,d,h), \quad (7-34)$$

in which the residual term $e(c,m,d,h)$ is assumed to be independent and normally distributed with mean zero. For simplicity and parsimony, the slope coefficient A was assumed to be the same for all microenvironments (m) and districts (d). Although the coefficient of determination (R^2) for this model was reasonably high (0.53), the model was found to be unacceptable because it did not properly reflect the strong correlations that were observed between concentrations measured outside the same location. Instead, this regression model assumes that the residuals associated with a particular residential location are independent. In other words, this model does not properly separate out the variation between locations from the variation within locations. Note that the R^2 goodness-of-fit statistic is not an appropriate measure of model adequacy when the true, underlying errors are highly correlated.

Analysts identified two other deficiencies in this model: (1) large negative values of the randomly-selected $e(c,m,d,h)$ term could produce impossible negative outdoor

concentrations and (2) the model did not generate outdoor concentrations characterized by lognormal distributions. Various researchers (e.g., Ott, 1995) have demonstrated that ambient CO concentrations tend to be characterized by lognormal distributions rather than normal distributions.

To better address these latter concerns, analysts evaluated an alternative model in which the natural logarithm of outdoor concentration was expressed as a linear function of the natural logarithm of monitor concentration:

$$\text{LN}[\text{CO}_{\text{out}}(c,m,d,h)] = a(m,d) + A \times \text{LN}[\text{CO}_{\text{mon}}(d,h)] + e(c,m,d,h), \quad (7-35)$$

In this equation and those that follow, LN[] indicates the natural logarithm of the quantity in brackets. To properly separate the variability between and within locations, the intercept term $a(m,d)$ was also permitted to vary with the cohort location, c , leading to the final model:

$$\text{LN}[\text{CO}_{\text{out}}(c,m,d,h)] = a(c,m,d) + A \times \text{LN}[\text{CO}_{\text{mon}}(d,h)] + e(c,m,d,h). \quad (7-35)$$

Exponentiating both sides of this equation yields the equivalent formulation:

$$\text{CO}_{\text{out}}(c,m,d,h) = M(m) \times L(c,m,d) \times T(c,m,d,h) \times [\text{CO}_{\text{mon}}(d,h)]^A, \quad (7-36)$$

in which

$$\text{LN}[M(m)] = \text{mean } [a(c,m,d)], \text{ averaged over cohorts,}$$

$$L(c,m,d) = \exp\{a(c,m,d) - \text{mean } [a(c,m,d)]\}, \text{ and}$$

$$T(c,m,d,h) = \exp[e(c,m,d,h)].$$

This equation is identical to the model formulation presented above in Equation 7-33.

Several alternative statistical models were considered during the development of the selected model formulation. Early in the process, analysts evaluated a series of autoregressive time series models, in which model predictions were influenced by the past history of CO concentrations at the monitor and outdoors of the microenvironment. These models were rejected for several reasons: (1) they were inherently complex, (2) they yielded a wide variation in model coefficients which did not always produce reasonable estimates when applied to specific California residences, and (3) they required microenvironment-specific time series data for coefficient estimation which were not readily available for non-residential microenvironments. Analysts also evaluated models similar to Equation 7-33 in which the exponent A varied with microenvironment. These models were rejected due to the need for parsimony and the lack of sufficient, suitable data for estimating microenvironment-specific values of A . A simpler model in which the exponent A is fixed at 1 was rejected because fits of Equation 7-33 to California data indicated that A differed significantly from 1, at statistical significance levels well below 1%. In addition, the assumption that A equals 1 produced unrealistically high predictions for outdoor CO concentrations when the model was applied to monitoring data obtained from the Denver Broadway site. These high

values were found to be a direct result of setting A equal to 1, which forced the geometric standard deviation of the estimated outdoor concentrations to significantly exceed the geometric standard deviation of the monitor values.

Analysts ultimately arrived at Equation 7-33 (equivalent to Equation 7-35), which permits the A exponent to differ from 1.0. This model, considered a reasonable compromise between model simplicity and performance, is completely specified by four parameters [M(m), σ_L , σ_T , and A]. Note that σ_L , σ_T , and A are defined to be independent of the microenvironment, whereas M(m) is microenvironment-specific. Researchers were unable to find a single data source capable of providing estimates of all four parameters. Consequently, values for σ_L , σ_T , and A were estimated by analyzing data obtained from the California study conducted by Wilson, Colome, and Tian (1995), whereas the M(m) values were based on data provided by the Denver Personal Monitoring Study (Akland et al, 1985; Johnson, 1984).

During the residential monitoring study described by Wilson, Colome, and Tian (1995), researchers measured 10-minute CO concentrations outside 293 residences throughout California in 1992. These residences were customers of Pacific Gas and Electricity (129 residences in Northern California), San Diego Gas and Electric Company (89 residences in the San Diego area), and Southern California Gas Corporation (75 residences in the Los Angeles area). After excluding the PG&E data (not part of the Los Angeles study area) and homes for which valid CO data were not available, analysts used the remaining subset of 156 residences, 70 from Los Angeles and 86 from San Diego, as the basis for estimating values of σ_L , σ_T , and A applicable to the Los Angeles study area. (These coefficient values were also applied to Denver, as researchers were unable to identify a usable data set specific to the Denver study area). The data subset contained 44,726 valid 10-minute averages measured outside of residences, of which less than 1% were negative (smallest value = -1.0 ppm), 14,817 (33 %) were equal to 0 ppm, and the remainder were positive (maximum = 68.7 ppm). The valid 10-minute values were averaged by clock hour to permit comparison with hourly-average CO concentrations reported by nearby fixed-site monitors.

Analysts determined that the negative values in the data set were most likely caused by the subtraction of an offset from all measured values to account for monitor drift. To adjust for this offset and to prevent the occurrence of negative and zero values (which could not be used in fitting Equation 7-33), analysts added a constant offset of 0.5 ppm to each hourly-average value measured outside a residence. In addition, seventeen (0.2%) of the original hourly averages less than or equal to -0.5 ppm were discarded. Each of the resulting one-hour outdoor CO concentrations was paired with the one-hour CO concentration measured simultaneously at the nearest fixed-site monitor [based on data obtained from EPA's Aerometric Information Retrieval System (AIRS)]. This approach yielded a final database containing 6,330 pairs of hourly average concentrations, in which each pair was indexed by date, time, residence identifier, fixed-site monitor identifier, and fixed-site monitor scale (e.g., neighborhood).

The proposed model (Equation 7-33) was fitted to the final database using statistical software for a mixed (random and fixed effects) model which employed restricted maximum likelihood estimation. The fit yielded estimates of $\sigma_L = 0.4208$, $\sigma_T = 0.4879$, and $A = 0.621$, the values subsequently used in the pNEM/CO runs described in this report. The fitted value of M(m), representing residences in Los Angeles during 1992, was 0.9706. An alternative value (1.034), based on the analyses described

below, was applied to the indoor - residence microenvironment in the pNEM/CO runs (Table 2-6).

Researchers conducted a series of sensitivity analyses to evaluate the potential effects on parameter estimates of variations in (1) region and (2) scale of the fixed-site monitor. Equation 7-33 was fitted to a series of data subsets defined by region (Los Angeles or San Diego) or by the scale of the fixed monitor (based on the estimated maximum distance from the monitor represented by the measured concentrations: micro, middle, neighborhood, or urban scale). The fitted values of σ_L , σ_T , A, and M(m) were very similar across the different subsets, supporting the assumption that these parameters can be assumed to be representative of concentration patterns outside residences in other regions and for other time periods, and can be chosen to be the same value for all monitoring scales. Due to a lack of suitable data, the values of σ_L , σ_T , and A are also assumed to be applicable to concentrations outside all other microenvironments, although M(m) varies with the microenvironment.

The M(m) values were based on data provided by the Denver Personal Monitoring Study (Akland et al, 1985; Johnson, 1984). During this study, each of approximately 450 subjects carried a personal exposure monitor (PEM) for two 24-hour periods. Each PEM measured CO concentration continuously. The PEM readings were averaged by exposure event such that each event was associated with a single microenvironment and a single clock hour (e.g., 1 pm to 2 pm). Event durations ranged from one minute to one hour. The microenvironment assigned to each PEM reading was determined from entries made in a real-time diary carried by the subject.

In Equation 7-33, the $CO_{out}(c, m, d, h)$ term represents the outdoor CO concentration associated with a particular microenvironment m, even when the microenvironment is an indoor location. Few of the outdoor PEM values reported by the Denver study could be reliably associated with particular indoor microenvironments. Consequently, researchers employed a simplified procedure for estimating M(m) values which assumed that the mean of the indoor PEM values associated with each indoor microenvironment was approximately equal to the mean of the outdoor concentration for the microenvironment. This assumption is consistent with the results of applying mass-balance modeling to non-reactive pollutants in enclosed spaces where the only source of the pollutant is the outside air. In such cases, the mean indoor concentration approximates the mean outdoor concentration, with the instantaneous indoor concentration exhibiting a lower degree of variability than the corresponding outdoor concentration.

Because the simplified approach was also less sensitive to the wide variation in averaging times exhibited by the PEM values (i.e., one minute to 60 minutes), analysts were able to use the majority of PEM values in the statistical analysis. Limiting the analysis to one-hour PEM values would have significantly reduced the pool of usable data.

Researchers created a data base in which each PEM value was matched to the corresponding hourly-average CO concentration reported by the nearest fixed-site monitor. The data were first processed by excluding cases with missing measurements, cases in which measurements failed a quality control check, and cases in which applicable diary data indicated the potential presence of smokers or gas stoves. Each PEM CO concentration was then assigned to a microenvironment, m, based on entries in the activity diary. In some cases, as indicated in the footnote to Table 2-6, the data

for two or more similar microenvironments were aggregated to provide more stable estimates than those based on the very limited amount of data available for specific microenvironments. For consistency with the analyses performed on the Wilson, Colome, and Tian (1995) database, all cases with a zero measurement from the personal exposure monitor were excluded, as were all cases in which the fixed site monitor concentration was zero after rounding to the nearest integer ppm. Note that the Denver fixed-site data were recorded to the nearest 0.1 ppm, whereas the Los Angeles fixed site data were only recorded to the nearest integer.

When Equation 7-33 is expressed in logarithmic form (i.e., as Equation 7-35) and averaged over cohorts, one obtains the equation

$$\begin{aligned} & \text{Mean}\{\text{LN}[\text{CO}_{\text{out}}(\text{c}, \text{m}, \text{d}, \text{h})]\} \\ & = \text{Mean}[a(\text{c}, \text{m}, \text{d})] + A \times \text{Mean}\{\text{LN}[\text{CO}_{\text{mon}}(\text{d}, \text{h})]\} + \text{Mean}[e(\text{c}, \text{m}, \text{d}, \text{h})] \\ & = \text{LN}[M(\text{m})] + A \times \text{Mean}\{\text{LN}[\text{CO}_{\text{mon}}(\text{d}, \text{h})]\}. \end{aligned}$$

Therefore, the value of $M(\text{m})$ equals

$$M(\text{m}) = \exp\{\text{Mean LN}[\text{CO}_{\text{out}}(\text{c}, \text{m}, \text{d}, \text{h})] - A \times \text{Mean LN}[\text{CO}_{\text{mon}}(\text{d}, \text{h})]\}, \quad (7-37)$$

where $A = 0.621$ (as above). This equation was used to obtain estimates of $M(\text{m})$ for each microenvironment, or aggregate of microenvironments, as indicated in Table 2-6. The same value of $M(\text{m})$ was applied to each specific microenvironment within an aggregate.

Equation 7-33 requires a complete (gapless) year of hourly average fixed-site monitoring values for each district. Subsection 7.3 describes the method used to fill in missing hourly-average values. The resulting filled-in data sets were assumed to represent existing conditions at each monitor.

7.6.2 Spatial Interpolation of Monitoring Data

Subsection 7.6.1 describes a method for simulating spatial variability in ambient pollutant concentrations by adding random components to monitoring data representing a relatively small number of geographic locations within the study area. Alternatively, analysts may apply interpolation techniques to monitoring data obtained from a small number of sites to yield ambient pollutant estimates at a large number of geographic locations (e.g., the geographic centroids of all census tracts). As this approach has not been used recently with any OAQPS exposure model, it will be illustrated by a recent application of SHEDS-PM, an exposure model applicable to $\text{PM}_{2.5}$ developed by EPA-ORD (Burke, Zufall, and Ozkaynak, 2001).

SHEDS-PM was used to estimate $\text{PM}_{2.5}$ exposures in a study area consisting of 482 census tracts in and around Philadelphia, PA. Hourly-average estimates of ambient $\text{PM}_{2.5}$ concentration were obtained from a database developed by the Department of Statistics of the University of British Columbia (UBC). A report by UBC (June 28, 2000) describes the methodology used by UBC in creating the database and presents an analysis of the spatial and temporal patterns in the data. Briefly, UBC began the research effort by acquiring the following data sets:

- daily average $PM_{2.5}$ and PM_{10} data collected at eight monitoring stations (Figure 7-6) in and around Philadelphia from May 1, 1992 to September 15, 1993;
- hourly PM_{10} and $PM_{2.5}$ TEOM data were collected at one site (“NE Airport”) for approximately 8 weeks during July and August, 1992, and
- meteorological data (temperature, relative humidity, barometric pressure, visibility, dew point, hours of rain, wind speed in north direction, and wind speed in east direction) collected at the NE Airport site (Figure 7-6).

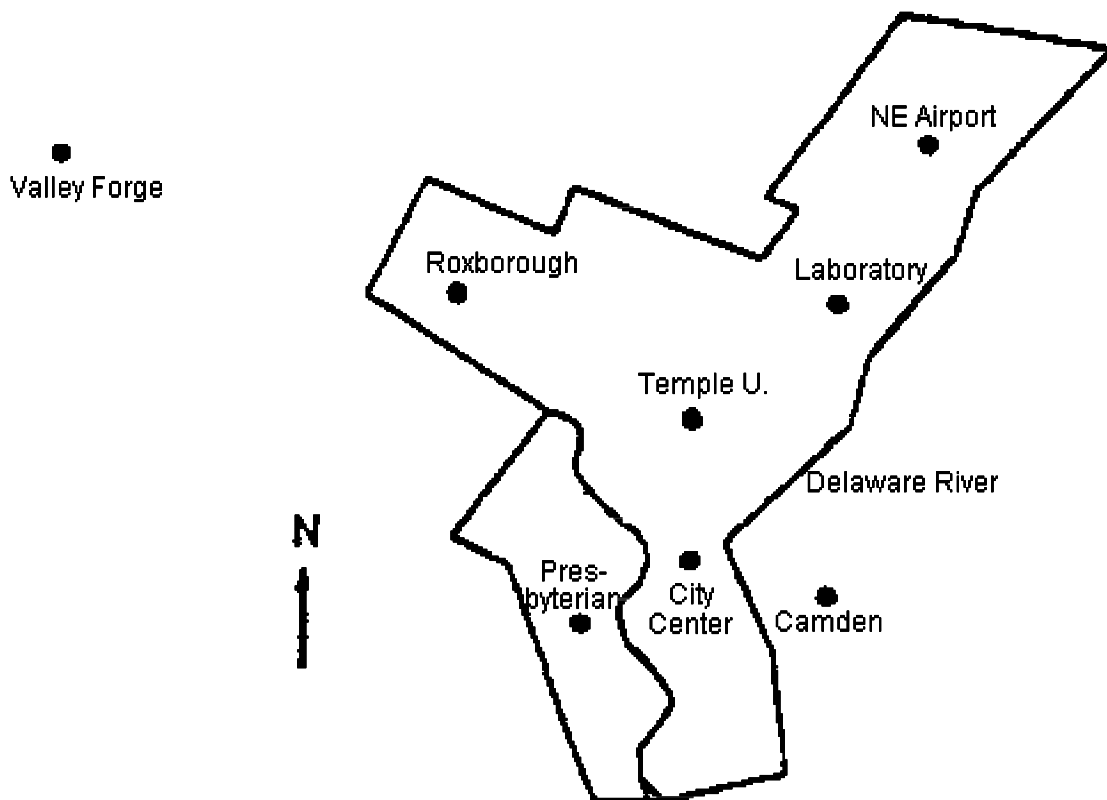


Figure 7-6. Location of eight PM monitoring stations providing data for development of interpolation method. Meteorological data provided by NE Airport station.

UBC performed a series of statistical analyses on these data sets to characterize temporal and spatial $PM_{2.5}$ patterns and to identify variables that could serve as

predictors of $PM_{2.5}$. They found that $PM_{2.5}$ levels displayed a marked "weekend/weekday" effect and a slight downward linear trend for the period under investigation. $PM_{2.5}$ levels were also found to be influenced by temperature, hours of rain, barometric pressure, visibility, and relative humidity. Based on the results of their analyses, UBC developed an interpolation scheme which they used to estimate hourly $PM_{2.5}$ values spanning the May 1992 - September 1993 period at the geographic centroid of each of 482 census tracts in the Philadelphia area. UBC organized these estimates into a database appropriate for input into the SHEDS-PM exposure model.

A more detailed description of the approach can be found in the following excerpts taken from the executive summary of the report prepared by UBC (June 28, 2000):

Based on methodology submitted earlier in a technical memo, investigators at the University of British Columbia, under contract with ManTech Environmental Technology, Inc., completed an analysis of Philadelphia pollution and meteorological data for the period May 1, 1999 to September 15, 1993, in the following stages:

Stage I. An analysis of the space-time structure of the data was conducted in the following steps:

1. preliminary examination of data quality from two sources [Harvard and the U.S. EPA's Aerometric Information Retrieval System (AIRS)], with special attention to the particulate variables -- PM_{10} and $PM_{2.5}$. (This examination and limitations of the spatial predictor used in the study, led to a decision to include only the Harvard data.)
2. data transformation of response variables to meet the normality assumption;
3. fitting a common deterministic process for the transformed data across monitoring sites using covariates such as time trend, meteorology and seasonality;
4. fitting a common auto-regressive process for the detrended residuals from Step 3 and imputing some missing values;
5. fitting an hourly model with the (less than) two months of $PM_{2.5}$ data available from the single monitoring site available-- 'NE Airport'.

Stage II. A spatial interpolation procedure for particulate levels was developed by:

6. using the deAR'd residuals -- the residuals after fitting the AR process in I(4) -- to fit a multivariate Gaussian Generalized Inverted Wishart model; this establishes the spatial correlation matrix between monitoring sites and the correlation matrix between pollutants in addition to estimating model parameters;

7. extending the covariance matrix between monitoring stations to the covariance matrix between all census tracts plus the existing stations via the Sampson-Guttorp method;
8. interpolating by methods explained later in this report, the residual values at the census tracts using the correlation structure from step 7;
9. calculating the interpolated particulate levels to the scale of the transformed responses in two steps: (a) adding the AR process from step I(4) to the interpolated residueal values; (b) adding to this result, the common deterministic process from I(3);
10. predicting hourly concentrations based on the model in I(5).

Stage III. Next, the interpolation procedure was assessed by:

11. using the interpolation procedure in Stage II to construct a predictive band for the $PM_{2.5}$ daily levels at Roxborough which had about a month of daily measurements of $PM_{2.5}$ during the summer of 1992. [These observations were excluded in building the interpolation model, because of the short period of operational time.]

Preliminary analysis showed particulate concentration distributions to be well approximated by a log-normal distribution. Thus, our analysis was carried out after logarithmically transforming the data.

The models for $\log PM_{2.5}$ and $\log PM_{10}$ proved to be very similar and for this summary, we will comment on only one of them, that for $\log PM_{2.5}$. The (BIC and AIC) methods used to fit the models tend to rule out the deleterious effects of collinearity and the significance of factors retained in the model is therefore unlikely be spurious.

Not surprisingly, daily $\log PM_{2.5}$ levels display a marked 'weekend/weekday' effect and levels revealed a slight downward linear trend for the period in under investigation. However increasing temperature tended to reverse that trend as did the number of hours of rain, and both horizontal components of wind speed when the wind was strong.

We also see distinct sinusoidal components in the series, these being enhanced by horizontal wind components.

Apart from its effect on trend, temperature has a direct positive impact on daily $\log PM_{2.5}$ levels and the degree of its impact increases quadratically.

Hours of rain also has a direct influence on daily $\log PM_{2.5}$ levels as did pressure and visibility. In fact, the latter has a quadratic effect. Relative humidity tends to reduce the level of daily $\log PM_{2.5}$ levels, and the size of its impact tends to be

quadratic. Over the range of the data, visibility had a negative effect on daily log $PM_{2.5}$ levels but that effect is less pronounced as the level of visibility increases.

Pressure tends to increase those levels. Interestingly, wind in the westerly direction tends to increase them, while wind in the northerly direction tends to decrease them.

The report also develops a trend model for hourly levels of log $PM_{2.5}$. Here the daily level of log $PM_{2.5}$ proves a good predictor although its effect can be reduced by both relative humidity and pressure. On its own the effect of pressure was found to be positive. That effect tends to decline quadratically although the rate of its decline can be reduced by relative humidity. Dew-point increases hourly levels of log $PM_{2.5}$ but that effect is reduced by visibility. Temperature, visibility and relative humidity all tend to pull down the positive effect of daily log $PM_{2.5}$ levels.

After trends have been removed from daily levels, we found the resulting series of residuals could be well described by an auto-regressive model of order three. That component too was then removed.

The resulting residual series then became the basis for the development of the spatial interpolator for daily levels. The latter proves to be quite accurate to the extent we were able to assess it. In particular, the relative frequency with which the associated prediction intervals cover the measured values is close to the nominal level.

To obtain a spatial predictor of hourly log $PM_{2.5}$, we added the trend component described above to daily log $PM_{2.5}$ levels. Finally, we transformed these values to original scale. A paucity of data for hourly $PM_{2.5}$ levels made an assessment of its accuracy impossible.

Note that UBC developed a "customized" interpolation approach based on the statistical characteristics and completeness of the available air quality and meteorological data. Because most applications of interpolation techniques will require sophisticated data analyses similar to those performed by UBC, researchers contemplating the use of such procedures should consult a qualified professional statistician for assistance.

7.7 References for Section 7

Akland, G., T. D. Hartwell, Ted R. Johnson, and Roy W. Whitmore. 1985. "Measuring Human Exposure to Carbon Monoxide in Washington, D.C., and Denver, Colorado, during the Winter of 1982-1983." Environmental Science and Technology, Volume 19, No. 10, pp. 911-918 (October 1985).

Barnett, V., and T. Lewis. 1978. Outliers in Statistical Data. John Wiley and Sons, New York.

Bloomfield, P. 1976. Fourier Analysis of Time Series: An Introduction. John Wiley and Sons, New York. Page 245.

Box, G. E. P., and G. M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Burke, J. M., M. J. Zufall, and H. Ozkaynak. 2001. "A Population Exposure Model for Particulate Matter: Case Study Results for PM_{2.5} in Philadelphia, PA." Journal of Exposure Analysis and Environmental Epidemiology, Vol. 11, No. 6, pp. 470-489.

Curran, T., W. Hunt, Jr., and R. Faoro. 1977. "Quality Control for Hourly Air Pollution Data." Presented at the 31st Annual Technical Conference for the American Society for Quality Control, Philadelphia. May 16 - 18.

Duff, M., R. L. Horst, and T. Johnson. 1998. "Quadratic Rollback: A Technique to Model Ambient Concentrations Due to Undefined Emission Controls." Paper No. 98-WA58.03, 91st Annual Meeting of the Air Pollution Control Association, San Diego. June.

Grant, E. L., and R. S. Leavenworth. 1996. Statistical Quality Control. McGraw-Hill, New York.

Grubbs, F. E., and G. Beck. 1972. "Extension of Sample Sizes and Percentage Points for Significance Test of Outlying Observations." Technometrics, Vol. 14, No. 4, November.

Hald, A. 1952. Statistical Theory with Engineering Applications. John Wiley and Sons, New York.

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. 1997a. "Sensitivity of Exposure Estimates to Air Quality Adjustment Procedure." Letter to Harvey Richmond, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. June 30.

Johnson, T. 1997b. "Evaluation of Quadratic and Weibull Air Quality Adjustment Procedures for pNEM/O₃." Letter to Harvey Richmond, Docket A95-58 (IV-A-11), Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. June 30.

Johnson, T. 1997c. "Evaluation of New Air Quality Adjustment Procedures for pNEM/O₃." Letter to Harvey Richmond, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. March 10.

Johnson, T. 1997d. "Exposure Estimates for the 8H7AV-84 Standard." Letter to Harvey Richmond, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. March 20.

Johnson, T., J. Capel, and M. McCoy. 1996. Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM and 1990 Population Data. Report prepared for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. April.

Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM. Report prepared by International Technology Air Quality Services under EPA Contract No. 68-D0-0062. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., G. Mihlan, J. Lapointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1). Report prepared for the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina.

Johnson, T., J. Mozier, and J. Capel. 1997. Supplement to Estimation of Ozone Exposures Experienced by Outdoor Children in Nine Urban Areas Using a Probabilistic Version of NEM (April 1996). Report prepared by TRJ Environmental, Inc., and IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., and L. Wijnberg. 1981. "Time Series Analysis of Hourly Air Quality Data." Paper No. 81-33.5, 74th Annual Meeting of the Air Pollution Control Association, Philadelphia. June.

Nelson, A.C., D. W. Armentrout, and T. R. Johnson. 1980. Validation of Air Monitoring Data. Publication No. EPA-600/4-80-030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, North Carolina 27711.

Ott, W. 1995. Environmental Statistics and Data Analysis. CRC Press, Boca Raton.

Paul, R. A., R. B. Faoro, and W. F. Hunt. 1981. "Screening SO₂ Data Anomalies from Four Test Cities." Technical Conference Transactions. American Society for Quality Control, Milwaukee, Wisconsin.

Remington, R., and M. Schork. 1970. Statistics with Applications to the Biological and Health Science. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Siegel, S. 1956. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New York.

University of British Columbia. 2000. Mapping Philadelphia's Hourly PM_{2.5} Field, 1992 - 93. Department of Statistics, University of British Columbia, Vancouver, British Columbia. June 28.

U.S. Department of Commerce. 1978. Computer Science and Technology: Performance Assurance and Data Integrity Practices. National Bureau of Standards, Washington, D.C. January.

USEPA. 1978. Screening Procedures for Ambient Air Quality Data. Publication No. EPA-450/2-78-037. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina. July.

USEPA. 2001. Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information. OAQPS Staff Paper. Report No. EPA-452/~~X-X-X~~. Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711. June draft.

Wilson, A.L., S.D. Colome, and Y. Tian. 1993. California Residential Indoor Air Quality Study, Volume 1, Methodology and Descriptive Statistics, Appendix. Integrated Environmental Services, Irvine, California.

SECTION 8

SIMULATION OF INDOOR EMISSIONS

This section describes the algorithms and data bases used in Version 2.1 of pNEM/CO to simulate CO emissions from two sources: gas stoves and passive smoking. Subsection 8.1 presents the methods used to determine the probability of gas stove use by time of day; the duration of use; the emission rates of burners and pilot lights as a function of fuel use, emission factors, and season; and prevalence rates for gas stoves in Denver and Los Angeles. Section 8.2 describes the mass balance models used to estimate the contribution of passive smoking in residences, bars, restaurants, and motor vehicles together with the algorithms used to determine the presence of smokers, number of cigarettes smoked per hour, and CO emission rate per cigarette.

8.1 Simulation of Gas Stove Operation

8.1.1 Probability of Stove Use

In Version 2.1 of pNEM/CO, the operation of gas stove burners in residences is simulated in the mass-balance model by specifying when the burners are on, the emission rate of the burners during operation, and the volume of the residence where the stove is located.

As discussed in Subsection 6.4.2.1, burner operation was assumed to occur in discrete BOPs such that use always began and ended within a single clock hour. BOP duration was assigned a value of either 30 or 60 minutes, depending on the time of day. These values were based on responses to a questionnaire administered by GEOMET to 4312 survey participants. Each participant provided data on the type of cooking facilities in the home, frequency of cooking, and average time spent in meal preparation (Koontz, Mehegan, and Nagda, 1992).

Table 8-1 presents a summary of data from this survey by type of meal (breakfast, lunch, and dinner). The values listed for average weekly time spent cooking breakfast, lunch, and dinner are 66, 71, and 288 minutes, respectively. The total time for all three meals is 425 minutes per week. The average daily cooking time based on this weekly value is $425/7$ or 61 minutes.

Table 8-1. Statistics on Gas Stove Use Obtained from a Survey by Koontz, Mehegan, and Nagda (1992)

Data item	Breakfast	Lunch	Dinner	Sum
Weekly duration of gas stove use, minutes	66	71	288	425
Weekly frequency of gas stove use	2.5	2.2	5.0	9.7
Average duration of use, minutes	26	32	58	

In addition to duration, the data in Table 8-1 provide an indication as to the fre-

quency that a gas stove is used to prepare meals in the typical residence. In one week, the stove will be used to prepare 2.5 breakfasts, 2.2 lunches, and 5.0 dinners -- a total of 9.7 meals per week. On an average day, the number of meals prepared on a gas stove is 9.7/7 or 1.4.

Dividing the weekly cooking time associated with each meal type by the average frequency of the meal yields average BOPs for breakfast, lunch, and dinner of 26, 32, and 58 minutes, respectively. Based on these results, pNEM/CO uses a value of 60 minutes for BOPs that occur during normal dinner hours and 30 minutes for BOPs that occur at other times.

In pNEM/CO, stove operation is determined on an hourly basis by comparing a randomly selected number between 0 and 1 with AP(h), the probability of a gas stove being operated during the indicated clock hour h (h = 1, 2, ..., 24). If the random number is less than AP(h), the stove is "on" for a duration of M(h) minutes, where M(h) is either 30 or 60 minutes, depending on the value of h. If the random number is greater than or equal to AP(h), the gas stove is "off" for the entire hour.

Table 8-2 lists the values of AP(h) and M(h) used in the pNEM/CO analysis by clock hour. These values were developed to (1) reflect diurnal patterns in gas stove usage specific to Denver, (2) yield an average daily duration for stove use of approximately 61 minutes, and (3) yield an average daily frequency of stove use of approximately 1.4.

Diurnal patterns in stove use were determined through an analysis of data from the Denver Personal Monitoring Study (Johnson, 1984). In this analysis, the diary entries and background questionnaire provided by each study subject were used to determine (1) when the subject was in a residence having a gas stove and (2) whether the stove was on. As working subjects would not always be present when other family members were operating a gas stove, it was assumed that workers would tend to under-report stove use in their residences. It was further assumed that nonworkers would use gas stoves more than the average person and that the diaries of nonworkers would tend to over-represent typical gas stove use. Consequently, the decision was made to average the worker and nonworker data and then adjust these results so that the adjusted P(h) values would yield 1.4 hours of stove use "events" per day, on average.

Table 8-3 presents the relevant data. For each clock hour, the table lists values of p(h) for workers and nonworkers calculated as

$$P(h) = [N(\text{stove on, GSR})]/[N(\text{GSR})] \quad (8-1)$$

where N(stove on, GSR) is the number of diary entries indicating the subject was in a gas stove residence when the stove was on and N(GSR) is the total number of diary entries indicating the subject was in a gas stove residence. In calculating these values, a stove was considered on during a particular clock hour if the subject's activity diary indicated at least 1 minute of use during the hour.

The column labeled "average P(h)" lists the arithmetic mean of the worker and nonworker P(h) values. These probabilities sum to 4.2 over 24 hours. It is desirable that the probabilities sum to 1.4, as this will produce an average of 1.4 BOPs per day. The values labeled "AP(h): adjusted P(h)" were calculated by multiplying the average Table 8-2. Probability of Gas Stove Use by Clock Hour and Assumed Burner Operation

Period

Clock hour	AP(h): probability of gas stove operation	M(h): assumed burner operation period, minutes	Product of AP(h) and M(h), minutes
1	0.025	30	0.76
2	0.023	30	0.68
3	0.023	30	0.69
4	0.023	30	0.70
5	0.023	30	0.70
6	0.026	30	0.77
7	0.049	30	1.46
8	0.058	30	1.73
9	0.081	30	2.43
10	0.073	30	2.20
11	0.062	30	1.86
12	0.075	30	2.25
13	0.085	30	2.54
14	0.071	30	2.14
15	0.067	60	4.01
16	0.064	60	3.86
17	0.107	60	6.41
18	0.130	60	7.80
19	0.091	60	5.49
20	0.058	60	3.45
21	0.052	60	3.11
22	0.047	60	2.79
23	0.040	30	1.21
24	0.035	30	1.04
Total	1.386		60.04

Table 8-3. Proportion of PEM Values in Gas Stove Residences with Stove in Operation by Clock Hour and Work Status

Clock hour	Nonworkers		Workers		Average P(h)	AP(h): adjusted P(h)
	n	P(h)	n	P(h)		
1	63	0.111	149	0.041	0.076	0.025
2	59	0.085	139	0.051	0.068	0.023
3	5	0.086	136	0.052	0.069	0.023
4	58	0.086	134	0.053	0.070	0.023
5	58	0.086	133	0.053	0.070	0.023
6	62	0.097	141	0.057	0.077	0.026
7	67	0.119	175	0.173	0.146	0.049
8	84	0.179	151	0.167	0.173	0.058
9	119	0.269	134	0.216	0.243	0.081
10	87	0.230	86	0.209	0.220	0.073
11	80	0.200	70	0.171	0.186	0.062
12	76	0.253	76	0.197	0.225	0.075
13	72	0.296	70	0.214	0.255	0.085
14	62	0.213	80	0.215	0.214	0.071
15	51	0.216	59	0.186	0.201	0.067
16	64	0.266	75	0.120	0.193	0.064
17	96	0.396	122	0.246	0.321	0.107
18	103	0.456	174	0.326	0.391	0.130
19	151	0.251	244	0.299	0.275	0.091
20	148	0.149	341	0.196	0.173	0.058
21	102	0.147	236	0.165	0.156	0.052
22	82	0.183	176	0.097	0.140	0.047
23	82	0.159	193	0.083	0.121	0.040
24	75	0.133	148	0.075	0.104	0.035
				Sum:	4.167	1.386

values by 0.333 (1.4/4.2). The adjusted values sum to 1.4.

The AP(h) values are listed also in Table 8-2. To the right of each AP(h) value is the assumed value of M(h); that is, the number of minutes the stove will be assumed to operate if the stove is determined to be "on" during the hour. The product of AP(h) and M(h) is listed in the far right column. Summing these values over all 24 hours provides an estimate of the average number of minutes per day that a gas stove will be operated according to the algorithm. The sum is approximately 60 minutes, a value very close to the desired value of 61 minutes.

8.1.2 Gas Stove Emission Rates

Version 2.1 of pNEM/CO differentiates between gas stoves with and without electronic ignitions. The Monte Carlo algorithm described in Subsection 8.1.4 was used to randomly determine pilot light status for each cohort defined as having gas stoves. Gas stoves without electronic pilot ignition were assumed to have continuously burning pilot lights. The mass of CO emitted by gas stove with a pilot light during a particular hour (h) was estimated by the equation

$$\text{MASSCO}(h) = (\text{ERBURN})[M(h)]/60 + (\text{ERPILOT})(1 \text{ hr}) \quad (8-2)$$

where MASSCO(h) is expressed in mg, ERBURN is the hourly burn emission rate in mg per hour, and ERPILOT is the hourly pilot light emission rate in mg per hour. M(h) is the duration of burner use during hour h expressed in minutes. The pilot light is assumed to on continuously during the one hour period.

The mass of CO emitted by gas stove with electronic ignition was estimated by the equation

$$\text{MASSCO}(h) = (\text{ERBURN})[M(h)]/60; \quad (8-3)$$

i.e., the ERPILOT term in Equation 8-2 was set equal to zero.

In both equations, M(h) is zero for each hour in which the algorithm assigns the stove a status of "off." If the stove status is "on" for a particular hour, M(h) is assigned a value of 30 or 60 minutes according to Table 8-2.

ERBURN was determined by the equation

$$\text{ERBURN} = (\text{AUB}/365.2)(\text{EFBURN})(Y') \quad (8-4)$$

where AUB is the annual fuel usage of the burners in kilojoules, 365.2 is the number of hours per year that the burners are operated assuming 60 minutes of use per day (Table 8-1), EFBURN is the burner emission factor in mg of CO per kilojoule, and Y' is an adjustment factor which varies sinusoidally throughout the year. The values of AUB, EFBURN, and Y' are specific to cohort.

Subsections 8.1.2.1 and 8.1.2.2 provide estimates for the parameter values in Equations 8-2 through 8-4 specific to Denver and Los Angeles, respectively. The development of Denver parameter values is discussed first, as many of the Los Angeles values were derived from the corresponding Denver values through the use of an adjustment factor.

8.1.2.1 Denver Estimates

Values of AUB for Denver were randomly selected from a lognormal distribution with geometric mean = 2.11 million kilojoules and geometric standard deviation = 1.48. This distribution is based on the distribution of annual burner gas use measured by the Northern Illinois Gas Company (NIGAS) in 57 homes (Menkedick, Niemuth, Hartford, and Landstrom, 1993). The value of AUB was not permitted to exceed 4.55 million kilojoules. This value represents the 97.5th percentile of the distribution.

The seasonal adjustment factor (Y') was determined by the equation

$$Y'(j) = 1.00 - (0.190)\{\sin[-1.616 + (2\pi)(j)/365]\}. \quad (8-5)$$

in which j is the Julian date. This equation was derived by Johnson (1998) from a sinusoidal pattern observed in a study conducted by NIGAS (Wilkes and Koontz, 1995). The NIGAS data indicate that gas use in the winter is approximately 46 percent higher than in the summer.

Values of EFBURN were randomly selected from a lognormal distribution with geometric mean = 0.0294 mg/kilojoule and geometric standard deviation = 2.77. Values of EFBURN were not permitted to exceed 0.400 mg/kilojoule. These values are based on the results of an analysis of data reported by Davidson, Borrazzo, and Hendrickson (1987) and represent a well-adjusted stove. As such, the assumed geometric mean is probably low with respect to the overall population of gas stoves.

Consistent with the above discussion, the following algorithm was used to estimate a value of ERBURN for each 24-hour period.

1. Randomly select a value for AUB from a lognormal distribution with geometric mean = 2.11 million kilojoules per year and geometric standard deviation = 1.48.
2. Randomly select a value of EFBURN from a lognormal distribution with geometric mean = 0.0294 mg/kilojoule and geometric standard deviation = 2.77.
3. Go to the next day of the current sequence. The Julian date of this day is j .
4. Use Equation 4-23 to calculate $Y'(j)$.
5. $ERBURN = (EFBURN)(AUB)(Y')/365.2$ for all hours of the day.
6. Go to Step 3. Repeat for all days of the year.

The resulting daily values of ERBURN were inserted into Equation 8-3 as required.

The ERPILOT value in Equation 8-2 was determined by the equation

$$ERPILOT = (AUP/8760)(EFPILOT) \quad (8-6)$$

where AUP is the annual fuel usage by all pilot lights in kilojoules, 8760 is the number of hours per year that the pilot lights are in operation, and EFPILOT is the pilot light emis-

sion factor in mg of CO per kilojoule. The value of AUP was held constant over the entire year and was randomly selected from a specified lognormal distribution. The value of EFPILOT was assumed to be constant over the entire year and was set equal to the value determined for EFBURN in Equation 8-4.

The distribution for gas usage by pilot lights (AUP) was based on data from the NIGAS study discussed previously. The total gas usage (burners plus pilot lights) for 33 stoves had an arithmetic mean of 57.1 therms and a standard deviation of 18.3 therms. Total gas use was also measured for 57 stoves that did not have pilot lights; the arithmetic mean gas use for stoves without pilot lights was 21.8 therms. The difference between the two samples, $57.1 - 21.8 = 35.3$ therms, provided an estimate of the mean fuel usage for pilot lights only. The square root of the differences in variances, $(18.3^2 - 8.9^2)^{1/2} = 16.0$ therms, provided an estimate of the standard deviation. Consequently, AUP was assumed to have an arithmetic mean of 35.3 therms (3,500 ft³) and an arithmetic standard deviation of 16.0 therms (1,600 ft³).

The ratio of standard deviation to mean is 0.45, indicating that the distribution is skewed. Consequently, analysts assumed that the underlying distribution was lognormal. The corresponding geometric mean and geometric standard deviation for this distribution were 3,215 ft³ = 3.37 million kilojoules per year and 1.84 (dimensionless), respectively. Values of AUP were randomly selected from this distribution as needed. Values of AUP were not permitted to fall below 1.02 million kilojoules per year or above 11.13 million kilojoules per year. These bounds correspond to the 2.5th and 97.5th percentiles, respectively, of the specified lognormal distribution.

The value of MASSCO(h) as determined by Equation 8-2 or 8-3 was used as the value of S for hour h in Equations 6-10 and 6-15, regardless of the value of M(h). This approach permitted the use of the hourly average exact solution (Equation 6-7) to the mass-balance equation (Equation 6-2). The practical result of this simplification was a slight smoothing in the simulated hour-to-hour variation in indoor CO concentrations with respect to the pattern which would be simulated by a model with finer time resolution.

8.1.2.2 Los Angeles Estimates

As discussed above, a study conducted by NIGAS was the principal source of the AUB and AUP distributions developed for Denver. Analysts obtained gas stove data specific to Los Angeles which provided a basis for adjusting the NIGAS-derived values for application to Los Angeles. Table 8-4 lists data for six utility districts provided by the Demand Analysis Office of the California Energy Commission (CEC, 1998). Three of the districts span the Los Angeles study area (LADWP, BDG, and SCE). The average gas use per stove in these three districts ranges from 45.0 to 49.2 therms with 47 therms representing a reasonable overall estimate. Note that these values specify total gas use by the stove; the CEC did not provide separate estimates for burners and pilot lights.

As previously discussed, the NIGAS study reported an average of total gas usage of 57.1 therms per stove based on 33 stoves. Based on the assumption that Los Angeles residents use less gas per stove than the population sampled by NIGAS (47 therms vs. 57.1 therms), analysts calculated an appropriate adjustment ratio by the

Table 8-4. Gas Stove Data Provided by Demand Analysis Office of the California Energy Commission.

Utility district	Approximate geographic area	Percent cooking w/gas	Number of gas stoves	Therms used by all stoves (millions)	Therms per stove ^a
Pacific Gas and Electric (PG&E)	northwest of Los Angeles	33.0	1,399,887	67.8	48.4
Sacramento Municipal Utility Division (SMUD)	Sacramento	30.5	130,319	6.2	47.6
Southern California Edison (SCE)	South coast outside downtown Los Angeles	68.6	2,772,522	131	47.3
Los Angeles Department of Water and Power (LADWP)	Los Angeles	78.6	982,891	48.4	49.2
San Diego Gas and Electric (SDG&E)	San Diego	46.5	486,537	20.5	42.1
Combined Municipal Districts of Burbank, Glendale, and Pasadena (BDG)	Burbank, Glendale, Pasadena	78.4	131,111	5.9	45.0

^aCalculated by Ted Johnson.

^bSource: California Energy Commission (1998).

expression

$$\text{Adjustment ratio} = (47 \text{ therms}) / (57.1 \text{ therms}) = 0.82. \quad (8-7)$$

The adjusted GM for annual burner use (AUB) in Los Angeles was 2.11 million kilojoules $\times 0.82 = 1.73$ million kilojoules. The corresponding GSD of 1.48 was unchanged, as this is a dimensionless quantity.

Using the same adjustment approach, the GM for annual pilot light use (AUP) was estimated as 3.37 million kilojoules $\times 0.82 = 2.76$ million kilojoules. Again, the GSD remained unchanged at 1.84. It should be noted that the application of the adjustment factor to the NIGAS pilot light data assumes that pilot light use is proportionally lower in Los Angeles, an assumption which cannot be currently verified.

Analysts assumed that the Denver estimates for EFBURN and EFPILOT could be applied to Los Angeles. Consistent with the Denver approach, Equations 8-2 and 8-3 were used to estimate MASSCO(h) values for Los Angeles gas stoves with and without pilot lights. The values of ERBURN and ERPILOT in this equation were similarly estimated by Equations 8-4 and 8-6, respectively. Equation 8-5 provided the estimate of $Y'(j)$ for Equation 8-4.

8.1.3 Gas Stove Prevalence Rates

The equations used to estimate cohort populations require an estimate of $F(h,f)$, defined as the fraction of homes in Home district h that use cooking fuel f . Analysts estimated that $F(h,f) = 19.6$ percent for the Denver study area based on data obtained from the Bureau of Census (1995). The demographic data listed in Table 8-4 was used to estimate $F(h,f) = 79$ percent for the Los Angeles study area. The data were compiled from three reports (Bureau of Census, 1996a, 1996b, 1997) listing statistics obtained from American Housing Survey (AHS). The area labeled Los Angeles in Table 8-5 includes the Los Angeles - Long Beach PMSA (i.e., Los Angeles County), a region which contains most of the census tracts included in the study area defined for Los Angeles pNEM/CO analysis. The remaining census tracts are located just outside Los Angeles County in Orange County or in Riverside County. Because of their similar demographics and close proximity to Los Angeles County, analysts assumed that these bordering census tracts were likely to be better represented by the prevalence rate of Los Angeles County (79.8 percent) than by rates listed for Orange County (61.9 percent) or Riverside County (77.4 percent).

The 79.8 percent prevalence rate for Los Angeles County obtained from the AHS is consistent with the gas stove use rates listed in Table 8-4 for LADWP (78.6 percent) and BDG (78.4 percent). Taken together, the three estimates support an estimate of about 79 percent for the gas stove prevalence rate.

8.1.4 Probability of Gas Stove Having Electronic Ignition

A Monte Carlo algorithm was used to randomly determine the pilot light status (pilot light or electronic ignition) for each cohort defined as having gas stoves. This approach required estimates of prevalence rates for pilot lights in Denver and Los Angeles. The CO criteria document (USEPA, 2000, p. 3-46) provided two sources of

such data:

1. Koontz, Mehegan, and Nagda. Distribution and Use of Cooking Appliances that can Affect Indoor Air Quality: Topical Report. 1992.
2. Wilson, Colome, and Tian. California Residential Indoor Air Quality Study, Volume 1, Methodology and Descriptive Statistics, Appendix. 1993.

Based on a national sample of homes, Koontz, Mehegan, and Nagda (1992) estimated that the fraction of U. S. gas stoves with electronic ignition was 20 percent in 1985 and 27 percent in 1991. Wilson, Colome, and Tian (1993) surveyed 293 households in California during 1991 and identified 142 homes as having gas stoves. Of the gas stove homes, 70 (49 percent) had electronic ignition. Note that the estimate obtained from Wilson, Colome, and Tian (1993) for California (49 percent) is significantly higher than the estimate obtained from Koontz, Mehegan, and Nagda (1992) for the nation (27 percent in 1991). Unable to find data specific to Denver, analysts decided to use the California value (49 percent) for Los Angeles and the national value (27 percent) for Denver.

Table 8-5. Fuel Use Statistics from the American Housing Survey (Bureau of Census, 1996a, 1996b, 1997).

Statistics	Anaheim (1994) ^a	Los Angeles (1995) ^b	Riverside (1994) ^c
Total Housing Units (HU)	918,000	3,276,000	1,121,400
Total HU w/cooking fuel	915,500	3,165,200	1,101,900
Total HU w/natural gas cooking fuel	568,700	2,614,000	867,500
Prevalence of HU w/gas cooking fuel ^d	61.9 percent	79.8 percent	77.4 percent

^aAnaheim = Orange County PMSA (includes all of Orange County).

^bLos Angeles = Los Angeles - Long Beach PMSA (includes all of Los Angeles County).

^cRiverside = Riverside-San Bernardino PMSA (Includes all of Riverside and San Bernardino Counties).

^dPrevalence rate = 100 x (total HU w/natural gas cooking fuel)/(total housing units).

These estimates were incorporated into a probabilistic algorithm for determining the pilot light status. The algorithm randomly selected a number between 0 and 1 for each cohort identified as having gas stoves. The selected number was compared to a city-specific value of X (0.49 for Los Angeles and 0.27 for Denver). If the selected number was X or less, the cohort was assumed to have gas stoves with electronic ignitions. Otherwise, the cohort was assumed to have gas stoves with pilot lights which burned continuously. Whenever a cohort was assigned to the electronic ignition

category, the value for “annual fuel use by gas stove pilot light” was set equal to zero in the gas stove emission algorithm (see Subsection 8.1.2).

8.2 Simulation of Passive Smoking

This subsection begins with a description of the method used to estimate CO emission rate from cigarettes in Version 2.1 of pNEM/CO. It then shows how this estimate is converted into CO levels in selected microenvironments.

8.2.1 Estimation of CO Emission Rate from Passive Smoking

Subsection 2.4.1 presented an overview of the methods used to estimate CO concentration in each microenvironment. Passive smoking was considered to be a significant source of CO in the following indoor and vehicle microenvironments: No. 1: indoors - residence, No. 4: restaurants (when permitted), No. 5: bars (when permitted), No. 12: automobiles, and No. 13: trucks. Subsections 8.2.2 through 8.2.4 describe the algorithms used to estimate the contribution of passive smoking in each of these microenvironments. Each of these algorithms requires an estimate of CO emission rate from smokers present in the microenvironment. This estimate was obtained from the equation

$$S = (n_{\text{smokers}})(n_{\text{cigs/smoker/hr}})(CO_{\text{cigarette}}) \quad (8-8)$$

in which S is the emission rate in $\mu\text{g/hr}$, n_{smokers} is the number of smokers present, $n_{\text{cigs/smoker/hr}}$ is the number of cigarettes smoked per hour per smoker, and $CO_{\text{cigarette}}$ is the mass of CO emitted per cigarette. The values used for n_{smokers} and $n_{\text{cigs/smoker/hr}}$ are presented by microenvironment in Subsections 8.2.2 through 8.2.4.

The values used for $CO_{\text{cigarette}}$ were randomly selected from a lognormal distribution with geometric mean equal to 71,400 $\mu\text{g/cigarette}$ and geometric standard deviation equal to 1.3. This distribution was developed by Traynor, Aceti, Apte, et al. (1989) based on data from six studies published between 1982 and 1988. The distribution appears to be generally consistent with other research in the scientific literature.

The 5th and 95th percentiles of the Traynor distribution cited above are 46,400 and 109,900 $\mu\text{g/cigarette}$. This range is generally consistent with the range of estimates reported in most studies cited in the Criteria Document (EPA, 2000). Values selected for $CO_{\text{cigarette}}$ from the Traynor distribution were not permitted to fall outside these bounds.

8.2.2 Residential Locations

Version 2.1 of pNEM/CO treats CO from ETS in the residence as an incremental addition to the CO concentration obtained from the one-hour mass-balance model. The incremental CO concentration is determined through the use of a one-minute mass-balance model that accounts for air exchange rate, residential volume, and CO emission rate. This mass-balance model determines the average CO concentration for each minute of the event. These average one-minute CO values are then averaged

over the duration of the event to determine the incremental CO concentration from ETS for the event.

As previously discussed in Subsection 6.4.1, the instantaneous indoor CO concentration at the end of minute m can be calculated as

$$C_{in}(m) = k_1 C_{in}(m - 1) + k_2 \text{CAVG}_{out}(m) + k_3 \quad (8-9)$$

where

$$k_1 = e^{-v} \quad (8-10)$$

$$k_2 = 1 - e^{-v} \quad (8-11)$$

$$k_3 = (S)(1 - e^{-v})/(vV) \quad (8-12)$$

The average indoor CO concentration for minute m can be calculated as

$$\text{CAVG}_{in}(m) = a_1 C_{in}(m - 1) + a_2 \text{CAVG}_{out}(m) + a_3 \quad (8-13)$$

$C_{in}(m - 1)$ is the instantaneous indoor concentration at the end of the preceding minute and $\text{CAVG}_{out}(m)$ is the average outdoor concentration during minute m . The other variables appearing in Equation 8-13 are defined by the following equations:

$$a_1 = z(m) \quad (8-14)$$

$$a_2 = 1 - z(m) \quad (8-15)$$

$$a_3 = (S)[1 - z(m)]/(vV) \quad (8-16)$$

$$z(m) = (1 - e^{-v})/v \quad (8-17)$$

Equations 8-9 and 8-13 can be used to construct a sequence of minute-average values for the duration of a particular exposure event.

With appropriate simplifications, this model was applied to each continuous sequence of exposure events in the residence in which each event indicated the occurrence of passive smoking. The CO from ETS was assumed to equal zero prior to the beginning of sequence. Thus the term $C_{in}(m - 1)$ in Equation 8-9 was zero for the first minute of the first event in the sequence. As the calculation was limited to the incremental effects of CO emitted within the residence, the outdoor term $\text{CAVG}_{out}(m)$ in Equations 8-9 and 8-13 was also set equal to zero. Consequently, Equation 8-9 simplified to

$$C_{in}(m) = k_1 C_{in}(m - 1) + k_3 \quad (8-18)$$

and Equation 8-13 simplified to

$$\text{CAVG}_{in}(m) = a_1 C_{in}(m - 1) + a_3 \quad (8-19)$$

To reduce computational time, analysts also assumed that all CO dissipated immediately after the end of the last smoking event in a sequence of smoking events; i.e., there was no residual CO from smoking during non-smoking exposure events. Consistent with this assumption, the one-minute mass-balance model was not run during non-smoking events. This approach tended to slightly under-estimate the total incremental CO contribution of passive smoking.

The following eight-step procedure was used to model each continuous sequence of smoking events within a residence.

1. The volume of the residence is determined for the cohort by an existing algorithm within pNEM/CO (see Subsections 6.4.2.1 and 6.4.5.1). This volume is held constant for all smoking events associated with the cohort.
2. The air exchange rate for the event is the value assigned to the hour containing the event by an existing algorithm (see Subsections 6.4.3.1 and 6.4.3.2) within pNEM/CO.
3. The smoking status of the event is determined by an existing code in the diary data base.
4. If smoking occurs during the event, the CO emission rate for ETS is determined by the algorithm in Table 8-6. This value is used for all passive smoking events associated with the cohort.
5. Equations 8-18 and 8-19 are used to determine the instantaneous CO concentration for the end of each minute and the average CO concentration for each minute, respectively.
6. The average CO concentration for the event is determined by averaging the minute-average CO concentrations for the event.
7. The inside CO concentration at the beginning of the next event is set equal to the instantaneous CO concentration calculated in Step 5 for the end of the last minute of the first event.
8. Repeat Steps 2 through 7 for each subsequent event in the sequence.

The one-minute CO concentrations determined by this algorithm were averaged by exposure event and added to the CO concentration estimated by the one-hour mass-balance model for the exposure event.

The algorithm for estimating CO emission rate incorporates the assumptions that there is only one smoker present during a passive smoking event, that the smoker consumes two cigarettes per hour, and that the CO emitted per hour per cigarette is characterized by a lognormal distribution with geometric mean = 71,400 μg and geometric standard deviation = 1.3. The assumptions concerning smoking prevalence and smoking rate are based on estimates presented by Repace, Jinot, Bayard, et al. (1998). These researchers estimated that chain smokers smoke approximately 6

cigarettes per hour and that average smokers smoke approximately 2 cigarettes. Repace, Jinot, Bayard, et al. (1998) also estimated that roughly 25 percent of Americans smoke tobacco products.

Table 8-6. Algorithm for Estimating CO Emission Rate from Passive Smoking in the Residential Microenvironment.

1.	If smoking status algorithm indicates that smoking occurs during exposure event, go to Step 2. Otherwise, CO emission rate from tobacco products is set at zero for exposure event.
2.	Assume there is one smoker in residence ($n_{\text{smoker}} = 1$).
3.	Assume that the smoker consumes two cigarettes per hour ($n_{\text{cigs/smoker/hr}} = 2$).
4.	Determine CO emitted per hour from each cigarette by randomly selecting a value from a lognormal distribution with geometric mean = 71,400 μg and geometric standard deviation = 1.3.
5.	The hour emission rate ($\mu\text{g/hr}$) is calculated by the equation $S_{\text{hr}} = (n_{\text{smokers}})(n_{\text{cigs/smoker/hr}})(\text{CO}_{\text{cigarette}}).$
6.	Convert S_{hr} value obtained in Step 5 to minute emission rate ($\mu\text{g/min}$) by the equation $S_{\text{min}} = (S_{\text{hr}})/60.$

8.2.3 Restaurants and Bars

The hour mass-balance model (Equation 6-7) was used estimate CO concentrations in restaurants and bars in both study areas. As discussed in Subsection 6.4.2.2, ETS was considered to be the only potential indoor source of CO in these microenvironments. Consequently, the a_3 term in Equation 6-7 was used solely to account for the effect of passive smoking. Because smoking is prohibited in Los Angeles bars and restaurants, a_3 was set to zero when pNEM/CO was applied to Los Angeles. Passive smoking was assumed to occur continuously in Denver bars and restaurants, however, as local regulations permit smoking in these locations. In Denver applications, pNEM/CO calculated a_3 using an alternative to Equation 6-10 which better utilized existing databases. This subsection describes the derivation of this alternative equation. The approach is similar to one employed by the California Air Resources Board in estimating pollution increments from ETS in public buildings (Miller, Branoff, Lim, et al., 1998).

The complete hour mass-balance model for CO is described in Subsection 6.4.1.

Equation 6-41 provides estimates of the average indoor pollutant concentration of hour h. The CO contribution of indoor sources such as ETS is represented by an emission rate variable (S) which appears in the a_3 term (see Equation 6-44). In estimating the CO levels in bars and restaurants, analysts assumed that cigarette smoke was the only potential source of CO and estimated the emission rate from passive smoking by the expression

$$S_{hr} = (n_{smokers})(n_{cigs/smoker/hr})(CO_{cigarette}). \quad (8-20)$$

in which $n_{smokers}$ is the number of smokers present, $n_{cigs/smoker/hr}$ is the average number of cigarettes smoked per hour by each smoker, and $CO_{cigarette}$ is the average CO emission rate of the smoked cigarettes.

Ordinarily, the a_3 term would be calculated as

$$a_3 = (S)[1 - z(h)]/(vV) \quad (8-21)$$

in which

$$z(h) = (1 - e^{-v})/v \quad (8-22)$$

and S is determined by Equation 8-20. This approach requires estimates for both the number of smokers ($n_{smokers}$) and the volume of the facility (V). Analysts were not able to identify reliable data for estimating these parameters specific to bars and restaurants.

Following the example of Miller, Branoff, Lim, et al. (1998), analysts implemented an alternative approach in which both the numerator and denominator of the a_3 term were divided by the number of occupants of the facility. Thus, the S term in the numerator of the a_3 term was calculated as

$$S_{hr} = (Fr)(n_{cigs/smoker/hr})(CO_{cigarette}), \quad (8-23)$$

in which Fr is the fraction of occupants who are smokers. The Fr parameter was estimated according to existing data on the prevalence rate of smokers in study area (Denver or Los Angeles). In the denominator of the a_3 term, the product of the variables v and V (m^3/hr) was replaced by the normalized ventilation rate ($m^3/hr/occupant$), hereafter denoted by NVR. NVR can be estimated from local building codes which specify average ventilation rate per occupant according to business type. With these substitutions, the a_3 term can be expressed as

$$a_3 = (Fr)(n_{cigs/smoker/hr})(CO_{cigarette})[1 - z(h)]/(NVR). \quad (8-24)$$

Consistent with this approach, the k_3 term in Equation 6-12 would be expressed as

$$k_3 = (Fr)(n_{cigs/smoker/hr})(CO_{cigarette})(1 - e^{-v})/(NVR). \quad (8-25)$$

Los Angeles does not permit smoking in restaurants and bars. Consequently, a_3 is set equal to zero for Los Angeles. Denver, which permits smoking in restaurants and bars, requires that these businesses meet the following 1989 ventilation standards developed by the American Society of Heating, Refrigeration, and Air Conditioning

Engineers (ASHRAE).

Restaurant: 20 cfm/person (34 m³/hour/person)
Bar: 30 cfm/person (51 m³/hour/person)

Analysts considered two methods for using this information. In Method A, analysts would use 34 and 51 m³/hour/person as point estimates for NVR when applying the special mass-balance model to Microenvironment No. 4 (restaurants) and No. 5 (bars), respectively. In Method B, analysts would treat these values as the geometric means of lognormal distributions. Each distribution would have a geometric mean of 1.81. Values of NVR would be selected from these distributions as required.

In developing Method B, analysts noted that the ASHRAE standard for restaurants (34 m³/hr/person) was consistent with the geometric mean derived by CARB from the Persily (1989) study of 14 office buildings (36.9 m³/hr/person). The geometric standard deviation for this data set was 1.81. In Method B, analysts assumed that restaurants had a lower geometric mean (34 m³/hr/person) but the same geometric standard deviation (1.81). Similarly, bars were assumed to have a higher geometric mean (51 m³/hr/person) but the same geometric standard deviation.

Consistent with EPA's goal of using probabilistic elements to represent the majority of parameters affecting exposure, analysts selected the distributional alternative (Method B) for the pNEM/CO analyses described in this report.

The CDC's Morbidity and Mortality Weekly Report of November 6, 1998 (Vol. 47, No. 43) indicates that the 1997 adult smoking prevalence rate for Colorado was 22.6 percent. Based on this value, analysts set Fr equal to 0.226 for Denver restaurants and bars.

To maintain consistency between the NVR value in the a₃ and k₃ terms and the air exchange rate (v) used elsewhere in hour mass-balance algorithm (i.e., the NVR should be relatively high when the air exchange rate is relatively high), the same distribution percentile was used in determining both values. For example, the 30th percentile of the NVR distribution was used when the 30th percentile of the air exchange rate distribution was selected. Table 8-7 presents the algorithm used to estimate the modified variables for the a₃ and k₃ terms of the hourly mass-balance algorithm for bars and restaurants in Denver.

8.2.4 Passenger Vehicles

The one-minute mass-balance model was used to estimate the combined minute-by-minute contribution of outdoor CO and in-vehicle ETS to CO levels in Microenvironments Nos. 12 (automobiles) and 13 (trucks). The model was implemented according to the 14-step procedure presented in Table 6-9. The equation in Step 2 of the algorithm provided an estimate of the CO emission rate from passive smoking in the vehicle. The equation assumes that the vehicle was occupied by only one smoker during each passive smoking event, the smoker consumed two cigarettes per hour on average, and the CO emission rate could be characterized by a lognormal distribution with geometric mean = 71,400 µg/hr and geometric standard deviation = 1.3.

Table 8-7. Algorithm for Estimating CO Emission Rate from Passive Smoking in Restaurants and Bars (Method B in Text).

1.	Assume the fraction of people in the bar/restaurant at any time is the same as the adult prevalence rate of smoking in Colorado for the year of simulation ($Fr = 0.226$).
2.	Assume that each smoker consumes two cigarettes per hour ($n_{\text{cigs/smoker/hr}} = 2$).
3.	Determine the CO emitted per hour from each cigarette by randomly selecting a value from a lognormal distribution with mean = 71,400 μg and geometric standard deviation = 1.3. Note percentile (P) of selected value.
4.	The hour emission rate per occupant ($\mu\text{g/hr/occupant}$) is calculated by the equation
	$S_{\text{hr}} = (Fr) (n_{\text{cigs/smoker/hr}})(CO_{\text{cigarette}}).$
5.	Determine the hour normalized ventilation rate ($\text{m}^3/\text{hr/occupant}$) as the value corresponding to percentile = P of a lognormal distribution with geometric standard deviation = 1.81 and geometric mean indicated below. P is the value previously determined in Step 3.
	Restaurant: 34 $\text{m}^3/\text{hr/person}$
	Bar: 51 $\text{m}^3/\text{hr/person}$

8.3 References for Section 8

Bureau of Census. 1995. "American Housing Survey for the Denver Metropolitan Area in 1995." Current Housing Reports. Report No. H170/95-46. U. S. Department of Commerce and U.S. Department of Housing and Urban Development.

Bureau of Census. 1996a. "American Housing Survey for Anaheim-Santa Ana Metropolitan Area in 1995." Current Housing Reports. Report No. H170/94-2. U.S. Department of Commerce and U.S. Department of Housing and Urban Development.

Bureau of Census. 1996b. "American Housing Survey for Riverside-San Bernardino-Ontario Metropolitan Area in 1995." Current Housing Reports. Report No. H170/94-37. U.S. Department of Commerce and U.S. Department of Housing and Urban Development.

Bureau of Census. 1997. "American Housing Survey for Los Angeles-Long Beach Metropolitan Area in 1995." Current Housing Reports. Report No. H170/95-7. U.S.

Department of Commerce and U.S. Department of Housing and Urban Development.

California Energy Commission. 1998. "Residential Demand Forecast Model, 1998 Baseline Energy Outlook." Demand Analysis Office.

Center for Disease Control. 1998. "Morbidity and Mortality Weekly Report." Vol. 47, No. 43 (November 6, 1998).

Davidson, C. L., J. E. Borrazzo, and C. T. Hendrickson. 1987. Pollutant Emission Factors for Gas Stoves. Report No. EPA/600/9-87-005. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T. 1984. A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Report No. EPA-600/54-84-014. U.S. Environmental Protection Agency, Research Triangle Park, N.C.

Johnson, T. 1998. "Analyses Supporting Proposed Modifications to the Mass-balance Algorithm and Gas Stove Algorithm in pNEM/CO." Memorandum dated September 23, 1998, prepared by TRJ Environmental, Inc., under EPA Contract No. 68-D6-0064.

Koontz, M. D., L. L. Mehegan, and N. L. Nagda. 1992. Distribution and Use of Cooking Appliances That Can Affect Indoor Air Quality. Report No. GRI-93/0013. Gas Research Institute, Chicago.

Menkedick, J., N. Niemuth, P. Hartford, and D. K. Landstrom. 1993. Metered Ranges, Cooktops, and Ovens in the Northern Illinois Gas Residential Load Study Data Base. GRI Report No. 93/0204. Report prepared by Battelle for the Gas Research Institute, 8600 West Bryn Mawr Avenue, Chicago, IL 60631.

Miller S.L., S. Branoff, Y. Lim, D. Liu, Van Loy, and W.W. Nazaroff. 1998. "Assessing Exposure to Air Toxicants from Environmental Tobacco Smoke." Contract No. 94-344. California Air Resources Board, Sacramento, California.

Persily, A. 1989. "Ventilation Rates in Office Buildings." IAQ '89 The Human Equation: Health and Comfort. American Society of Heating, Refrigeration, and Air Conditioning Engineers, Inc. Atlanta. pp. 128 - 136.

Repace, J.L., J. Jinot, S. Bayard, K. Emmons, and S. K. Hammond. 1998. "Air Nicotine and Saliva Cotinine as Indicators of Workplace Passive Smoking Exposure and Risk." Risk Analysis, Volume 18, pp. 71 - 83.

Traynor, W.T., J. Aceti, M. Apte, B. Smith, L. Green, A. Smith-Reiser, K. Novak, and D. Moses. 1989. Macromodel for Assessing Residential Concentrations of Combustion-Generated Pollutants: Model Development and Preliminary Predictions for CO, NO₂, and Respirable Suspended Particles. Report No. LBL-25211. Lawrence Berkeley Laboratory, Berkeley, CA 94720.

U. S. Environmental Protection Agency. 2000. Air Quality Criteria for Carbon Monoxide. Report No. EPA/600/P-99/001F. U. S. Environmental Protection Agency, National Center for Environmental Assessment, Research Triangle Park, NC.

Wilkes, C.R., and M.D. Koontz. 1995. "Analysis of Sampling Strategies for Estimating Annual Average Indoor NO₂ Concentrations in Residences with Gas Stoves." Journal of the Air and Waste Management Association, Volume 46, pp. 853 - 860.

Wilson, A.L., S.D. Colome, and Y. Tian. 1993. California Residential Indoor Air Quality Study, Volume 1, Methodology and Descriptive Statistics, Appendix. Integrated Environmental Services, Irvine, California.

SECTION 9

ESTIMATION OF HUMAN VENTILATION RATE

In pNEM-type models, each population unit (cohort or individual) is represented by an exposure event sequence derived from activity diary data (see Subsection 2.3). Each exposure event is typically defined by a start time, a duration, assignments to a geographic location and microenvironment, and an indication of activity level. Since 1998, the various pNEM models have defined activity level using the activity classification coding scheme incorporated into CHAD. A probabilistic module within each pNEM model converts the activity classification code of each exposure event to an energy expenditure rate, which in turn is converted into an estimate of oxygen uptake rate. The oxygen uptake rate is then converted into an estimate of ventilation rate (V_E), expressed in liters min^{-1} . In Version 2.1 of pNEM/CO, the oxygen uptake rate is also converted into an estimate of alveolar ventilation rate (V_A). V_A represents the portion of V_E that is involved in gaseous exchange with the blood and is also expressed in liters min^{-1} . Note that the estimates of V_E and V_A are both derived from oxygen uptake rate.

This section focuses primarily on the estimation of V_E with supplemental material concerning the estimation of V_A . Subsection 9.1 provides a brief overview of the physiological principles incorporated into the algorithms used in pNEM to convert each activity classification code to an oxygen uptake rate. Subsection 9.2 describes the additional steps required to convert oxygen uptake to V_E . Subsection 9.3 presents tests for identifying unrealistic values of oxygen uptake rate. Subsection 9.4 presents a probabilistic algorithm for estimating V_E that incorporates the physiological principles discussed in the preceding subsections.

Subsection 9.5 provides a discussion of the physiological principles relating V_A to oxygen uptake. It also presents a probabilistic algorithm for estimating V_A based on these principles that is consistent with the algorithm used to estimate V_E . Subsection 9.6 discusses limitations of the estimation algorithms that should be considered by researchers who plan to use the algorithms in other models.

9.1 Estimating Oxygen Uptake Rate

McCurdy (2000) has recommended that ventilation rate be estimated as a function of energy expenditure rate. The energy expended by an individual during a particular activity can be expressed as

$$EE = (\text{MET})(\text{RMR}) \quad (9-1)$$

in which EE is the average energy expenditure rate (kcal min^{-1}) during the activity and RMR is the resting metabolic rate of the individual expressed in terms of number of energy units expended per unit of time (kcal min^{-1}). MET (the “metabolic equivalent of work”) is a ratio specific to the activity and is dimensionless. If RMR is specified for an individual, then Equation 9-1 requires only an activity-specific estimate of MET to produce an estimate of the energy expenditure rate for a given activity. McCurdy, Glen, Smith, and Lakkadi (2000) have developed distributions of MET for the activity

classifications appearing in the CHAD database (Appendix A). As discussed below, equations for estimating RMR as a function of body mass (BM) can be obtained from the literature for various combinations of age and gender.

For convenience, let $EE_a(i,j,k)$ indicate the energy expenditure rate associated with the i -th activity of day j for person k . Equation 9-1 can now be expressed as

$$EE_a(i,j,k) = [MET(i,j,k)][RMR(k)] \quad (9-2)$$

in which $RMR(k)$ is the average value for resting metabolic rate specific to person k . Note that $MET(i,j,k)$ is specific to a particular activity performed by person k .

Energy expenditure requires oxygen which is supplied by ventilation (respiration). Let $ECF(k)$ indicate an energy conversion factor defined as the volume of oxygen required to produce one kilocalorie of energy in person k . The oxygen uptake rate (VO_2) associated with a particular activity can be expressed as

$$VO_2(i,j,k) = [ECF(k)][EE_a(i,j,k)], \quad (9-3)$$

in which $VO_2(i,j,k)$ has units of liters oxygen min^{-1} , $ECF(k)$ has units of liters oxygen kcal^{-1} , and $EE(i,j,k)$ has units of kcal min^{-1} . The value of $VO_2(i,j,k)$ can now be determined from $MET(i,j,k)$ by substituting Equation 9-2 into Equation 9-3 to produce the relationship

$$VO_2(i,j,k) = [ECF(k)][MET(i,j,k)][RMR(k)]. \quad (9-4)$$

The program algorithm described in Subsection 9.4 provides estimates of $ECF(k)$ and $RMR(k)$ for person k consistent with the person's demographic characteristics as specified by the program user. Subsection 9.4 also identifies the sources of these estimates.

It is often advantageous to express oxygen uptake rate as liters oxygen min^{-1} per kg of body mass (BM). In this case, Equation 9-4 can be restated as

$$VO_2(i,j,k)/BM = [ECF(k)][MET(i,j,k)][RMR(k)]/BM. \quad (9-5)$$

This expression is used below to convert oxygen uptake rate to ventilation rate.

9.2 Estimating Ventilation Rate from Oxygen Uptake Rate

Ventilation rate (V_E) tends to increase as VO_2 increases up to the point of maximum oxygen uptake ($VO_{2\text{max}}$). The relationship is known to be non-linear, with the slope of the relationship usually increasing at higher values of VO_2 . The ratio of V_E to VO_2 at a particular value of VO_2 is referred to as the ventilatory equivalence ratio (VER). The maximum value of VER (VER_{max}) tends to occur at $VO_{2\text{max}}$.

In current pNEM models, the relationship between $V_E(i,j,k)$ and $VO_2(i,j,k)$ is modeled by the empirical equation

$$\ln[V_E(i,j,k)/BM(k)] = a + (b)\{\ln[VO_2(i,j,k)/BM(k)]\} + d(k) + e(i,j,k) \quad (9-6)$$

in which $V_E(i,j,k)$ is the V_E value associated with the i th event of day j for person k , $BM(k)$ is the body mass assigned to person k , and a and b are constants determined by the age and gender of person k according to Tables 9-1, 9-2, and 9-3. The term $d(k)$ is a random variable selected for each person from a normal distribution with mean equal to zero and standard deviation equal to σ_d . The term $e(i,j,k)$ is a random variable selected for each individual event from a normal distribution with mean equal to zero and standard deviation equal to σ_e . Tables 9-1 through 9-3 list values of σ_d and σ_e by age and gender. The value of $VO_2(i,j,k)/BM(k)$ is obtained from Equation 9-5.

The values of a , b , σ_d , and σ_e listed in Table 9-1 are specific to adults (18+ years) and were determined through a statistical analysis of data obtained from 32 clinical studies performed by Dr. William Adams (1998). During each study, a panel of subjects were put through a test sequence of increasing levels (steps) of exertion on either a cycle ergometer or a treadmill. At each step, researchers measured test time, machine setting, the subject's ventilation rate (V_E), the subject's oxygen uptake rate (VO_2), and the subject's heart rate (HR). In most cases, researchers also established a maximum value of VO_2 (VO_{2max}) for each sequence of steps. Subject-specific data on age, gender, body mass, lean body mass, height, and other relevant parameters were also recorded.

Table 9-1. Regression-Based Estimates for Parameters of Equation 9-6 Obtained from Statistical Analyses of Data Provided by Adams (1998).

Subject characteristics		Regression results ^a				
Age group	Gender	Number of test values	Intercept (a)	Slope (b)	Standard deviation	
					Person-level errors (σ_d)	Test-level errors (σ_e)
18 - 44	Female	1473	4.357	1.276	0.1351	0.1182
	Male	3145	3.991	1.197	0.1228	0.1395
45 - 64	Female	60	3.454	1.021	0.1106	0.0769
	Male	641	4.018	1.165	0.1107	0.1112
65+	Female	45	2.956	0.908	0.0886	0.0338
	Male	317	3.730	1.071	0.1082	0.0632

^aBased on regression analysis of data provided by Dr. William Adams. See Adams (1998) and Appendix B of the report by Johnson and Capel (2001) for details.

Johnson (1998) performed a series of statistical analyses on these data to determine whether the data would support the development of subject-specific equations relating ventilatory equivalence ratio (VER) -- the ratio of V_E to VO_2 -- to VO_2 expressed as a percentage of VO_{2max} . The data for most subjects were found to lack sufficient measurements of V_E at low values of VO_2 to properly "anchor" the lower end of

the assumed relationship, although the relationship between V_E and VO_2 at high exertion rates was usually well characterized. Supplemental analyses of the Adams data by EPA researchers suggested that a strong relationship existed between V_E/BM and VO_2/BM (where BM was the body mass of the subject providing the values of V_E and VO_2) when data from multiple subjects were combined according to age and gender.

Based on these findings, EPA elected to develop an equation relating V_E/BM to VO_2/BM applicable to each of six groups of adult subjects defined by the age and gender categories listed in Table 9-1. Dr. Jonathan Cohen of ICF Consulting used regression analysis to fit Equation 9-6 to cross-sectional data for the subjects belonging to each of the six groups. Appendix B of the report by Johnson and Capel (2001) describes these analyses in detail. The resulting regression coefficients appear as estimates of a , b , σ_d , and σ_e in Table 9-1.

Table 9-2. Estimates for Parameters of Equation 9-6 Applicable to Males Aged 0 through 17.

Age group, years	Average physiological characteristics for age group		Parameter Estimates for Equation 9-6			
	$(VO_{2max})/BM^a, l \text{ min}^{-1} \text{ kg}^{-1}$	VER_{max}	Intercept (a)	Slope (b)	Standard deviation	
					σ_d	σ_e
0 - 2	0.044	25.3	3.846	1.197	0.1228	0.1395
3	0.046	33.8	4.127	1.197	0.1228	0.1395
4	0.048	39.3	4.269	1.197	0.1228	0.1395
5	0.050	41.1	4.306	1.197	0.1228	0.1395
6	0.052	39.5	4.258	1.197	0.1228	0.1395
7	0.054	38.0	4.212	1.197	0.1228	0.1395
8	0.056	36.5	4.165	1.197	0.1228	0.1395
9	0.057	35.0	4.119	1.197	0.1228	0.1395
10	0.057	35.5	4.134	1.197	0.1228	0.1395
11	0.058	34.5	4.102	1.197	0.1228	0.1395
12	0.058	33.5	4.072	1.197	0.1228	0.1395
13	0.058	32.0	4.026	1.197	0.1228	0.1395
14	0.057	32.0	4.030	1.197	0.1228	0.1395
15	0.057	32.0	4.030	1.197	0.1228	0.1395
16	0.055	31.0	4.005	1.197	0.1228	0.1395
17	0.053	32.8	4.069	1.197	0.1228	0.1395

^aBM = body mass

Table 9-3. Estimates for Parameters of Equation 9-6 Applicable to Females Aged 0 through 17.

Age group, years	Average physiological characteristics for age group		Parameter Estimates for Equation 9-6			
	$(VO_{2max})/BM^a, l \text{ min}^{-1} \text{ kg}^{-1}$	VER_{max}	Intercept (a)	Slope (b)	Standard deviation	
					σ_d	σ_e
0 - 2	0.043	23.6	4.028	1.276	0.1351	0.1182
3	0.044	30.8	4.288	1.276	0.1351	0.1182
4	0.046	36.4	4.443	1.276	0.1351	0.1182
5	0.047	37.5	4.467	1.276	0.1351	0.1182
6	0.050	38.0	4.463	1.276	0.1351	0.1182
7	0.052	38.5	4.466	1.276	0.1351	0.1182
8	0.053	39.0	4.473	1.276	0.1351	0.1182
9	0.052	38.5	4.466	1.276	0.1351	0.1182
10	0.051	37.0	4.431	1.276	0.1351	0.1182
11	0.050	36.0	4.409	1.276	0.1351	0.1182
12	0.049	35.0	4.387	1.276	0.1351	0.1182
13	0.047	33.5	4.354	1.276	0.1351	0.1182
14	0.046	33.0	4.345	1.276	0.1351	0.1182
15	0.046	34.2	4.381	1.276	0.1351	0.1182
16	0.045	33.5	4.366	1.276	0.1351	0.1182
17	0.044	32.5	4.342	1.276	0.1351	0.1182

^aBM = body mass

The Adams database did not provide adequate clinical data for testing the applicability of Equation 9-6 to children. However, it seemed reasonable to assume that the shape and variability of the relationship between $\ln(V_E/BM)$ and $\ln(VO_2/BM)$ for children would be similar to that of young adults, especially as these two parameters have been normalized with respect to body mass. Consequently, researchers assumed that (1) Equation 9-6 could be applied to children and (2) that the gender-specific values listed for slope (b), σ_d , and σ_e in Table 9-1 for adults 18 to 44 could be applied to children aged 0 to 17.

In Tables 9-2 and 9-3, three of parameters in Equation 9-6 [slope (b), variance between persons (σ_d), and variances between activities (σ_e)] are considered to be constant for all ages under 18. Researchers assumed that the relationship between $\ln(V_E/BM)$ and $\ln(VO_2/BM)$ varied with age primarily with respect to the intercept (a)

parameter. The intercept estimates in Tables 9-2 and 9-3 were obtained by solving an equation which specified the relationship between $\ln(V_E/BM)$ and $\ln(VO_2/BM)$ when $VO_2 = VO_{2max}$. Ignoring the $d(k)$ and $e(i,j,k)$ terms (which have average values of zero), the relationship between V_E/BM and VO_2/BM can be expressed as

$$\ln(V_E/BM) = a + (b)[\ln(VO_2/BM)] \quad (9-7)$$

At VO_{2max} , the value of V_E is equal to $VER_{max} \times VO_{2max}$. Equation 9-7 can now be written as

$$\ln[(VER_{max} \times VO_{2max})/(BM)] = a + (b)[\ln(VO_{2max}/BM)]. \quad (9-8)$$

for conditions at VO_{2max} . This equation can be rearranged to yield

$$a = \ln(VER_{max}) + (1 - b)[\ln(VO_{2max}/BM)]. \quad (9-9)$$

Note that the value of the intercept term (a) is now entirely determined by VER_{max} and VO_{2max}/BM , parameters which are known to vary by age and gender. Tables 9-2 and 9-3 provide age-specific estimates of VER_{max} and VO_{2max}/BM for male and female children, respectively. The VER_{max} estimates for ages 5 through 17 were taken directly from the curves for males and females provided in Figure 7-10 of Astrand and Rodahl (1977). The ratios for children under 5 were determined by extending the male and female curves tangentially downward. The VO_{2max}/BM estimates for ages 6 through 17 were taken directly from curves provided by Astrand and Rodahl (1977) in Figure 9-13. The ratios for ages 0 to 5 were obtained by extending the curves downward. In all cases, the estimates of VER_{max} and VO_{2max}/BM are considered to be average values for the indicated gender and age group.

As indicated above, researchers assumed that $b = 1.197$ for male children and $b = 1.276$ for female children. These values were entered into Equation 9-9 together with the appropriate estimates of VER_{max} and VO_{2max}/BM to determine the estimates of the intercept term (a) listed in Tables 9-2 and 9-3. For example, the estimate for 17-year-old females ($a = 4.3421$) in Table 9-3 was determined by solving Equation 9-9 when $b = 1.276$, $VER_{max} = 32.5$, and $VO_{2max}/BM = 0.044$. It is reassuring to note that this estimate for females aged 17 years (4.3421) is similar to the estimate for females aged 18 to 44 years (4.3568) presented in Table 9-1.

Although there are uncertainties concerning the various assumptions made in estimating the parameters of Equation 9-5 (a , b , σ_d , and σ_e) when applied to children, the approach described above is expected to provide reasonable values for $\ln(V_E/BM)$ as $\ln(VO_2/BM)$ approaches $\ln(VO_{2max}/BM)$, as each proposed curve is anchored to realistic values at this point. The uncertainty associated with estimates of $\ln(V_E/BM)$ for children increases as $\ln(VO_2/BM)$ approaches its minimum value, which corresponds to a very low exertion rate. *All other things being equal, analysts tend to accept more uncertainty relating to estimates of ventilation rate occurring at low exertion rates than to estimates of ventilation rate at high exertion rates.*

In summary, Equation 9-4 was used to convert event-specific values of MET to corresponding values of VO_2 in adults and children. Equation 9-6 was then used to convert the VO_2 value to a value of V_E . Subsection 9.3 describes two tests for

reasonableness that were performed on the VO_2 estimates prior to converting them to V_E values. Subsection 9.4 presents a step-by-step description of the algorithm that was used to probabilistically determine the various parameter values required by Equations 9-4 and 9-6.

9.3 Tests to Identify Unrealistic Values of Oxygen Uptake Rate

A person's maximum ventilation rate is determined by his or her maximum oxygen uptake rate (VO_{2max}) and the V_E/VO_2 ratio in effect under maximum oxygen uptake conditions. As work increases, energy is provided primarily by aerobic (oxygen-based) processes up to the point of VO_{2max} , referred to as the point of maximal aerobic power (MAP). The additional energy required for higher work rates is provided primarily by anaerobic processes. Consequently, the work rate where VO_{2max} is reached is less than a person's maximum work rate.

Astrand and Rodahl (1977) state that most individuals cannot maintain a work rate equal to 100 percent of MAP (i.e., a work rate where VO_2 equals VO_{2max}) for more than about five minutes. As the duration of work increases, there is a progressive decrease in the average VO_2 level that can be maintained. Astrand and Rodahl also state that a VO_2 level equal to 50 percent of VO_{2max} cannot be maintained for a whole working day.

Erb (1981) provide estimates of the percentage of "maximum work capacity" that can be maintained by young and middle-aged adults for durations of one to nine hours (Table 9-4). These values -- which apply to normally active, non-trained adults -- appear to be functionally equivalent to the percentage of VO_{2max} (designated $PCTVO_{2max}$) that can be maintained for the indicated time period and are so labeled in Table 9-4. According to Erb, a person can maintain 64 percent of VO_{2max} for one hour and 33 percent of VO_{2max} for nine hours without straining.

The following expression provides a close fit to values Erb proposed for durations of one to nine hours:

$$PCTVO_{2max}(t) = 121.2 - (14.0)[\ln(t)]. \quad (9-10)$$

Note that t is duration in minutes and \ln indicates the natural (base e) logarithm. Equation 9-10 provides an estimate of approximately 100 percent for $t = 5$ minutes, consistent with the statement by Astrand and Rodahl (1977) that 100 percent of VO_{2max} can be maintained for up to 5 minutes. These findings suggest that it is reasonable to assume that (1) $PCTVO_{2max}$ should not exceed 100 percent for events with durations between 0 and 5 minutes, (2) Equation 9-10 can be used to determine the upper limit of $PCTVO_{2max}$ for events of durations between 5 minutes and 540 minutes (9 hours), and (3) the $PCTVO_{2max}$ values in Table 9-4 can be used as upper limits for VO_2 averaged over multi-hour periods from one to nine hours in duration. A conservative assumption (i.e., one which may permit unrealistically high VO_2 values) is that the value for nine hours (33 percent) applies to longer time periods.

The concepts discussed above were the basis of two tests applied to VO_2 estimates to ensure their reasonableness. The first test was applied to VO_2 values associated with individual activity events. The second test was applied to running-average values of VO_2 with durations of 1 to 24 hours.

Table 9-4. Values of the Upper Limit of PCTVO_{2max} for Specified Averaging Times.

Averaging time (hours)	Upper limit of PCTVO _{2max} , percent
1	64
2	54
3	48
4	44
5	41
6	39
7	37
8	35
9	33
10 to 24	33 ^a

^aConservative estimate based on nine-hour value proposed by Erb (1981). All other values in this column are identical to values proposed by Erb (1981).

The first test was based on the assumption that the PCTVO_{2max} value associated with an individual (event-specific) VO₂ value could not exceed an upper limit calculated by Equation 9-10. This test was carried out by comparing each event-specific VO₂ value generated for a person with a permitted upper limit (PUL) value obtained from the following equation.

$$\text{Permitted upper limit of VO}_2 = (\text{Upper limit of PCTVO}_{2\text{max}})(\text{VO}_{2\text{max}})/100. \quad (9-11)$$

When the event duration ranged between 5 minutes and nine hours, the value for the upper limit of PCTVO_{2max} in Equation 9-11 was obtained from Equation 9-10 using the value of VO_{2max} assigned to the person by the physiological profile generator. Outside this range, PCTVO_{2max} was assumed to equal 100 percent for durations less than 5 minutes and to equal 33 percent for durations greater than nine hours. If the VO₂ value exceeded the PUL determined by Equation 9-10, the VO₂ value was set equal to the calculated PUL. Otherwise, the value of VO₂ was not affected by Test 1.

The second test assumed that “running-average” VO₂ values (expressed as a percentage of the VO_{2max} value) could not exceed the values specified by Erb in Table 9-4. This test was implemented by first averaging the event-specific VO₂ values generated for a cohort by clock hour to produce 24 one-hour VO₂ values. Running-average VO₂ values for all possible sequential periods of 1 to 24 hours in duration were then calculated from these one-hour values. Each running-average VO₂ value was used to calculate the value

$$\text{Test Ratio} = (\text{running average } VO_2)/(\text{permitted upper limit of } VO_2) \quad (9-12)$$

in which the “permitted upper limit” (PUL) is a value specific to the indicated averaging time. If any test ratio exceeded 1.0, then all event-specific VO_2 values for the person-day were proportionally reduced so that the largest test ratio equaled exactly 1.0.

Equation 9-11 was used to determine PULs for running-average VO_2 values. In this application, the upper limit of $PCTVO_{2max}$ was obtained from Table 9-4 according to the period of the running average.

The value for VO_{2max} required by the two tests was determined by the equation

$$VO_{2max} = (NVO_{2max})(BM), \quad (9-13)$$

in which BM was body mass in kg and NVO_{2max} was maximum oxygen uptake rate per kg of body mass. As discussed in the next section, the values of NVO_{2max} and BM were taken from the physiological profile generated by the program to represent the cohort.

9.4 A Probabilistic Algorithm for Estimating V_E

Recent versions of pNEM have included a probabilistic algorithm for estimating ventilation rate which incorporates the physiological principles discussed above. The most up-to-date version of the algorithm has been incorporated into the “ V_E Estimator Program”, a stand-alone PC-based program (Johnson and Capel, 2001). The program includes a sorter module that selects 24-hour activity patterns from CHAD based on demographic specifications provided by the user. The user then has the option of constructing multi-day activity patterns through the use of a special sequencer module. Each activity pattern (24-hour or multi-day) is composed of activity events that are functionally equivalent to the exposure events produced by pNEM models.

Tables 9-5 and 9-6 present the probabilistic algorithm for estimating ventilation rate as it was implemented in the V_E Estimator Program together with the parameter values required by the algorithm. Table 9-7 lists the principal parameters appearing in the algorithm and indicates the functional form and source of data for each parameter.

To run the model, the algorithm requires distributions for BM, NVO_{2max} , and ECF specific to age and gender. Table 9-8 lists distributions for BM by age and gender based on articles by Brainard and Burmaster (1992) and by Burmaster and Crouch (1997). Table 9-9 lists distributions for NVO_{2max} obtained through a review of the literature. Table 9-10 lists distributions of ECF based on data provided by Esmail, Bhambhani, and Brintnell (1995). Analysts reviewed these data and selected the most appropriate distribution for each parameter for each combination of gender and age ($0 < \text{age} < 100$ years). These distributions (listed in Appendix C) were incorporated into the ventilation rate algorithm.

The ventilation rate algorithm also requires an equation for estimating RMR for each combination of age and gender. Analysts reviewed a list of equations previously compiled by McCurdy (1998) and determined that a set of equations developed by Schofield (1985) provided good coverage of all age and gender combinations. These equations were determined through regression analyses and have the functional form

Table 9-5. Probabilistic Method for Generating Estimates of Ventilation Rates as a

Function of Energy Expenditure Rate.

1. User specifies characteristics of demographic group containing person k. Sorter and sequencer modules select one or more person-days from CHAD master database to represent person k.

Result: sequence of person-days representing person k.
2. Based on demographic characteristics specified in Step 1, perform algorithms in Table 9-6 to determine physiological profile of person k.

Result: gender(k), age(k) in years, BM(k) in kg, ECF(k), RMR(k) in kcal min⁻¹, NVO_{2max}(k) in liters O₂ min⁻¹ kg⁻¹, VO_{2max}(k) in liters O₂ min⁻¹, and regression parameters for Equation 9-6 (a, b, σ_d, and σ_e).
3. Go to first/next day. Day = j.
4. Go to first/next activity event. Event = i. Read activity classification code for event. Determine appropriate parameters of MET distribution from Appendix A based on activity code, age, and occupation (if applicable). Randomly select a value of MET for event i from distribution.

Result: MET(i,j,k)
5. Note ECF(k) and RMR(k) from Step 2. Based on MET(i,j,k) from Step 4, calculate oxygen uptake of event as

$$VO_2(i,j,k) = [ECF(k)][MET(i,j,k)][RMR(k)].$$

Result: VO₂(i,j,k) in liters O₂ min⁻¹
6. Determine upper limit of PCTVO_{2max}(i,j,k) by Equation 9-10. If duration is less than 5 minutes, upper limit of PCTVO_{2max}(i,j,k) equals 100 percent. If duration is greater than 540 minutes (9 hours), upper limit of PCTVO_{2max}(i,j,k) equals 33 percent.

Result: upper limit of PCTVO_{2max}(i,j,k) in percent.
7. Test 1: Determine the permitted upper limit for VO₂ by Equation 9-11 using the value of PCTVO_{2max}(i,j,k) determined in Step 6 and VO_{2max}(k) from Step 2. If VO₂ for event exceeds upper limit, set VO₂ equal to the upper limit.

Result: adjusted VO₂(i,j,k) in liters O₂ min⁻¹
8. If last activity event of day, go to Step 9. Otherwise, go to Step 4.
9. Test 2: Average event-specific VO₂ values for day by clock hour to produce 24 one-hour VO₂ values. Calculate running-average VO₂ values for all possible sequences of 1 to 24 hours in duration from these one-hour values. Obtain PCTVO_{2max} value for each duration from Table 9-4. Use Equation 9-11 to calculate permitted upper limit of VO₂ for each duration. Use Equation 9-12 to determine test ratio for each running-average VO₂ value. If any test ratio exceeds 1.0, reduce all event-specific VO₂ values proportionally so that largest test ratio of resulting adjusted VO₂ values equals exactly 1.0.

Result: adjusted VO₂(i,j,k) in liters O₂ min⁻¹ for each event of day.
10. Following completion of Test 2 (Step 9), use parameter values obtained in Step 2 to convert each event-specific VO₂ value of day from Step 9 to a corresponding V_E value through the use of the following relationship:

$$\ln[V_E(i,j,k)/BM(k)] = a + (b)\{\ln[VO_2(i,j,k)/BM(k)]\} + d(k) + e(i,j,k)$$

Result: V_E(i,j,k) estimate for each event of day.
11. If last day, go to Step 12. Otherwise, go to Step 2 if generating individual person-days or go to Step 3 if creating a year-long sequence.
12. End.

Table 9-6. Probabilistic Algorithms Used to Determine Physiological Parameters Associated with Person k.

1.	<p>User specifies characteristics of demographic group containing person k. Based on specified characteristics and table of population statistics, randomly determine gender and age of person k to be simulated.</p> <p>Result: age(k) in years and gender(k).</p>
2.	<p>Based on gender(k) and age(k) from Step 1, determine appropriate parameters in Appendix C for distribution of body mass. Randomly select BM(k) from distribution.</p> <p>Result: BM(k) in kg.</p>
3.	<p>Determine appropriate parameters in Appendix C for distribution of energy conversion factor (ECF) for person k. Randomly select ECF(k) from distribution.</p> <p>Result: ECF(k).</p>
4.	<p>Based on gender(k) and age(k) from Step 1, determine appropriate regression parameters (a, b, σ_e) in Appendix C for estimating resting metabolic rate (RMR). Using BM(k) from Step 2, determine RMR for person k by formula</p> $\text{RMR}(k) = (0.166)\{a + (b)[\text{BM}(k)] + e\}$ <p>in which 0.166 converts MJ day⁻¹ to kcal min⁻¹ and e is a randomly selected value from a normal distribution with mean equal to zero and standard deviation equal to σ_e.</p> <p>Result: RMR(k) in kcal min⁻¹.</p>
5.	<p>Based on gender(k) and age(k) from Step 1, determine appropriate parameters from Appendix C for distribution of normalized maximum oxygen uptake rate (NVO_{2max}) of person k. Randomly select NVO_{2max}(k) from distribution. Divide by 1000 to convert from ml O₂ min⁻¹ kg⁻¹ to liters O₂ min⁻¹ kg⁻¹.</p> <p>Result: NVO_{2max}(k) in liters O₂ min⁻¹ kg⁻¹.</p>
6.	<p>Based on BM(k) from Step 2 and NVO_{2max}(k) from Step 5, determine maximum oxygen uptake rate (VO_{2max}) for person k by equation</p> $\text{VO}_{2\text{max}}(k) = [\text{NVO}_{2\text{max}}(k)][\text{BM}(k)].$ <p>Result: VO_{2max}(k) in liters O₂ min⁻¹.</p>
7.	<p>Based on gender(k) and age(k) from Step 1, determine appropriate regression parameters (a, b, σ_a, σ_e) in Table 9-1, 9-2, or 9-3 for use in Equation 9-5.</p> <p>Result: parameter values for Equation 9-6.</p>

Table 9-7. Principal Parameters Used in Probabilistic Algorithm for Estimating Ventilation Rates (Tables 9-5 and 9-6).

Parameter	Abbreviation	Functional Form	Source of Data
Body mass	BM	Lognormal distribution	Brainard and Burmaster, 1992; Burmaster and Crouch, 1997 (see Table 9-8)
Energy Conversion Factor	ECF	Point estimate	Esmail et al., 1995 (see Table 9-10)
Metabolic Equivalence	MET	Distribution specified in CHAD Database	McCurdy, 1998 (see Appendix A)
Resting metabolic rate	RMR	Regression equations specific to age and gender	Schofield, 1985, as compiled by McCurdy, 1998 (see Table 9-11)
Normalized oxygen uptake rate	NVO _{2max}	Normal distribution	Research summarized in Table 9-9

Table 9-8. Parameters for Lognormal Distributions Fitted to Body Mass (BM) Data in Kilograms Organized by Age and Gender.

Age, years	Males				Females			
	ln(BM)		BM		ln(BM)		BM	
	μ	σ	GM	GSD	μ	σ	GM	GSD
0.5 - 1	2.23	0.132	9.3	1.141	2.16	0.145	8.7	1.156
1	2.46	0.119	11.7	1.126	2.38	0.128	10.8	1.137
2	2.60	0.120	13.5	1.127	2.56	0.112	12.9	1.119
3	2.75	0.114	15.6	1.121	2.69	0.137	14.7	1.147
4	2.87	0.133	17.6	1.142	2.83	0.133	16.9	1.142
5	2.99	0.138	19.9	1.148	2.98	0.163	19.7	1.177
6	3.13	0.145	22.9	1.156	3.10	0.174	22.2	1.190
7	3.21	0.151	24.8	1.163	3.19	0.174	24.3	1.190
8	3.33	0.181	27.9	1.198	3.31	0.156	27.4	1.169
9	3.43	0.165	30.9	1.179	3.46	0.214	31.8	1.239
10	3.59	0.195	36.2	1.215	3.57	0.199	35.5	1.220
11	3.69	0.252	40.0	1.287	3.71	0.226	40.9	1.254
12	3.78	0.224	43.8	1.251	3.82	0.213	45.6	1.237
13	3.88	0.215	48.4	1.240	3.92	0.216	50.4	1.241
14	4.02	0.181	55.7	1.198	3.99	0.187	54.1	1.206
15	4.09	0.159	59.7	1.172	4.00	0.156	54.6	1.169
16	4.20	0.168	66.7	1.183	4.06	0.167	58.0	1.182
17	4.19	0.167	66.0	1.182	4.08	0.165	59.1	1.179
18	4.25	0.159	70.1	1.172	4.07	0.147	58.6	1.158
19	4.26	0.154	70.8	1.166	4.10	0.149	60.3	1.161
18 - 74 ^a	4.34	0.17	76.7	1.19	4.17	0.20	64.7	1.22

^aDerived from Brainard and Burmaster (1992). All other statistics derived from Burmaster and Crouch (1997).

Table 9-9. Descriptive Statistics for VO₂ and V_E Measured at Maximal Exertion by Various Researchers.

Population group	n	VO _{2max} , liters/min			NVO _{2max} , ml/min per kg			Maximum ratio of V _E to VO ₂			Source
		Mean	S.D.	C.V. ^a	Mean	S.D.	C.V.	Mean	S.D.	C.V.	
Females, 20-29	8	2.23	0.26	0.12	39.9	4.7	0.12	34.1	6.0	0.18	Åstrand (1960)
Females, 30-39	12	2.13	0.28	0.13	37.3	5.2	0.14	35.2	5.7	0.16	
Females, 40-49	8	2.01	0.19	0.09	32.5	2.7	0.08	31.8	4.7	0.15	
Females, 50-65	16	1.85	0.25	0.14	28.4	2.7	0.10	33.1	4.0	0.12	
Females, 20-25	32	2.88	0.24	0.08	48.4	2.8	0.06	32.3	2.8	0.09	
Males, 20-29	4	4.19	NR	NR	52.2	NR	NR	31.8	NR	NR	
Males, 30-39	13	3.01	0.54	0.18	39.8	7.3	0.18	34.6	6.0	0.17	
Males, 40-49	9	2.99	0.32	0.11	39.2	5.5	0.14	33.8	5.3	0.16	
Males, 50-59	66	2.54	0.36	0.14	33.1	4.9	0.15	26.9	4.6	0.17	
Males, 60-69	8	2.23	0.29	0.13	31.4	5.3	0.17	33.1	5.8	0.18	
Males, 20-33	29	4.16	0.39	0.09	58.6	4.5	0.08	29.4	3.0	0.10	Mercier, Varray, Ramonatxo, et. al. (1991)
Males, 11	18	1.65	0.30	0.18	45.4	8.06	0.18	37.4	7.00	0.19	
Males, 12	15	1.85	0.31	0.17	47.4	8.13	0.17	35.0	5.42	0.15	
Males, 13	15	2.26	0.39	0.17	46.0	6.97	0.15	31.0	3.02	0.10	
Males, 14	15	2.62	1.12	0.43	45.7	4.26	0.09	33.2	7.28	0.22	
Males, 15	13	2.70	0.50	0.19	47.5	4.69	0.10	34.4	5.05	0.15	Katch and Park (1975)
Males, 21-27	13	3.91	0.52	0.13	54.5	7.61	0.14	NR	NR	NR	
Males and Females, 20-29	80	3.09	0.83	0.27	45.3	7.54	0.17	NR	NR	NR	Heil, Freedson, Ahliquist, et. al. (1995)
Males and Females, 30-39	81	3.19	0.86	0.27	43.8	8.15	0.19	NR	NR	NR	

Population group	n	VO _{2max} , liters/min			NVO _{2max} , ml/min per kg			Maximum ratio of V _E to VO ₂			Source
		Mean	S.D.	C.V. ^a	Mean	S.D.	C.V.	Mean	S.D.	C.V.	
Males and Females, 40-49	79	3.13	0.92	0.29	42.9	9.04	0.21	NR	NR	NR	Heil, Freedson, Ahlquist, et. al. (1995)
Males and Females, 50-59	78	2.84	0.91	0.32	36.8	8.93	0.24	NR	NR	NR	
Males and Females, 60-69	74	2.31	0.72	0.31	30.7	7.98	0.26	NR	NR	NR	
Males and Females, 70-79	47	1.91	0.56	0.29	27.2	5.67	0.21	NR	NR	NR	
Males, 20-79	210	3.54	0.71	0.20	44.0	9.42	0.21	NR	NR	NR	
Females, 20-79	229	2.14	0.51	0.24	33.8	8.65	0.26	NR	NR	NR	
Males, 10-17	6	NR	NR	NR	46.6	5.8	0.12	NR	NR	NR	Mermier, Samet, Lambert, et. al. (1993)
Males, 18-72	15	NR	NR	NR	45.7	16.7	0.37	NR	NR	NR	
Females, 7-17	6	NR	NR	NR	38.0	5.0	0.13	NR	NR	NR	
Females, 21-72	16	NR	NR	NR	32.2	8.9	0.28	NR	NR	NR	
Male, 23-33	20	NR	NR	NR	48.3	4.9	0.10	NR	NR	NR	Rowland, Auchinachie, Keenan, et. al. (1987)
Male, 9-13	20	NR	NR	NR	57.9	6.9	0.12	NR	NR	NR	

^aC.V. = (std. dev.)/(mean), dimensionless.

Table 9-10. Estimates of the Energy Conversion Factor (ECF) Based on Data in Esmail, Bhambhani, and Brintnell (1995).

Group	Number of subjects	Test	Mean VO ₂ , liters min ⁻¹	Mean GEC ^a , kcal min ⁻¹	Ratio of means ^b
Women	20	wheel-turn	0.81	4.1	0.198
		push-pull	0.80	4.0	0.200
		overhead-reach	0.87	4.4	0.198
Men	Not reported	wheel-turn	1.13	5.7	0.198
		push-pull	1.16	5.6	0.207
		overhead-reach	1.13	5.7	0.198

^aGEC: gross energy cost.

^bData were not available for calculating the mean of subject-specific ratios.

$$\text{RMR} = a + (b)(\text{BM}) + e, \quad (9-17)$$

in which e is assumed to be normally distributed with mean = zero and standard deviation = σ_e . Table 9-11 lists Schofield's values of a , b , and σ_e for 12 age/gender combinations. These values are the basis of the RMR equations listed in Appendix C which have been incorporated into the ventilation rate algorithm.

9.5 Estimation of Alveolar Ventilation Rate (V_A)

Alveolar ventilation (V_A) represents the portion of the minute ventilation that is involved in gaseous exchange with the blood. VO_2 is the oxygen uptake that occurs during this exchange. The absolute value of V_A is known to be affected by total lung volume, lung dead space, and respiration frequency -- parameters which vary according to person and/or exercise rate. However, it is reasonable to assume that the ratio of V_A to VO_2 is relatively constant regardless of a person's physiological characteristics or energy expenditure rate. Consistent with this assumption, Version 2.1 of pNEM/CO converted each estimate of $VO_2(i,j,k)$ to an estimate of $V_A(i,j,k)$ by the proportional relationship

$$V_A(i,j,k) = (19.63)[VO_2(i,j,k)] \quad (9-18)$$

in which both V_A and VO_2 are expressed in units of liters min^{-1} . This relationship was obtained from an article by Joumard, Chiron, Vidon, et al. (1981), who based it on research by Galetti (1959). Equation 9-18 was applied to all population units (cohorts) under all energy expenditure rates.

The V_A algorithm included a method for identifying "impossible" values which were occasionally generated by the estimation process. This method determined a maximum VO_2 value for each exposure event which accounted for the duration of the activity and for the age, weight, and gender of the person. No estimate of VO_2 (and the corresponding estimate of V_A) was permitted to exceed this limit. Subsection 9.4 provides a more detailed description of this procedure.

In summary, Equation 9-4 was used to convert event-specific values of MET to corresponding values of VO_2 . Equation 9-18 was then used to convert the VO_2 value to a value of V_A . Table 9-12 presents a probabilistic algorithm for estimating V_A that is identical to the algorithm in Table 9-5 with one major exception. VO_2 is converted to V_A in Step 10 of Table 9-12, whereas VO_2 is converted to V_E in Table 9-5.

9.6 Limitations of the Algorithm

In a typical application of the V_E Estimator Program, each activity (exposure) event is characterized by a numerical code indicating the activity associated with each event (e.g., 11220 = "clean house"). Section 9.4 describes a probabilistic algorithm that can be used to convert the activity code for each event to a corresponding estimate of V_E for the event. Table 9-13 lists all parameters appearing in this algorithm, including the parameters provided as input to the algorithm (labeled "input variables") and the parameters that are estimated as functions of other parameter values (labeled "intermediate results"). The V_E Estimator Program provides a library of values for each

Table 9-11. Regression Equations for Predicting Basal Metabolic Rate Provided by Schofield (1985) as Compiled by McCurdy (1998).

Gender	Age, years	Regression coefficients ^a		
		a	b	σ_e
Female	< 3	-0.130	0.244	0.25
Female	3 - 9.9	2.033	0.085	0.29
Female	10 - 17.9	2.898	0.056	0.47
Female	18 - 29.9	2.036	0.062	0.50
Female	30 - 59.9	3.538	0.034	0.47
Female	≥ 60	2.755	0.038	0.45
Male	< 3	-0.127	0.244	0.29
Male	3 - 9.9	2.110	0.095	0.28
Male	10 - 17.9	2.754	0.074	0.44
Male	18 - 29.9	2.896	0.063	0.64
Male	30 - 59.9	3.653	0.048	0.70
Male	≥ 60	2.459	0.049	0.69

^aRegression equation: $BMR(MJ/day) = a + (b)(BM) + e$, σ_e = standard deviation of e. Basal metabolic rate (BMR) is assumed to be equivalent to resting metabolic rate (RMR).

Table 9-12. Probabilistic Method for Generating Estimates of Ventilation Rates as a Function of Energy Expenditure Rate.

1.	User specifies characteristics of demographic group containing person k. Sorter and sequencer modules select one or more person-days from CHAD master database to represent person k. Result: sequence of person-days representing person k.
2.	Based on demographic characteristics specified in Step 1, perform algorithms in Table 9-6 to determine physiological profile of person k. Result: gender(k), age(k) in years, BM(k) in kg, ECF(k), RMR(k) in kcal min ⁻¹ , NVO _{2max} (k) in liters O ₂ min ⁻¹ kg ⁻¹ , and VO _{2max} (k) in liters O ₂ min ⁻¹ .
3.	Go to first/next day. Day = j.
4.	Go to first/next activity event. Event = i. Read activity classification code for event. Determine appropriate parameters of MET distribution from Appendix A based on activity code, age, and occupation (if applicable). Randomly select a value of MET for event i from distribution. Result: MET(i,j,k)
5.	Note ECF(k) and RMR(k) from Step 2. Based on MET(i,j,k) from Step 4, calculate oxygen uptake of event as $VO_2(i,j,k) = [ECF(k)][MET(i,j,k)][RMR(k)].$ Result: VO ₂ (i,j,k) in liters O ₂ min ⁻¹
6.	Determine upper limit of PCTVO _{2max} (i,j,k) by Equation 9-10. If duration is less than 5 minutes, upper limit of PCTVO _{2max} (i,j,k) equals 100 percent. If duration is greater than 540 minutes (9 hours), upper limit of PCTVO _{2max} (i,j,k) equals 33 percent. Result: upper limit of PCTVO _{2max} (i,j,k) in percent.
7.	<u>Test 1:</u> Determine the permitted upper limit for VO ₂ by Equation 9-11 using the value of PCTVO _{2max} (i,j,k) determined in Step 6 and VO _{2max} (k) from Step 2. If VO ₂ for event exceeds upper limit, set VO ₂ equal to the upper limit. Result: adjusted VO ₂ (i,j,k) in liters O ₂ min ⁻¹
8.	If last activity event of day, go to Step 9. Otherwise, go to Step 4.
9.	<u>Test 2:</u> Average event-specific VO ₂ values for day by clock hour to produce 24 one-hour VO ₂ values. Calculate running-average VO ₂ values for all possible sequences of 1 to 24 hours in duration from these one-hour values. Obtain PCTVO _{2max} value for each duration from Table 9-4. Use Equation 9-11 to calculate permitted upper limit of VO ₂ for each duration. Use Equation 9-12 to determine test ratio for each running-average VO ₂ value. If any test ratio exceeds 1.0, reduce all event-specific VO ₂ values proportionally so that largest test ratio of resulting adjusted VO ₂ values equals exactly 1.0. Result: adjusted VO ₂ (i,j,k) in liters O ₂ min ⁻¹ for each event of day.
10.	Following completion of Test 2 (Step 9), convert each event-specific VO ₂ value of day from Step 9 to a corresponding V _A value through the use of the following relationship: $V_A(i,j,k) = (19.63)[VO_2(i,j,k)]$ Result: V _A (i,j,k) estimate for each event of day.
11.	If last day, go to Step 12. Otherwise, go to Step 2 if generating individual person-days or go to Step 3 if creating a year-long sequence.
12.	End.

Table 9-13. Parameters Appearing in Algorithms Used to Convert Activity Codes to Estimates of Ventilation Rate.

Parameters	Definition	Type	Values	Comments
RMR regression coefficients (a, b, and σ_e)	Coefficients of Equation 9-17	Input variable	Pages C-22 through C-27 (also Table 9-11)	Used to estimate RMR as a function of BM, classified by gender and age
EE	Average energy expenditure rate for activity	Intermediate result	Estimated by Equation 9-2	Function of MET and RMR
RMR	Resting metabolic rate	Intermediate result	Estimated by Equation 9-11	Function of BM, a, b, and σ_e
MET	Metabolic equivalent of work: ratio of energy expenditure rate to resting metabolic rate	Input variable	Appendix A	Classified by activity code, age, and occupation (if applicable).
VO ₂	Oxygen uptake rate	Intermediate result	Estimated by Equation 9-4	Function of ECF, MET, and RMR
ECF	Energy conversion factor: ratio of VO ₂ to EE	Input variable	Pages C- 16 through C-21	Classified by age and gender
BM	Body mass	Input variable	Pages C-10 through C-15 (also Table 9-8)	Classified by age and gender
VER regression coefficients (a, b, σ_d , σ_e)	Coefficients of Equation 9-6	Input variable	Tables 9-1, 9-2, and 9-3	Used to estimate relationship between V _E , VO ₂ , and BM in Equation 9-6; classified by age and gender
NVO _{2max}	Normalized oxygen uptake rate (ratio of VO ₂ to BM)	Input variable	Pages C-4 through C-9 (also Table 9-9)	Classified by gender and age
VO _{2max}	Maximum oxygen uptake rate	Intermediate result	Estimated by equation in Step 6 (Table 9-6)	Function of NVO _{2max} and BM
VER _{max}	Maximum ratio of V _E to VO ₂	Input variable		
VER	Ratio of V _E to VO ₂	Intermediate result	Estimated by Equation 9-6	

input parameter in the form of distributions or point estimates, with most parameters being represented by distributions. A particular parameter may be represented by multiple distributions that vary with the activity code of the event or the demographic characteristics of the cohort. For example, the program provides 42 distributions for body mass values that vary with age and gender.

Most of the steps in converting the activity code for a particular event to a V_E value employ equations and parameter values that are relatively well-supported by clinical data (see Subsection 9.4). Perhaps the weakest link in the algorithm is the step which requires the analyst to provide a distribution of possible MET values for each activity code. These distributions (Appendix A) are currently based on distributions provided by the developers of CHAD (McCurdy et al., 2000). Because available data were often insufficient to accurately define a distribution for each activity code, the developers tended to follow a conservative approach and over-estimate the variability of each distribution. Consequently, the V_E values produced by the ventilation rate algorithm may exhibit an excessive degree of variability. To prevent the occurrence of “impossible” values arising from this variability, the estimator program includes test routines that identify and adjust VO_2 values which exceed limits based on activity duration and the physiological characteristics of the cohort. Note that these test routines also require parameter values (e.g., VER_{max}) that are themselves subject to a certain degree of uncertainty.

The V_E value estimated for a particular activity event is not explicitly affected by the ventilation rates estimated for preceding events. Consequently, the algorithm may not adequately account for excess post-exercise oxygen consumption (EPOC), a condition experienced when individuals are engaged in strenuous exertion that results in an oxygen debt that impacts oxygen uptake and ventilation rates after cessation of the strenuous exercise. Glen (1999) has developed a method for adjusting MET values to account for EPOC. McCurdy et al. (2000) show the application of this method to a partial person-day of activity data. The authors of this report believe that the method proposed by Glen should be evaluated further before it is incorporated into algorithms for estimating V_E or V_A .

9.7 References for Section 9

Adams, W. C. 1998. Letter to Tom McCurdy, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. August 21.

Astrand, I. 1960. “Aerobic Work Capacity in Men and Women with Special Reference to Age. Acta. Physiol. Scand., Volume 49, Supplement 169.

Astrand, P. O. and K. Rodahl, 1977. Textbook of Work Physiology. 2nd ed. McGraw-Hill, New York, New York.

Brainard, J, and D. Burmaster. 1992. “Bivariate Distributions for Height and Weight of Men and Women in the United States.” Risk Analysis, Vol. 12, No. 2, pp. 267 - 275.

Burmaster, D. E., and E. A. C. Crouch. 1997. “Lognormal Distributions of Body Weight

as a Function of Age for Males and Females in the United States, 1976 - 1980," Risk Analysis, Vol. 17, No. 4.

Erb, B. D., 1981. "Applying Work Physiology to Occupational Medicine," Occupational Health Safety, Vol. 50, pp. 20-24.

Esmail, Bhambhani, and Brintnell. 1995. "Gender Differences in Work Performance on the Baltimore Therapeutic Equipment Work Simulator." Amer. J. Occup. Therapy, Volume 49, pp. 405 - 411.

Galetti, P. M. "Respiratory Exchanges During Muscular Effort." Helv. Physiol. Acta. Vol. 17, pp. 34 - 61 (1959).

Glen, G. 1999. Revisions to the Metabolic Rate (MET) Values and Excess Post-Exercise Oxygen Consumption (EPOC) Adjustments in the CHAD Database. ManTech Environmental Technology, Inc., Research Triangle Park, North Carolina.

Heil, D. P., P. S. Freedson, L. E. Ahlquist, J. Price, and J. M. Rippe. 1995. "Nonexercise Regression Models to Estimate Peak Oxygen Consumption." Med. Sci. Sports Exer., Volume 27, pp. 599 - 606.

Johnson, T. 1998. "Analysis of Clinical Data Provided by Dr. William Adams and Revisions to Proposed Probabilistic Algorithm for Estimating Ventilation Rate in the 1998 Version of pNEM/CO." Memorandum submitted to the U.S. Environmental Protection Agency under EPA Contract No. 68-D6-0064. TRJ Environmental, Inc.

Johnson, T., and J. Capel. 2001. User's Guide. Software for Estimating Ventilation (Respiration) Rates for Use in Dosimetry Models. Report prepared for National Center for Environmental Assessment, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Joumard, R, M. Chiron, R. Vidon, M. Maurin, and J-M Rouzioux. 1981. "Mathematical Models of the Uptake of Carbon Monoxide on Hemoglobin at Low Carbon Monoxide Levels." Environmental Health Perspectives. Vol. 41, pp. 277 - 289 (October 1981).

Katch, V. L., and M. W. Park. 1975. "Minute-by-Minute Oxygen Requirement and Work Efficiency for Constant-Load Exercise of Increasing Duration." Res. Q., Volume 46, pp. 38 - 47.

McArdle, W. D., F. I. Katch, and V. L. Katch, 1991. Exercise Physiology: Energy, Nutrition, and Human Performance. Lea and Febiger, Malvern, Pennsylvania.

McCurdy, T. 1998. Personal Communication. Information on distributions available upon request to author. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, North Carolina 27711.

McCurdy, T., 2000. "Conceptual Basis for Multi-Route Intake Dose Modeling Using an

Energy Expenditure Approach.” Journal of Exposure Analysis and Environmental Epidemiology, Volume 10, pp. 1 - 12.

McCurdy, T., G. Glen, L. Smith, and Y. Lakkadi. 2000. “The National Exposure Research Laboratory’s Consolidated Human Activity Database.” Journal of Exposure Analysis and Environmental Epidemiology, Volume 10, Number 6, pp. 566 - 578.

Mercier, J., A. Varray, M. Ramonatxo, B. Mercier, and C. Préfaut. 1991. “Influence of Anthropomorphic Characteristics on Changes in Maximal Exercise Ventilation and Breathing Pattern During Growth in Boys.” Eur. J. Appl. Physiol., Volume 63, pp. 235 - 241.

Mermier, J., J. M. Samet, W. E. Lambert, and T. W. Chick. 1993. “Evaluation of the Relationship Between Heart Rate and Ventilation for Epidemiologic Studies.” Arch. Environ. Health, Volume 48, pp. 263 - 269.

Rowland, T. W., J. A. Auchinachie, T. J. Keenan, and G. M. Green. 1987. “Physiologic Responses to Treadmill Running in Adult and Prepubertal Males.” Inter. J. Sports Med., Volume 8, pp. 292 - 297.

Schofield, W. N. 1985. “Predicting Basal Metabolic Rate, New Standards, and Review of Previous Work.” Hum. Nutr. Clin. Nutr., Volume 39C, Supplement 1, pp. 5 - 41.

SECTION 10

ESTIMATION OF CARBOXYHEMOGLOBIN LEVELS

This section describes the probabilistic COHb module included in Version 2.1 of pNEM/CO. The material is taken from Appendix E of the report by Johnson, Mihlan, LaPointe et al. (2000). The approach is based primarily on an earlier COHb module described by Biller and Richmond in an Appendix to the report describing the 1992 version of pNEM/CO (Johnson, Capel, Paul, and Wijnberg, 1992).

10.1 The Coburn-Forster-Kane Model

The COHB module in the original CO-NEM model (Johnson and Paul, 1983) used as its basic model the differential equation derived by Coburn, Forster, and Kane (1965) which described the dynamic relationship between instantaneous blood levels of COHb, inspired CO, and other physiological variables. This model, which will be referred to here as the CFK model, continues to be the most widely-used method for estimating COHb and is the basic model for COHb computations in pNEM/CO. The mathematical structure of the CFK model is described in Subsection 10.2.

The CFK model describes the rate of change of COHb blood levels as a function of the following quantities:

1. Inspired CO pressure
2. COHb level
3. Oxyhemoglobin (O_2Hb) level
4. Hemoglobin (Hb) content of blood
5. Blood volume
6. Alveolar ventilation rate
7. Endogenous CO production rate
8. Mean pulmonary capillary oxygen pressure
9. Pulmonary diffusion rate of CO
10. Haldane coefficient (M)
11. Barometric pressure
12. Vapor pressure of water at body temperature (47 torr)

If all of the listed quantities except COHb level are constant over some time interval, the CFK equation has a linear form over the interval and is readily integrated. The solution to the linear form gives reasonably accurate results for lower levels of COHb. However, CO and oxygen compete for the available hemoglobin and therefore are not independent of each other. If this dependency is taken into account, the resulting differential equation is no longer linear. Peterson and Stewart (1975) proposed a heuristic approach to account for with this dependency which assumed the linear form and then adjusted the O_2Hb level iteratively based on the assumption of a linear relationship between COHB and O_2Hb . This approach was used in the COHb module of the original CO-NEM. Alternatively, it is possible to determine COHb at any time by

numerical integration of the nonlinear CFK equation (e.g. by use of the Runge-Kutta method) if one assumes a particular relationship between COHb and O₂Hb. Muller and Barton (1987) demonstrated that assuming a linear relationship between COHb and O₂Hb leads to a form of the CFK equation equivalent to the Michaelis-Menton kinetic model that can be integrated. However, the analytical solution in this case cannot be solved explicitly for COHb. Muller and Barton demonstrated a binary search method for determining the COHb value.

The COHb module developed for pNEM/CO employs a linear relationship between COHb and O₂Hb that is consistent with the basic assumptions of the CFK model but differs from the linear forms used by other modelers. The Muller and Barton (1987) solution is employed. However, instead of the simple binary search described in the Muller and Barton paper, a combination of the binary search and Newton-Raphson root finding methods was used to solve for COHb (Press et al., 1986). Using the Muller and Barton solution increased computation time compared to the Peterson-Stewart method but was shown to be faster than fourth order Runge-Kutta numerical integration.

10.2 Mathematical Structure of the CFK Model Used in pNEM/CO

Table 10-1 defines the variables which appear in the equations of this section. Coburn, Forster, and Kane (1965) derived the following differential equation governing COHb levels in the blood upon exposure to CO.

$$\frac{d[COHb]}{dt} = \frac{\dot{V}_{CO}}{V_b} + \frac{P_{Ico}}{BV_b} - \frac{\bar{P}_{CO_2}[COHb]}{MBV_b[O_2Hb]} \quad (10-1)$$

where

$$B = \frac{1}{D_{Lco}} + \frac{(P_B - P_{H_2O})}{\dot{V}_A} \quad (10-2)$$

If the only quantity in this equation that can vary with time is [COHb], the CFK equation is linear and can be readily integrated. However, since oxygen (O₂) and CO compete for the available HB, [COHb] and [O₂Hb] must be related. Increasing [COHb] will result in decreasing [O₂Hb]. Thus the CFK equation is not linear and requires the relationship between the two quantities to be known if it is to be accurately integrated over a wide range of COHb levels.

Various linear relationships between [COHb] and [O₂Hb] have been used (see Marcus, 1980; McCartney, 1990; Muller and Barton, 1987; and Tikuisis et al., 1987). A relationship not previously used follows directly from the basic assumptions of the CFK model. The CFK model employs the Haldane coefficient, which is the equilibrium constant associated with the following reaction representing the replacement of O₂ in O₂Hb by CO:



Table 10-1. Definitions Of CFK Model Variables.

Variable	Definition
t	Time from start of an exposure event, min
[COHb]	Concentration of carboxyhemoglobin (COHb) in blood at time, t, ml CO per ml blood at STPD
[O ₂ Hb]	Concentration of oxyhemoglobin (O ₂ Hb) in blood at time t, ml O ₂ per ml blood at STPD
{RHb}	Concentration of reduced hemoglobin in blood as equivalent ml CO per ml of blood at STPD
[COHb] ₀	%[COHb] at t = 0
[THb] ₀	[RHb] + [COHb] + [O ₂ Hb]
%[COHb]	[COHb] expressed as percent of [RHb] ₀
%[O ₂ Hb]	[O ₂ Hb] expressed as percent of [RHb] ₀
%[COHb] ₀	%[COHb] at t = 0
%[COHb] _∞	%[COHb] at t = ∞
P ₁₀₀	Pressure of inspired CO in air saturated with water vapor at body temperature, torr
$\bar{P}_{C_{CO}}$	Mean pulmonary capillary CO pressure, torr
$\bar{P}_{C_{O_2}}$	Mean pulmonary capillary O ₂ pressure, torr
P _B	Barometric pressure, torr
P _{H₂O}	Vapor pressure of water at body temperature, torr (47 torr)
\dot{V}_A	Alveolar ventilation rate, ml/min STPD
\dot{V}_{CO}	Endogenous CO production rate, ml/min STPD
D _{L_{CO}}	Pulmonary CO diffusion rate, ml/min/torr STPD
M	Haldane coefficient
k	Equilibrium constant for reaction O ₂ + RHb = O ₂ Hb
V _b	Blood volume, ml
Hb	Total hemoglobin in blood, g/100ml
%MetHb	Methemoglobin as weight percent of Hb

The following equation, the Haldane relationship, applies approximately at equilibrium conditions.

$$\frac{\bar{P}_{CO_2}[COHb]}{\bar{P}_{CO}[O_2Hb]} = M \quad (10-4)$$

The Haldane coefficient, M, is the chemical equilibrium constant for reaction (3)

The above reaction can also be viewed as the difference between two competing chemical reactions:



Subtracting Equation 10-6 from Equation 10-5 yields Equation 10-3. If Equation 10-3 is in equilibrium, then Equations 10-5 and 10-6 are in equilibrium. If k is the equilibrium constant for Equation 10-6 then:

$$\frac{[O_2Hb]}{\bar{P}_{CO_2}[RHb]} = k \quad (10-7)$$

It is known that an individual breathing air free of CO for an extended period will have about 97% of the reactive hemoglobin tied up as O₂Hb and the rest (3%) as RHb. It is also known that at one atmosphere barometric pressure the mean pulmonary capillary oxygen pressure is approximately 100 torr. Substituting into Equation 10-7 yields 0.32 as the approximate value of k at body temperature. From mass balance considerations:

$$[O_2Hb] + [COHb] + [RHb] = [THb]_o \quad (10-8)$$

Eliminating [RHb] between Equations 10-7 and 10-8 and solving for [O₂Hb] yields:

$$[O_2Hb] = \frac{k\bar{P}_{CO_2}}{1 + k\bar{P}_{CO_2}} ([THb]_o - [COHb]) \quad (10-9)$$

This equation is the desired linear relationship. It has the same form as a relationship given without explanation by McCartney (1990), but replaces the constant in the McCartney equation by the term in Equation 10-9 involving the mean pulmonary capillary oxygen pressure and the equilibrium constant k. Substituting Equation 10-9 into Equation 10-1 yields a CFK equation free of [O₂Hb] and fully consistent with Coburn, Forster, and Kane's original derivation.

$$\frac{d[COHb]}{dt} = \frac{\dot{V}_{CO}}{\dot{V}_b} + \frac{P_{I_{CO}}}{BV_b} - \frac{[COHb]}{[THb]_o - [COHb]} \frac{1 + k\bar{P}_{CO_2}}{kMBV_b} \quad (10-10)$$

In working with the CFK model it is convenient to express COHb as a percent of [RHb]₀. Multiplying Equation 10-10 by 100 and dividing by [RHb]₀:

$$\frac{d\%[COHb]}{dt} = \frac{100}{[THb]_o} \left(\frac{\dot{V}_{CO}}{V_b} + \frac{P_{I_{CO}}}{BV_b} \right) - \frac{\%[COHb]}{100 - \%[COHb]} \frac{100(1 + k\bar{P}_{c_{O_2}})}{k[RHb]_o MBV_b} \quad (10-11)$$

Equation 10-11 can be written in the form suggested by Muller and Barton (1987):

$$\frac{d\%[COHb]}{dt} = C_o - C_1 \frac{\%[COHb]}{100 - \%[COHb]} \quad (10-12)$$

where

$$C_o = \frac{100}{[THb]_o} \left(\frac{\dot{V}_{CO}}{V_b} + \frac{P_{I_{CO}}}{BV_b} \right) \quad (10-13)$$

$$C_1 = \frac{100(1 + k\bar{P}_{c_{O_2}})}{k[THb]_o MBV_b} \quad (10-14)$$

Given values for the atmospheric pressure and the physiological variables in Equations 10-12 through 10-14, the value of % [COHb] at time t can be found by numerical integration using such techniques as the fourth order Runge-Kutta method (Press et al., 1986).

Muller and Barton (1987) demonstrated that an equation of the form of Equation 10-12 is equivalent to a Michaelis-Menton kinetics model which can be integrated. Integration yields:

$$-(C_o + C_1)t + \%[COHb] - \%[COHb]_0 - (100 - \%[COHb]_\infty) \ln \frac{(\%[COHb]_\infty - \%[COHb])}{(\%[COHb]_\infty - \%[COHb]_0)} = 0 \quad (10-15)$$

The equation for % [COHb]_∞ is obtained by setting equation (10-12) equal to zero and solving for % [COHb] which is now equal to % [COHb]_∞:

$$\%[COHb]_\infty = \frac{100C_o}{(C_o + C_1)} \quad (10-16)$$

Equation 10-15 cannot be solved explicitly for % [COHb]. The Muller and Barton paper

suggests the binary search method as one way to find the value of %[COHb]. Press and coauthors (1986) contend a combination of the binary search and Newton-Raphson methods is faster on average.

10.3 Implementation of the CFK Model in pNEM/CO

10.3.1 The COHb Module

As described in Subsection 2.3, Version 2.1 of pNEM/CO processes time/activity data derived from CHAD to create an exposure event sequence for each cohort. Each event in the sequence specifies a start time, duration, activity, microenvironment, and home/work status. Algorithms within pNEM/CO use this information to estimate an average CO concentration and alveolar ventilation rate (V_A) for each exposure event. A profile generator algorithm assigns physiological attributes to the cohort consistent with its demographic characteristics. The COHb module processes these data according to the COHb model described above to produce an estimate of COHb level at the end of each exposure event. The COHb estimates calculated for individual cohorts are then weighted according to the cohort populations to produce COHb distributions for larger population groups of particular interest.

The COHb module employs the Muller and Barton (1987) integration of the CFK model as represented by Equations 10-12 through 10-14 to compute the COHb level of a cohort at the end of each exposure event. To perform this computation, the COHb module requires information on each of the quantities listed in the section describing the CFK model. In addition, the COHb level at the beginning of the exposure event must be known. This latter quantity is usually the COHb level computed at the end of the previous contiguous exposure event. To obtain the initial COHb at the start of the exposure period, the computation is started one day before the beginning of the period. The effect of the initial COHb value on the end value is negligible after about 15 hours. The program stores the COHb levels at the end of each clock hour and outputs distributions of COHb levels for the sensitive population.

10.3.2 Overview of Input Data Requirements

Subsection 10.4 describes the equations and procedures used by the pNEM/CO COHb module to obtain the values of the input variables for Equations 10-2 and 10-13 through 10-16. A brief overview is given here.

The actual inspired CO level can change significantly during an exposure event. The model supplies an average exposure concentration for the event, which is used as the CO input. The time constant for the change in COHb is sufficiently large that the use of concentrations based on averaging times up to one hour can be used in place of the instantaneous concentrations over the averaging time period with little loss of accuracy in estimating the COHb level at the end of the exposure event. Furthermore, applying the average concentrations to a contiguous sequence of exposure events does not cause an accumulation of error.

The COHb model presently used in pNEM/CO does not account for changing barometric pressure. It uses a constant barometric pressure which is a function of the average elevation of an area above sea level. The pressure at sea level is taken to be

760 torr.

The remaining input variables to the CFK model are all physiological parameters. While the Haldane coefficient, the equilibrium constant k , and average pulmonary capillary oxygen pressure are treated as having the same constant values for all cohorts, the remaining physiological input variables will vary among individuals. Subsection 10.4 describes the methods used to generate the various physiological input variables for each combination of cohort and calendar day processed by pNEM/CO.

10.4 Computation of Input Data for the COHb Module

As discussed in the previous section and in Subsections 2.43 through 2.45, the algorithms used to estimate V_A and COHb require values for various physiological parameters such as body mass, blood volume, and pulmonary diffusion rate. Table 10-2 provides a complete list of these parameters. A special algorithm within pNEM/CO probabilistically generates a value for each parameter on the list (collectively referred to as a “physiological profile”) for each combination of cohort and calendar day processed by pNEM/CO. Figure 10-1 is a flow diagram showing the process by which each physiological profile was generated. Each of the generated physiological profiles is internally consistent, in that the functional relationships among the various parameters are maintained. For example, blood volume is determined as a function of weight and height, where height is estimated as a function of weight. Weight in turn is selected from a distribution specific to gender and age. Table 10-2 provides a brief summary of the method used to estimate values for each parameter in the application of pNEM/CO to Denver.

For each cohort, as defined above, pNEM/CO computes exposure for a contiguous sequence of exposure events spanning the total time period of the computation. This multi-day sequence of exposure events is determined by random sampling day-long event sequences from a set of pools of 24-hour activity patterns. An individual 24-hour pattern in one of these pools is referred to as a unit exposure sequence (UES). Each pool consists of a collection of UESs that are specific to the cohort demographic group, day type, and average daily temperature (see Subection 2.3).

Each UES is a contiguous set of exposure events spanning 24 hours. As noted previously, each event is characterized by start time, duration in minutes, home/work status, microenvironment, and activity. Each exposure event is constrained to occur entirely within a clock hour.

The CFK model within the COHb module is called for each exposure event. For each event it requires the following data.

- Time duration of event, min
- Inspired CO partial pressure averaged over the event, torr
- Percent COHb at the start of the event
- Alveolar ventilation rate, ml/min STPD
- Average pulmonary capillary oxygen pressure, torr
- Haldane Coefficient

Table 10-2. Parameters Included in Physiological Profile for Adults in Version 2.1 of pNEM/CO.

Parameter	Algorithm(s) Containing Parameter	Other Parameters Required for Calculating Parameter	Method Used to Estimate Parameter Value
Age	COHb Ventilation rate	Demographic group	Randomly selected from population-weighted distribution specific to demographic group
Gender	COHb Ventilation rate	Demographic group	Randomly selected from population-weighted distribution specific to demographic group
Weight (body mass)	COHb Ventilation rate	Gender Age	Randomly selected from population-weighted distribution specific to age and gender based on Brainard and Burmaster (1992).
Height	COHb	Weight Gender	<p>Estimated by the following equations:</p> <p>males: height = 34.43 inches + (6.67)[ln(weight)] + (2.38 inches)(z) females: height = 48.07 inches + (3.07)[ln(weight)] + (2.48 inches)(z)</p> <p>The z term was randomly selected from a unit normal [N(0,1)] distribution. Units: height (inches), weight (lbs).</p> <p>The estimation equations are based on the results of a statistical analysis by Johnson (1998) of height and weight data provided by Brainard and Burmaster (1992).</p>
Menstrual phase	COHb	Gender Age	<p>If gender = female, menstrual phase was randomly assigned in alternating 14-day cycles according to the following age-specific probabilities.</p> <p>Age < 12 or >50: 100% premenstrual Age 12 through 50: 50% premenstrual, 50% postmenstrual.</p>

Parameter	Algorithm(s) Containing Parameter	Other Parameters Required for Calculating Parameter	Method Used to Estimate Parameter Value
Blood volume	COHb	Gender Weight Height	<p>Blood volume (V_b) was determined according to gender by the following equations which are based on work by Allen et al. (1956) which was modified to accept the units used for height and weight.</p> <p>Men: $V_b = (20.4)(\text{weight}) + (0.00683)(H^3) - 30$ Women: $V_b = (14.6)(\text{weight}) + (0.00678)(H^3) - 30$</p> <p>Units: blood volume (ml), weight (lbs), height (inches).</p>
Hemoglobin content of the blood, Hb	COHb	Gender Age	<p>Randomly selected from normal distribution with arithmetic mean (AM) and arithmetic standard deviation (ASD) determined by gender and age based on data obtained from the 1976-1980 NHANES study (USDHHS, 1982) as follows.</p> <p>Males, 18 - 44: AM = 15.3, ASD = 1.0 Males, 45 - 64: AM = 15.1, ASD = 1.2 Males, 65+: AM = 14.8, ASD = 1.4 Females, 18 - 44: AM = 13.3, ASD = 1.1 Females, 45 - 64: AM = 13.6, ASD = 1.2 Females, 65+: AM = 13.7, ASD = 1.2</p> <p>Units: grams of Hb per deciliter of blood</p>

Parameter	Algorithm(s) Containing Parameter	Other Parameters Required for Calculating Parameter	Method Used to Estimate Parameter Value
Pulmonary CO diffusion rate, DL_{CO}	COHb	Gender Height Age	<p>Pulmonary CO diffusion rate (DL) was determined according to gender, height, and age according to the following equations obtained from a paper by Salorinne (1976) and modified to conform to the units used in the COHb module.</p> <p>Males: $DL_{CO} = (0.361)(\text{height}) - (0.232)(\text{age}) + 16.3 \text{ ml/min/torr}$</p> <p>Females: $DL_{CO} = (0.556)(\text{height}) - (0.115)(\text{age}) - 5.97 \text{ ml/min/torr}$</p> <p>Units: DL_{CO} (ml/min/torr), height (inches), age (years).</p>
Endogenous CO production rate	COHb	Gender Age Menstrual phase	<p>Endogenous CO production rate was randomly selected from a lognormal distribution with geometric mean (GM) and geometric standard deviation (GSD) determined according to the following equations specific to age, gender, and menstrual phase.</p> <p>Males, 18 - 64: GM = 0.473, GSD = 1.316 Males, 65+: GM = 0.473, GSD = 1.316 Females, 18 - 64, premenstrual: GM = 0.497, GSD = 1.459 Females, 18 - 64, postmenstrual: GM = 0.311, GSD = 1.459 Females, 65+: GM = 0.497, GSD = 1.459</p> <p>Units: GM (ml/hr), GSD (dimensionless).</p>
Resting metabolic rate (RMR)	Ventilation rate	Gender Age Weight (body mass)	See Subsection 9.4.
Energy conversion factor (ECF)	Ventilation rate	Gender	See Subsection 9.4.

Parameter	Algorithm(s) Containing Parameter	Other Parameters Required for Calculating Parameter	Method Used to Estimate Parameter Value
NVO _{2max}	Ventilation rate	Gender Age	See Subsection 9.4.
VO _{2max}	Ventilation rate	NVO _{2max} Weight (body mass)	See Subsection 9.4.

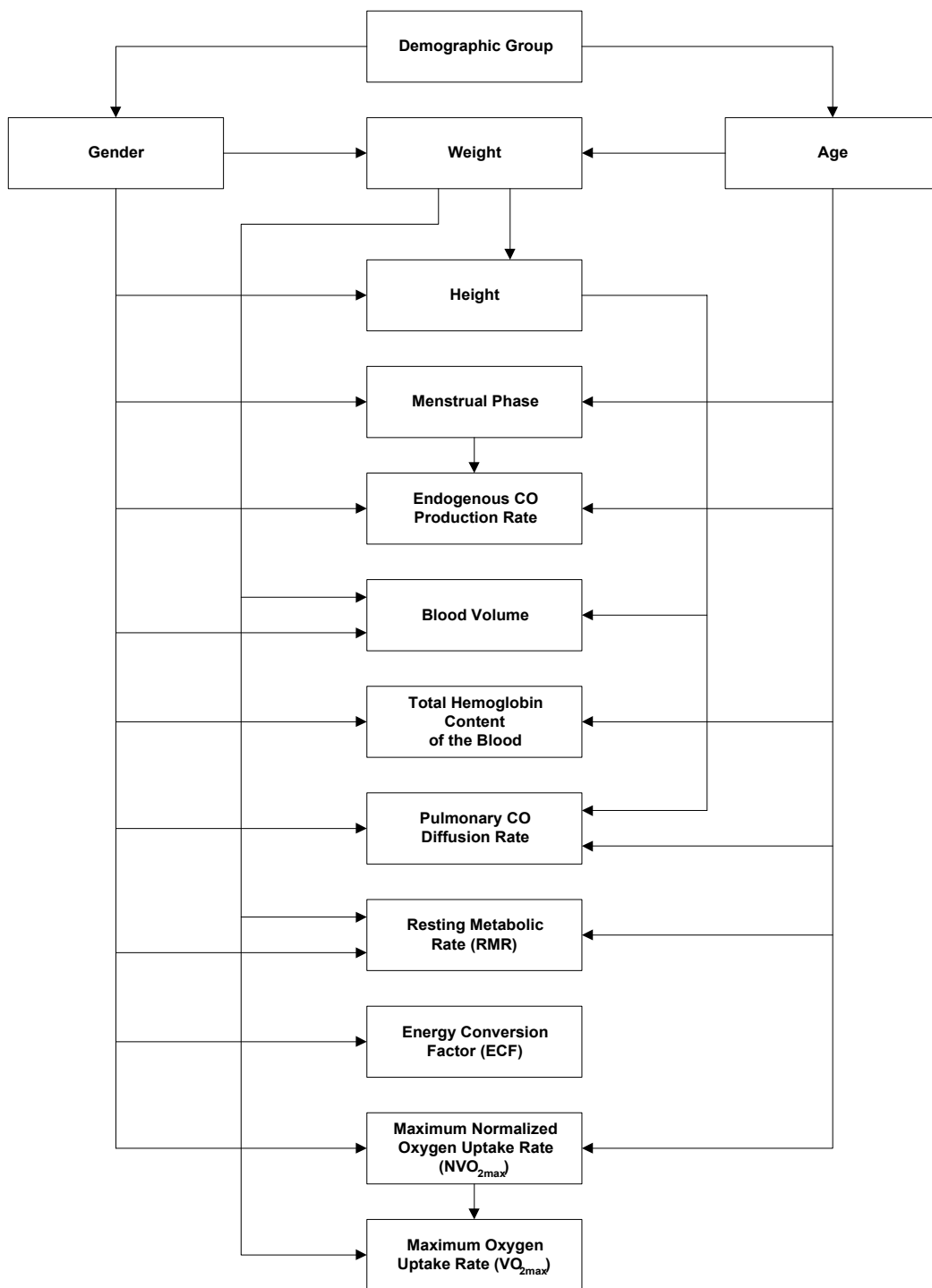


Figure 2-1. Flow Diagram for Physiological Profile Generator. Equilibrium constant for the reaction of O₂

Atmospheric pressure, torr
 Blood volume, ml
 Total potential reduced hemoglobin content of blood, ml CO/ml STPD
 Pulmonary CO diffusion rate, ml/min/torr STPD
 Endogenous CO production rate, ml/min STPD

Given these data as inputs, the module computes the percent COHb at the end of the exposure event. This value is used by the module as the initial percent COHb for the next contiguous exposure event. The main program retains only those COHb values at the end of each clock hour.

Some of the above data do not change during a pNEM/CO computer run and, therefore, need to be supplied to the computer program only once at the start of the run. Some data items vary with the cohort and therefore need to be supplied at the beginning of each activity day. Other data tend to change with the exposure event and therefore need to be supplied for each new exposure event.

10.4.1 Input Data Supplied at Start of the pNEM/CO Computation

Barometric Pressure

A constant barometric pressure is assumed for the study area based on the average height above sea level:

$$P_B = 760 * \exp(-0.0000386 * \textit{Altitude}) \quad (10-17)$$

where altitude is the average height (in feet) of the study area above sea level (USEPA, 1978). The altitude was set at 5183 feet for Denver and 328 feet for Los Angeles.

Average Pulmonary Capillary Oxygen Pressure

The equation employed is based on an approximation used by Peterson and Stewart (1975) in which the 49 torr is subtracted from the partial pressure of inspired oxygen. This leads to the following approximate relationship:

$$\bar{P}_{C_{O_2}} = 0.209(P_B - 47) - 49 \quad (10-18)$$

The constant 0.209 is the mole fraction of O₂ in dry air. The constant 47 is the vapor pressure of water at body temperature. This expression was used in an investigation of the CFK equation by Tikuisis et al. (1987). Modelers have tended to use the value 100 torr. Equation 10-18 gives the value 100 torr for a barometric pressure of 760 torr.

Haldane Coefficient

Measured values in the range 210 to 270 have been reported for the Haldane Coefficient in the literature. Modelers have tended to use values in the range 210 to 240. In the early 1980's, the Clean Air Scientific Advisory Committee (CASAC)

expressed the opinion to EPA (Friedlander, 1982) that the most careful work done in this area was that by Rodkey (1969), who determined a value of 218. This value was used in the COHb module of the earlier CO NEM version. Other modelers using values in the range 218 to 220 are Peterson and Stewart, 1970; Marcus, 1980; Collier and Goldsmith, 1983; and Muller and Barton, 1987. As the value 218 falls within the range currently used by modelers, EPA analysts have decided to continue using this value in Version 2.1 of pNEM/CO.

Equilibrium Constant for the Reaction of O₂ and RHb

Using methods described in Subsection 10.2, researchers estimated the value of this quantity to be 0.32 based on (1) the observation that %[RHb] is about 3% in individuals breathing air which is free of CO and (2) an assumed value of 100 torr for $\bar{P}_{C_{O_2}}$.

Total Reduced Hemoglobin in the Absence of O₂ and CO

The quantity [THb]₀ is expressed as equivalent milliliters of O₂ or CO at STPD per milliliter of blood. Total Hb blood levels are customarily expressed as grams per deciliter of blood. The total Hb level in the absence of COHb and O₂Hb would consist principally of RHb which can react with O₂ or CO and MetHb which cannot. Total Hb blood levels also tend to be higher in people living at higher altitudes. To relate [THb]₀ to Hb, it is therefore necessary to correct for the MetHb present, adjust for the effect of altitude, and convert to equivalent milliliters of CO at STPD. The later conversion is based on the observation that a gram of reduced Hb can react with a maximum of 1.39 ml of O₂ or CO at STPD. The application of these three factors yields the equation:

$$[THb]_0 = 1.39 * Hb(100 - \% MetHb) * \frac{(1 + HbAlt * Altitude)}{100} \quad (10-19)$$

where HbAlt is a regression constant. Hb in Equation 10-19 is a sea level value. Hb level in a human population is normally distributed with the mean Hb and standard deviation both dependent on gender and age class (see entry in Table 10-2 for the distributions of Hb by age and gender). Given the hemoglobin content of the blood based on the distributions listed in Table 10-2, [THb]₀ is calculated using Equation 10-19. The weight percent MetHB, %MetHB, is taken to be 0.5% of the weight of Hb (Muller and Barton, 1987).

The altitude correction factor, HbAlt, was developed by application of simple regression analyses to Hb data obtained in 17 U.S. cities (USEPA, 1973).

Men:	0.000161	S.E. = 0.000064
Women:	0.000115	S.E. = 0.000043

Two cities (Phoenix and Houston) were eliminated in the regression analysis because the measured Hb levels were substantially below that of the other cities. The altitude factor is small. It predicts about a 5% increase in Hb for residents of Denver over that

for people living at sea level.

Determination of Weight

Body mass or weight (in kg) was determined by fitting lognormal distributions to data organized by age and gender based on work by Brainard and Burmaster (1992) and Burmaster and Crouch (1997). Table 9-7 summarizes the parameters for the lognormal distributions obtained.

Determination of Height

The following equations were used to estimate height as a function of gender and weight.

$$\text{males: } \text{height} = 34.43 \text{ inches} + (6.67)[\ln(\text{weight})] + (2.38 \text{ inches})(z) \quad (10-20)$$

$$\text{females: } \text{height} = 48.07 \text{ inches} + (3.07)[\ln(\text{weight})] + (2.48 \text{ inches})(z) \quad (10-21)$$

The z term was randomly selected from a unit normal [N(0,1)] distribution. Equations 20 and 21 are based on the results of a statistical analysis by Johnson (1998) of height and weight data provided by Brainard and Burmaster (1992).

Base Pulmonary Diffusion Rate of CO

A base lung diffusivity of CO for the cohort is calculated as follows:

males:

$$D_{L_{CO}} = 0.556 * \text{height} - 0.115 * \text{age} - 5.97 \quad (10-22)$$

females:

$$D_{L_{CO}} = 0.361 * \text{height} - 0.232 * \text{age} = 16.3 \quad (10-23)$$

where height is in inches and age is in years.

The regression equations were obtained from a paper by Salorinne (1976) and modified to conform to the units used in the COHb module. The Salorinne data were obtained for non-exercising individuals. Tikuisis et al. (1992), working with eleven male subjects at various exercise levels, showed significant increase in lung diffusivity of CO with increasing alveolar ventilation rate. Regression analyses of data provided by Tikuisis for the individual subjects in the study showed the relationship to be linear. From this relationship and the heights and ages of the subjects in the Tikuisis et al. study, it was determined that the Salorinne equations for male subjects correspond to an alveolar ventilation rate of 6.69 l/min STPD. In the absence of other data it is assumed that this same value applies to women. Thus, for each twenty-four hour

period Equations 10-26 and 10-27 are used to compute lung diffusion rates of CO for a base case alveolar ventilation rate of 6.69 l/min STPD. As will be seen, this value is adjusted to account for the actual ventilation rate experienced by the cohort during each individual exposure event.

Endogenous Rate of CO Production

The endogenous CO production rates taken from a number of sources (Table 10-3) show the rate to be distributed lognormally in the population. The distribution is different for men and women. For a woman there is a further difference depending on whether she is in her premenstrual or postmenstrual phase. Table 10-2 presents these distributions classified by class, gender, and menstrual phase.

For each male cohort, pNEM/CO specifies a single value for endogenous CO production rate and uses it for all days of the year. For each female cohort between 18 and 64 years of age, pNEM/CO specifies one value of endogenous CO production rate to represent premenstrual days and one value to represent postmenstrual days. Females cohorts under 12 years and older than 50 are assumed to be premenstrual; consequently, pNEM/CO specifies a single value for endogenous CO production rate to be used for all days of the year. The specified values are randomly selected from the appropriate distributions presented in Table 10-2. A random number, z , is sampled from the standardized normal distribution, $N(0,1)$ to make each selection. The appropriate endogenous CO production rate is then obtained from:

$$\dot{V}_{CO} = 0.01667 * (\text{geom.mean}) * (\text{geom.S.D.})^z \quad (10-24)$$

The constant term converts ml/hr to ml/min.

A probabilistic algorithm within pNEM/CO assigns a menstrual phase to each day of the year for females cohorts aged 12 to 50 years. The algorithm randomly assigns a number between 1 and 28 to January 1. The number is increased by one for each successive day until number 28 is reached. The next day is numbered 1 and the 28-day numbering cycle is repeated until each day of the year has been assigned a number between 1 and 28. Days numbered 1 through 14 are identified as post-menstrual days; days numbered 15 through 28 are identified as pre-menstrual days.

10.4.2 Input Data Supplied with Each Exposure Event

Duration of Exposure Event

The duration of the exposure event in minutes is supplied by the main program to the COHb module.

Partial Pressure of Inspired Carbon Monoxide

The main program supplies the inspired CO concentration averaged over the duration of the exposure expressed as ppm. This quantity is converted to pressure via:

Table 10-3. Literature Data Used to Derive Geometric Mean and Standard Deviation for Lognormal Distribution of Endogenous CO Production Rate

ENDOGENOUS CO PRODUCTION RATE FOR MEN	
V _{co} (ml/hr)	REFERENCE
0.35	Coburn et al., 1963
0.35	"
0.4	"
0.39	"
0.43	"
0.35	"
0.51	"
0.42	"
0.57	"
0.45	"
0.4	Lynch and Moede, 1972
0.81	"
0.26	"
0.65	"
0.51	"
0.62	"
0.44	"
0.43	Berk et al., 1974
0.58	"
0.52	"
0.59	"
0.8	"
0.72	"
0.54	"

ENDOGENOUS CO PRODUCTION RATE FOR MEN	
V_{CO} (ml/hr)	REFERENCE
0.45	Delivoria-Papadopoules et al., 1974
0.26	“
0.6	“
0.45	“
0.39	“
0.4	“
0.81	Brouillard et al., 1975
0.57	“
0.33	“
0.7	“
0.58	Coltman and Dudley, 1969
0.38	“
0.51	“
0.55	“
0.37	“
0.49	“
0.45	“
0.5	“
0.33	“
0.45	“
0.36	“
0.54	Werner and Lindahl, 1980
0.76	“
0.48	“
0.31	“
0.7	“

ENDOGENOUS CO PRODUCTION RATE FOR MEN	
V _{co} (ml/hr)	REFERENCE
0.36	“
0.65	“
0.38	Luomanmaki and Coburn, 1969
0.42	“
0.41	“
0.54	“
0.38	“
0.72	Lynch and Moede, 1972
0.37	“
0.23	“
0.33	“
0.42	“
0.44	“
0.29	“
0.48	“
0.57	Delivoria-Papadopoulos et al., 1974
0.54	“
0.72	“
0.99	“
0.48	“
0.53	“
0.43	“
0.64	Merke et al., 1975
0.86	“
0.35	“
0.52	“

ENDOGENOUS CO PRODUCTION RATE FOR MEN	
V _{co} (ml/hr)	REFERENCE
0.8	“
0.54	“
0.68	“
0.28	“
0.48	Lynch and Moede, 1972
0.23	“
0.25	“
0.2	“
0.22	“
0.15	“
0.21	“
0.23	Delivoria-Papadopoulos et al., 1974
0.51	“
0.34	“
0.41	“
0.26	“
0.16	“
0.3	“
0.4	Merke et al., 1975
0.47	“
0.23	“
0.24	“
0.55	“
0.32	“
0.43	“
0.35	“

$$P_{I_{CO}} = (CO) * (P_b - 47) * 10^{-6} \quad (10-25)$$

Initial Percent COHb Level at Start of Exposure Event

The program retains the percent COHb computed at the end of the previous exposure event and uses this value as the initial percent COHb for the present event. The starting COHb at the beginning of an activity day is the final COHb level at the end of the preceding activity day. This latter procedure is used for the first activity day of the overall computation since the program starts the day before the overall period covered by the pNEM/CO computation.

Alveolar Ventilation Rate

The main program supplies the COHb module with ventilation rate derived from the algorithm discussed in Subsections 2.4.2 and 9.5.

Adjusted Pulmonary Diffusion Rate of CO

Given the alveolar ventilation rate for the exposure event the associated adjusted pulmonary diffusion rate can be calculated from:

$$D_{L_{CO}} (Adjusted) = D_{L_{CO}} (Base) + 0.000845\dot{V}_A - 5.65 \quad (10-26)$$

(See discussion of base pulmonary diffusion rate.)

10.5 References for Section 10

- Biller, W. F. and H.M. Richmond. 1982. "Sensitivity Analysis on Coburn Model Predictions of COHb Levels Associated with Alternative CO Standards." Report to Strategies and Air Standards Division of the Office of Air quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, November, 1982.
- Brouillard, R. P., M.E. Contrad, and T.A. Bensinger. 1975. Blood, Vol.45, pp. 67-69.
- Brainard, J. and D. Burmaster. 1992. "Bivariate Distributions for Height and Weight of Men and Women in the United States," Risk Analysis, Vol. 12, No.2, pp. 267-275.
- Burke, P. D., F.L. Rodkey, T.F. Blaschke, H.A. Collison, and J.G. Waggoner. 1974. "Comparison of Plasma Bilirubin Turnover and Carbon Monoxide Production in Man." J. Lab. Clin. Med., Vol. 83, pp. 29-37.
- Burmaster, D. E., and E. A. C. Crouch. 1997. "Lognormal Distributions of Body Weight as a Function of Age for Males and Females in the United States, 1976 - 1980," Risk Analysis, Vol. 17, No. 4.

Coburn, R. F., R.E. Forster, and R.B. Kane. 1965. "Considerations of the Physiology and Variables That Determine Blood Carboxyhemoglobin Concentration in Man." J. Clin. Invest., Vol. 44, pp. 1899-1910.

Coburn, R. F., W.S. Blakemore, and R.E. Forster. 1963. "Endogenous Carbon Monoxide Production in Man." J. Clin. Invest., Vol. 42, pp.1172-1178.

Coltman, C. A., and G.M. Dudley III. 1969. "The Relationship Between Endogenous Carbon Monoxide and Total Heme Mass in Normal and Abnormal Subjects." Am. J. Med. Sci., Vol. 258, pp. 374-385.

Collier, C. and J.R. Goldsmith. 1983. "Interactions of Carbon Monoxide and Hemoglobin at High Altitude." Atmos. Environ., Vol. 17, pp. 723-728.

Delivoria-Papadoppulos, M., R.F. Coburn, and R.E. Forster. 1974. "Cyclic Variation of Rate of Carbon Monoxide Production in Normal Women." J. Appl. Physiol., Vol. 36, pp. 49-71.

Friedlander, S. K. 1982. "Letter from The Clean Air Scientific Advisory Committee to USEPA Administrator." August 31, 1982.

Johnson, T. and R.A. Paul. 1983. "The NAAQS Model (NEM) Applied to Carbon Monoxide." EPA 450/5-83-003, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Johnson, T. 1998. "Memo No. 5: Equations for Converting Weight to Height Proposed for the 1998 Version of pNEM/CO." Memorandum Submitted to U.S. Environmental Protection Agency. TRJ Environmental, Inc., 713 Shadylawn Road, Chapel Hill, North Carolina 27514.

Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM. Report prepared by International Technology Air Quality Services under EPA Contract No. 68-D0-0062. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Johnson, T., G. Mihlan, J. LaPointe, et al. 2000. Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Residents of Denver and Los Angeles Using pNEM/CO (Version 2.1), prepared for the Risk and Exposure Assessment Division of U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, June.

Luomanmaki, K. and R.F. Coburn. 1969. "Effects of Metabolism and Distribution of Carbon Monoxide on Blood and Body Stores." Am. J. Physiol., Vol. 217, No.2, pp. 354-362.

Lynch, S. R. and A.L. Moede. 1972. "Variation in the Rate of Endogenous Carbon

Monoxide Production in Normal Human Beings.” J. Lab. Clin. Med., Vol. 79, pp. 85-95.

Marcus, A. H. 1980. “Mathematical Models for Carboxyhemoglobin.” Atmos. Environ. , Vol. 14, pp. 841-844.

McCartney, M. L. 1990. “Sensitivity Analysis Applied to Coburn-forster-kane Models of Carboxhemoglobin Formation.” Am. Ind. Hyg. Assoc. J., Vol. 51, No. 3, pp.169-177.

Merke, C., E. Cavallin-Stahl, and B. Lundh. 1975. “Carbon Monoxide Production and Reticulocyte Count in Normal Women: Effect of Contraceptive Drugs and Smoking.” Acta Med. Scan., Vol. 198, pp.155-160.

Muller, K. E. and C.N. Barton. 1987. “A Nonlinear Version of the Coburn, Forster and Kane Model of Blood Carboxyhemoglobin.” Atmos. Environ., Vol. 21, pp.1963-1967.

Peterson, J. E. and R.D. Stewart. 1975. “Predicting the Carboxhemoglobin Levels Resulting from Carbon Monoxide Exposures.” J. Appl. Physiol., Vol. 39, pp. 633-638.

Press, W. H., B.P. Flannery, S.A. Teukolsky, and W.T. Vettering. 1986. NUMERICAL RECIPES, Cambridge University Press.

Rodkey, F. L., J. D. O’Neal, and H. A. Collison. 1969. “Oxygen and Carbon Monoxide Equilibria of Human Adult Hemoglobin at Atmospheric and Elevated Pressure.” Blood, Vol. 33, pp. 57 - 65.

Salorinne, Y. 1976. “Single-breath Pulmonary Diffusing Capacity.” Scand. J. Resp. Diseases, Supplementum 96.

Tikuisis, P., H.D. Madill., B.J. Gill, W.F. Lewis, K.M. Cox, and D.M. Kane, D. M. (1987) A critical analysis of the use of the CFK equation in predicting COHb formation. Am. Ind. Hyg. Assoc. J. 48(3):208-213.

Tikuisis, P., Kane, D. M., McLellan, T. M., Buick, F., and S.M. Fairburn. 1992. “Rate of Formation of Carboxyhemoglobin in Exercising Humans Exposed to Carbon Monoxide.” J. Appli. Physiol., Vol. 72, No. 4.

U.S. Department of Health and Human Services. 1982. “Hematological and Nutritional Biochemistry Reference Data for Persons 6 Months-74 Years of Age: United States, 1976-80.” DHHS Publication No. (PHS)83-1682.

U.S. Environmental Protection Agency. 1978. “Altitude as a Factor in Air Pollution.” EPA-600/9-78-015, U.S. Environmental Protection Agency, Research Triangle Park, NC.

U. S. Environmental Protection Agency. 1990. “Air Quality Criteria for Carbon Monoxide.” EPA-600/8-90-045A, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Werner, B. and J. Lindahl. 1980. "Endogenous Carbon Monoxide Production after Bicycle Exercise in Healthy Subjects and in Patients with Hereditary Spherocytosis." Scan. J. Lab. Invest., Vol. 40, pp. 319-324.

SECTION 11

USEFUL RECENT RESEARCH

This section provides summaries of recent research in four areas that may be useful in enhancing existing exposure models and developing new models.

Subsection 11.1 describes advanced methods for calculating heat indices and wind chill factors that may be useful in developing improved algorithms for sequencing time/activity data and simulating the opening and closing of windows.

Subsection 11.2 summarizes the results of recent field studies in Texas and North Carolina in which technicians monitored the positions of windows in residences and motor vehicles under varying conditions. Data from these studies may be useful in improving the algorithms employed by various OAQPS models to simulate the opening and closing of windows in residential and vehicular microenvironments.

As noted in Subsection 6.4.3.2, the literature provides few measurements of air exchange rates in residences with windows open. Subsection 11.3 presents results from a recent study in which researchers measured air exchange rates under varying air flow conditions (including open windows) in a residence in Columbus, Ohio.

Ambient ozone concentrations are typically determined by fixed-site monitors that vary with respect to height above ground and distance to nearby roadways. Subsection 11.4 presents the results of studies conducted in Cincinnati that provide data useful in accounting for the effects of these factors.

Note: Subsections 11.2.1, 11.2.3, and 11.3 are currently incomplete because the source studies are still in progress.

11.1 Advanced Temperature-Related Comfort Indices

Subsection 5.1.2 describes a method used in pNEM, HAPEM, and other OAQPS exposure models to sequence diary-derived data into multi-day sequences according to daily maximum temperature. Temperature is also used as a factor in the probabilistic algorithms that are used to determine window position (and thus air exchange rate) in pNEM/O₃ and pNEM/CO (see Subsections 6.3.4 and 6.4.3). In developing these methods, researchers assumed that people vary their activities and environmental conditions according to a “comfort” index related to the ambient (outdoor) temperature.

Ideally, future exposure models will employ more advanced comfort indices that account for the effects of humidity and wind speed as well as temperature. Under conditions of high humidity and high temperature, low rates of heat and moisture transfer from the body cause the perceived “apparent” temperature to be higher than

the ambient temperature. The heat index parameter described in Subsection 11.1.1 can be used to quantify this effect. Apparent temperature is perceived to be lower than ambient temperature when temperature is low and wind velocity is high. Subsection 11.1.2 describes the use of the wind chill factor to quantify this effect.

11.1.1 Heat Index

Steadman (1979a) has developed an extensive biometeorological model to quantify the effect of humidity on apparent temperature, and the results of this model have been adopted by the US National Weather Service (NWS) as the Heat Index (HI). This subsection describes Steadman's apparent temperature model, presents the NWS HI table, and provides a second-order regression of the table values suitable for calculation of HI in model algorithms.

11.1.1.1 The Steadman Model

The Steadman apparent temperature model [Steadman (1979a)] combines physiological variables, clothing science, and meteorological factors to determine the body's overall resistance to heat and moisture transfer under various temperature and humidity conditions. The apparent temperature is equivalent to the ambient temperature that would be required to produce these transfer resistances at a standard "temperate" absolute humidity of 1.6 kPa water vapor pressure. This humidity corresponds to a sea level dew point of 57.2 degrees F (14.0 degrees C), which is the standard humidity at which testing labs are held in the temperate zone.

Tables 11-1 and 11-2 present the physiological, clothing, and meteorological parameters included in the Steadman model. These parameters are used to calculate the value of primary model variables describing heat transfer from the skin and clothing to the environment. Table 11-3 lists the primary heat transfer variables in the model. The values of these primary heat transfer variables are then used as model inputs, and the model is solved again to determine apparent temperature. The model assumes that a short-sleeved shirt and long trousers are worn under "mild" conditions, when the apparent temperature is 77 degrees F or less, and that negligible clothing is worn under "severe" conditions (apparent temperature > 77 degrees F).

The model is solved once at ambient temperature and vapor pressure to determine the skin resistance (severe conditions) or clothing thickness (mild conditions) required to achieve thermal equilibrium. Thermal equilibrium is reached when heat production from activity equals total heat loss from the lungs and from covered and uncovered skin. The resulting skin resistance or clothing thickness value is then used to solve the model a second time for apparent temperature at the reference water vapor pressure of 1.6 kPa.

Human dimensions and activity level are based on physiological data representing typical adults. The sweating rate is assumed to be sufficient to reach thermal equilibrium, but not such that sweat drips from the body. The model scope is designed to include a range of weather conditions exceeded on less than 1% of the earth's surface and for less than 1% of the time. This covers ambient temperatures of 68–122 degrees F and vapor pressures from 0–4.6 kPa (dew points up to 88 degrees F). Under low humidity conditions, the apparent temperature can be less than the

Table 11-1. Assumed Parameter Values Used in Steadman Apparent Temperature Model.

Parameter	Description	Value
Average Wind Speed	Wind speed at human height (1.7 m), calculated from 10 m anemometer wind speed of 2.5 m/s	1.3 m s ⁻¹
Body Core Temperature	Internal body temperature	37 degrees C
Body Core Vapor Pressure	Determined from core temperature and salinity	5.65 kPa
Clothing Cover	Fraction of body covered by clothing	Mild conditions: 0.84 Severe conditions: 0
Clothing Resistance to Heat Transfer	Assumes clothing consists of 20% fiber and 80% air	0.167 m ² K W ⁻¹ cm ⁻¹
Clothing Resistance to Moisture Transfer	Considers vapor diffusion through air portion of clothing to be dominant	0.021 m ² kPa W ⁻¹ cm ⁻¹
Effective Radiation Area	Fraction of skin surface area exchanging heat with the surroundings	Mild conditions, clothed portion of body: 0.79 Mild conditions, bare portion of body: 0.85 Severe conditions: 0.80
Human Activity	Steady-state heat output resulting from walking at 1.4 m/s	180 W m ⁻² skin surface
Human Height	Used in calculating skin surface area	1.7 m
Human Significant Diameter	Used in standard heat transfer engineering calculations; a function of the body's volume and density	15.3 cm
Human Weight	Used in calculating skin surface area	67 kg
Skin Surface Area		1.78 m ²

Table 11-2. Derived Quantities Calculated in Steadman Apparent Temperature Model.

Parameter	Description	Equation Form
Convective Heat Transfer Coefficient (h_c)	Quantifies convective heat loss from skin and clothing surfaces. Vector average determined from wind speed, walking speed, and properties of moist air. Found to be nearly independent of temperature and humidity.	Mild, unclothed portion: $17.4 \text{ W m}^{-2} \text{ K}^{-1}$ Mild, clothed portion: $11.6 \text{ W m}^{-2} \text{ K}^{-1}$ Severe: $12.3 \text{ W m}^{-2} \text{ K}^{-1}$
Effective Wind Speed	Relative air movement calculated from vector sum of average wind speed and walking speed	Calculation incorporated in Convective Heat Transfer Coefficient calculation (see above)
Radiative Heat Transfer Coefficient (h_r)	Quantifies radiation from skin and clothing surfaces. Can be described by a linear function of ambient temperature over the meteorological conditions covered by this model.	Mild, unclothed portion: $(4.18 + 0.036 \cdot T) \text{ W m}^{-2} \text{ K}^{-1}$ Mild, clothed portion: $(3.35 + 0.049 \cdot T) \text{ W m}^{-2} \text{ K}^{-1}$ Severe: $(4.10 + 0.028 \cdot T) \text{ W m}^{-2} \text{ K}^{-1}$
Skin Resistance to Heat Transfer (R_s)	Primary variable in calculating apparent temperature under severe conditions. Under mild conditions, it is treated as a constant determined from physiological data, primarily activity.	Mild: $0.0387 \text{ m}^2 \text{ K W}^{-1}$ Severe: see Table 6-X3
Skin Resistance to Moisture Transfer (Z_s)	For mild conditions, calculated from vapor pressure (humidity) difference across skin; for severe conditions, calculated as a fitted function of Skin Resistance to Heat Transfer	Mild: $0.0521 \text{ m}^2 \text{ kPa W}^{-1}$ Severe: $Z_s = (6.0 \times 10^5) \times R_s$
Surface Resistance to Heat Transfer (R_a)	A function of Radiative and Convective Heat Transfer Coefficients.	$R_a = 1 / (h_r + h_c)$

Parameter	Description	Equation Form
Surface Resistance to Moisture Transfer (Z_a)	Calculated by analogy to convective heat transfer. The ratio between moisture transfer and convective heat transfer was found to be nearly constant over the full range of temperatures and vapor pressures at sea level.	$Z_a = 0.0606 / h_c$
Surface Temperatures and Vapor Pressures of Skin and Clothing	Affects heat and moisture transfer from the skin's surface	Iterative calculation involving convective, radiative, and evaporative heat transfer coefficients; skin and clothing resistances; and skin surface conditions.
Ventilation Heat Loss	Heat loss from lungs as a fraction of total heat loss; dependent upon ambient temperature and vapor pressure	$0.143 - 0.00112 * T - 0.0168 * P$

Table 11-3. Primary Heat Transfer Variables in Steadman Apparent Temperature Model.

Parameter	Description	Equation Form
Skin Resistance to Heat Transfer (R_s)	Primary variable in calculating apparent temperature under severe conditions. Under mild conditions, it is treated as a constant determined from physiological data, primarily activity.	Mild: $0.0387 \text{ m}^2 \text{ K W}^{-1}$ Severe: calculated by model to provide heat loss sufficient to establish thermal equilibrium
Clothing Thickness (d_c)	Primary variable in calculating apparent temperature under mild conditions. Negligible clothing assumed under severe conditions.	Mild: calculated by model to provide heat loss sufficient to establish thermal equilibrium Severe: 0.0 mm

ambient temperature.

The effects of solar radiation and wind on apparent temperature are considered separately in a companion paper [Steadman (1979b)]. While direct noontime solar radiation can raise the apparent temperature by as much as 9–13 degrees F, NWS Heat Index values do not include radiation or wind effects. The model is not designed to apply to buildings, animals, or plants, which have different heat transfer characteristics than humans.

11.1.1.2 NWS Heat Index

The NWS uses the apparent temperature determined by the Steadman model as its Heat Index [Rothfus (1990)]. HI values are most often presented as a lookup table [National Weather Service (2001)], although a nomograph is also available. Both the table and the nomograph express HI as a function of only air temperature and relative humidity, although all other variables in the Steadman model are implied. The NWS HI table is provided in Table 11-4.

Table 11-4. Heat Index Table from National Weather Service Showing HI in degrees F.

Temperature, deg. F	Relative Humidity, %												
	40	45	50	55	60	65	70	75	80	85	90	95	100
110	136												
108	130	137											
106	124	130	137										
104	119	124	131	137									
102	114	119	124	130	137								
100	109	114	118	124	129	136							
98	105	109	113	117	123	128	134						
96	101	104	108	112	116	121	126	132					
94	97	100	103	106	110	114	119	124	129	135			
92	94	96	99	101	105	108	112	116	121	126	131		
90	91	93	95	97	100	103	106	109	113	117	122	127	132
88	88	89	91	93	95	98	100	103	106	110	113	117	121
86	85	87	88	89	91	93	95	97	100	102	105	108	112
84	83	84	85	86	88	89	90	92	94	96	98	100	103
82	81	82	83	84	84	85	86	88	89	90	91	93	95
80	80	80	81	81	82	82	83	84	84	85	86	86	87

11.1.1.3 Approximate Calculation of Heat Index

In order to provide a form of the Heat Index suitable for use in calculations and modeling applications, NWS performed a multiple regression analysis on the data from Table 11-4. The resulting equation enables approximate calculation of HI using only air temperature and relative humidity as independent variables. HI is estimated using the expression

$$\text{HI} = -42.4 + (2.049)(T) + (10.14)(\text{RH}) - (0.2248)(T)(\text{RH}) - (6.838 \times 10^{-3})(T)^2 - (5.482 \times 10^{-2})(\text{RH})^2 + (1.229 \times 10^{-3})(T)^2(\text{RH}) + (8.528 \times 10^{-4})(T)(\text{RH})^2 - (1.99 \times 10^{-6})(T)^2(\text{RH})^2 \quad (11-1)$$

where T is ambient dry bulb temperature (degrees F) and RH is relative humidity (percent).

The HI value determined from Equation 11-1 has an error of +/- 1.3 degrees F compared with the original Steadman model. Over the more limited range of conditions included in the NWS table (Table 11-4), the error is much lower, on the order of +/- 0.5 degrees F.

Equation 11-1 may therefore be used to calculate HI in model algorithms for temperature values of 79 degrees F and higher over the entire range of relative humidity conditions (0-100%). If HI values are desired for mild conditions when temperatures are in the range 68-78 degrees F, the original Steadman model table should be consulted [Steadman (1979a)]. HI is not applicable to temperatures below 68 degrees F.

11.1.2 Wind Chill Factor

As discussed above, the developers of the NEM and HAPEM-series exposure models have assumed that people vary their activities and environmental conditions according to a "comfort" index related primarily to the ambient (outdoor) temperature. Subsection 11.1.1 describes an improved heat index that may be useful in characterizing conditions according to temperature and humidity. This subsection describes a recently-developed wind chill factor that may be employed to characterizing conditions in terms of temperature and wind speed.

At low temperatures, windy conditions increase the rate of heat transfer from the body and cause the perceived temperature to be lower than the ambient temperature. High winds combined with sub-freezing temperatures can lead to frostbite within minutes. A wind chill formula developed in the 1940's to quantify this effect based on the cooling rate of water in plastic cylinders has recently been updated by a joint committee formed by the U.S. and Canadian weather services. The new formula, known as the Wind Chill Temperature (WCT) Index, is derived from the results of an algorithm modeling heat transfer from the human face. Both the U.S. National Weather Service (NWS) and the Meteorological Service of Canada (MSC) have implemented WCT as their cold weather apparent temperature index for the 2001/2002 winter season. The remainder of this subsection describes the development of the WCT Index model, outlines the features of the WCT Index algorithm, presents the NWS WCT formula and table, and discusses projected work to further improve WCT and other apparent temperature indices.

11.1.2.1 Development of the WCT Index

The 2001/2002 WCT Index model was developed by the Joint Action Group for Temperature Indices (JAG/TI), a group composed of representatives of U.S. Federal agencies, Canadian national ministries, the academic research community, and the International Society of Biometeorology (ISB) (Nelson et al. 2002). The group held discussions during 2000 and 2001 to review the existing NWS and MSC indices and consider proposed improvements to the index methods. The JAG/TI agreed that the indices in use at that time overstated the effect of the wind, particularly at very low temperatures and high wind speeds. This bias resulted in reported wind chill temperatures that did not pose the same risk of frostbite and hypothermia as the actual temperature; i.e., a wind chill of -40°F was not as severe as a dry-bulb temperature of -40°F . Several methodological problems contributed to this discrepancy, including (1) the combined modeling of radiative and convective heat losses; (2) ignoring thermal resistance of the skin; (3) using wind speed measured at a height of 10 meters rather than human face height; and (4) using an unrealistically high value for the assumed skin temperature.

The JAG/TI decided to use two recently published wind chill models addressing these problems (Bluestein and Zecher 1999; Osczevski 1995, 2000) as the basis for the new WCT Index model. These models use the human face to evaluate wind chill effects, as facial exposure to the wind presents the greatest discomfort and hazard of frostbite (Osczevski 2000). The group also recommended that human studies be carried out to validate and improve the WCT Index. These human trials were conducted in early summer 2001. Volunteers walked on a treadmill in a chilled wind tunnel while thermal transducers measured heat loss from various parts of their faces.

The completed algorithm for the WCT Index was presented at the JAG/TI meeting in August 2001 and accepted by the group for implementation by NWS, MSC, and the U.S. Department of Defense (DOD) during autumn 2001. The next section describes the features of the WCT Index model.

11.1.2.2 The WCT Index Model

The 2001/2002 WCT Index model currently uses wind speed and air temperature as the primary variables, with solar radiation to be added for the 2002/2003 winter season. An iterative method must be used to solve for WCT, since heat transfer coefficients depend on the skin surface temperature, which is unknown (Bluestein and Zecher 1999). The estimated skin surface temperature and WCT are continually revised using updated heat transfer rates until the WCT value converges.

The new WCT Index model includes several features designed to improve the accuracy of the apparent temperature estimate compared to the previous wind chill index (Nelson et al. 2002). Specifically, the WCT Index

- is based on a human face model;
- corrects wind speed measured at the standard height of 33 ft (10 m) to the speed at 5 ft (1.5 m), the height of an average adult's face;

- assumes the face-level wind speed under “calm” conditions is equal to walking speed, 3 miles/hour (4.8 km/hr);
- incorporates heat transfer theory to estimate heat loss from the body to its surroundings under a range of wind conditions;
- uses skin tissue heat transfer resistance values consistent with the cold cheek of an inactive individual;
- uses solar radiation conditions producing the maximum chilling effect (clear night sky).

11.1.2.3 WCT Index Algorithms

To eliminate the computational effort that would be required to repeatedly perform iterative calculations of WCT for the weather services’ hourly forecasts and other uses, researchers developed non-iterative equations to estimate WCT in English and metric units. Researchers performed more than 800 runs of the full iterative model with different combinations of wind speed and ambient air temperature, and then conducted a multiple regression analysis of the results. The regression equations are presented below.

In English units:

$$\text{WCT } (^\circ\text{F}) = 35.74 + 0.6215(T) - 35.75(V)^{0.16} + 0.4275(T)(V)^{0.16} \quad (11-2)$$

T = temperature in °F

V = observed wind speed at 33 ft elevation in miles/hr

In metric units:

$$\text{WCT } (^\circ\text{C}) = 13.12 + 0.6215(T) - 11.37(V)^{0.16} + 0.3965(T)(V)^{0.16} \quad (11-3)$$

T = temperature in °C

V = observed wind speed at 10 m elevation in km/hr

Although frostbite will not occur at temperatures above the freezing point of water (32°F, 0°C), Equations 11-2 and 11-3 may be used to calculate apparent temperature for ambient temperatures at or below 50°F (10°C) (National Weather Service 2002). Tables 11-5 and 11-6 present the WCT Index in chart format.

11.1.2.4 Ongoing Work on Temperature Indices

The JAG/TI is preparing to adjust the WCT index model to account for solar radiation effects prior to the 2002/2003 winter season. Currently, WCT is calculated for

Table 11-5. 2001/2002 WCT Index Table in English Units (National Weather Service 2002).

Wind speed (mph) at 33 ft elevation	Temperature (degrees F)																	
	40	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40	-45
5	36	31	25	19	13	7	1	-5	-11	-16	-22	-28	-34	-40	-46	-52	-57	-63
10	34	27	21	15	9	3	-4	-10	-16	-22	-28	-35	-41	-47	-53	-59	-66	-72
15	32	25	19	13	6	0	-7	-13	-19	-26	-32	-39	-45	-51	-58	-64	-71	-77
20	30	24	17	11	4	-2	-9	-15	-22	-29	-35	-42	-48	-55	-61	-68	-74	-81
25	29	23	16	9	3	-4	-11	-17	-24	-31	-37	-44	-51	-58	-64	-71	-78	-84
30	28	22	15	8	1	-5	-12	-19	-26	-33	-39	-46	-53	-60	-67	-73	-80	-87
35	28	21	14	7	0	-7	-14	-21	-27	-34	-41	-48	-55	-62	-69	-76	-82	-89
40	27	20	13	6	-1	-8	-15	-22	-29	-36	-43	-50	-57	-64	-71	-78	-84	-91
45	26	19	12	5	-2	-9	-16	-23	-30	-37	-44	-51	-58	-65	-72	-79	-86	-93
50	26	19	12	4	-3	-10	-17	-24	-31	-38	-45	-52	-60	-67	-74	-81	-88	-95
55	25	18	11	4	-3	-11	-18	-25	-32	-39	-46	-54	-61	-68	-75	-82	-89	-97
60	25	17	10	3	-4	-11	-19	-26	-33	-40	-48	-55	-62	-69	-76	-84	-91	-98

Table 11-6. 2001/2002 WCT Index Table in Metric Units (Nelson et al. 2002).

Wind speed (km/hr) at 10 m elevation	Temperature (degrees C)											
	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50	
5	-2	-7	-13	-19	-24	-30	-36	-41	-47	-53	-58	
10	-3	-9	-15	-21	-27	-33	-39	-45	-51	-57	-63	
15	-4	-11	-17	-23	-29	-35	-41	-48	-54	-60	-66	
20	-5	-12	-18	-24	-30	-37	-43	-49	-56	-62	-68	
25	-6	-12	-19	-25	-32	-38	-44	-51	-57	-64	-70	
30	-6	-13	-20	-26	-33	-39	-46	-52	-59	-65	-72	
35	-7	-14	-20	-27	-33	-40	-47	-53	-60	-66	-73	
40	-7	-14	-21	-27	-34	-41	-48	-54	-61	-68	-74	
45	-8	-15	-21	-28	-35	-42	-48	-55	-62	-69	-75	
50	-8	-15	-22	-29	-35	-42	-49	-56	-63	-69	-76	
55	-8	-15	-22	-29	-36	-43	-50	-57	-63	-70	-77	
60	-9	-16	-23	-30	-36	-43	-50	-57	-64	-71	-78	

clear nighttime conditions, a conservative assumption. Other solar radiation conditions

to be included in the updated WCT are daytime clear, daytime cloudy, and nighttime cloudy. This will provide a more realistic and useful estimate of apparent temperature for a variety of conditions.

ISB has established a commission (Commission 6) charged with development of a single comprehensive temperature index, tentatively titled the Universal Thermal Climate Index (UTCI) (International Society of Biometeorology 2001). The UTCI will be designed to apply to year-round outdoor weather conditions, eliminating the need for separate heat and cold indices. It will also bridge the gap between the upper limit of the WCT Index (50°F) and the lower limit of the NWS Heat Index (80°F). The UTCI will be developed using physiologically appropriate heat transfer calculations for both heating and cooling of the body. Applications of the UTCI will include both forecasting and climatological purposes. Commission 6 expects the UTCI to be fully developed by 2005.

11.2 Factors Affecting Window Position

11.2.1 Residences (Durham, NC)

*****field study not complete*****

11.2.2 Motor Vehicles (Houston, TX)

The exposure models developed by OAQPS typically estimate exposures by simulating the movement of specific population groups through defined microenvironments. The accuracy of the resulting exposure estimates is highly dependent on the validity of the probabilistic algorithms used to estimate pollutant concentrations in each microenvironment. In the more sophisticated models, a mass balance model is used to calculate the pollutant concentration within an enclosed microenvironment as a function of outside concentration, air exchange rate, decay rate, and deposition rate, as appropriate. In motor vehicles, air exchange rate is significantly affected by vehicle speed, window position, vent status, and the use of air conditioners and heaters. Two of these factors, vehicle speed and window position, can be monitored by external observers. A pilot study was conducted in Houston, Texas, during September 2000 to determine whether useful information on window position as a function of speed and other factors could be obtained through the use of video cameras. An article by Long, Johnson, and Ollison (2002) describes the procedures used in this study, presents results of a statistical analysis of the data, and presents recommendations for follow-up studies. This subsection presents potentially useful material excerpted from this article.

11.2.2.1 Study Methodology

The primary purpose of the Houston pilot study was to test an approach for gathering data on window position that could be employed in later, more comprehensive studies. Researchers also hoped to obtain data that would be immediately useful to exposure modelers requiring window position data representative of high temperature/humidity conditions in metropolitan areas with heavy vehicle use. In

addition, researchers desired a location where older vehicles with non-functioning air conditioning systems would be well-represented within the vehicle mix. The Houston metropolitan area was selected as the general area for the pilot study because (1) the city tends to be hot and humid during the summer, (2) the population of Houston contains a substantial population of residents in the lower socioeconomic stratum, and (3) the metropolitan area includes outlying communities not subject to vehicle emission inspection and maintenance (I &M) programs. Researchers expected these outlying communities to have relatively large populations of older, poorly-maintained vehicles without functioning air conditioners. The pilot study was conducted in mid-September 2000 during a period when the Heat Index (HI) – a function of temperature and relative humidity – varied between 80 and 101 degrees F.

The Houston I & M district includes all of surrounding Harris County, but not adjacent counties. Researchers selected three videotaping locations near the Harris County border (Figure 11-1) with the goal of including vehicles from outside the I & M district. These sites included one location along an interstate highway and two locations along arterial roads. Each site permitted an elevated view of vehicles as they proceeded through a turn, thereby exposing all windows to the stationary video camera.

Katy Mills. The interstate highway site, hereafter referred to as “Katy Mills”, was located along the westbound exit ramp of Interstate 10 at Pin Oak Road, near the town of Katy, TX. This section of I-10 has three westbound lanes and three eastbound lanes. The exit ramp filmed in this study has two lanes. This exit experiences high traffic volume because it serves the Katy Mills outlet mall. The video camera was situated at the end of the exit ramp, near the Pin Oak Road overpass.

The Katy Mills location can be characterized as a rural interstate site. The nearby Katy Mills outlet mall appeared to have been recently constructed and was located away from other development. Field staff noted that most of the vehicles filmed for this study were traveling toward the outlet mall.

Katyland Drive. One arterial road site, referred to as “Katyland Drive”, was located at the intersection of Katyland Drive and US Route 90 in the town of Katy, TX. The camera was set up at a safe location on a railroad grade, one of the few elevated spots in Katy. Vehicles traveling south on Katyland Drive were videotaped as they turned left onto US 90. Katyland Drive has one lane in each direction for through traffic and one left turn lane in the southbound direction.

Katyland Drive runs along the edge of town, away from the downtown area of Katy. Land use in the vicinity of the videotaping site includes light and medium industry, while a school and athletic fields are located along Katyland Drive in the opposite direction of traffic flow, upstream of the camera. Vehicles filmed by the video camera were heading out of town toward an Interstate 10 interchange approximately 1 mile away.

Top of Figure 11-1

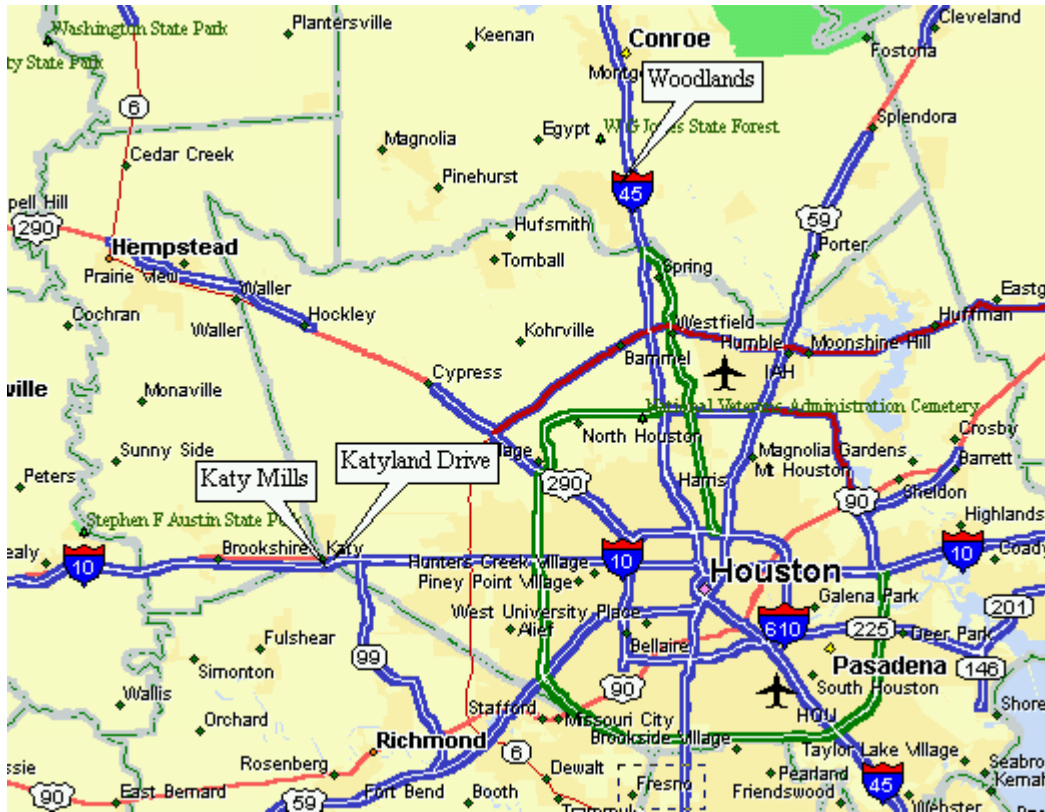


Figure 11-1. Map of Houston study area showing videotaping locations

Woodlands. The other arterial road site, called “Woodlands”, was located at the intersection of Grogan’s Mill Road and Research Forest Drive in the community of The Woodlands. The Woodlands is situated in Montgomery County, north of Houston along Interstate 45. Traffic was videotaped turning left from Grogan’s Mill Rd. onto Research Forest Dr. Grogan’s Mill Road is a one-way street with one left turn lane and one combination left/right turn lane. The video camera was again set up on a slightly elevated site near the intersection.

Grogan’s Mill Road is a suburban parkway carrying local traffic. Field staff did not observe schools, industry, or retail locations in the area. Vehicles on the videotape were heading away from Interstate 45.

Diagrams and detailed descriptions of the sites are provided in the article by Long, Johnson, and Ollison (2002).

Field staff conducted three videotaping sessions at the Katy Mills (interstate) location, and one session each at the Katyland Drive and Woodlands sites. Table 11-7 lists the date and nominal time period of each taping session as well as the observed ranges of vehicle speed and HI.

Videotaping was performed mainly during afternoon and early evening hours, when HI values were likely to be highest. Vehicles used for window position determination were videotaped between the hours of 11 am to 2 pm and 4 pm to 7 pm. Vehicle speeds varied from 30 to 74 mph during the five taping sessions. HI was calculated by Equation 11-1. During the taping sessions, temperature varied from 81 to 95 degrees F, relative humidity varied from 28 to 55 percent, and HI varied from 80 to 101 degrees F. No precipitation occurred during any of the taping sessions, and roadways were dry in all cases.

11.2.2.2 Results

Analysts initially constructed a database listing data on 868 vehicles monitored during the five sessions listed in Table 11-7. The database contained a variety of parameters relating to monitoring conditions such as location, date, time of day, temperature, relative humidity, and HI. Data specific to each vehicle included vehicle type (e.g., minivan), color, number of occupants (if observable), number of windows fully open, number of windows partially open, number of windows closed, and number of windows not visible. The database also included data on the position of sunroofs (open/closed) and convertible tops (up/down). Prior to analyzing the database, analysts identified 107 vehicles (12 percent) in which one or more windows were not visible because of sun angle, another vehicle, or attached visor. These vehicles were omitted from subsequent analyses. Three motorcycles were also omitted. The remaining 758 vehicles were used as the master database for the statistical analyses that follow.

Table 11-8 presents frequency distributions for selected parameters relating to monitoring conditions based on the 758 vehicles in the master database. Sixty percent of the vehicles were videotaped at the Katy Mills interstate location, with the remaining vehicles evenly split between the two arterial road locations (Katyland Drive and Woodlands). Sixty percent of the vehicles were monitored on Friday, September 15. The remaining vehicles were monitored on Saturday, September 16. Of the six clock

Table 11-7. Characteristics of Videotaping Sessions.

Location	Date	Time period	Range of vehicle speeds, mph	Range of Heat Index, degrees F	Number of vehicles initially included in master database	Number of vehicles with valid data for statistical analysis
Katy Mills	Friday 9/15/2000	11:50 – 12:13	59 to 70	100	178	152
	Friday 9/15/2000	18:29 – 18:53	63 to 70	97 to 100	174	152
	Saturday 9/16/2000	11:04 – 11:23	64 to 74	80 to 81	176	149
Katyland Drive	Friday 9/15/2000	16:14 – 17:07	30 to 48	97 to 101	169	153
Woodlands	Saturday 9/16/2000	13:33 – 13:59	37 to 48	85 to 88	171	152
Total					868	758

hours represented in the database, the clock hour beginning at 1100 (i.e., 1100 to 1159) had the largest number of vehicles (215) in 30 total minutes of taping and the clock hour beginning at 1700 had the smallest number (31) in 8 total minutes of taping. HI ranges of 80, 85 to 87, and 97 to 101 degrees F. were associated with 149, 152, and 457 vehicles, respectively. Approximately 38 percent of the vehicle speeds were between 30 and 49 mph; the remaining 62 percent were between 50 and 79 mph with 39 percent occurring in the interval of 60 to 69 mph.

Table 11-9 presents frequency distributions for selected parameters relating to vehicle characteristics. Of the seven vehicle types appearing in the database, passenger cars with hardtops occur most frequently (52.1 percent) and commercial trucks least frequently (1.1 percent). With respect to the seven window configurations listed in Table 11-9, vehicles with four windows occur most frequently (58.6 percent) followed by vehicles with two windows (30.9 percent). Note that these categories do not include vehicles with sunroofs or convertible tops which are classified separately.

Although vehicle color was included in the database primarily for quality assurance purposes, subsequent statistical analyses detected a weak correlation between color and window position. Table 11-9 provides frequency distributions for 10 color categories. Note that the colors are characterized by hue only. In hindsight, color should also have been characterized by shade (e.g., light, medium, dark), as shade is more likely to affect internal vehicle temperature than hue.

Table 11-9 does not provide a frequency distribution for number of vehicle occupants, because this statistic was determined for only 170 (22 percent) of the 758 vehicles in the master database. In reviewing the videotapes, researchers found that it was difficult to accurately determine the number of passengers in many of the vehicles

Table 11-8. Frequency Distributions for Selected Variables Relating to Monitoring Conditions.

Variable	Category	Number of valid cases	Percent of total valid cases	Percent of vehicles in category with one or more open windows ^a
Location	Katy Mills	453	59.8	15.7
	Katyland Drive	153	20.2	25.5
	Woodlands	152	20.1	19.7
Date	9/15/00 (Friday)	457	60.3	17.9
	9/16/00 (Saturday)	301	39.7	19.3
Time (total taping duration in this clock hour, minutes)	1100 – 1159 (30)	215	28.4	16.7
	1200 - 1259 (14)	86	11.3	7.0
	1300 - 1359 (27)	152	20.1	19.7
	1600 - 1659 (46)	122	16.1	27.1
	1700 - 1759 (8)	31	4.1	19.4
	1800 - 1859 (25)	152	20.1	19.1
Temperature, degrees F	81	124	16.4	17.7
	82	25	3.3	24.0
	88	60	7.9	23.3
	90	92	12.1	17.4
	91	207	27.3	11.1
	93	145	19.1	23.5
	95	105	13.9	23.8
Relative humidity, percent	28	92	12.1	17.4
	29	85	11.2	23.5
	30	76	10.0	19.7
	31	48	6.3	14.6
	42	31	4.1	32.3
	43	122	16.1	23.8
	47	97	12.8	20.6
	50	55	7.3	16.4
	55	152	20.1	9.2

Variable	Category	Number of valid cases	Percent of total valid cases	Percent of vehicles in category with one or more open windows ^a
Heat Index, degrees F	80	149	19.7	18.8
	85	60	7.9	23.3
	87	92	12.1	17.4
	97	103	13.6	22.3
	99	249	32.8	13.7
	100	31	4.1	32.3
	101	74	9.8	20.2
Vehicle speed, mph	30 to 39	118	15.6	26.3
	40 to 49	167	22.0	21.0
	50 to 59	29	3.8	10.3
	60 to 69	292	38.5	15.1
	70 to 79	152	20.1	17.8

^aIncludes windows, sunroofs, and convertible tops.

Table 11-9. Frequency Distributions for Selected Variables Relating to Vehicle Characteristics.

Variable	Category	Number of valid cases	Percent of total valid cases	Percent of vehicles in category with one or more open windows ^a
Vehicle Type	Passenger car (hardtop)	395	52.1	20.0
	Minivan/passenger van	55	7.3	16.4
	Sport utility vehicle (SUV)	160	21.1	10.6
	Pickup truck	127	16.8	22.1
	Commercial truck	8	1.1	25.0
	Convertible	13	1.7	38.5
Window configuration	2 windows	234	30.9	22.2
	2 windows + sunroof	14	1.8	21.4
	2 windows + convertible top	9	1.2	22.2
	3 windows	14	1.8	35.7
	4 windows	444	58.6	14.4
	4 windows + sunroof	37	4.9	29.7
	4 windows + convertible top	4	0.5	75.0
	Other	2	0.3	0
Color	White	180	23.7	17.2
	Black	210	27.7	22.9
	Silver/gray	139	18.3	14.4
	Red	105	13.9	15.2
	Blue	62	8.2	16.1
	Green	39	5.1	25.6
	Brown	3	0.4	33.3
	Tan	16	2.1	25.0
	Yellow	1	0.1	0
	Other	3	0.4	0

^aIncludes windows, sunroofs, and convertible tops.

due to window reflections, tinted windows, and dark vehicle interiors. Researchers had particular difficulty in seeing rear seat occupants.

Each vehicle in the master database was assigned a value for the variable OPEN in which OPEN = 1 indicated that one or more windows were fully or partially open, a sunroof was open, and/or a convertible top was down. OPEN = 0 indicated the vehicle was completely closed with respect to windows, sunroofs, and convertible tops. Hereafter, vehicles with OPEN values equal to 1 and 0 will be referred to as "open vehicles" and "closed vehicles", respectively. The far-right columns in Tables 11-8 and 11-9 indicate the percentage of open vehicles in each category. The overall rate (n = 758) for this statistic was 18.5 percent.

Table 11-10 provides frequency distributions and open vehicle statistics for categories defined by vehicle type and window configuration. Table 11-11 provides similar results for categories defined by HI and vehicle speed. Results for categories defined by location and date are presented in Table 11-12.

Researchers reviewed the results in Tables 11-8 through 11-12 with the goal of identifying variable categories associated with open-vehicle percentages that were noticeably smaller or larger than the average value for all vehicles (18.5 percent). Based on this review, researchers defined the 16 variables in Table 11-13 as candidate predictor variables for stepwise linear regression (SLR) analyses in which OPEN was the dependent variable. Table 11-14 provides results of the SLR analyses for all vehicles and for six subsets of vehicles based on vehicle type. Table 11-15 presents SLR results for vehicles grouped by videotaping location.

Each SLR analysis used all variables in Table 11-13 as candidate predictor variables, with the exception of any variable that had the same value for all vehicles in the data classification. In reviewing the results of each analysis, researchers checked appropriate diagnostics provided in the SLR outputs (e.g., the variance inflation factor) to determine whether the pool of candidate variables exhibited significant collinearity. In performing the SLR analyses, analysts specified p-to-enter and p-to-exit values of 0.05 for adding and removing variables, consistent with recommendations by Draper and Smith (1981).

The first listing in Table 11-14 provides results for a SLR analysis performed on all 758 vehicles in the master database. The four parameters selected into the regression equation are listed according to the order in which they were selected. The table lists the regression coefficient of each parameter and indicates the cumulative R² value. Definitions of the parameters can be found in Table 11-13. The resulting regression equation can be expressed as

$$\text{OPEN} = 0.184 - (0.115)(\text{TIME12}) - (0.101)(\text{SUV}) + (0.077)(\text{BLACK}) \\ + (0.083)(\text{TIME16}) + e(11-4)$$

in which e is the residual term.

The sign of each regression coefficient provides a directional indicator of the associated variable's effect on OPEN, with the caveat that the value of the regression coefficient may vary according to the other variables included in the regression equation. Factors tending to decrease OPEN when all vehicles are considered (i.e., n = 758) include TIME12 (time between 1200 and 1259) and SUV (vehicle is a sport utility vehicle). Factors tending to increase OPEN include BLACK (vehicle color is black) and

Table 11-10. Frequency Distributions for Specific Combinations of Vehicle Type and Window Configuration.

Category		Number of valid cases	Percentage of total valid cases	Percent of vehicles in category with one or more open windows ^a	Percent of vehicles in category with open sunroof or convertible top
Vehicle Type	Window configuration				
Passenger car (hardtop)	2 windows	62	8.2	29.0	--
	2 windows + sunroof	13	1.7	23.1	15.4
	4 windows	288	38.0	16.7	--
	4 windows + sunroof	32	4.2	31.3	21.9
Minivan/ passenger van	2 windows	45	5.9	15.6	--
	3 windows	7	0.9	28.6	--
	4 windows	2	0.3	0	--
	Other	1	0.1	0	--
Sport utility vehicle	2 windows	7	0.9	14.3	--
	2 windows + sunroof	1	0.1	0	0
	4 windows	147	19.4	10.2	--
	4 windows + sunroof	5	0.7	20.0	20.0
Pickup truck	2 windows	112	14.8	21.4	--
	3 windows	7	0.9	14.3	--
	4 windows	7	0.9	0	--
	Other	1	0.1	42.9	--
Commercial truck	2 windows	8	1.1	25.0	--
Convertible	2 windows + conv. top	9	1.2	22.2	22.2
	4 windows + conv. top	4	0.5	75.0	75.0
All vehicles		758	100.0	18.5	--

^aIncludes windows, sunroofs, and convertible tops.

Table 11-11. Frequency Distributions for Specific Combinations of Heat Index and Vehicle Speed.

Category		Number of valid cases	Percentage of total valid cases	Percent of vehicles in category with one or more open windows ^a
Heat Index, degrees F.	Vehicle speed, mph			
80	60 to 69	79	10.4	21.5
	70 to 79	70	9.2	15.7
85	40 to 49	60	7.9	23.3
87	40 to 49	92	12.1	17.4
97	30 to 39	48	6.3	29.2
	60 to 69	55	7.3	16.4
99	50 to 59	9	1.2	0.0
	60 to 69	158	20.8	11.4
	70 to 79	82	10.8	19.5
100	30 to 39	16	2.1	31.3
	40 to 49	15	2.0	33.3
101	30 to 39	54	7.1	22.2
	50 to 59	20	2.6	15.0
All vehicles		758	100.0	18.5

^aIncludes windows, sunroofs, and convertible tops.

Table 11-12. Frequency Distributions for Specific Combinations of Location, Date, and Time of Day.

Category			Number of valid cases	Percentage of total valid cases	Mean speed, mph	Mean HI, degrees F.	Percent of vehicles in category with one or more open windows ^a
Location	Date	Clock hour					
Katy Mills	9/15/00	1100 to 1159	66	8.7	66	100	12.1
		1200 to 1259	86	11.3	66	100	7.0
		1800 to 1859	152	20.1	67	99	19.1
	9/16/00	1100 to 1159	149	19.7	70	80	18.8
Katyland Drive	9/15/00	1600 to 1659	122	16.1	40	100	27.1
		1700 to 1759	31	4.1	39	101	19.4
Woodlands	9/16/00	1300 to 1359	152	20.1	47	87	19.7
All vehicles			758	100.0	58	93	18.5

^aIncludes windows, sunroofs, and convertible tops.

Table 11-13. Variables Created for Stepwise Linear Regression Analyses.

Variable	Conditions when variable equals 1 (variable equals 0 otherwise)	Percent of vehicles with one or more open windows ^a	
		When variable equals 1	When variable equals 0
BLACK	Vehicle color = black	22.9	16.8
CONVERT	Convertible	38.5	18.1
EARTHSTONE	Vehicle color is green, brown, or tan	25.9	17.9
KATYLAND	Location = Katyland Drive	25.5	16.7
KATYMILLS	Location = Katy Mills	15.7	22.6
SATURDAY	Date = 9/16/00 (Saturday)	19.3	17.9
SILVER	Vehicle color is silver or gray	14.4	19.4
SLOWHOT	Speed < 50 mph and HI > 99.9 degrees	25.9	17.5
SPEEDLT50	Speed < 50	23.2	15.6
SUV	Sport utility vehicle	10.6	20.6
HIGT99	HI > 99.9 degrees	23.8	17.6
TIME12	Time between 1200 and 1259	7.0	19.9
TIME16	Time between 1600 and 1659	27.1	16.8
VEHWSUNROOF	Vehicle with sunroof	27.5	17.8
WHITE	Vehicle color is white	17.2	18.9
WIN3	Vehicle has three windows	35.7	18.2

^aIncludes windows, sunroofs, and convertible tops.

Table 11-14. Results of Stepwise Linear Regression Analyses by Vehicle Type When Dependent Variable = OPEN Using Candidate Predictor Variables in Table 11-13 as Appropriate.

Vehicle types included in analysis	Sample size (n)	Regression results ^a		
		Variables included in regression equation	Regression coefficient	Cumulative R ²
All vehicles	758	Constant	0.184	0.0000
		TIME12	-0.115	0.0112
		SUV	-0.101	0.0222
		BLACK	0.077	0.0290
		TIME16	0.083	0.0350
Passenger car (hard top)	395	Constant	0.172	0.0000
		BLACK	0.114	0.0152
Minivan/passenger van	55	Constant	0.104	0.0000
		SLOWHOT	0.467	0.1772
Sport utility vehicle	160	Constant	0.049	0.0000
		SPEEDLT50	0.159	0.0607
Pickup truck	127	Constant	0.255	0.0000
		TIME12	-0.255	0.0437
Commercial truck	8	Constant	0.250	0.0000
Convertible	13	Constant	0.384	0.0000

^aDependent variable is OPEN for all analyses. OPEN equals 1 when one or more windows, a sunroof, and/or a convertible top is open. OPEN equals 0 otherwise.

Table 11-15. Results of Stepwise Linear Regression Analyses by Location When Dependent Variable = OPEN Using Candidate Predictor Variables in Table 11-13 as Appropriate.

Location	Sample size (n)	Regression results ^a		
		Variables included in regression equation	Regression coefficient	Cumulative R ²
Katy Mills	453	Constant	0.209	0.0000
		SUV	-0.140	0.0250
		TIME12	-0.110	0.0390
Katyland Drive	153	Constant	0.281	0.0000
		SILVER	-0.226	0.0279
Woodlands	152	Constant	0.117	0.0000
		BLACK	0.255	0.0572
		EARTHTONE	0.241	0.0869

^aDependent variable is OPEN for all analyses. OPEN equals 1 when one or more windows, a sunroof, and/or a convertible top is open. OPEN equals 0 otherwise.

TIME16 (time between 1600 and 1659). Note that the regression model based on these four variables explains less than 4 percent of the total variability in the OPEN values ($R^2 = 0.0350$). While the effect of each of the four variables on OPEN is statistically significant, the regression equation does not provide a very powerful model for predicting whether a particular vehicle in the database will be open or closed.

Table 11-14 also provides results of SLR analyses performed on subsets of vehicles defined by vehicle type. The regression equation for passenger car (hardtop) includes only one variable (BLACK), which has a positive coefficient. SLOWHOT (vehicle speed is less than 50 mph and HI is greater than 99.9 degrees F) is the only variable appearing in the regression equation for minivan/passenger vans; the variable has a positive coefficient. The R^2 value for this equation (0.1772) is the largest value listed in Table 11-14.

The regression equation for sport utility vehicles also includes a single variable (SPEEDLT50). The regression coefficient is positive, indicating that these vehicles tend to open windows when speed is less than 50 mph. The effect is quite weak, however, as the R^2 value is only 0.0607. The R^2 value for pickup trucks is even lower (0.0437). The regression equation for pickup trucks includes TIME12 with a negative coefficient. No predictor variables were identified by the SLR analyses performed on commercial trucks or convertibles, possibly because of the small sample sizes available for these two vehicle types.

Table 11-15 provides results of SLR analyses specific to videotaping location. The regression equation for Katy Mills includes the SUV and TIME12 variables, both with negative coefficients. The Katyland Drive equation includes a single variable (SILVER = vehicle color is silver or gray). The Woodlands equation contains two variables (BLACK and EARTHSTONE), both with positive coefficients. The R^2 values for all three regression equations are low, ranging from 0.0250 (Katy Mills) to 0.0869 (Woodlands).

Overall, the SLR results suggest that the probability of a vehicle in the master database being "open" is weakly affected by time of day, vehicle type, vehicle color, vehicle speed, and HI. It is important to note, however, that the predictor variables selected for inclusion in the regression equation vary with vehicle type and videotaping location.

11.2.2.3 Discussion and Recommendations

The vehicles included in the Houston pilot study represent a set of special conditions that should be carefully considered in evaluating the study results. The vehicles were videotaped at locations on the perimeter of the Harris County in or near areas not subject to I&M. Consequently, the proportion of older vehicles with poorly maintained or non-operating air conditioners was likely higher than would have been observed in central portions of Harris County where vehicles are subject to I&M.

HI varied from 80 to 101 degrees F during the videotaping sessions. It would be expected that most drivers with working air conditioning systems would have the air conditioner on and windows closed under these conditions. The exception would be cases in which the driver had windows open to cool a vehicle during the first few minutes of driving. Such cases may have occurred more frequently at the Katyland

Drive site where taping occurred late in the afternoon in an area near schools and light industry.

We would expect windows to be open more frequently when HI falls between 65 and 75 degrees, F. As the HI falls below 65 degrees F, we would expect windows to be closed more frequently as drivers begin to use their heaters. To test these assumptions, researchers plan to conduct a follow-up study during 2001 in which vehicles will be videotaped under a full range of HI values.

The database includes vehicle speeds between 30 and 74 mph. In general, we would expect passengers to open windows at lower speeds because of the reduced flow-through ventilation. They would tend to close windows at higher speeds to reduce wind noise and buffeting. The results of this pilot study weakly support these expectations, as open windows showed a slight tendency to occur more frequently when speeds fell within the 30 to 49 mph range, particularly when HI exceeded 99.9 degrees F and the vehicle was a minivan or passenger van. In the planned follow-up study, researchers will attempt to obtain window position data for vehicles moving at speeds below 30 mph.

Field staff noticed a number of drivers knocking ash from cigarettes through a partially open driver's window. In other cases, field staff observed drivers smoking in vehicles with a partially open driver's window, possibly to allow secondary smoke to escape. This anecdotal evidence indicates that vehicles containing a smoker may have a higher proportion of partially open windows than the general population, especially under conditions expected to favor the use of air conditioning (e.g., high HI).

One reason for uncertainty in determining window and sunroof positions was the presence of visors or shields near the edge of the windows and sunroofs. These structures are designed to provide protection from wind at high speeds. On the videotapes, the effect of the visors is to obscure a portion of the window or sunroof from view, preventing researchers from determining whether the window or sunroof is completely closed or partially open. Fully-open windows and sunroofs are not affected by the visors. Therefore, vehicles excluded from the statistical analysis because of window or sunroof visors are likely to have a higher proportion of closed and partially open windows than vehicles included as valid cases.

Overall, the pilot study demonstrated that data on factors affecting vehicle window position could be acquired through a relatively simple experimental protocol using a single video camera. Limitations of the study requiring further research include the inability to determine the status of the vehicle air conditioning system; lack of a wide range of weather, vehicle speed, and road type conditions; the potential effect of smoking on window position, and the need to exclude some vehicles from statistical analyses due to ambiguous window positions. Keeping in mind these limitations, the results of the study may be useful to researchers who are constructing or refining algorithms for estimating window position (and associated air exchange rate) in motor vehicle microenvironments. The results can be briefly summarized as follows:

Of the 758 vehicles included in the database, 140 (18.5 percent) were labeled as "open," indicating a window, sunroof, or convertible top was fully or partially open. The results of a series of stepwise linear regression analyses indicate that the probability of a vehicle in the master database being "open" was weakly affected by time of day, vehicle type, vehicle color, vehicle speed, and HI. In

particular, open windows occurred more frequently when vehicle speed was less than 50 mph during periods when HI exceeded 99.9 degrees F and the vehicle was a minivan or passenger van.

PRESENT POSSIBLE ALGORITHM — CONSIDER COMBINING WITH CHAPEL HILL RESULTS TO BE SUMMARIZED IN SUBSECTION 11.2.3.

11.2.3 Motor Vehicles (Chapel Hill, NC)

field study not complete

11.3 Air Exchange Rates in Residences

field study not complete

11.4 Variation in Ozone Concentration with Height Above Ground and Distance from Roadways

11.4.1 Introduction

Over 1500 fixed-site ozone monitors are listed in the Aerometric Information Retrieval System (AIRS) maintained by the U.S. Environmental Protection Agency. Ozone concentrations measured by currently active monitors are routinely used to determine the attainment status of defined urban areas with respect to the National Ambient Air Quality Standard (NAAQS) for ozone. In addition, researchers frequently use ozone data obtained from selected monitors to represent outdoor (ambient) ozone concentrations in applications of the probabilistic NAAQS Exposure Model for ozone (pNEM/O3) and similar computer-based exposure models. Both uses of fixed-site monitoring data are based on the implicit assumption that the data provide an unbiased estimate of the ozone concentrations in air respired by people at outdoor locations near the monitor. This assumption may not be justified, as many monitor probe inlets are located at heights greater than 2 m, the upper bound for the breathing zone of a typical pedestrian.

To evaluate the effect of probe height on measured ozone concentration, IT Air

Quality Services (ITAQS) performed a series of vertical measurements of ozone concentration in the lower 80 meters of the surface boundary layer. The measurements were made during August 1995 at an inoperative radio tower located in Cincinnati, Ohio. A report by Wisbith, Meiners, Ollison, and Johnson (1996) describes the procedures used in the study, summarizes the collected data, and provides an evaluation of the results. Subsection 11.4.2 summarizes the results of this study and presents supplemental analyses conducted by TRJ Environmental (TRJ) that indicate ozone concentrations measured at 10 meters and above tend to be higher than concentrations measured at 2 meters.

A second study conducted by ITAQS provides data useful in evaluating the effects of probe height and horizontal position. During the 1995 phase of the Cincinnati Roadside Ozone Study, ITAQS measured ozone concentrations at 2 and 10 meter elevations at various distances from selected roadways. Subsection 11.4.3 summarizes the results of this study and presents supplemental analyses of the data performed by TRJ that indicate that the vertical ozone gradient varies with distance from roadways.

11.4.2 The Cincinnati Tower Study

11.4.2.1 Test Procedures

The Cincinnati Tower Study was conducted on August 16, 1995 at the Bethany Relay Station of the Voice of America. The Bethany Station is located 20 miles north of downtown Cincinnati and is situated midway between two ozone monitoring stations operated by the Southwest Ohio Air Pollution Control Agency. These ozone monitors typically report the highest ozone concentrations measured in the Greater Cincinnati area. The Bethany Station consists of one square mile of open, level land, with numerous inoperative radio towers. Testing was conducted on a 91-meter (300-foot) tower located near the center of the facility. This rural location provided a grassy, unobstructed test area 0.5 mile from the nearest roadway. There were no other known sources of ozone precursors or scavengers within 0.5 mile of the site.

During testing, atmospheric measurements were made at inlet probe heights of 2, 10, 20, 40, and 80 meters above ground level, providing a progression of probe heights between 2 and 80 meters. Carbon monoxide (CO) and temperature measurements were also made at each test height.

A pulley-mounted instrument skip was used to move the measurement equipment between the sampling points on the tower. The instrument skip rider pulley trolleyed up and down a one-inch rope that served as both support rope and drive rope. The support rope was attached to the rider pulley, passed up through a pulley attached near the top of the tower, back down through the rider pulley, and attached to the rear of a Ford Explorer. As the vehicle moved away from the tower, the instrument skip trolleyed up the support rope, permitting the instrument to be moved between points in 60 seconds or less. The vehicle engine was operated only during skip movement.

To assure maximum data capture and minimize instrument bias, duplicate sets of instruments were used to measure ozone, CO, and temperature. Ozone concentrations were measured using two portable ozone analyzers: an Analytical Instruments Development, Inc. (AID) Model 560 and a Thermo Environmental Instruments (TECO) Model 560. Both instruments utilize the vapor-phase chemiluminescent reaction of

ozone and ethylene for the measurement of ozone concentrations.

11.4.2.2 Test Results

Six vertical gradient measurements (soundings) were made during the study. Each sounding included measurement periods of five or six minutes at each of five sequential heights (2, 10, 20, 40, and 80 m). The six soundings or "runs" were performed in sequence such that the Run A measurements were made as the equipment skip ascended, Run B as the skip descended, Run C as the skip again ascended, and so forth. Table 11-16 lists the average ozone value measured at each height during each run. The values were calculated by averaging the three central one-minute ozone concentration values measured during each five or six-minute period when the skip was stationary at a specified height. This procedure ensured that any values measured during movement of the skip between levels were omitted from the analysis.

A comparison of data obtained from the AID and TECO monitors showed that the two monitors tracked each other closely. However, as the post-test calibration span data for the TECO monitor were lost through equipment failure, the ozone concentration values listed in Table 11-16 were calculated using the AID monitor data only.

Table 11-16 lists the value of the quantity

$$\text{DIFF}(h) = \text{ozone}(h) - \text{ozone}(2 \text{ m}) \quad (11-5)$$

by run for each height h between 10 and 80 meters. In this equation, $\text{ozone}(h)$ is the ozone concentration measure at height h during a particular run and $\text{ozone}(2 \text{ m})$ is the ozone concentration measured at 2 meters during the same run. Of the 24 values of DIFF listed in Table 11-16, 22 (92 percent) are positive, indicating that ozone concentrations at heights of 10 m and greater tend to exceed the ozone concentration at 2 meters. The DIFF values range from +53 ppb to -2 ppb. The minimum, median, and maximum values of DIFF are listed by height in Table 11-17.

Table 11-16 also lists the value of the quantity

$$\text{RATIO}(h) = [\text{ozone}(2 \text{ m})]/[\text{ozone}(h)] \quad (11-6)$$

by run for each height h between 10 and 80 meters. Consistent with the values calculated for DIFF, 22 of the RATIO values are less than 1.00. Table 11-17 lists the minimum, median, and maximum values of RATIO by height.

Figure 3 of the report by Wisbith, Meiners, Ollison, and Johnson (1996) provides a plot of the ozone concentrations measured at each height by run. The following patterns can be observed in the plotted data.

1. The 2-meter concentration tends to be the lowest concentration measured during each run.
2. The ordering of the concentrations measured at 10 through 80 meters

Table 11-16. Three-minute average ozone concentrations measured at specified heights above ground by the Cincinnati Tower Study.

Run	Height above ground, m	Start time of three-minute period	Ozone conc., ppb	DIFF ^a , ppb	RATIO ^b
A	2	1309	111	---	---
	10	1314	118	7	0.941
	20	1319	120	9	0.925
	40	1325	134	24	0.828
	80	1331	139	29	0.799
B	2	1359	82	---	---
	10	1354	97	15	0.845
	20	1349	131	49	0.626
	40	1343	135	53	0.607
	80	1336	135	53	0.607
C	2	1416	89	---	---
	10	1421	104	15	0.856
	20	1426	106	17	0.840
	40	1431	115	26	0.774
	80	1437	125	36	0.712
D	2	1502	113	---	---
	10	1457	124	11	0.911
	20	1452	120	7	0.942
	40	1447	123	10	0.919
	80	1442	120	7	0.942
E	2	1513	131	---	---
	10	1518	162	31	0.809
	20	1523	129	-2	1.016
	40	1528	170	39	0.771
	80	1534	129	-2	1.016

Run	Height above ground, m	Start time of three-minute period	Ozone conc., ppb	DIFF ^a , ppb	RATIO ^b
F	2	1559	114	---	---
	10	1554	137	23	0.832
	20	1549	136	22	0.838
	40	1544	153	39	0.745
	80	1539	133	19	0.857

^aDIFF(h) = ozone(h) - ozone(2 m), h = height.

^bRATIO(h) = [ozone(2 m)]/[ozone(h)].

Table 11-17. Median values of DIFF^a and RATIO^b by height above ground.

Height above ground, m	Minimum DIFF, ppb	Median DIFF, ppb	Maximum DIFF, ppb	Minimum RATIO	Median RATIO	Maximum RATIO
10	31	15	7	0.809	0.851	0.941
20	49	13	-2	0.626	0.882	1.016
40	53	32.5	10	0.607	0.772	0.919
80	53	24	-2	0.607	0.828	1.016

^aDIFF(h) = ozone(h) - ozone(2 m), h = height.

^bRATIO(h) = [ozone(2 m)]/[ozone(h)].

tends to vary from run to run.

3. In most of the runs, clusters of values are observed at one or two concentration levels. For example, three of the ozone concentrations measured during Run B are clustered between 131 and 135 ppb. In Run D, four of the concentrations are clustered between 120 and 124 ppb. Three concentrations are clustered between 133 and 137 ppb in Run F.
4. In the five runs in which the 2-meter concentration is less than 120 ppb (i.e., Runs A, B, C, D, and F), at least one of the concentrations measured between 10 m and 80 m exceeds 120 ppb.

The fourth observation is particularly important. The data in Figure 3 show that ozone measurements made at elevations between 10 and 80 meters may indicate ozone concentrations exceed the current one hour NAAQS level (120 ppb) when ozone concentrations at the 2-meter "breathing zone" are significantly less than 120 ppb.

11.4.2.3 Supplemental Analyses of Tower Data

TRJ performed supplemental analyses of the data obtained from the Cincinnati Tower Study. Table 11-18 presents results of these analyses. Consistent with the discussion above, $RATIO(h)$ is defined as the ozone concentration at 2 meters divided by the concentration at height h (both values measured during the same run).

Table 11-18. Results of regression analyses in which the dependent variable was the ozone concentration at 2 meters and the predictor variable was ozone concentration at indicated height.

Probe height, m	Mean value of $RATIO(h)^a$	Regression results			
		Intercept	Slope	R ² value	p value
10	0.8656	17.02	0.725	0.898	0.004
20	0.8643	39.30	0.545	0.105	0.532
40	0.7740	20.92	0.620	0.479	0.128
80	0.8220	140.00	-0.256	0.010	0.853

^aSee Equation 11-4.

Table 11-18 lists the arithmetic means values of $RATIO(h)$ by probe height. There is not much difference between the means for 10 m and 20 m. In fact, there is not much difference between any of the means. To test this observation, TRJ performed an analysis of variance on the $RATIO(h)$ means by height and found that height in the range of 10 to 80 meters did not have a statistically significant effect on the mean values ($p = 0.486$). TRJ also performed a student t test comparing the highest and lowest $RATIO$ means (10 m and 40 m). The difference between these two means was almost, but not quite, significant ($p = 0.08$).

Table 11-18 also lists results of regression analyses (n = 6) in which the dependent variable was the ozone concentration at 2 meters and the predictor variable was the ozone concentration at the indicated height. The results indicate that the ozone concentrations at 10 and 40 meters are the best predictors based on a comparison of the R² values listed in Table 11-18. However, only the regression equation for 10 m has a significant R² value (p < 0.05). The ozone concentration at 80 meters is a very poor predictor of 2 meter ozone (R² = 0.010, p = 0.853). This is a significant result, as it suggests that concentrations measured by probes higher than 10 m are poor indicators of 2-meter ozone.

TRJ also performed a series of regression analyses in which the dependent variable was a function of the concentration at 2 meters and the predictor variables were functions of (1) concentration at height h and (2) height h, where h was not equal to 2 meters. The analyses were performed using data for all heights (i.e., 10 through 80 meters) and using only data for 10 and 20 meters. Height was found to be a poor predictor in all cases.

The results of the Cincinnati Tower Study suggest that a majority of the observed decrease in ozone concentration occurred between 2 m and 10 m. The shape of this gradient cannot be determined, however, as no ozone measurements were made between 2 and 10 meters. A reasonable, conservative model for the effect is to assume that RATIO(h) falls off in a linear manner between 2 and 10 meters and then is constant at 10 meters and above. The formula for this adjustment can be expressed as

$$\text{RATIO}(h) = 1 - (0.01675)(h - 2) \qquad 2 \leq h \leq 10 \text{ m} \qquad (11-7)$$

$$\text{RATIO}(h) = 0.866 \qquad h > 10 \text{ m} \qquad (11-8)$$

Note that the adjustment factor for a height that exactly equals 10 meters is 0.866, a reduction of 13.4 percent.

11.4.2.4 Sample Application of Tower Study Results to pNEM/O₃ Analyses

Table 11-19 lists the 12 monitors that provided ambient ozone data for a recent application of pNEM/O₃ to outdoor workers in Chicago (Johnson, Capel, McCoy, and Mozier, 1996). As the probe heights listed in Table 11-19 fell between 4 and 16 m, analysts assumed that the adjustment procedure proposed above could be applied to data reported by all 12 monitors.

The parameter "Adjusted ozone - Method 1" in Table 11-19 was calculated by applying the above formula to the highest one-hour ozone concentration reported by each Chicago monitor for 1991 (listed in the column labeled "1991 max. ozone conc."). The mean of the maximum values before adjustment is 124 ppb; the mean after adjustment is 112 ppb (a reduction of about 10 percent).

11.4.3 The 1995 Cincinnati Roadside Ozone Study

11.4.3.1 Test Procedures and Results

Another recent study provides data useful in evaluating the effects of monitor

probe height. As part of the 1995 Cincinnati Roadside Ozone Study described by Johnson, Pakrasi, Wisbith, Meiners, and Ollison (1995), ozone measurements were made at two heights (2 and 10 meters) by ozone monitors in a mobile van that was parked at various downwind distances from selected roadways. Tables 11-20 through 11-28 list mean values of one-minute ozone and nitric oxide (NO) concentrations measured at each distance during each of nine tests. Most of the mean values are based on 10 or more one-minute values; the values in parentheses were calculated from less than 10 one-minute values and are not used in the analyses discussed hereafter.

Each table provides values for the ratio of ozone concentrations measured at 2 and 10 meters. Table 11-29 summarizes these data by listing the ratio values from all nine tests according to distance from roadway. The far right column in Table 11-29 lists the arithmetic mean of the values in each row, i.e., the mean of the values reported for the specified distance. (The individual ratio values in parentheses were ignored when calculating the row means.)

Listed below are ranges and means of the row means in Table 11-29 when classified into three distance ranges with approximately equal sample sizes.

- o Range A: 25 to 86 feet (7.6 to 26.2 meters)
range: 0.738 to 0.847 (n = 7)
mean: 0.815
- o Range B: 99 to 182 feet (30.2 to 55.5 meters)
range: 0.836 to 1.031 (n = 6)
mean: 0.937
- o Range C: 222 to 546 feet (67.7 to 166.4 meters)
range: 0.882 to 1.186 (n = 7)
mean: 0.985

In general, the mean ratio value increases as the distance from the road increases. The mean of all 20 arithmetic means is 0.911.

Fourteen of the Chicago monitor locations listed in Table 11-29 are classified by road type (local, major, arterial, or expressway). The reported traffic counts (vehicles/day) are listed below by roadway type with calculated median values.

Local: 150, 500, 1000, 2000, 8000 (median = 1000)
Major: 3000, 6500, 8250 (median = 6500)
Arterial: 3,400, 13,800, 16,800, 30,000 (median = 15,300)
Expressway: 37,000, 100,200 (median = 68,600)

As expected, median traffic counts increase as one moves down the list from local roads to expressways.

The reported distance to nearest roadway of each monitors is listed below by

Table 11-19. Characteristics of Chicago ozone monitoring sites and 1991 maximum ozone concentrations.

District	Site ID	Probe height, m	Distance to roads near monitor, m	Traffic, vehicles per day	Road type	1991 max. ozone conc., ppb	Adjusted ozone - Method 1, ppb ^a	Adjusted ozone - Method 2, ppb ^b
1	0001	10	160	8250	major	108	94	88
			20	1000	local			
2	0032	14	300	6500	major	120	104	118
3	1003	16	65 m	4800	local	134	116	126
			65 m	5000	through			
			30 m	146,000	I-90 express			
4	1601	4	40	500	local	152	147	150
5	4002	4	150 m	5000	through	125	121	124
6	4003	11	34 m	7100	major	119	103	112
7	7002	10	24	2000	local	123	107	100
8	8003	4	1000	100,200	expressway	109	105	104
			700	16,800	arterial			
			13	3,400	arterial			
9	6001	14	80	30,000	arterial	118	102	116
			450	37,000	expressway			
10	0005	7	52	13,800	arterial	128	117	123
11	0001	5	198	3000	major	124	118	121
			37	150	local			
12	1002	11	27	8000	local	126	109	103

^aAdjustment according to Equations 11-5 and 11-6. ^bAdjustment according to Equations 11-7 through 11-12.

Table 11-20. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 11 on August 10, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
35	1555	56	11/11	11/11	82.455	93.818	0.879	9.909	4.182	2.369
61	1530	44	11/11	11/11	91.182	95.727	0.953	3.364	2.818	1.194
105	1505	39	10/12	10/12	90.000	90.667	0.993	2.200	1.250	1.760
182	1438	38	10/13	10/13	80.800	77.692	1.040	6.200	3.539	1.752
315	1413	48	10/11	10/11	68.100	73.091	0.932	15.300	4.000	3.825
546	1350	43	10/6	10/6	77.700	(79.667)	(0.975)	-0.600	(-0.667)	---

Table 11-21. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 12 on August 10, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
35	1904	43	10/10	10/10	33.900	55.200	0.614	22.400	7.800	2.872
61	1840	36	10/11	10/11	44.400	53.182	0.835	11.800	7.546	1.564
105	1816	59	11/7	11/7	53.727	57.571	0.933	7.000	10.857	0.645
182	1750	51	11/11	11/11	57.364	67.545	0.849	3.273	1.182	2.769
315	1736	45	10/11	10/11	65.100	66.091	0.985	0.800	1.455	0.550
546	1701	41	11/11	11/11	67.273	71.455	0.941	1.273	0.818	1.556

Table 11-22. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 21 on August 11, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
35	1543	75	12/11	12/11	70.333	93.455	0.753	58.583	48.182	1.216
61	1517	72	11/13	11/13	65.273	89.231	0.732	54.182	54.923	0.987
122	1452	63	10/11	10/11	79.500	95.182	0.835	59.100	52.909	1.117
182	1429	67	11/10	11/10	87.636	86.300	1.015	8.000	-17.600	---
314	1405	61	10/11	10/11	91.100	85.182	1.069	-20.100	-19.818	---
544	1340	66	11/11	11/11	84.273	79.000	1.067	-18.273	-16.727	---

Table 11-23. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 22 on August 11, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
35	1812	83	12/14	12/14	64.750	93.786	0.690	28.167	46.357	0.608
61	1749	59	10/11	10/11	73.600	78.909	0.933	52.900	19.091	2.771
122	1725	68	11/11	11/11	66.818	72.364	0.923	32.273	45.182	0.714
182	1703	62	10/10	10/10	72.700	77.400	0.939	46.600	12.500	3.728
314	1637	70	14/10	14/10	80.071	80.600	0.993	25.214	41.300	0.611
544	1612	53	12/11	12/11	86.167	93.273	0.924	21.417	-2.546	---

Table 11-24. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 31 on August 12, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
35	1601	62	12/10	12/10	77.750	103.10	0.754	35.083	3.200	10.963
61	1538	54	11/10	11/10	73.273	101.90	0.719	17.909	3.100	5.777
122	1513	47	11/12	11/12	83.818	93.750	0.894	9.546	4.167	2.291
182	1450	69	11/10	11/10	78.636	86.900	0.905	7.546	3.600	2.096
314	1424	46	11/11	11/11	82.727	94.545	0.875	4.455	2.727	1.634
544	1400	49	12/10	12/10	83.833	86.045	0.974	4.750	-8.100	---

Table 11-25. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 41 on August 14, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
25	1558	30	12/13	12/13	96.750	118.69	0.815	1.583	-14.308	---
43	1532	26	11/11	11/11	107.18	125.45	0.845	-11.364	-11.455	---
86	1504	26	4/11	4/11	(124.00)	140.18	(0.885)	(-7.750)	-13.273	---
129	1440	28	12/10	12/10	133.75	140.30	0.953	-13.917	-13.600	---
222	1414	19	13/11	13/11	138.77	140.73	0.986	-15.308	-14.364	---
385	1345	24	12/14	12/14	127.67	129.93	0.983	-15.583	-14.929	---

Table 11-26. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 42 on August 14, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
25	1926	21	12/10	12/10	67.250	79.500	0.846	-12.083	-18.100	---
43	1904	28	10/10	10/10	64.700	85.900	0.753	-8.000	-18.500	---
86	1840	29	11/10	11/10	85.182	104.30	0.817	-14.818	-16.900	---
129	1817	34	11/10	11/10	108.27	118.90	0.911	-18.545	-17.300	---
222	1755	34	10/10	10/10	118.30	126.40	0.936	-17.100	-15.500	---
385	1732	35	11/10	11/10	90.273	107.30	0.841	-14.364	-14.700	---

Table 11-27. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 51 on August 17, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
40	1522	48	11/11	11/11	88.727	104.73	0.847	34.182	10.909	3.133
63	1458	61	11/11	11/11	85.455	102.18	0.836	33.727	12.273	2.748
99	1433	45	11/11	11/11	91.364	95.909	0.953	21.091	13.182	1.600
156	1409	40	11/11	11/11	81.909	96.182	0.852	19.818	8.818	2.247
245	1345	58	11/11	11/11	90.818	95.545	0.951	10.455	10.182	1.027
385	1320	53	11/11	11/11	108.09	112.73	0.959	7.091	6.455	1.099

Table 11-28. Mean ozone and nitric oxide (NO) concentrations measured during Test No. 52 on August 17, 1995 in Cincinnati, Ohio.

Distance from roadway, feet	Start time	Vehicles counted during interval	2 m/10 m ozone sample sizes	2 m/10 m NO sample sizes	Mean 2-m ozone conc., ppb	Mean 10-m ozone conc., ppb	Ratio of 2-m to 10-m ozone means	Mean 2-m NO conc., ppb	Mean 10-m NO conc., ppb	Ratio of 2-m to 10-m NO means
40	1747	58	11/4	11/4	49.455	(77.500)	(0.638)	43.091	(2.000)	(21.546)
63	1723	63	0/6	0/6	---	(76.667)	---	---	(13.33)	---
99	1700	58	11/10	11/10	70.909	84.800	0.836	21.364	10.400	2.054
156	1636	60	11/11	11/11	77.182	63.818	1.209	13.455	17.364	0.775
245	1612	66	11/11	11/11	75.545	53.182	1.420	9.000	21.273	0.423
385	1548	50	11/11	11/11	65.636	88.091	0.745	14.636	8.636	1.695

roadway type with calculated median values for each group.

Local: 20, 40, 24, 27, 37 (median = 27 meters)

Major: 160, 198, 300 (median = 198 meters)

Arterial: 13, 52, 80, 700 (median = 66 meters)

Expressway: 450, 1000 (median = 725 meters)

With the exception of the arterial roads, median separation distance tends to increase as we move down the list from low-traffic roadways to high-traffic roadways (i.e., from local roads to expressways). Surprisingly, three of the Chicago ozone monitors are located less than 100 meters from arterial roads.

TRJ used these results to develop three formulae that can be used for adjusting ozone data with respect to probe height and distance to nearby roadways.

Formula A: monitor to road distance = 3 to 30 meters

$$\text{RATIO}(h) = 1 - (0.02313)(h - 2) \quad 2 \leq h \leq 10 \text{ m} \quad (11-9)$$

$$\text{RATIO}(h) = 0.815 \quad h > 10 \text{ m} \quad (11-10)$$

Formula B: monitor to road distance = 31 to 67 meters

$$\text{RATIO}(h) = 1 - (0.00788)(h - 2) \quad 2 \leq h \leq 10 \text{ m} \quad (11-11)$$

$$\text{RATIO}(h) = 0.937 \quad h > 10 \text{ m} \quad (11-12)$$

Formula C: monitor to road distance = 68+ meters

$$\text{RATIO}(h) = 1 - (0.00188)(h - 2) \quad 2 \leq h \leq 10 \text{ m} \quad (11-13)$$

$$\text{RATIO}(h) = 0.985 \quad h > 10 \text{ m} \quad (11-14)$$

Each formula assumes that RATIO(h) decreases from 1 to the indicated value as h increases from 2 meters to 10 meters. Above 10 meters the value of RATIO(h) equals the 10-meter value. The three 10-meter values (0.815, 0.937, and 0.985) have been set equal to the mean values determined above for Ranges A, B, and C, respectively. The choice of formula would depend on the distance of the monitor to the nearest upwind road.

The use of a constant RATIO(h) for h values above 10 is consistent with the analysis of the Cincinnati tower data discussed earlier in this section. In that analysis, researchers found that there was no statistically significant change in RATIO(h) values as the probe height varied from 10 to 80 meters. The main fall-off in ozone concentration appeared to have occurred between 2 meters and 10 meters, although the tower study did not actually measure ozone inside this range.

11.4.3.2 Sample Application of Roadside Study Results to pNEM/O3 Analyses

Table 11-30 lists the 12 monitors that provided ambient ozone data for the recent application of pNEM/O3 to outdoor workers in Chicago (Johnson, Capel, McCoy, and Mozier, 1996). The table also indicates which of the three adjustment formulae (A, B, or C) is applicable to each monitor based on the distance between the monitor and the nearest roadway. Monitors with multiple roadway listings were assigned the smallest distance value. The column in Table 11-19 labeled "Adjusted ozone - Method 2" shows the effects of applying the indicated formula to the highest 1991 one-hour value reported by each monitor.

The mean of the 1991 maximum values before adjustment is 124 ppb; the mean after adjustment by Method 2 is 115 ppb (a reduction of about 7 percent). With four exceptions, each of the adjusted values obtained from Method 2 is greater than the corresponding value obtained from Method 1.

Table 11-31 lists the 16 monitors that provide ambient ozone data for the application of pNEM/O3 to Los Angeles. Note that one site (Santa Monica) was missing data on distances to nearby roads. In this case, analysts assumed that the distance to the nearest road was 54 meters, the median value of the nearest-road distances. The distances to the nearest roadway (reported or assumed) range from 5 meters to 400 meters; all three adjustment formulas (A, B, and C) were used in determining downwind adjustment factors. As previously indicated, the upwind factor was always determined by Formula C.

11.4.4 Development of the Final Adjustment Algorithm

Researchers considered Method 2 to be superior to Method 1 because

- (1) Method 2 accounted for the effects of probe height and proximity to roadways, and
- (2) Method 2 was based on data collected at paved, urban locations representative of typical monitoring stations.

Method 1 did not account for proximity to roadways; the method was based on ozone measurements collected at least 0.5 mile from the nearest roadway in a field containing knee-high vegetation.

Researchers also assumed that Formulae A and B were only applicable to a monitoring site when the site was downwind of the nearest road. On occasions when the monitor was upwind of the nearest roadway, Formula C was substituted for Formula A or B, as the site was assumed to be at least 222 feet (68 meters) downwind from any roadway.

The two far-right columns in Table 11-30 list the downwind and upwind values of $RATIO(h)$ determined for each of the Chicago monitors. The downwind value was calculated by applying Formula A, B, or C based on the distance to the nearest roadway. The upwind value was determined using Formula C. The same procedure was used to calculate the values in the two far-right columns in Table 11-31.

TRJ incorporated these assumptions into an algorithm for adjusting the Chicago

Table 11-30. Characteristics of Chicago ozone monitoring sites used in pNEM/O3 and proposed values of RATIO(h) adjustment factor.

District	Site ID	Probe height, m	Distance to nearby roads, m	Road type	Assumed distance to nearest road, m	Downwind distance category	Downwind value of RATIO(h)	Upwind value of RATIO(h)
1	0001	10	160	major	20	A	0.815	0.985
			20	local				
2	0032	14	300	major	300	C	0.985	0.985
3	1003	16	65	local	65	B	0.937	0.985
			65	through				
			200	express				
4	1601	4	40	local	40	B	0.984	0.996
5	4002	4	150	through	150	C	0.996	0.996
6	4003	11	34	major	34	B	0.937	0.985
7	7002	10	24	local	24	A	0.815	0.985
8	8003	4	1000	express	13	A	0.954	0.996
			700	arterial				
			13	arterial				
9	6001	14	80	arterial	80	C	0.985	0.985
			450	express				
10	0005	7	52	arterial	52	B	0.961	0.991
11	0001	5	198	major	37	B	0.976	0.994
			37	local				
12	1002	11	27	local	27	A	0.815	0.985

and Los Angeles monitoring data used in pNEM/O₃ exposure assessments. To account for varying wind directions in the algorithm, researchers assumed that the downwind/upwind designation of a monitor would be determined daily with an equal probability assigned to each designation.

The resulting algorithm can be expressed by the following sequence of steps.

1. Note the downwind and upwind values of RATIO(h) in the last two columns of Table 11-30.
2. Go to first/next calendar day.
3. Draw a random number RN from a uniform (0,1) distribution.
4. If $RN \leq 0.5$, then the monitor is considered to be downwind from the nearest roadway for the entire day. Calculate adjusted one-hour concentrations for the day by the equation

Adjusted ozone conc. = [downwind RATIO(h)](unadjusted ozone conc.).

5. If $RN > 0.5$, then the monitor is considered to be upwind of all nearby roadways for the entire day. Calculate adjusted one-hour concentrations for the day by the equation

Adjusted ozone conc. = [upwind RATIO(h)](unadjusted ozone conc.).

6. End if last day complete. Otherwise, go to Step 2.

Section 11 of a report by Johnson (1997) describes the results of sensitivity analyses which used this algorithm.

11.5 References for Section 11

Bluestein, M., and J. Zecher. 1999. "A new approach to an accurate wind chill factor." Bulletin of the American Meteorological Society, 80(9):1893-1899.

Draper, N. R., and H. Smith. 1981. Applied Regression Analysis. John Wiley and Sons: New York, NY. Page 311.

International Society of Biometeorology. 2001. <http://www.med.uni-muenchen.de/arbmed/biomet/Commission6.htm>. Updated November 9, 2001. Accessed January 22, 2002.

Johnson, T., J. Capel, M. McCoy, and J. Mozier. 1996. Estimation of Ozone Exposures Experienced by Outdoor Workers in Nine Urban Areas Using a Probabilistic Version of NEM. Report prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Table 11-31. Characteristics of Los Angeles ozone monitoring sites used in pNEM/O3 and proposed values of RATIO(h).

District code	AIRS code	Location	Probe height, m	Reported distances to nearby roads, m	Road type code ^a	Assumed distance to nearest road, m ^d	Downwind distance category	Downwind value of RATIO(h)	Upwind value of RATIO(h)
1	037-0016	Glendora	4.9	54	?	54	B	0.977	0.995
				108	?				
2	037-1103	Los Angeles	12.4	71	?	71	C	0.980	0.980
3	037-1301	Lynwood	6.7	19	1	19	A	0.891	0.991
4	037-1601	Pico Rivera	6.5	65	5	65	B	0.965	0.992
				122	6				
5	037-1902	Santa Monica	4	?	?	54 ^e	B	0.984	0.996
6	037-2005	Pasadena	4	80	5	80	C	0.996	0.996
7	037-4002	Long Beach	11	5	?	5	A	0.792	0.983
8	037-5001	Hawthorne	3.6	16	?	16	A	0.963	0.997
9	059-0001	Anaheim	4.2	160	?	160	C	0.996	0.996
10	059-1001 ^b	Costa Mesa	6.1	26	?	26	A	0.905	0.992
11	059-3002	Los Alamitos	4.6	400	?	400	C	0.995	0.995
12	059-5001	La Habra	4.3	49	?	49	B	0.982	0.996
13	071-8001 ^c	Rubidoux	2	111	?	111	C	1.000	1.000
				162	?				
14	071-1004	Upland	6	69	5	69	C	0.992	0.992
				122	5				
15	071-4003	Redlands	4	18	5	18	A	0.954	0.996

District code	AIRS code	Location	Probe height, m	Reported distances to nearby roads, m	Road type code ^a	Assumed distance to nearest road, m ^d	Downwind distance category	Downwind value of RATIO(h)	Upwind value of RATIO(h)
16	071-9004	San Bernardino	5	30	?	30	A	0.931	0.994

^a1: arterial

4: major street or highway

5: through street or highway

6: local street or highway

^bCurrently listed as 059-1003.

^cCurrently listed as 065-8001.

^dDistance was assumed to be 54 meters when actual distance was not reported.

^eAssumed value (median of reported probe heights).

Johnson, T., A. Pakrasi, A. Wisbith, G. Meiners, and W. Ollison. 1995. "Ozone Exposures Within Motor Vehicles -- Results of a Field Study in Cincinnati, Ohio," Paper No. 95-WA84A.02, 88th Annual Meeting of the Air and Waste Management Association, June.

Johnson, T. 1997. Sensitivity Testing of pNEM/O3 Exposure Estimates to Changes in the Model Algorithms, Report No. FR2, American Petroleum Institute, Washington, DC., May.

Long, T., T. Johnson, and W. Ollison. 2002. Determining the Frequency of Open Windows in Motor Vehicles: a Pilot Study Using a Video Camera in Houston, Texas During High Temperature Conditions," Journal of Exposure Analysis and Environmental Epidemiology, Vol. 12, pp. 214 - 225.

National Weather Service. 2001. <http://weather.noaa.gov/weather/hwave.html>. Accessed December 7, 2001.

National Weather Service. 2002. <http://www.nws.noaa.gov/om/windchill/index.shtml>. Updated January 8, 2002. Accessed January 22, 2002.

Nelson, C.A., M. Tew, G. Phetteplace, et al. 2002. "Review of the federal interagency process used to select the new Wind Chill Temperature (WCT) Index." Paper No. P1.9, 18th International Conference on Interactive Information and Processing Systems, January.

Osczevski, R.J. 1995. "The basis of wind chill." Arctic, 48(4):372-382.

Osczevski, R.J. 2000. "Windward cooling: an overlooked factor in the calculation of wind chill." Bulletin of the American Meteorological Society, 81(12):2975-2978.

Rothfusz, L.P. 1990. The Heat Index "equation" (or, more than you ever wanted to know about Heat Index). Scientific Services Division, National Weather Service Southern Region Headquarters, Forth Worth, TX. Technical Attachment SR 90-23, 7/1/90.

Steadman, R.G. 1979. "The assessment of sultriness. Part I: A temperature-humidity index based on human physiology and clothing science." Journal of Applied Meteorology, 18:861-873.

Steadman, R.G. 1979. "The assessment of sultriness. Part II: Effects of wind, extra radiation and barometric pressure on apparent temperature." Journal of Applied Meteorology, 18:874-885.

Wisbith, A., G. Meiners, T. Johnson, and W. Ollison. 1996. "Effect of Monitor Probe Height on Measured Ozone Concentration," Paper No. A310, 89th Annual Meeting of the Air and Waste Management Association, June.

