Accuracy of Petroleum Supply Data

by Tammy G. Heppner and Carol L. French

Overview

Petroleum supply data collected by the Petroleum Division (PD) in the Office of Oil and Gas (OOG) of the Energy Information Administration (EIA) showed an improvement in the accuracy of the 2002 data from good, to better, to best, for initial estimates to final values. These data were presented in a series of PD products: the *Weekly Petroleum Status Report* (WPSR), *This Week in Petroleum* (TWIP), the *Petroleum Supply Monthly* (PSM), and the *Petroleum Supply Annual* (PSA). Weekly estimates in the *WPSR* and *TWIP* were the first values available.

Figure FE1 illustrates that as reporting time increases from the weekly estimates to the interim monthly values to the final petroleum supply values, there is more in-depth review resulting in an improvement in the accuracy of the data. For the monthly-from-weekly (MFW) data, respondents have the shortest reporting time, and the data are least accurate but "good." For the *PSM* data, respondents have a longer reporting time than the weekly, and the data are more accurate or "better." For the *PSA* data, respondents have the longest reporting time, and the data are the most accurate or "best." For 2002, 66 petroleum supply data series were analyzed to determine how close the *PSM* values were to the final *PSA* values. For these series, 44 out of the 66 were within 1 percent of the *PSA* values in terms of mean absolute percent error as compared to 36 in 2001. Sixty-one petroleum supply data series were analyzed to see how close the MFW estimates were to the final *PSA* values. For these 61 series, 27 were within 2 percent of the *PSA* values in terms of mean absolute percent error and, of those, 12 were within 1 percent, compared to 25 and 11, respectively, for 2001.

Two major factors that contribute to the *PSM* values being more accurate than the MFW estimates are: (1) the greater length of time between the close of the reference period and the publication date of the *PSM*; and, (2) most MFW values (weekly data converted to a monthly value) are based on company's operational records whereas *PSM* values are generally extracted from company's accounting systems, the later being more accurate. The greater length of time allows more in-depth review of the data by the respondents and EIA. Within 2 months of the close of a reference month, interim values are published in the *PSM*. The weekly data are more quickly available. The *WPSR* and *TWIP* are available electronically 5 days after the close of the reference week (excluding holiday weeks). About 5 months after the end of the reference year, final monthly values, reflecting resubmissions, are published in the *PSA*.

Figure FE1.   The 2002 Weekly and Monthly Data Did Not Stack Up to the Annual Data
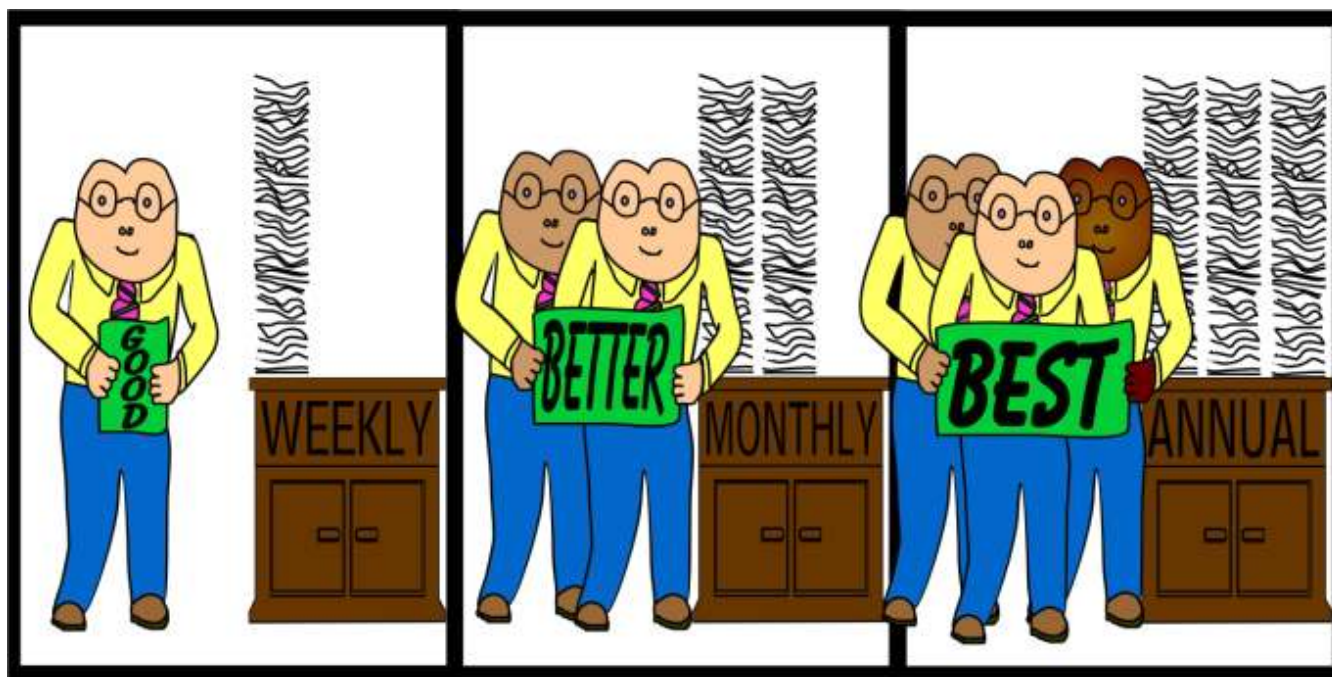
Historically, the weekly publication (*WPSR*) and the monthly publication (*PSM*) provided volumes of crude oil and petroleum products data at relatively increasing levels of accuracy. This article provides petroleum analysts with a measure of the degree to which, on average, estimates and interim values vary from their final values.

The Petroleum Supply Reporting System

The 15 surveys in the Petroleum Supply Reporting System (PSRS) track the supply and disposition of crude oil, petroleum products, and natural gas liquids in the United States. To maintain a database with historically accurate observations and current estimates from the petroleum industry, EIA administers three survey series: weekly, monthly, and annual.

The PSRS is organized into two data collection subsystems, the Weekly Petroleum Supply Reporting System (WPSRS) and the Monthly Petroleum Supply Reporting System (MPSRS). The WPSRS processes data from the six weekly surveys. The MPSRS includes eight monthly surveys and one annual survey.

Figure FE2 displays the petroleum supply and distribution system and indicates the points at which petroleum supply data are collected. Both weekly and monthly surveys are administered at five key points along the petroleum production and supply path: (1) refineries, (2) bulk terminals, (3) product pipelines, (4) crude oil stock holders, and (5) importers of crude oil and products.

Annual U.S. refinery capacity data are collected on the Form EIA-820, "Annual Refinery Report." These data were collected and published in Volume 1 of the *PSA* for 2002, available in hardcopy and electronically. Volume 2 of the *PSA* is only available electronically.

The Weekly Petroleum Supply Reporting System

The WPSRS contains the data collected from the six weekly surveys. Each weekly survey is distributed to a sample of the corresponding monthly survey's universe. In Figure FE2, the icons represent the target population of the monthly and weekly surveys of the PSRS. For example, the target population for the survey Forms EIA-801 and EIA-811 is bulk terminal stocks. Thus, the respondents to the Form EIA-801 are a sample of the respondents who report on Form EIA-811.

Figure FE2.   Petroleum Supply Reporting System: Surveys and Subsystems
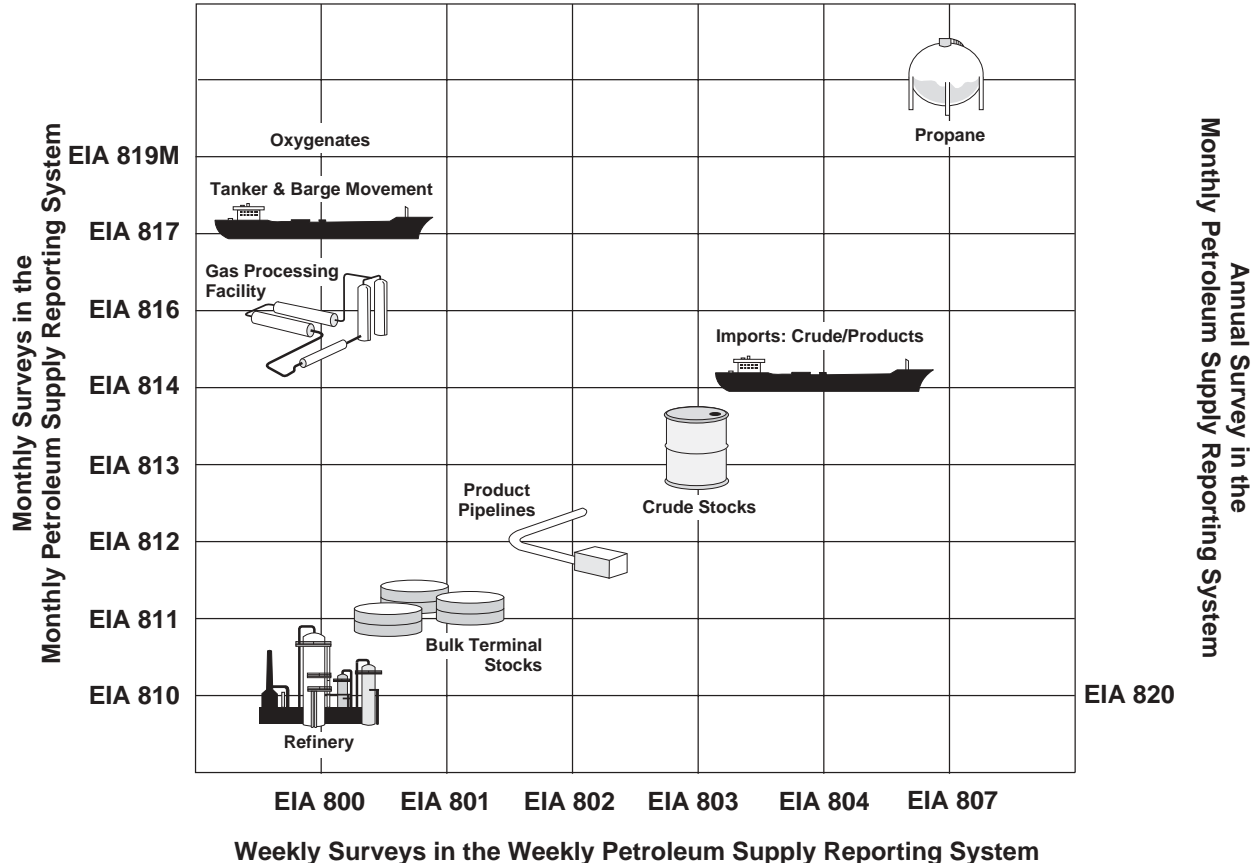
Source: Energy Information Administration, Petroleum Supply Reporting System.

For the weekly surveys, EIA aims for a minimum 90-percent multi-attribute-cutoff sample from the respondents to the corresponding monthly survey. In choosing the sample for each product, companies are ranked in descending order by volume. Respondents are chosen in order, down the list until the sample includes those companies contributing at least 90 percent of a variable's total volume. For example, for distillate fuel oil stocks, the weekly sample includes those respondents whose combined volumes of stocks for distillate fuel oil from refineries, bulk terminals, and pipelines constitute at least 90 percent of the total volume of distillate fuel oil stocks as reported in the corresponding monthly surveys.

These surveys enable EIA to provide timely, relatively accurate snapshots of the U.S. petroleum industry every week. The weekly surveys collect information on the supply and disposition of selected petroleum products and crude oil. The reference period for each weekly survey begins at 7:01 a.m. each Friday and ends at 7:00 a.m. the following Friday. Respondents report their data via telephone, facsimile, electronic spreadsheets, or EIA's electronic data collection software package, the Personal Computer Electronic Data Reporting Option (PEDRO). All respondents must submit their data by 5:00 p.m. on the Monday following the end of the reference period. During 2 working days, quality control procedures are performed. Cell values determined to be unusual or inconsistent with other cell values are flagged. The validity of the value of each flagged cell is investigated. Some flagged values are verified by the respondent to be correct; other flagged cells are corrected; and the remaining flagged values are referred to as unresolved. Nonrespondent and unresolved flagged data are imputed using an exponentially smoothed mean of the respondents' historical data.

As a new weekly web product, *This Week in Petroleum* (TWIP) provides analysis, data, and charts of the latest weekly petroleum supply and price data. Weekly propane data previously collected during the heating season on Form EIA-807, "Propane Telephone Survey," are collected and published on a weekly basis year round, beginning with the October 11, 2002 *WPSR*.

Within 5 days of the close of the reference week, weekly data are made available to the public on the EIA's internet web site (http://www.eia.doe.gov) through the *WPSR* and *TWIP*. Except when holidays delay data processing schedules, values for the weekly variables, with the exception of propane, are available via the internet at 10:30 a.m. Eastern Time on the Wednesday following the close of the reference week. *TWIP* and propane data are generally available at 1:00 p.m. on Wednesdays at http://tonto.eia.doe.gov/oog/info/twip/twip.asp.

The Monthly Petroleum Supply Reporting System

The reference period for the monthly surveys starts on the first day of the month at 12:01 a.m. and ends on the last day of the

month at midnight. Except for the Form EIA-819M, the deadline for filing monthly surveys is the 20th calendar day following the end of the report month. Data collection for the Form EIA-819M begins on the seventh working day of the month. Form EIA-819M data are solicited by telephone or received by facsimile or electronic mail. Data for the other monthly surveys are reported via mail, telephone, facsimile, electronic spreadsheets, or PEDRO.

During the period of data editing, either the respondent or EIA staff may identify an error. If the respondent discovers an error, the EIA representative for a particular survey is notified and the value is corrected. If EIA's edits diagnose an unusual value, an EIA representative will determine if the value is correct or incorrect by calling the company and/or reviewing historical data.

Within 60 days of the close of the reference month, all of the interim monthly data are published in the *PSM* and on the internet. Throughout the year, EIA accepts data revisions of monthly data. If a revision is made after the *PSM* has been published, it is referred to as a resubmission. The impact of resubmissions to previous months published data are presented in Appendix C of the *PSM*. Additionally, preliminary company-level imports data are released electronically between the 7th and 10th of each month.

Beginning with the February 1994 *PSM*, Table H1, "Petroleum Supply Summary" was included to show early estimates of monthly data. The current-month values in Table H1 are preliminary estimates based on weekly submissions. These monthly-from-weekly estimates are published in the *WPSR* and on the internet on the Wednesday following the first Friday of each month.

Within 5 months of the end of the calendar year, the final monthly values for the previous year are published in the *PSA*. These values reflect all *PSM* resubmissions and other data corrections. The values contained in the *PSA* are EIA's most accurate measures of petroleum supply activity.

Factors Affecting Data Accuracy

Maintaining an accurate database is a major goal of EIA. The quality of the data drives the quality of all qualitative and quantitative analyses conducted using these data. Accuracy and timeliness are primary attributes of high quality data. Accuracy of survey data is measured as the closeness of the published values to the true values (i.e., those values that would be obtained if the entire target population had been surveyed and all the data had been precisely recorded).

Respondents to the monthly surveys have more time to file than the weekly respondents, enabling them to collect, review,

Table FE1. Average Coverage for Weekly Surveys, 2002 and 2001 (Percent of Final Monthly Volumes Included in Monthly-from-Weekly Sample)

Product	Stocks						Production		Imports	
	Refinery		Bulk Terminal		Pipeline					
	2002	2001	2002	2001	2002	2001	2002	2001	2002	2001
Total Motor Gasoline	98	98	93	93	97	97	98	99	90	91
Jet Fuel	98	98	91	92	98	100	99	99	93	91
Distillate Fuel Oil	96	96	87	87	98	98	97	97	94	96
Residual Fuel Oil	95	96	90	89	—	—	94	95	94	96
Crude Oil	96	96	—	—	—	—	—	—	95	94

— = Not Applicable.

Source: Energy Information Administration, Petroleum Supply Reporting System.

and revise their data more carefully than the weekly respondents. Additionally, EIA has more time to edit the monthly data. Also, some weekly respondents report estimates while many monthly respondents extract actual data from accounting systems. Thus, the monthly data are typically more accurate.

Some sources of error, such as nonresponse, are not totally preventable. Other errors, such as sampling errors, are unique to a particular type of survey. One situation where sampling error occurs is if the group of sampled respondents is dissimilar to the full population. Within the PSRS, only weekly surveys, and the Form EIA-819M, "Monthly Oxygenate Telephone Report," are at risk of having sampling errors. However, all surveys in the PSRS are at risk for nonsampling errors, such as: (1) insufficient coverage of respondents (the survey frame does not include all members of the target population); (2) nonresponse; (3) response error; and (4) errors due to lack of survey clarity. A detailed discussion of factors influencing data accuracy and how they are minimized in the PSRS follows.

Samples and Sampling Error

A sample is a subsection of a universe identifying members of a target population. The weekly surveys are administered to samples of the monthly populations to reduce respondent burden and to expedite the turnaround of data from survey respondents to the public. As with any sample, the values obtained are different from those obtained if the full universe had been surveyed. Sampling error is the difference between a sample estimate and a population value.

There are six samples, one for each weekly petroleum supply survey, in the WPSRS. For these surveys, the sampling error is minimized by using a minimum 90-percent multi-attribute-cutoff sample from the corresponding monthly survey's frame. At the end of each month, updates are made to

the samples and survey frames if a 90-percent coverage was not obtained.

For the weekly surveys, better coverage will most likely reduce sampling error. As shown in Table FE1, 2002 coverage was comparable to 2001. Of the 21 product and supply type combinations, 18 had coverage of 90 percent or above in 2002. For 15 of the 21 combinations, 2002 coverage slightly decreased from 2001. Crude oil imports had the largest percentage increase from 2001 to 2002, at 1.47 percent. The largest percentage decrease from 2001 to 2002 was for residual fuel oil imports, at 1.70 percent. Tabulations were done before rounding of the coverage values.

Nonsampling Error

Unlike sampling errors, all survey data, even those from a census survey, are at risk of incurring nonsampling errors. There are two categories of nonsampling errors, random and systematic. With random error, on average, and over time, values will be overestimated by the same amount they are underestimated. Therefore, over time, random errors do not bias the data, but they will give an inaccurate portrayal at any point in time. On the other hand, systematic error is a source of bias in the data, since these patterns of errors are made repeatedly. The following is a discussion of how the four most frequently occurring types of nonsampling error are minimized within the PSRS.

Frame Updates

The list of all companies identified as members of the target population is called a frame. If members of the target population are not included in the frame, there is an undercount of the aggregate data. To diminish the chance of undercounting, the PSRS frames are continually updated. New companies are identified through continual review of petroleum industry periodicals, newspaper articles, and correspondence from respondents.

Table FE2.   Average Response Rates for Monthly and Weekly Surveys, 2002

Survey Site	Respondents to Monthly Surveys			Respondents to Weekly Surveys		
	Average Universe Size	Average Number of Respondents	Percent [1]	Average Weekly Sample Size	Average Number of Respondents	Percent [2]
Refinery	298	288	97.2	194	189	97.5
Bulk Terminal	246	240	97.6	65	62	95.4
Pipeline	83	83	100.0	40	40	98.4
Crude Oil Stocks	154	152	98.6	63	62	97.9

[1] The average response rates for monthly surveys are calculated by summing the individual monthly response rates and dividing by 12.
[2] The average response rates for weekly surveys are calculated by summing the individual weekly response rates and dividing by 52.
Note: Percents are calculated before rounding.
Source: Energy Information Administration, Petroleum Supply Reporting System.

Maintaining a Low Nonresponse

Survey respondents are required by law to report to EIA (see Explanatory Note 6 of the *PSM* for a description of action for chronic nonresponse). The 2002 response rates for the weekly surveys and their corresponding monthly surveys are enumerated in Table FE2.   With the exception of the monthly refinery survey, which only decreased by one half of a percent, all of the 2002 response rates for each of the EIA weekly and monthly surveys increased from 2001. The largest increase in response rate was for the weekly refinery survey, increasing from 93.1 percent in 2001 to 97.5 percent in 2002.

To mitigate the effect of nonresponse, imputed values are calculated for all nonreported values except monthly imports. Weekly imputed values are the exponentially smoothed mean of that respondent's historical values for that variable. Monthly imputed values are the previous month's value for the particular respondent and variable. For imports, however, there is a great deal of fluctuation from one reference period to another, with respondents frequently having no imports of a particular product. As a result, the values are not imputed. In addition, the monthly imports are collected and published at a much greater level of detail than the weekly imports, which makes imputation impractical.

Reducing Response Error

Improvements to the PSRS system are continuously being made to reduce response error. To satisfy customer needs and meet the particular requirements of some respondents, computerized spreadsheets that resemble the actual survey forms have been developed, and are available for respondent reporting.   Another improvement has been the increased participation in the PEDRO system, which permits all weekly and monthly survey data, except the Form EIA-819M and Form EIA-807, to be submitted to EIA electronically. A respondent entering values via PEDRO may execute edit routines prior to transmission of the survey responses. These routines include consistency and outlier (extreme value) checks of the data. Unusual or nonreported cells are flagged and, prior to transmission of the data, a representative of the company is able to review and verify or correct data in the flagged cells.

Even with sophisticated edit checks, response error (the difference between the reported value and the actual value) remains the most likely cause of data inaccuracy.   The weekly surveys are more susceptible to response error since some of their values are estimates or based on operational records. Many monthly respondents abstract their monthly data from accounting systems and thus are generally more accurate.

Maintaining accurate accounting records, however, does not ensure against response error. For example, numbers can be transposed within the correct cell; an otherwise correct value may be entered in the wrong cell; a respondent may misinterpret the intent of a question; or the wrong units may be used.

Survey Clarity

The terms, layout, and definitions on all survey forms are periodically reviewed for completeness, clarity, and consistency across surveys. At regular intervals, survey intent, as well as what data are collected, are subject to industry and government review. To the extent possible, industry changes in terminology and practice are incorporated into the PSRS on an ongoing basis.

Data Assessment

Each of the variables included in these analyses is of current and historical interest. Of the 66 variables for which both *PSM* and *PSA* values were published, only 61 of them were published weekly throughout 2002. For each variable, six measures of accuracy were calculated to compare the differences between the MFW and *PSM* values relative to the *PSA* values.

Error is the difference between the estimate (MFW) or interim (*PSM*) value and the final (*PSA*) value for a given month. For inputs, production, stock change, imports,

exports, and product supplied, values are expressed in units of thousands of barrels per day. For stocks, values are expressed in units of thousands of barrels.

MFW Error = MFW Volume - *PSA* Volume

PSM Error = *PSM* Volume - *PSA* Volume

Percent Error is the error for a given month divided by the final value for a given month, and multiplied by 100.

$$\text{MFW Percent Error} = \frac{\text{MFW Error}}{PSA \text{ Volume}} \times 100$$

$$PSM \text{ Percent Error} = \frac{PSM \text{ Error}}{PSA \text{ Volume}} \times 100$$

Mean absolute error is the weighted average over the 12 months of the year of the absolute values of the errors for each month. The mean absolute error measures the average magnitude of the revisions that took place over a year. Outliers increase the mean absolute error. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months.

Mean absolute percent error is the weighted average over the 12 months of the year of the absolute values of the percent errors. It provides a measure of the average magnitude of the revisions relative to final values. The mean absolute percent error has an inverse relationship with data accuracy; i.e., the smaller the mean absolute error, the closer the interim data are to the final data; conversely, the larger the mean absolute percent error, the greater the difference in the interim value and the final value. Outliers inflate the mean absolute percent error.

Range is the difference between the smallest and largest percent errors. The range shows the dispersion of the percent differences between interim and final values.

Median of the percent errors is the point at which half the values are higher and half are lower. Unlike the mean, the median is not affected by an outlier. In these analyses, each distribution has 12 observations. The median is the average of the sixth and seventh ordered observation.

The average final absolute volumes and the mean absolute percent error for MFW estimates and *PSM* interim values for 2002 and 2001 are presented in Table FE3. The average final absolute volumes are presented to give the reader an idea of the magnitude of these volumes. Variables with very small volumes are prone to larger percent changes because a modest volume change is being compared to a small final volume. The mean absolute error and the size of the volumes involved must both be included in the interpretation of data accuracy.

The 2002 MFW mean absolute percent errors which were within 2 percent of their respective *PSA* values (27 of the 61 MFW series), and the 2002 *PSM* mean absolute percent errors which were within 1 percent of their *PSA* values (44 of the 66 *PSM* series), are distinguished by a single asterisk. Mean absolute percent errors that were greater than 10 percent are marked by a double asterisk. There were 12 such MFW series and 4 *PSM* series, compared to 16 and 8, respectively, for 2001.

For 2002, 8 of the 11 weekly production series decreased in mean absolute percent error from 2001. Thirteen of the 14 production series have a single asterisk in the *PSM* column, indicating a mean absolute percent error of less than 1 percent from the *PSA*. Additionally, all of the 14 *PSM* production series in 2002 decreased or stayed the same in mean absolute percent error from 2001. Weekly fuel ethanol supply and disposition data are not available; therefore, the weekly oxygenated motor gasoline field production is based on the latest available monthly value.

The single asterisks in Table FE3 by the stock series show that, as in prior years, the stock values for both MFW estimates and *PSM* interim values are very close to the final *PSA* values. A major exception is the double asterisk shown by the MFW percent error for oxygenated motor gasoline stocks. The increase is related to the average absolute volume. Fuel ethanol and methyl tertiary butyl ether stocks are not collected weekly, but are collected on the Form EIA-819M, "Monthly Oxygenate Telephone Report." The survey provides production data and preliminary stock data from a sample of respondents reporting on the monthly surveys and from the universe of oxygenate producers. These data are displayed in Appendix D of the *PSM*. Interim data are collected later on the monthly surveys and published in the *PSM*. Fourteen of the 19 weekly stock series and 17 of the 19 monthly stock series for 2002 decreased or stayed the same in mean absolute percent error from 2001.

Stock change is the difference between stocks at the beginning of the month and stocks at the end of the month. Since the monthly change in stock levels is small compared to the stock levels themselves, a large percent error in stock change can occur even when the percent errors in stock levels are small.

Crude oil stock change is one of the components in the calculation of unaccounted for crude oil (calculated disposition minus calculated supply of crude oil). For both the MFW and the *PSM* numbers, the volume of the unaccounted for crude oil may be increased by a combination of factors including an understatement of imports, an overstatement of exports, an understatement of crude oil production, an understatement of stock withdrawals, and an overstatement of crude oil inputs. The overstatement of crude oil inputs can be caused by injections along crude oil pipelines of natural gas liquids. When refiners receive this mixture, they process it as crude oil. As seen in Table FE3, the production, imports, and refinery

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 2002 and 2001

Variable	PSA Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		PSM Mean Absolute Percent Error	
	2002	2001	2002	2001	2002	2001
Crude Oil Production (thousand barrels/day)	5,745	5,801	* 1.17	0.64	1.50	0.92
Refinery Operations						
Refinery Crude Oil Inputs (thousand barrels/day)	14,947	15,128	* 0.34	0.57	* 0.16	0.06
Operating Utilization Rate (percent)	91	93	* 1.58	1.17	* 0.39	0.34
Production (thousand barrels/day)						
Total Production	19,571	19,537	—	—	* 0.13	0.35
Refinery Production	17,273	17,285	* 0.98	4.17	* 0.13	0.33
Finished Motor Gasoline	8,475	8,312	* 1.19	0.71	* 0.24	0.28
Reformulated Motor Gasoline	2,690	2,597	2.68	1.30	* 0.74	1.09
Oxygenated Motor Gasoline	926	762	** 17.54	24.64	4.68	13.92
Other Motor Gasoline	4,859	4,953	3.01	4.13	* 0.68	1.69
Jet Fuel	1,514	1,530	* 0.62	0.82	* 0.05	0.05
Distillate Fuel Oil	3,592	3,695	* 0.65	1.15	* 0.13	0.14
Low Sulfur Distillate Fuel Oil	2,606	2,617	* 0.95	1.26	* 0.05	0.42
High Sulfur Distillate Fuel Oil	986	1,077	2.48	2.69	* 0.40	0.78
Residual Fuel Oil	601	721	3.88	1.98	* 0.30	0.30
Other Products	5,389	5,281	—	—	* 0.46	0.90
Propane	1,121	1,095	—	—	* 0.19	0.59
Other Products Refinery Production	3,383	3,318	9.10	27.16	* 0.27	0.35
Stocks (thousand barrels)						
Total Stocks	1,586,337	1,543,493	* 0.60	0.78	* 0.07	0.20
Total Stocks, excl. SPR	1,009,412	999,365	* 0.90	1.20	* 0.11	0.31
Total Crude Stocks	883,482	854,131	* 0.30	0.49	* 0.06	0.27
Crude Oil Stocks, excl. SPR	306,557	310,003	* 0.83	1.33	* 0.18	0.73
SPR Stocks	576,925	544,128	* 0.08	0.00	* 0.00	0.00
Refined Products Stocks	702,855	689,362	* 1.11	1.24	* 0.11	0.15
Total Motor Gasoline Stocks	211,486	206,421	* 1.14	0.59	* 0.15	0.19
Reformulated Motor Gasoline Stocks	42,390	43,271	2.01	2.00	1.71	0.65
Oxygenated Motor Gasoline Stocks	449	658	** 17.89	30.84	1.57	14.12
Other Motor Gasoline Stocks	119,294	113,700	* 1.60	0.87	* 0.12	0.45
Jet Fuel Stocks	40,517	41,851	* 1.64	1.92	* 0.31	0.42
Distillate Fuel Oil Stocks	128,645	121,063	* 1.41	1.57	* 0.34	0.19
Low Sulfur Distillate Fuel Oil Stocks	74,717	70,535	2.01	1.92	* 0.37	0.46
High Sulfur Distillate Fuel Oil Stocks	53,928	50,528	* 1.38	1.39	* 0.37	0.58
Residual Fuel Oil Stocks	34,568	38,896	* 1.93	2.03	* 0.16	0.99
Other Products Stocks	287,639	281,131	2.37	2.37	* 0.05	0.42
Propane Stocks	56,073	50,330	* 1.72	2.17	* 0.28	1.07
Fuel Ethanol Stocks	5,901	3,568	3.78	7.06	* 0.56	4.19
Methyl Tertiary Butyl Ether Stocks	6,980	7,866	* 1.98	3.16	* 0.28	0.56
Stock Change (thousand barrels/day)						
Total Stock Change	397	485	** 42.53	105.63	** 24.00	33.78
Crude Stock Change	321	310	** 39.58	103.24	** 11.76	37.05
Refined Products Stock Change	437	453	** 64.84	235.07	** 14.69	35.60
Imports (thousand barrels/day)						
Total Imports	11,530	11,871	3.16	3.77	1.50	2.11
Total Crude Imports	9,124	9,318	2.65	2.37	1.03	1.94
Crude Oil Imports, excl. SPR	9,140	9,328	2.67	2.38	1.03	1.94
SPR Imports	0	0	* 0.00	0.00	* 0.00	0.00
Refined Products Imports	2,390	2,543	5.17	8.63	3.37	2.55
Finished Motor Gasoline Imports	498	455	3.09	9.31	1.17	2.24
Reformulated Motor Gasoline Imports	233	217	8.81	8.72	* 0.41	2.49
Oxygenated Motor Gasoline Imports	0	1	* 0.00	22.60	* 0.00	0.00
Other Motor Gasoline Imports	265	236	7.04	14.71	1.87	2.89
Jet Fuel Imports	107	148	** 18.82	15.56	3.18	2.24

See footnotes at end of table.

Variable	PSA Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		PSM Mean Absolute Percent Error	
	2002	**2001**	**2002**	**2001**	**2002**	**2001**
Distillate Fuel Oil Imports..	267	344	6.26	15.58	1.88	5.41
Low Sulfur Distillate Fuel Oil Imports...........................	107	130	** 14.86	18.69	4.01	6.73
High Sulfur Distillate Fuel Oil Imports	161	215	** 13.45	14.98	* 0.96	8.07
Residual Fuel Oil Imports	249	295	** 21.22	28.74	** 17.07	33.54
Other Products Imports	1,268	1,301	6.68	16.04	2.96	12.60
Propane Imports	145	145	—	—	1.04	5.07
Exports (thousand barrels/day)						
Total Exports	984	971	** 10.42	7.68	* 0.34	2.03
Crude Oil Exports	9	20	**312.31	540.90	* 0.00	5.66
Refined Products Exports...	975	951	** 10.13	6.95	* 0.34	1.80
Total Net Imports (thousand barrels/day)	10,547	10,900	3.62	4.18	1.59	2.38
Products Supplied (thousand barrels/day)						
Total Products Supplied	19,761	19,649	* 1.01	1.63	* 0.56	0.44
Finished Motor Gasoline Supplied..................................	8,848	8,610	* 1.34	0.82	* 0.23	0.37
Jet Fuel Supplied...	1,614	1,655	* 1.98	2.58	* 0.42	0.60
Distillate Fuel Oil Supplied..	3,776	3,847	2.26	2.67	* 0.70	0.84
Residual Fuel Oil Supplied	700	811	8.81	14.91	6.37	15.95
Other Products Supplied	4,824	4,725	3.69	6.36	1.01	2.62
Propane Supplied	1,248	1,142	—	—	* 0.95	1.58

— = Not Applicable.
 * = For MFW values, mean absolute percent error less than or equal to 2; for PSM values, mean absolute percent error less than or equal to 1.
 ** = Mean absolute percent error greater than or equal to 10.
 SPR = Strategic Petroleum Reserve
 Notes: Error is the difference between Monthly-from-Weekly estimates or interim monthly data published in the Petroleum Supply Monthly and the final value as published in the Petroleum Supply Annual. Percent error is the error multiplied by 100 and divided by the final published value. Mean absolute error is the weighted average of the absolute errors. Mean absolute percent error is the weighted average of the absolute percent errors. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months.
 •Totals may not equal sum of components due to independent rounding.
 Source: Energy Information Administration, Petroleum Supply Reporting System.

inputs of crude oil have a small mean absolute percent error relative to crude oil stock change.

For petroleum products, stock change is a component in the calculation of product supplied (representing the consumption of petroleum products). Unlike the other variables, stock change values can be negative. Stock change thus has an added dimension by which to evaluate accuracy; this is the correctness of the direction of the change. Table FE4 provides a measure of accuracy of the direction of MFW and *PSM* stock change values for 2002 and 2001. All of the six stock change values for 2002 decreased the number of months that differed from the direction of the *PSA* values compared to 2001. All of the *PSM* stock change values were the same direction as the *PSA* values.

For imports, one reason for the large mean absolute percent errors in the MFW values is that shipments do not always arrive during the week in which they were expected. This has a greater impact

Table FE4.   Number of Months In Which the Direction of NonFinal Stock Change Values Differed From PSA

	Number of Months	
	2002	**2001**
Total Stock Change		
MFW and PSA Values	0	1
PSM and PSA Values	0	1
Crude Stock Change		
MFW and PSA Values	1	2
PSM and PSA Values	0	1
Refined Products Stock Change		
MFW and PSA Values	2	2
PSM and PSA Values	0	1

 Source: Energy Information Administration, Petroleum Supply Reporting System.

when the end of the month occurs in the middle of the week. Eleven of the 15 MFW import series in Table FE3 showed a decrease or stayed the same in mean absolute percent error from 2001 to 2002, similar to last year's decrease of 11 series from 2000 to 2001. For the *PSM*, 14 of the 16 import series decreased or stayed the same in mean absolute percent error compared to last year's decrease of 11 import series.

With the exception of refinery receipts in the U.S. Territories, EIA does not collect export data. They are gathered by the U.S. Bureau of the Census on a monthly basis. They are received by EIA on a monthly basis approximately 7 weeks after the close of the reporting month. The weekly estimates for exports are projections based on past monthly data. Because the export data are highly variable, it is difficult to obtain estimates of comparable quality to domestic estimates.

Products supplied is the calculation of field production, plus refinery production, plus imports, plus unaccounted for crude oil, minus stock change, minus crude oil losses, minus refinery inputs, minus exports. Therefore, the accuracy of products supplied is affected by the individual components.

Box and Whisker Plots

Example 1 in the shaded box titled "Structure of Box and Whisker Plots," is a simplified illustration of the box and whisker plots that follow. The box and whisker plots map the 5-year trends in historical accuracy of weekly estimates and monthly interim values. The details provided by the box and whisker plots include: historical trends, the range of monthly percent errors, direction of the error (i.e., overestimation or underestimation), and the identification of unusual values.

Each box and whisker plot is placed on a graph, where the horizontal axis represents the year and the vertical axis represents the percent error. The center horizontal line for all the box and whisker plots is zero percent error. For each variable studied, a pair of charts, each containing five box and whisker plots (one for each year, from 1998 through 2002), are presented side-by-side; the chart on the left contains the percent errors for the MFW estimates, and the chart on the right contains the percent errors for the *PSM* values. To facilitate the comparison of MFW percent errors and the *PSM* percent errors, the plots have the same scale.

The position of the box along the y-axis denotes whether the MFW or *PSM* values are predominantly overestimates or underestimates of the *PSA* values. For example, if the majority of the MFW values were overestimates, more than half of the box would be above the zero percent error line.

Crude Oil Production and Crude Oil Inputs

Crude oil production data are not collected through any of EIA's surveys. EIA's Dallas Field Office assembles data collected from State agencies responsible for measuring crude oil production. Based on historical trends and data reported on Form EIA-182,

"Domestic Crude Oil First Purchase Report," EIA estimates weekly and monthly production. Final estimates based on revised Form EIA-182 data, State government agencies, and the U.S. Department of Interior's Minerals Management Service data are published in the *PSA*. Figure FE3 presents errors of MFW and *PSM* values relative to *PSA* values for crude oil production and crude oil inputs. All but one of the 2002 MFW estimates for crude oil production overestimated the final *PSA* values. Except for the two outliers in May (-0.54) and October (3.97), the percent errors were tightly distributed around the median of 1.05 percent. Those outliers were caused by company revisions. In contrast, the 2002 *PSM* percent errors ranged from -0.61 to 5.74 percent. October 2002 (5.74) had the largest percent error over the 60 months studied. Most of the *PSM* interim values overestimated the final *PSA* values.

For refinery crude oil inputs, the range (1.53) of the 2002 MFW percent errors was the smallest range of all other MFW plots analyzed for 2002. As in prior years, the 2002 *PSM* refinery crude oil inputs were extremely close to their final values, with percent errors within 0.51 percent.

Product Production

As expected, *PSM* interim values for production of each of the four major petroleum products were superior to their comparable MFW estimates. Figures FE4 and FE5 contain the box and whisker plots for motor gasoline and distillate fuel oil production, and residual fuel oil and jet fuel production, respectively.

The 2002 MFW motor gasoline production percent errors, displayed in Figure FE4, ranged from -2.46 to 0.94 percent. The median of -0.99 percent had the largest absolute percent error over the 5-year period. Unlike prior years, most of the 2002 *PSM* interim values for motor gasoline production overestimated the final *PSA* values but the percent errors were within 0.71 percent. The median of 0.08 percent was the only positive median over the past 5 years.

The 2002 range (2.41) of MFW percent errors for distillate fuel oil production, ranging from -0.90 to 1.51 percent, was the smallest range over the 5-year period. Similarly, the 2002 range of 0.56 percent for the *PSM* percent errors was the smallest range over the past 5 years.

The box and whisker plots for residual fuel oil production and jet fuel production are shown in Figure FE5. The range (17.57) of the 2002 MFW percent errors for residual fuel oil production was the largest range over the 5-year period, ranging from -11.27 to 6.30 percent. The large range was caused by an outlier in November (-11.27) due to misreporting. In contrast, the range (1.62) of the 2002 *PSM* percent errors for residual fuel oil production was the smallest range over the 5-year period, ranging from -1.62 to 0.0 percent. One outlier in March (-1.62) was due to company revisions.

Structure of Box and Whisker Plots

All box and whisker plots discussed in this article are the visual presentation of a variable's distribution of 12 values of percent errors for either MFW or *PSM* values relative to *PSA* values for a given year. In general, box and whisker plots group data, ordered from smallest to largest, into four areas of equal frequency, quartiles, and show the range and dispersion of data within the quartiles. Sometimes the values of quartiles must be interpolated, i.e., if there are two values that meet the criteria of a quartile, then the average of the two must be taken. Presented below is a discussion of components of box and whisker plots and how they apply to the 12-value distribution illustrated in Example 1: -35, -20, -11, -9, 0, 0, 0, 0, 4.5, 5.5, 15, and 20.

First Quartile

Twenty-five percent of the values are equal to or below the first quartile. In Example 1, the first quartile is the average of the third and fourth ordered observations, i.e., (-11+(-9))/2=-10. The first quartile demarcates the lower boundary of the box.

Second Quartile

The second quartile is the median, and it intersects the box. Fifty percent of the observations are equal to or below the median; in our example, the values of these six observations are: 0, 0, -9, -11, -20, and -35. Also, for this example, the median is the average of the sixth and seventh value, 0, i.e., (0+0)/2. The plot provides the value of the median (the second quartile) as well as information on how the median compares in magnitude to the rest of the observations. Outliers distort the magnitude of the mean, whereas a median is not distorted since it is the actual value that falls in the middle of the distribution. Since outliers have occurred in the distributions of values of PSRS variables, a median is preferred to a mean when assessing accuracy.

Third Quartile

Seventy-five percent of the observations (9 in this case) have values equal to or below the third quartile. In Example 1, the third quartile is 5, i.e., (4.5+5.5)/2. The third quartile demarcates the upper boundary of the box.

Box

The box contains half of all the values. In Example 1, as well as in each box found in Figures FE3-FE11, a minimum of six values are contained within the box. The interquartile range is the length of the box, the difference between the first and third quartiles. The interquartile range for Example 1 is 15, i.e., 5-(-10).

Whiskers

Each whisker extends out from the box, one from the first quartile and the other from the third quartile, to the most extreme value that still falls within 1.5 times the interquartile range. In Example 1, a whisker extends from the third quartile, 5, to 20, which is the maximum value and is within 1.5 interquartile ranges of 5 (as it is less than 5+(1.5*15)=27.5). Also in Example 1, the lower whisker extends from the first quartile -10, to -20, which is the lowest value of the distribution within 1.5 interquartile ranges of the first quartile.

Fourth Quartile

The fourth quartile is the maximum value of the distribution. In Example 1, the fourth quartile, 20, also demarcates the upper value of the top whisker as it is within 1.5 interquartile ranges of the third quartile.

Outlier

An outlier, identified as an asterisk, is an observation that is more than 1.5 interquartile ranges greater than the third quartile, or more than 1.5 interquartile ranges less than the first quartile. In Example 1, there is one outlier, -35. It is less than the lower whisker's threshold value, which is -32.5 (-10-(1.5*15)). The importance of the occurrence of an outlier depends on the distribution of the variable. If the interquartile range is very tight and the outlier is in close proximity, then there is little concern about the occurrence of that outlier. (See Figure FE3, MFW vs *PSA* of Crude Oil Production for 2000.)
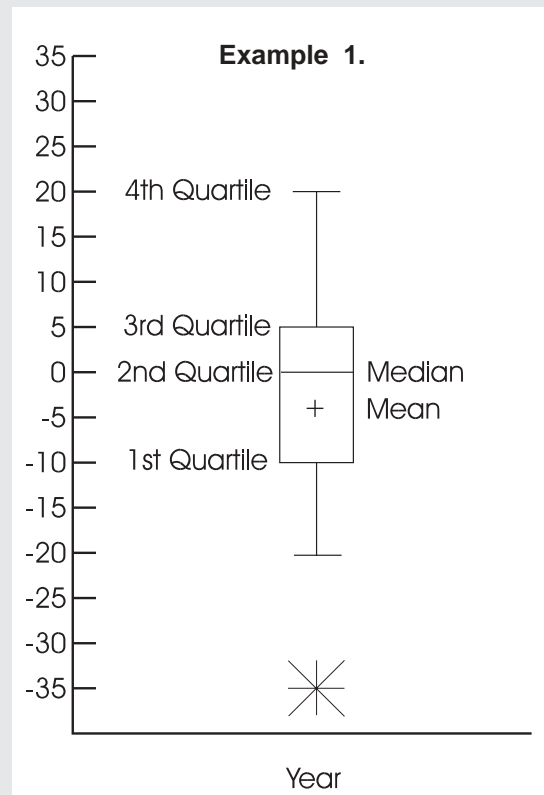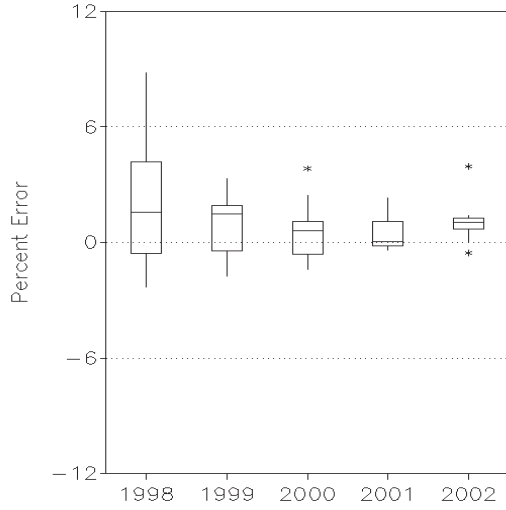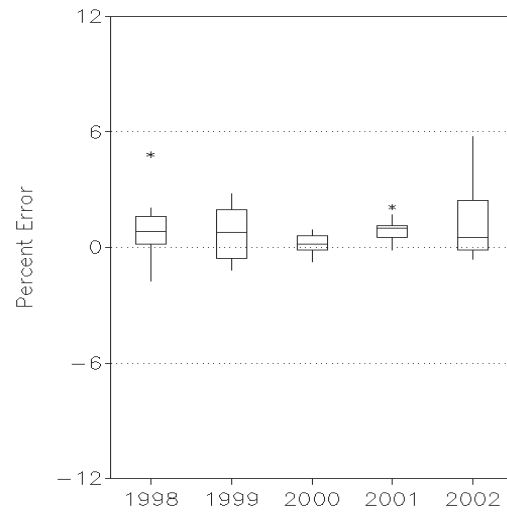

Example 1.

Figure FE3. **Range of Percent Errors for MFW and *PSM* Crude Oil Production and Refinery Crude Oil Inputs Data, 1998 - 2002**
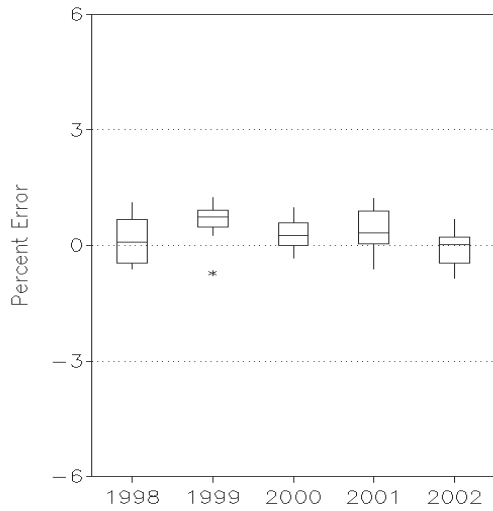
Crude Oil Production
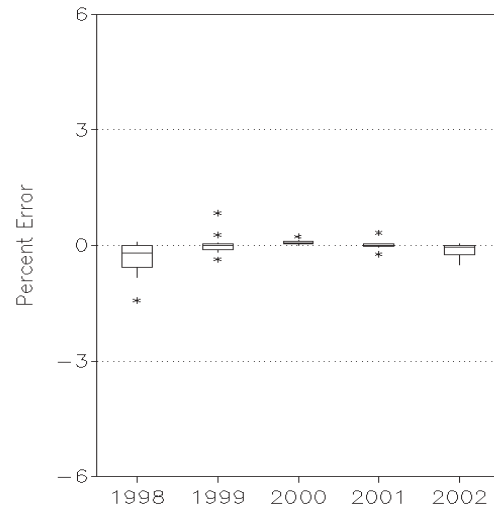
MFW vs. *PSA*

PSM vs. *PSA*

Refinery Crude Oil Inputs
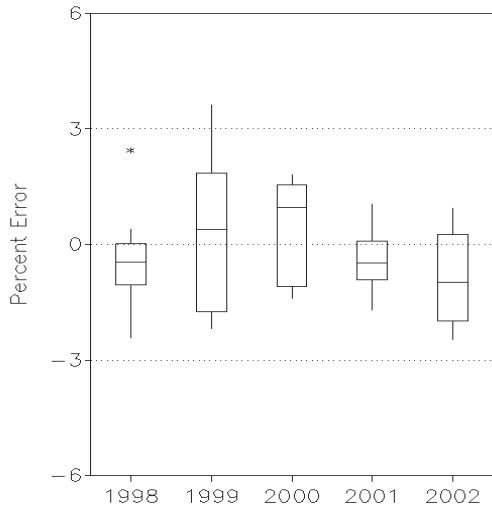
MFW vs. *PSA*

PSM vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.
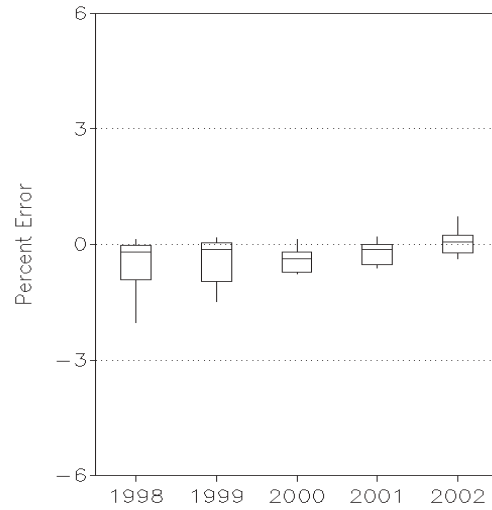
Figure FE4.   Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Production Data, 1998 - 2002
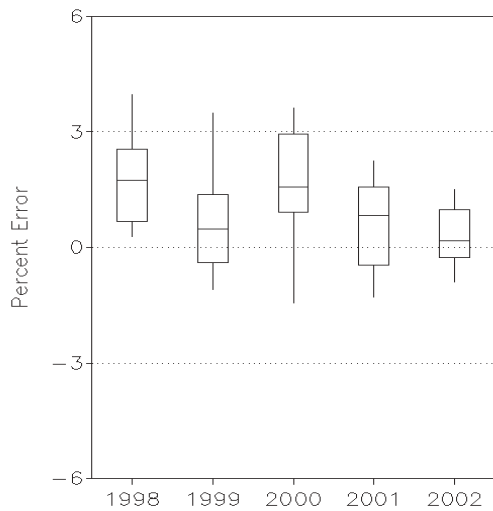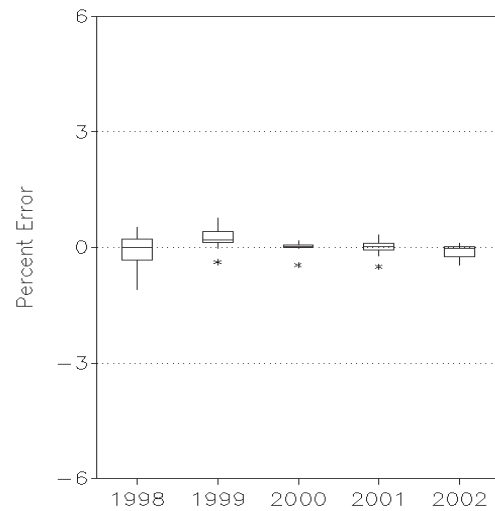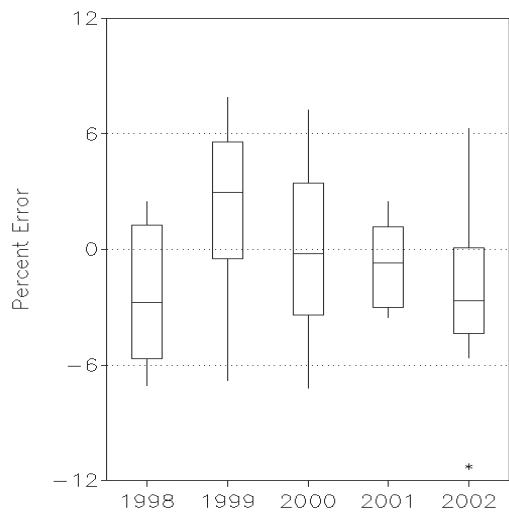
Motor Gasoline Production

MFW vs. *PSA*

PSM vs. *PSA*

Distillate Fuel Oil Production

MFW vs. *PSA*

PSM vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.

Energy Information Administration/Petroleum Supply Monthly, October 2003

Figure FE5.   Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Production Data, 1998 - 2002

Residual Fuel Oil Production
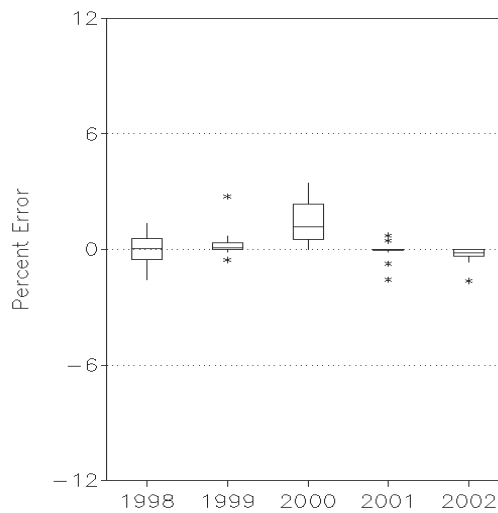
MFW vs. *PSA*   *PSM* vs. *PSA*

Jet Fuel Production

MFW vs. *PSA*   *PSM* vs. *PSA*
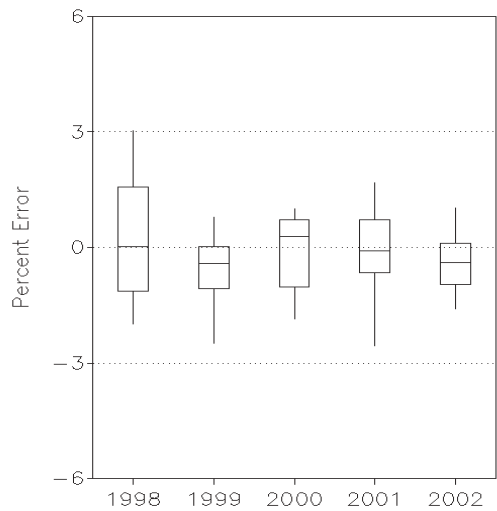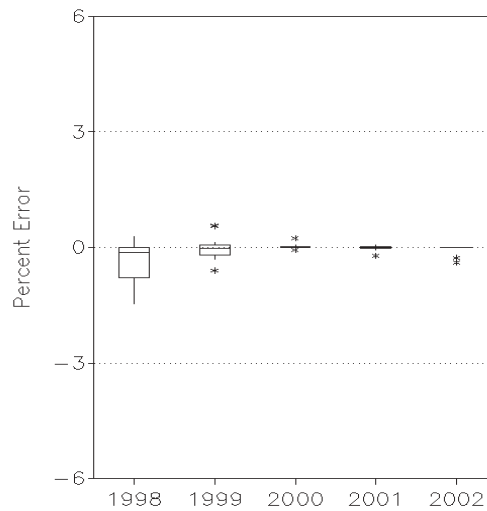
Source: Energy Information Administration, Petroleum Supply Reporting System.

The 2002 range (2.62) of MFW percent errors for jet fuel production, ranging from -1.59 to 1.03 percent, was the smallest range over the 5 years studied. The range (0.39) of the 2002 *PSM* percent errors was the smallest range of all other *PSM* plots analyzed for 2002. Two outliers in March (-0.27) and November (-0.39) were the only resubmissions that year.

Stocks

Figures FE6, FE7, and FE8 show the yearly distribution of percent errors for stocks of crude oil, motor gasoline, distillate fuel oil, residual fuel oil, jet fuel, and propane. Figure FE6 shows the box and whisker plots for crude oil stocks and motor gasoline stocks. The 2002 range (3.58) of MFW percent errors for crude oil stocks was the smallest range over the 5-year period. Similarly, the range (0.71) of the 2002 *PSM* percent errors for crude oil stocks was the smallest range over the 5 years. The 2002 median of -0.18 percent had the smallest absolute percent error over the 5-year period. One outlier in March (-0.62) was due to respondent revisions.

As in prior years, most of the 2002 MFW estimates for motor gasoline stocks underestimated the final *PSA* values. In contrast, most of the 2002 *PSM* interim values were overestimates. The percent errors for 2002 were closely distributed around the median of 0.03 percent. The outlier in December (0.78) was due to misreporting.

Figure FE7 shows box and whisker plots for distillate and residual fuel oil stocks. The 2002 median of 0.21 percent for the 2002 MFW percent errors for distillate fuel oil stocks was the smallest median and the first positive median over the 5 years studied. The 2002 median (0.04) of the *PSM* percent errors for distillate fuel oil stocks was also the smallest median over the 5-year period. One outlier in June (-1.67) was due to reporting problems.

Residual fuel oil stocks typically have larger percent errors than other stock series. The range (9.46) of the 2002 MFW percent errors was the largest range for the 5 years analyzed, but the median of 0.67 percent was the closest to zero. One outlier in November (-5.02) was due to company revisions. In contrast, the range (0.69) of the 2002 *PSM* percent errors for residual fuel oil stocks was the smallest range over the 5-year period. The 2002 percent errors were tightly distributed around the median of 0.06 percent.

The box and whisker plots for jet fuel stocks and propane stocks are shown in Figure FE8. The range (6.55) of the 2002 MFW percent errors for jet fuel stocks was similar to prior years. The 2002 median of -0.14 percent was the closest to zero. Similarly, the median (0.01) of the 2002 *PSM* percent errors for jet fuel stocks was the closest to zero. One outlier in December (1.44) was due to misreporting.

The range (6.75) of the 2002 MFW percent errors for propane stocks was the smallest range over the 5-year period, ranging from -2.72 to 4.03 percent. The median (-0.02) of the 2002 *PSM* percent errors for propane stocks was the closest to zero. Three outliers in January, September, and November were due to respondent revisions.

Imports

Figures FE9, FE10, and FE11 show the yearly distributions of percent errors for the imports of crude oil and four products: motor gasoline, distillate fuel oil, residual fuel oil, and jet fuel. Because of the irregularity of imports for crude oil and petroleum products, the magnitude and range of percent errors for both the MFW and the *PSM* imports numbers can be expected to be much larger and wider than for production and stocks.

Figure FE9 shows that the majority of the 2002 MFW estimates of crude oil imports underestimated the final *PSA* values. The 2002 range of 5.24 percent was the smallest range over the 5-year period. The one outlier in December (1.18) was probably a result of end-of-year adjustments. Similarly, the 2002 *PSM* interim values for crude oil imports underestimated the final *PSA* values. The smallest range over the 5-year period was 1.90 percent in 2002.

The distributions of percent errors of the MFW estimates and *PSM* interim values for 1998 through 2002 of motor gasoline and distillate fuel oil imports are shown in Figure FE10. Most of the 2002 MFW and *PSM* percent errors for motor gasoline imports were smaller than the percent errors for the prior 4 years and their medians were closest to zero. The ranges of the MFW (20.92) and *PSM* (5.05) percent errors were the smallest ranges over their respective 5-year periods.

Most of the 2002 MFW estimates for distillate fuel oil imports were underestimated. The 2002 median of -2.97 percent was the closest to zero. Unlike 2001, most of the 2002 *PSM* interim values for distillate fuel oil imports underestimated the final *PSA* values. The 2002 range of 8.99 percent was the smallest range in the past 5 years. Two outliers in March (2.14) and February (-6.85) were due to company misreporting.

Figure FE11 shows the box and whisker plots for residual fuel oil imports and jet fuel imports. Unlike 2001, most of the 2002 MFW estimates for residual fuel oil imports underestimated the final *PSA* values. All of the 2002 *PSM* interim values underestimated the final *PSA* values. One outlier in December (-0.39) was due to reporting problems.

The 2002 MFW range of 82.94 percent for jet fuel imports was the largest range of all other MFW plots analyzed for 2002. One outlier in December (60.00) was due to internal adjustments. The range (16.79) of the 2002 *PSM* percent errors for jet fuel imports was also the largest range of all other *PSM* plots analyzed for 2002. Two outliers in March (-13.76) and July (-13.04) were due to respondent revisions.
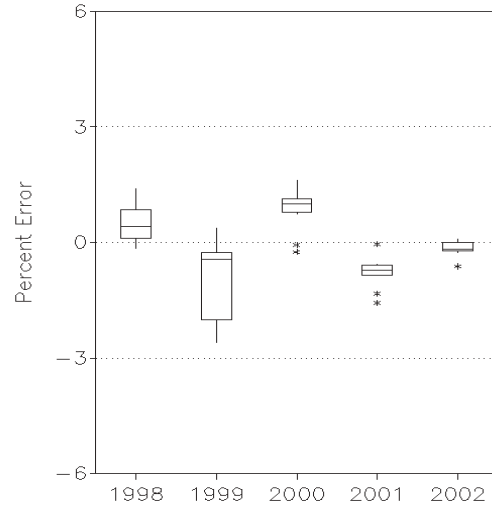
Figure FE6.   Range of Percent Errors for MFW and *PSM* Crude Oil Stocks Excluding Strategic Petroleum Reserve (SPR) and Motor Gasoline Stocks Data, 1998 -2002
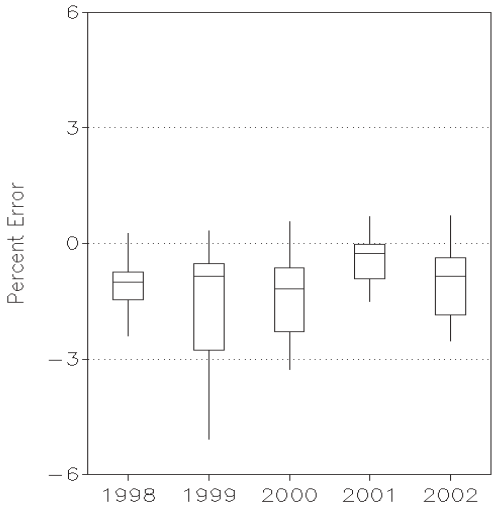
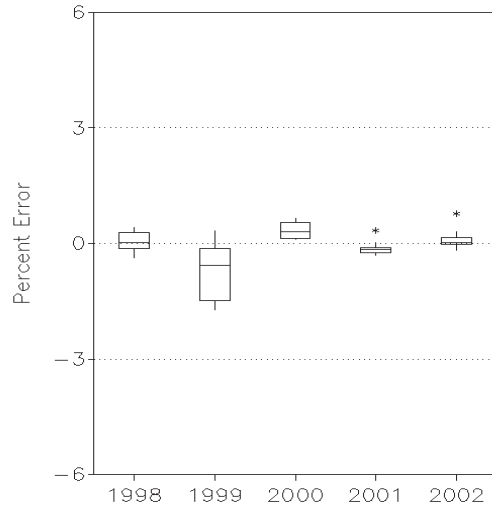**Crude Oil Stocks Excluding SPR**

MFW vs. *PSA*
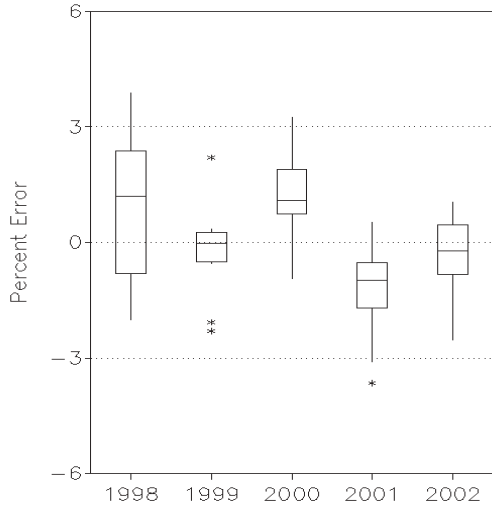
PSM vs. *PSA*

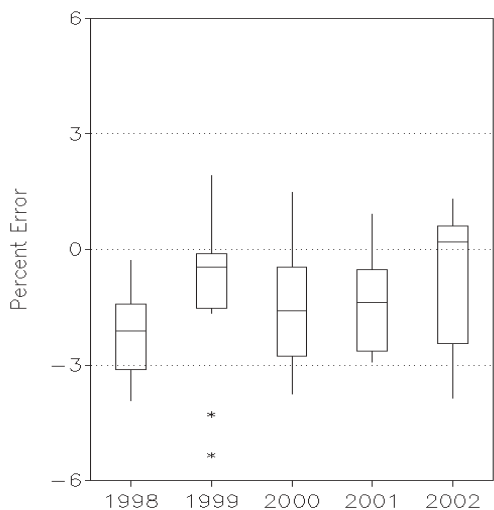**Motor Gasoline Stocks**

MFW vs. *PSA*

PSM vs. *PSA*

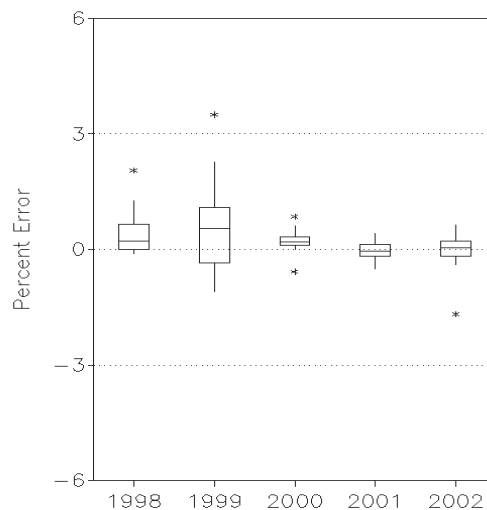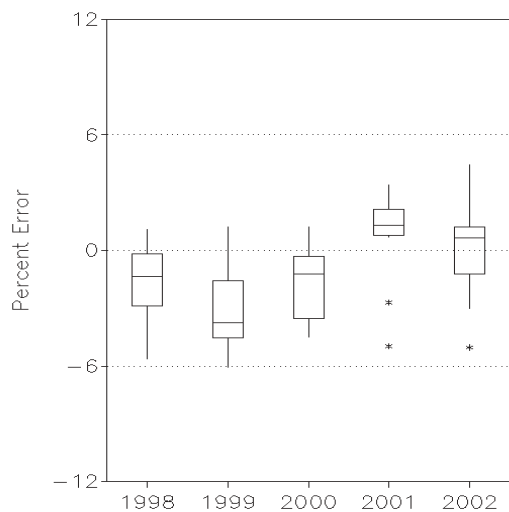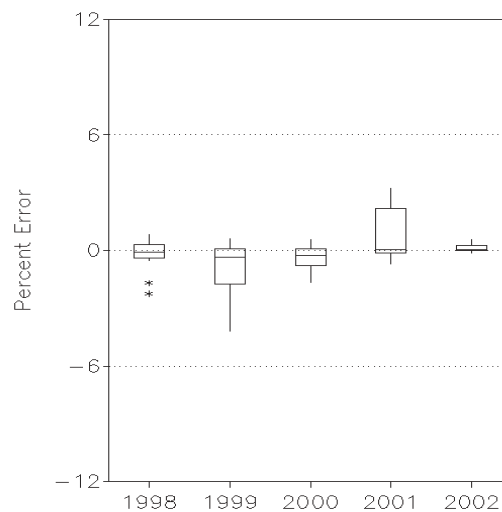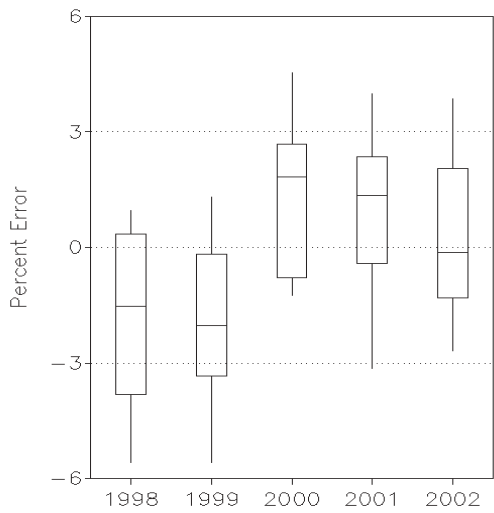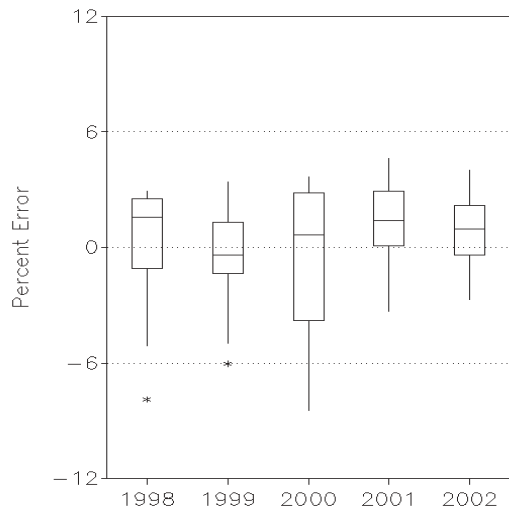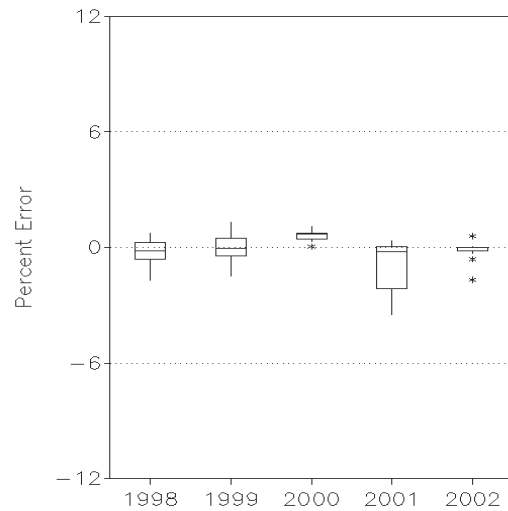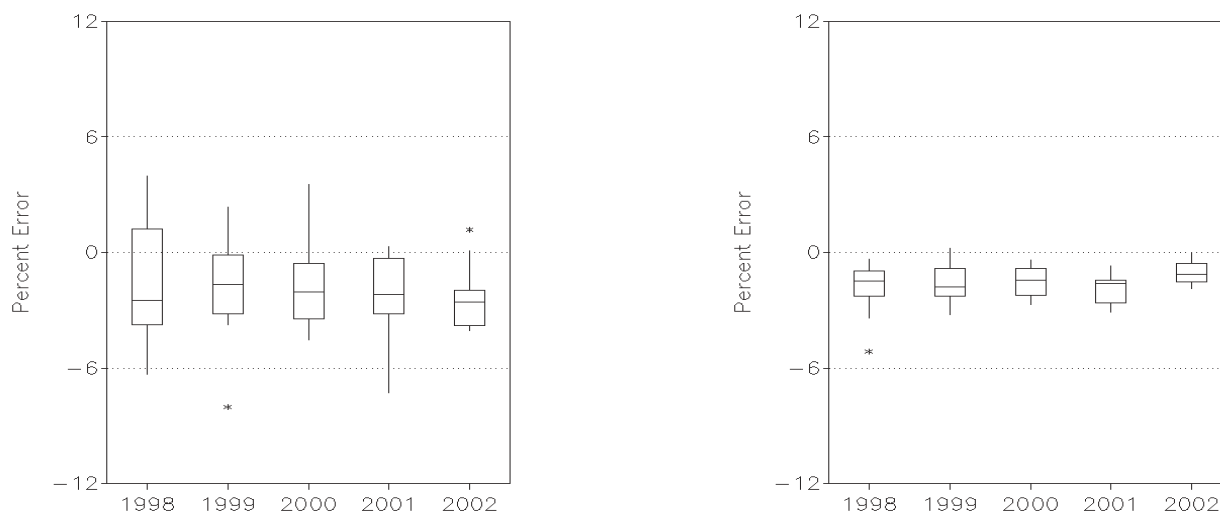Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE7.   Range of Percent Errors for MFW and *PSM* Distillate Fuel Oil and Residual Fuel Oil Stocks Data, 1998 - 2002

Distillate Fuel Oil Stocks

MFW vs. *PSA*

PSM vs. *PSA*

Residual Fuel Oil Stocks

MFW vs. *PSA*

PSM vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.

Energy Information Administration/Petroleum Supply Monthly, October 2003

Figure FE8.   Range of Percent Errors for MFW and *PSM* Jet Fuel Stocks and Propane Stocks Data, 1998 - 2002

Jet Fuel Stocks

MFW vs. *PSA*   *PSM* vs. *PSA*
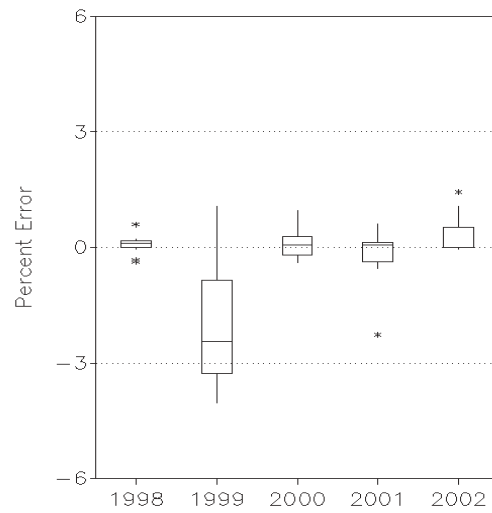
Propane Stocks

MFW vs. *PSA*   *PSM* vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE9.   Range of Percent Errors for MFW and *PSM* Crude Oil Imports Excluding SPR Data, 1998 - 2002

Conclusion

In summary, similar to previous years, the interim *PSM* data were closer in value to the final *PSA* volumes than the MFW estimates. This is largely a result of the longer time period provided to process the monthly data and monthly respondents' accounting systems.

In 2002, 44 of 66 *PSM* interim values were within 1 percent (mean absolute percent error) of the final values; 27 of 61 MFW estimates were within 2 percent (mean absolute percent error) of the final values; and 12 of those 27 were within 1 percent. As in previous years, the accuracy of 2002 preliminary and interim values varied by product and by petroleum supply type. As a group, stocks continued to have the most accurate MFW estimates and *PSM* interim values.

The good coverage for weekly surveys across petroleum supply type and product combinations has contributed to the accuracy of weekly estimates. In 2002, for 18 of the 21 categories, coverage was 90 percent or above. All but one of the 2002 response rates for the weekly and monthly surveys increased from the 2001 response rates.

To successfully maintain and improve the accuracy of these data, the Petroleum Division (PD) is participating in several Office of Oil and Gas initiatives in the areas of data collection, survey processing, and data dissemination. Some of the specific areas during 2002 included the development of the Data Collection Module (DCM) which allows data to be collected in a common system, the continuation of the development of the Standard Energy Processing System (STEPS) which is designed to handle different surveys with different needs using generalized programs and data structures to process survey data, the continuation of

nonresponse follow-up, the extension of customer outreach including industry brochures, the expansion and improvement of electronic data dissemination on the EIA web site, including many new user-friendly information retrieval options; and the continuation of efforts to insure compliance with reporting requirements.

The PD is continuously looking at other government agencies and private industry for best practices in the field of data collection and processing systems. The PD decided to implement STEPS with the goal of developing a system that would upgrade and unify legacy systems by incorporating state-of-the-art technology. The code was developed by the U.S. Census Bureau, and is written in SAS® and operated in a UNIX environment. The system includes integrated code to perform various survey processing activities including edit/imputation, data review and correction, estimation, variance estimation, and other data collection activities.

In 2002, the PD continued to expand the Survey Information System (SIS) which contains information needed for data validation and ad hoc queries. The system now includes both weekly imputed and reported values to help study and improve imputation techniques. The web products were also expanded and improved, including *This Week in Petroleum*. The PD dedicated resources to developing a 2004 forms clearance package that included more detailed breakouts for certain fuel categories, the addition of two new survey forms, the EIA-805, "Weekly Terminal Blenders Report," and the EIA-815, "Monthly Terminal Blenders Report," and the consolidation of the EIA-807 form into existing weekly forms. The EIA-805 and EIA-815 were specifically designed to collect blending activity on a terminal basis. Isolating this reporting from existing surveys should provide more accurate data. The results of all of these efforts should enable the PD to continue to improve the accuracy of weekly and monthly petroleum supply data estimates.
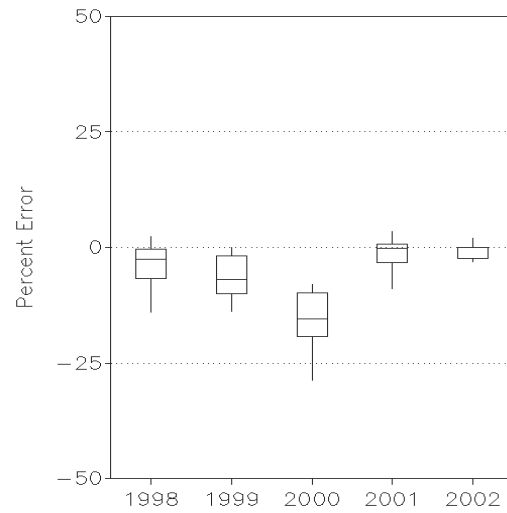
Figure FE10.   Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Imports Data, 1998 - 2002
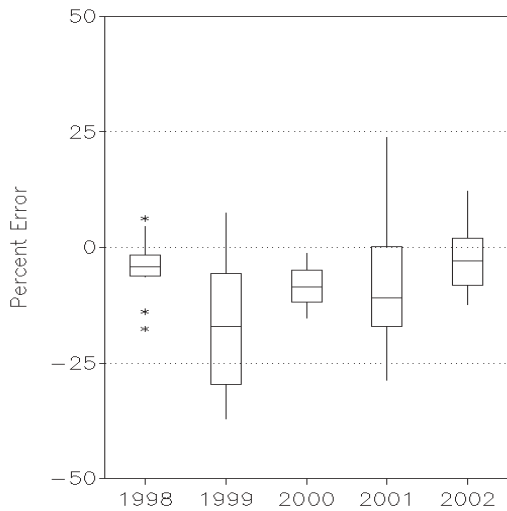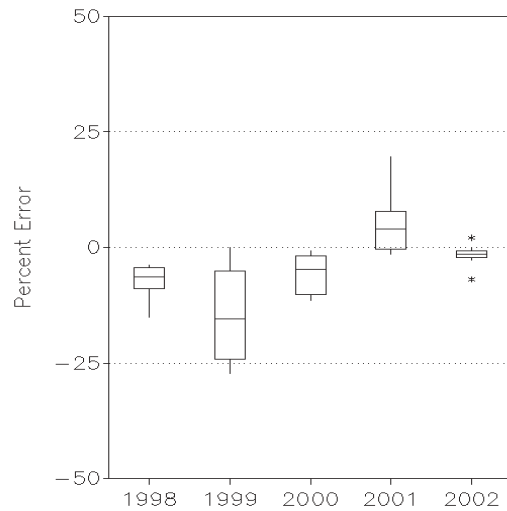
Motor Gasoline Imports

MFW vs. *PSA*
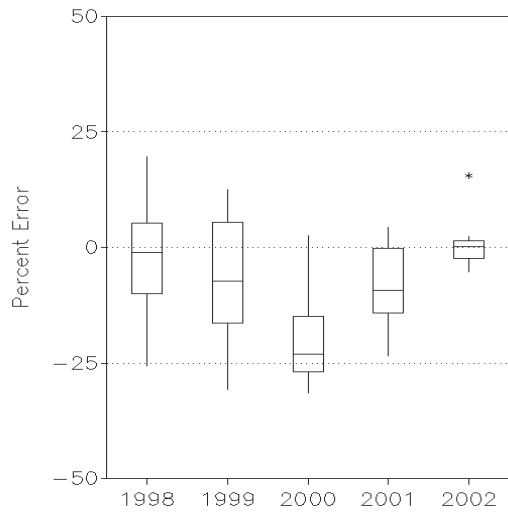
PSM vs. *PSA*

Distillate Fuel Oil Imports

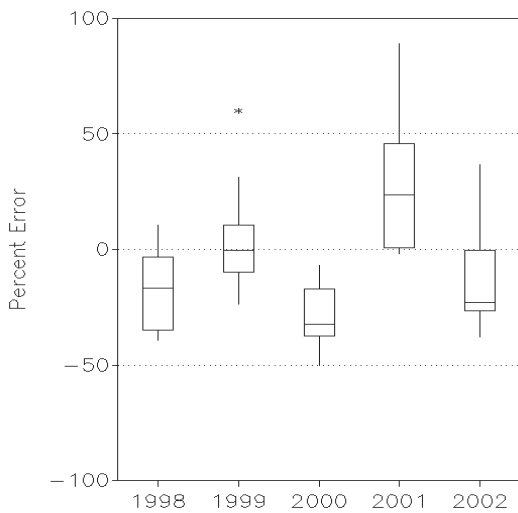#### MFW vs. *PSA*

PSM vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.
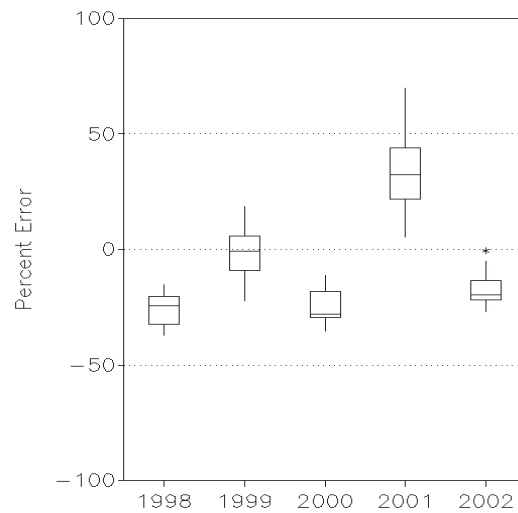
Figure FE11.   Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Imports Data, 1998 - 2002
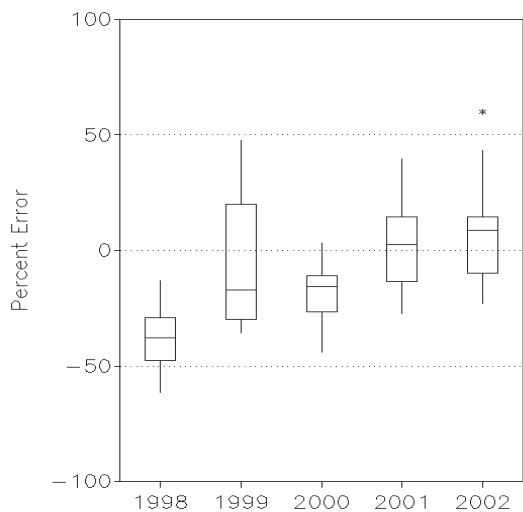
Residual Fuel Oil Imports
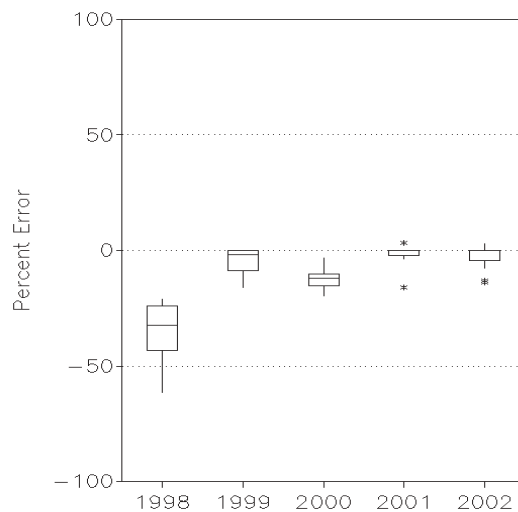
MFW vs. *PSA*

PSM vs. *PSA*

Jet Fuel Imports

MFW vs. *PSA*

PSM vs. *PSA*

Source: Energy Information Administration, Petroleum Supply Reporting System.

Energy Information Administration/Petroleum Supply Monthly, October 2003