Evaluation of Models
Environmental Protection Agency

Mr. John S. Irwin, NOAA Meteorologist
EPA Office of Air Quality Planning and Standards,
Air Quality Modeling Group

ABSTRACT

The EPA has a long history of model evaluation. The Clean Air Act of 1977 required the EPA to define models for air quality. As part of this effort, EPA developed model evaluation methods (Oreskes et al, 1994, Science (263), 641-646).

Initially the EPA ran into trouble by focusing on the intended use of the models. This is because of the uncertainties involved. When you evaluate a model for extreme values against observations then you are comparing garbage to reality. The answer to the problem was to create an evaluation procedure that uses statistical evaluation test methods to do three things:

- assure that the physics are right in the model (label the model as a "best performer")
- determine whether differences seen in the performance of the models are statistically significant (identify other models that may be performing as well, i.e., a "set of best performers")
- compare the models' results to each other in an open and fair competition in which the rules are known and the conclusions reached are objectively determined

Once the "set of best performers" has been defined, then a new set of statistical evaluation test methods would be used to determine which of these models best performs the user-defined tasks. The following test methods provide promising results:

- grouped data
- decomposed time series
- process analyses.

The American Society for Testing and Materials (ASTM) is developing model evaluation standard guidelines. Those interested in model evaluation should consider joining the ASTM and to contributing to this effort.