

"known" conditions that are input to the models, e.g., wind speed, mixed layer height, surface heat flux, emission characteristics, etc. However, in addition to the known conditions, there are unmeasured or unknown variations in the conditions of this event, e.g., unresolved details of the atmospheric flow such as the turbulent velocity field. These unknown conditions, may vary among repetitions of the event. As a result, deviations in observed concentrations from their ensemble average, and from the concentrations estimated by the model, are likely to occur even though the known conditions are fixed. Even with a *perfect model* that predicts the correct ensemble average, there are likely to be deviations from the observed concentrations in individual repetitions of the event, due to variations in the unknown conditions. The statistics of these concentration residuals are termed "inherent" uncertainty. Available evidence suggests that this source of uncertainty alone may be responsible for a typical range of variation in concentrations of as much as ± 50 percent.¹²⁴

b. Moreover, there is "reducible" uncertainty¹¹⁵ associated with the model and its input conditions; neither models nor data bases are perfect. Reducible uncertainties are caused by: (1) Uncertainties in the input values of the known conditions (i.e., emission characteristics and meteorological data); (2) errors in the measured concentrations which are used to compute the concentration residuals; and (3) inadequate model physics and formulation. The "reducible" uncertainties can be minimized through better (more accurate and more representative) measurements and better model physics.

c. To use the terminology correctly, reference to model accuracy should be limited to that portion of reducible uncertainty which deals with the physics and the formulation of the model. The accuracy of the model is normally determined by an evaluation procedure which involves the comparison of model concentration estimates with measured air quality data.¹²⁵ The statement of accuracy is based on statistical tests or performance measures such as bias, noise, correlation, etc.¹⁶ However, information that allows a distinction between contributions of the various elements of inherent and reducible uncertainty is only now beginning to emerge.²⁴ As a result most discussions of the accuracy of models make no quantitative distinction between (1) limitations of the model versus (2) limitations of the data base and of knowledge concerning atmospheric variability. The reader should be aware that statements on model accuracy and uncertainty may imply the need for improvements in model performance that even the "perfect" model could not satisfy.

10.1.2 Studies of Model Accuracy

a. A number of studies^{126,127} have been conducted to examine model accuracy, particularly with respect to the reliability of short-term concentrations required for ambient standard and increment evaluations. The results of these studies are not surprising. Basically, they confirm what expert atmospheric scientists have said for

some time: (1) Models are more reliable for estimating longer time-averaged concentrations than for estimating short-term concentrations at specific locations; and (2) the models are reasonably reliable in estimating the magnitude of highest concentrations occurring sometime, somewhere within an area. For example, errors in highest estimated concentrations of ± 10 to 40 percent are found to be typical^{128,129}, i.e., certainly well within the often quoted factor-of-two accuracy that has long been recognized for these models. However, estimates of concentrations that occur at a specific time and site, are poorly correlated with actually observed concentrations and are much less reliable.

b. As noted above, poor correlations between paired concentrations at fixed stations may be due to "reducible" uncertainties in knowledge of the precise plume location and to unquantified inherent uncertainties. For example, Pasquill¹³⁰ estimates that, apart from data input errors, maximum ground-level concentrations at a given hour for a point source in flat terrain could be in error by 50 percent due to these uncertainties. Uncertainty of five to 10 degrees in the measured wind direction, which transports the plume, can result in concentration errors of 20 to 70 percent for a particular time and location, depending on stability and station location. Such uncertainties do not indicate that an estimated concentration does not occur, only that the precise time and locations are in doubt.

10.1.3 Use of Uncertainty in Decision-Making

a. The accuracy of model estimates varies with the model used, the type of application, and site specific characteristics. Thus, it is desirable to quantify the accuracy or uncertainty associated with concentration estimates used in decision-making. Communications between modelers and decision-makers must be fostered and further developed. Communications concerning concentration estimates currently exist in most cases, but the communications dealing with the accuracy of models and its meaning to the decision-maker are limited by the lack of a technical basis for quantifying and directly including uncertainty in decisions. Procedures for quantifying and interpreting uncertainty in the practical application of such concepts are only beginning to evolve; much study is still required.^{115,116,117,131,132}

b. In all applications of models an effort is encouraged to identify the reliability of the model estimates for that particular area and to determine the magnitude and sources of error associated with the use of the model. The analyst is responsible for recognizing and quantifying limitations in the accuracy, precision and sensitivity of the procedure. Information that might be useful to the decision-maker in recognizing the seriousness of potential air quality violations includes such model accuracy estimates as accuracy of peak predictions, bias, noise, correlation, frequency distribution, spatial extent of high concentration, etc. Both space/time pairing of estimates and measurements and unpaired comparisons are recommended. Emphasis should be on the

highest concentrations and the averaging times of the standards or increments of concern. Where possible, confidence intervals about the statistical values should be provided. However, while such information can be provided by the modeler to the decision-maker, it is unclear how this information should be used to make an air pollution control decision. Given a range of possible outcomes, it is easiest and tends to ensure consistency if the decision-maker confines his judgement to use of the "best estimate" provided by the modeler (i.e., the design concentration estimated by a model recommended in the *Guideline* or an alternate model of known accuracy). This is an indication of the practical limitations imposed by current abilities of the technical community.

c. To improve the basis for decision-making, EPA has developed and is continuing to study procedures for determining the accuracy of models, quantifying the uncertainty, and expressing confidence levels in decisions that are made concerning emissions controls.^{133,134} However, work in this area involves "breaking new ground" with slow and sporadic progress likely. As a result, it may be necessary to continue using the "best estimate" until sufficient technical progress has been made to meaningfully implement such concepts dealing with uncertainty.

10.1.4 Evaluation of Models

a. A number of actions have been taken to ensure that the best model is used correctly for each regulatory application and that a model is not arbitrarily imposed. First, the *Guideline* clearly recommends the most appropriate model be used in each case. Preferred models, based on a number of factors, are identified for many uses. General guidance on using alternatives to the preferred models is also provided. Second, the models have been subjected to a systematic performance evaluation and a peer scientific review. Statistical performance measures, including measures of difference (or residuals) such as bias, variance of difference and gross variability of the difference, and measures of correlation such as time, space, and time and space combined as recommended by the AMS Woods Hole Workshop¹⁶, were generally followed. Third, more specific information has been provided for justifying the site specific use of alternative models in previously cited EPA guidance^{22,25}, and new models are under consideration and review.²⁴ Together these documents provide methods that allow a judgement to be made as to what models are most appropriate for a specific application. For the present, performance and the theoretical evaluation of models are being used as an indirect means to quantify one element of uncertainty in air pollution regulatory decisions.

b. EPA has participated in a series of conferences entitled, "Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes,"¹³⁵ for the purpose of promoting the development of improved methods for the characterization of model performance. There is a consensus developing on what should be considered in the evaluation of air quality models¹³⁶,