Designation: D5157 – 97 (Reapproved 2014)

Standard Guide for Statistical Evaluation of Indoor Air Quality Models[1]

This standard is issued under the fixed designation D5157; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide provides quantitative and qualitative tools for evaluation of indoor air quality (IAQ) models. These tools include methods for assessing overall model performance as well as identifying specific areas of deficiency. Guidance is also provided in choosing data sets for model evaluation and in applying and interpreting the evaluation tools. The focus of the guide is on end results (that is, the accuracy of indoor concentrations predicted by a model), rather than operational details such as the ease of model implementation or the time required for model calculations to be performed.

1.2 Although IAQ models have been used for some time, there is little guidance in the technical literature on the evaluation of such models. Evaluation principles and tools in this guide are drawn from past efforts related to outdoor air quality or meteorological models, which have objectives similar to those for IAQ models and a history of evaluation literature.(**1**)[2] Some limited experience exists in the use of these tools for evaluation of IAQ models.

2. Referenced Documents

2.1 *ASTM Standards:*[3]

D1356 Terminology Relating to Sampling and Analysis of Atmospheres

3. Terminology

3.1 *Definitions:* For definitions of terms used in this standard, refer to Terminology D1356.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *IAQ model, n*—an equation, algorithm, or series of equations/algorithms used to calculate average or time-varying pollutant concentrations in one or more indoor chambers for a specific situation.

3.2.2 *model bias, n*—a systematic difference between model predictions and measured indoor concentrations (for example, the model prediction is generally higher than the measured concentration for a specific situation).

3.2.3 *model chamber, n*—an indoor airspace of defined volume used in model calculations; IAQ models can be specified for a single chamber or for multiple, interconnected chambers.

3.2.4 *model evaluation, n*—a series of steps through which a model developer or user assesses a model's performance for selected situations.

3.2.5 *model parameter, n*—a mathematical term in an IAQ model that must be estimated by the model developer or user before model calculations can be performed.

3.2.6 *model residual, n*—the difference between an indoor concentration predicted by an IAQ model and a representative measurement of the true indoor concentration; the value should be stated as positive or negative.

3.2.7 *model validation, n*—a series of evaluations undertaken by an agency or organization to provide a basis for endorsing a specific model (or models) for a specific application (or applications).

3.2.8 *pollutant concentration, n*—the extent of the occurrence of a pollutant or the parameters describing a pollutant in a defined airspace, expressed in units characteristic to the pollutant (for example, mg/m^3, ppm, Bq/m^3, area/m^3, or colony forming units per cubic metre).

4. Significance and Use

4.1 Using the tools described in this guide, an individual seeking to apply an IAQ model should be able to (*1*) assess the performance of the model for a specific situation or (*2*) recognize or assess its advantages and limitations.

4.2 This guide can also be used for identifying specific areas of model deficiency that require further development or refinement.

5. Components of Model Evaluation

5.1 The components of model evaluation include the following: (*1*) stating the purpose(s) or objective(s) of the evaluation, (*2*) acquiring a basic understanding of the specification and underlying principles or assumptions, (*3*) selecting

[1] This guide is under the jurisdiction of ASTM Committee D22 on Air Quality and is the direct responsibility of Subcommittee D22.05 on Indoor Air.

Current edition approved Sept. 1, 2014. Published September 2014. Originally approved in 1991. Last previous edition approved in 2008 as D5157 – 97 (2008). DOI: 10.1520/D5157-97R14.

[2] The boldface numbers in parentheses refer to the list of references at the end of this standard.

[3] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

data sets as inputs to the evaluation process, and (*4*) selecting and using appropriate tools for assessing model performance. Just as model evaluation has multiple components, model validation consists of one or more evaluations. However, model validation is beyond the scope of this document.

5.1.1 *Establishing Evaluation Objectives:*

5.1.1.1 IAQ models are generally used for the following: (*1*) to help explain the temporal and spatial variations in the occurrences of indoor pollutant concentrations, (*2*) to improve the understanding of major influencing factors or underlying physical/chemical processes, and (*3*) to predict the temporal/spatial variations in indoor concentrations that can be expected to occur in specific types of situations. However, model evaluation relates only to the third type of model use—prediction of indoor concentrations.

5.1.1.2 The most common evaluation objectives are (*1*) to compare the performance of two or more models for a specific situation or set of situations and (*2*) to assess the performance of a specific model for different situations. Secondary objectives include identifying specific areas of model deficiency. Determination of specific objectives will assist in choosing appropriate data sets and quantitative or qualitative tools for model evaluation.

5.1.2 *Understanding the Model(s) to be Evaluated:*

5.1.2.1 Although a model user will not necessarily know or understand all details of a particular model, some fundamental understanding of the underlying principles and concepts is important to the evaluation process. Thus, before evaluating a model, the user should develop some understanding of the basis for the model and its operation. IAQ models can generally be distinguished by their basis, by the range of pollutants they can address, and by the extent of temporal or spatial detail they can accommodate in inputs, calculations, and outputs.

5.1.2.2 Theoretical models are generally based on physical principles such as mass conservation. (**2, 3**) That is, a mass balance is maintained to keep track of material entering and leaving a particular airspace. Within this conceptual framework, pollutant concentrations are increased by emissions within the defined volume and by transport from other airspaces, including outdoors. Similarly, concentrations are decreased by transport exiting the airspace, by removal to chemical/physical sinks within the airspace, or for reactive species, by conversion to other forms. Relationships are most often specified through a differential equation quantifying factors related to contaminant gain or loss.

5.1.2.3 Empirical models (**3**) are generally based on approaches such as least-squares regression analysis, using measurements under different conditions across a variety of structures, at different times within the same structure, or both. Theoretical models will generally be suitable for a wide range of applications, whereas empirical models will generally be applicable only within the range of measurements from which they were developed.

5.1.2.4 Some combination of theoretical and empirical components is also possible. Specific parameters of a theoretical model may have relationships with other factors that can be more easily quantified than the parameters themselves. For example, the rate of air infiltration into a structure could depend on outdoor windspeed and the indoor-outdoor temperature difference, or the emission rate from a cigarette could depend on the combustion rate and the constituents of the particular brand smoked. Given sufficient data, such relationships could be estimated through techniques such as regression analysis.

5.1.2.5 IAQ models may be specified for a particular pollutant or in general terms; this distinction is important, for example, because particle-phase pollutants behave differently from gas-phase pollutants. Particulate matter is subject to coagulation, chemical reaction at surfaces, gravitational settling, diffusional deposition, resuspension and interception, impaction, and diffusional removal by filtration devices; whereas some gaseous pollutants are subject to sorption and, in some cases, desorption processes.

5.1.2.6 Dynamic IAQ models predict time-varying indoor concentrations for time steps that are usually on the order of seconds, minutes, or hours; whereas integrated models predict time-averaged indoor concentrations using average values for each input parameter or averaging these parameters during the course of exercising the model. Models can also differ in the extent of partitioning of the indoor airspace, with the simplest models treating the entire indoor volume as a single chamber or zone assumed to have homogeneous concentrations throughout; more complex models can treat the indoor volume as a series of interconnected chambers, with a mass balance conducted without each chamber and consideration given to communicating airflows among chambers.

5.1.2.7 Generally speaking, as the model complexity grows in terms of temporal detail, number of chambers, and types of parameters that can be used for calculations, the user's task of supplying appropriate inputs becomes increasingly demanding. Thus users must have a basic understanding of the underlying principles, nature and extent of inputs required, inherent limitations, and types of outputs provided so that they can choose a level of model complexity providing an appropriate balance between input effort and output detail.

5.1.2.8 A number of assumptions are usually made when modeling a complex environment such as the indoor airspace. These assumptions, and their potential influence on the modeling results, should be identified in the evaluation process. One method of gaining insights is by performing sensitivity analysis. An example of this technique is to systematically vary the values of one input parameter at a time to determine the effect of each on the modeling results; each parameter should be varied over a reasonable range of values likely to be encountered for the specific situation(s) of interest.

5.1.3 *Choosing Data Sets for Model Evaluation:*

5.1.3.1 A fundamental requirement for model evaluation is that the data used for the evaluation process should be independent of the data used to develop the model. This constraint forces a search for available data pertinent to the planned application or, if no appropriate data sets can be found, collection of new data to support the evaluation process. Such data should be collected according to commonly recognized and accepted methods, such as those given in the compendium developed by the U.S. Environmental Protection Agency (**4**).

5.1.3.2 The following series of steps should be used in choosing data sets for model evaluation: (*1*) select situations for applying and testing the model; (*2*) note the model input parameters that require estimation for the situations selected; (*3*) determine the required levels of temporal detail (for example, minute-by-minute or hour-by-hour) and spatial detail (that is, number of chambers) for model application as well as variations of the contaminants within each chamber; and (*4*) find or collect appropriate data for estimation of the model inputs and comparison with the model outputs.

5.1.3.3 Thus, the information required for the evaluation process includes not only measured indoor concentrations at an appropriate level of temporal detail, but also suitable estimates for required input parameters. Among the inputs typically required are outdoor concentrations, indoor emission and sink rates, coagulation coefficients, deposition rates and diffusion coefficients for particles, and rates of airflow between indoor and outdoor airspaces (as well as flows among multiple indoor airspaces, if a multichamber model is used). If suitable data to support the choice of inputs are not available, the alternatives are as follows: (*1*) to compress the level of temporal detail for model application to that for which suitable data can be obtained; (*2*) to provide best estimates for model inputs, recognizing the limitations imposed by this particular approach; or (*3*) to collect the additional data required to enable proper estimation of inputs.

5.1.4 *Tools for Assessing Model Performance:*

5.1.4.1 The tools to be used in assessing the performance of IAQ models all involve comparisons between indoor concentrations predicted by the model, C_p, and observed concentrations, C_o, comprising the data set(s) used for evaluation. These tools can be quantitative, involving various types of statistical indexes, or qualitative, involving plots of C_p, C_o, or differences between the two (that is, model residuals). The tools presented below are classified by use for (*1*) assessing the general agreement between predicted and observed concentrations and (*2*) assessing bias in the mean or variance of predicted values relative to that for observed values.

5.1.4.2 The following tools are to be used for assessing the general agreement between C_p and C_o:

(*1*) Correlation coefficient, *r*, ranging from −1 to 1, with 1 indicating a strong, direct relationship between C_p and C_o, 0 indicating no relationship, and − 1 indicating a strong but inverse relationship. The formula to be used for calculating this coefficient (**5, 6**) is as follows:

$$r = \sum_{i=1}^{n}\left[\left(C_{oi}-\bar{C}_o\right)\left(C_{pi}-\bar{C}_p\right)\right] / \sqrt{\sum_{i=1}^{n}\left[\left(C_{oi}-\bar{C}_o\right)^2\right]\left[\sum_{i=1}^{n}\left(C_{pi}-\bar{C}_p\right)^2\right]} \quad (1)$$

where the summation extends across all C_p and C_o pairs and $\bar{C}_o$ and $\bar{C}_p$ are averages (that is, $\bar{C}_o = \sum_{i=1}^{n} C_{oi}/n$, where *n* is the number of observed values).

(*2*) Line of regression, the best-fit relationship between C_p and C_o, ideally exhibiting a slope, *b*, of one and an intercept, *a*, of zero. Formulas to be used in calculating the slope and intercept are as follows:

$$b = \sum_{i=1}^{n}\left[\left(C_{oi}-\bar{C}_o\right)\left(C_{pi}-\bar{C}_p\right)\right] / \sum_{i=1}^{n}\left[\left(C_{oi}-\bar{C}_o\right)^2\right] \quad (2)$$

$$a = \bar{C}_p - \left[(b)\left(\bar{C}_o\right)\right] \quad (3)$$

(*3*) Normalized mean square error (NMSE), a measure of the magnitude of prediction error relative to C_p and C_o. The formula to be used for calculating this measure[2] is as follows:

$$NMSE = \overline{(C_p - C_o)^2} / \left[\left(\bar{C}_o\right)\left(\bar{C}_p\right)\right] \quad (4)$$

where:

$$\overline{(C_p - C_o)^2} = \sum_{i=1}^{n}(C_{pi} - C_{oi})^2/n.$$

The NMSE will have a value of 0 when there is perfect agreement for all pairs of C_p and C_o and will tend toward higher values as C_p and C_o differ by greater magnitudes. For example, if C_p and C_o differ consistently by 50 %, the NMSE value will be near 0.2; for differences of 100 %, the NMSE value will be near 0.5; for differences of one order of magnitude, the NMSE value will be near 8.0. In addition to these quantitative tools, a qualitative tool to be used is a plot of C_p and C_o over time. This plot will indicate not only the general extent of agreement between C_p and C_o but also the specific areas of disagreement. Model residuals can also be plotted over time or against predicted or observed concentrations (after ordering the concentrations from lowest to highest); such a plot should indicate no distinct trend or pattern. If a trend or pattern is discerned, possible reasons for the trend should be identified and investigated.

5.1.4.3 The following tools are to be used for assessing bias:

(1) Normalized or fractional bias (FB) of the mean concentrations. This statistic is to be calculated as follows:

$$FB = 2\cdot\left(\bar{C}_p - \bar{C}_o\right)/\left(\bar{C}_p + \bar{C}_o\right) \quad (5)$$

The FB will have a value of 0 when $\bar{C}_p$ and $\bar{C}_o$ agree perfectly and will tend towards −2 or 2 as these quantities differ by greater magnitudes.

(2) A similar index of bias (FS) based on the variance, σ^2, of the concentrations. This statistic is to be calculated as follows:

$$FS = 2\cdot\left(\sigma^2_{C_p} - \sigma^2_{C_o}\right)/\left(\sigma^2_{C_p} + \sigma^2_{C_o}\right) \quad (6)$$

(3) Bias in the mean of the highest 10 % of concentrations, FB_{10}. This statistic is to be calculated in the same manner as FB but using only the highest decile of observed concentrations.

6. Considerations in Applying Model Evaluation Tools

6.1 The results obtained from applying the model evaluation tools can be used for the following: (*1*) to compare the performance of two or more models for a single situation, (*2*) to compare the performance of a single model for a variety of situations, or (*3*) to compare the performance of multiple models for multiple situations. In reaching final conclusions, the evaluator must determine the situations and performance aspects that are of greatest importance.

6.2 In evaluating model performance, the collective evidence provided by all model evaluation tools should be considered; otherwise, misleading conclusions could result. For example, if model predictions and measured concentrations differed systematically by a factor of two, a correlation coefficient near unity would be obtained, but the regression slope, FB, and FS would reflect the systematic differences. Similarly, if the predictions and measurements coincided on the average but diverged widely for a subset of data pairs, the FB would be close to zero and the regression slope could be close to unity, but the correlation coefficient and NMSE would reflect the cases of divergence.

6.3 Discrepancies between model predictions and measured concentrations can be caused by uncertainties in the measurement process as well as by incorrect model predictions. For example, if the model predictions matched the true concentrations exactly, but the measurement process had a random-error component causing deviations up to ±25 % from the true concentrations, the correlation coefficient would be near 0.95, the NMSE near 0.03, the FB could range from −0.1 to 0.1, and the FS could range from −0.3 to 0.3. Similarly, if the model predictions matched the true concentrations exactly, but the measurement process had a negative bias causing systematic deviations of −10 %, the correlation coefficient would be unity, but the regression slope would be near 0.9, the NMSE near 0.015, the FB near 0.1, and the FS near 0.2.

6.4 Considering the potential consequences of measurement uncertainties, the following values can be taken as generally indicative of adequate model performance:

(1) Correlation coefficient of 0.9 or greater,
(2) Regression slope between 0.75 and 1.25,
(3) Regression intercept 25 % or less of the average measured concentration,
(4) NMSE of 0.25 or lower,
(5) FB of 0.25 or lower, and
(6) FS of 0.5 or lower.

As the community of IAQ model developers and users gains experience in conducting model evaluations with these tools, it may be possible to reach a consensus on the range of values associated with excellent, good, marginal, or unsatisfactory model performance.

7. Keywords

7.1 indoor air quality; model; model performance; qualitative evaluation; quantitative evaluation; statistical

REFERENCES

(1) Hanna, S. R., “Air Quality Model Evaluation and Uncertainty,” *Journal of the Air Pollution Control Association*, Vol 38, 1988, pp. 406–412.

(2) Nagda, N. L., Rector, H. E., and Koontz, M. D., *Guidelines for Monitoring Indoor Air Quality* , Chapter 3: “Factors Affecting Indoor Air Quality,” Hemisphere Publishing Corporation, New York, NY, 1987.

(3) Wadden, R. A., and Scheff, P. A., *Indoor Air Pollution: Characterization, Prediction, and Control*, Chapter 6: “Air Quality Models,” John Wiley and Sons, New York, NY, 1983.

(4) U.S. Environmental Protection Agency, 1990, *Compendium of Methods for the Determination of Air Pollutants in Indoor Air* , Report No. EPA/600/4-90-010, Research Triangle Park, NC.

(5) W. J. Dixon and F. J. Massey, Jr., 1969. *Introduction to Statistical Analysis*, McGraw-Hill Book Company, New York, NY.

(6) N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, Inc., New York, NY.