



Designation: D6589 – 23

Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance¹

This standard is issued under the fixed designation D6589; 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 techniques that are useful for the comparison of modeled air concentrations with observed field data. Such comparisons provide a means for assessing a model's performance, for example, bias and precision or uncertainty, relative to other candidate models. Methodologies for such comparisons are yet evolving; hence, modifications will occur in the statistical tests and procedures and data analysis as work progresses in this area. Until the interested parties agree upon standard testing protocols, differences in approach will occur. This guide describes a framework, or philosophical context, within which one determines whether a model's performance is significantly different from other candidate models. It is suggested that the first step should be to determine which model's estimates are closest on average to the observations, and the second step would then test whether the differences seen in the performance of the other models are significantly different from the model chosen in the first step. An example procedure is provided in [Appendix X1](#) to illustrate an existing approach for a particular evaluation goal. This example is not intended to inhibit alternative approaches or techniques that will produce equivalent or superior results. As discussed in [Section 6](#), statistical evaluation of model performance is viewed as part of a larger process that collectively is referred to as model evaluation.

1.2 This guide has been designed with flexibility to allow expansion to address various characterizations of atmospheric dispersion, which might involve dose or concentration fluctuations, to allow development of application-specific evaluation schemes, and to allow use of various statistical comparison metrics. No assumptions are made regarding the manner in which the models characterize the dispersion.

1.3 The focus of this guide is on end results, that is, the accuracy of model predictions and the discernment of whether differences seen between models are significant, rather than

operational details such as the ease of model implementation or the time required for model calculations to be performed.

1.4 This guide offers an organized collection of information or a series of options and does not recommend a specific course of action. This guide cannot replace education or experience and should be used in conjunction with professional judgment. Not all aspects of this guide may be applicable in all circumstances. This guide is not intended to represent or replace the standard of care by which the adequacy of a given professional service must be judged, nor should it be applied without consideration of a project's many unique aspects. The word "Standard" in the title of this guide means only that the document has been approved through the ASTM consensus process.

1.5 This standard applies to gaussian plume models; it may not be applicable to non-point sources, heavy gas models from evaporation from pool (for example, liquid spills), as well as near-field receptors.

1.6 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this guide.

1.7 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.8 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*²
[D1356 Terminology Relating to Sampling and Analysis of Atmospheres](#)

¹ This guide is under the jurisdiction of ASTM Committee [D22](#) on Air Quality and is the direct responsibility of Subcommittee [D22.11](#) on Meteorology.

Current edition approved Sept. 1, 2023. Published September 2023. Originally approved in 2000. Last previous edition approved in 2015 as D6589 – 05 (2015). DOI: 10.1520/D6589-23.

² 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.

3. Terminology

3.1 *Definitions*—For definitions of terms used in this guide, refer to Terminology [D1356](#).

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *atmospheric dispersion model, n*—an idealization of atmospheric physics and processes to calculate the magnitude and location of pollutant concentrations based on fate, transport, and dispersion in the atmosphere. This may take the form of an equation, algorithm, or series of equations/algorithms used to calculate average or time-varying concentration. The model may involve numerical methods for solution.

3.2.2 *dispersion, absolute, n*—the characterization of the spreading of material released into the atmosphere based on a coordinate system fixed in space.

3.2.3 *dispersion, relative, n*—the characterization of the spreading of material released into the atmosphere based on a coordinate system that is relative to the local median position of the dispersing material.

3.2.4 *evaluation objective, n*—a feature or characteristic, which can be defined through an analysis of the observed concentration pattern, for example, maximum centerline concentration or lateral extent of the average concentration pattern as a function of downwind distance, which one desires to assess the skill of the models to reproduce.

3.2.5 *evaluation procedure, n*—the analysis steps to be taken to compute the value of the evaluation objective from the observed and modeled patterns of concentration values.

3.2.6 *fate, n*—the destiny of a chemical or biological pollutant after release into the environment.

3.2.7 *model input value, n*—characterizations that must be estimated or provided by the model developer or user before model calculations can be performed.

3.2.8 *regime, n*—a repeatable narrow range of conditions, defined in terms of model input values, which may or may not be explicitly employed by all models being tested, needed for dispersion model calculations. It is envisioned that the dispersion observed should be similar for all cases having similar model input values.

3.2.9 *uncertainty, n*—refers to a lack of knowledge about specific factors or parameters. This includes measurement errors, sampling errors, systematic errors, and differences arising from simplification of real-world processes. In principle, uncertainty can be reduced with further information or knowledge ([1](#)).³

3.2.10 *variability, n*—refers to differences attributable to true heterogeneity or diversity in atmospheric processes that result in part from natural random processes. Variability usually is not reducible by further increases in knowledge, but it can in principle be better characterized ([1](#)).

4. Summary of Guide

4.1 Statistical evaluation of dispersion model performance with field data is viewed as part of a larger process that collectively is called model evaluation. Section [6](#) discusses the components of model evaluation.

4.2 To statistically assess model performance, one must define an overall evaluation goal or purpose. This will suggest features (evaluation objectives) within the observed and modeled concentration patterns to be compared, for example, maximum surface concentrations, lateral extent of a dispersing plume. The selection and definition of evaluation objectives typically are tailored to the model's capabilities and intended uses. The very nature of the problem of characterizing air quality and the way models are applied make one single or absolute evaluation objective impossible to define that is suitable for all purposes. The definition of the evaluation objectives will be restricted by the limited range conditions experienced in the available comparison data suitable for use. For each evaluation objective, a procedure will need to be defined that allows definition of the evaluation objective from the available observations of concentration values.

4.3 In assessing the performance of air quality models to characterize a particular evaluation objective, one should consider what the models are capable of providing. As discussed in Section [7](#), most models attempt to characterize the ensemble average concentration pattern. If such models should provide favorable comparisons with observed concentration maxima, this is resulting from happenstance, rather than skill in the model; therefore, in this discussion, it is suggested a model be assessed on its ability to reproduce what it was designed to produce, for at least in these comparisons, one can be assured that zero bias with the least amount of scatter is by definition good model performance.

4.4 As an illustration of the principles espoused in this guide, a procedure is provided in [Appendix X1](#) for comparison of observed and modeled near-centerline concentration values, which accommodates the fact that observed concentration values include a large component of stochastic, and possibly deterministic, variability unaccounted for by current models. The procedure provides an objective statistical test of whether differences seen in model performance are significant.

5. Significance and Use

5.1 Guidance is provided on designing model evaluation performance procedures and on the difficulties that arise in statistical evaluation of model performance caused by the stochastic nature of dispersion in the atmosphere. It is recognized there are examples in the literature where, knowingly or unknowingly, models were evaluated on their ability to describe something which they were never intended to characterize. This guide is attempting to heighten awareness, and thereby, to reduce the number of “unknowing” comparisons. A goal of this guide is to stimulate development and testing of evaluation procedures that accommodate the effects of natural variability. A technique is illustrated to provide information from which subsequent evaluation and standardization can be derived.

³ The boldface numbers in parentheses refer to the list of references at the end of this standard.