



Designation: D 5490 – 93<sup>ε1</sup>

## Standard Guide for Comparing Ground-Water Flow Model Simulations to Site-Specific Information<sup>1</sup>

This standard is issued under the fixed designation D 5490; 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 ( $\epsilon$ ) indicates an editorial change since the last revision or reapproval.

<sup>ε1</sup> NOTE—Paragraph 1.9 was added editorially October 1998.

### 1. Scope

1.1 This guide covers techniques that should be used to compare the results of ground-water flow model simulations to measured field data as a part of the process of calibrating a ground-water model. This comparison produces quantitative and qualitative measures of the degree of correspondence between the simulation and site-specific information related to the physical hydrogeologic system.

1.2 During the process of calibration of a ground-water flow model, each simulation is compared to site-specific information such as measured water levels or flow rates. The degree of correspondence between the simulation and the physical hydrogeologic system can then be compared to that for previous simulations to ascertain the success of previous calibration efforts and to identify potentially beneficial directions for further calibration efforts.

1.3 By necessity, all knowledge of a site is derived from observations. This guide does not address the adequacy of any set of observations for characterizing a site.

1.4 This guide does not establish criteria for successful calibration, nor does it describe techniques for establishing such criteria, nor does it describe techniques for achieving successful calibration.

1.5 This guide is written for comparing the results of numerical ground-water flow models with observed site-specific information. However, these techniques could be applied to other types of ground-water related models, such as analytical models, multiphase flow models, noncontinuum (karst or fracture flow) models, or mass transport models.

1.6 This guide is one of a series of guides on ground-water modeling codes (software) and their applications. Other standards have been prepared on environmental modeling, such as Practice E 978.

1.7 The values stated in SI units are to be regarded as the standard.

1.8 *This standard does not purport to address all of the safety problems, if any, associated with its use. It is the*

*responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

1.9 *This guide offers an organized collection of information or a series of options and does not recommend a specific course of action. This document 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 ASTM standard 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 this document be applied without consideration of a project's many unique aspects. The word "Standard" in the title of this document means only that the document has been approved through the ASTM consensus process.*

### 2. Referenced Documents

#### 2.1 ASTM Standards:

D 653 Terminology Relating to Soil, Rock, and Contained Fluids<sup>2</sup>

E 978 Practice for Evaluating Environmental Fate Models of Chemicals<sup>3</sup>

### 3. Terminology

#### 3.1 Definitions:

3.1.1 *application verification*—using the set of parameter values and boundary conditions from a calibrated model to approximate acceptably a second set of field data measured under similar hydrologic conditions.

3.1.1.1 *Discussion*—Application verification is to be distinguished from code verification which refers to software testing, comparison with analytical solutions, and comparison with other similar codes to demonstrate that the code represents its mathematical foundation.

3.1.2 *calibration*—the process of refining the model representation of the hydrogeologic framework, hydraulic properties, and boundary conditions to achieve a desired degree of correspondence between the model simulations and observations of the ground-water flow system.

<sup>1</sup> This guide is under the jurisdiction of ASTM Committee D-18 on Soil and Rock and is the direct responsibility of Subcommittee D18.21 on Ground Water and Vadose Zone Investigations.

Current edition approved Nov. 15, 1993. Published January 1994.

<sup>2</sup> Annual Book of ASTM Standards, Vol 04.08.

<sup>3</sup> Annual Book of ASTM Standards, Vol 11.04.

3.1.3 *censored data*—knowledge that the value of a variable in the physical hydrogeologic system is less than or greater than a certain value, without knowing the exact value.

3.1.3.1 *Discussion*—For example, if a well is dry, then the potentiometric head at that place and time must be less than the elevation of the screened interval of the well although its specific value is unknown.

3.1.4 *conceptual model*—an interpretation or working description of the characteristics and dynamics of the physical system.

3.1.5 *ground-water flow model*—an application of a mathematical model to represent a ground-water flow system.

3.1.6 *hydrologic condition*—a set of ground-water inflows or outflows, boundary conditions, and hydraulic properties that cause potentiometric heads to adopt a distinct pattern.

3.1.7 *residual*—the difference between the computed and observed values of a variable at a specific time and location.

3.1.8 *simulation*—in ground-water flow modeling, one complete execution of a ground-water modeling computer program, including input and output.

3.1.8.1 *Discussion*—For the purposes of this guide, a simulation refers to an individual modeling run. However, simulation is sometimes also used broadly to refer to the process of modeling in general.

3.2 For definitions of other terms used in this guide, see Terminology D 653.

#### 4. Summary of Guide

4.1 Quantitative and qualitative comparisons are both essential. Both should be used to evaluate the degree of correspondence between a ground-water flow model simulation and site-specific information.

4.2 Quantitative techniques for comparing a simulation with site-specific information include:

4.2.1 Calculation of residuals between simulated and measured potentiometric heads and calculation of statistics regarding the residuals. Censored data resulting from detection of dry or flowing observation wells, reflecting information that the head is less than or greater than a certain value without knowing the exact value, should also be used.

4.2.2 Detection of correlations among residuals. Spatial and temporal correlations among residuals should be investigated. Correlations between residuals and potentiometric heads can be detected using a scattergram.

4.2.3 Calculation of flow-related residuals. Model results should be compared to flow data, such as water budgets, surface water flow rates, flowing well discharges, vertical gradients, and contaminant plume trajectories.

4.3 Qualitative considerations for comparing a simulation with site-specific information include:

4.3.1 Comparison of general flow features. Simulations should reproduce qualitative features in the pattern of ground-water contours, including ground-water flow directions, mounds or depressions (closed contours), or indications of surface water discharge or recharge (cusps in the contours).

4.3.2 Assessment of the number of distinct hydrologic conditions to which the model has been successfully calibrated. It is usually better to calibrate to multiple scenarios, if the scenarios are truly distinct.

4.3.3 Assessment of the reasonableness or justifiability of the input aquifer hydrologic properties given the aquifer materials which are being modeled. Modeled aquifer hydrologic properties should fall within realistic ranges for the physical hydrogeologic system, as defined during conceptual model development.

#### 5. Significance and Use

5.1 During the process of calibration of a ground-water flow model, each simulation is compared to site-specific information to ascertain the success of previous calibration efforts and to identify potentially beneficial directions for further calibration efforts. Procedures described herein provide guidance for making comparisons between ground-water flow model simulations and measured field data.

5.2 This guide is not meant to be an inflexible description of techniques comparing simulations with measured data; other techniques may be applied as appropriate and, after due consideration, some of the techniques herein may be omitted, altered, or enhanced.

#### 6. Quantitative Techniques

6.1 Quantitative techniques for comparing simulations to site-specific information include calculating potentiometric head residuals, assessing correlation among head residuals, and calculating flow residuals.

6.1.1 *Potentiometric Head Residuals*—Calculate the residuals (differences) between the computed heads and the measured heads:

$$r_i = h_i - H_i \quad (1)$$

where:

$r_i$  = the residual,

$H_i$  = the measured head at point  $i$ ,

$h_i$  = the computed head at the approximate location where  $H_i$  was measured.

If the residual is positive, then the computed head was too high; if negative, the computed head was too low. Residuals cannot be calculated from censored data.

NOTE 1—For drawdown models, residuals can be calculated from computed and measured drawdowns rather than heads.

NOTE 2—Comparisons should be made between point potentiometric heads rather than ground-water contours, because contours are the result of interpretation of data points and are not considered basic data in and of themselves.<sup>4</sup> Instead, the ground-water contours are considered to reflect features of the conceptual model of the site. The ground-water flow model should be true to the essential features of the conceptual model and not to their representation.

NOTE 3—It is desirable to set up the model so that it calculates heads at the times and locations where they were measured, but this is not always possible or practical. In cases where the location of a monitoring well does not correspond exactly to one of the nodes where heads are computed in the simulation, the residual may be adjusted (for example, computed heads may be interpolated, extrapolated, scaled, or otherwise transformed) for use in calculating statistics. Adjustments may also be necessary when the times of measurements do not correspond exactly with the times when heads are calculated in transient simulations; when many observed heads

<sup>4</sup> Cooley, R. L., and Naff, R. L., "Regression Modeling of Ground-Water Flow," *USGS Techniques of Water Resources Investigations*, Book 3, Chapter B4, 1990.

are clustered near a single node; where the hydraulic gradient changes significantly from node to node; or when observed head data is affected by tidal fluctuations or proximity to a specified head boundary.

**6.1.2 Residual Statistics**—Calculate the maximum and minimum residuals, a residual mean, and a second-order statistic, as described in the following sections.

**6.1.2.1 Maximum and Minimum Residuals**—The maximum residual is the residual that is closest to positive infinity. The minimum residual is the residual closest to negative infinity. Of two simulations, the one with the maximum and minimum residuals closest to zero has a better degree of correspondence, with regard to this criterion.

**NOTE 4**—When multiple hydrologic conditions are being modeled as separate steady-state simulations, the maximum and minimum residual can be calculated for the residuals in each, or for all residuals in all scenarios, as appropriate. This note also applies to the residual mean (see 6.1.2.2) and second-order statistics of the residuals (see 6.1.2.4).

**6.1.2.2 Residual Mean**—Calculate the residual mean as the arithmetic mean of the residuals computed from a given simulation:

$$R = \frac{\sum_{i=1}^n r_i}{n} \quad (2)$$

where:

$R$  = the residual mean and  
 $n$  = the number of residuals.

Of two simulations, the one with the residual mean closest to zero has a better degree of correspondence, with regard to this criterion (assuming there is no correlation among residuals).

**6.1.2.3** If desired, the individual residuals can be weighted to account for differing degrees of confidence in the measured heads. In this case, the residual mean becomes the weighted residual mean:

$$R = \frac{\sum_{i=1}^n w_i r_i}{\sum_{i=1}^n w_i} \quad (3)$$

where  $w_i$  is the weighting factor for the residual at point  $i$ . The weighting factors can be based on the modeler's judgment or statistical measures of the variability in the water level measurements. A higher weighting factor should be used for a measurement with a high degree of confidence than for one with a low degree of confidence.

**NOTE 5**—It is possible that large positive and negative residuals could cancel, resulting in a small residual mean. For this reason, the residual mean should never be considered alone, but rather always in conjunction with the other quantitative and qualitative comparisons.

**6.1.2.4 Second-Order Statistics**—Second-order statistics give measures of the amount of spread of the residuals about the residual mean. The most common second-order statistic is the standard deviation of residuals:

$$s = \left\{ \frac{\sum_{i=1}^n (r_i - R)^2}{(n - 1)} \right\}^{\frac{1}{2}} \quad (4)$$

where  $s$  is the standard deviation of residuals. Smaller values of the standard deviation indicate better degrees of correspon-

dence than larger values.

**6.1.2.5** If weighting is used, calculate the weighted standard deviation:

$$s = \left\{ \frac{\sum_{i=1}^n w_i (r_i - R)^2}{(n - 1) \sum_{i=1}^n w_i} \right\}^{\frac{1}{2}} \quad (5)$$

**NOTE 6**—Other norms of the residuals are less common but may be revealing in certain cases.<sup>5,6</sup> For example, the mean of the absolute values of the residuals can give information similar to that of the standard deviation of residuals.

**NOTE 7**—In calculating the standard deviation of residuals, advanced statistical techniques incorporating information from censored data could be used. However, the effort would usually not be justified because the standard deviation of residuals is only one of many indicators involved in comparing a simulation with measured data, and such a refinement in one indicator is unlikely to alter the overall assessment of the degree of correspondence.

**6.1.3 Correlation Among Residuals**—Spatial or temporal correlation among residuals can indicate systematic trends or bias in the model. Correlations among residuals can be identified through listings, scattergrams, and spatial or temporal plots. Of two simulations, the one with less correlation among residuals has a better degree of correspondence, with regard to this criterion.

**6.1.3.1 Listings**—List residuals by well or piezometer, including the measured and computed values to detect spatial or temporal trends. Figures X1.1 and X1.2 present example listings of residuals.

**6.1.3.2 Scattergram**—Use a scattergram of computed versus measured heads to detect trends in deviations. The scattergram is produced with measured heads on the abscissa (horizontal axis) and computed heads on the ordinate (vertical axis). One point is plotted on this graph for each pair. If the points line up along a line with zero intercept and 45° angle, then there has been a perfect match. Usually, there will be some scatter about this line, hence the name of the plot. A simulation with a small degree of scatter about this line has a better correspondence with the physical hydrogeologic system than a simulation with a large degree of scatter. In addition, plotted points in any area of the scattergram should not all be grouped above or below the line. Figures X1.3 and X1.4 show sample scattergrams.

**6.1.3.3 Spatial Correlation**—Plot residuals in plan or section to identify spatial trends in residuals. In this plot, the residuals, including their sign, are plotted on a site map or cross section. If possible or appropriate, the residuals can also be contoured. Apparent trends or spatial correlations in the residuals may indicate a need to refine aquifer parameters or boundary conditions, or even to reevaluate the conceptual model (for example, add spatial dimensions or physical processes). For example, if all of the residuals in the vicinity of a no-flow boundary are positive, then the recharge may need to

<sup>5</sup> Ghassemi, F., Jakeman, A. J., and Thomas, G. A., "Ground-Water Modeling for Salinity Management: An Australian Case Study," *Ground Water*, Vol 27, No. 3, 1989, pp. 384–392.

<sup>6</sup> Konikow, L. F., *Calibration of Ground-Water Models, Proceedings of the Specialty Conference on Verification of Mathematical and Physical Models in Hydraulic Engineering*, ASCE, College Park, MD, Aug. 9–11, 1978, pp. 87–93.

be reduced or the hydraulic conductivity increased. Figure X1.5 presents an example of a contour plot of residuals in plan view. Figure X1.6 presents an example of a plot of residuals in cross section.

6.1.3.4 *Temporal Correlation*—For transient simulations, plot residuals at a single point versus time to identify temporal trends. Temporal correlations in residuals can indicate the need to refine input aquifer storage properties or initial conditions. Figure X1.7 presents a typical plot of residuals versus time.

6.1.4 *Flow-Related Residuals*—Often, information relating to ground-water velocities is available for a site. Examples include water budgets, surface water flow rates, flowing well discharges, vertical gradients, and contaminant plume trajectories (ground-water flow paths). All such quantities are dependent on the hydraulic gradient (the spatial derivative of the potentiometric head). Therefore, they relate to the overall structure of the pattern of potentiometric heads and provide information not available from point head measurements. For each such datum available, calculate the residual between its computed and measured values. If possible and appropriate, calculate statistics on these residuals and assess their correlations, in the manner described in 5.1 and 5.2 for potentiometric head residuals.

6.1.4.1 *Water Budgets and Mass Balance*—For elements of the water budget for a site which are calculated (as opposed to specified in the model input) (for example, base flow to a stream), compare the computed and the measured (or estimated) values. In addition, check the computed mass balance for the simulation by comparing the sum of all inflows to the sum of all outflows and changes in storage. Differences of more than a few percent in the mass balance indicate possible numerical problems and may invalidate simulation results.

6.1.4.2 *Vertical Gradients*—In some models, it may be more important to accurately represent the difference in heads above and below a confining layer, rather than to reproduce the heads themselves. In such a case, it may be acceptable to tolerate a correlation between the head residuals above and below the layer if the residual in the vertical gradient is minimized.

6.1.4.3 *Ground-Water Flow Paths*—In some models, it may be more important to reproduce the pattern of streamlines in the ground-water flow system rather than to reproduce the heads themselves (for example, when a flow model is to be used for input of velocities into a contaminant transport model). In this case, as with the case of vertical gradients in 6.1.4.2 it may be acceptable to tolerate some correlation in head residuals if the ground-water velocity (magnitude and direction) residuals are minimized.

## 7. Qualitative Considerations

7.1 *General Flow Features*—One criterion for evaluating the degree of correspondence between a ground-water flow model simulation and the physical hydrogeologic system is whether or not essential qualitative features of the potentiometric surface are reflected in the model. The overall pattern of flow directions and temporal variations in the model should correspond with those at the site. For example:

7.1.1 If there is a mound or depression in the potentiometric surface at the site, then the modeled contours should also

indicate a mound or depression in approximately the same area.

7.1.2 If measured heads indicate or imply cusps in the ground-water contours at a stream, then these features should also appear in contours of modeled heads.

7.2 *Hydrologic Conditions*—Identify the different hydrologic conditions that are represented by the available data sets. Choose one data set from each hydrologic condition to use for calibration. Use the remaining sets for verification.

7.2.1 *Uniqueness (Distinct Hydrologic Conditions)*—The number of distinct hydrologic conditions that a given set of input aquifer hydrologic properties is capable of representing is an important qualitative measure of the performance of a model. It is usually better to calibrate to multiple conditions, if the conditions are truly distinct. Different hydrologic conditions include, but are not limited to, high and low recharge; conditions before and after pumping or installation of a cutoff wall or cap; and high and low tides, flood stages for adjoining surface waters, or installation of drains. By matching different hydrologic conditions, the uniqueness problem is addressed, because one set of heads can be matched with the proper ratio of ground-water flow rates to hydraulic conductivities; whereas, when the flow rates are changed, representing a different condition, the range of acceptable hydraulic conductivities becomes much more limited.

7.2.2 *Verification (Similar Hydrologic Conditions)*—When piezometric head data are available for two times of similar hydrologic conditions, only one of those conditions should be included in the calibration data sets because they are not distinct. However, the other data set can be used for model verification. In the verification process, the modeled piezometric heads representing the hydrologic condition in question are compared, not to the calibration data set, but to the verification data set. The resulting degree of correspondence can be taken as an indicator or heuristic measure of the ability of the model to represent new hydrologic conditions within the range of those to which the model was calibrated.

NOTE 8—When only one data set is available, it is inadvisable to artificially split it into separate “calibration” and “verification” data sets. It is usually more important to calibrate to piezometric head data spanning as much of the modeled domain as possible.

NOTE 9—Some researchers maintain that the word “verification” implies a higher degree of confidence than is warranted.<sup>7</sup> Used here, the verification process only provides a method for estimating confidence intervals on model predictions.

7.3 *Input Aquifer Hydraulic Properties*—A good correspondence between a ground-water flow model simulation and site-specific information, in terms of quantitative measures, may sometimes be achieved using unrealistic aquifer hydraulic properties. This is one reason why emphasis is placed on the ability to reproduce multiple distinct hydrologic stress scenarios. Thus, a qualitative check on the degree of correspondence between a simulation and the physical hydrogeologic system should include an assessment of the likely ranges of hydraulic properties for the physical hydrogeologic system at

<sup>7</sup> Konikow, L. F., and Bredehoeft, J. D., “Ground-Water Models Cannot Be Validated,” *Adv. Wat. Res.* Vol 15, 1992, pp. 75–83.