



# Standard Practice for Estimation of Measurement Uncertainty for Data from Regression-based Methods<sup>1</sup>

This standard is issued under the fixed designation D7366; 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.

## 1. Scope

1.1 This practice establishes a standard for computing the measurement uncertainty for applicable test methods in Committee D19 on Water. The practice does not provide a single-point estimate for the entire working range, but rather relates the uncertainty to concentration. The statistical technique of regression is employed during data analysis.

1.2 Applicable test methods are those whose results come from regression-based methods and whose data are intra-laboratory (not inter-laboratory data, such as result from round-robin studies). For each analysis conducted using such a method, it is assumed that a fixed, reproducible amount of sample is introduced.

1.3 Calculation of the measurement uncertainty involves the analysis of data collected to help characterize the analytical method over an appropriate concentration range. Example sources of data include: 1) calibration studies (which may or may not be conducted in pure solvent), 2) recovery studies (which typically are conducted in matrix and include all sample-preparation steps), and 3) collections of data obtained as part of the method's ongoing Quality Control program. Use of multiple instruments, multiple operators, or both, and field-sampling protocols may or may not be reflected in the data.

1.4 In any designed study whose data are to be used to calculate method uncertainty, the user should think carefully about what the study is trying to accomplish and much variation should be incorporated into the study. General guidance on designing studies (for example, calibration, recovery) is given in Appendix A. Detailed guidelines on sources of variation are outside the scope of this practice, but general points to consider are included in Appendix B, which is not intended to be exhaustive. With any study, the user must think carefully about the factors involved with conducting the analysis, and must realize that the computed measurement uncertainty will reflect the quality of the input data.

<sup>1</sup> This practice is under the jurisdiction of ASTM Committee D19 on Water and is the direct responsibility of Subcommittee D19.02 on Quality Systems, Specification, and Statistics.

Current edition approved Jan. 1, 2013. Published January 2013. DOI: 10.1520/D7366-08R13.

1.5 Associated with the measurement uncertainty is a user-chosen level of statistical confidence.

1.6 At any concentration in the working range, the measurement uncertainty is plus-or-minus the half-width of the prediction interval associated with the regression line.

1.7 It is assumed that the user has access to a statistical software package for performing regression. A statistician should be consulted if assistance is needed in selecting such a program.

1.8 A statistician also should be consulted if data transformations are being considered.

1.9 *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 and health practices and determine the applicability of regulatory limitations prior to use.*

## 2. Referenced Documents

- 2.1 *ASTM Standards*:<sup>2</sup>  
D1129 [Terminology Relating to Water](#)

## 3. Terminology

### 3.1 *Definitions of Terms Specific to This Standard*:

3.1.1 *confidence level*—the probability that the prediction interval from a regression estimate will encompass the true value of the amount or concentration of the analyte in a subsequent measurement. Typical choices for the confidence level are 99 % and 95 %.

3.1.2 *fitting technique*—a method for estimating the parameters of a mathematical model. For example, ordinary least squares is a fitting technique that may be used to estimate the parameters  $a_0, a_1, a_2, \dots$  of the polynomial model  $y = a_0 + a_1x + a_2x^2 + \dots$ , based on observed  $\{x,y\}$  pairs. Weighted least squares is also a fitting technique.

3.1.3 *lack-of-fit (LOF) test*—a statistical technique when replicate data are available; computes the significance of

<sup>2</sup> For referenced ASTM standards, visit the ASTM website, [www.astm.org](http://www.astm.org), or contact ASTM Customer Service at [service@astm.org](mailto:service@astm.org). For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

residual means to replicate  $y$  variability, to indicate whether deviations from model predictions are reasonably accounted for by random variability, thus indicating that the model is adequate; at each concentration, compares the amount of residual variation from model prediction with the amount of residual variation from the observed mean.

3.1.4 *least squares*—fitting technique that minimizes the sum of squared residuals between observed  $y$  values and those predicted by the model.

3.1.5 *model*—mathematical expression (for example, straight line, quadratic) relating  $y$  (directly measured value) to  $x$  (concentration or amount of analyte).

3.1.6 *ordinary least squares (OLS)*—least squares, where all data points are given equal weight.

3.1.7 *prediction interval*—a pair of prediction limits (an “upper” and “lower”) used to bracket the “next” observation at a certain level of confidence.

3.1.8 *p-value*—the statistical significance of a test; the probability value associated with a statistical test, representing the likelihood that a test statistic would assume or exceed a certain value purely by chance, assuming the null hypothesis is true (a low  $p$ -value indicates statistical significance at a level of confidence equal to 1.0 minus the  $p$ -value).

3.1.9 *regression*—an analysis technique for fitting a model to data; often used as a synonym for OLS.

3.1.10 *residual*—error in the fit between observed and modeled concentration; response minus fit.

3.1.11 *root mean square error (RMSE)*—an estimate of the measurement standard deviation (that is, inherent variation in the measurement system).

3.1.12 *significance level*—the likelihood that a measured or observed result came about due to simple random behavior.

3.1.13 *uncertainty (of a measurement)*—the lack of exactness in measurement (for example, due to sampling error, measurement variation, and model inexactness); a statistical interval within which the measurement error is believed to occur, at some level of confidence.

3.1.14 *weight*—coefficient assigned to observations in order to manipulate their relative influence in subsequent calculations. For example, in weighted least squares, noisy observations are weighted downwards, while precise data are weighted upwards.

3.1.15 *weighted least squares (WLS)*—least squares, where data points are weighted inversely proportional to their variance (“noisiness”).

## 4. Summary of Practice

4.1 Key points of the statistical protocol for measurement uncertainty are:

4.1.1 Within the working range of the method’s data set, the estimate of the method uncertainty at any given concentration is calculated to be plus-or-minus the half-width of the prediction interval.

4.1.2 The total number of data points in any designed study should be kept high. Blanks may or may not be included, depending on the data-quality objectives of the test method.

4.1.3 In applying regression to any applicable data set, the proper fitting technique (for example, ordinary least squares (OLS) or weighted least squares (WLS)) must be determined (for fitting the proposed model to the data).

4.1.4 The residual pattern and the lack-of-fit test are used to evaluate the adequacy of the chosen model.

4.1.5 The magnitude of the half-width of the prediction interval must be evaluated, remembering that accepting or rejecting the amount of uncertainty is a judgment call, not a statistical decision.

## 5. Significance and Use

5.1 Appropriate application of this practice should result in an estimate of the test-method’s uncertainty (at any concentration within the working range), which can be compared with data-quality objectives to see if the uncertainty is acceptable.

5.2 With data sets that compare recovered concentration with true concentration, the resulting regression plot allows the correction of the recovery data to true values. Reporting of such corrections is at the discretion of the user.

5.3 This practice should be used to estimate the measurement uncertainty for any application of a test method where measurement uncertainty is important to data use.

## 6. Procedure

### 6.1 Introduction

6.1.1 For purposes of this practice, only regression-based methods are applicable. An example of a module that is not regression-based is a balance. If an object is placed on a balance, the readout is in the desired units; that is, in units of mass. No user intervention is required to get to the needed result. However, for an instrument such as a chromatograph or a spectrometer, the raw data (for example, peak area or absorbance) must be transformed into meaningful units, typically concentration. Regression is at the core of this transformation process.

6.1.2 One additional distinction will be made regarding the applicability of this protocol. This practice will deal only with intralaboratory data. In other words, the variability introduced by collecting results from more than one lab is not being considered. The examples that are shown here are for one method with one operator. If the user wishes, additional operators may be included in the design, to capture multiple-operator variability.

6.1.3 A brief example will help illustrate the importance of estimating measurement uncertainty. A sample is to be analyzed to determine if it is under the upper specification limit of 5 (the actual units of concentration do not matter). The final test result is 4.5. The question then is whether the sample should pass or fail. Clearly, 4.5 is less than 5. If the numbers are treated as being absolute, then the sample will pass. However, such a judgment call ignores the variability that always exists with a measurement. The width of any measurement’s uncertainty interval depends not only on the noisiness of the data, but also on the confidence level the user wishes to assume. This

latter consideration is not a statistical decision, but a reasoned decision that must be based on the needs of the customer, the intended use of the data, or both. Once the confidence level has been chosen, the interval can be calculated from the data. In this example, if the uncertainty is determined to be  $\pm 1.0$ , then there is serious doubt as to whether the sample passes or not, since the true value could be anywhere between 3.5 and 5.5. On the other hand, if the uncertainty is only  $\pm 0.1$ , then the sample could be passed with a high level of comfort. Only by making a sound evaluation of the uncertainty can the user determine how to apply the sample estimate he or she has obtained. The following protocol is designed to answer questions such as:  $4.5 \pm ?$

6.2 Regression Diagnostics for Recovery Data

6.2.1 Analysts who routinely use chromatographs and spectrometers are familiar with the basics of the regression process. The final results are: 1) a plot that visually relates the responses (on the y-axis) to the true concentrations (on the x-axis) and 2) an equation that mathematically relates the two variables.

6.2.2 Underlying these results are two basic choices: 1) a model, such as a straight line or some sort of curved line, and 2) a fitting technique, which is a version of least squares. The modeling choices are generally well known to most analysts, but the fitting-technique choices are typically less well understood. The two most common forms of least-squares fitting are discussed next.

6.2.2.1 Ordinary least squares (OLS) assumes that the variance of the responses does not trend with concentration. If the variance does trend with concentration, then weighted least squares (WLS) is needed. In WLS, data are weighted according to how noisy they are. Values that have relatively low uncertainty are considered to be more reliable and are subsequently afforded higher weights (and therefore more influence on the regression line) than are the more uncertain values.

6.2.2.2 Several formulas have been used for calculating the weights. The simplest is  $1/x$  (where  $x$  = true concentration), followed by  $1/x^2$ . At each true concentration, the reciprocal square of the actual standard deviation has also been used.

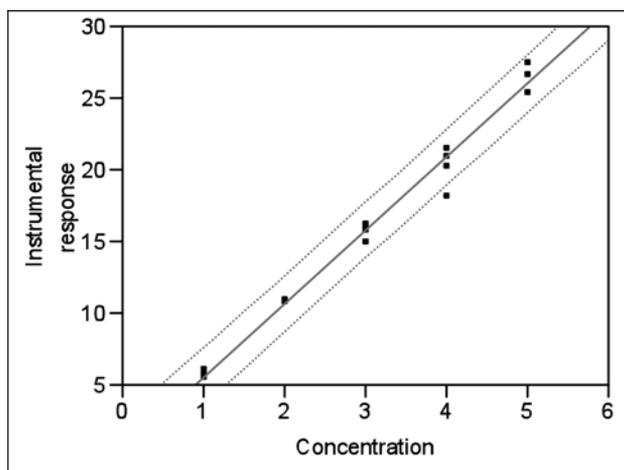
However, the preferred formula comes from modeling the standard deviation. In other words, the actual standard-deviation values are plotted versus true concentration; an appropriate model is then fitted to the data. The reciprocal square of the equation for the line is then used to calculate the weights. The simplest model is a straight line, but more precise modeling should be done if the situation requires it. (In practice, it is best to normalize the weight formula by dividing by the sum of all the reciprocal squares. This process assures that the root mean square error is correct.)

6.2.2.3 In sum, two choices, which are independent of each other, must be made in performing regression. These two choices are a model and a fitting technique. In practice, the options for the model are typically a straight line or a quadratic, while the customary choices for the fitting technique are ordinary least squares and weighted least squares.

6.2.2.4 However, a straight line is not automatically associated with OLS, nor is a quadratic automatically paired with WLS. The fitting technique depends solely on the behavior of the response standard deviations (that is, do they trend with concentrations). The model choice is not related to these standard deviations, but depends primarily on whether the data points exhibit some type of curvature.

6.2.3 Once an appropriate model and fitting technique have been chosen, the regression line and plot can be determined. One other very important feature can also be calculated and graphed. That feature is the prediction interval, which is an “envelope” around the line itself and which reports the uncertainty (at the chosen confidence level) in a future measurement predicted from the line. An example is given in Fig. 1. The solid red line is the regression line; the dashed red lines form the prediction interval.

6.2.4 While the concept of a model is familiar to most analysts, the statistically sound process for selecting an adequate model typically is not. A series of regression diagnostics will guide the user. The basic steps are as follows, and can be carried out with most statistical software packages that are commercially available:



NOTE 1—The interval in the above plot is nearly parallel to the regression line. This geometry will typically occur when OLS is the appropriate fitting technique and when the number of data points is high. However, if WLS is needed, the interval will flare. This WLS phenomenon makes sense, since the uncertainty in relatively noisy data will be larger than will the uncertainty in “tight” data.

FIG. 1 Example of a Regression Line with its Associated Prediction Interval

- (1) Plot y vs. x
- (2) Determine the behavior of y's standard deviation
- (3) Fit proposed model
- (4) Examine residuals
- (5) Conduct lack-of-fit (LOF) test
- (6) Evaluate prediction interval

Step 1 generates a scatterplot. This graph is helpful for spotting potential errant data points (which may simply be due to typographical errors in the data table), as well as for getting a general sense about the behavior of the response standard deviation and any curvature in the data. Step 2 will show which fitting technique (that is, OLS or WLS) is needed. Steps 3 through 5 allow for the selection of an adequate model. Step 6 provides the information needed to decide if the uncertainty in the measurements is at an acceptable level.

6.2.5 These steps can best be illustrated with an example, which will show how an appropriate model and fitting technique are found for simulated recovery data, using the diagnostic steps outlined above. (Although this example is for recovery data, it must be emphasized that the illustrated techniques are generic and can be used with data from applicable test methods as described in the Scope.) **Table 1** contains the simulated data for this example. The associated scatterplot is shown in **Fig. 2**.

6.2.6 To determine the behavior of the standard deviation of the responses, a plot of the standard deviations versus concentration is constructed (see **Fig. 3**). A straight line is fitted using ordinary least squares. The p-value for the slope of the line is 0.0045, which is significant. Thus, weighted least squares is needed to fit any model to the recovery data themselves. The formula for the weights is the reciprocal square of the line's expression of  $[-0.317326 + (0.5206949 \times x)]$ , divided by the mean of all such reciprocal squares.

6.2.7 The regression diagnostics reveal that a straight line is an adequate model. The final plot (that is, a straight line fitted with WLS), with the prediction interval at 95 % confidence, is shown in **Fig. 4**.

**TABLE 1 Simulated Recovery Data**

True or spiked concentration	Recovered concentration	Weight
1	5.76	4.4375
1	6.29	4.4375
1	5.58	4.4375
1	5.54	4.4375
2	12.24	0.3501
2	10.64	0.3501
2	11.94	0.3501
2	11.44	0.3501
3	14.62	0.1185
3	15.82	0.1185
3	17.08	0.1185
3	14.89	0.1185
4	21.48	0.0589
4	23.13	0.0589
4	20.10	0.0589
4	19.69	0.0589
5	28.60	0.0351
5	27.11	0.0351
5	24.31	0.0351
5	23.11	0.0351

6.2.8 Evidence for the adequacy of the model is indicated by the fact that the LOF p-value was 0.4358, which is insignificant (the starting hypothesis is that there is no lack of fit with the candidate model). The residual plot (see **Fig. 5**), with its nearly random scatter of points about the zero line, also supports the choice of a straight line. The trumpet shape of the pattern is characteristic of data where the response standard deviations trend with concentration.

6.2.9 Any concentration that is estimated from the recovery plot has an uncertainty of  $\pm$  the half-width of the prediction interval (at the chosen confidence level), thereby answering the question (that is,  $4.5 \pm ?$ ) posed in Section 6.1.3.

6.2.10 Results should be reported by stating: 1) the estimate of the value itself, 2) the uncertainty, and 3) the confidence level. An example is:  $4.5 \pm 0.2$  ppb, at 95 % confidence.

## 7. Keywords

7.1 measurement uncertainty; regression-based methods; calibration; prediction interval; confidence level

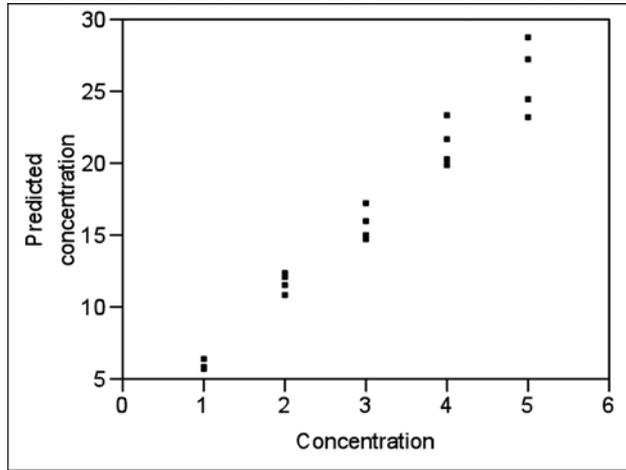


FIG. 2 Scatterplot of Simulated Recovery Data

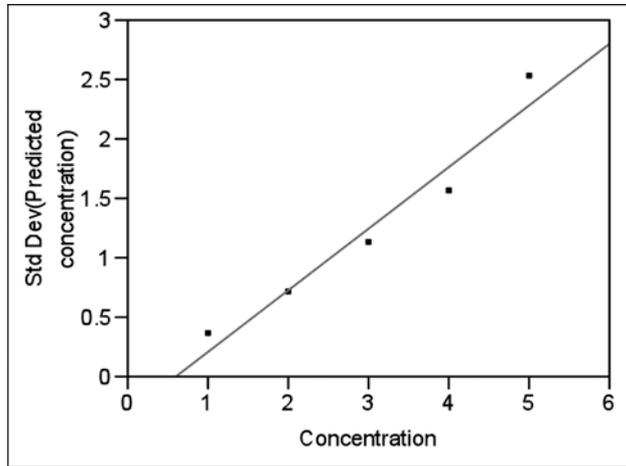


FIG. 3 Plot of Standard Deviation of Responses Versus Concentration

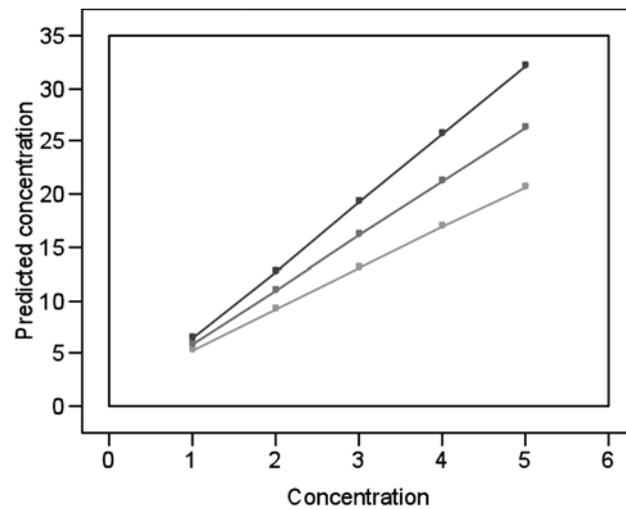


FIG. 4 Recovery Plot with its Associated Prediction Interval

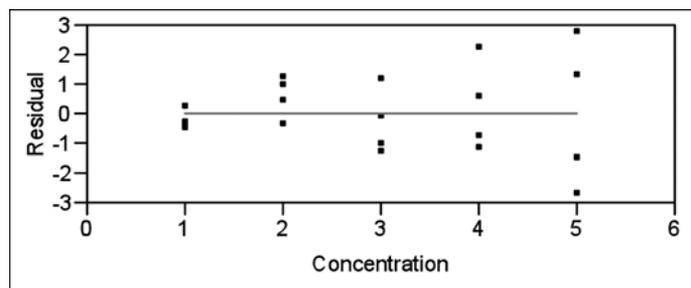


FIG. 5 Residuals Plot for the Straight-line Model Fitted to the Recovery Data, using WLS

## APPENDIXES

### (Nonmandatory Information)

#### X1. Guidance for Designing Studies for Regression-Based Test Methods

X1.1 With the study design, the ultimate goal is to decide what concentrations (or levels) will be included, and how many replicates of each solution will be analyzed. To make these decisions, several questions should be addressed. First, what is the concentration range of interest? Some prior knowledge is needed of the levels expected in the samples that eventually will have to be tested. This range should be wide enough to prevent having to extrapolate the calibration curve. Second, will the sensitivity of the method be challenged? Are reliable data necessary in the low-end region, meaning that sufficient levels and replicates are needed in this area? For work in this region, a well chosen blank is typically necessary. Third, will high precision be needed in at least some portions of the working range, indicating that an adequate number of replicates are required at each concentration? Fourth, are the data expected to exhibit curvature? If so, then an adequate number of concentrations should be assigned to the suspect portion of the range. Fifth, are there specification limits that are of concern? Such critical concentrations should be included in the

design and also should be bracketed tightly.

X1.2 Once the above questions (and any others that are of concern) have been answered, the actual concentration range, along with the number of concentrations and the number of replicates, can be selected. It is not mandatory that the same number of replicates be analyzed for each concentration. Also, the confidence level should be set, since that determination must be made before data can be analyzed properly. Finally, within each set of replicates, the set of concentrations should be randomized. This process allows for the determination of such phenomena as carryover.

X1.3 There is no “magic” design that works for all calibration studies. However, a good starting place is a 5×5 arrangement (that is, five replicates of each of five concentrations). The numbers can and should be adapted to fit the needs of the study (and, ultimately, the analytical method). It is good to keep in mind that having a high number of data points is desirable.

#### X2. Guidance on Sources of Variation

X2.1 In designing a calibration or recovery study, every effort should be made to capture as much variation as is reasonably expected to occur in the day-to-day use of a given test method.

X2.2 While the following paragraphs are not intended to be inclusive, several typical sources of variation are discussed. The user should use these ideas as a starting place for assessing “problem areas” with his or her method.

X2.2.1 *Analyst*—With some low-level methods (for example, trace levels of ammonia), the analyst himself/herself can be a source of contamination, which can vary from one day to the next.

X2.2.2 *Method*—Start-up and shut-down procedures can affect the stability of a method.

X2.2.3 *Environment and time-varying influences*— Factors such as temperature, power fluctuations, humidity, and air-borne contaminants may affect some procedures.

X2.2.4 *Chemicals*—Some reagents and standards may have a limited shelf life, especially at low concentrations.

X2.2.5 *Sample preparation*—This arena is perhaps the largest source of variation in many test methods.

X2.2.6 *Sample containers*—The cleanliness of all laboratory glassware/plasticware is of utmost importance in low-level analyses.

 **D7366 – 08 (2013)**

*ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.*

*This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.*

*This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or [service@astm.org](mailto:service@astm.org) (e-mail); or through the ASTM website ([www.astm.org](http://www.astm.org)). Permission rights to photocopy the standard may also be secured from the ASTM website ([www.astm.org/COPYRIGHT/](http://www.astm.org/COPYRIGHT/)).*