Procedures for Establishing Geotechnical Design Parameters from Two Data Sources

Prepared by
   J. Erik Loehr, Ph.D., P.E.
   Andrew Z. Boeckmann, P.E.
   Paul L. Speckman, Ph.D.
   Dan Ding
   University of Missouri Department of Civil and Environmental Engineering and
      Department of Statistics

MoDOT

PROCEDURES FOR ESTABLISHING GEOTECHNICAL DESIGN PARAMETERS FROM TWO DATA SOURCES

prepared for

Missouri Department of Transportation

by

J. Erik Loehr, Andrew Z. Boeckmann, Paul L. Speckman, and Dan Ding

University of Missouri
Department of Civil and Environmental Engineering

and

Department of Statistics

July 2013

TECHNICAL REPORT DOCUMENTATION PAGE.

1. Report No.: cmr 14-002	2. Government Accession No.	3. Recipient's Catalog No.:
4. Title and Subtitle: Procedures for Establishing Geotechnical Design Parameters from Two Data Sources		5. Report Date: July, 2013
		6. Performing Organization Code:
7. Author(s):Loehr, J.E., A.Z. Boeckmann, P.L. Speckman, and D. Ding		8. Performing Organization Report No.:
9. Performing Organization Name and Address: University of Missouri E2509 Lafferre Hall Columbia, MO 65211		10. Work Unit No.:
		11. Contract or Grant No.: TRyy1204
12. Sponsoring Agency Name and Address: Missouri Department of Transportation Research, Development and Technology PO BOX 270, JEFFERSON CITY MO 65102		13. Type of Report and Period Covered: Research Report
		14. Sponsoring Agency Code:

15. Supplementary Notes:
The investigation was conducted in cooperation with the U. S. Department of Transportation, Federal Highway Administration.

16. Abstract:
The Missouri Department of Transportation (MoDOT) recently adopted new provisions for geotechnical design that require that the mean value and the coefficient of variation (COV) for the mean value of design parameters be established in order to determine appropriate values for resistance factors to be used for design of specific projects. Current guidelines include provisions to establish appropriate values for the mean and COV of the mean for relevant design parameters based on lab or field measurements made for a specific project. This report describes procedures developed to establish geotechnical design parameters and their variability when there are two sources of data that can be used to estimate values for the design parameter. The proposed procedures include methods for establishing the mean and COV of the mean for a design parameter from "surrogate" measurements that can be related to the design parameter of interest as well as procedures for establishing the mean and COV for a design parameter from combined direct and surrogate measurements. The procedures are demonstrated through application to direct measurements of uniaxial compressive strength and surrogate Standard Penetration Test (SPT) measurements in Missouri shales. Results of analyses presented demonstrate that use of combined direct and surrogate measurements produces smaller values of COV when compared to use of direct measurements alone but that the magnitude of the reduction in COV depends on the number of direct and surrogate measurements available and on the variability and uncertainty present in the relation between the surrogate measurements and the design parameter of interest. Results of the analyses also show that SPT measurements are generally a poor predictor of uniaxial compressive strength and drilled shaft capacity in shale when compared to direct measurements of uniaxial compressive strength.

17. Key Words: Geotechnical Engineering, Site Characterization, Load and Resistance Factor Design (LRFD)	18. Distribution Statement: No restrictions. This document is available to the public through National Technical Information Center, Springfield, Virginia 22161.
19. Security Classification (of this report): Unclassified.	20. Security Classification (of this page): Unclassified.

19. Security Classification (of this report): Unclassified.	20. Security Classification (of this page): Unclassified.	21. No of Pages: 22 pages	22. Price:

Form DOT F 1700.7 (06/98).

Procedures for Establishing Geotechnical Design Parameters from Two Data Sources

1. Introduction

The Missouri Department of Transportation (MoDOT) recently adopted new provisions for geotechnical design that implement Load and Resistance Factor Design (LRFD) techniques (e.g. Loehr et al., 2011a; Loehr et al., 2011b, Loehr et al., 2011c). These new provisions generally require that the mean value and the coefficient of variation (COV) for the mean value of design parameters be established in order to determine appropriate values for resistance factors to be used for design of specific projects. Current guidelines include provisions to establish appropriate values for the mean and COV of the mean for relevant design parameters based on lab or field measurements made for a specific project. However, the existing provisions do not provide procedures where design parameters may be estimated from multiple data sources. Since the variability and uncertainty in a specific design parameter generally decreases with the number of available measurements, use of measurements from multiple sources will tend to reduce the variability and uncertainty in established design parameters. Reduced variability and uncertainty in the design parameters, in turn, often allows for use of greater resistance factors and, thus, more economical design even if the mean or nominal value for the parameter does not change. A research project was therefore undertaken to develop procedures for practically establishing design parameters (both mean values and the COV) from multiple data sources so that the benefits of having multiple types of measurements can be realized.

This report documents procedures developed to establish geotechnical design parameters and their variability when there are two sources of data that can be used to estimate values for the design parameter. Application of the procedures produces a combined estimate for the mean value of the design parameter and the coefficient of variation (COV) of the mean value to be used for design. The procedures are applicable for data from many different sources, but are primarily expected to be applied in cases where a design parameter has been characterized using "direct" measurements (e.g. laboratory measurement of uniaxial compressive strength, q_u) and using "indirect" or "surrogate" measurements that can be correlated to the design parameter (e.g. Standard Penetration Test measurements). While there are many potential applications for these procedures, the primary focus of the work has been on combining measurements of q_u from laboratory tests and measurements of N_{eq}-values from Standard Penetration Tests (SPT) in shales because the opportunity for applying these measurements for design of drilled shafts is commonly encountered in current MoDOT practice.

This report first provides a summary of important concepts related to variability and uncertainty in different types of measurements. The mathematical and statistical methods used to develop the procedures are then described. The proposed procedures are then demonstrated using results obtained from numerous SPT and q_u measurements taken from extensive site investigations performed at several research sites that include Missouri shales. Finally, conclusions drawn from development and evaluation of the proposed procedures and recommendations for implementation and for future work are provided.

2. Background and Concepts

It is common in geotechnical engineering practice to estimate values for design parameters from both direct and indirect, or "surrogate" measurements. Examples of common direct measurements used by MoDOT include:

- use of uniaxial compression tests to establish the uniaxial compressive strength (q_u) of rock,
- use of unconfined compression tests and/or triaxial compression tests to establish the undrained shear strength (s_u) of soils,
- use of direct shear tests and/or triaxial compression tests to establish effective stress shear strength parameters ($\bar{c}$ and $\tan \bar{\phi}$) for soils, and

- use of one-dimensional consolidation tests to establish the compression (c_c) and recompression (c_r) indices and pre-consolidation stress (σ'_p) for soils.

These measurements are effectively direct measurements of the respective design parameter(s) that do not require any correlation or manipulation to produce estimates of the design parameter. These direct measurements are subject to variability and uncertainty introduced due to variability and uncertainty in the samples acquired for testing and due to variability and uncertainty in the measurement techniques. As such, the variability and uncertainty in design parameters established from these direct measurements is controlled by these same sources of variability and uncertainty.

In contrast, indirect, or "surrogate" measurements are measurements that generally require some manipulation or correlation to produce an estimate of the actual design parameter to be used. Common surrogate measurements used by MoDOT include:

- use of the Standard Penetration Test N-values (or N_{eq}-values) to estimate the undrained shear strength and/or the effective stress friction angle for soils, and to estimate the uniaxial compressive strength for rock;
- use of the Cone Penetration Test tip resistance (q_t) to estimate the undrained shear strength and/or the effective stress frictional angle for soils;
- use of the Pocket Penetrometer or Torvane Tests to estimate the undrained shear strength for soils; and
- use of Atterberg Limits (LL and PL) to estimate the effective stress friction angle for soils.

The manipulations required to "convert" such surrogate measurements to actual design parameters often involve empirical or theoretical correlations. Like direct measurements, surrogate measurements are subject to variability and uncertainty due to variability and uncertainty in the measurements and variability and uncertainty in the soil/rock being tested. However, unlike direct measurements, the variability and uncertainty in design parameters established from surrogate measurements also includes variability and uncertainty attributed to the relation used to "convert" the surrogate measurement to a design parameter value. Thus, surrogate measurements are subject to an additional source of variability and uncertainty that is not present when making direct measurements. Some surrogate measurements correlate very closely with some design parameters, in which case the variability and uncertainty in such design parameters may be very close to those established from direct measurements. In other cases, surrogate measurements are not accurate or reliable predictors of some design parameters, in which case the variability and uncertainty in design parameters established from surrogate measurements may be substantially greater than those established from direct measurements.

The considerations just described suggest that surrogate measurements will often be inferior to direct measurements of design parameters since they are subject to an additional source of variability and uncertainty. However, this inaccurate perspective neglects the influence of the quantity of tests on variability and uncertainty in design parameters. Variability and uncertainty in geotechnical design parameters, as reflected by the COV of the mean value of the parameter, is highly dependent on the quantity of tests performed (Loehr et al., 2013). Thus, it is easily possible for the variability and uncertainty of a design parameter established from surrogate measurements to be less than the variability and uncertainty of the design parameter from direct measurements when greater numbers of surrogate measurements are available. This is, in fact, a common situation since most surrogate measurements can be made more quickly and at less cost than many direct measurements. The expedience and cost of surrogate tests are in fact the primary advantage of such tests. Whether direct measurements, surrogate measurements, or some combination of both direct and surrogate measurements are more likely to produce the least variability and uncertainty for a specific case depends predominantly on the number of direct and surrogate tests performed, the reliability of the correlation between the surrogate measurement and the design parameter of interest, and the specific design parameter being considered. The methods and procedures described in this report provide a sound basis upon which to judge the relative benefits of direct and surrogate tests, both on a site or project specific basis, as well as more generally.

3. Procedures for Establishing Mean Values and Coefficients of Variation for Design Parameters from Direct, Surrogate, and Combined Measurements

The procedures described in this section were developed considering two arbitrary sets of independent measurements: one including only "direct" measurements of a design parameter (e.g. laboratory measurements of soil strength) and the other including only "surrogate" measurements that can be related to the design parameter of interest (e.g SPT *N*-value as a surrogate for soil strength). Procedures for computing the mean and the variance of the mean for design parameters from direct measurements are first described. Methods for performing regression analyses to relate surrogate measurements to a specific design parameter are then described, followed by description of procedures for computing the mean and variance for a design parameter based exclusively on surrogate measurements. Finally, procedures for estimating the mean and variance for a design parameter from combined direct and surrogate measurements are described. The equations and methods presented in this section are based primarily on inferential statistics with the application of several simplifying approximations (e.g. Taylor Series). The methods presented also presume constant values for the design parameter within the stratum of interest.

In all cases, the variance described in this section refers to the variance of the mean values for the design parameters. Current MoDOT design provisions are generally based on use of the coefficient of variation of the mean value for a parameter. The coefficient of variation of the mean value is a normalized form of the variance of the mean value that can be computed as

$$COV = \frac{\sqrt{\sigma^2}}{\bar{y}}$$
(consistent units)   Eq. 1

where σ^2 is the variance of the mean value for a design parameter and $\bar{y}$ is the mean value for the parameter. The equations presented subsequently are presented in terms of variance for purposes of clarity. These equations can easily be converted to produce the coefficient of variation by direct substitution into Eq. 1.

3.1 Procedure for Computing Mean and Variance for Design Parameters from Direct Measurements

The mean value and variance of the mean value for a specific design parameter can be calculated using equations provided in current MoDOT design provisions. Using the notation adopted for this report, the mean value ($\bar{y}_d$) and variance of the mean value (σ_d^2) for the design parameter of interest are respectively computed from a collection of direct measurements in the stratum of interest as

$$\bar{y}_d = \frac{\sum_{i=1}^{n} \hat{y}_i}{n_d}$$
(consistent units)   Eq. 2

$$\sigma_d^2 = \frac{\sigma_{\hat{y}}^2}{n_d} = \frac{1}{n_d} \cdot \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_d)^2}{n_d - 1}$$
(consistent units)   Eq. 3

where $\hat{y}_i$ is a direct measurement of the parameter of interest, n_d is the number of direct measurements of the parameter of interest, and $\sigma_{\hat{y}}^2$ is the variance of the direct measurements. The subscript "d" is used to indicate that the design parameter is established from direct measurements.

3.2 Establishing Correlations Between Design Parameters and Surrogate Measurements Using Linear Regression

Establishing the mean and variance of design parameters from surrogate measurements is not as straightforward as for direct measurements. In general, the mean value is estimated directly from an established correlation between direct and surrogate measurements, and the variance is estimated by accounting for the variability inherent to the established correlation as well as the variability of the surrogate measurements themselves. Before describing these calculations, it is helpful to provide some background information on potential forms of the correlations and how they are established. Geotechnical practice and literature are replete with examples of correlations between direct and surrogate measurements. Many of the established correlations fall into the categories described below and therefore can be used for the procedures described in this report with appropriate estimates of parameters associated with the regression equation.

Correlations between direct and surrogate measurements can take many forms, but the procedures detailed in this report are limited to linear correlations of the form

$$y = \beta_0 + \beta_1 x \qquad \text{(consistent units)} \qquad \text{Eq. 4}$$

where $y = f(x)$ is the value for the design parameter of interest, x is the value of the surrogate measurement, β_0 is the y-intercept of the linear relation between surrogate measurements and the parameter of interest, and β_1 is the slope of the linear relation between surrogate measurements and the parameter of interest.

Values for β_0 and β_1 are commonly established by applying least squares regression techniques to a data set that includes direct measurements of a design parameter of interest and corresponding surrogate measurements. In this report, such data sets will be referred to as regression data sets with direct measurements being denoted as $\hat{y}_{r-i}$ and corresponding surrogate measurements being denoted as $\hat{x}_{r-i}$. Throughout this report, the "hat" accent is used to indicate measured data while the subscript "r" denotes that the data are part of the regression data set (to distinguish it from data used for a specific site). Regression data sets are commonly composed of measurements collected from many different sites that presumably reflect the entire population of possible conditions where the regression may be applied. Results from analyses on such data sets are generally applicable across the range of conditions that are reflected in the regression data set. In some cases, regression data sets may be composed of a more restrictive collection of data, such as from a specific site, in which case the results can be considered as "site specific". It is often true that site specific correlations between surrogate measurements and a design parameter will have less variability and uncertainty than correlations developed for more general application. However, development of site specific correlations is generally only practical for relatively large projects with ample direct and surrogate measurements.

Least squares regression capabilities are commonly automated in computer software, and can be performed using the LINEST function in Microsoft Excel© or similar functions in other computer programs. Such analyses seek to find the regression parameters, β_0 and β_1, that will minimize the sum of the squared residuals for the entire regression data set. When the residuals are left unweighted, the regression is generally referred to as Ordinary Least Squares (OLS) while the term Weighted Least Squares (WLS) is used to refer to analyses performed to minimize the sum of "weighted" residuals, as described subsequently in more detail. It is noteworthy that the OLS technique produces a constant conditional variance, which is appropriate when the scatter about the regression line is practically independent of the value of x, whereas WLS produces a conditional variance that varies with x.

In some cases, it can be beneficial to enforce the constraint that $\beta_0 = 0$ to force the regression to pass through the origin. If such a regression is based on the OLS regression technique, the slope of the regression line is computed as

$$\hat{\beta}_1 = \frac{\sum \hat{x}_{r-i}\hat{y}_{r-i}}{\sum \hat{x}_{r-i}^2} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 5}$$

where $\hat{y}_{r-i}$ is a direct measurement of the parameter of interest, $\hat{x}_{r-i}$ is a corresponding surrogate measurement, and $\hat{\beta}_1$ is the estimated value of β_1 determined from the regression. The regression parameter from Eq. 5 is identical to the result generated using the LINEST function in Microsoft Excel© when the intercept is forced to be zero.

In cases where the scatter of the data about the regression line is not independent of x (i.e. when the scatter tends to increase or decrease with the value of x), weighted least squares regression can be performed to produce a regression that better reflects the observed data. Weighted least squares regression will often improve the fit to experimental data, especially when the data points are far from the origin, as is often the case for geotechnical correlations. Weighted least squares analysis introduces a weighting factor, w_i, to the regression. For a regression passing through the origin, the slope of the regression line is calculated as

$$\hat{\beta}_1 = \frac{\sum w_i \hat{x}_{r-i}\hat{y}_{r-i}}{\sum w_i \hat{x}_{r-i}^2} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 6}$$

where

$$w_i \propto \frac{1}{\text{Var}(y_i)} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 7}$$

Various assumptions can be made for the weighting factor to produce a regression that best reflects the available measurements and the scatter of the measurements about the regression line. One commonly used weighting method is to assume that the variance of the calibration measurements ($\text{Var}(y_i)$) is proportional to the square of the expected value of y from the correlation, $\bar{y}^2 = (\beta_1 x_i)^2$. This choice of weights is equivalent to assuming that the regression produces a constant coefficient of variation across the entire range of the regression. If this assumption is made, then

$$w_i = \frac{1}{\hat{x}_{r-i}^2} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 8}$$

and

$$\hat{\beta}_1 = \frac{\sum \frac{1}{\hat{x}_{r-i}^2}\hat{x}_{r-i}\hat{y}_{r-i}}{\sum \frac{\hat{x}_{r-i}^2}{\hat{x}_{r-i}^2}} = \frac{\sum \frac{\hat{y}_{r-i}}{\hat{x}_{r-i}}}{m} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 9}$$

where m is the number of data points in the regression data set.

Values for β_0 and β_1 from the regression analyses are used to estimate the mean value for a design parameter for a given value of the surrogate measurement (x) via Eq. 1. To compute the variance of the mean value for the design parameter, additional results from the regression are also needed. These include the mean square error

$$s_y^2 = \frac{\sum(\hat{y}_{r-i} - \bar{y}_i)^2}{m-2} \qquad\qquad \text{(consistent units)} \qquad \text{Eq. 10}$$

and

$$s_{xx} = \sum(\hat{x}_{r-i} - \bar{x}_r)^2 \qquad \text{(consistent units)} \qquad \text{Eq. 11}$$

where s_y^2 is the mean square error from the regression, $\hat{y}_{r-i}$ are the direct measurements from the regression data set, $\bar{y}_i$ is the predicted value from the regression at the corresponding $\hat{x}_{r-i}$, m is the number of data points in the regression data set, s_{xx} is a quantity related to the variance of the surrogate measurements in the regression data set, $\hat{x}_{r-i}$ are the surrogate measurements from the regression data set, and $\bar{x}_r$ is the mean value of the surrogate measurements in the regression data set.

In cases where the relationship between a surrogate measurement and the parameter of interest is notably non-linear, least squares regression can often be applied to transformed values of the surrogate and/or parameter of interest (e.g. $\log x_i$, $\sqrt{y_i}$, etc.) that serve to "linearize" the relationship and result in a correlation that takes the form of Eq. 4. Equations for a log transformation in y are provided in Section 3.3 and an example application for these equations is provided in Section 4 of this report.

3.3   Procedure for Computing Mean and Variance for Design Parameters from Surrogate Measurements

Once regression analyses are completed to establish the relation between surrogate measurements and the design parameter of interest, the mean value of the design parameter of interest within a particular stratum can be computed from the mean value for independent surrogate measurements within the same stratum using the established regression as

$$\bar{y}_s = \beta_0 + \beta_1 \bar{x} \qquad \text{(consistent units)} \qquad \text{Eq. 12}$$

where $\bar{y}_s$ is the mean or design value for the design parameter of interest established from surrogate measurements, $\bar{x}$ is the mean value of the surrogate measurements within the stratum of interest, and β_0 and β_1 are, respectively, the intercept and slope of the regression relation that have been previously established for the particular design parameter and surrogate measurement. Note that Eq. 12 can be applied regardless of whether the term β_0 is forced to be zero and that the subscript "s" is used to indicate that the design parameter is established from surrogate measurements.

The variance of the mean value for the design parameter of interest for a particular stratum can be established from surrogate measurements in the same stratum by considering the variability and uncertainty in the established relation between the specific design parameter and surrogate measurements (determined as part of the regression analyses described previously) as well as the variability and uncertainty of the surrogate measurements themselves. This follows from a well-known theorem of probability that is sometimes referred to as the "law of total variance" (e.g. Bain and Engelhardt, 1992):

$$\sigma_s^2 = \mathrm{E}[\mathrm{Var}(y|x)] + \mathrm{Var}[\mathrm{E}(y|x)] \qquad \text{(consistent units)} \qquad \text{Eq. 13}$$

where σ_s^2 is the variance of the mean value for the design parameter of interest from surrogate measurements, $\mathrm{E}[\mathrm{Var}(y|x)]$ is the expected value of the conditional variance from the regression, and $\mathrm{Var}[\mathrm{E}(y|x)]$ is the variance of the expected value of the regression. The first term in Eq. 13 represents variability and uncertainty from the established empirical relation while the second term represents variability and uncertainty in the surrogate measurements for the specific stratum being considered. Note that the magnitude of σ_s^2 varies with the specific value of the surrogate measurement, x. For a linear OLS regression of the form shown in Eq. 4, and assuming that the "model" variance of the surrogate measurements in the particular stratum are appropriate, this can be approximated as

$$\sigma_s^2 \approx s_y^2 \left[1 + \frac{1}{m} + \frac{(\bar{x} - \bar{x}_r)^2 + \frac{\sigma_x^2}{n_s}}{s_{xx}} \right] + \beta_1^2 \frac{\sigma_x^2}{n_s}$$ (consistent units)   Eq. 14

where s_y^2 is the mean squared error from the regression analyses (Eq. 10), m is the number of data in the regression data set, $\bar{x}$ is the mean value of the surrogate measurements in the stratum of interest, $\bar{x}_r$ is the mean value of surrogate measurements in the regression data set, σ_x^2 is the variance of the surrogate measurements in the stratum of interest, n_s is the number of surrogate measurements in the stratum of interest, s_{xx} is computed from Eq. 11, and β_1 is the slope of the regression relation relating the surrogate measurement to the design parameter of interest. Note that all terms in Eq. 14 except for $\bar{x}$, σ_x^2, and n_s are produced from the regression analyses on the regression data set. The mean and variance for the surrogate measurements in the stratum of interest are computed as

$$\bar{x} = \frac{\sum_{i=1}^{n} \hat{x}_i}{n_s}$$ (consistent units)   Eq. 15

$$\sigma_x^2 = \sigma_{\hat{x}}^2 = \frac{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2}{n_s - 1}$$ (consistent units)   Eq. 16

where $\hat{x}_i$ are surrogate measurements in the stratum of interest.

For a linear WLS regression forced to pass through the origin, and again assuming that the "model" variance of the surrogate measurements in the particular stratum is appropriate, the variance of the mean value for the design parameter can be approximated as

$$\sigma_s^2 \approx \left(\bar{x}^2 + \frac{\sigma_x^2}{n_s} \right) \left(1 + \frac{1}{m-1} \frac{\sum w_i (\hat{y}_{r-i} - \hat{\beta}_1 \hat{x}_{r-i})^2}{\sum w_i \hat{x}_{r-i}^2} \right) + \beta_1^2 \frac{\sigma_x^2}{n_s}$$ (consistent units)   Eq. 17

For a logarithmic transformation of the direct measurements using OLS, where the regression data set includes the natural logarithm of the direct measurements rather than the actual direct measurements, the mean value and the variance of the mean of the design parameter of interest are calculated as

$$\bar{y}_s = e^{\beta_0 + \beta_1 \bar{x}}$$ (consistent units)   Eq. 18

$$\sigma_s^2 \approx e^{2(\beta_0 + \beta_1 \bar{x})} \cdot \left\{ s_y^2 \left[1 + \frac{1}{m} + \frac{(\bar{x} - \bar{x}_r)^2 + \frac{\sigma_x^2}{n_s}}{s_{xx}} \right] + \beta_1^2 \frac{\sigma_x^2}{n_s} \right\}$$ (consistent units)   Eq. 19

where the y-values used to calculate s_y^2, β_0, and β_1 are the logarithm of the direct measurements in the regression data set. Transformation of the surrogate measurements (x values) does not affect the form of Equations 12 through 19, but if a regression is based on transformed values of surrogate measurements, the same transformation should be applied to the surrogate data within a particular stratum before calculating the mean ($\bar{x}$) and variance (σ_x^2) to be used in the above equations. Also, note that the mean of a set of transformed numbers is generally not the same as applying the transformation to the mean of the same set of numbers.

3.4   Mean and Variance from Combined Direct and Surrogate Measurements

Given the mean values and coefficients of variation of the mean values for a design parameter taken from direct and surrogate measurements as respectively described in Sections 3.1 and 3.3, the mean value and variance of the mean value for the design parameter can be computed for the combined set of direct

and surrogate measurements for the stratum of interest. The mean value for the design parameter of interest is computed as

$$\bar{y} = \frac{\frac{\bar{y}_d}{\sigma_d^2} + \frac{\bar{y}_s}{\sigma_s^2}}{\frac{1}{\sigma_d^2} + \frac{1}{\sigma_s^2}} = \frac{\bar{y}_d \cdot \sigma_s^2 + \bar{y}_s \cdot \sigma_d^2}{\sigma_d^2 + \sigma_s^2}$$ (consistent units)   Eq. 20

where $\bar{y}$ is the mean or design value for the design parameter of interest from both direct and surrogate measurements, $\bar{y}_d$ and σ_d^2 are respectively the mean value of the design parameter and the variance of the mean value based on the available direct measurements (Eqs. 2 and 3), and $\bar{y}_s$ and σ_s^2 are respectively the mean value of the design parameter and the variance of the mean value based on the available surrogate measurements (e.g. Eqs. 12 and 14). Similarly, the variance of the mean value of the design parameter from the combined data set is computed as

$$\sigma^2 = \frac{\sigma_d^2 \cdot \sigma_s^2}{\sigma_d^2 + \sigma_s^2}$$ (consistent units)   Eq. 21

where σ^2 is the variance of the mean value of the design parameter of interest based on both the direct and surrogate measurements.

3.5   Summary

Equations 20 and 21 can be used to calculate design values for the mean and the variance of the mean, respectively, for a parameter of interest when both direct and surrogate measurements are available. The required inputs for Eqs. 20 and 21 are the mean and variance of the design parameter of interest from both sets of measurements (direct and surrogate). Defining the mean and variance for the direct measurements is straightforward via Eq. 2 and Eq. 3, respectively. The mean and variance from surrogate measurements depends on the type of regression and the assumed form of the relation between surrogate measurements and the design parameter of interest. Several different alternatives were described in the section including: Eq. 12 and Eq. 14 that apply when the regression is based on unweighted (OLS) linear regression; Equations 12 and 17 that apply when the regression is based on weighted linear regression through the origin; and Equations 18 and 19 that apply when the relation is based on linear OLS regression using logarithm transformed values for the direct measurements.

4.   Application of Procedures to Establish Uniaxial Compressive Strength from Direct Measurements and Standard Penetration Test Measurements in Missouri Shale

4.1   Background

The procedures described in Section 3 were applied and evaluated using data collected as part of a comprehensive site characterization program conducted in 2008-2010. The data considered consisted of uniaxial compressive strength values from uniaxial compression tests (direct measurements) and N_{eq}-values from SPT measurements (surrogate measurements) for 17 shale layers from five different test sites – Frankford, Grandview, kcICON, Lexington, and Warrensburg. The data were collected from "companion" borings consisting of a side-by-side pair of one boring with split spoon sampling and SPT measurements and a second boring using "best practice" drilling and sampling methods to retrieve the samples for laboratory testing (direct measurement). Each site had at least one set of companion borings; several had two. For each stratum, the mean value of N_{eq} (N_{eq-i}) was calculated and paired with the mean value of uniaxial compressive strength (q_{u-i}). Regression analyses were then performed on the collective data set to develop three different forms of correlation equations. Each form of correlation was

used along with the direct measurements to establish a mean value and variance of the mean value for the combined data set. These steps and the associated results are detailed in the following sections.

4.2 Regression Analyses and Design Values from Surrogate Measurements

The regression data set consisting of paired (N_{eq-i}, q_{u-i}) data were used to perform three regression analyses: "Method 1" refers to linear OLS regression of the untransformed data, "Method 2" refers to linear OLS regression of transformed data (ln N_{eq-i}, ln q_{u-i}), and "Method 3" refers to weighted regression for the untransformed data with the regression forced to pass through the origin. The resulting correlations are referred to as "ecological" correlations since the values of N_{eq-i} and q_{u-i} are based on average values for each layer instead of values from individual tests (i.e. one blow count per N_{eq-i}, one laboratory measurement per q_{u-i}).

Method 1: OLS of Untransformed Data, q_{u-i} vs. N_{eq-i}

Microsoft Excel© was used to perform an unweighted linear regression analysis of the regression data. Results of these analyses are shown in Figure 1. The coefficient of determination, R^2, for the regression is 0.43, indicating that the model explains 43 percent of the total variance. This is a modest, but less than ideal, correlation. The regression equation and model parameters shown in the figure and listed below were used with the procedures described in Section 3.3 to develop the mean and variance equations listed respectively as Equations 22 and 23.

$y = 0.1775 \cdot x + 11.26$ with parameters $s_y^2 = 1230$; $m = 17$; $\bar{x}_r = 242$, and $s_{xx} = 444{,}896$; therefore

$$\bar{y}_s = 0.1775 \cdot \bar{x} + 11.26 \qquad \text{(ksf)} \qquad \text{Eq. 22}$$

$$\sigma_s^2 \approx 1463 + 0.03427 \frac{\hat{\sigma}_x^2}{n_s} - 1.338 \cdot \bar{x} + 0.00276 \cdot \bar{x}^2 \qquad \text{(ksf)}^2 \qquad \text{Eq. 23}$$
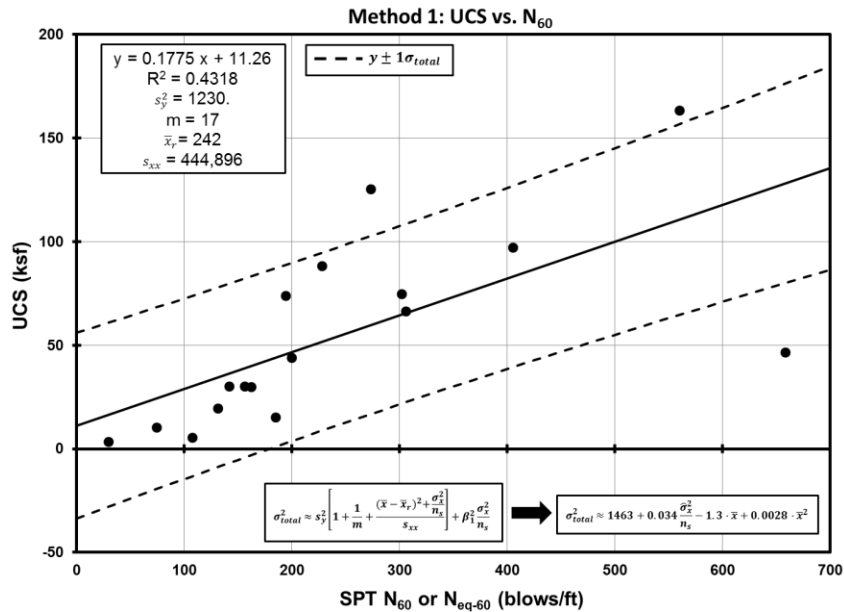

Figure 1. Regression "Method 1" using uniaxial compressive strength vs. N_{eq-60} from companion borings where each data point represents the mean of measurements from one stratum. N_{eq-60} refers to blow counts extrapolated linearly based on penetration of less than 12 in.

Method 2: OLS of Log Transformed Data, $\ln q_{u-i}$ vs. $\ln N_{eq-i}$

To improve the fit from Method 1, both the direct and surrogate data sets were transformed by calculating natural logarithms of each q_{u-i} and N_{eq-i}. Microsoft Excel© then was used to perform an unweighted linear regression analysis of the transformed data. Results are shown in Figure 2. The coefficient of determination, R^2, for the regression is 0.71, indicating the model explains 71 percent of the total variance. This is a substantial improvement over Method 1. The improvement is evident from the reduced scatter, as well. The regression equation shown in the figure was rearranged to produce Equation 24 for the shale-specific mean strength. Relevant parameters for the variance calculations are also shown on the figure and listed below and were used to develop the shale-specific variance equation shown as Equation 25. The $\ln$ modifiers are used for x and y to indicate that both sets of data were transformed. Note that transformation of the x variable (N_{eq-i}) affects the input of both Equations 24 and 25, which use the mean natural logarithm of blow count, $\overline{lnx}$, and the variance of the natural logarithm of blow count, $\hat{\sigma}^2_{lnx}$. Also, note that the mean of the natural logarithm of a set of numbers is not the same as the natural logarithm of the mean of the same set of numbers.

$\ln y = 1.354 \ln x - 3.587$ with parameters $s^2_{lny} = 0.3412$; $m = 17$; $\overline{lnx}_r = 5.135$, and $s_{lnx-lnx} = 6.962$;

therefore

$$\bar{y}_s = \exp{(1.354 \cdot \overline{lnx} - 3.587)} \hspace{4cm} \text{(ksf)} \hspace{2cm} \text{Eq. 24}$$

$$\sigma^2_s \approx \bar{y}^2_s \cdot (1.654 + 1.882 \frac{\hat{\sigma}^2_{lnx}}{n_s} - 0.503 \cdot \overline{lnx} + 0.0490 \cdot \overline{lnx}^2) \hspace{1cm} \text{(ksf)}^2 \hspace{1cm} \text{Eq. 25}$$
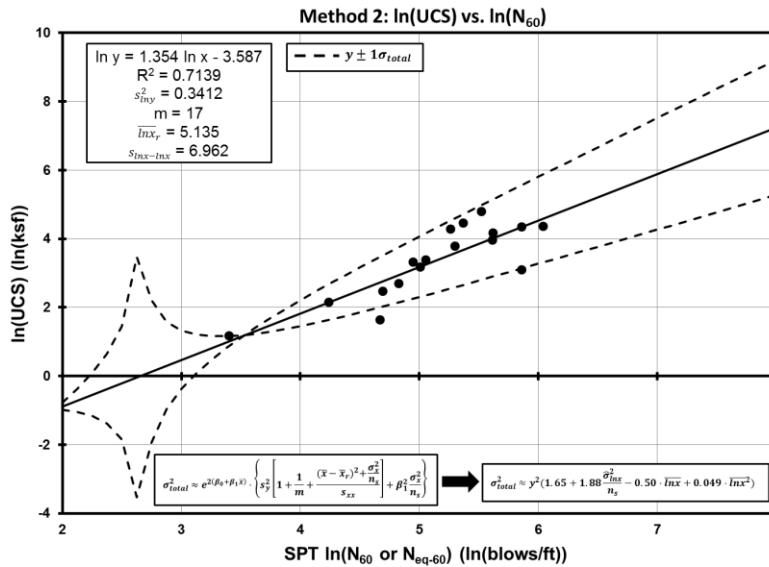

Figure 2. Regression "Method 2" using natural log of uniaxial compressive strength vs. natural log of N_{eq-60} for companion borings where each of the data points represents the mean of the log of measurements for one stratum. N_{eq-60} refers to blow counts extrapolated linearly based on penetration of less than 12 in.

Method 3: Weighted Least Squares Regression through the Origin of Untransformed Data, q_{u-i} vs. N_{eq-i}

A weighted least squares linear regression analysis of the data was performed with the weighting factor, $w_i = \frac{1}{x_i^2}$. Results are shown in Figure 3. The scatter shown is similar to that for Method 1, indicating a fit of similar quality, although this correlation is perhaps more practical since it is common to assume zero intercept for estimates of strength from blow counts in practice. The regression equation and model parameters are shown in the figure and listed below and were used with the equations of Section 3.3 to develop the mean and variance equations listed respectively as Equations 26 and 27.

$y = 0.2132 \cdot x$ with parameters $\frac{\sum w_i(\hat{y}_{r-i} - \hat{\beta}_1 \hat{x}_{r-i})^2}{\sum w_i \hat{x}_{r-i}^2} = 0.01248$, and $m = 17$;

therefore

$$\bar{y}_s = 0.2132 \cdot \bar{x} \qquad\qquad\qquad \text{(ksf)} \qquad\qquad \text{Eq. 26}$$

$$\sigma_s^2 \approx 1.001 \cdot \bar{x}^2 + 1.046 \frac{\hat{\sigma}_{\bar{x}}^2}{n_s} \qquad\qquad \text{(ksf)}^2 \qquad\qquad \text{Eq. 27}$$
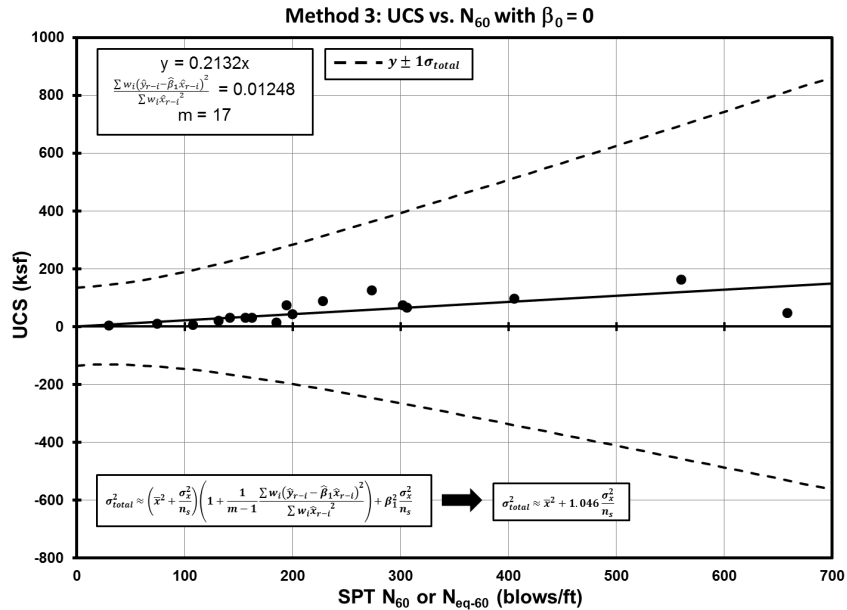
Figure 3. Regression "Method 3" using uniaxial compressive strength vs. N_{eq-60} for companion borings where each of the data points represents the mean of measurements for one stratum. N_{eq-60} refers to blow counts extrapolated linearly based on penetration of less than 12 in.

4.3   Design Values from Combined Direct and Surrogate Measurements

The methods proposed in Section 3 were used to estimate design values for the mean and variance of the mean for uniaxial compressive strength using both the direct measurements (uniaxial compression tests) and surrogate measurements (SPT measurements) for each stratum. Results of these analyses are summarized in Table 1. Means and variances for uniaxial compressive strength from the direct (uniaxial compression test) measurements were calculated according to Equations 2 and 3. The correlations and equations developed in Section 4.2 were used to estimate the means and variances for uniaxial compressive strength from the surrogate (SPT) measurements for each of the three correlation methods.

Table 1. Summary of mean and variance for uniaxial compressive strengths from direct, surrogate, and combined measurements. bpf, ksf, (ksf)^2, ln ksf, (ln ksf)^2

Site	Stratum	SPT $\bar{N}_{60}$ or $\bar{N}_{eq-60}$	n_s	Direct $\bar{y}_d$	σ_d^2	n_d	Surrogate Method 1 $\bar{y}_s$	σ_s^2	Method 2 $\bar{y}_s$	σ_s^2	Method 3 $\bar{y}_s$	σ_s^2	Combined Method 1 $\bar{y}$	σ^2	Method 2 $\bar{y}$	σ^2	Method 3 $\bar{y}$	σ^2
Frankford	Maquoketa Formation A	30	1	3.3	0.04	6	N/A – only one blow count in this layer						N/A – only one blow count in this layer					
	Maquoketa Formation B	75	2	10.4	2.5	20	24.5	1406	8.6	51	15.9	6346	10.4	2.5	10.3	2.4	10.4	2.5
	Maquoketa Formation C	306	5	66.2	4.7	41	65.6	1527	55.9	1434	65.2	100240	66.2	4.7	66.2	4.7	66.2	4.7
Grandview	Chanute Shale	131	6	19.4	22.4	12	34.6	1348	19.1	148	28.0	17656	19.7	22.1	19.4	19.5	19.5	22.4
	Quivira Shale	156	4	30.0	25.6	12	39.0	1346	24.4	250	33.3	25202	30.2	25.1	29.5	23.2	30.0	25.6
	Wea Shale	194	6	73.7	18.4	31	45.7	1311	34.5	437	41.4	37826	73.3	18.1	72.1	17.6	73.7	18.3
KC Icon	Weathered Shale	200	2	43.9	674	1	46.7	1307	36.1	474	42.6	39951	44.8	444.5	39.3	278	43.9	662.4
	Shale	273	8	125.3	30.6	22	59.8	1391	49.1	988	58.3	77469	123.8	30.0	123.0	29.7	125.2	30.6
Lexington	Bevier Formation	228	2	88.3	43.4	15	51.7	1501	39.9	940	48.6	58068	87.2	42.2	86.1	41.5	88.2	43.4
	Verdigris Formation	162	2	29.8	5.4	4	40.0	1376	26.1	329	34.6	28027	29.8	5.4	29.7	5.3	29.8	5.4
	Croweburg Formation	659	2	46.4	1183	4	128.2	12430	77.6	19819	140.4	759161	53.5	1080	48.1	1116	46.5	1181
	Fleming Formation A	405	2	97.2	328.9	11	83.2	2784	77.6	5747	86.4	207396	95.8	294.1	96.2	311	97.2	328.3
Warrensburg	Croweburg Formation B	185	7	15.0	4.0	43	44.1	1592	16.0	166	39.4	42702	15.1	4.0	15.0	3.9	15.0	4.0
	Croweburg Formation C	107	2	5.4	0.7	6	30.3	1359	15.4	96	22.9	11756	5.4	0.7	5.5	0.7	5.4	0.7
	Fleming Formation	302	6	74.7	90.2	34	64.9	1460	55.4	1356	64.5	95978	74.1	84.9	73.5	84.6	74.7	90.1
	Mineral Formation A	142	3	30.1	12.3	10	36.4	1334	22.6	190	30.2	20249	30.2	12.2	29.7	11.6	30.1	12.3
	Mineral Formation B	560	4	163.3	2030	20	110.7	3531	99.0	7545	119.4	373581	144.1	1289	149.7	1600	163.0	2019

The results shown in Table 1 were interpreted by computing the percent difference between the mean and variance of the mean established from the direct (lab) measurements alone and the mean and variance of the mean established using the combined direct and surrogate (SPT) measurements. These differences are shown in Table 2. The comparison of mean values shown in Table 2 indicates that estimates of the mean q_u were not significantly affected by combining direct and surrogate measurements. This is not surprising since estimates from the combined data sets were generated using the same data that was used to develop the correlations. Consideration of the change in variance of the mean is therefore more informative.

For all three correlation methods and for all strata, the combined estimate has a lower variance than established using either the direct or surrogate measurements alone. This is intuitive; consideration of additional data from the surrogate measurements should of course reduce the variance. Variances computed from the combined estimates show the greatest reduction in variance for Method 2, somewhat less reduction for Method 1, and virtually zero reduction for Method 3. These changes in the variance of the mean are consistent with the correlations described in Section 4.2 and with the error bounds shown in Figures 1 through 3. The reduction in variance is greatest for Method 2 because the variance inherent to the correlation between the log transformed data sets is less than the variance inherent to the other methods. Likewise, the reduction in variance is negligible for Method 3 because the variance inherent to the correlation is too great to provide substantive value. The practical implications of the reductions in variance are considered in more detail by a set of calculations of factored axial load capacity for a drilled shaft in Section 5.

Table 2. Percent difference between variance established from combined direct and surrogate measurements and variance established from direct measurements alone, relative to variance from direct measurements alone.

Site	Stratum	Method 1		Method 2		Method 3	
		$\bar{y}$	σ^2	$\bar{y}$	σ^2	$\bar{y}$	σ^2
Frankford	Maquoketa Formation A	N/A – Only one blow count in this layer.					
	Maquoketa Formation B	0	0	-1	-5	0	0
	Maquoketa Formation C	0	0	0	0	0	0
Grandview	Chanute Shale	1	-2	0	-15	0	0
	Quivira Shale	1	-2	-2	-10	0	0
	Wea Shale	-1	-1	-2	-4	0	0
KC Icon	Weathered Shale	2	-52	-12	-142	0	-2
	Shale	-1	-2	-2	-3	0	0
Lexington	Bevier Formation	-1	-3	-2	-5	0	0
	Verdigris Formation	0	0	0	-2	0	0
	Croweburg Formation	13	-10	4	-6	0	0
	Fleming Formation A	-2	-12	-1	-6	0	0
Warrensburg	Croweburg Formation B	0	0	0	-2	0	0
	Croweburg Formation C	0	0	1	-1	0	0
	Fleming Formation	-1	-6	-2	-7	0	0
	Mineral Formation A	0	-1	-2	-6	0	0
	Mineral Formation B	-13	-57	-9	-27	0	-1

5. Example Calculations: Shaft TS-W8 at MoDOT Test Site in Warrensburg, MO

Axial load capacity was calculated for test shaft W8 at the MoDOT test site in Warrensburg to demonstrate application of the procedures and equations outlined in this report. The calculation procedure consists of (1) summarizing the direct and surrogate data sets, (2) calculating means and variances for the direct measurements, (3) calculating means and variances for the surrogate measurements, (4) applying regression equations that relate SPT measurements to q_u to determine means and variances for q_u from surrogate measurements, (5) using the results of steps (3) and (4) to determine means and variances for q_u from the combined measurements, and (6) using the established

design parameters (from direct, surrogate, or combined measurements) to calculate axial resistance per established MoDOT EPG guidelines.

5.1   Site, Shaft, and Subsurface Data

The Warrensburg test site has a variable stratigraphy that generally consists of 10 to 15 ft of silty clay overburden overlaying Pennsylvanian bedrock. The Croweburg formation is the topmost rock formation, which consists of a discontinuous sandstone layer (Croweburg A; not present at TS-W8) up to 5-ft thick, a sandy shale (Croweburg B) approximately 20-ft thick, and a soft shale (Croweburg C) about 5-ft thick. The Croweburg Formation is underlain by the Fleming Formation, a hard shale approximately 15 to 20 ft thick. The Fleming Formation is underlain by the Mineral and Scammon Formations.

TS-W8 is a nominally 36-in. diameter drilled shaft with a permanent casing extending down to the top of the Croweburg Formation. The casing diameter is 42 in., but contributions to axial capacity from the silty clay overburden are neglected for the purpose of this example, so the calculations use a shaft diameter of 36 in. The tip of the shaft is at Elev. 735.1 in the Fleming Formation.

Subsurface data for the site is extensive, but for the purpose of this example, only q_u and SPT N_{eq} data are considered. The data used for the example calculation are listed in Table 3. The comprehensive site characterization program described in Section 4.1 included a total of 83 q_u measurements for the layers contributing to the axial capacity of TS-W8. This is more measurements than typically would be collected for a bridge project, so approximately two-thirds of the q_u measurements from each layer were ignored in order to illustrate a project more representative of standard practice. As a result, the mean q_u values for the three layers in this example are slightly different from those listed in Table 1, and the variances of the mean values are notably greater.

Table 3.   Subsurface data used for TS-W8 example calculation. Note the assumption of constant layer properties (with depth) makes elevations of individual measurements inconsequential.

Stratum	Elevation	Direct Measurements, $q_{u-i} = y_i$ (ksf)	Surrogate Measurements	
			Untransformed $N_{eq-i} = x_i$ (bpf)	Natural Log Transformation $\ln(N_{eq-i}) = \ln(x_i)$, (ln(bpf))
Croweburg Formation B	768.3 to 750.3	17.4, 4.9, 8.1, 5.1, 3.2, 9.0, 6.0, 7.7, 3.3, 40.8, 3.1, 9.5, 11.2, 82.1, 14.3	102, 704, 47, 243, 41, 55, 101	4.6, 6.6, 3.9, 5.5, 3.7, 4.0, 4.6
Croweburg Formation C	750.3 to 744.3	8.3, 2.8	122, 93	4.8, 4.5
Fleming Formation	744.3 to 735.1	67.1, 5.6, 46.8, 24.7, 11.1, 155.2, 143.7, 106.1, 56.7, 103.2, 132.0	608, 203, 243, 152, 304, 304	6.4, 5.3, 5.5, 5.0, 5.7, 5.7

5.2   Calculation of Means and Variances for q_u from Direct Measurements

Equations 1, 2, and 3 were applied to the direct measurements in Table 3 to calculate the mean value, variance of the mean value, and coefficient of variation for the mean value of q_u from direct measurements in each stratum. The results are shown in Table 4. These values are used subsequently to calculate the means and variances from the combined measurements as well as to calculate the factored axial shaft resistance without consideration of surrogate measurements.

Table 4.   Mean and variance of the mean value for q_u from direct measurements.

Stratum	Elevation	Number of Direct Measurements, n_d	Mean Value, $\bar{y}_d$ (ksf)	Variance of the Mean Value, σ_d^2 (ksf)^2	Coefficient of Variation of the Mean Value, COV
Croweburg Formation B	768.3 to 750.3	15	15.1	28.9	0.36
Croweburg Formation C	750.3 to 744.3	2	5.5	7.5	0.50
Fleming Formation	744.3 to 735.1	11	77.5	261.1	0.21

5.3   Calculation of Means and Variances for Surrogate Measurements

Equations 15 and 16 were applied to the surrogate measurements in Table 3 to calculate the mean and variance of surrogate measurements for each stratum. The results are shown in Table 5. These values are used subsequently as inputs for the regression equations presented in Section 4.2.

Table 5.   Mean and variance of surrogate measurements, with and without natural log transformation.

Stratum	Elevation	Number of Surrogate Measurements, n_s	Untransformed		Natural Log Transformation	
			Mean Value, $\bar{x}$ (bpf)	Variance, σ_x^2 (bpf)^2	Mean Value, $\overline{lnx}$ ln(bpf)	Variance, σ_{lnx}^2 (ln(bpf))^2
Croweburg Formation B	768.3 to 750.3	7	185	57,238	4.7	1.04
Croweburg Formation C	750.3 to 744.3	2	107	403	4.7	0.037
Fleming Formation	744.3 to 735.1	6	302	25,894	5.6	0.22

5.4   Application of Regression Equations to Determine Means and Variances for q_u from Surrogate Measurements

Equations 24 through 27 were applied to the results presented in Table 5 to determine the mean value and the variance of the mean value for q_u from SPT measurements using Methods 2 and 3 from Section 4.2. The results are presented in Table 6. These results are subsequently used with similar results obtained from the direct measurements (Table 4) to calculate means and variances for q_u from the combined measurements for each stratum. They are also used to calculate the factored axial shaft resistance based on the surrogate measurements alone, without consideration of direct measurements.

Table 6.   Mean and variance of the mean value for q_u from surrogate measurements.

Stratum	Elevation	Method 2: Natural Log Transformed			Method 3: Untransformed through Origin		
		Mean Value, $\bar{y}_s$ (ksf)	Variance of the Mean Value, σ_s^2 (ksf)^2	Coefficient of Variation of the Mean Value, COV	Mean Value, $\bar{y}_s$ (ksf)	Variance of the Mean Value, σ_s^2 (ksf)^2	Coefficient of Variation of the Mean Value, COV
Croweburg Formation B	768.3 to 750.3	16.0	166	0.81	39.4	42,702	5.25
Croweburg Formation C	750.3 to 744.3	15.4	96	0.64	22.9	11,756	4.73
Fleming Formation	744.3 to 735.1	55.4	1356	0.66	64.5	95,978	4.81

5.5   Calculation of Means and Variances for q_u from Combined Measurements

The data from Table 4 (established from direct measurements) and Table 6 (established from SPT measurements) were used with Equations 20 and 21 to determine the means and variances of the mean for q_u from the combined direct and surrogate measurements. The results are shown in Table 7. Two sets of parameters are included: one established using Method 2, which utilized the natural log transformation, and one using Method 3, which utilized the untransformed data with the regression forced through the origin. These results are subsequently used to calculate factored axial capacities considering both direct and surrogate measurements.

Table 7.   Mean and variance of the mean value for q_u from combined direct and surrogate measurements.

Stratum	Elevation	Method 2: Natural Log Transformed			Method 3: Untransformed through Origin		
		Mean Value, $\bar{y}$ (ksf)	Variance of the Mean Value, σ^2 (ksf)^2	Coefficient of Variation of the Mean Value, COV	Mean Value, $\bar{y}$ (ksf)	Variance of the Mean Value, σ^2 (ksf)^2	Coefficient of Variation of the Mean Value, COV
Croweburg Formation B	768.3 to 750.3	15.2	24.6	0.33	15.1	28.8	0.36
Croweburg Formation C	750.3 to 744.3	6.3	7.0	0.42	5.6	7.5	0.49
Fleming Formation	744.3 to 735.1	73.9	73.9	0.20	77.4	260	0.21

5.6   Calculation of Factored Axial Load Capacity

Values from Table 4 through Table 7 were used to calculate the factored axial capacity for TS-W8 according to existing MoDOT design guidelines for drilled shaft capacity in weak rock (EPG 751.37.3.2 and EPG 751.37.3.3). Factored shaft resistance was calculated in six different ways:

(1) using direct/laboratory measurements only,
(2) using surrogate measurements only and Method 2 (log transformation) regression,
(3) using surrogate measurements only and Method 3 (untransformed through origin) regression,
(4) using combined direct and surrogate measurements and Method 2 regression,
(5) using combined direct and surrogate measurements and Method 3 regression, and
(6) directly from N_{eq-60} without correlation to q_u (as provided in EPG 751.37.3.3)

The factored shaft resistances established using each of these methods are presented in this section. For all calculations, the terms are consistent with those used in EPG 751.37.3:

$\overline{q_u}$ = mean value of uniaxial compressive strength of rock core along the shaft segment (ksf),

q_s = nominal unit side resistance for the shaft segment (ksf),

φ_{qs} = resistance factor for unit side resistance along shaft (dimensionless),

R_{sR} = factored side resistance (consistent units of force),

q_p = nominal unit tip resistance (consistent units of stress),

φ_{qp} = resistance factor for unit tip resistance (dimensionless),

R_{pR} = factored tip resistance (consistent units of force), and

R_R = factored axial shaft resistance (consistent units of force).

Not shown explicitly are the shaft areas, which are consistent for all calculations. For TS-W8, the unit side area is 9.42 ft^2/ft, and the tip area is 7.07 ft^2 based on a nominal diameter of 3 ft.

(1) Shaft Resistance from Direct/Laboratory Measurements of q_u

Factored axial resistance was calculated based on laboratory measurements of q_u alone following the provisions of EPG 751.37.3.2 using the values from Table 4. The results are shown in Table 8. The factored axial resistance for TS-W8 is 2100 kips when considering only direct measurements of q_u.

Table 8.   Factored axial resistance calculations using only direct measurements of q_u.

Stratum	$\overline{q_u} = \bar{y}_d$, ksf	COV_{qu}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{SR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	15.1	0.36	6.5	0.275	248	96	0.495	337	584
Croweburg Formation C	5.5	0.50	2.9	0.185	31	47	0.40	133	411
Fleming Formation	77.5	0.21	23.6	0.26	533	307	0.59	1281	2092

[1]Side resistance values listed are total within the stratum; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

(2) Shaft Resistance from Surrogate Measurements with Method 2 (Log Transformation) Regression

Factored axial resistance was calculated using SPT measurements as a surrogate for q_u following the provisions of EPG 751.37.3.2 with the values from Table 6 for the Method 2 (log transformation) regression. The results are shown in Table 9. The factored axial resistance for TS-W8 is 960 kips when using only the SPT data with log-transformed regression.

Table 9.   Factored axial resistance calculations using only SPT measurements as a surrogate for q_u with Method 2 regression.

Stratum	$\overline{q_u} = \bar{y}_d$, ksf	COV_{qu}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{SR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	16.0	0.81	6.8	0.12	138	100	0.245	174	312
Croweburg Formation C	15.4	0.64	6.6	0.155	58	98	0.320	221	417
Fleming Formation	55.4	0.66	18.1	0.15	236	242	0.31	531	962

[1]Side resistance values listed are total within the layer; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

(3) Shaft Resistance from Surrogate Measurements with Method 3 (Untransformed through Origin) Regression

Factored axial resistance was calculated based using SPT measurements as a surrogate for q_u following the provisions of EPG 751.37.3.2 with the values from Table 6 for the Method 3 (untransformed through origin) regression. The results are shown in Table 9. The factored axial resistance for TS-W8 is 660 kips when using only the SPT data with regression through the origin.

Table 10. Factored axial resistance calculations using only SPT measurements as a surrogate for q_u with Method 3 regression.

Stratum	$\overline{q_u} = \bar{y}_d$, ksf	COV_{qu}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{sR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	39.4	5.24	13.8	0.08	188	190	0.15	202	389
Croweburg Formation C	22.9	4.73	9.0	0.08	41	129	0.15	137	366
Fleming Formation	64.5	4.80	20.4	0.08	142	270	0.15	286	656

[1]Side resistance values listed are total within the layer; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

(4) Shaft Resistance from Combined Measurements with Surrogate by Method 2 (Log Transformation) Regression

Factored axial resistance based on the combined (direct and surrogate) estimates of q_u was calculated according to the provisions of EPG 751.37.3.2 with the data from Table 7 and with Method 2 (log transformed) regression applied to the surrogate measurements. The results are shown in Table 11. The factored axial resistance for TS-W8 is 2200 kips when using the combined measurements with the log-transformed regression.

Table 11. Factored axial resistance calculations using combined estimates for q_u with Method 2 regression.

Stratum	$\overline{q_u} = \bar{y}_d$, ksf	COV_{qu}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{sR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	15.2	0.33	6.5	0.235	260	97	0.515	352	612
Croweburg Formation C	6.3	0.42	3.3	0.21	39	52	0.45	165	463
Fleming Formation	73.9	0.12	22.8	0.275	543	297	0.64	1344	2185

[1]Side resistance values listed are total within the layer; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

(5) Shaft Resistance from Combined Measurements with Surrogate by Method 3 (Untransformed through Origin) Regression

Factored axial resistance based on the combined (direct and surrogate) estimates of q_u was calculated according to the provisions of EPG 751.37.3.2 with the data from Table 7 and with Method 3 (untransformed through origin) regression applied to the surrogate measurements. The results are shown in Table 12. The factored axial resistance for TS-W8 is 2100 kips when using the combined measurements with the regression forced to pass through the origin.

Table 12. Factored axial resistance calculations using combined estimates for q_u with Method 3 regression.

Stratum	$\overline{q_u} = \overline{y}_d$, ksf	COV_{qu}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{sR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	15.1	0.36	6.5	0.225	248	96	0.495	337	584
Croweburg Formation C	5.6	0.49	2.7	0.19	32	48	0.405	136	416
Fleming Formation	77.4	0.21	23.6	0.26	532	307	0.59	1280	2092

[1]Side resistance values listed are total within the layer; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

(6) Shaft Resistance Directly from SPT Measurements Without Correlation to q_u

Factored axial resistance was also calculated directly from SPT measurements without correlation to q_u according to the provisions of EPG 751.37.3.3 using the values from Table 5. The results are shown in Table 13. The factored axial resistance for TS-W8 is 640 kips when using only the SPT data without correlation to q_u.

Table 13. Factored axial resistance calculations for direct use of SPT measurements without correlation to q_u based on EPG 751.37.3.3.

Stratum	$\overline{N}_{eq}$, bpf	COV_{Neq}	Side Resistance[1]			Tip Resistance[1]			R_R, kip[1]
			q_s, ksf	φ_{qs}	R_{sR}, kip	q_p, ksf	φ_{qp}	R_{pR}, kip	
Croweburg Formation B	185.0	1.29	13.2	0.05	112	116	0.055	45	157
Croweburg Formation C	107.0	0.19	7.6	0.22	95	67	0.24	113	321
Fleming Formation	302.0	0.53	21.6	0.13	243	189	0.14	187	637

[1]Side resistance values listed are total within the layer; tip resistance values are for the bottom of the layer, and the total resistance values are cumulative at the bottom of the layer.

5.7   Summary of Axial Load Capacities

Results of calculations presented in Section 5.6 are summarized in Table 14. The results were also used to produce the plot of factored axial load capacity versus elevation ("design curves") shown in Figure 4. Comparison of the factored capacity values is informative. The "baseline" factored capacity using just the direct (laboratory) measurements of q_u is 2100 kips based on Calculation (1). The only estimate that provides any improvement (higher capacity) is Calculation (4), which is based on use of the combined measurements with surrogate values correlated via the log transformed regression. That estimate resulted in a factored capacity of 2200 kips, which represents a 4% improvement. This suggests the value of SPT measurements as a surrogate for q_u is limited; the variance inherent in the correlation between SPT measurements and q_u is simply too great for the SPT measurements to provide much benefit even when the number of SPT measurements is relatively large. The limited value of SPT as a surrogate for q_u is confirmed by Calculations (2) and (3), which indicate the factored capacity is less than half of the baseline value (direct measurements only) when only the surrogate measurements are considered alone. Comparison of Calculations (2) and (3) and of Calculations (4) and (5) also confirm that the improvement offered by using the log transformed regression (Method 2) over the regression through the origin (Method 3) is significant, resulting in a nearly 50% increase in factored capacity when q_u data are not considered. Results of Calculation (6) provide further evidence that SPT measurements are not particularly valuable as a predictor of axial load capacity. The direct calculation of factored axial load

capacity from SPT measurements via EPG 751.37.3.3 (rather than indirectly through correlation with q_u) resulted in the lowest factored capacity of all calculations, 640 kips. This value is similar to the value predicted by Calculation (3), which uses the same data and is based on similar assumptions but calculates capacity through correlations with q_u.

Table 14. Summary of factored axial resistance calculations for six different methods.

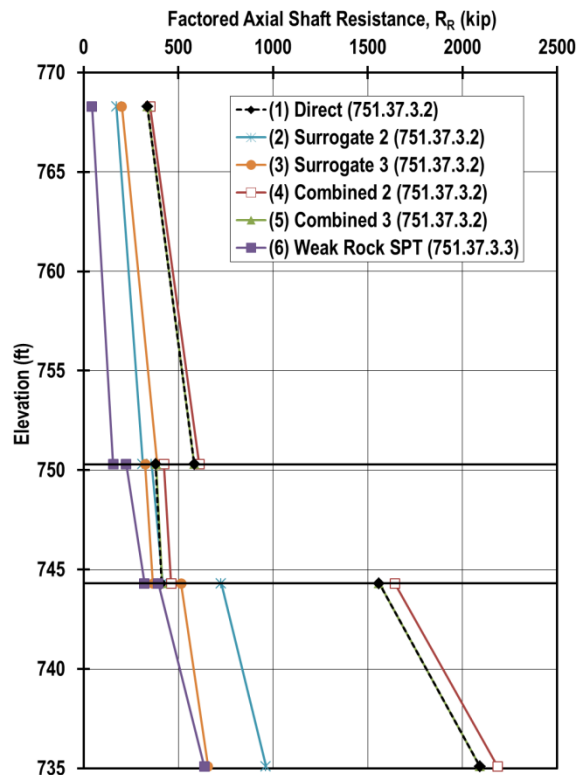| Calculation Number	Measurements Considered	Calculation Method	Regression Method	R_R, kip[1]
(1)	q_u	EPG 751.37.3.2	N/A	2092
(2)	$N_{eq\text{-}60}$	EPG 751.37.3.2	2 (Log Transformed)	962
(3)	$N_{eq\text{-}60}$	EPG 751.37.3.2	3 (Untransformed through Origin)	656
(4)	q_u and $N_{eq\text{-}60}$	EPG 751.37.3.2	2 (Log Transformed)	2185
(5)	q_u and $N_{eq\text{-}60}$	EPG 751.37.3.2	3 (Untransformed through Origin)	2092
(6)	$N_{eq\text{-}60}$	EPG 751.37.3.3	N/A	637

Figure 4. Factored axial resistance values versus elevation for six different methods.

6. Conclusions & Recommendations

This report documents proposed procedures for calculating means and variances for a design parameter of interest when there are two different sources of data for the parameter. The report also presents procedures for developing correlations between direct and surrogate data sources (Section 3.2), and for

how to estimate variance introduced from these correlations (Section 3.3). The report also documents application of these procedures for the case of using SPT measurements to establish design values for uniaxial compressive strength in Missouri shales (Section 4) to result in a useful set of correlations and equations for estimating the mean and variance of uniaxial compressive strength values in shale based on SPT measurements. Results of these analyses were applied in sample calculations for the factored axial capacity of a drilled shaft at the Warrensburg test site (Section 6).

The results of developing the procedures and applying them to SPT measurements in Missouri shale leads to several notable conclusions:

- The equations for estimating the mean (Eq. 20) and variance (Eq. 21) for a design parameter based on combined direct and surrogate measurements provide a useful, practical, statistically appropriate, and widely applicable method for considering combined measurements from two different data sources in the context of reliability calculations.
- Similarly, the equations developed for estimating the variance of design parameters derived from surrogate measurements that can be correlated to the parameters from linear regressions (Section 3.3) provide a practical method for considering the variability of not only the surrogate measurements, but also the variability of design parameter values derived from the surrogate measurements. Both are necessary considerations for reliability-based analysis and design and for use of both direct and surrogate data within existing MoDOT LRFD guidelines.
- Application of the general procedure to the specific case of uniaxial compressive strength for Missouri shale resulted in equations that allow for appropriate utilization of both direct and surrogate measurements within existing EPG provisions for shale sites characterized by both lab and SPT measurements.
- The Missouri shale application case also resulted in notable conclusions about the procedure for establishing design values for parameters from two different types of measurements:
  - Consideration of multiple data sources always resulted in a reduction of variance for the design parameter, which reduces the coefficient of variation (COV) and therefore increases the resistance factor to improve design efficiency.
  - Stronger correlations (i.e. better fit from regression analysis) lead to more accurate estimates of the mean.
  - Logarithmic transformation of the SPT and q_u measurements dramatically improves the quality of the correlation between the measurements and results in a greater reduction in variance when using the combined measurements for design.
- Calculations of drilled shaft capacity using direct, surrogate, and combined measurements demonstrate application of the procedures of this report start-to-finish.
- Results of the drilled shaft capacity calculations suggest SPT measurements are not a strong predictor of q_u or shaft capacity; the variance inherent in the correlation between SPT and q_u is substantial, so estimates of q_u based on SPT are not especially valuable.

The methods presented in this report also provide practical and useful means for incorporating multiple data sources into load and resistance factor design methods. The methods and results presented open the door to a substantial amount of potential future work that includes:

- Developing procedures and relations similar to those presented for SPT measurements in shales but for cone penetration test (CPT) measurements. Such procedures will not only provide another useful correlation but also make the general procedure more robust by addressing cases where individual surrogate measurements are not independent (as is the case for CPT data). Other correlations (potentially based on pocket penetrometer, torvane, etc.) should also be considered, depending on MoDOT's experience and needs.
- Evaluating the shale-specific procedure presented in Section 4 with an independent set of data to evaluate how the procedure and correlations perform for data not used in the regression analyses. This could be accomplished using the data from the new I-64 design-build project in St. Louis.

- Developing similar procedures for strata without constant layer properties (e.g. a layer with linearly increasing strength).
- Expanding the generalized procedure (Section 3) for the case of three (or potentially even more than three) data sources.
- Applying the general procedure to different types of surrogate measurements that lends itself to a transformation that is not based on logarithms to further evaluate the effect of transformed correlations on estimated variance.

7. References

Bain, L.J. and M. Engelhardt (1992), *Introduction to Probability and Mathematical Statistics*, Brooks/Cole, Sectond Edition, 644 pp.

Loehr, J.E., J.J. Bowders, L. Ge, W.J. Likos, R. Luna, N. Maerz, B.L. Rosenblad, and R.W. Stephenson (2011a), *Engineering Policy Guidelines for Design of Drilled Shafts*, Missouri Department of Transportation, Report cmr12003, 75 pp.

Loehr, J.E., J.J. Bowders, L. Ge, W.J. Likos, R. Luna, N. Maerz, B.L. Rosenblad, and R.W. Stephenson (2011b), *Engineering Policy Guidelines for Design of Earth Slopes*, Missouri Department of Transportation, Report cmr12004, 19 pp.

Loehr, J.E., J.J. Bowders, L. Ge, W.J. Likos, R. Luna, N. Maerz, B.L. Rosenblad, and R.W. Stephenson (2011c), *Engineering Policy Guidelines for Design of Spread Footings*, Missouri Department of Transportation, Report cmr12005, 42 pp.

Loehr, J.E., J.J. Bowders, B.L. Rosenblad, R. Luna, N. Maerz, R.W. Stephenson, W.J. Likos and L. Ge (2013), "Implementation of LRFD Methods to Quantify Value of Site Characterization Activities," *Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering*, Paris, France.