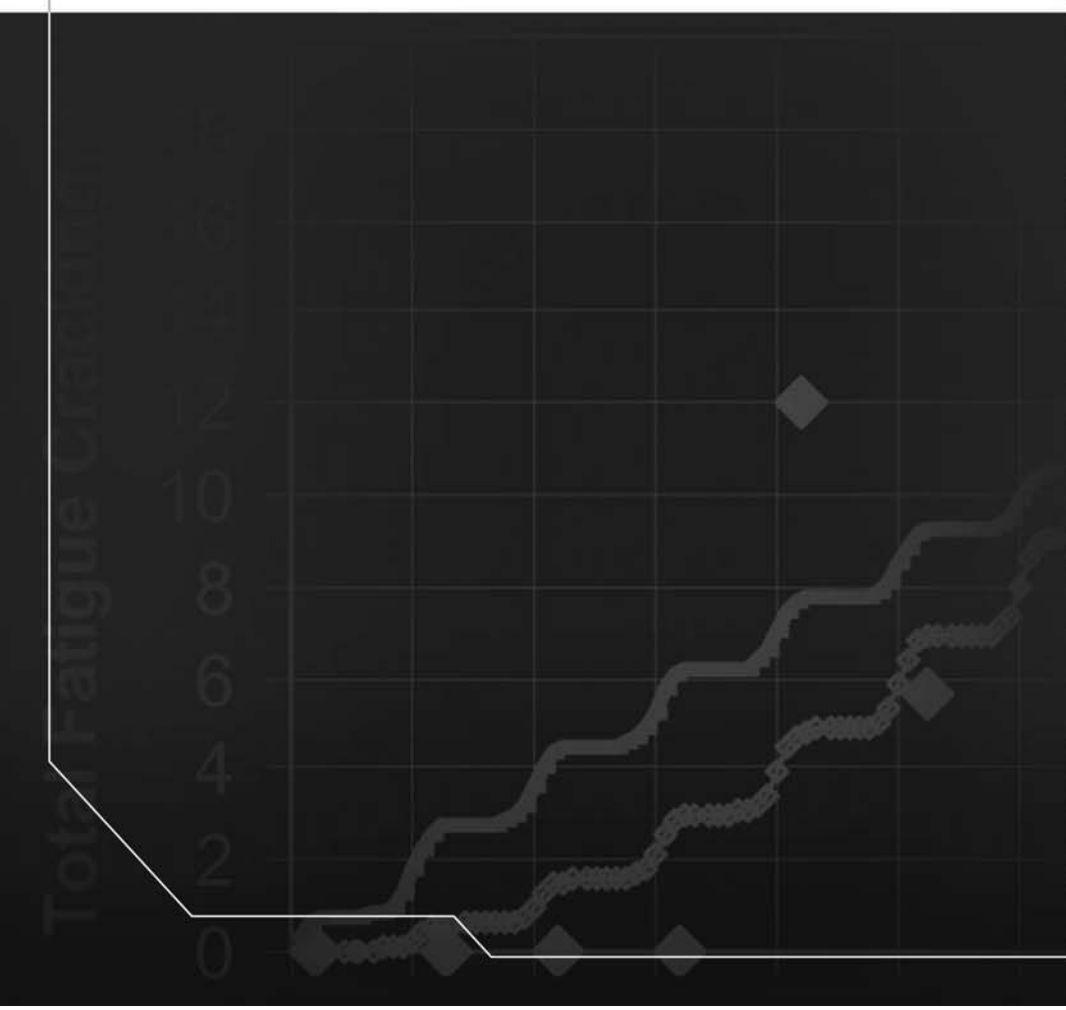
Guide for the Local Calibration of the Mechanistic-Empirical Pavement Design Guide

November 2010



Publication Code: LCG-1
ISBN: 978-1-56051-449-7

Preface

This guide is to provide guidance to calibrate the *Mechanistic-Empirical Pavement Design Guide* (MEPDG) software to local conditions, policies, and materials and to conduct the local calibration process. The guide does not provide guidance for determining the inputs and running the MEPDG software. A separate document, the *Mechanistic-Empirical Pavement Design Guide—A Manual of Practice*, provides guidance for using the MEPDG software to analyze and design new pavements and rehabilitation strategies. The *Manual of Practice* is referenced throughout this guide.

Version 1.0 of the MEPDG software is currently available. It should be noted that version 2.0 of the MEPDG software is in the process of being developed. Version 2.0 may include different transfer functions for selected distresses based on the results and recommendations from other on-going NCHRP projects. If any of the transfer functions are revised, the *Guide for Local Calibration* and the *Mechanistic-Empirical Pavement Design Guide—A Manual of Practice* for the MEPDG software may need to be revised accordingly.

Table of Contents

1.0 Introduction

The overall objective of the *Mechanistic-Empirical Pavement Design Guide* (MEPDG) is to provide the highway community with a state-of-the-practice tool for the design of new and rehabilitated pavement structures, based on mechanistic-empirical (M-E) principles. This means that the design procedure calculates pavement responses (stresses, strains, and deflections) and uses those responses to compute incremental damage over time. The procedure empirically relates the cumulative damage to observed pavement distresses. This M-E based procedure is shown in flowchart form in Figure 1-1. "MEPDG," as used in this guide, refers to the documentation and software package (NCHRP 2007).

Pavement distress prediction models, or transfer functions, are the key components of any M-E design and analysis procedure. The accuracy of performance prediction models depends on an effective process of calibration and subsequent validation with independent data sets. Pavement engineers gain confidence in the procedure by seeing an acceptable correlation between observed levels of distress in the field and those levels predicted with the performance model or transfer function. The validation of the performance prediction model is a mandatory step in their development to establish confidence in the design and analysis procedure and facilitate its acceptance and use. It is also necessary to establish the design reliability procedure. It is essential that distress prediction models be properly calibrated prior to adopting and using them for design purposes.

The term calibration refers to the mathematical process through which the total error (often termed residual) or difference between observed and predicted values of distress is minimized. The term validation refers to the process to confirm that the calibrated model can produce robust and accurate predictions for cases other than those used for model calibration. A successful validation process requires that the bias and precision statistics of the model for the validation data set be similar to those obtained during calibration. This calibration-validation process is critical for potential users to have confidence in the design procedure.

All performance models in the MEPDG were calibrated on a global level to observed field performance over a representative sample of pavement test sites throughout North America. The Long Term Pavement Performance (LTPP) test sections were used extensively in the calibration process, because of the consistency in the monitored data over time and the diversity of test sections spread throughout North America. Other experimental test sections were also included such as

MnRoad and Vandalia. However, policies on pavement preservation and maintenance, construction and material specifications, and materials vary across the United States and are not considered directly in the MEDPG. These factors can be considered indirectly through the local calibration parameters included in the MEPDG. The purpose of this guide is to provide guidance in calibrating the MEPDG to local conditions and materials that may not have been included in the global calibration process.

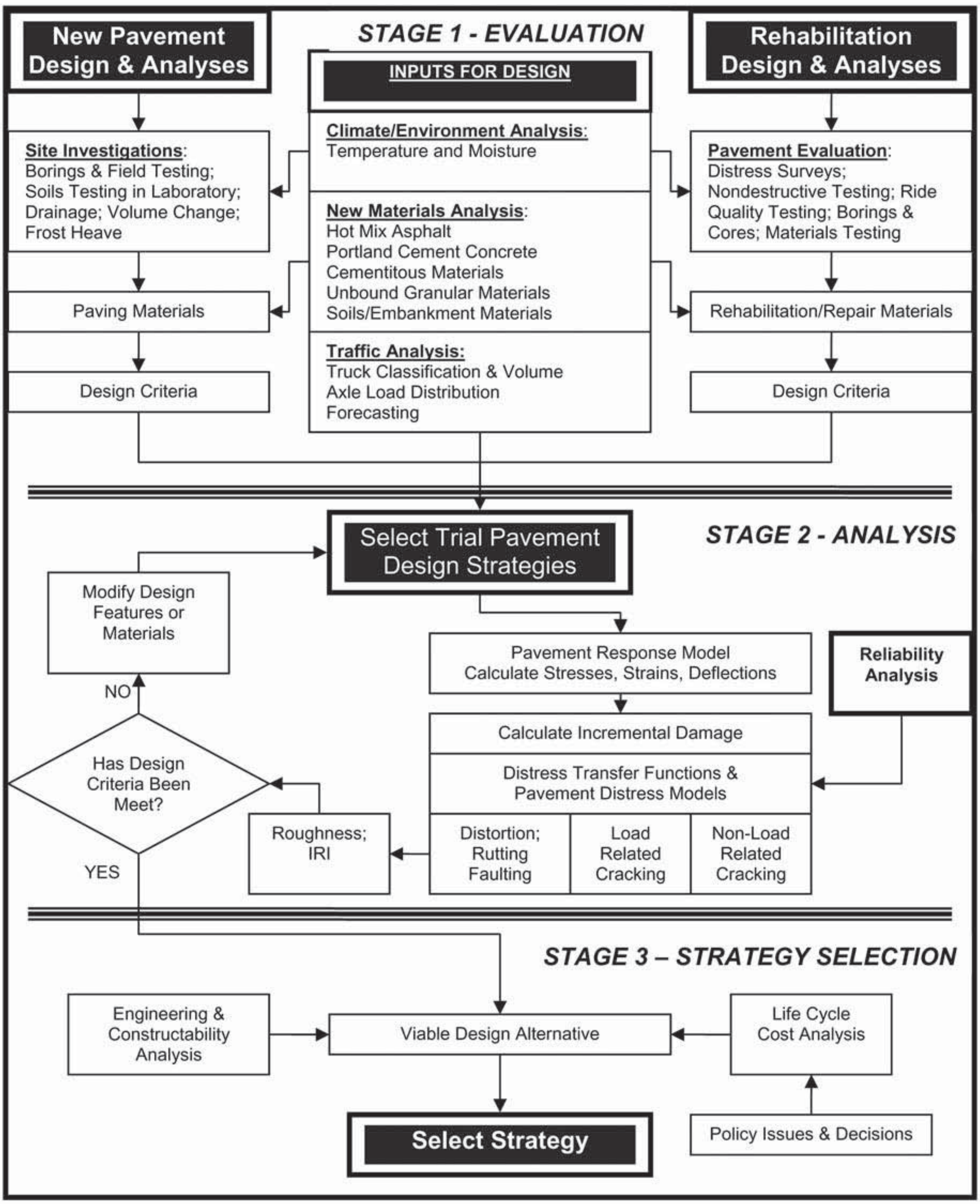


Figure 1-1. Conceptual Flow Chart of the Three-Stage Design/Analysis Process for the MEPDG

2.0 Terminology and Definition of Terms

This section provides the definition of selected terms that are used in the MEPDG local calibration process.

2.1 Statistical Terms

Accuracy—The exactness of a prediction to the observed or "actual" value. The concept of accuracy encompasses both precision and bias.

Bias—An effect that prevents predictions from simulating "real world" observations by systematically distorting it, as distinct from a random error that may distort on any one occasion but balances out on the average. A prediction model that is "biased" is significantly over- or under-predicting observed distress or roughness (as measured by the International Roughness Index [IRI]).

Calibration—A systematic process to eliminate any bias and minimize the residual errors between observed or measured results from the real world (e.g., the measured mean rut depth in a pavement section) and predicted results from the model (e.g., predicted mean rut depth from a permanent deformation model). This is accomplished by modifying empirical calibration parameters or transfer functions in the model to minimize the differences between the predicted and observed results. These calibration parameters are necessary to compensate for model simplification and limitations in simulating actual pavement and material behavior.

Model, Mathematical—A model that is derived from fundamental engineering principles that represent exact, error-free assumed relationships among the variables. The JULEA program is the mathematical or structural response model used for flexible pavements to calculate pavement responses (deflections, stresses, and strains), while the ISLAB2000 program is used for rigid pavements. A stress dependent finite element program is also available for flexible pavement analyses using Input Level 1 for unbound materials, but that model is intended for research purposes only. The Integrated Climatic Model (ICM) is also considered a mathematical model within the MEPDG.

Model, Statistical—A model that is derived from data that are subject to various types of observational, experimental, and measurement errors. The statistical models in the MEPDG include the distress transfer functions and the IRI regression equations. The time-dependent material property models for hot mix asphalt (HMA) and Portland cement concrete (PCC) are also regression or statistical relationships. These models, however, are assumed to be correct in the MEPDG model formulation or computational methodology. Adjustments to the coefficients of these relationships are not permitted within Version 1.0 of the MEPDG.

Model, Simulation or Prediction—Prediction models take two related forms. First the real-world system under investigation is approximated by a conceptual model. A conceptual model is a series of mathematical and logical relationships concerning the components and the structure of the system. Then the conceptual model is coded into a computer-recognizable form, the operation model, which is an approximate representation of the real-world system. The MEPDG operation models combine mathematical and statistical models.

Precision—The ability of a model to give repeated estimates that correlate strongly with the observed values. They may be consistently higher or lower but they correlate strongly with observed values.

Residual Error—The difference between the observed or measured and predicted distress and IRI values (e.g., measured minus predicted values). The residuals contain the available information about how well the model predicts the observed distress and IRI.

Standard Error of the Estimate (s_e)—The standard deviation of the residual errors for the pavement sections included in the calibration data set for each prediction model. The standard error is usually obtained by taking the positive square root of the variance of the statistic.

Validation—A systematic process that re-examines the recalibrated model to determine if the desired accuracy exists between the calibrated model and an independent set of observed data. The calibrated model required inputs such as the pavement structure, traffic loading, and environmental data. The simulation model must predict results (e.g., rutting, fatigue cracking) that are reasonably close to those observed in the field. Separate and independent data sets should be used for calibration and validation. Assuming that the calibrated models are successfully validated, the models can then be recalibrated using the two combined data sets without the need for additional validation to provide a better estimate of the residual error.

Verification—Verification of a model examines whether the operational model correctly represents the conceptual model that has been formulated. Verification can be achieved for simple models by comparing the model predictions (e.g., stress) against other analytical solutions for specific cases. Verification can also be accomplished by entering typical materials, structural, environmental, and traffic data into the distress and performance models, and then determining through parameter studies whether the program operates rationally and provides outputs that meet the criterion of engineering reasonableness. If this criterion is not met, the computer code maybe erroneous or the conceptual model may be unsatisfactory. In either case, these problems must be remedied before the model enhancement process continues. No field data are needed in either of the verification approaches described. Verification is primarily intended to confirm the internal consistency or reasonableness of the model. The issue of how well the model predicts reality is addressed during calibration and validation. As such, verification of the MEPDG prediction models is not included within this *Local Calibration Guide*.

2.2 MEPDG Calibration Terms

Calibration Factors—Two calibration factors are used in the MEPDG: global and local. These calibration factors are adjustments applied to the coefficients and/or exponents of the transfer function to eliminate bias between the predicted and measured pavement distresses and IRI. The combination of calibration factors (coefficients and exponents for the different distress prediction equations) can also be used to minimize the standard error of the prediction equation. The standard error of the estimate (s_e) measures the amount of dispersion of the data points around the line of equality between the observed and predicted values. The *MEPDG Manual of Practice* presents the calibration parameters for each distress prediction model included in the MEPDG (AASHTO, 2008).

Damage, Incremental—Incremental damage (ΔDI) is a ratio defined by the actual number of axle load applications (n) for a specified axle load and type within an interval of time divided by the allowable number of axle load applications (N) to some design criteria defined for the same axle load and type for the conditions that exist within the same specific period of time. The incremental damage indices are summed to determine the cumulative damage index over time.

Reliability—The probability that the predicted performance indicator of the trial design will not exceed the design criteria within the design-analysis period. The design reliability (R) is similar, in concept, to that in the current AASHTO *Pavement Design Guide*—the probability that the pavement will not exceed specific failure criteria over the design traffic. For example, a design reliability of 90 percent represents the probability (9 out of 10 projects) that the mean faulting for the project will not exceed the faulting criteria over the analysis period. The reliability of a particular design analyzed by the MEPDG is dependent on the standard errors of the transfer functions. The *User Manual* provides a more complete discussion on reliability and its use in the MEPDG for analyzing a trial design or pavement structure (AASHTO, 2008). However, a reliability level of 50 percent should always be used for predicting distresses to confirm or adjust the local calibration coefficients in accordance with this manual. This is further explained in Step 7 of Section 6 (Step-by-Step Procedure for Local Calibration).

Structural Response Model—*See model, mathmatical.*

Transfer Function—*See model, statistical.*

2.3 Hierarchical Input Level Terms

The hierarchical input level included in the MEPDG is an input scheme that is used to categorize the designer's knowledge of the input parameter. Three levels are available for determining the input values for most of the material and traffic parameters. The *MEPDG Manual of Practice* provides more detailed discussion on the purpose, use, and selection of the hierarchical input level for pavement design (AASHTO, 2008). The following defines each hierarchical input level that can be used by the designer:

Input Level 1—Input parameter is measured directly; it is site- or project-specific. This level represents the greatest knowledge about the input parameter for a specific project but has the highest testing and data collection costs to determine the input value. Level 1 should be used for pavement designs having unusual site features, materials, or traffic conditions that are outside the inference-space used to develop the correlations and defaults included for Input Levels 2 and 3.

Input Level 2—Input parameter is estimated from correlations or regression equations. The input value is calculated from other site specific data or parameters that are less costly and/or easier to measure. Input Level 2 can also represent measured regional values that are not project-specific.

Input Level 3—Input parameter is based on "best-estimated" or default values. Level 3 inputs are based on global or regional default values—the median value from a group of data with similar characteristics. This input level has the least knowledge about the input parameter for the specific project but has the lowest testing and data collection costs.

2.4 Distress or Performance Indicator Terms

This subsection provides a definition of each distress and performance indicator predicted by the MEPDG. It also provides the standard errors of the estimate for each transfer function that are considered reasonable, and are similar to the values included in the *MEPDG Manual of Practice* (AASHTO, 2008). A reasonable standard error of the estimate, however, will be dependent on the design or threshold value used by the agency in their day-to-day management practices.

Hot Mix Asphalt (HMA)-Surfaced Pavements

Alligator Cracking—A form of fatigue or load-related cracking defined as a series of interconnected cracks (characteristically with a "chicken wire/alligator" pattern) that initiate at the bottom of the HMA layers. Alligator cracks initially show up as multiple short, longitudinal, or transverse cracks in the wheel path that become interconnected laterally with continued truck loadings. Alligator cracking is calculated as a percent of total lane area in the MEPDG. The MEPDG does not predict the severity of alligator cracking, but includes low, medium, and high in the definition. A reasonable standard error of the estimate for alligator or bottom-up cracking is seven percent.

Longitudinal Cracking—A form of fatigue or load-related cracking that occurs within the wheel path, defined as cracks parallel to the pavement centerline. Longitudinal cracks initiate at the surface of the HMA pavement and initially show up as short longitudinal cracks that become connected longitudinally with continued truck loadings. Raveling or crack deterioration can occur along the edges of these cracks but they do not form an alligator cracking pattern defined above. The unit of longitudinal cracking calculated by the MEPDG is feet per mile (meters per kilometer). The MEPDG does not predict severity of the longitudinal cracks, but includes low, medium, and high in the definition. A reasonable standard error of the estimate for longitudinal or top-down cracking is 600 ft/mi.

Unless an agency cuts cores or trenches through the HMA surface to confirm where the cracks initiated, it is recommended that the local calibration refinement be confined to total cracking that combines alligator and longitudinal cracks. To combine percent total lane area fatigue cracks with linear or longitudinal fatigue cracks, the total length of longitudinal cracks should be multiplied by 1 ft and that area divided by the total lane area. When an agency decides to combine alligator and longitudinal cracks, the alligator transfer function should be the one used in the local calibration process for determining the local calibration values. If an agency recovers cores or cuts trenches, but cannot determine where the cracks initiated, it is recommended that the agency assume all cracks initiated at the bottom of the HMA layer.

Reflective Cracking—Fatigue cracks in HMA overlays of flexible pavements and of semi-rigid and composite pavements, plus transverse cracks that occur over transverse cracks and joints and cracks in jointed PCC pavements. The unit of reflective cracking calculated by the MEPDG is feet per mi (meters per kilometer). The MEPDG does not predict the severity of reflective cracks but includes low, medium, and high in the definition. Unless an agency cuts cores or trenches through the HMA overlay of flexible pavements to confirm reflective cracks, it is recommended that the local calibration refinement be confined to total cracking of HMA overlays. In this case, all surface cracks in the wheel path (reflective, alligator, and longitudinal cracks) should be combined, using the recommendation for longitudinal cracking listed above. If all cracks are combined, the alligator and reflection cracking transfer functions can be used in the local calibration process.

Rutting or Rut Depth—A longitudinal surface depression in the wheel path resulting from plastic or permanent deformation in each pavement layer. The rut depth is representative of the maximum vertical difference in elevation between the transverse profile of the HMA surface and a wire-line across the lane width. The unit of rutting calculated by the MEPDG is inches (millimeters). A reasonable standard error of the estimate for total rutting is 0.10 in. The MEPDG also computes the rut depths within the HMA, unbound aggregate layers, and foundation. Unless an agency cuts trenches through pavement sections, however, it is recommended that the calibration refinement be confined to the total rut depth predicted with the MEPDG.

Transverse Cracking—Non-wheel load-related cracking that is predominately perpendicular to the pavement centerline and caused by low temperatures or thermal cycling. The unit of transverse cracking calculated by the MEPDG is feet per mile (meters per kilometer) or spacing of transverse cracks in feet. The MEPDG does not predict the severity of transverse cracks but includes low, medium, and high in the definition. A reasonable standard error of the estimate for transverse cracking is 250 ft/mi.

Portland Cement Concrete (PCC)-Surfaced Pavements

Faulting, Mean Transverse Joint (Jointed Plain Concrete Pavement [JPCP])—Transverse joint faulting is the differential elevation across the joint measured approximately 1 ft from the slab edge (longitudinal joint for a conventional lane width), or from the rightmost lane paint stripe for a widened slab. Since joint faulting varies significantly from joint to joint, the mean faulting of all transverse joints in a pavement section is the parameter predicted by the MEPDG. A reasonable standard error of the estimate for faulting is 0.05 in.

Faulting is an important deterioration mechanism of JPCP because of its impact on ride quality. Transverse joint faulting is the result of a combination of repeated applications of moving heavy axle loads, poor load transfer across the joint, free moisture beneath the PCC slab, erosion of the supporting base/subbase, subgrade, or shoulder base material, and upward curling of the slab.

Punchouts, Continuously Reinforced Concrete Pavement (CRCP)—When truck axles pass along near the longitudinal edge of the slab between two closely spaced transverse cracks, a high-tensile stress occurs at the top of the slab, some distance from the edge (typically 48 in. from the edge), transversely across the pavement. This stress increases greatly when there is loss of load transfer across the transverse cracks or loss of support along the edge of the slab. Repeated loading of heavy axles results in fatigue damage at the top of the slab, which results first in micro-cracks that initiate at the transverse crack and propagate longitudinally across the slab to the other transverse crack resulting in a punchout. The punchouts in CRCP are predicted considering the loss of crack load transfer efficiency (LTE) and erosion along the edge of the slab over the design life, and the effects of permanent and transitory moisture and temperature gradients. The transverse crack width is the most critical factor affecting LTE and, therefore, punchout development. Only medium- and high-severity punchouts, as defined by LTPP (FHWA, 2003), are included in the MEPDG model global calibration. A reasonable standard error of the estimate for the number of punchouts is four per mile.

Transverse Cracking, Bottom-Up (JPCP)—When the truck axles are near the longitudinal edge of the slab, midway between the transverse joints, a critical tensile bending stress occurs at the bottom of the slab under the wheel load. This stress increases greatly when there is a high positive temperature gradient through the slab (the top of the slab is warmer than the bottom of the slab). Repeated loadings of heavy axles under those conditions result in fatigue damage along the bottom edge of the slab, which eventually result in a transverse crack that propagates to the surface of the pavement. A reasonable standard error of the estimate for total transverse cracking or total percent slabs cracked is seven percent. The MEPDG predicts the total percent slabs cracked which includes both bottom-up and top-down cracking of JPCP.

Transverse Cracking, Top-Down (JPCP)—Repeated loading by heavy truck tractors with certain axle spacing when the pavement is exposed to high negative temperature gradients (the top of the slab cooler than the bottom of the slab) result in fatigue damage at the top of the slab. This stress eventually results in a transverse or diagonal crack that is initiated on the surface of the pavement. The critical wheel loading condition for top-down cracking involves a combination of axles that loads the opposite ends of a slab simultaneously. In the presence of a high negative temperature gradient, such load combinations cause a high-tensile stress at the top of the slab near the critical pavement edge. This type of loading is most often produced by the combination of steering and drive axles of truck tractors and other vehicles with similar axle spacing. Multiple trailers with relatively short trailer-to-trailer axle spacing are the other source of critical loadings for top-down cracking.

3.0 Significance and Use

Predicting pavement distress is a very complex process that involves uncertainty, variability, and approximations of all input parameters. Mechanistic concepts do provide a more rational and realistic methodology for accounting for variations and approximations, but all prediction models have errors associated with them. The overall error is termed the standard error of the estimate (s_e).

The goal of any calibration-validation process is to confirm that the prediction model can predict, without bias, pavement distress and smoothness, and to determine the standard error associated with the prediction equations. The standard error is used to establish confidence intervals for the prediction model which is used in the design reliability procedure. The standard error estimates the scatter of the data around the line of equality between the predicted and observed values of distress.

All prediction models in the MEPDG were globally calibrated using a representative sample of pavement test sites around North America. Most of these test sites are included in the LTPP program and were used because of the consistency in the monitored data over time and the diversity of test sections spread throughout North America. Policies on pavement preservation and maintenance, construction and material specifications, and materials, however, vary across the United States and can significantly affect distress and performance. These factors are not considered directly in the MEDPG, but can be considered indirectly through the local calibration parameters included in the MEPDG determined through local calibration.

This guide can be used to determine if local policies and practices result in a significant bias in the predicted values, and to recalibrate the MEPDG to local conditions and materials that were not considered in the global calibration process. Eliminating any significant bias and decreasing the standard error of the estimate will reduce construction costs at the same reliability level.

The local calibration process only relates to the transfer functions or statistical models (refer to Figure 1-1). The supporting, mathematical models within the MEPDG simulation model are assumed to be accurate and a correct simulation of real-world conditions. These supporting models are used to compute specific parameters that are needed to predict pavement distress, and include the Integrated Climatic Model (ICM), the structural response model (JULEA for flexible pavements and ISLAB200 for rigid pavements), and time-dependent material property models (strength-gain model of PCC and the age-hardening model for HMA). The time-dependent material property models are statistical models, but were assumed to be mathematical models in the global calibration process. Any error resulting from inaccuracies in the supporting statistical models, however, will be translated into lack-of-fit or model errors of the transfer functions.

4.0 Defining Accuracy of MEPDG Prediction Models

This section provides an overview of the general calibration and validation process of mechanistic-based simulation models for pavement design.

4.1 Calibration

The primary objective of model calibration is to reduce bias. A biased model will consistently produce either over-designed or under-designed pavements, both of which have important cost consequences. The secondary objective of calibration is to increase precision of the model predictions. A model that lacks precision is undesirable because it leads to inconsistency in design effectiveness, including some premature failures. As part of the calibration process, predicted distress is compared against measured distress and appropriate calibration adjustment factors are applied to eliminate significant bias and maximize precision in the model predictions.

In model calibration, a fitting process produces model constants that are evaluated based on goodness-of-fit criteria to decide on the best set of values for the coefficients of the statistical model formulated. The methods of evaluation are either: 1) an analytical process for models that suggest a linear relationship, or 2) the use of numerical optimization for models that suggest a non-linear relationship. The analytical calibration is based on the method of least squares using multiple regression analysis, stepwise regression analysis, principal components analysis, and/or principal component regression analysis. *NCHRP Results Digest #283* provides limited discussion on calibration and the use of different analytical and statistical techniques to reduce bias and determine the standard error for a particular transfer function (NCHRP, 2003).

Numerical optimization, consisting of unconstrained minimization techniques, can be defined as rank ordered or positive pattern search methods. Rank ordered or positive pattern search methods are not equivalent, but both will generally find an optimal set of calibration parameters for a specific set of site conditions and design features. Rank ordered based search methods consider the potential for the interrelationship between different calibration parameters of the transfer function. Pattern search methods are more commonly used, and rely on the steep descent or ascent procedure of nonlinear minimization and are gradient related. The step length or size of the gradients is important in the pattern search techniques to find the global minimum, rather than a localized minimum of mulitple calibration parameters. Numerical optimization methods, however, require a larger number of sections or observations and many more runs of the MEPDG to deterimine the set of calibration parameters that result in the lowest global error of each transfer function. For example, numerical optimization can require more than four times the number of runs needed for the analytical process.

Use of the analytical process within a constrained area or set of boundary conditions should provide reasonable results for the MEPDG transfer functions considering the measurement errors for each of the distresses and performance indicators predicted by the MEPDG. Measurement error and other components of the standard error of the estimate term are discussed in the next section of this document.

Two different calibration approaches may be required depending upon the nature of the distress being predicted through the transfer function. One approach is used for those models that directly calculate the magnitude of the surface distress, while the other approach is used for those models that calculate the incremental damage index rather than the actual distress magnitude. Both are briefly defined below.

Computation of Actual Distress Magnitude from Pavement Response. The term calibration refers to the mathematical process by which the difference between an observed result (e.g., the measured mean rut depth in a pavement section) and a predicted result (e.g., predicted mean rut depth from a permanent deformation model) is reduced to a minimum value for all available sections. The pavement response parameter is used to compute the incremental distress in a direct relationship. This fitting of the predicted to the observed results is most often accomplished by minimizing some function of the differences between the predicted and observed results (normally written as ε_i) by modifying the values of empirical parameters that are part of the model. These empirical parameters are necessary to compensate for the model simplification and limitations in simulating actual pavement behavior and distress development. This difference term $\varepsilon_i(b)$ can be defined by Eq. 4-1.

$$\varepsilon_i(b) = y_i - n(x_i; b) \tag{4-1}$$

Where y_i is the ith observed response and $n(x_i; b)$ is the ith predicted response, the x_i are the independent variables that govern the predicted response, and b represents the calibration parameters or coefficients that are chosen such that the predicted responses are as close as possible to the observed responses.

For example, the MEPDG permanent deformation model directly predicts the magnitude of the actual pavement distress, the rutting measured at the pavement surface. The difference $\varepsilon_i(\theta)$ between the field rutting measurements and the model rutting predictions can be defined as $\varepsilon_i(\theta) = y_i - \theta$ $(x_i; \theta)$, where y_i is the ith observed response and η $(x_i; \theta)$ is the ith predicted response, the x_i are the independent variables that govern the predicted response, and θ represents the calibration parameters or coefficients that are chosen such that the predicted responses are as close as possible to the observed responses; i.e., that minimize $\varepsilon_i(\theta)$ in some overall sense.

Computation of Incremental Damage from Pavement Response. The incremental damage index is computed using a mathematical process describing the development of the distress in terms of accumulated damage. The pavement response is used to compute damage and damage is then correlated to the observed amount of distress. The field test sections were used to adjust or relate the cumulative incremental damage computations to the actual distress measured along the test sections at different points in time. Thus, the calibration proceeds by regressing the damage indices to the actual observations of distress. This approach determines one calibration or transfer function, which can include various site and pavement design features.

For example, the MEPDG model for fatigue cracking is based on an incremental damage index rather than the actual distress magnitude; i.e., the area of cracking. In this case, the empirical calibration coefficients attempt to relate measured cracked pavement area (the actual field distress) to the cumulative damage values (i.e., the model predictions).

Data collected from field test sections are used with both approaches to establish calibration coefficients such that the standard error is minimized between the predicted response (n) and the observed response (y).

4.2 Validation

The objective of model validation is to demonstrate that the calibrated model can produce robust and accurate predictions of pavement distress for cases other than those used for model calibration. Validation typically requires an additional and independent set of in-service pavement performance data. Successful model validation requires that the bias and precision statistics of the model when applied to the validation data set are similar to those obtained from model calibration. The purpose of validation is to determine whether the calibrated conceptual model is a reasonable representation of the real-world system, and if the desired accuracy or correspondence exists between the model and the real-world system.

The success of the validation process can be gauged based on the biases in predicted values and the standard error of estimate, s_e. The s_e for the validation may not be equal to the s_e for calibration; generally, it is higher. To test if it is significantly higher at a given level of significance, which would suggest that the validation failed, a chi-square test is typically used. Conversely, an operational definition of "reasonable correlation" is that the null hypothesis (of equality) is accepted when the paired t-test is used to compare the observed and predicted responses at a confidence interval of 95 percent ($\alpha = 0.05$).

The validation factorials should be designed to statistically test the null hypothesis for each performance indicator. The null hypothesis is that the predicted distress is not statistically different from the actual measured distress. If the null hypothesis is true, then the error is determined using all data for that distress type (both the calibration and validation data sets). If the validation process results in the rejection of the null hypothesis at the chosen significance level, the soundness and completeness of the conceptual and the operational models must be re-evaluated. Further changes in the models require another round of calibration and validation to assure that the revised models are sufficiently accurate.

The benefit of a stringent, independent test on the accuracy of the calibrated model far outweighs the increased costs associated with obtaining two independent data sets. Thus, the split sample approach is typically used in the calibration and validation of statistical and simulation models. A typical split of a sample is 80/20 with 80 percent of the data used in calibration and 20 percent used for verification (chosen randomly of course).

4.3 General Approach to Local Calibration-Validation

Two approaches can be used to improve on the accuracy of the prediction model, including the MEPDG global calibration coefficients for local conditions, policies, and materials. These two approaches are termed the split-sample approach, which is the traditional approach used and discussed in Subsection 4.3.1, and an alternative procedure called jack-knifing.

4.3.1 Traditional Approach—Split-Sample

Some type of organized subdivision of pavement conditions is usually employed for model development, calibration, and validation, because of the wide range of materials (e.g., natural subgrades, local aggregates), truck traffic levels, and environments (e.g., temperature ranges, rainfall levels) for which pavements must be designed. Typically, each cell within an experimental matrix would contain several field sites for which in-service pavement performance data exist for use in model calibration and validation. Alternatively, the same underlying performance model can be used for all cells with each cell being calibrated separately using a portion of the field sites for the cell, with the remaining field sites reserved for subsequent model validation.

Unfortunately, the most common procedure for model development is to use all of the data for calibrating the coefficients and to then take the resulting goodness-of-fit statistics (e.g., the correlation coefficient) as indicators of the prediction accuracy of the model. The calibration statistics consequently only reflect the accuracy of the model for regenerating the calibration data and may not accurately reflect prediction accuracy over the full population. This procedure ignores proper model validation and may produce misleading results unless the size of the calibration data set is exceedingly large, which is rarely the case for pavement performance data.

Those who recognize that calibration goodness-of-fit may not be a good indicator of prediction accuracy have often used split-sample testing for model validation. In the traditional application of split-sample testing, a portion of the data (typically half or more) is used for calibrating the coefficients while the remainder is used to validate accuracy. While split-sample calibration and validation is an improvement over no validation, it has some of the same limitations and can produce a misleading indication of model accuracy for small sample sizes (refer to Note 1). For small sample sizes, the jack-knifing approach is recommended as an alternate to the traditional split-sample approach.

Note 1—A small sample size is a relative term that depends on the number of factors included in the sampling template. In summary, a small sample size is defined as a partial factorial with less than 25 percent of the cells filled with a project but without replication.

4.3.2 Jack-Knife Testing—An Experimental Approach to Refine Model Validation

Jack-knifing is a procedure that provides more reliable assessments of model prediction accuracy than either traditional split-sample validation or the use of the calibration goodness-of-fit statistics. Jack-knifing provides goodness-of-fit statistics that are based on predictions, unlike the calibration statistics that depend on the data used for fitting the model parameters. Thus, the model validation statistics are developed independently of the data used for calibration. Multiple jack-knifing is used to assess the sensitivity of the validation goodness-of-fit statistics to sample size.

To develop jack-knife statistics from a sample of n sets of measured values, the data matrix is divided into two groups, one part for calibration and the other for prediction. These sets are selected at random. Assume that the data matrix includes measurements of p predictor variables X_{ij}, j=1…p and a single criterion variable Y_i, with i=1…n sets of measured values. Thus, the data matrix has n rows and p+1 columns.

For an n-1 jack-knife validation, the procedure begins by removing one set of measurements from the data matrix and calibrating the model with the remaining n-1 sets of measurements. The kth set of measurements that was withheld is then used to predict the criterion variable Y_k from which the standard error (e_1) is computed as the difference between the predicted (Y_k) and measured (Y_k) values of the criterion variable. A second set of measurements is removed while replacing the first set, and the new n-1 set is used to calibrate a new model. This new calibrated model is then used with the withheld set of X values to predict Y and compute the standard error, e_2.

The process of withholding, calibrating, and predicting is repeated until all n sets have been used for prediction. This yields n values of the standard error, from which the jack-knifing goodness-of-fit statistics can be computed. While both the calibration statistics based on all n sets and the jack-knifing statistics are computed from n measures of the error, the jack-knifed errors are computed from measured X values that were not used in calibrating the model coefficients. Thus, the jack-knifing goodness-of-fit statistics are considered to be independent measures of model accuracy.

Because sample sizes of most pavement engineering data sets are limited, one objective of model validation is to assess the sensitivity, or stability, of the accuracy of the model to sample size. To assess the stability of the jack-knifed goodness-of-fit statistics, multiple jack-knifing can be performed by withholding two sets of X, while calibrating on the remaining n-2 sets. Two errors are computed for each calibration based on the n-2 withheld sets of X. For small samples, the goodness-of-fit statistics for the n-2 jack-knifing may be quite different from those for the n-1 jack-knifing. If however the n-1 and n-2 jack-knifing goodness-of-fit statistics are similar, this indicates that the n-1 jack-knifing statistics are not sensitive to the sample size and the statistics are stable. Stable statistics are reliable indicators of goodness-of-fit or prediction accuracy.

The primary advantage of jack-knifing is that the goodness-of-fit statistics are based on predictions from data that are independent of the calibration data. Thus, they more likely indicate the accuracy of future predictions than the statistics based on calibration of all n data vectors. The use of multiple jack-knifing to assess the stability of the prediction statistics is a second advantage of jack-knifing. A third advantage is that the method is easy to apply.

Split-sample validation differs from jack-knifing in that the goodness-of-fit statistics for both calibration and prediction are based on n/2 values (for symmetric split sampling the usual case) rather than n values. Traditional split-sample validation has the distinct disadvantage that, if n is small relative to the inference space being simulated, then n/2 is even smaller, which produces inaccurate calibrations, inaccurate coefficients, and less reliable prediction accuracy.

To overcome in part this deficiency, a method was proposed as part of the NCHRP Project 9-30 experimental plan that combines jack-knifing and split-sample testing (NCHRP, 2003b). It is essentially an $n/2$ jack-knifing scheme and will be termed split-sample jack-knifing. Split-sample jack-knifing provides somewhat better measures of prediction accuracy than the traditional split-sample validation. This approach and process is recommended for the local calibration-validation process because the sample size for local calibration will probably be much smaller than used for the global calibration process.

5.0 Components of the Standard Error of the Estimate

The standard error of the estimate of a prediction model is an important factor that must be understood and quantified in making a decision on whether to try and increase the precision of a simulation model. This section defines and describes the four major components of variance, which are listed below and in Table 5-1.

1. Measurement error.
2. Input error.
3. Model or lack-of-fit error.
4. Pure error.

These components of the total standard error can be mathematically expressed by the sum of the error variances, as shown by Eq. 5-1, assuming that no correlation exists between the contributing errors to the overall error.

$$(V_{Total})^2 = (V_m)^2 + (V_{Input})^2 + (V_{Pure})^2 \tag{5-1}$$

where:

V_{Total} = Total variance of the residual error of prediction associated with the "actual" versus "predicted" performance quantity, sometimes referred to as the calibration error variance,

V_{Input} = Portion of the total variance caused by variations in laboratory and field measurements to estimate the model inputs,

V_m = Portion of the total variation caused by inaccuracies in measuring the distress along the test section used in the calibration process,

V_{Purel} = Portion of the total variance due to replication, referred to as pure error, and

V_l = Portion of the total variance caused by inadequate theory, algorithms, and/or an incorrect model form, typically referred to as lack-of-fit or model variance.

Table 5-1. Summary of Major Components of Calibration Error Dependency

Components of Calibration Error	Error Is:		
	Distress/IRI Dependent	**Input Level Dependent**	**Prediction Model Dependent**
Measurement Error	Yes	No	No
Input Error	No	Yes	No
Model Error	No	No	Yes
Pure Error	Yes	Yes	Yes

This section discusses the importance of the error terms resulting from the calibration of the MEPDG distress prediction equations or transfer functions. The separation and quantification of the sources of variability is important when refining the calibration-validation process to reduce the total standard error. As an example, decreasing the input error of pavement material properties will have little effect on reducing the total standard error or uncertainty of the predictions for the condition if the majority of the total error is caused by measurement error of the distress observations. Most likely, one would want to implement the calibration refinement that has the greatest effect on reducing the total error for the least cost. Each of the four error components is discussed in the following sections.

5.1 Distress/IRI Measurement Error

Errors associated with measuring the distress quantity and IRI for a pavement section are defined as measurement errors. For example, the mean rut depth measured along a project is not the true value, but an estimate of the true mean of the population or test section at a particular point in time. The greater the number of measurements within a test section, the lower the potential difference between the sample and population means and the lower the measurement error. Measurement errors are dependent on the performance measure being calibrated, but are independent of the input level and prediction equation (refer to Table 5-1). The variance in the measured value is composed of different parts, which are listed below.

- V_{mr}—The variance in measurement at a point determined by taking multiple readings at the same location. This component of the measurement variance decreases as the number of repeat measurements increase at the same point on the roadway.

- V_{ms}—The variance in taking a measurement at the same test point (location) but at different times, or the expected difference in taking readings at the same point of the pavement's surface.

- V_{mv}—The variance in taking measurements along the project or the inherent error—sample versus population mean. This component of the measurement error decreases as the number of point measurements increase along a roadway segment.

All of these variances are assumed to be independent and can be added together to determine the total variance in the measured value. These variances are also constant in developing and/or comparing different prediction models using the same database. In general, the sample and test components of the measurement variance are small relative to the inherent variance along the project. However, the measured error or variance for each distress must be representative of how the data were used in the calibration-validation process. For example, if smoothed and cleaned data were used, then the variance associated with the smoothed and cleaned data must be used with the calibration error or variance. The measured data should be used to determine the true bias and standard error of the estimate. The use of smoothed data is not recommended. The bias and precision components (repeatability error) of the distress measured within LTPP were evaluated and documented by Rada, et al. (Rada, 1999).

5.2 Estimated Input Error

The errors associated with estimating each input parameter needed to predict the performance indicator and used in the calibration process are defined as input errors. For example, the mean asphalt content by volume is a required HMA property for predicting mean rut depth or fatigue cracking over time. The mean asphalt content by volume for each HMA layer within an LTPP test section is determined by averaging no more than two test values for that HMA layer. The mean asphalt content of these two tests is not the true mean value for that layer, it is only an estimate based on the results from two tests. The input error is dependent on the material property (or input level, because the input level defines the material properties and parameters to be used) required to predict the performance measure, and is independent of distress type and prediction equation.

The input variance component is composed of three basic parts; testing error, sampling error, and inherent variation of the material along a project. All of these variance components are independent and additive to determine the total input error component. In general, the test and sample error parts of this component are small relative to the inherent material variance or error along the project or test section.

5.3 Model or Lack-of-Fit Error

The inability of a model to predict the actual or true value of the performance measure due to deficiencies in the transfer function or inappropriate assumptions included in the mathematical model, and its inability to model real world conditions, are defined as model or lack-of-fit errors. This type of error is a result of inappropriate assumptions, model simplicity, or inadequate damage algorithm including functional form for the predicted performance indicator. For example, assuming uniform contact pressure under the wheel loads on a flexible pavement is not reality—it is a simplification of reality. Lack-of-fit errors are dependent on the prediction and response (mathematical) models and are independent of the input level and performance measure.

5.4 Pure Error

The random or normal variation between distress values of supposedly “identical” roadway segments is defined as pure error (refer to Note 2). Pure error is dependent on the input level, distress type, and prediction equation. Pure error is difficult to quantify unless replicate test sections within each experimental cell are included in the calibration factorial, which was not the case for the MEPDG distress prediction models.

Note 2—Identical roadway segments are usually restricted to test sections built under highly controlled conditions—test tracks or test pads placed at accelerated pavement testing (APT) facilities. Identical as used in this context is a relative term, which implies that multiple projects or roadway segments with similar experimental factors fall within the same cell and have the same range of secondary properties or factors. Replicate roadway segments do not necessarily have to have the same properties to be considered identical.

6.0 Step-by-Step Procedure for Local Calibration

This section lists and defines the steps that are suggested for calibrating the MEPDG to local conditions, policies, and materials. These steps are shown in the form of a flow chart in Figures 6-1 and 6-2.

Step 1—Select Hierarchical Input Level for Each Input Parameter

The first step in the local calibration process is to select the hierarchical input level for the inputs that will be used by an agency for pavement design and analysis. This step will likely be a policy decision, influenced by the agency's current field and laboratory testing capabilities, material and construction specifications, and traffic data collection procedures and equipment. The *MEPDG Manual of Practice* provides recommendations on selecting the hierarchical input level for each input parameter (AASHTO, 2008).

Selecting the hierarchical input level can be important because decisions made in this step may have a significant impact on the final standard error of each distress prediction model, which affects material quality requirements and construction costs. If the input error only has a minor effect on the standard error of the estimate for the transfer function, the input level may only have a minor effect on the design and the construction costs. (Refer to discussion in Section 5, Components of the Standard Error of the Estimate.) The highest level of input data available from the LTPP database were used to determine the inputs for the global recalibration effort under NCHRP Project 1-40D (NCHRP, 2006), and resulting standard error of the estimate. Agencies will probably elect to use different input levels for some of the input parameters. The bias and standard error of the estimate should be determined for the input levels that will be typically used by an agency for pavement design.

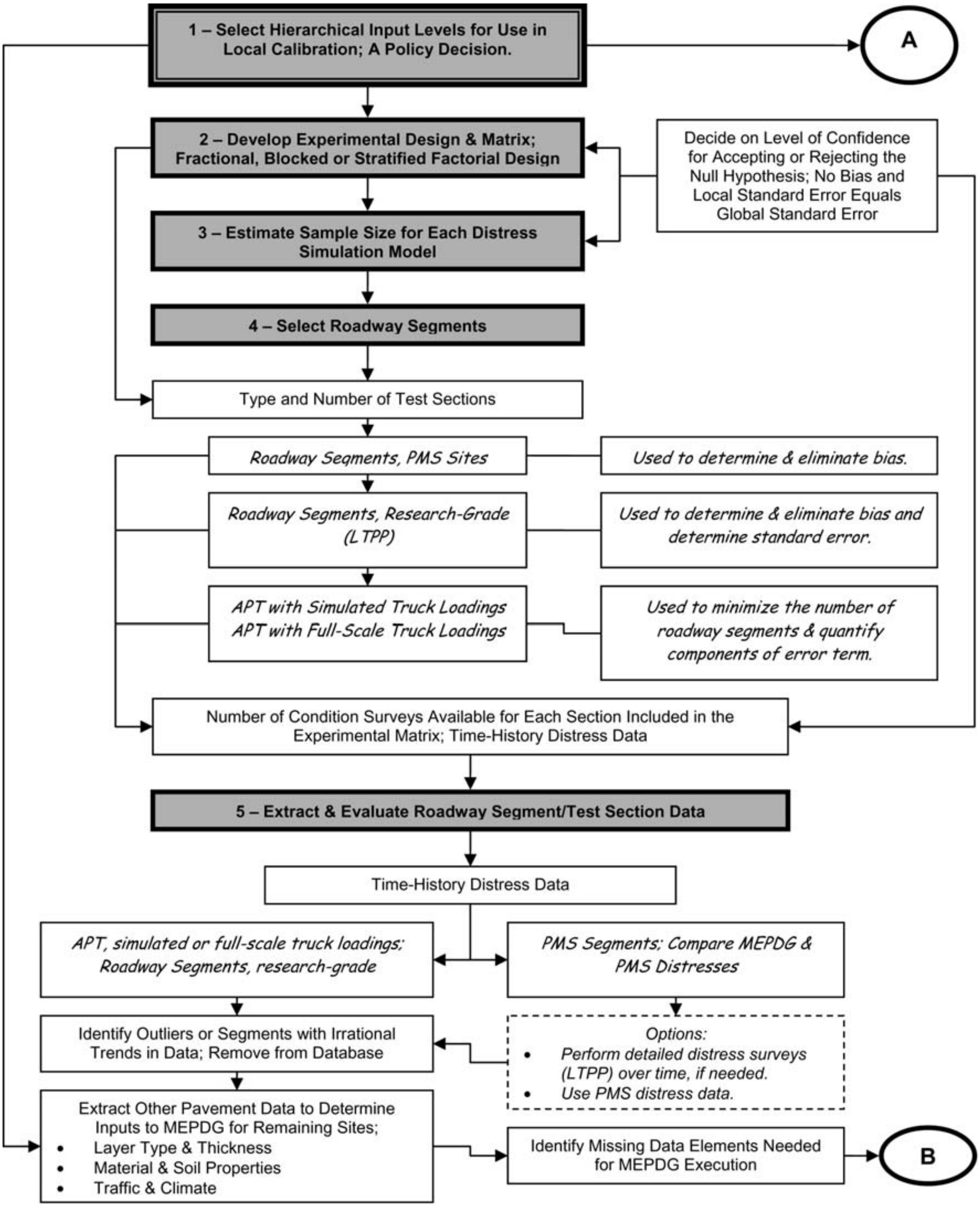


Figure 6-1. Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 1 Through 5

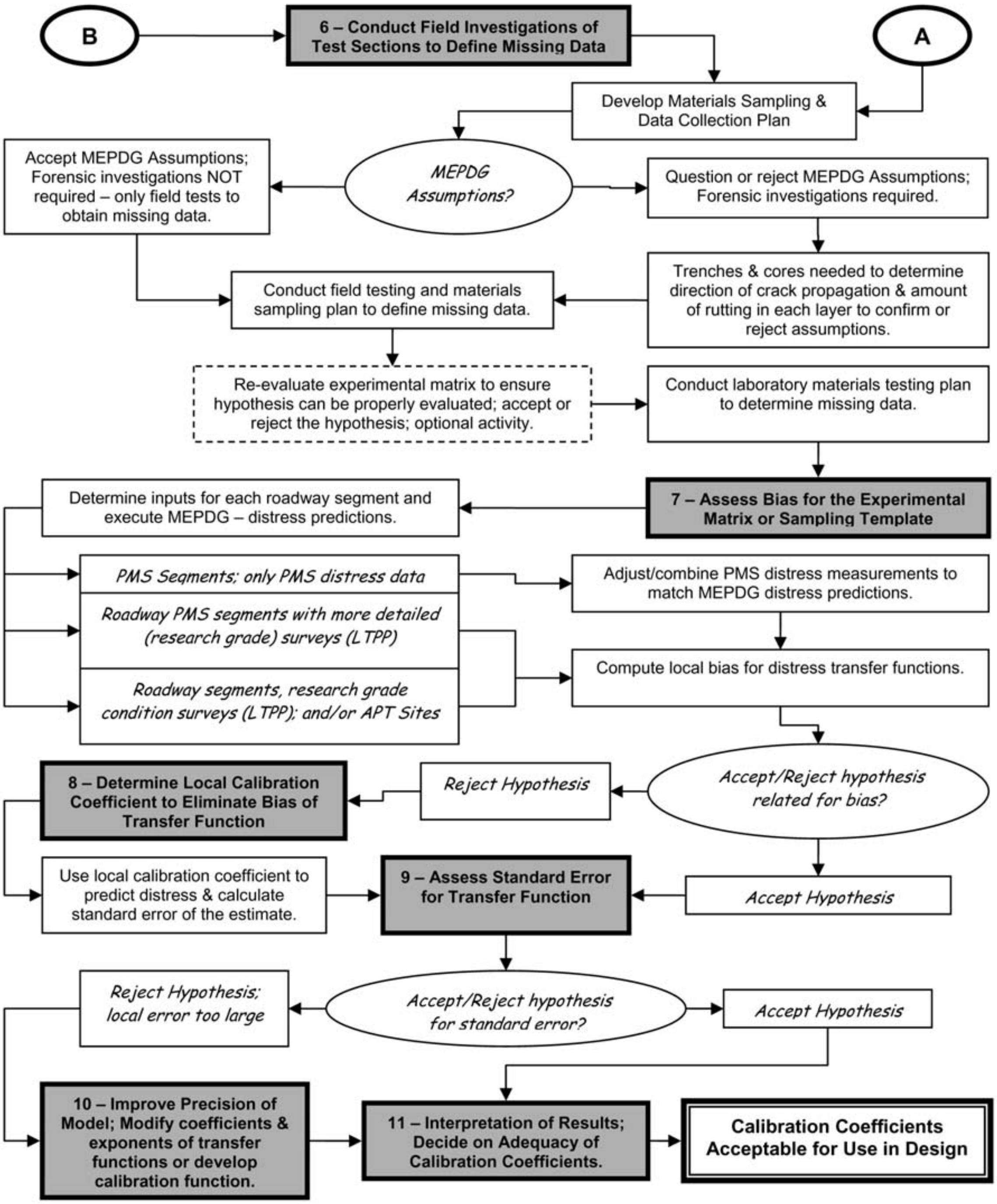


Figure 6-2. Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 6 Through 11

Step 2—Develop Local Experimental Plan and Sampling Template

The second step is to develop a detailed, statistically sound experimental plan or sampling template to refine the calibration of the MEPDG distress and IRI prediction models based on local conditions, policies, and materials, if required. The local or regional calibration factorial for each distress simulation model should be designed to accomplish three objectives:

1. Determine whether there is any local bias in the MEPDG distress predictions or simulation model.
2. Establish the cause of any bias, if it is found through the local validation process.
3. Determine the local calibration coefficients or function for each distress and IRI prediction models.

The primary and secondary tiers within the sampling template should be based on the agency's standard practice and specifications for the more common new construction and rehabilitation strategies. The primary tier parameters in the sampling template should be distress dependent, and likely include pavement type, surface layer type and thickness, and subgrade soil type. Each column in the matrix should represent the effect of changing structure (layer thickness), mixtures, and foundation soils. The secondary tier parameters should include climate (temperature), traffic, and other design features that are pavement type dependent. These are considered secondary parameters because traffic is probably interrelated to surface thickness and climate is interrelated to asphalt binder grade. In some areas, climate may need to be included as a primary tier parameter because of large variations in climate within the same region (e.g., the western part of the United States that includes mountains and plains).

The sampling template should be designed as a fractional factorial matrix as much as possible. Not all cells will likely be filled with or without replicate roadway segments. The matrix should be a balanced design that can be blocked for specific design features or site conditions for each type of pavement and distress. Blocking the fractional factorial will determine whether the bias and standard error of the transfer function is dependent on any of the primary tier parameters of the matrix. This type of matrix is recommended because the experiment needs to evaluate the effect of pavement type and local conditions and materials on reducing the bias and standard error term. When this is completed, answer the question: "Is the bias and/or standard error of the transfer function dependent on type of soil, surface mixture properties, climate, etc.?"

Most cells within the sampling template or matrix should contain two replicate projects to provide an estimate of the pure error.

Step 3—Estimate Sample Size for Specific Distress Prediction Models
The numbers in each cell of the sampling template should include replicate projects, as noted in Step 2. This step is used to estimate the sample size or number of roadway segments to confirm the adequacy of the global calibration coefficients and determine the local calibration coefficients for a specific distress prediction model, if needed.

Both bias and precision are important, thus the number of model evaluations (i.e., the sample size) needed to properly validate the prediction model is evaluated for both bias and precision. The bias is the average residual error; therefore, the confidence interval on the mean can be used to relate the sample size and the bias. Letting e_t be the tolerable bias, the confidence interval on the mean yields the following expression:

$$N \geq \left(\frac{ts_y}{e_t}\right)^2 \qquad (6\text{-}1)$$

Where s_y is the standard deviation of the "true" values of *Y* and *t* is based on *n*-1 degrees of freedom. For accuracy, the standard error of estimate will be used. Since the square of s_e is a variance, the confidence interval on the variance can be used to show the relationship between sample size and the relative error variance (s_e/s_y). The basic equation for the confidence interval is:

$$\sigma^2 \geq (n-1)\frac{s^2}{x_a^2} \tag{6-2}$$

Where the chi-square statistic is *n*-1 degrees of freedom and level of significance, α. Inserting s_e and s_y yields:

$$\frac{s_e}{s_y} \geq \left[\frac{x_\alpha^2}{n-1}\right]^{0.5} \tag{6-3}$$

By selecting a level of significance, the relative deviation s_e/s_y can be determined for a selected sample size. Three levels of significance can be used in estimating the sample size for each distress: 75, 90, and 95 percent. A level of significance of 90 percent is suggested as a practical level in determining the sample size to be used in the experiment.

The same test sections can and should be used for all distresses, because of the coupling effect between different distresses. The MEPDG assumes that all distress transfer functions are uncoupled (distress occurrence and magnitude is independent of the other distresses), with the exception of the IRI regression equation or statistical relationships. The following provides guidance for the minimum number of total test sections for each distress.

- Distortion (Total Rutting or Faulting)—20 roadway segments
- Load-Related Cracking—30 roadway segments
- Non-Load-Related Cracking—26 roadway segments
- Reflection Cracking (HMA surfaces only)—26 roadway segments

The bias and precision components of the distresses reported within the LTPP program were used in estimating the number of test sections listed above (Rada, 1999). If repeatedability errors are unavailable to the agency or user, those reported within the LTPP program for each distress may be used.

Step 4—Select Roadway Segments

This step is used to select roadway projects to obtain maximum benefit of existing information and data to keep sampling and field testing costs to a minimum.

As noted above, replicate projects should be included in specific cells of the sampling template for a specific distress. One of the replicate test sections can come from one of the other distress matrix or factorials. Test sections for refining the validation process will be used for multiple purposes or distresses for efficiency. For example, the test sections that exhibit relatively high amounts of fatigue cracking can be used in the rutting matrix or thermal cracking matrix for low magnitudes of those distresses.

Although three types of experimental test sections can be used in the local validation-calibration refinement plan, roadway segments or long-term field experiments need to be used to determine the standard error of the estimate for each distress simulation model. The three types of experimental test sections are listed and defined below.

1. Long-term, full-scale roadway segments or test sections should be used to fully validate and calibrate the distress prediction models and confirm the superposition of the environmental, aging, and wheel-load effects on the predictions of distress. All of the test sections included in the NCHRP Project 1-40D calibration efforts were from this category of pavement test sites (NCHRP, 2006). Long-term roadway segments can be grouped into two types—those that are PMS segments and those that are research-grade roadway segments (e.g., LTPP sites). Although most of the roadway segments included in the NCHRP Project 1-40D were LTPP test sections (research grade sites), it is expected that many agencies will use PMS segments for judging the adequacy of the global calibration coefficient to their local policies, conditions, and materials. Both types of roadway segments are discussed further in Step 7, and are recommended for use in determining the standard error of the estimate for all distress prediction models.

2. Accelerated Pavement Testing (APT) pads with simulated truck loadings can be used for rapid verification of the form of the distress growth models (transfer function) and selected factor effects on the occurrence of distress. Results from APT test pads are independent of climatic-related factors and time-dependent properties of the pavement materials, so fewer tests are needed to determine the effect of selected factors. APT sites can be used to supplement the roadway segments used in the local calibration process, but should not be used to determine the standard error of the estimate. Use of APT test pads will result in much lower standard errors of the estimate, because traffic and climate parameters are highly controlled and time-dependent properties are excluded from these short-term loading conditions. Thus, APT test pads should only be used to determine bias and to quantify the variance components of the distress prediction model.

3. APT experiments with full-scale truck loadings (test tracks) can be used to calibrate and validate the effects of wheel load on the distress predictions without the added complexity of long-term aging and extensive environmental variations. Results from this type of experiment are slightly dependent on the climatic factors and time-dependent material properties. Use of these full-scale APT experiments will also result in lower standard errors of the estimate, because many of the factors are controlled. Results from full-scale APT experiments should be used to supplement and reduce the number of roadway segments required for local calibration in determining bias and to quantify the error term components of the distress prediction mode.

A listing of some factors that should be considered in selecting roadway segments for use in the local validation-calibration refinement plan includes the following:

- Roadway segments should be selected with the fewest number of structural layers and materials (e.g., one PCC layer, one or two HMA layers, one unbound base layer, and one subbase layer) to reduce the amount of testing and input required for material characterizations. These roadway segments, however, need to include the types of new construction and rehabilitation strategies typically used or specified by the agency. The roadway segments used to define the standard error of the estimate should include the range of materials and soils that are common to an area or region and the physical condition of those materials and soils.

- Roadway segments with and without overlays are needed for the validation-calibration sampling template. Those segments that have detailed time-history distress data prior to and after rehabilitation should be given a higher priority for use in the experiment because these segments can serve in dual roles as both new construction and rehabilitated pavements.

- Roadway segments that include non-conventional mixtures or layers should be included in the experimental plan to ensure that the model forms and calibration factors are representative of these mixtures. Non-conventional mixtures can include: stone matrix asphalt (SMA), polymer modified asphalt (PMA), open-graded drainage layers, cement-aggregate mixtures, and high-strength PCC mixtures. Many of the LTPP test sections included in the NCHRP Project 1-37A calibration factorial were built with conventional HMA and PCC mixtures. The flexible sections excluded open-graded drainage, SMA, and PMA layers. There were numerous sections with open-graded mixtures in the JPCP sections. The *MEPDG Manual of Practice* provides a more completed listing on the limitations of the MEPDG and those design features not considered directly by the MEPDG (AASHTO, 2008).

It is recommended that at least three condition surveys be available for each roadway segment to estimate the incremental increase in distress over time. The interval between the distress measurements should be similar between all of the test sections. It is also suggested that this time-history distress data represent at least a 10-year period, if available. This time period will ensure that all time-dependent material properties and the occurrence of distress are properly taken into account in the determination of any bias and the standard error of the estimate. If available, repeat condition surveys should be planned for those roadway segments that exhibit higher levels of distress to reduce the inherent variability of distress measurements and estimate the measurement error for a particular distress. A similar number of observations per age, per project should be considered in selecting roadway segments for the sampling template.

Step 5—Extract and Evaluate Distress and Project Data

This step of the local calibration process is to collect all data and identify any missing data elements that are needed to execute the MEPDG. All data should be entered into a calibration database, similar to the one that was developed under NCHRP Project 9-30 for flexible pavements (NCHRP, 2003a), or at least filed for future reference. This step is grouped into four activities, as discussed in the following paragraphs.

Step 5.1

The first activity under Step 5 is to extract, review, and convert the measured distress data into the values predicted by the MEPDG, if needed. It is imperative that a consistent definition and measurement protocol of surface distress be used throughout any calibration-validation process. If possible, all flexible pavement distress data should be measured in accordance with AASHTO R 55 *Standard Practice for Quantifying Cracks in Asphalt Pavement Surface* and AASHTO R 48 *Standard Practice for Determining Rut Depth in Pavements* or the FHWA LTPP publication *Data Collection Guide for Long-Term Pavement Performance.* All rigid pavement distress data should be measured in accordance with the FHWA LTPP publication *Data Collection Guide for Long-Term Pavement Performance* and the AASHTO R 36 S*tandard Practice for Evaluating Faulting of Concrete Pavements.* Distress measurements should be made to ensure consistency with the MEPDG predictions of distress and smootheness (FHWA, 2003). Pavement smoothness should be measured in accordance with AASHTO R 57 S*tandard Practice for Operating Inertial Profilers and Evaluating Pavement Profiles.*

Many agencies, however, will need to use their PMS distress data for the local validation-calibration effort, which may differ from LTPP. For this condition, two options are available for use by the agency.

1. The first is to select PMS segments (refer to Step 4) and complete distress surveys in accordance with the *LTPP Distress Identification Manual* (FHWA, 2003). This option is time consuming to collect sufficient distress data for completing the local calibration. Agencies that select this option will need to have at least a five-year implementation plan in place for the MEPDG to ensure a minimum of three observations per project. Few agencies are expected to select this option.

   Distress surveys can be completed on the selected PMS segments within one season to reduce the time, but time-history distress data will be unavailable for any one PMS segment. This single-distress point is not suggested for use, because the incremental increase in distress over time will not be included in the evaluation of bias and in determining of standard error of the estimate for the distress prediction models. The lack-of-fit error will not be well-defined for this option.

2. The second, and probably preferred, option is to select PMS segments and use the PMS condition survey data in the local validation-calibration effort. This option requires less time and cost, but the PMS measurements may need to be adjusted or modified to be consistent with the MEPDG distress predictions (refer to Step 7).

Step 5.2
The second activity under this step is to compare the maximum measured distress values to the trigger values or design criteria used by the agency for each distress. The average maximum distress values from the sampling template should exceed 50 percent of the design criteria, as a minimum. This consideration becomes important when evaluating the bias and standard error terms of the prediction model under Steps 7 and 9, respectively. If the maximum distress values are significantly lower than the agency's design criteria for that distress (less than 50 percent of the design criteria), the accuracy and bias of the transfer function may not be well defined at the values that trigger major rehabilitation.

Step 5.3
The measured distress data for all roadway segments should be evaluated and checked for anomalies and outliers—observations that have irrational trends in the distress data. This evaluation can be limited to visual inspection of the data over time to ensure that the distress data are reasonable, or include a detailed statistical comparison of the performance data. Multiple distress surveys and profile measurements are used to establish the performance trends for each roadway segment. Any segment with irrational trends in the distress data should be considered for removal from the local calibration database. As a minimum, the following two questions should be asked in evaluating the measured distress data.

1. Does the data make sense within and between each roadway segment? Obviously, any zeros that represent non-entry values should be removed from the local validation-calibration database. Distress data that return to zero values within the measurement period may indicate some type of maintenance or rehabilitation activity.

- Measurements taken after structural rehabilitation should be removed from the database or the observation period should end prior to the rehabilitation activity.

- Distress values that are zero as a result of some maintenance or pavement preservation activity, which is a part of the agency's management policy, should be removed but future distress observation values after that activity should be used.

2. Are there roadway segments with anomalies, outliers, or blunders in the data? If the outliers or anomalies can be explained and are a result of some non-typical condition, they should be removed. If the outliers or anomalies cannot be explained, they should remain in the database.

For the roadway segments that remain, all data should be extracted for use in determining the required inputs for the hierarchical input levels selected (refer to Step 1). Data sources that will likely be used by at least some agencies to determine the MEPDG inputs are construction records, acceptance tests in a quality assurance (QA) program, and as-built construction plans. Use of QA and historical data provides overall project or lot averages that can be different from the layer properties of individual PMS segments. The difference between PMS segments and lot average values will increase the input error component of the total standard error term (see Section 5).

Each agency will need to consider these sources of errors and make a judgment decision on whether to increase the effort and costs in determining the inputs to the MEPDG for local calibration. The final standard error of the estimate for each distress simulation model will impact this judgment decision in the long-term because of its effect on construction costs.

Any missing or questionable data to determine the MEPDG inputs should be identified. The missing or questionable data elements should be determined through field investigations. For the PMS segments selected, the falling weight deflectometer (FWD) deflection basin and other field tests should be performed to confirm layer thickness and estimate the in-place modulus values for each structural layer. FWD testing should be performed by AASHTO T 256 *Standard Method of Test for Pavement Deflection Measurements.* The *MEPDG Manual of Practice* includes recommends for a field test program for pavement evaluation and rehabilitation studies.

Step 6—Conduct Field and Forensic Investigations

Step 6.1
The first activity of this step is to develop a materials sampling and testing plan to determine any missing data element or to validate some key inputs for the roadway segments selected. The *MEPDG Manual of Practice* provides recommended guidelines for field investigations.

The pavement materials should be recovered and tested in accordance with the agency's standard practice that is used during pavement evaluation for rehabilitation design. The *MEPDG Manual of Practice* does provide recommendations for both field and laboratory testing for measuring the layer properties in accordance with the hierarchical input level selected. If the agency's standards differ from the *MEPDG Manual of Practice*, the agency's standards should be followed because those are the ones that will be used for day-to-day new pavement and rehabilitation designs.

Step 6.2

As part of this step and any field investigation, an agency needs to decide whether forensic investigations are required to confirm the assumptions embedded in the MEPDG. As an example, the portion of total rutting measured at the surface that can be assigned to each pavement layer and the location of where cracks initiated (top-down versus bottom-up, load-related cracking). Two options are available that have a significant impact on costs and time to conduct any field investigation.

If the agency elects to accept the MEPDG assumptions for layer rutting and the location of crack initiation, no forensic investigations are required. For this option, the agency should restrict the local calibration to total rut depth and total load related cracking—combining longitudinal and alligator cracks within the wheel path (refer to Section 2.4).

If the agency rejects or questions the MEPDG assumptions under the first option, then trenches and cores will be needed to measure the rut depths within each pavement layer and estimate the direction of crack propagation. This option will likely require additional roadway segments and/or APT sections for confirming or adjusting the local calibration values for rutting and fatigue cracking. It should be noted that no trenches and cores were taken under NCHRP Projects 1-37A and 1-40D to verify and confirm the amount of rutting in each pavement layer, as well as where the cracks initiated or the direction of crack propagation.

Trenches or test pits are recommended so that individual pavement layer rutting can be measured, but are only needed for projects that have exhibited levels of rutting greater than 0.35 in. at the surface. It is difficult at best to measure the permanent deformation in subsurface layers for rut depths shallower than 0.35 in.

To determine the percentage of load-related cracks that start at the top of the pavement and propagate downward (as opposed to the classical assumption of bottom-up cracking), 6-in. diameter cores can be drilled directly on top of load-related cracks and extracted to observe the depth of the crack. Crack initiation and propagation direction should be reported for each core. Crack width at the initiation point and depth of crack from the initiation point should also be reported. If the crack extends completely through the HMA layers and it is impossible to determine the direction of crack propagation or where the cracks initiated, it is recommended that the agency assume that the cracks initiated at the bottom of the HMA layers.

Step 6.3

Prior to going to Step 7, the number of roadway segments remaining with all data needed to execute the MEPDG should be re-evaluated to ensure that a sufficient number of segments are available for the local validation-calibration effort. If too many of the roadway segments have been removed for one reason or the other, additional segments may need to be added to the sampling template.

Step 7—Assess Local Bias: Validation of Global Calibration Values to Local Conditions, Policies, and Materials

The MEPDG and global calibration values should be used to calculate the performance indicators for each roadway segment (new pavement and rehabilitation strategies). The predicted values are compared to the measured values to determine bias and the standard error of the estimate to validate each distress prediction model for local conditions, policies, specifications, and materials. The distresses predicted by the MEPDG for calibration puposes should be based on average values for each input parameter. In assessing and eliminating the bias, if needed, the predicted distresses at a 50 percent reliability level should always be used. In other words, the average input values and the distresses at a 50 percent reliability level should be used within this step, as well as within Steps 8, 9, and 10.

Step 7.1

The bias and standard error of the estimate should be determined for this full set of data for each distress simulation model. Compare the predictions for each performance indicator to the measurements (or adjusted observations; refer to Step 5), and compute the residual errors, bias, and standard error of the estimate for each distress prediction model. A plot of the predicted values and measured data should be prepared to compare the general location of the data points to the line of equality.

Step 7.2

Evaluate the null hypothesis for the sampling template or experimental factorial (refer to Steps 2 and 3). The null hypothesis for this initial assessment is that there is no bias or no systematic difference between the measured and predicted values of distress. The null hypothesis should be evaluated for the entire sampling template and individual blocks within the sampling template. A paired *t*-test can be used to determine if there is a significant difference between the sets of predicted and measured distress and IRI values. The null hypothesis is as follows:

$$H_0 : \sum\left(y_{Measured} - x_{Predicted}\right) = 0 \tag{6-4}$$

where:

$y_{Measured}$ = Measured value, and
$x_{Predicted}$ = Predicted value using the model.

It is possible that the above hypothesis could be accepted (the sum of the residual errors are indifferent to zero), but the model still be biased. Two other model parameters (termed intercept and slope estimators) should be used to fully evaluate model bias using the following fitted regression model between the measured (*y*) and predicted (*x*) values, as well as the variability in the measured value associated with the distributed errors for each predicted value.

$$\hat{y}_i = b_o + m\left(x_i\right) \tag{6-5}$$

This regression model is used to provide estimators of the mean measured values ($\hat{y}_i$). The intercept (b_o) and slope (m) are used in hypothesis testing as follows:

H_o: $b_o = 0$
H_o: $m = 1.0$

In summary:

- If any of the null hypothesis is rejected, the specific distress prediction model should be recalibrated to the local conditions and materials—proceed to Step 8. The results from the three hypothesis tests can be used to make decisions during the re-calibration process.

- If the null hypothesis is accepted (no bias), the standard error of the estimate for the local data set should be compared to the global calibration data set—proceed to Step 9. The global standard errors are provided in the "Tools" section of the MEPDG software for each distress and in the *MEPDG Manual of Practice*.

Step 8—Eliminate Local Bias of Distress and IRI Prediction Models
The process used to eliminate the bias found to be significant from using the global calibration values depends on the cause of the bias and accuracy desired by the agency. In general, there are three possibilities, which are listed below.

1. The residual errors are, for the most part, always positive or negative with a low standard error of the estimate in comparison to the trigger value, and the slope of the residual errors versus predicted values is relatively constant and close to zero. The precision of the prediction model is reasonable but the accuracy is poor (large bias). In this case, the local calibration coefficient is used to reduce the bias. This condition generally requires the least level of effort and the fewest number of runs or iterations of the MEPDG to reduce the bias.

2. The bias is low and relatively constant with time or number of loading cycles, but the residual errors have a wide dispersion varying from positive to negative values. The accuracy of the prediction model is reasonable, but the precision is poor. In this case, the coefficient of the prediction equation is used to reduce the bias but the value of the local calibration coefficient is probably dependent on some site feature, material property, and/or design feature included in the sampling template. This condition generally requires more runs and a higher level of effort to reduce the bias.

3. The residual errors versus the predicted values exhibit a significant and variable slope that appears to be dependent on the predicted value. The precision of the prediction model is poor and the accuracy is time or number of loading cycles dependent—there is poor correlation between the predicted and measured values. This condition is the most difficult to evaluate because the exponent of the number of loading cycles needs to be considered. This condition also requires the highest level of effort and many more runs to reduce the bias.

The agency needs to first decide on whether to use the agency specific values or the local calibration parameters that are considered as inputs in the MEPDG software. Either one can be used with success. The following provides general guidance.

Compute the bias within each block of the sampling template (refer to Section 5) to determine whether the local bias is dependent on any primary or secondary tier parameter of the sampling template. Results from this analysis of local bias can be used to make revisions to specific calibration parameters to eliminate the local bias.

Adjust the local calibration values (agency specific values) for the distress transfer functions to eliminate the bias. Tables 6-1 and 6-2 list the local calibration parameters of the MEPDG transfer functions or distress and IRI prediction models that should be considered for revising the predictions to eliminate bias for flexible and rigid pavements, respectively. These tables are provided for guidance only in eliminating any local bias in the predictions. In addition, the local calibration values could be dependent on site factors, layer parameters, or policies established by the agency. If the local calibration values (agency specific values) are found to be dependent on some site factor, design feature, or material property, those types of adjustments or corrections need to be made external to the MEPDG.

Table 6-1. Recommendation for the Flexible Pavement Transfer Function Calibration Parameters to Be Adjusted for Eliminating Bias and Reducing the Standard Error

Distress		Eliminate Bias	Reduce Standard Error
Total Rutting	Unbound Materials and HMA Layers	k_{r1}, β_{s1}, or β_{r1}	kr_2, kr_3, and β_{r2}, β_{r3}
Load-Related Cracking	Alligator Cracking	C_2 or k_{f1}	k_{f2}, k_{f3}, and C_1
	Longitudinal Cracking	C_2 or k_{f1}	k_{f2}, k_{f3}, and C_1
	Semi-Rigid Pavements	C_2 or β_{c1}	C_1, C_2, C_4
Non-Load-Related Cracking	Transverse Cracking	β_{t3}	β_{t3}
IRI		C_4	C_1, C_2, C_3

Table 6-2. Recommendation for the Rigid Pavement Transfer Function Calibration Coefficients to Be Adjusted for Eliminating Bias and Reducing the Standard Error

Distress		Eliminate Bias	Reduce Standard Error
Faulting		C_1	C_1
Fatigue Cracking		C_1 or C_4	$C_{2,}$ C_5
CRCP Punchouts	Fatigue	C_1	C_2
	Punchouts	C_3	$C_{4,}$ C_5
	Crack Widths	C_6	C_6
IRI	JPCP	C_4	C_1
	CRPC	C_4	C_1, C_2

After the bias has been eliminated, compute the standard error of the estimate using the local calibration values (or agency specific values) based on local conditions; compare that standard error to the global standard error reported under NCHRP Project 1-40D and included in Section 5 of the *MEPDG Manual of Practice* (AASHTO, 2008)—proceed to Step 9.

Step 9—Assess the Standard Error of the Estimate

Compare the standard error determined from the sampling template to the standard error derived from the global data set, which are included in Section 5 of the *MEPDG Manual of Practice* for each transfer function (AASHTO, 2008). Reasonable values of the standard error for each distress transfer function are included in Subsection 2.4 (Distress or Performance Indicator Terms) of this document. The standard errors of the estimate for the IRI regression equations for different pavement types are also included in Section 5 of the *MEPDG Manual of Practice.* A reasonable standard error of the estimate for the IRI is 17 in./mi.

Evaluate the null hypothesis for the sampling template relative to the standard error (refer to Step 6). The null hypothesis for this initial assessment is that there is no significant difference between the standard error for the global and local calibration efforts at the selected level of confidence.

- If the null hypothesis is rejected, there is a significant difference between the standard error terms resulting from use of the global and local calibration values.
  - If the local calibration has a higher standard error term, it is recommended that the distress simulation model be recalibrated in an attempt to lower the standard error—proceed to Step 10. The agency can decide, however, to just accept the higher standard error or default standard error determined from the original calibration process using LTPP test sections. If this is the case, proceed to Step 11.
  - If the local calibration has a lower standard error term, these calibration coefficients are recommended for use—proceed to Step 11.
- If the null hypothesis is accepted, the local and global standard errors are considered the same; these calibration coefficients can be used for pavement design—proceed to Step 11.

Step 10—Reduce Standard Error of the Estimate
If the user decides that the standard error is too large, resulting in overly conservative designs at higher reliability levels, revisions to the local calibration values of the transfer function or statistical model may be needed. This step can be complicated and will probably require external revisions to the local calibration parameters or agency specific values to improve on the prediction model's precision. The following provides some general guidance to accomplish this step.

Step 10.1
Prior to the recalibration or modification process to local conditions, the standard error components should be quantified to estimate the potential reduction in the total standard error term (refer to Section 5). The lack-of-fit or model error is the only portion of the total standard error that can be reduced through the local calibration process, after the hierarchical input level has been selected (refer to Step 1). The measurement error should be quantified and compared to the total error to estimate the potential increase in precision of the prediction model. The measurement error is probably the larger of the error components and making changes to the local calibration (or agency specific) values will not change the magnitude of that error component. The agency needs to decide whether additional costs and effort will significantly reduce the total standard error of the specific distress and IRI prediction models.

- If it is expected that the total standard error cannot be significantly reduced, proceed to Step 11.
- On the other hand, if it is expected that the model precision can be significantly improved, continue with this step.

Step 10.2
Compute the standard error within each block of the sampling template (refer to Section 5) to determine whether the local standard error term is dependent on any primary or secondary tier parameter of the matrix. Results from the analysis of local standard errors within each block can be used to make revisions to specific local calibration parameters.

Step 10.3
Adjust the local calibration values (agency specific values) of the distress transfer functions to reduce the standard error of the recalibration data set. Tables 6-1 and 6-2 list the coefficients of the MEPDG

transfer functions or distress and IRI prediction models that should be considered for revising the predictions to minimize the standard error for flexible and rigid pavements, respectively.

A fitting process of the model constants is evaluated based on a goodness-of-fit criteria on the best set of values for the coefficients of the model. The methods of evaluation make use of either the analytical process for models that suggest linear relationship or make use of numerical optimization for models that suggest non-linear relationship. The analytical approach is based on least squares using multiple regression analysis, stepwise regression analysis, principal components analysis, and/or principal component regression analysis. The numerical optimization includes methods such as the steepest descent or pattern search.

These local calibration values that result in the lowest standard error should be used for pavement design—proceed to Step 11.

Step 11—Interpretation of Results, Deciding on Adequacy of Calibration Parameters
Step 11.1
The local standard error of the estimate for each distress and IRI prediction model should be evaluated to determine the impact on the resulting designs at different reliability levels. The sampling template can be used to determine the design life of typical site features and pavement structures or rehabilitation strategies for different reliability levels. An agency should review the expected pavement/rehabilitation design life within each cell of the sampling template. The agency now has three options to consider.

1. The expected design life is believed to be "reasonable" for the reliability levels used by the agency—not resulting in overly conservative designs based on historical data. For this option or condition, proceed to the last activity; Step 11.2. To define reasonable expected design life, survivability or probability of failure curves should be prepared from the PMS data and compared to the calculated reliability of each segment included in the sampling template.

2. The expected design life is believed to be "too short" for the reliability levels used by the agency, resulting in very conservative and costly designs. The agency should try to reduce the standard error of the estimate for the specific distress simulation model—proceed back to Step 10 for reducing the standard error of the estimate.

3. The expected design life is believed to be "too short" for the reliability levels used by an agency, because the measurement and pure error components are too large relative to the lack-of-fit and input error components. Therefore, making model revisions, adding more validation-calibration roadway segments, using Level 1 input parameters, completing field forensic investigations, and other costly field activities are expected to have a minor impact on the total standard error of the estimate. For this condition, the agency should consider increasing the failure criteria or trigger values for new pavement and rehabilitation designs and proceed to Step 11.2.

Step 11.2
The local calibration values and new standard error of the estimate should be entered into the MEPDG software for use in new pavement and rehabilitation designs.

7.0 Referenced Documents and Standards

7.1 Referenced Documents

AASHTO. *Mechanistic-Empirical Pavement Design Guide—A Manual of Practice.* Publication code: MEPDG-1, ISBN: 978-1-56051-423-7. American Association of State Highway and Transportation Officials, Washington, DC, Interim Edition, July 2008.

FHWA. *Distress Identification Manual for Long-Term Pavement Performance Program* (Fourth Revised Edition). Publication No. FHWA-RD-03-031, Federal Highway Administration, Washington, DC, 2003.

NCHRP. *Refining the Calibration and Validation of Hot Mix Asphalt Performance Models: An Experimental Plan and Database*. NCHRP Results Digest Number 284, National Cooperative Highway Research Program, Transportation Research Board of the National Academies, Washington, DC, December 2003a.

NCHRP, *Jack-Knife Testing—An Experimental Approach to Refine Model Calibration and Validation*, NCHRP Results Digest Number 283, National Cooperative Highway Research Program, Transportation Research Board of the National Academies, Washington, DC, December 2003b.

NCHRP. Changes to the Mechanistic-Empirical Pavement Design Guide Software Through Version 0.900. *NCHRP Research Results Digest 308*, National Cooperative Highway Research Program, Transportation Research Board of the National Academies, Washington, DC, September 2006.

NCHRP. *Version 1.0—Mechanistic-Empirical Pavement Design Guide Software*. National Cooperative Highway Research Program, National Academy of Sciences, Washington, DC, April 2007.

Rada, G. R., et al. *Study of LTPP Distress Variability,* Volume 1. Publication No. FHWA-RD-99-074, Federal Highway Administration, Office of Infrastructure Research and Development, McLean, Virginia, September, 1999.

7.2 Test Protocols and Standards

AASHTO R 55 *Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces*

AASHTO R 57 *Standard Practice for Operating Inertial Profilers and Evaluating Pavement Profiles*

AASHTO R 36 *Standard Practice for Evaluating Faulting of Concrete Pavements*

AASHTO R 48 *Standard Practice for Determining Rut Depth in Pavements*

AASHTO T 256 *Standard Method for Pavement Deflection Measurements*

List of Figures

Attachments

List of Tables

Appendix: Examples and Demonstrations for Local Calibration

A1. Background

All performance indicator prediction models in the *Mechanistic-Empirical Pavement Design Guide* (MEPDG) were calibrated to observed field performance from a representative sample of pavement test sites located throughout North America. These models are defined as being globally calibrated. Data from the Long-Term Pavement Performance (LTPP) test sections were used extensively in the global calibration process, because of their consistency in the monitored data over time and the diversity of test sections spread throughout North America. Other experimental test sections, such as MnRoad and Vandalia, were also included in the global calibration process.

Policies on pavement preservation and maintenance, construction and material specifications, and design features vary across the United States and are not considered directly in the MEDPG. These factors can be considered indirectly through the local or agency specific calibration coefficients included in the MEPDG, if found to cause bias in the performance predictions. The purpose of this appendix is to provide examples using the *Local Calibration Guide* for validating and/or revising the MEPDG global calibration factors to account for local conditions and materials that were not considered in the global calibration process.

It is impossible to cover all conditions and scenarios that user agencies may encounter in validating the acceptability of the global prediction models. The objective of this appendix is to provide a listing of the steps, with examples of judgments and decision-making criteria, in deciding whether to accept the global calibration factors or develop local calibration or agency specific factors in predicting the performance indicators.

This appendix is divided into three parts, including this Background. Appendix Two (A2) is focused on flexible pavements and common rehabilitation strategies for flexible pavements (hot mix asphalt [HMA] overlays), while Appendix Three (A3) is focused on rigid pavements (specifically, new jointed plain concrete pavements). All demonstrations were focused on validating and/or revising the global calibration coefficients for the common design strategy used by two agencies (the Kansas and Missouri Departments of Transportation).

Each pavement type uses two data sources for demonstrating the local validation-calibration effort—segments within a pavement management system (PMS) and LTPP test sections. Both data sets are used to demonstrate the steps and potential differences between measures of distress that are consistent and inconsistent with the definitions used during the global calibration effort under the NCHRP studies. These two different sets for the flexible pavements and HMA overlays were also used to demonstrate differences in the quality and quantity of the input data.

The PMS segments used in the examples include those from the Kansas Department of Transportation (KSDOT) for flexible pavements and the Missouri Department of Transportation (MODOT) for rigid pavements. The KSDOT segments are focused on their common design strategy or full-depth flexible pavements and HMA overlays, while the MODOT segments are focused on plain jointed concrete pavements (JPCP). The LTPP SPS-1 and SPS-5 experiments are used for flexible pavements, while the SPS-2 experiment is used for rigid pavements.

Another objective of this appendix is to demonstrate how data from different sources (PMS and LTPP databases) can be combined to establish regional, as well as local calibration factors for each performance measure. The *Guide for Local Calibration* is used in determining the local calibration values from each data set and pavement type.

A2. New Flexible Pavements and Rehabilitation of Flexible Pavements

A2.1 Demonstration 1—PMS Data and Local Calibration

A2.1.1 Description of PMS Segments

Sixteen projects from the KSDOT PMS database were selected for demonstrating use of the *Local Calibration Guide*. Eleven of the projects were HMA full-depth new construction/reconstruction and six were HMA overlays of flexible pavements. Table A2-1 provides general information about the pavement structures, while Table A2-2 provides project descriptions and their locations. Figure A2-1 shows the general location of each PMS segment. Table A2-3 summarizes the material types and layer thicknesses that were extracted from construction files. The binder types used in the projects included conventional or neat HMA, polymer modified asphalt (PMA), and Superpave mixtures. More detailed descriptions of each project are provided in Attachment A2.4.A.

A2.1.2 Step-by-Step Procedure

The steps included in the *Local Calibration Guide* were followed for this demonstration using PMS data. Many of the decisions made by KSDOT were based on an expedited time frame to collect the necessary data for the demonstration. KSDOT would likely make different decisions given a longer time frame for the local calibration process.

Table A2-1. General Structure Information for the Selected Kansas PMS Projects

No.	Payment Type	Project ID (KSDDOT PMS ID)	Binder Type	Number and Length of Homogenous Sections	Construction and Maintenance and Rehabilitation History	Comments: Initial IRI (in./mi)	Comments: Analysis Period (yr)
1	Full-Depth HMA	FDAC-C-2 (0673047017180)	Conventional HMA Mix	1 section 0.94 mi	1992. Reconstruction. (1" HMA + 8" Recycle HMA + 6" Lime subgrade + subgrade) 2001. Overlay (1.6" HMA overlay)	85	1993–2001
2		FDAC-C-4 (0223120008080)	Conventional HMA Mix	1 section 0.83 mi	1989. Reconstruction. (1.5" HMA + 6.5" HMA + subgrade) 1999. Overlay (1.6 HMA overlay)	110	1990–1999
3		FDAC-P-1 (0552083020210)	PMA Mix	1 section 0.91mi	1999. Reconstruction. (1" HMA + 4" HMA + 8.7" HMA + 4" Lime subgrade + subgrade)	36	2000–2006
4		FDAC-P-2 (0552083026270)	PMA Mix	1 section 0.50 mile	1999. Reconstruction. (1" HMA+4" HMA + 8.7"HMA + 4" Lime subgrade + subgrade)	35	2000–2006
5		FDAC-P-3 (0552083028290)	PMA Mix	2 sections: I: 0.34 mi II: 0.13 mi	Section I: 1999. Reconstruction. (1.0" HMA + 4.0" HMA + 4.7" HMA + 3.5" Lime subgrade + subgrade Section II: 1999. Reconstruction. (1.0" HMA + 4.0" HMA+ 8.7" HMA + 4" Lime subgrade + subgrade	40.5	2000–2006
6		FDAC-P-4 (0552083021220)	PMA Mix	1 section 0.45 mi	1999. Reconstruction. (1" HMA + 4" HMA + 8.7" HMA + 4" Lime subgrade + subgrade)	39.0	2000–2006
7		FDAC-P-5 (0552083023240)	PMA Mix	1 section 0.38mi	1999. reconstruction (1" HMA + 4" HMA + 8.7"HMA + 4" Lime subgrade + subgrade)	37.0	2000–2006
8		FDAC-S-1 (0372054012120)	Superpave	1 section, 0.67 mi	1998. Reconstruction. (1" HMA + 15" HMA + 6" Lime subgrade + subgrade)	44.0	1999–2006
9		FDAC-S-3 (0702075007080)	Superpave Mix	2 sections, same structure—0.84 mi	1998 Reconstruction. (1" HMA + 4" HMA + 6" HMA + 6" Lime subgrade + subgrade)	55.0	1999–2006

Continued on next page

Table A2-1—Continued

No.	Payment Type	Project ID (KSDDOT PMS ID)	Binder Type	Number and Length of Homogenous Sections	Construction and Maintenance and Rehabilitation History	Comments	
						Initial IRI (in./mi)	Analysis Period (yr)
10	Full-Depth HMA	FDAC-S-4 (0632075032330)	Superpave Mix	1 section 0.81mi	1998. Reconstruction. (1" HMA + 12" HMA + 6" Lime subgrade + subgrade)	54.0	1999–2006
11		FDAC-S-5 (0702075008090)	Superpave Mix	1 section 4.2 mi	1998 Reconstruction (1" HMA + 4" HMA + 6" HMA + 6" Lime subgrade + subgrade)	40.0	1999–2006
12	HMA Overlay	HMA_HMA-C-1 (0083177008090)	Conventional HMA Mix	1 section 1.39 mi	• 1979 New construction. (1" HMA + 7" HMA + 8" Lime subgrade + subgrade) • 1989. Overlay (¾" overlay + 1" recycled HMA) • 1997. Overlay (3" HMA Overlay)	70.5	1998–2005
13		HMA_HMA-C-2 (0083177006070)	Conventional HMA Mix	1 section 2.16 mi	• 1979 New construction. (1" HMA + 5" HMA + 8" Lime subgrade + subgrade) • 1989. Overlay (¾" overlay + 1" recycle HMA) • 1997. Overlay (3" HMA Overlay)	61.5	1998–2006
14		HMA_HMA-C-4 (0132160005060)	Conventional HMA Mix	1 section 0.29 mi	• 1964 New construction. (7.5" HMA + subgrade) • 1996. Overlay. (1" + 4.5" HMA overlay)	37.0	1997–2006
15		HMA_HMA-P-3 (0233010007080)	PMA Mix	1 section 1.10 mi	• 1996. Reconstruction. (1" HMA + 8"HMA + 6" Lime subgrade + subgrade) • 2000 Overlay. (1.6" + 2.0" HMA overlay)	49.5	2001–2006
16		HMA_HMA-S-3 (0052281018200)	Superpave Mix	2 sections I: 0.04 mi II: 0.78 mi	Section II: • 1965 New Construction. (1.5" + 3.0" HMA + subgrade) • 1999. Overlay (1.5" HMA + 4.0" recycled HMA) • 2005. Overlay (1" overlay superpave)	53.5	2000–2005

Table A2-2. General Project Information for the Kansas PMS Segments

Project Name	KSDOT PMS ID	Length (mi)	Direction	Route	Begin Milepost	End Milepost	County Number	County Name	District Number
FDAC-C-2	0673047017180	0.94	Northbound	State 47	32.913	33.855	67	Neosho	4
FDAC-C-4	0223120008080	0.83	Eastbound	State 120	8.098	8.933	22	Doniphan	1
FDAC-P-1	0552083020210	0.91	Northbound	US 83	144.713	145.625	55	Logan	3
FDAC-P-2	0552083026270	0.50	Northbound	US 83	150.360	150.857	55	Logan	3
FDAC-P-3 (Section I)	0552083028290	0.34	Northbound	US 83	153.092	153.436	55	Logan	3
FDAC-P-3 (Section II)	0552083028290	0.13	Northbound	US 83	153.436	153.566	55	Logan	3
FDAC-P-4	0552083021220	0.45	Northbound	US 83	145.625	146.075	55	Logan	3
FDAC-P-5	0552083023240	0.38	Northbound	US 83	148.185	148.566	55	Logan	3
FDAC-S-1	0372054012120	0.67	Eastbound	US 54	282.514	283.193	37	Greenwood	4
FDAC-S-3	0702075007080	0.84	Northbound	US 75	120.227	121.068	70	Osage	1
FDAC-S-4	0632075032330	0.81	Northbound	US 75	32.679	33.493	63	Montgomery	4
FDAC-S-5	0702075008090	4.21	Northbound	US 75	121.068	125.275	70	Osage	1
HMA_HMA-C-1	0083177008090	1.39	Northbound	State 177	7.414	8.807	8	Butler	5
HMA_HMA-C-2	0083177006070	2.16	Northbound	State 177	5.255	7.414	8	Butler	5
HMA_HMA-C-4	0132160005060	0.29	Eastbound	US 160	135.261	135.550	13	Clark	6
HMA_HMA-P-3	0233010007080	1.10	Eastbound	State 10	7.332	8.430	23	Douglas	1
HMA_HMA-S-3 (Section I)	0052281018200	0.04	Northbound	US 281	115.497	115.539	5	Barton	5
HMA_HMA-S-3 (Section II)	0052281018200	0.78	Northbound	US 281	115.539	116.315	5	Barton	5

Table A2-3. Material Types and Layer Thickness for trhe Kansas PMS Segments

Project ID	Layer Information			Initial AADTT; Two Way	Climate	
	No.	Material Type	Thickness, in.		Latitude	Longitude
FDAC-P-2	1	Fine-Grained Soil	—	218	39.07	–100.85
	2	Lime Modified Soil	4			
	3	Neat HMA Base Mixture	8.7			
	4	PMA Binder and Wearing Surface	5			
FDAC-P-3	1	Fine-Grained Soil; A-7-6	—	210	39.10	–100.85
	2	Lime Modified Soil	3.5			
	3	Neat HMA Base Mixture	4.7			
	4	PMA Binder and Wearing Surface	5			
FDAC-P-4	1	Fine-Grained Soil; A-7-6	—	148	39.00	–100.85
	2	Lime Modified Soil	4			
	3	Neat HMA Base Mixture	8.7			
	4	PMA Binder and Wearing Surface	5			
FDAC-P-5	1	Fine-Grained Soil; A-7-6	—	156	39.00	–100.85
	2	Lime Modified Soil	4			
	3	Neat HMA Base Mixture	8.7			
		PMA Binder and Wearing Surface	5			
FDAC-S-1	1	Fine-Grained Soil; A-7-6	—	155	37.81	–96.31
	2	Lime Modified Soil	6			
	3	Superpave HMA Mixtures	16			
FDAC-S-3	1	Fine-Grained Soil; A-7-6	—	114	38.52	–95.70
	2	Lime Modified Soil	6			
	3	Superpave HMA Base Mixture	6			
	4	Superpave Binder and Wearing Course	5			
FDAC-S-4	1	Fine-Grained Soil; A-7-6	—	129	37.34	–95.71
	2	Lime Modified Soil	6			
	3	Superpave HMA Mixtures	13			
FDAC-S-5	1	Fine-Grained Soil; A-7-6	—	99	38.54	–95.69
	2	Lime Modified Soil	6			
	3	Superpave Base Mixture	6			
	4	Superpave Binder and Wearing Course	5			
HMA_ HMA-C-1	1	Fine-Grained Soil; A-7-6	—	7	37.92	–96.75
	2	Lime Modified Soil	8			
	3	HMA Mixture—Existing	6.75			
	4	Neat HMA Overlay	4			

Table A2-3—*Continued*

Project ID	Layer Information			Initial AADTT; Two Way	Climate	
	No.	Material Type	Thickness, in.		Latitude	Longitude
HMA_ HMA-C-2	1	Fine-Grained Soil; A-7-6	—	7	37.92	–96.75
	2	Lime Modified Soil	8			
	3	HMA Mixture—Existing	6.75			
	4	Neat HMA Overlay	3			
HMA_ HMA-C-4	1	Fine-Grained Soil; A-7-6	—	64	37.29	–100.01
	2	HMA Mixture—Existing	7.5			
	3	Neat HMA Overlay	5.5			
HMA_ HMA-P-3	1	Fine-Grained Soil; A-7-6	—	100	38.92	–95.29
	2	Lime Modified Soil	6			
	3	HMA Mixture—Existing	9			
	4	Neat HMA Overlay	3.6			
HMA_ HMA-S-3	1	Fine-Grained Soil; A-7-6	—	61	38.52	–98.71
	2	HMA Mixture—Existing (recycled)	5.5			
	3	Neat HMA Overlay	1.5			

Step 1—Select Hierarchical Input Level
The hierarchical input level to be used in the local validation-calibration process should be consistent with the way the agency intends to determine the inputs for day-to-day use. This demonstration using PMS roadway segments is for the condition for which minimum data are available.

Input Levels 2 and 3 were used for all input parameters for the PMS segments—most were input Level 3. Data needed to determine Level 1 inputs were unavailable for the PMS segments. The general information from which the inputs were determined for each input category is discussed in Step 5.

Step 2—Experimental Factorial and Matrix or Sampling Template
Creating a detailed sampling template was not completed within this example, because only two pavement cross sections or design strategies were used within this example; full-depth HMA and HMA overlays of flexible pavements. It was decided that the global calibration values would be validated for the HMA design strategies and materials commonly used in Kansas. Table A2-4 shows the simplified sampling template used for this demonstration, along with the number of PMS segments or projects within each cell (refer to Step 4).

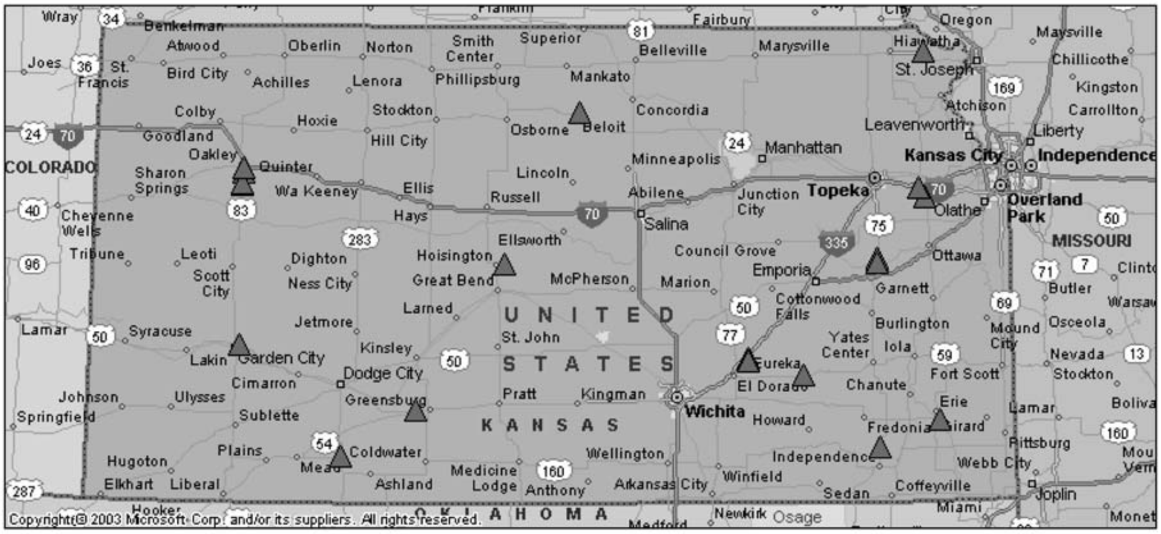


Figure A2-1. General Location of the Roadway Segments Selected for Demonstrating the Local Validation-Calibration Process Using Kansas PMS Data

Table A2-4. Simplified Sampling Template for the Demonstration Using PMS Data

HMA-Mixture Type	Pavement Type		Total PMS Segments
	Full-Depth Reconstruction	HMA Overlays	
Conventional Neat HMA Mixtures	2	3	5
Superpave Mixtures	4	1	5
PMA Mixtures	5	1	6
Total PMS Segments	11	5	16

Extreme climate variations and soil conditions do not occur across Kansas, with the exception of some localized areas. The only other primary tier in the factorial is HMA mixture type—neat HMA, PMA, and Superpave mixtures. KSDOT has adopted Superpave and has been moving towards the use of PMA mixtures to reduce rutting and thermal cracking since the late 1990s. Thus, pavement type (full-depth and HMA overlays) and HMA mixture type were the only two tiers of the sampling template for the PMS segments.

The number of roadway segments selected for the sampling template should result in a balanced factorial with the same number of replicates within each cell. As shown in Table A2-4, the factorial is unbalanced and some cells do not have replication.

Step 3—Estimate Sample Size for Each Performance Indicator Prediction Model

A 90-percent level of significance was used for estimating the sample size (total number of roadway segments or projects). Higher confidence levels can be used, but that will increase the number of segments needed. Table A2-5 summarizes the estimated sample size for each measure of performance. A minimum of four observations per project was assumed. The number of distress observations per segment is dependent on the measurement error or within segment data variability over time (i.e., the higher the within project data dispersion or variability, the larger the number of observations needed for each distress). The number of distress measurements made within a roadway segment is also dependent on the within project variability of the design features and site conditions.

Table A2-5. Estimated Number of PMS Segments Needed for the Local Validation-Calibration Process

Performance Indicator	Design Criteria or Magnitude	Tolerable Bias, e_t	s_e/s_y	Minimum Number of Samples	
				Projects	Observations
Rutting, in.	0.40	0.05	0.50	8	32
Fatigue Cracking, %	20	2.5	0.50	8	32
Thermal Cracking, ft/mi	1,500	300	0.50	18	72
Roughness, in./mi	130	10	0.50	12	48

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E} \right)^2$$

where:
z_α = 1.282 for a 90 percent confidence interval;
s_y = standard deviation of the maximum true or observed values; and
e_t = tolerable bias.

The tolerable bias was estimated from the levels that are expected to trigger some major rehabilitation activity (see Table A2-5), which are agency dependent. The s_e/s_y value (ratio of the standard error and standard deviation of the measured values) will also be agency dependent. A value of 0.50 was selected for this demonstration, which is fairly low for PMS data.

Step 4—Select Roadway Segments

Projects should be selected to cover a range of distress values that are of similar ages within the sampling template. Roadway segments exhibiting premature or accelerated distress levels, as well as those exhibiting superior performance (low levels of distress over long periods of time), can be used, but with caution. The roadway segments selected for the sampling template when using hierarchical input Level 3 should represent average performance conditions.

A limited number of potential PMS segments were available for this example. Many of the PMS segments had insufficient construction histories or insufficient distress data to be included in the local validation-calibration procedure.

Sixteen segments with multiple distress measurements within each segment were selected for this demonstration (see Figure A2-1 and Table A2-2). The 16 segments, however, are believed to be sufficient for the two pavement structures and different surface mixtures considered within the demonstration (see Tables A2-4 and A2-5). It is important that the same number of performance observations per age per project be available in selecting roadway segments for the sampling template. It would not be good practice to have some segments with ten observations over 10 years with other segments having only two or three observations over 10 years. The segments with one observation per year would have a greater influence on the validation-calibration process than the segments with 0.25 observations per year. The number of observations per year for the 16 PMS segments selected vary from 1.0 observation per year for the full-depth reconstruction projects to about 0.75 observations per year for the HMA-Overlay Projects. This range is considered acceptable.

Step 5—Extract and Evaluate Distress and Project Data
This step is grouped into four activities: (1) extracting and reviewing the performance data; (2) comparing the performance indicator magnitudes to the trigger values; (3) evaluating the distress data to identify anomalies and outliers; and (4) determining the inputs to the MEPDG.

Step 5.1—Extract, Review, and Convert Measured Values to the Values Predicted by the MEPDG, if Needed.
First, the distress or performance indicator measurements included in the KSDOT PMS database were reviewed to determine whether the measured values are consistent with the values predicted by the MEPDG. For the KSDOT PMS data, the measured cracking values are different, while the rutting and IRI values are similar and assumed to be the same. The cracking values and how they were used in the local calibration process are defined below.

- **Fatigue Cracking**. KSDOT measures fatigue cracking in number of wheel path feet per 100-ft sample by crack severity, but does not distinguish between alligator cracking and longitudinal cracking in the wheel path. In addition, reflection cracks are not distinguished separately from the other cracking distresses. The PMS data were converted to a percentage value similar to what is reported in the HPMS system from Kansas. In summary, the following equation was used to convert KSDOT cracking measurements to a percentage value that is predicted by the MEPDG.

$$FC = \left(\frac{FCR_1(0.5) + FCR_2(1.0) + FCR_3(1.5) + FCR_4(2.0)}{8.0} \right) \tag{A2-1}$$

All load related cracks are included in one value. Thus, the MEPDG predictions for load related cracking were combined into one value by simply adding the length of longitudinal cracks and reflection cracks for HMA overlays, multiplying by 1.0 ft, dividing that product by the area of the lane and adding that value to the percentage of alligator cracking predicted by the MEPDG.

- **Transverse Cracking**. Another difference is that KSDOT records thermal or transverse cracks as the number of cracks by severity level. The following equation has been used by KSDOT to convert their measured values to the MEPDG predicted value of ft/mi.

$$TC = \left(\frac{TCR_o + TCR_1 + TCR_2 + TCR_3}{(10)(12)(52.8)} \right) \tag{A2-2}$$

The value of 10 in the above equation is needed because the data are stored with an implied decimal. The value of 12 is the typical lane width, and the value of 52.8 coverts from 100-ft sample to a per mi basis. Prior to 1999, KSDOT did not record the number or amount of sealed transverse cracking (TCR_0). As a result, the amount of transverse cracks sometimes goes to "0".

The average measured value should be determined for each measurement period for each PMS segment. The time-history data should not be smoothed but can be cleaned to remove data errors. As an example, measured distresses that are recorded as zero after multiple values of distress have been recorded. Plots of the average time-history data are included in Attachment A2.4.B. Some important observations of the data that have an impact on the validation-calibration process are listed below.

- Large measurement errors are present for all performance indicators. The measured values significantly increase and decrease with time. In addition, there are abrupt changes in the rutting and IRI data over time. Thus, improving on the precision of the prediction model is not likely.
- Few of the PMS segments have any measured fatigue cracking. Thus, confirming the fatigue cracking global calibration values is not likely.
- Rut depths are low for all PMS segments included within this demonstration. Thus, confirming the rut depth global calibration values will be limited to rut depths significantly less than the design criteria or trigger value.

Step 5.2—Compare Distress Magnitudes to Trigger Values

The next activity of this step is to compare the distress magnitudes to the trigger values for each distress. Then answer the question—Does the sampling template include values close to the design criteria or trigger value? Table A2-6 summarizes the average, maximum, and minimum distress values for each performance indicator as compared to the trigger values (design criteria) for some major rehabilitation activity (see Table A2-5).

Table A2-6. Summary of the Maximum Values of Different Performance Indicators in Comparison to the Design Criteria or Trigger Values (Number of Sites = 16)

Distress or Performance Indicator	Design Criteria	Maximum Values Measured for Each Segment				Probability of Exceeding Trigger Value, %
		Average Max. Value	Lowest Max. Value	Largest Max. Value	Stand. Dev. of Max. Values	
Rut Depth, in.	0.40	0.22	0.17	0.36	0.0754	0.8
Fatigue Cracking, %	20	2.7	0	20.8	4.44	0.0
Transverse Cracks, ft/mi	1,500[1]	969	0	4,689	1,237	33.4
Roughness, in./mi	130	87	55	154	26.9	5.5

Note 1: The 1,500 ft/mi corresponds to an average crack spacing of about 40 ft.

As tabulated, most of the observed or measured distress values are significantly less than the design criteria. In fact, the average distress magnitudes are more than two standard deviations below the design criteria, with the exception of transverse cracking. Table A2-6 also summarizes the expected probability that the trigger values will be exceeded using the PMS local calibration data set. As shown, the probability of exceeding the trigger values is low for all performance indicators, with the exception of transverse cracking.

This comparison suggests that the values used as KSDOT's trigger values are too high or the flexible pavements and HMA overlays are being rehabilitated for other reasons. This observation becomes important when evaluating the bias and standard error terms of the prediction models under Steps 7 and 9, respectively. More importantly, the maximum area of fatigue cracking measured is less than 3 percent. This level of fatigue cracking is too low to validate and accurately determine the local calibration values or adjustments for predicting the increase in cracking over time; especially when 20 percent fatigue cracking was selected for the design criteria.

Step 5.3—Evaluate Distress Data to Identify Anomalies and/or Outliers
The distress data for each roadway segment included in the sampling template should be reviewed prior to determining all of the MEPDG input parameters. This evaluation can be limited to visual inspection of the data over time to ensure that the distress data are reasonable—time-history plots or include a detailed statistical comparison of the performance data. As a minimum, the following questions should be asked (Attachment A2.4.B includes graphs that show the distress values over time for the roadway segments).

– Does the data make sense within and between each roadway segment? All of the data extracted from the Kansas PMS segments looked reasonable and appears to represent typical performance characteristics and conditions. Obviously, any zeros that represent non-entry values should be removed from the local validation-calibration database. Distress data that return to zero values within the measurement period may indicate some type of maintenance or rehabilitation activity.

– Measurements taken after structural rehabilitation should be removed from the database or the observation period should end prior to the rehabilitation activity.

– Distress values that are zero as a result of some maintenance or pavement preservation activity, which is a part of the agency's management policy, should be removed but future distress observation values after that activity should be used.

- Are there segments with anomalies, outliers, or blunders in the data? If the outliers or anomalies can be explained and are a result of some non-typical condition, they should be removed. If the outlier or anomaly cannot be explained, they should remain in the database. No outliers or anomalies were found in the KSDOT data. The magnitude of the performance indicators, however, do increase and then decrease exhibiting high within project variability in the measured values (refer to Attachment A2.4.B). The number of measurements per segment was increased from four to a minimum of 10, because of the high within segment variability.

Step 5.4—Inputs to the MEPDG for Each Input Category
The following provides a brief discussion on the information extracted from the KSDOT databases and files for determining the inputs needed to execute the MEPDG for each PMS segment.

- **Initial IRI**—As noted in the above project descriptions, the initial IRI was determined from the measured values within one or two years after construction. This value is believed to be reasonable, because only minor magnitudes of distress were recorded for the first couple of years after construction.

- **Construction Histories, Cross Sections, and Layer Thicknesses**—As-built plans that were available from KSDOT records were used to determine the material types and layer thicknesses for each PMS segment. It was assumed that all layers were placed in accordance with the project specifications. Material properties consistent with the specifications were used for inputs with minimum construction data.

- The construction date for the full-depth pavements and HMA overlays was obtained from the as-built plans or construction database files.

  - For all full-depth reconstruction projects, it was assumed that the first lift of HMA was placed one month following preparation of the subgrade and that date was entered as the construction date.
  - For the HMA-Overlay Projects, the construction date or time that the first overlay lift was placed was assumed to be the start of construction. The construction date of the existing flexible pavement for the HMA-Overlay Projects was taken from the construction database.
  - The opening month to traffic for all projects was assumed to be one month following the HMA placement.

- **Rehabilitation Inputs**—The condition of the HMA pavement prior to overlay was determined from the distress values included in the PMS database prior to overlay. All layers were assumed to be fully bonded.
- **Traffic**—Default values were used for all input with the exception of speed, number of lanes, traffic growth, truck traffic classification groups, and average annual daily truck traffic (AADTT).
  - The posted speed limit was used as the input for all PMS segments.
  - The AADTT was taken from the KSDOT traffic database for each roadway segment.
  - The AADTT or annual equivalent single axle loads (ESALs) included in KSDOT's database were used to estimate the average growth in truck traffic for each PMS segment.
- **Climate**—The longitude and latitude of each PMS segment was used to create a virtual weather station for that segment of roadway. The weather stations in Kansas and adjacent states were used to create the virtual weather stations.
- **Materials**—The material and layer properties for each pavement layer and subgrade were taken from construction records, when available. If adequate data were unavailable, the mean value from the specifications was used or the average value determined for the specific input from other projects with similar material.
  - Dynamic modulus data were unavailable for all HMA mixtures and resilient modulus data were unavailable for all soils. Thus, Level 3 or the default values included in the MEPDG were used in all cases.
  - The creep compliance and indirect tensile strength for all HMA mixtures were unavailable for all projects. Thus, the default values (Input Level 3) were used in all cases.

Step 6—Conduct Field and Forensic Investigations
If the assumptions in the MEPDG are challenged by the agency, forensic investigations are needed to measure the rutting in the individual layers and to determine where the cracks initiated or the direction of crack propagation. Steps 7 through 11 are executed for each specific MEPDG predicted distress after the forensic investigations are completed.

For this demonstration, KSDOT decided to accept the assumptions and conditions included in the MEPDG for the global calibration effort. Thus, no field and forensic investigations were planned or conducted to determine the location of crack initiation and the amount of rutting within each pavement layer and foundation.

Step 7—Assess Local Bias from Global Calibration Factors

The MEPDG was executed using the global calibration values to predict the performance indicators for each PMS segment. The predicted performance measures are shown in Attachment A2.4.B relative to the measured values for the PMS segments. The null hypothesis is first checked for the entire sampling matrix. The null hypothesis is that the average residual error or bias is zero for a specified confidence level or level of significance. A 90-percent confidence level was used in this demonstration.

$$H_O : \sum_{i=1}^{n} \left(y_{Measured} - x_{Predicted} \right)_i = 0 \qquad (A2\text{-}3)$$

Table A2-7 lists the bias for each performance indicator for the entire sampling template and whether the null hypothesis is rejected or accepted, while Figures A2-2 to A2-5 compare the predicted and measured values for each performance indicator. Figures A2-2 and A2-5 also include a comparison between the residual errors (e_r) and predicted values (x_i). This same comparison was excluded from Figures A2-3 and A2-4, because no to nil fatigue and transverse cracks were predicted for most of the PMS segments that exhibited measurable cracks.

Table A2-7. Summary of the Statistical Parameters—Global Calibration Values Used for Predicting Performance Indicators for the Kansas PMS Sections

Performance Indicator	Project	Bias, $e_{r\,(Mean)}$	Standard Error, s_e	s_e/s_y	R^2 (see Note)	Hypothesis; H_o: $y_i - x_i = 0$	Comment
Rutting	New	–0.00412	0.1004	1.06	Poor	Accept; $p = 0.484$	Extensive variability.
	Rehab.	+0.0352	0.0767	1.33	Poor	Accept; $p = 0.269$	Extensive variability; over predicting rut depths.
Fatigue Cracking	New	–0.807	2.65	1.00	Poor	Accept; $p = 0.302$	Limited fatigue cracking was predicted and only a few sections exhibited fatigue cracks. Insufficient number of sections to complete calibration process. No fatigue cracks were predicted for any of the HMA overlays.
	Rehab.	–0.657	1.46	0.948	Poor	Accept; $p = 0.334$	

Table A2-7—*Continued*

Performance Indicator	Project	Bias, $e_{r\,(Mean)}$	Standard Error, s_e	s_e/s_y	R^2 (see Note)	Hypothesis; H_o: $y_i - x_i = 0$	Comment
Transverse Cracking	New	−64.4	702.6	1.40	Poor	Accept; $p = 0.413$	Thermal cracks were predicted for only two PMS segments, while high amounts of transverse cracks were exhibited on multiple segments. No transverse cracks were predicted for any of the HMA overlays.
	Rehab.	−448.0	685.3	1.02	Poor	Accept; $p = 0.252$	
IRI	New	−9.65	26.15	0.927	0.22	Accept; $p = 0.367$	The hypothesis should be checked and evaluated after any bias has been reduced for the other distresses.
	Rehab.	−8.13	27.60	1.66	Poor	Accept; $p = 0.312$	

Note: Poor means that the model did not explain variation in the measured data within and between the PMS segments.
Residual Error = $e_r = y_i - x_i$
y_i = Measured or Observed Value; Standard Deviation of the observed values.
x_i = Predicted Value

As shown in Table A2-7, the hypothesis is accepted for the transfer functions using all of the PMS segments included within the sampling template. The reason that the hypothesis was accepted is that the bias is low in comparison to the mean measured value and within project variability of the measured distress values. Two other model parameters, however, were used to evaluate model bias—the intercept (b_o) and slope (m) estimators using the following fitted linear regression model between the measured (y_i) and predicted (x_i) values.

$$\hat{y}_i = b_o + m(x_i) \qquad (A2\text{-}4)$$

A non-linear regression model could also be considered and used to reduce the standard error between the predicted mean measured values and predicted values. If the predicted mean values ($\hat{y}_i$) fall along the line of equality, $b_o = 0.0$, $m = 1.0$, and $\hat{y}_i = x_i$.

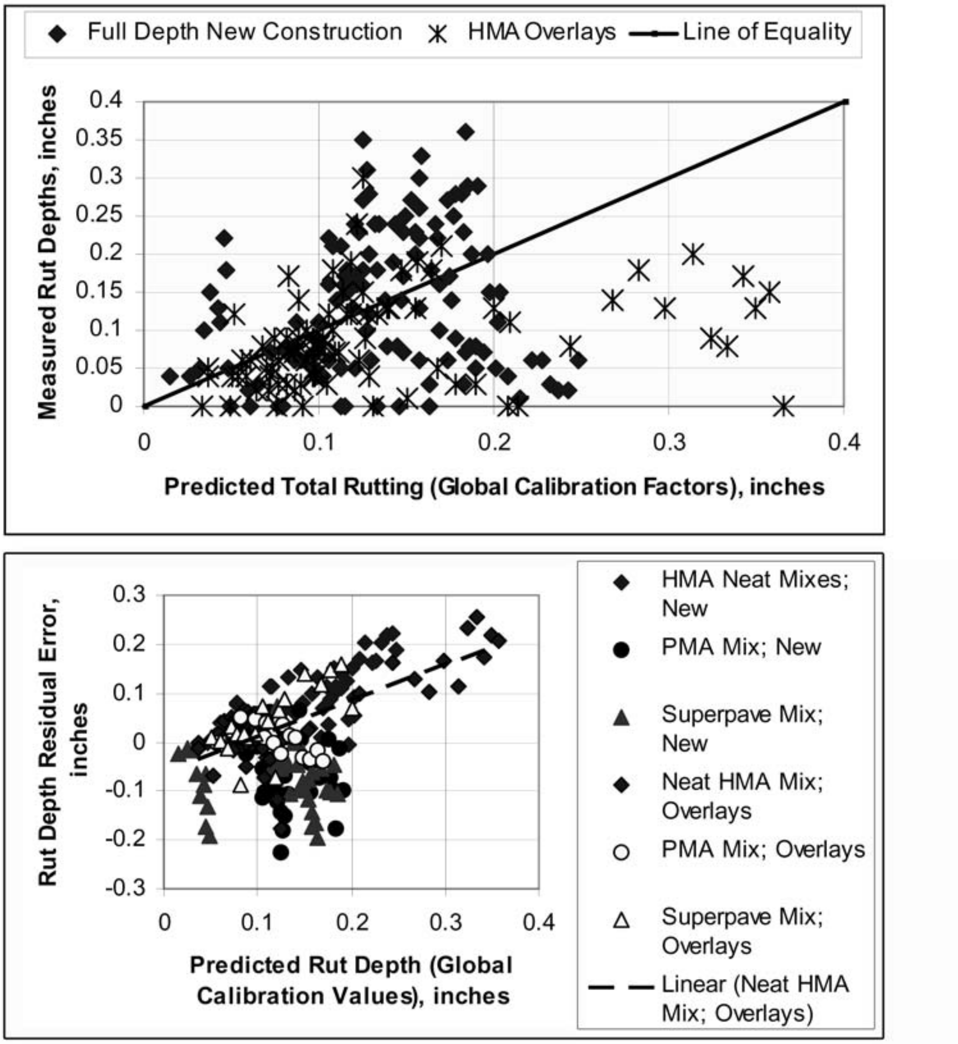


Figure A2-2. Comparison of Predicted and Measured Rut Depths Using the Global Calibration Values and Local Calibration Values of Unity

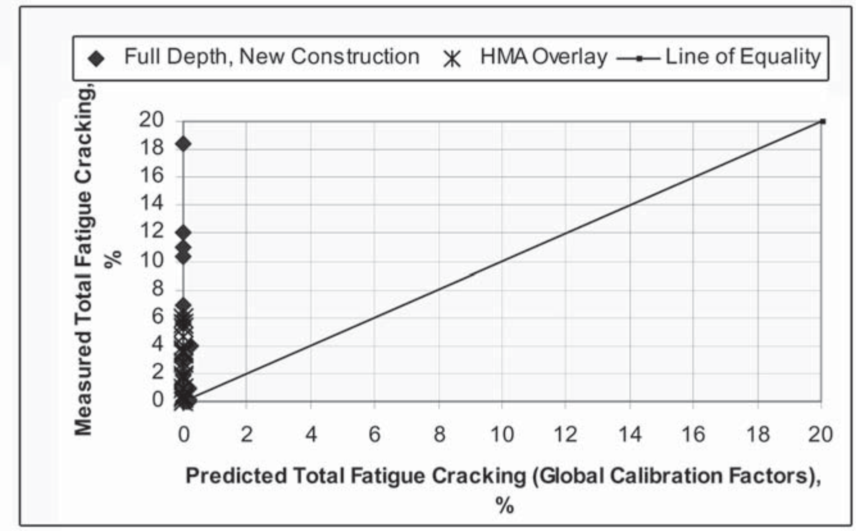


Figure A2-3. Comparison of Predicted and Measured Fatigue Cracking Using the Global Calibration Values and Local Calibration Values of Unity

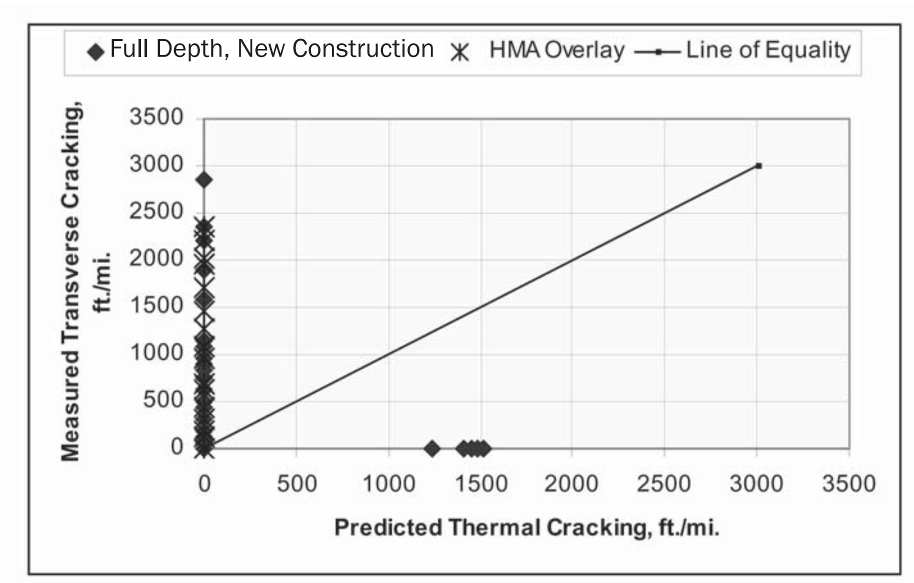


Figure A2-4. Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Global Calibration Values and a Local Calibration Value of Unity

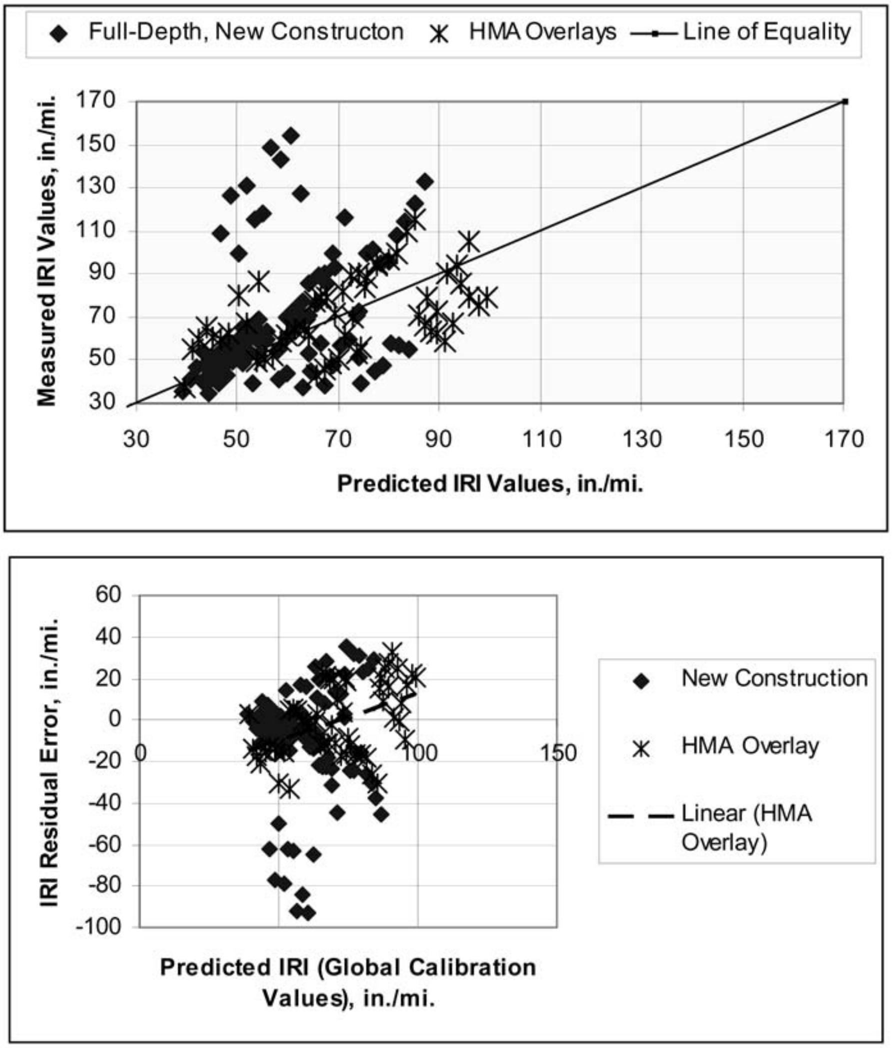
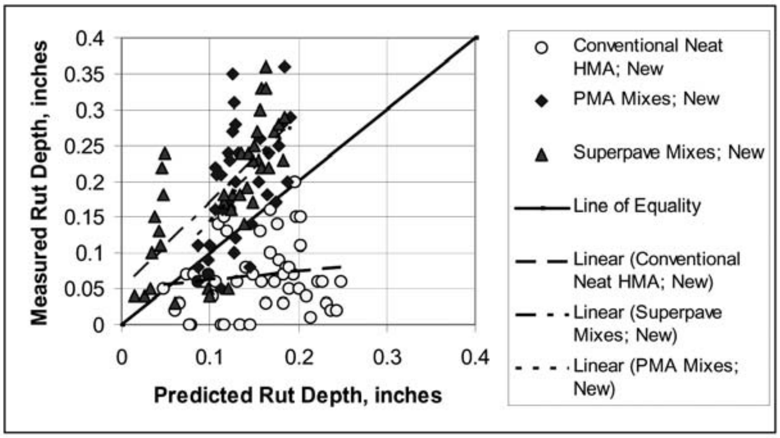
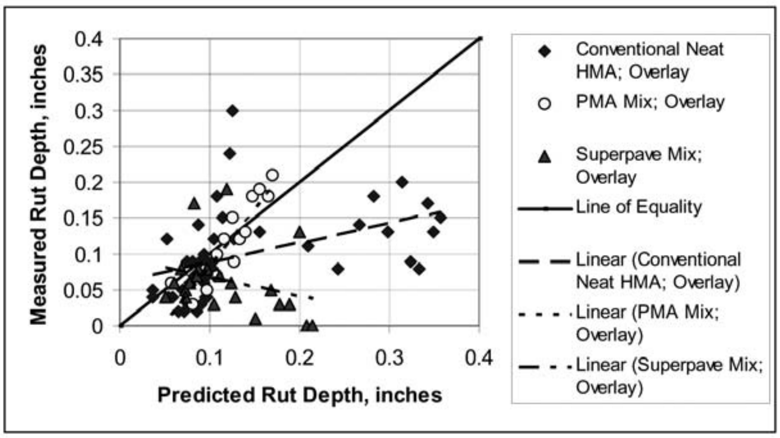


Figure A2-5. Comparison of Predicted and Measured IRI Using the Global Calibration Values and Local Calibration Values of Unity

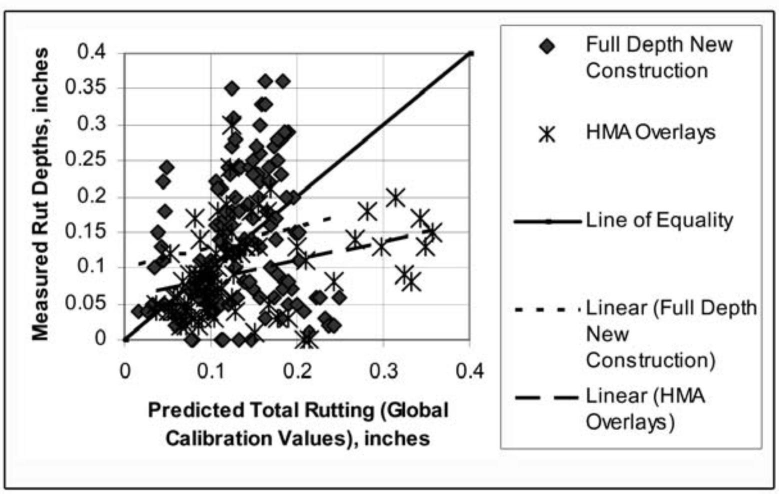
Figures A2-6 and A2-7 show examples of using these estimators of the mean measured values for rut depth and IRI for the primary cells of the sampling template. As shown, the intercept estimator is significantly different from 0, and/or the slope estimator is significantly different from 1.0. More importantly, the intercept and/or slope estimators are dependent on the primary tiers of the sampling template. In summary, all transfer functions exhibited similar trends or bias and that bias is related to pavement structure (new construction versus rehabilitation) and mixture type (conventional neat HMA versus PMA versus Superpave mixtures).



a. Intercept and slope estimators that are dependent on mixture type for the new construction PMS segments.

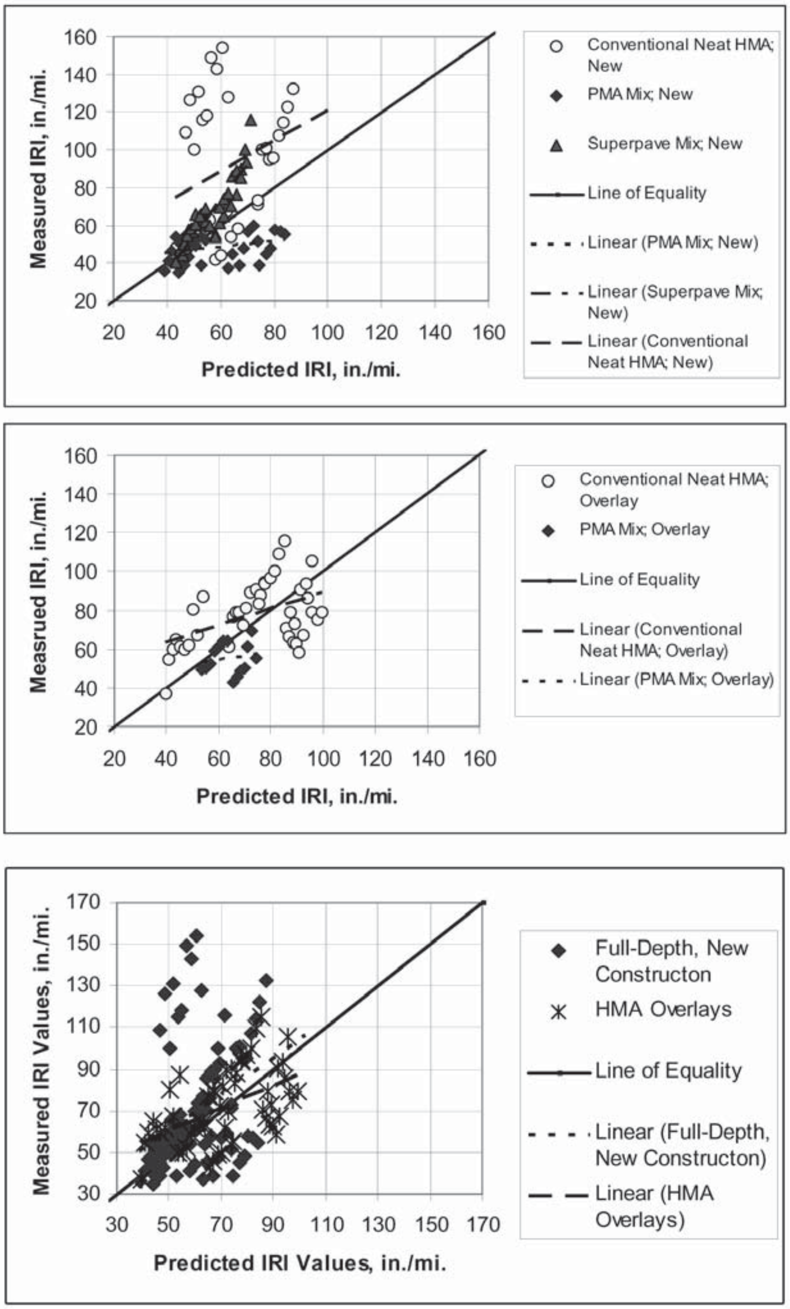


b. Intercept and slope estimators that are dependent on mixture type for the rehabilitation PMS segments.



c. Intercept and slope estimators that are structure dependent for the PMS segments.

Figure A2-6. Comparison of the Intercept and Slope Estimators to the Line of Equality for the Predicted and Measured Rut Depths Using the Global Calibration Values



a. Intercept and slope estimators that are dependent on mixture type for the new construction PMS segments.

b. Intercept and slope estimators that are dependent on mixture type for the rehabilitation PMS segments.

c. Intercept and slope estimators that are structure dependent for the PMS segments.

Figure A2-7. Comparison of the Intercept and Slope Estimators to the Line of Equality for the Predicted and Measured IRI Using the Global Calibration Values

Thus, the MEPDG global calibration values resulted in bias for rutting, transverse cracking, and IRI. The bias for fatigue cracking is low relative to the tolerable bias (refer to Table A2-6), because the measured and predicted values are low. The hypothesis, however, was rejected for all distresses at a 90-percent confidence level based on the slope and/or interceptor estimators. The following summarizes the findings for each transfer function or performance indicator.

- **Rut Depth**
  - Extensive dispersion exists between the predicted and measured rut depths within both sets of data—full-depth new construction and HMA overlays. Poor correlation exists between the predicted and measured rut depths (refer to Figure A2-2 and Attachment A2.4.B).
  - The MEPDGG over predicted the measured rut depths for the HMA overlays (refer to Figure A2-2). More importantly, the residual error is dependent on the pavement structure and type of mixture (refer to Figures A2-2 and A2-6).
  - The maximum rutting predicted in the subgrade and unbound layers varied from 0.1 to 0.3 in. A value of 0.3 in. already exceeds the measured rut depth for most of these PMS segments recorded in the database. In addition, previous forensic studies completed in Kansas on full-depth or depth strength projects with and without stabilized layers have suggested that the rutting in the unbound soils and materials is nil. Thus, it is hypothesized that the subgrade and unbound layer rut depths are over predicted. For this demonstration, it was assumed that the maximum rutting in the subgrade under thick HMA layers would be limited to a value of 0.1 in. in determining the local calibration value for these roadway segments. That local calibration value was then used to predict unbound layer rutting for the PMS segments with thinner HMA layers. This assumption is considered acceptable under the condition that all unbound layers were constructed in accordance with the project specifications.
  - The slope of the rut depth versus time or pavement age is lower for the HMA overlays than for the new construction full-depth segments. The truck traffic (Average Annual Daily Truck Traffic [AADTT]) is low for all of the PMS segments, with AADTT values being less than 100 for nearly half of the segments. Low truck traffic results in minimal increases in rutting over time, after the first couple of years.
- **Total Fatigue (Alligator) Cracking**
  - There are too few PMS segments with measurable fatigue cracking to validate or confirm the global calibration values and determine the local calibration values, if needed (refer to Figure A2-3). Twelve of the 16 PMS segments (75 percent) have none to less than 2 percent fatigue cracking over the monitoring period of time. The MEPDG consistently under predicted the measured fatigue cracks for those PMS segments exhibiting fatigue cracks.
- **Thermal (Transverse) Cracking**
  - Thermal cracking was predicted for only two (87 percent) of the full-depth PMS segments and those segments did not exhibit any transverse cracking, while large amounts of transverse cracks were recorded in the PMS database for many of the PMS segments of both data sets—full-depth, new construction and HMA overlays (refer to Figure A2-4).
  - A negative bias exists for the thermal cracking prediction model, even though the statistical analysis suggests that the bias is insignificant (refer to Table A2-7).
- **IRI or Roughness**
  - The IRI values are heavily dependent on the other distresses calculated by the MEPDG and the site factor. Changing the local calibration factor from unity will affect the IRI values. Thus, the IRI predictions should be evaluated for bias only after the bias has been removed from the other prediction models.

Step 8—Eliminate Local Bias of Distress Prediction Models
All of the globally calibrated transfer functions were found to be biased based on the intercept and slope estimators from Step 7. The process used to eliminate the bias depends on the cause of that bias and the accuracy desired by the agency. In general, there are three possibilities which are listed below.

1. The residual errors are, for the most part, always positive or negative with a low standard error of the estimate in comparison to the trigger value, and the slope of the residual errors versus predicted values is relatively constant and close to zero. The precision of the prediction model is reasonable but the accuracy is poor. In this case, the local calibration coefficient is used to reduce the bias. This condition generally requires the least level of effort and the fewest number of runs or iterations of the MEPDG to reduce the bias.

2. The bias is low and relatively constant with time or number of loading cycles, but the residual errors have a wide dispersion varying from positive to negative values. The accuracy of the prediction model is reasonable, but the precision is poor. In this case, the coefficient of the prediction equation is used to reduce the bias but the value of the local calibration coefficient is probably dependent on some site feature, material property, and/or design feature included in the sampling template. This condition generally requires more runs and a higher level of effort to reduce dispersion of the residual errors.

3. The residual errors versus the predicted values exhibit a significant and variable slope that is dependent on the predicted value. The precision of the prediction model is poor and the accuracy is time or number of loading cycles dependent—there is poor correlation between the predicted and measured values. This condition is the most difficult to evaluate because the exponent of the number of loading cycles needs to be considered. This condition also requires the highest level of effort and many more runs to reduce bias and dispersion.

The third one applies to this demonstration. An analysis of variance (ANOVA) was completed to determine whether e_r, b_o, and/or m are dependent on factors included in the sampling matrix or some other design feature and site condition factor of the PMS segments. As shown in Figures A2-2 and A2-5 the residual error is dependent on pavement structure and mixture type.

To eliminate the bias, the agency should first decide on whether to use the agency specific values or the local calibration factors that are considered as inputs in the MEPDG software. Either one can be used with success—for this demonstration the local calibration parameters were used. Time-history plots of each performance indicator should be prepared to determine if one or multiple calibration factors need to be evaluated, as noted above. The following describes the process used to eliminate the bias using the performance indicators stored in the Kansas PMS database.

Rut Depth Transfer Function
Poor correlation was found between the predicted and measured rut depths using the global calibration values, even though the bias is insignificant for both new construction and rehabilitation strategies (refer to Table A2-7 and Figure A2-2). One possible reason for the poor correlation is the large measurement error in rut depths. There is more variability in the measured rut depths within a PMS segment than between the segments. As an example, the standard deviation of the average maximum rut depth between all of the segments is 0.0754 in. (refer to Table A2-6), while the standard deviation of the average rut depths within the monitoring period for a specific PMS segment can be as high as 0.085 in. Varying the local calibration values to represent different site

conditions and materials will not increase model precision (i.e., reducing the data measurement errors). The following points or observations were identified in completing an analysis of the residual errors relative to the sampling template.

- The maximum predicted rutting in the unbound layers varied from 0.1 to about 0.3 in. This level of rutting in the unbound layers and foundation is inconsistent with previous forensic studies conducted by KSDOT. All PMS segments were constructed in accordance with KSDOT specifications. Thus, it is hypothesized that the rutting in the unbound layers is over predicted for these PMS segments.

The local calibration value for the unbound layers (β_{s1}) was estimated by making repeat runs of the MEPDG with varying values for a limited number of segments (four were used for this demonstration) to reduce the bias within each PMS segment. An average value of 0.50 was estimated for both the fine and coarse-grained soils (new construction and HMA overlays). An insufficient number of PMS segments with different soils were available to determine whether the local calibration value is soil type dependent. [Note: Version 1.0 of the MEPDG only allows the local calibration factor for the subgrade to be altered from unity; changes from unity cannot be made to the unbound aggregate base layers.]

- A review of the comparisons between the predicted and measured rut depths included in Attachment A2.4.B and in Figure A2-2 found that the MEPDG over predicted the measured rut depths for the HMA overlays, under predicted the rutting of the segments with Superpave and PMA mixtures, and over predicted the rutting of the conventional, neat HMA mixtures. Thus, the residual errors or bias on a sampling template basis is cell specific.

The local calibration exponents of temperature (β_{r2}) and number of load cycles (β_{r3}) in the rut depth transfer function were considered adequate, because of the variability in the measured rut depths with time—believed to be measurement error. However, the rut depth versus age relationship for some of the PMS segments has a greater slope than predicted with the MEPDG (see plots in Attachment A2.4.B). Thus, the local calibration parameter for the coefficient (β_{r1}) and number of load cycles (β_{r3}) were the terms considered in the local calibration process. The following summarizes the results from the ANOVA and local calibration process.

- For the full-depth HMA pavements, the residual error was found to be mixture dependent—conventional dense-graded versus polymer modified asphalt (PMA) versus Superpave coarse or gap-graded mixtures. However, correspondence or correlation was not identified between the residual error and volumetric properties of the HMA, layer thickness, and other input values. Thus, constant values for the HMA local calibration values were determined for each mixture type, which are listed below.
  - Conventional, Dense-Graded HMA Mixtures: $\beta_{r1} = 1.5$ and $\beta_{r3} = 0.9$
  - Superpave, Dense-Graded Mixtures: $\beta_{r1} = 1.5$ and $\beta_{r3} = 1.2$
  - Polymer Modified, Dense-Graded Mixtures: $\beta_{r1} = 2.5$ and $\beta_{r3} = 1.15$

- This mixture effect was not found for the HMA overlays, because all of the overlaid segments exhibited much lower rut depths. The average maximum rut depth measured along the new construction projects (full-depth segments) was 0.26 in., while an average maximum value of 0.175 in. was measured along the overlaid segments (refer to Attachment A2.4.B). The values determined for the two local calibration parameters for the HMA-Overlay Projects were: $\beta_{r1} = 1.5$ and $\beta_{r3} = 0.95$.

Table A2-8 lists the rut depth bias using the local calibration values listed above for the full-depth new construction and HMA-Overlay Projects. As shown, the hypothesis is accepted, and the statistical parameters indicate a more precise rut depth prediction model for the Kansas PMS segments (Table A2-7 compared to Table A2-8). Figure A2-8 compares the predicted and measured rut depth using the local calibration values, and shows increased precision, as compared to the use of the global calibration values (refer to Figure A2-2), especially for the new construction—full-depth segments. The rut depth transfer function still does not adequately explain the measured values for the HMA overlays.

Fatigue (Alligator) Cracking Transfer Function

Low- to no-fatigue cracking was predicted for the PMS segments, which exhibited little to no fatigue cracks. Thus, the bias is low. For those limited PMS segments with fatigue cracking, the local calibration coefficient (β_{f1}) was used to reduce the bias to the minimum value possible for those segments and that value was used to predict the fatigue cracking for all PMS segments.

The resulting local calibration value was less than 0.00005 for the HMA-Overlay Projects. This value is low and would result in much greater amounts of fatigue cracking if used for the full-depth new construction projects. Additional runs were made with the MEDPG assuming that bond had been lost between the existing surface and HMA overlay. Using that assumption, the resulting local calibration value was similar to the value determined for the full-depth new construction projects for the same type of mixtures that had exhibited low levels of fatigue cracking (β_{f1} =0.05). Thus, the local calibration value was determined using the condition of zero bond between the existing surface and HMA overlay. The local calibration value was found to be mixture dependent for the full-depth pavements.

Table A2-8. Summary of the Statistical Parameters—Local Calibration Values Used for Predicting the Performance Indicators for the Kansas PMS Sections

Performance Indicator	Project	Bias	Standard Error	s_e/s_y	R^2	Hypothesis; H_o:	Comment
Rutting	New	+0.0249	0.00397	0.522	0.650	Accept; p = 0.316	Transfer function is adequate.
	Rehab.	+0.0278	0.0725	1.26	Poor		Transfer function does not explain variation in measured data.
Fatigue Cracking	New	+0.383	2.154	0.814	0.322	Accept; p = 0.204	Transfer function does not explain variation in measured data.
	Rehab	+1.272	1.441	0.937	Poor		
Transverse Cracking	New	–59.4	313.6	0.626	0.595	Accept; p = 0.453	Transfer function is adequate.
	Rehab	–43.7	410.2	0.610	0.736		
IRI	New	–3.99	9.8	0.348	0.703	Accept; p =0.444	Regression model is adequate.
	Rehab	+0.38	14.3	0.864	0.402		
Rutting	SPS-1	+0.0059	0.209	0.877	0.252	Accept	Significantly under predicting total rutting; slope estimator significantly different than 1.0.
	SPS-5	–0.081	0.0989	1.46	Poor	Accept	

Table A2-8—Continued

Performance Indicator	Project	Bias	Standard Error	s_e/s_y	R^2	Hypothesis; H_o:	Comment
Fatigue Cracking	SPS-1	–8.63	17.34	1.05	Poor	Accept	Significantly under predicting fatigue cracking; slope and intercept estimators significantly different than line of equality
	SPS-5	–2.76	8.57	0.720	Poor	Accept	
Transverse Cracking	SPS-1	–173.2	408.3	0.857	Poor	Accept	Significantly under predicting transverse cracking; slope and intercept estimators significantly different than line of equality
	SPS-5	–638.2	1017.3	1.12	Poor	Accept	
IRI	SPS-1	–6.44	19.43	0.790	0.350	Accept	The hypothesis should be checked and evaluated after any bias has been reduced for the other distresses.
	SPS-5	1.08	8.507	0.973	0.372	Accept	

Note 1: Poor means that the model did not explain variation in the measured data within and between the LTPP test sections.
Residual Error = $e_r = y - x_i$
y_i = Measured or Observed Value; S_y = Standard Deviation of the observed values.
x_i = Predicted Value

The local calibration exponents of tensile strain (β_{f2}) and dynamic modulus (β_{f3}) in the allowable number of load applications equation were considered adequate because of the variability in the measured fatigue cracking with time—believed to be measurement error. The C_2 parameter in the bottom-up fatigue cracking prediction equation was also excluded from the evaluation. The C_2 value of unity determined from the global calibration effort was assumed to be appropriate for the Kansas PMS segments.

That assumption for C_2, however, is probably incorrect. The growth in fatigue cracking with time can be much steeper than predicted by the MEPDG using the global calibration values. This condition is illustrated in Figure A2-9 for the two PMS segments with higher amounts of fatigue cracking. This difference between the measured and predicted values with time will decrease the precision of the fatigue cracking prediction model for the Kansas PMS segments. Unfortunately, there are too few PMS segments with appreciable fatigue cracking to determine a reliable estimate of C_2. In addition, the measurement error and combining longitudinal cracks in the wheel path with the area fatigue cracks (alligator cracks) make it difficult to reliably estimate both C_2 and β_{f1}. The following summarizes the results from the local calibration process.

- For the full-depth HMA pavement (new construction), the HMA local calibration coefficient was found to be mixture dependent—conventional dense-graded versus PMA versus Superpave coarse or gap-graded mixtures. A constant value was determined for each mixture type, which are listed below.

- Conventional, Dense-Graded HMA Mixtures: $\beta_{f1} = 0.05$
- Polymer Modified, Dense-Graded Mixtures: $\beta_{f1} = 0.005$
- Superpave, Dense-Graded Mixtures: $\beta_{f1} = 0.0005$

- This mixture effect was not found for the HMA overlays, similar to the finding for rutting. The value determined for the local calibration parameter for the overlay projects is: $\beta_{f1} = 0.05$. One reason hypothesized for this finding is that inadequate bond exists between the existing surface and HMA overlay. However, other construction/material anomalies (i.e., segregation, stripping, etc.) may explain this difference. Forensic investigations would need to be completed to confirm or reject the inadequate bond hypothesis.

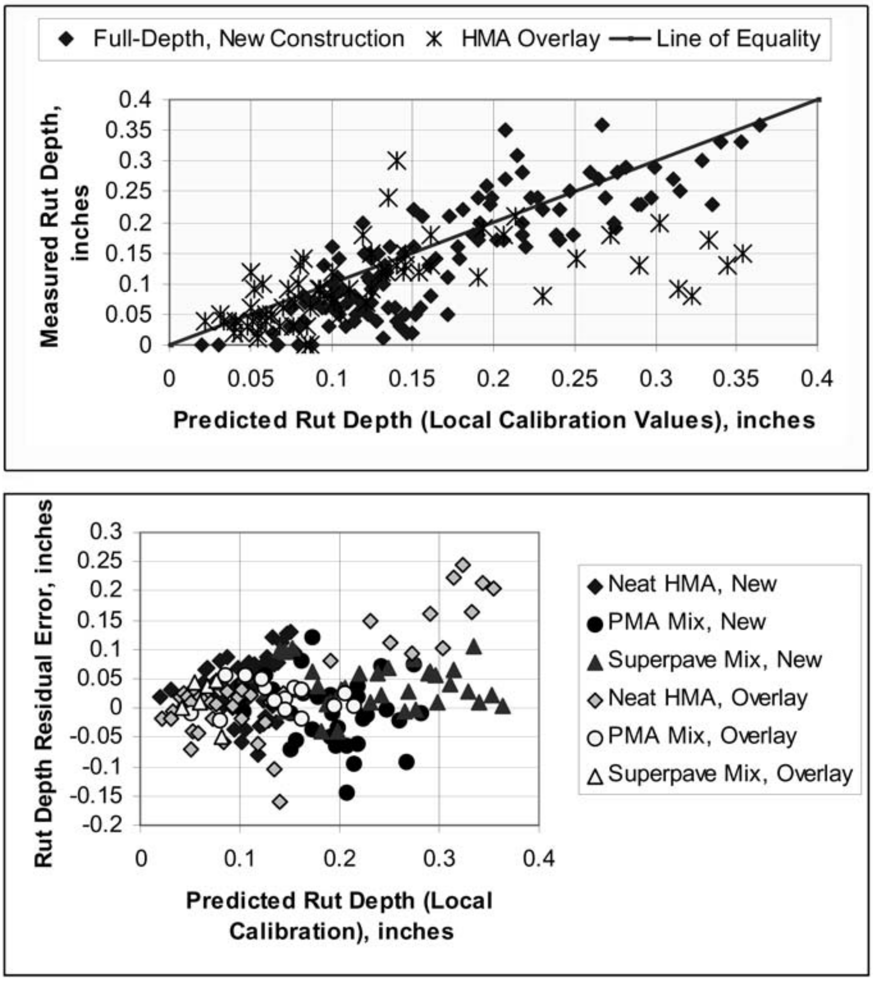


Figure A2-8. Comparison of Predicted and Measured Rut Depths Using the Subgrade and HMA Local Calibration Values for the PMS Segments

Table A2-8 lists the fatigue cracking bias using the local calibration values listed above. As shown, the hypothesis is accepted, but the statistical parameters still indicate a poor correlation between the predicted and measured values (Table A2-7 compared to Table A2-8). Figure A2-10 compares the predicted and measured fatigue cracking using the local calibration values, and shows that there is increased accuracy of the transfer function, as compared to the use of the global calibration values (refer to Figure A2-3).

Although the hypothesis for the local calibration values was accepted, these values would not be recommended for use from a practical engineering standpoint without more segments exhibiting fatigue cracking approaching the trigger value or design criteria. As noted above, 75 percent of the PMS segments have yet to exhibit measurable magnitudes of fatigue cracking and those with measurable cracking were significantly less than 10 percent, with the exception for one PMS segment (refer to Attachment A2.4.B).

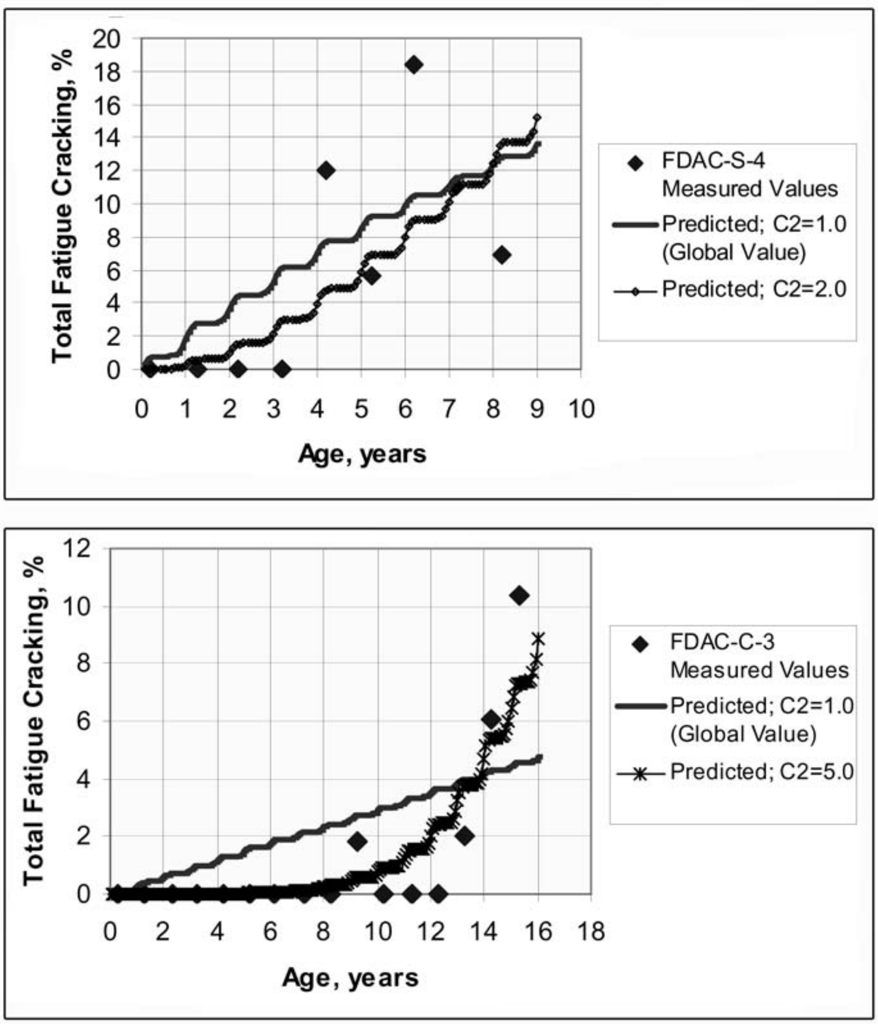


Figure A2-9. Comparison of Measured and Predicted Values of Fatigue Cracking Using Different Value for the C_2 and β_{f1} Parameters for PMS Segments FDAC-C-3 and FDAC-S-4

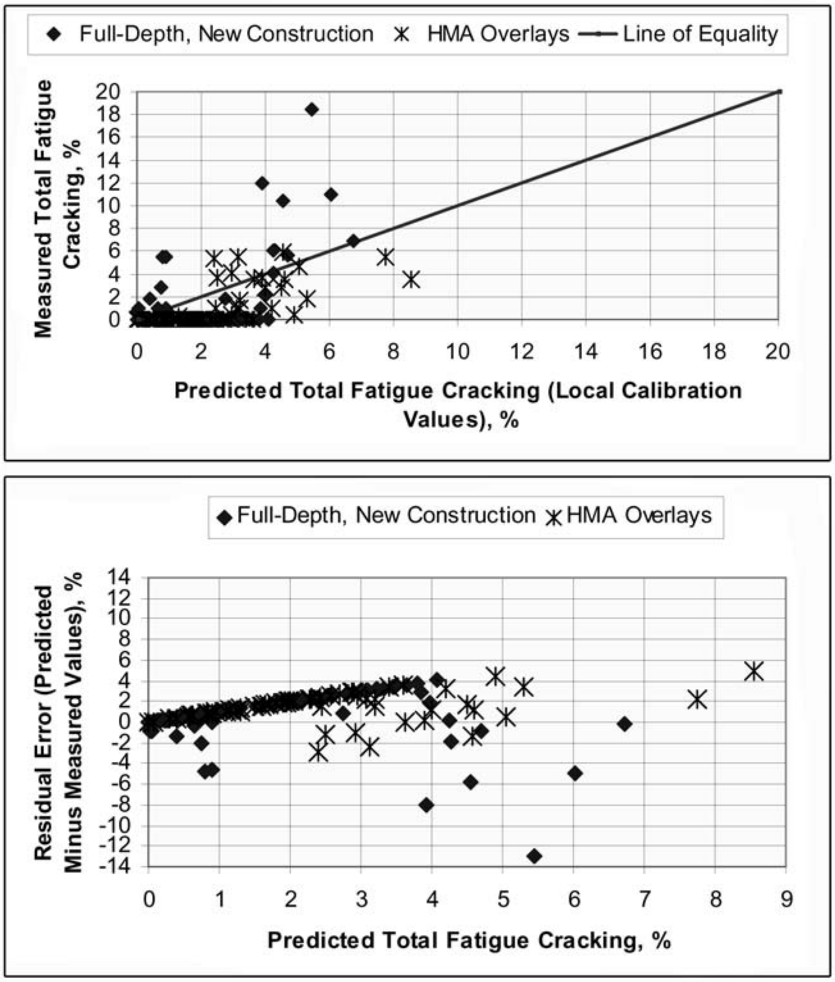


Figure A2-10. Comparison of Predicted and Measured Fatigue Cracking Using the Local Calibration Values for the PMS Segments

Transverse Cracking Transfer Function

The maximum length of thermal cracks predicted by the MEPDG is 2,200 ft/mi, which corresponds to about a 30-ft spacing of transverse cracks. Some of the PMS segments exhibited transverse cracking exceeding that maximum limit. Thus, only those measured responses less than about 2,500 ft/mi should be used in determining the local calibration factor to reduce bias and dispersion.

Figure A2-4 compared the observed and predicted transverse cracks using the global calibration values. As shown, the length of transverse cracks was under predicted for nearly all of the PMA segments for new construction and HMA overlays. In fact, thermal cracking was predicted for only two of the full-depth projects and none for the HMA-Overlay Projects. The local calibration parameter (β_{t3}) was used to reduce that bias (refer to Table A2-7).

In summary, the residual error was found to be mixture and structure dependent—conventional dense-graded versus PMA versus Superpave coarse or gap-graded mixtures; and new construction versus HMA overlays. The thermal cracking local calibration values (β_{t3}) to reduce model bias are listed below for the different mixtures and structures as shown in Table A2-9 (as defined by the sampling template [refer to Table A2-4]).

Table A2-9. Thermal Cracking Local Calibration Values

Mixture Type	Pavement Type	
	New Construction	HMA Overlay
Polymer Modified, Dense-Graded Mixtures	2.0	2.0
Conventional, Dense-Graded HMA Mixtures	2.0	7.5
Superpave, Dense-Graded Mixtures	3.5	7.5

Table A2-8 lists the thermal cracking bias using the local calibration values listed above. As shown, the hypothesis is accepted, and the statistical parameters indicate a more accurate and precise thermal cracking prediction model for the Kansas PMS segments (Table A2-7 compared to Table A2-8). Figure A2-11 compares the predicted and measured thermal cracking using the local calibration value, and shows that there is an increase in model accuracy and precision, as compared to the use of the global calibration values (refer to Figure A2-4). However, the null hypothesis for the intercept and slope estimators for the HMA overlays would still be rejected for the overlaid segments (refer to Figure A2-11).

Roughness or IRI Regression Model

The IRI values predicted by the MEPDG using the local calibration values for the other distresses are within acceptable limits of the measured values. Figure A2-12 compares the measured and predicted IRI values, while Table A2-8 summarizes the statistical information. As shown, the hypothesis was accepted—the IRI regression prediction equation is confirmed for the Kansas PMS segments.

If the hypothesis has been rejected, however, the agency would first identify the distress causing the higher residual errors or if the residual error is heavily time dependent (time and site factor related). The coefficients of the distress and/or site factor terms included in the IRI prediction equation would be determined to reduce local bias.

Step 9—Assess Standard Error of the Estimate

After the bias is reduced or eliminated for each of the transfer functions, the standard error of the estimate (refer to Table A2-8) is evaluated. The Standard Error of the Estimate (SEE) for each globally calibrated transfer function is included under the "Tools" section of the MEPDG software. Figure A2-13 compares the SEE for the globally calibrated transfer functions to the SEE for the locally calibrated transfer functions. For the runs using the local calibration values, the SEE was found to be statistically different in comparison to the SEE included in the MEPDG for each performance indicator. The following summarizes the comparison of the values between the global and local calibration.

- **Rut Depth Transfer Function (Total Rut Depths)**—The SEE values are lower for the locally calibrated transfer function than for the globally calibrated model. The PMS segments, however, exhibited much smaller rut depths than for the global calibration database.
- **Alligator Cracking Transfer Function**—SEE values based on the local calibration are lower than the values determined from the global calibration process. The PMS segments, however, exhibited less than 5 percent fatigue cracks for most of the test sections. This amount of cracking is too low in comparison to the trigger value (refer to Table A2-5) to develop a valid relationship between the SEE and predicted fatigue cracking values.

- **Thermal Cracking Transfer Function**—SEE values based on the local calibration are consistently higher than the values determined from the global calibration process.
- **IRI Regression Model**—SEE values for the IRI regression equation are not included on the MEPDG software screens and can not be changed.

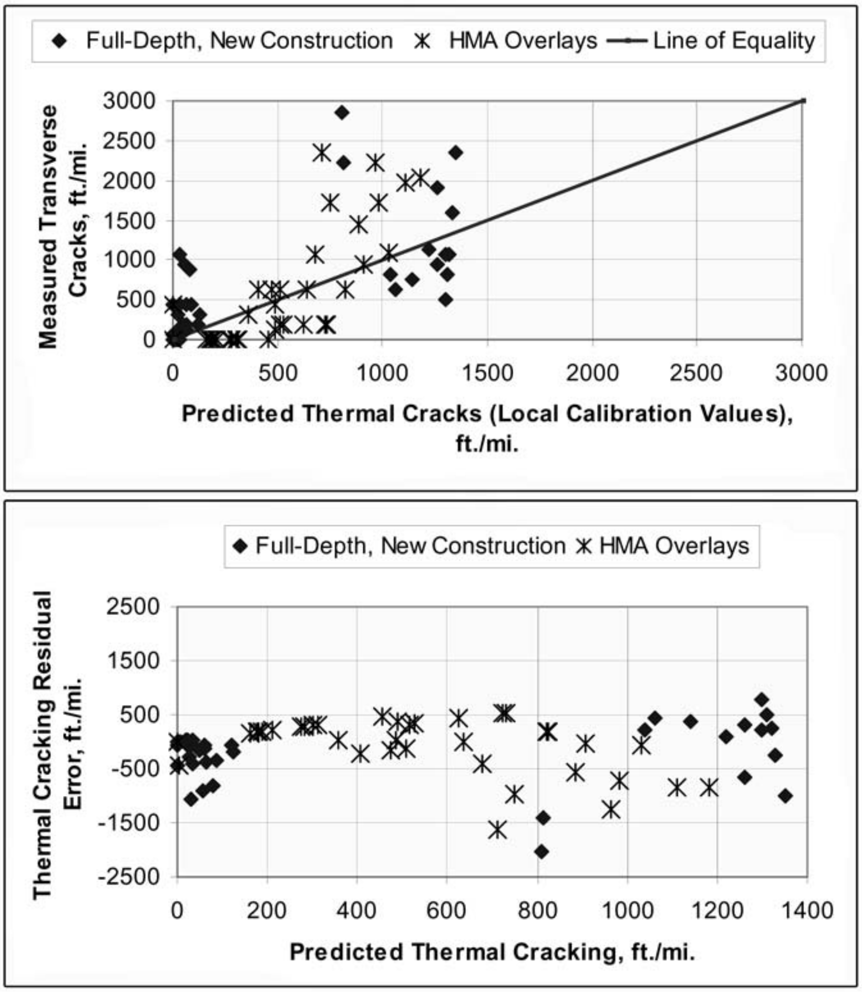


Figure A2-11. Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Local Calibration Value for the PMS Segments

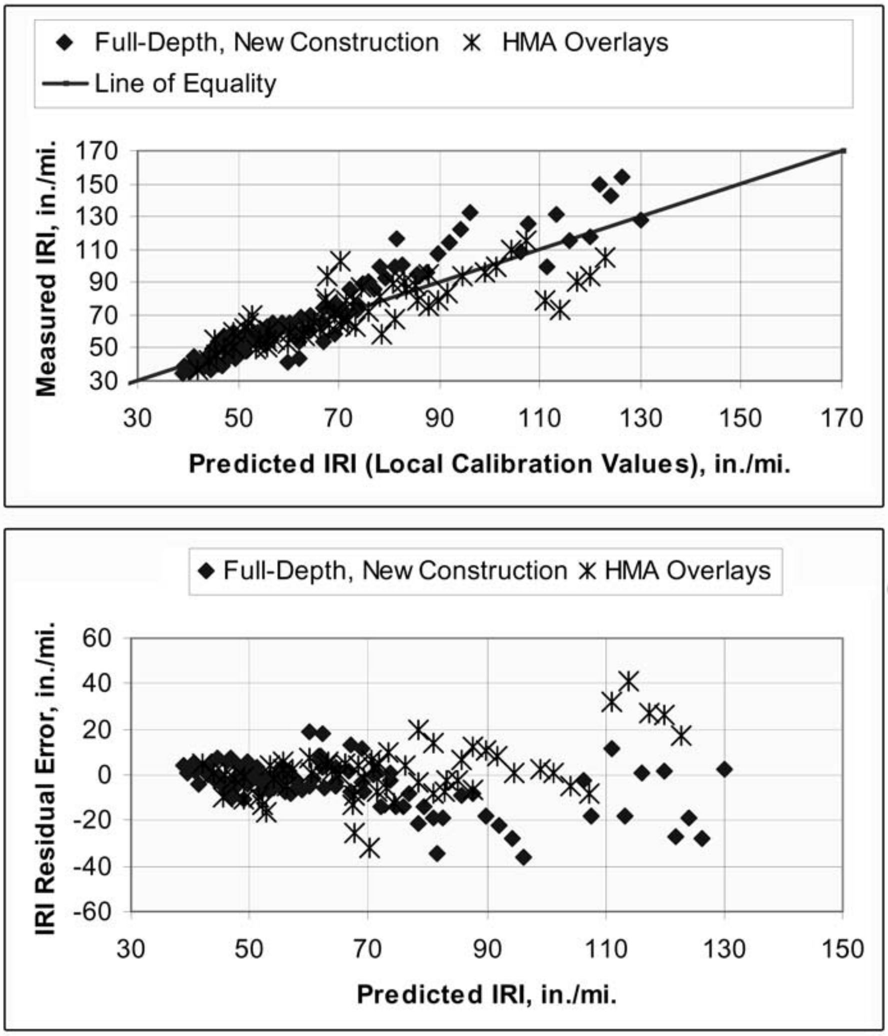


Figure A2-12. Comparison of Predicted and Measured IRI Values Using the Global Calibration Values

Step 10—Reduce Standard Error of the Estimate

As noted in Step 9 and shown in Figure A2-13, the SEE from the local calibration process was found to be different than the SEE relationships included in the MEPDG software for rutting, fatigue cracking, and thermal cracking. An ANOVA can be completed to determine if the residual error is dependent on some other parameter or material/layer property for the PMS segments. No correlation was identified, so the SEE values shown in Figure A2-13 and the local calibration factors summarized in Step 8 are believed to be the final values for the PMS segments included in the sampling matrix.

Step 11—Interpretation of Results and Deciding on Adequacy of Calibration Factors

For this demonstration, the global calibration values did result in a bias for all distresses. The MEPDG did not accurately explain the differences in performance between the different HMA mixtures and pavement structures. To reduce that bias required local calibration values that were different from unity. The MEPDG IRI regression equation was the only model that was confirmed using the KSDOT PMS data. The purpose of this step is to decide whether to adopt the local calibration values or continue to use the global values that were based on data included in the LTPP program from around the United States.

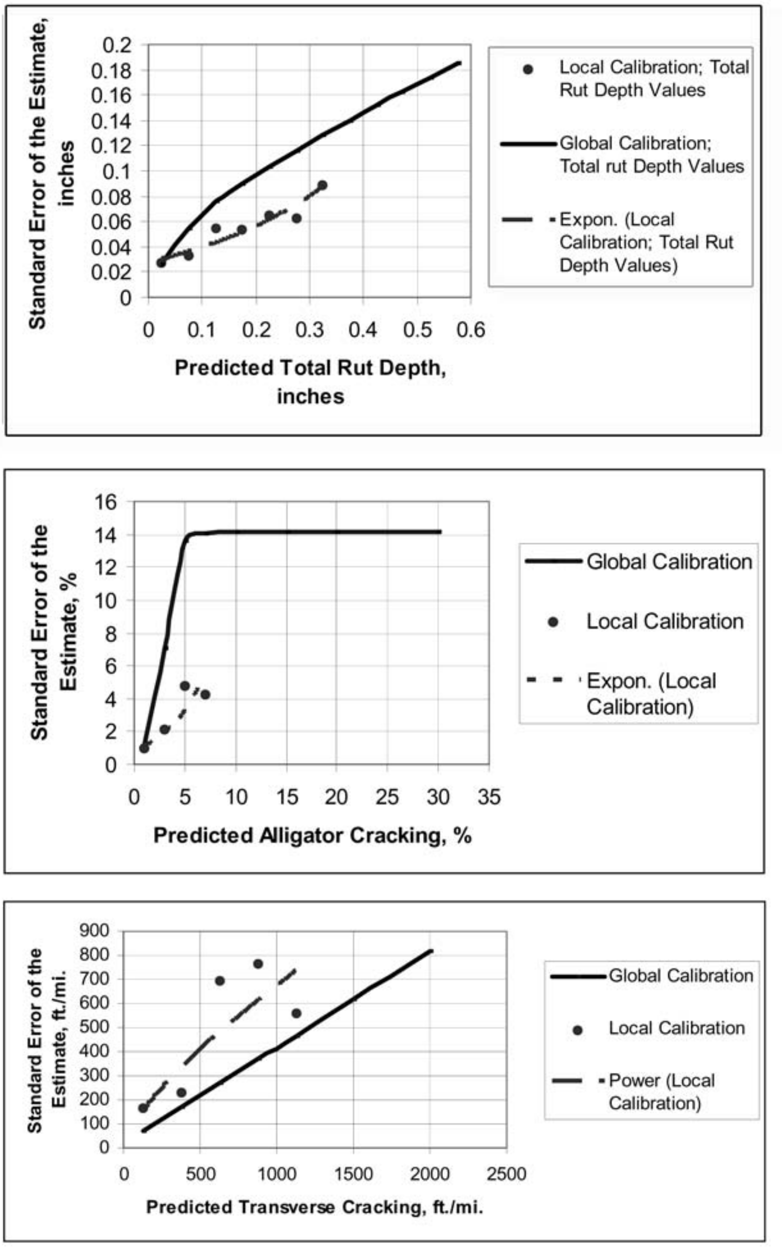


Figure A2-13. Comparison of the Standard Error of the Estimate for the Global-Calibrated and Local-Calibrated Transfer Function

To make that decision, an agency should identify major differences between the LTPP projects and the standard practice of the agency to specify, construct, and maintain their roadway network. Section 3.5 of the *MEPDG Manual of Practice* (*Design Features and Factors Not Included Within the MEPDG Process*) lists the factors and/or features that were excluded from the global calibration process. More importantly, the agency should determine whether the local calibration values can explain those differences. The agency should evaluate any change from unity for the local calibration parameters to ensure that the change provides engineering reasonableness.

Next Page

The interpretation of results is discussed further in Section A2.3 (Summary for Local/Regional Calibration Values) using the two different data sets: PMS segments and selected LTPP SPS projects in and adjacent to Kansas. The following briefly interprets the results using PMS data.

- The IRI regression equation was found to be a reasonable simulation of the IRI values measured on the PMS segments in Kansas. Thus, the MEPDG IRI regression model is believed to be adequate for Kansas climate, materials, and other site features for their more common design strategies and mixtures.

- All HMA mixtures included in the PMS segments exhibited less resistance to fracture or are much more susceptible to fracture than included in the global calibration process. These mixtures are brittle in comparison to those used to determine the global calibration values. Although only small amounts of fatigue cracking have occurred, the HMA layers are thick and the truck traffic low. It is expected that many of these sections may have exhibited surface initiated cracking or the cracking recorded as fatigue in the Kansas PMS database is some other type of cracking caused by a combination of environmental conditions and wheel loads. Since the amount of cracking is low, the fatigue cracking local calibration values should not be used without additional sections exhibiting higher amounts of fatigue cracks. The C_2 parameter seems to be significantly different from unity (refer to Figure A2-9). However, the global calibration value for C_2 (unity) should continue to be used until more segments with higher amounts of fatigue cracking become available to confirm or dispute that observation. There are an insufficient number of PMS segments with higher levels of fatigue cracking to confirm the SEE at the trigger value (design criteria).

- All mixtures are also more susceptible to thermal cracking than those included in the global calibration process, similar to the finding for fatigue cracking. Substantial lengths of transverse cracking were exhibited on many of the PMS segments. Thus, it would be recommended that the local calibration value for thermal cracking be used for design. It would also be recommended that the SEE values determined from the local calibration process be used for design (refer to Figure A2-13).

- The Superpave and PMA mixtures exhibit more rutting potential than the conventional neat HMA mixtures used in Kansas. In summary, the neat HMA mixtures exhibit lower rutting potential than the mixtures included in the global calibration database, while the Superpave and PMA mixtures exhibit slightly higher levels of rutting. However, the magnitude of rutting is low and would not trigger any type of rehabilitation. The HMA local calibration values for rutting would be recommended for use. The SEE values derived from the local calibration are lower than the SEE values derived from the global calibration, but the measured rut depths for the PMS segments are significantly lower than the trigger value (refer to Table A2-5). There are an insufficient number of PMS segments with greater rut depths to confirm the SEE at the trigger value (design criteria).

- The subgrade rutting local calibration value is believed to be reasonable because of the findings from previous forensic studies and would be recommended for use.

In summary, other results or observations from the local validation-calibration process are listed below.