

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

November 2010



Total Fatigue Cracking

18
16
14
12
10
8
6
4
2
0



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

November 2010



Documents

Document Sharing Hub

© 2010, by the American Association of State Highway and Transportation Officials. All rights reserved. This book, or parts thereof, may not be reproduced in any form without written permission of the publisher. Printed in the United States of America.

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

American Association of State Highway and Transportation Officials

Executive Committee 2009/2010

President: Larry L. “Butch” Brown, Sr., Mississippi

Vice President: Susan Martinovich, Nevada

Secretary-Treasurer: Carlos Braceras, Utah

REGIONAL REPRESENTATIVES

REGION I

Joseph Marie, Connecticut, One-Year Term

Gabe Klein, District of Columbia, Two-Year Term

REGION II

Dan Flowers, Arkansas, One-Year Term

Mike Hancock, Kentucky, Two-Year Term

REGION III

Nancy J. Richardson, Iowa, One-Year Term

Thomas K. Sorel, Minnesota, Two-Year Term

REGION IV

Paula Hammond, Washington, One-Year Term

Amadeo Saenz, Jr., Texas, Two-Year Term

NON-VOTING MEMBERS

Immediate Past President: Allen Biehler, Pennsylvania

AASHTO Executive Director: John Horsley, Washington, DC



Documents

Document Sharing Hub

Joint Technical Committee on Pavements 2008/2009

Chair: David Nichols, Missouri
Vice Chair: Judith Corley-Lay, North Carolina
Secretary: Pete Stephanos, FHWA

REGION I

Delaware
Robin Davis (Design)

Maryland
Tim Smith (Materials)

New York
Wes Yang (Design)
Vacant (Design)
Vacant Design

REGION II

Alabama
Larry Lockett (Materials)

Arkansas
Phillip McConnell (Design)

Louisiana
Jeff Lambert (Design)

North Carolina
Judith Corley-Lay (Vice Chair)

South Carolina
Andy Johnson (Design)
Kentucky
Paul Looney (Design)

REGION III

Kansas
Andy Gisi (Design)

Minnesota
Curt Turgeon (Materials)

Missouri
Jay F. Bledsoe (Design)

Missouri
David Nichols (Chair)

Ohio
Aric Morse (Design)

Iowa
Chris Brakke (Design)

REGION IV

California
Bill Farnbach (Design)

Colorado
Richard Zamora (Design)

Oklahoma
Jeff Dean (Design)

Wyoming
Rick Harvey (Materials)
Vacant (Design)



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	1-1
2.0 TERMINOLOGY AND DEFINITION OF TERMS	2-1
2.1 Statistical Terms	2-1
2.2 MEPDG Calibration Terms	2-3
2.3 Hierarchical Input Level Terms	2-3
2.4 Distress or Performance Indicator Terms	2-4
3.0 SIGNIFICANCE AND USE	3-1
4.0 DEFINING ACCURACY OF MEPDG PREDICTION MODELS	4-1
4.1 Calibration	4-1
4.2 Validation	4-3
4.3 General Approach to Local Calibration-Validation	4-4
4.3.1 Traditional Approach—Split-Sample	4-4
4.3.2 Jack-knife Testing—An Experimental Approach to Refine Model Validation	4-4
5.0 COMPONENTS OF THE STANDARD ERROR OF THE ESTIMATE	5-1
5.1 Distress/IRI Measurement Error	5-2
5.2 Estimated Input Error	5-3
5.3 Model or Lack-of-Fit Error	5-3
5.4 Pure Error	5-3
6.0 STEP-BY-STEP PROCEDURE FOR LOCAL CALIBRATION	6-1
7.0 REFERENCED DOCUMENTS AND STANDARDS.....	7-1
7.1 Referenced Documents	7-1
7.2 Test Protocols and Standards	7-1

APPENDIX: EXAMPLES AND DEMONSTRATIONS FOR LOCAL CALIBRATION A-1

A1. Background	A-1
A2. New Flexible Pavements and Rehabilitation of Flexible Pavements	A-2
A2.1 Demonstration 1—PMS Data and Local Calibration	A-2
A2.1.1 Description of PMS Segments	A-2
A2.1.2 Step-by-Step Procedure	A-2
A2.2 Demonstration 2—LTPP Data and Local Calibration	A-34
A2.2.1 Description of LTPP Test Sections Used in Demonstration.....	A-34
A2.2.2 Step-by-Step Procedure	A-35
A2.3. Summary for Local/Regional Calibration Values	A-62
A2.3.1 Comparison of Results: PMS Segments and LTPP SPS Test Sections	A-62
A2.3.1.1 Alligator (Fatigue) Cracking Transfer Function	A-63
A2.3.1.2 Rut Depth Transfer Function	A-64
A2.3.1.3 Thermal Cracking Transfer Function	A-65
A2.3.1.4 IRI Regression Model	A-66
A2.3.2 Application of Results from Local Calibration Process for Pavement Design	A-66

A2.4 Attachments.....	A-68
A2.4.A Attachment A—Description of PMS Segments	A-68
A2.4.A.1 HMA Full-Depth New Construction/Reconstruction Projects	A-68
A2.4.A.2 HMA Overlay of Flexible Pavement Projects	A-72
A2.4.B Attachment B—Plots of Time-History Performance Data	A-73
A2.4.B.1 Full-Depth HMA, New Construction	A-74
A2.4.B.2 HMA Overlays of Flexible Pavements, Rehabilitation	A-80
A2.4.C Attachment C—Description of LTPP Projects	A-84
A2.4.C.1 Full-Depth and Conventional New Construction—LTPP SPS-1 Projects	A-84
A2.4.C.2 HMA Overlays of Flexible Pavement—LTPP SPS-5 Projects	A-85
A2.4.D Attachment D—Plots of Time-History Performance Data for the LTPP SPS Projects	A-86
A2.4.D.1 Full-Depth and Conventional HMA, New Construction	A-86
A2.4.D.2 HMA Overlays of Flexible Pavements, Rehabilitation	A-90
A3. New Rigid Pavements—Jointed Plain Concrete Pavements.....	A-94
A3.1 Demonstration 3—LTPP and PMS Data and Local Calibration.....	A-94
A3.1.1 Description of LTPP And PMS Segments.....	A-94
A3.1.2 Step-by-Step Procedure.....	A-94
A3.2 Attachments.....	A-117
A3.2.A Attachment A—Description of MODOT LTPP and PMS JPCP Segments	A-117
A3.2.A.1 Design (Analysis) Life	A-117
A3.2.A.2 Analysis Parameters	A-118
A3.2.A.3 Traffic	A-119
A3.2.A.4 Climate	A-124
A3.2.A.5 Pavement Surface Layer Thermal Properties	A-124
A3.2.A.6 Design Features for JPCP Sections	A-124
A3.2.A.7 Pavement Structure Definition	A-126
A3.2.B Attachment B—Plots of Time-History Performance Data	A-126

List of Figures

Figure 1-1.	Conceptual Flow Chart of the Three-Stage Design/Analysis Process for the MEPDG.....	1-2
Figure 6-1.	Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 1 Through 5	6-2
Figure 6-2.	Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 6 Through 11	6-3
Figure A2-1.	General Location of the Roadway Segments Selected for Demonstrating the Local Validation-Calibration Process Using Kansas PMS Data	A-8
Figure A2-2.	Comparison of Predicted and Measured Rut Depths Using the Global Calibration Values and Local Calibration Values of Unity	A-16
Figure A2-3.	Comparison of Predicted and Measured Fatigue Cracking Using the Global Calibration Values and Local Calibration Values of Unity.....	A-17
Figure A2-4.	Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Global Calibration Values and a Local Calibration Value of Unity.....	A-17
Figure A2-5.	Comparison of Predicted and Measured IRI Using the Global Calibration Values and Local Calibration Values of Unity	A-18
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-19
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	A-20
Figure A2-8.	Comparison of Predicted and Measured Rut Depths Using the Subgrade and HMA Local Calibration Values for the PMS Segments ..	A-26
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	A-27
Figure A2-10.	Comparison of Predicted and Measured Fatigue Cracking Using the Local Calibration Values for the PMS Segments.....	A-28
Figure A2-11.	Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Local Calibration Value for the PMS Segments.....	A-30
Figure A2-12.	Comparison of Predicted and Measured IRI Values Using the Global Calibration Values	A-31
Figure A2-13.	Comparison of the Standard Error of the Estimate for the Global-Calibrated and Local-Calibrated Transfer Function	A-32
Figure A2-14.	Rut Depths Measured Over Time for the Kansas SPS-1 Project	A-38

Figure A2-15. Determination of the Resilient Modulus of the Subgrade Using Laboratory Test Data Included in the LTPP Database	A-41
Figure A2-16. Determination of the Resilient Modulus of the Unbound Aggregate Base Layer Using Laboratory Test Data Included in the LTPP Database.....	A-41
Figure A2-17. Comparison of Predicted and Measured Rut Depths Using the Global Calibration Values.....	A-43
Figure A2-18. Comparison of Predicted and Measured Fatigue Cracking Using the Global Calibration Values.....	A-44
Figure A2-19. Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Global Calibration Values.....	A-45
Figure A2-20. Comparison of Predicted and Measured IRI Using the Global Calibration Values.....	A-46
Figure A2-21. Comparison of Predicted and Measured Rut Depths Using the Local Calibration Values for the Subgrade, Unbound Aggregate, and HMA Layers.....	A-53
Figure A2-22. Comparison of Measured and Predicted Values of Fatigue Cracking for Specific Test Sections	A-54
Figure A2-23. Comparison of Predicted and Measured Fatigue Cracking Using a Local Calibration Values for the HMA Mixture That Is Air Void Dependent	A-55
Figure A2-24. Comparison of Predicted Thermal Cracking and Measured Transverse Cracking Using the Local Calibration Values for the HMA Mixture.....	A-57
Figure A2-25. Comparison of Predicted and Measured IRI Values Using the Global Calibration Values.....	A-59
Figure A2-26. Comparison of the Standard Error of the Estimate from the Global and Local Calibration Process	A-60
Attachments	
Figure A2.4.B.1-1. Full-Depth HMA, New Construction.....	A-74
Figure A2.4.B.2-1. HMA Overlays of Flexible Pavements, Rehabilitation.	A-80
Figure A2.4.D.1-1. Rut Depth Measurements and Predictions with the Global Calibration Values.....	A-86
Figure A2.4.D.1-2. Fatigue Cracking Measurements and Predictions with the Global Calibration Values.....	A-87
Figure A2.4.D.1-3. Transverse Cracking Measurements and Thermal Cracking Predictions with the Global Calibration Values.....	A-88
Figure A2.4.D.1-4. IRI Measurements and Predictions with the Global Calibration Values.....	A-89
Figure A2.4.D.2-1. Rut Depth Measurements and Predictions with the Global Calibration Values.....	A-90
Figure A2.4.D.2-2. Fatigue Cracking Measurements and Predictions with the Global Calibration Values.....	A-91

Figure A2.4.D.2-3. Transverse Cracking Measurements and Thermal Cracking Predictions with the Global Calibration Values.....	A-92
Figure A2.4.D.2-4. IRI Measurements and Predictions with the Global Calibration Values	A-93
Figure A3-1. General Location of the Missouri LTPP and PMS Roadway Segments Selected for Demonstrating the Local Validation-Calibration Process	A-96
Figure A3-2. LTPP Transverse Cracking	A-101
Figure A3-3. Illustration of Initial IRI Backcasting for G01-S-S1	A-105
Figure A3-4. Illustration of Field Measurement of Permanent Curl/Warp Effective Temperature	A-106
Figure A3-5. Comparison of Predicted and Measured Transverse Joint Faulting Using the Global Calibration Values	A-109
Figure A3-6. Comparison of Predicted and Measured IRI Using the Global Calibration Values.....	A-110
Figure A3-7. Different Stages of Faulting Development.....	A-112
Figure A3-8. Effect of Parameter C_1 on Faulting Prediction.....	A-112
Figure A3-9. Effect of Parameter C_3 on Faulting Prediction.....	A-113
Figure A3-10. Effect of Parameter C_7 on Faulting Prediction.....	A-113
Figure A3-11. Effect of Parameter C_8 on Faulting Prediction.....	A-114
Figure A3-12. Comparison of Predicted and Measured Faulting Using the New Faulting Local Calibration Values for MODOT LTPP and PMS Segments	A-116
Figure A3-13. Locations of WIM Sites in Missouri from Which Traffic Data Were Obtained for Validation-Recalibration	A-121
Figure A3-14. Summary of Years of WIM Data Available.....	A-122
Figure A3-15. Monthly Truck Volume Adjustment Factors for 29-1005.....	A-122
Figure A3-16. Cumulative Single Axle-Load Distribution for Class 5 Trucks on MODOT Heavy Duty Pavements Pertaining to TTC 1	A-123
Figure A3-17. Cumulative Tandem Axle-Load Distribution for Class 9 Trucks on MODOT Heavy Duty Pavements Pertaining to TTC 1	A-123
Figure A3-18. Lateral Truck Wander and Mean Number Axles/Truck for 29-1005	A-124
Attachments	
Figure A3.2.B.-1. PMS Sites.....	A-127

List of Tables

Table 5-1.	Summary of Major Components of Calibration Error Dependency	5-2
Table 6-1.	Recommendation for the Flexible Pavement Transfer Function Calibration Parameters to Be Adjusted for Eliminating Bias and Reducing the Standard Error	6-13
Table 6-2.	Recommendation for the Rigid Pavement Transfer Function Calibration Coefficients to Be Adjusted for Eliminating Bias and Reducing the Standard Error	6-13
Table A2-1.	General Structure Information for the Selected Kansas PMS Projects	A-3
Table A2-2.	General Project Information for the Kansas PMS Segments	A-5
Table A2-3.	Material Types and Layer Thicknesses for the Kansas PMS Segments	A-6
Table A2-4.	Simplified Sampling Template for the Demonstration Using PMS Data	A-8
Table A2-5.	Estimated Number of PMS Segments Needed for the Local Validation-Calibration Process	A-9
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)	A-11
Table A2-7.	Summary of the Statistical Parameters—Global Calibration Values Used for Predicting Performance Indicators for the Kansas PMS Sections	A-14
Table A2-8.	Summary of the Statistical Parameters—Local Calibration Values Used for Predicting the Performance Indicators for the Kansas PMS Sections	A-24
Table A2-9.	Thermal Cracking Local Calibration Values	A-29
Table A2-10.	LTPP SPS-1 and SPS-5 Site Locations	A-34
Table A2-11.	Sampling Template for the Demonstration Using LTPP Data	A-35
Table A2-12.	Summary of the Maximum Values of Different Performance Indicators in Comparison to the Design Criteria or Trigger Values (Number of Sections = 56)	A-37
Table A2-13.	Summary of the Statistical Parameters—Global Calibration Values Used to Predict the Performance Indicators of the LTPP SPS-1 and SPS-5 Projects	A-42
Table A2-14.	Summary of the Performance Indicator Predictions Using the Global Calibration Values	A-47
Table A2-15.	Local Calibration Parameter of Unbound Layers	A-51
Table A2-16.	An Analysis of Residual Errors from the Use of Global Calibration Values	A-51
Table A2-17.	Summary of the Statistical Parameters—Local Calibration Values Used for Predicting	A-52
Table A2-18.	Local Calibration Values for Ranges of Air Voids in Relation to Thermal Cracking	A-56
Table A2-19.	Comparison of the Two Demonstrations for Flexible Pavements and HMA Overlays	A-63
Table A2-20.	HMA Layers/Mixture Local Calibration Parameters, β_{f1} , C_2	A-64
Table A2-21.	Subgrade/Unbound Layer Local Calibration Parameter, β_{s1}	A-65
Table A2-22.	HMA Layers/Mixture Local Calibration Parameters, β_{r1} , β_{r3}	A-65

Table A2-23. Summary β_{fi} Values from Kansas PMS and LTPP Data	A-66
Table A3-1. General Information for the Selected Missouri LTPP and PMS Projects	A-95
Table A3-2. Material Types and Layer Thicknesses for the Missouri LTPP and PMS Segments .A-	96
Table A3-3. Design Features for the Missouri LTPP and PMS Segments.....	A-99
Table A3-4. Simplified Sampling Template for the Demonstration Using LTPP and PM Data .A-	99
Table A3-5. Estimated Number of PMS Segments Needed for the Local Validation-Calibration Process	A-100
Table A3-6. Summary of the Maximum Values of Different Performance Indicators in Comparison to the Design Criteria or Trigger Values (Number of Sites = 24)	A-102
Table A3-7. Summary of the Statistical Parameters—Global Calibration Values Used for Predicting Performance Indicators for the Missouri LTPP and PMS Sections.....	A-107
Table A3-8. Comparison of Measured and Predicted Transverse Slab Cracking (Percentage of All Data Points)	A-108
Table A3-9. Summary of New Local Coefficients for Faulting Model.....	A-115
Table A3-10. Summary of the Statistical Parameters—Global Calibration Values Used for Predicting Performance Indicators for the Missouri LTPP and PMS Sections .A-	115
Table A3-11. Summary of Construction Dates and Analysis Periods for All New JPCP	A-117
Table A3-12. Summary of Backcast Initial IRI Values	A-118
Table A3-13. Summary of Traffic Inputs for All New JPCP Projects	A-120
Table A3-14. JPCP Project Design Features	A-125



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 MEDPG. The purpose of this guide is to provide guidance in calibrating the MEDPG to local conditions and materials that may not have been included in the global calibration process.

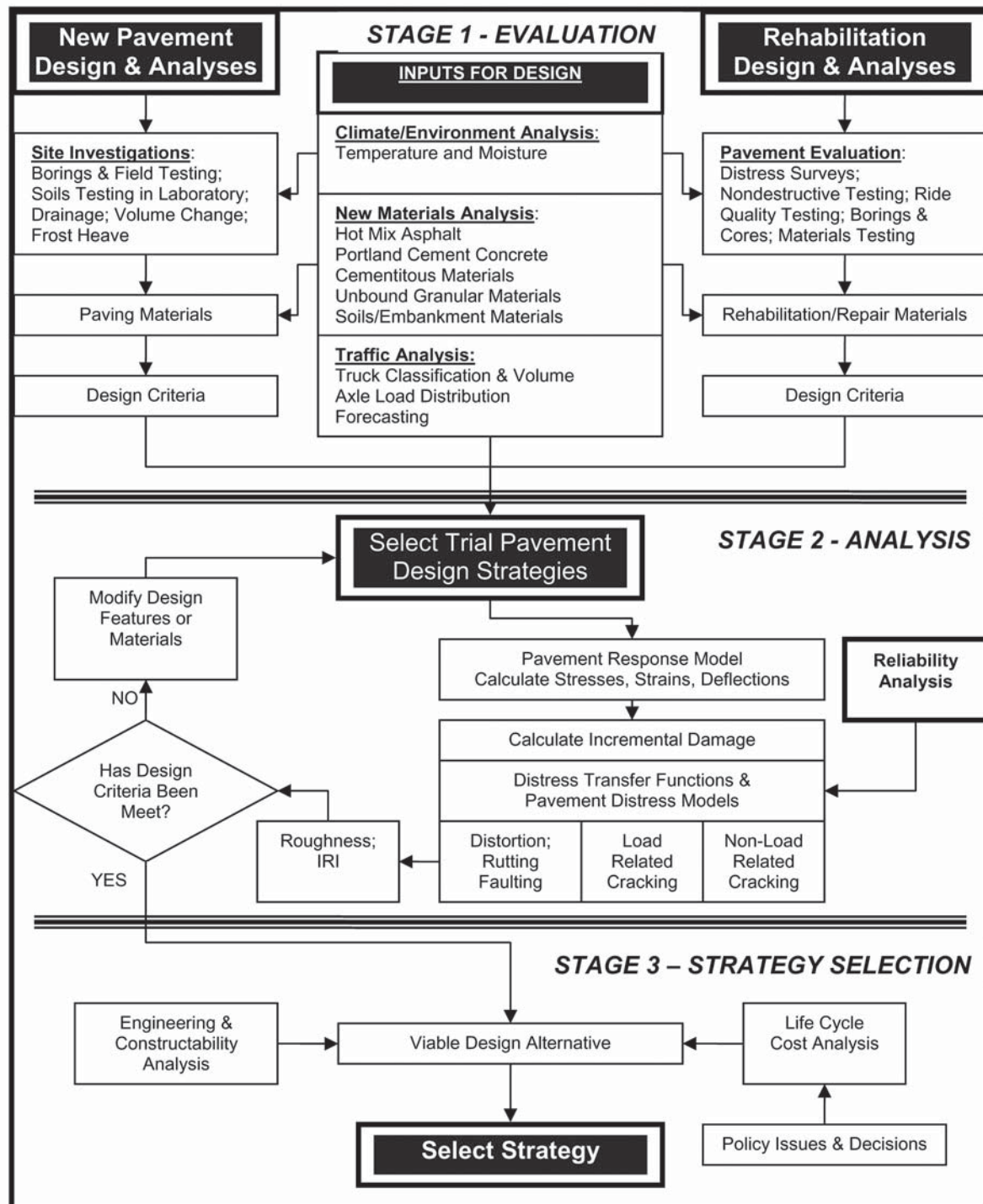


Figure 1-1. Conceptual Flow Chart of the Three-Stage Design/Analysis Process for the MEDPG



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, mathematical.*

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 MEPDG, 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 multiple calibration parameters. Numerical optimization methods, however, require a larger number of sections or observations and many more runs of the MEPDG to determine 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.

$$\varepsilon_i(b) = y_i - n(x_i; b) \quad (4-1)$$

Where y_i is the i th observed response and $n(x_i; b)$ is the i th 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 $\varepsilon_i(\theta) = y_i - \eta(x_i; \theta)$, where y_i is the i th observed response and $\eta(x_i; \theta)$ is the i th 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 ($\hat{y}$) 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} , $j=1 \dots p$ and a single criterion variable Y_i , with $i=1 \dots n$ 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 k th 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 ($\hat{Y}_k$) 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.

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

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_{Pure} = 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.

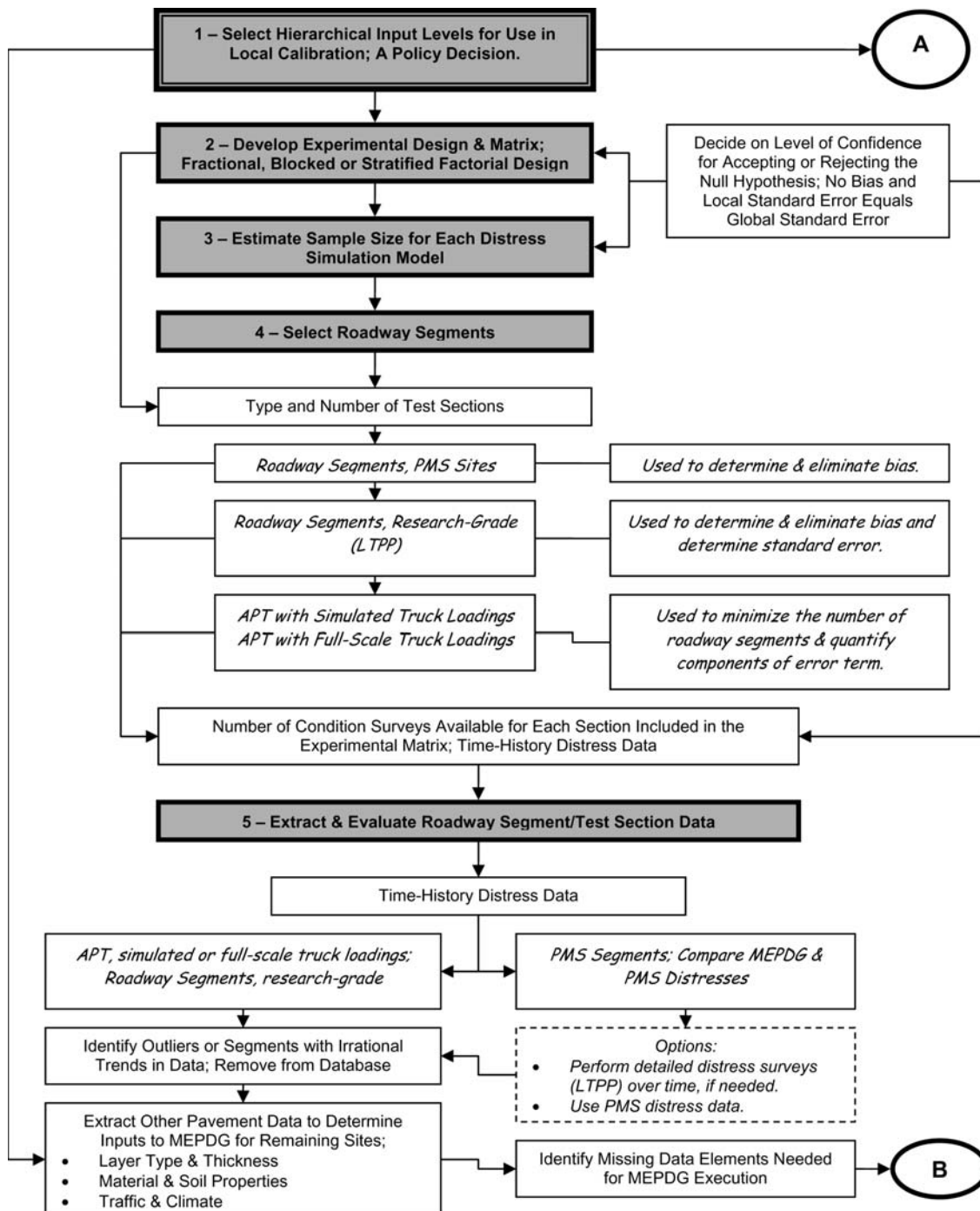


Figure 6-1. Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 1 Through 5

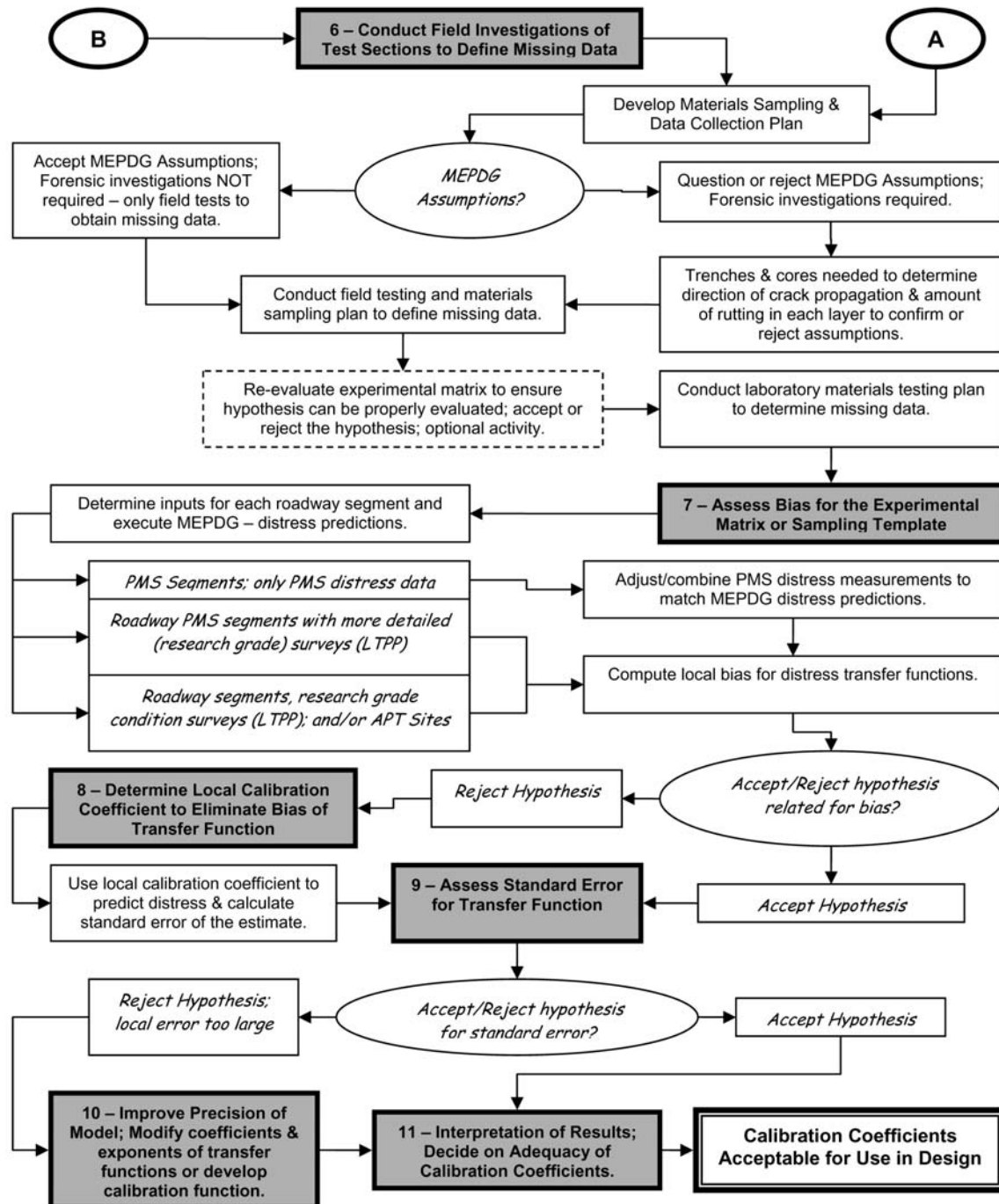


Figure 6-2. Flow Chart of the Procedure and Steps Suggested for Local Calibration; Steps 6 Through 11

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:

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

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:

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

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

$$\frac{s_e}{s_y} \geq \left[\frac{x_a^2}{n-1} \right]^{0.5} \quad (6-3)$$

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 repeatability 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 *Standard Practice for Evaluating Faulting of Concrete Pavements*. Distress measurements should be made to ensure consistency with the MEPDG predictions of distress and smoothness (FHWA, 2003). Pavement smoothness should be measured in accordance with AASHTO R 57 *Standard 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 purposes 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:

$$H_0: \sum(y_{\text{Measured}} - x_{\text{Predicted}}) = 0 \quad (6-4)$$

where:

y_{Measured} = Measured value, and
 $x_{\text{Predicted}}$ = 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.

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

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*



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.