Canadian Journal of Civil Engineering
Revue canadienne de génie civil

Local Calibration of Flexible Pavement Performance Models in Michigan

Journal:	<i>Canadian Journal of Civil Engineering</i>
Manuscript ID	cjce-2015-0556.R1
Manuscript Type:	Article
Date Submitted by the Author:	12-Aug-2016
Complete List of Authors:	Haider, Syed; Michigan State University, Civil and Environmental Engineering Brink, Wouter; Applied Research Associates, Inc. Buch, Neeraj; Michigan State University, Department of Civil and Environmental Engineering
Keyword:	Calibration, Fatigue cracking, Rutting, IRI, bootstrapping

Local Calibration of Flexible Pavement Performance Models in Michigan

Syed Waqar Haider, Ph.D., P.E.
Associate Professor (Corresponding Author)
Department of Civil and Environmental Engineering, Michigan State University, Engineering
Building, 428 S. Shaw Lane, Room 3546,
East Lansing, MI 48824,
Phone: 517-353-9782; Fax: 517-432-1827; e-mail: syedwaqa@egr.msu.edu

Wouter C. Brink, Ph.D.
Staff Civil Engineer
Applied Research Associates, Inc.
Transportation Infrastructure Division, 100 Trade Centre Drive, Suite 200
Champaign, IL 61820-7233,
Phone: 217-356-4500 Ext 273; Fax 217-356-3088;
e-mail: wbrink@ara.com

Neeraj Buch, Ph.D.
Professor and Associate Dean for Undergraduate Studies
Department of Civil and Environmental Engineering
428 S. Shaw Lane, 1415 Engineering Building, Michigan State University
East Lansing, MI 48824
Phone: 517-355-5128; Fax: 517-432-1356; e-mail: buch@egr.msu.edu

Text Count = 4,323

Number of Figures and Tables = 13 (Equivalent word count = 3,250)

Total Word Count = 7,573

Haider, Brink, and Buch

Local Calibration of Flexible Pavement Performance Models in Michigan

ABSTRACT

The performance prediction models in the Pavement-ME are nationally calibrated using in-service pavement material properties, pavement structure, climate and truck loading conditions, and performance data obtained from the Long Term Pavement Performance (LTPP) program. Generally, the nationally calibrated models may not be accurate if the inputs and performance data used to calibrate do not represent a state local conditions and practices. Therefore, each state highway agency (SHA) should evaluate the nationally calibrated performance models to determine the adequacy of predicted field performance before implementing the new M-E design procedure. If the predictions are not satisfactory, local calibration of the Pavement-ME performance models is recommended to improve the performance prediction capabilities reflecting the unique field conditions and design practices. The commonly used calibration technique such as split sampling does not necessarily provide adequate results, especially with small sample sizes. Consequently, there is a need to employ statistical methodologies that are more efficient and robust for model calibrations given the data related challenges encountered by SHAs. The bootstrap is a nonparametric and robust resampling technique for estimating standard errors and confidence intervals of a statistic. The main advantage of resampling methodologies like bootstrapping includes estimation of parameter without making distribution assumptions. This paper presents the use of resampling techniques to locally calibrate the flexible pavement performance models and how the locally calibrated Pavement-ME models improved the performance prediction accuracy. The results of the local calibration show that the validation standard error (SEE) and bias obtained from bootstrapping were much lower than other resampling techniques. In addition, the validation statistics were similar to that of the model

Haider, Brink, and Buch

calibration, which indicates robustness of the local model coefficients. The reliability equations after local calibration are a better representation of measured pavement performance in Michigan.

Keywords: Local calibrations, fatigue cracking, rutting, IRI, bootstrapping, and jackknifing.

Draft

Haider, Brink, and Buch

INTRODUCTION

The performance prediction models in the Pavement-ME are nationally calibrated using the in-service pavement material properties, pavement structure, climate and truck loading conditions, and performance data obtained from the Long Term Pavement Performance (LTPP) program. Due to the limited availability of Level 1 inputs, the nationally calibrated models are primarily based on Level 2 and Level 3 inputs (AASHTO 2008). Generally, the nationally calibrated models may not perform well if the inputs and performance data used to calibrate those do not represent state design practices. Therefore, it is recommended that each state highway agency (SHA) conduct an evaluation to determine how well the nationally calibrated performance models predict local field performance. The local calibration of the Pavement-ME performance models is recommended to improve the pavement performance prediction capabilities reflecting the unique field conditions and design practices. Such calibration is performed to (a) confirm that the models can predict pavement distress and smoothness with minimum bias, and (b) determine the standard error associated with the prediction equations.

The local calibration of the performance models in the new mechanistic-empirical pavement design guide (MEPDG) is a challenging task, especially due to the deficiencies in the needed data. For the selected set of pavement sections used in models calibration, the data requirements include (a) a wide range of inputs related to traffic, climate, design and material characterization, (b) a reasonable extent and occurrence of measured performance data over time. In addition to data requirements, the split sampling used for model calibration and validation further add to these complications. In the split sampling approach, generally 70% of the data are used for calibration and remaining 30% are utilized for validation. However, most of the States have a limited number of identified pavement sections for the local calibration. Therefore, there

Haider, Brink, and Buch

is a need to employ statistical methodologies that are more efficient and robust for model calibrations given the data related challenges encountered by SHAs.

In model calibration, a non-linear optimization fitting process estimates empirical model coefficients that are evaluated based on the goodness-of-fit criteria. The success of the local calibration depends on the dataset used. The local calibration guide (NCHRP Project 1-40B 2009) suggests using statistical techniques to validate the adequacy of the performance prediction models. Traditional split sampling provides one method to calibrate the performance predictions models. Furthermore, resampling methods such as jackknifing and bootstrapping are recommended because they provide more reliable and robust assessment of the model prediction accuracy than the split sampling methods. Different resampling techniques can improve the confidence in the local calibration coefficients. While traditional split sample approach uses a two-step process for calibration and validation, advanced resampling approaches can simultaneously consider both steps. Moreover, the goodness-of-fit statistics are based on predictions rather than on data used for fitting the model parameters. The efficiency and robustness of such approaches become more important when the sample size is limited. Different resampling methods can be used for the local calibration of the distress models including (a) split sampling, (b) bootstrapping, and (c) jackknifing. The main objectives of this paper are to (a) demonstrate the local calibration process for flexible pavement performance models, and (b) show the benefits of resampling approaches such as bootstrapping and jackknifing for local calibration of the flexible pavement performance models. The local calibration for flexible pavement performance models for alligator cracking, rutting and roughness in terms of international roughness index (IRI) in the State of Michigan are presented in this paper.

Haider, Brink, and Buch

BACKGROUND

The performance models for alligator cracking, rutting and IRI along with the calibration coefficients are summarized below.

Alligator Cracking

Alligator cracking is defined as cracks that initiate at the bottom of the HMA layers and propagate to the surface (bottom-up) with continued truck traffic in the wheel-path. Equation (1) shows the transfer function for bottom-up fatigue cracking in the Pavement-ME and the required calibration coefficients (AASHTO 2008).

$$FC_{Bottom} = \left(\frac{1}{60} \right) \left(\frac{C_4}{1 + e^{C_1 C_1^* + C_2 C_2^* \log(DI_{Bottom} - 100)}} \right) \quad (1)$$

where:

FC_{Bottom} = Area of alligator cracking, percent of total lane area.

DI_{Bottom} = Cumulative damage index at the bottom of the HMA layers.

$C_{1,2,4}$ = Transfer function calibration coefficients

The C_1^* and C_2^* coefficients in Equation(1) can be determined as:

$$C_1^* = -2C_2^* \quad (2)$$

$$C_2^* = -2.40874 - 39.748(1 + H_{HMA})^{-2.856} \quad (3)$$

Surface Rutting

The rutting model predicts the permanent deformation in each layer/sub-layer incrementally for the entire analysis period. The rutting is predicted based on an incremental approach and

Haider, Brink, and Buch

cumulative rutting over time is included in the output. The average vertical resilient strain is computed for each analysis period over the entire design life of the pavement. Equation (4) shows the current rutting model for the HMA layers. The model indicates that there are three local calibration coefficients in this function (i.e., $\beta_{r1}, \beta_{r2}, \beta_{r3}$) Equation (5) shows the rutting model for unbound layers. This model has one calibration coefficient. Since, the rutting is predicted separately for the HMA, base, and subgrade, the total surface rutting is calculated by summing rut depths in all pavement layers [see Equation(6)].

$$\Delta_{p(HMA)} = \varepsilon_{p(HMA)} h_{HMA} = \beta_{r1} k_z \varepsilon_{r(HMA)} 10^{k_{1r}} n^{k_{2r} \beta_{r2}} T^{k_{3r} \beta_{r3}} \quad (4)$$

where:

- $\Delta_{p(HMA)}$ = Accumulated permanent or plastic vertical deformation in the HMA layer/sub-layer, in.
- $\varepsilon_{p(HMA)}$ = Accumulated permanent or plastic axial strain in the HMA layer/sub-layer, in/in.
- $\varepsilon_{r(HMA)}$ = Resilient or elastic strain calculated by the structural response model at the mid-depth of each HMA sub-layer, in/in.
- $h_{(HMA)}$ = Thickness of the HMA layer/sub-layer, in.
- n = Number of axle load repetitions.
- T = Mix or pavement temperature, °F.
- k_z = Depth confinement factor.
- $k_{1r, 2r, 3r}$ = Global field calibration parameters.
- $\beta_{r1}, \beta_{r2}, \beta_{r3}$ = Local or mixture field calibration constants; for the global calibration, these constants were all set to 1.0.

$$\Delta_{Base/Subgrade} = \beta_{s1/sg1} k_{s1} \varepsilon_v h_{soil} \left(\frac{\varepsilon_o}{\varepsilon_r} \right) e^{-\left(\frac{\rho}{n} \right)^\beta} \quad (5)$$

where:

- $\Delta_{Base/Subgrade}$ = Permanent or plastic deformation for the base/subgrade layer, in.

Haider, Brink, and Buch

- n = Number of axle load applications.
- ε_o = Intercept determined from laboratory repeated load permanent deformation tests, in/in.
- ε_r = Resilient strain imposed in laboratory test to obtain material properties ε_o , β , and ρ , in/in.
- ε_v = Average vertical resilient or elastic strain in the layer/sub-layer and calculated by the structural response model, in/in.
- h_{Soil} = Thickness of the unbound layer/sub-layer, in.
- k_{sl} = Global calibration coefficients; $k_{sl}=1.673$ for granular materials and 1.35 for fine-grained materials.
- $\beta_{sl/sgl}$ = Local calibration constant for the rutting in the unbound base and subgrade layers; the local calibration constant was set to 1.0 for the global calibration effort.

$$\text{Rut depth}_{Total} = \Delta_{HMA} + \Delta_{Base/subbase} + \Delta_{Subgrade} \quad (6)$$

Surface Roughness (IRI)

Equation (7) shows the IRI prediction model in the Pavement-ME. There are four calibration coefficients in this model:

$$IRI = IRI_o + C_1(RD) + C_2(FC_{Total}) + C_3(TC) + C_4(SF) \quad (7)$$

where:

- IRI_o = Initial IRI after construction, in/mi.
- RD = Average rut depth, in.
- FC_{Total} = Area of fatigue cracking (combined alligator, longitudinal, and reflection cracking in the wheel path), percent of total lane area. All load related cracks are combined on an area basis – length of cracks is multiplied by 1 foot to convert length into an area basis.
- TC = Length of transverse cracking (including the reflection of transverse cracks in existing HMA pavements), ft/mi.
- SF = Site factor

Haider, Brink, and Buch

$$\begin{array}{l} C_1, C_2, \\ C_3, C_4 \end{array} = \text{Calibration coefficients } C_1 = 40.0, C_2 = 0.400, C_3 = 0.008, C_4 = 0.0150$$

Currently, the following equation is documented in most of the literature for the site factors (SF):

$$SF = Age \left[0.02003 (PI + 1) + 0.007947 (Precip + 1) + 0.000636 (FI + 1) \right] \quad (8)$$

where:

Age = Pavement age, years.

PI = Percent plasticity index of the soil.

FI = Average annual freezing index, degree F days.

$Precip$ = Average annual precipitation or rainfall, in.

However, during the local calibration in Michigan, it was found that the site factor was coded based on the following equations in the Pavement-ME software:

$$SF = (Frost + Swell) \times Age^{1.5} \quad (9)$$

$$Frost = Ln \left[(Rain + 1) \times (FI + 1) \times P_4 \right] \quad (10)$$

$$Swell = Ln \left[(Rain + 1) \times (FI + 1) \times P_{200} \right] \quad (11)$$

where:

SF = Site factor

Age = Pavement age (years)

FI = Freezing index, °F-days.

$Rain$ = Mean annual rainfall (in.)

P_4 = Percent subgrade material passing No. 4 sieve

P_{200} = Percent subgrade material passing No. 200 sieve.

Haider, Brink, and Buch

Several SHAs' have performed local calibration of the flexible pavement performance prediction models. Table 1 summarizes the local calibration coefficients for a sample of States (Banerjee et al. 2009, Glover and Mallela 2009, Hall et al. 2010, Kim et al. 2014, Li et al. 2008, Mallela et al. 2009a, Mallela et al. 2009b, Mallela et al. 2013, Pierce and McGovern 2014, Quintus and Moulthrop 2007, Velasquez et al. 2009). The results of local calibration show that most of the SHAs are relying on the global models, especially for alligator cracking and IRI. The reasons for adopting the global models include: (a) lack of adequate number of pavement sections for local calibration, (b) scarcity of the measured performance data, (c) difficulties in finding appropriate inputs related to pavement cross sections, material properties and traffic, and (d) lower confidence level for locally calibrated models.

In addition, the IRI model is fundamentally different from other distress models (e.g., fatigue cracking and rutting) because of its empirical nature (i.e., regression model) and dependence on the other distress models. Consequently, since IRI is related to construction and generally not adopted as design criteria, this might also be a reason that most of the States are adopting global IRI model.

PROCESS OF MODEL CALIBRATION AND DATA NEEDS

The NCHRP Project 1-40B (NCHRP Project 1-40B 2009) documented the recommended practices for the local calibration of the Pavement-ME performance models. The guide outlines the significance of the calibration process and the general approach for local calibrations of the models. The local calibration process is used to: (a) confirm that the performance models can predict pavement distress and smoothness with minimal bias, and (b) determine the standard error associated with the prediction equations.

Haider, Brink, and Buch

The model standard error shows the scatter of the data around the line of equality between predicted and measured values of distresses. The model bias indicates if there is any consistent under or over-prediction by the prediction models. In general, the local calibration process involves the following steps:

1. Select the appropriate number of pavement sections based on the selection criteria for each performance model. The final list of the candidate pavement sections should be refined based on the magnitude and extent of the measured performance.
2. Collect traffic, climate, pavement cross-section, and materials data for all the selected pavement sections.
3. Execute the Pavement-ME software to predict the pavement performance for each selected pavement section.
4. Extract the predicted distresses and compare with the measured distresses.
5. Test the accuracy of the global model predictions and determine if local calibration is required.
6. Adjust the local calibration coefficients to eliminate bias and reduce standard error by using different sampling and resampling techniques, if local calibration is required. It should be noted that different subset of data representing reconstruction and rehabilitation can be analyzed separately to determine the need for distinct calibration coefficients.
7. Validate the local calibration coefficients with pavement sections not included in the calibration dataset.
8. Modify the reliability equations for each performance model based on the final calibrated models.

Haider, Brink, and Buch

The local calibration process includes the selection of an adequate number of pavement sections representing the state-of-the-practice for local conditions. A subsequent, essential step is to collect the required data for the selected in-service pavement sections. The data includes the information about the measured pavement condition over time, and several Pavement-ME inputs, for each project. The inputs directly affect the performance predictions. These predictions are compared to the measured performance of the as-constructed pavement sections. A pavement section is defined as a specific length of roadway corresponding to a construction project.

Pavement Management Data Needs

The first step in the calibration process was to evaluate the Michigan Department of Transportation (MDOT) Pavement Management System (PMS) database to extract the necessary measured performance data for the model calibrations. The measured performance data were obtained for the identified pavement sections. It is important to ensure that the local PMS distress data are consistent with the outputs of the Pavement-ME. Therefore, unit conversions may be required prior to comparing the measured and predicted distress. The MDOT PMS distresses were compared to the Pavement-ME predicted distresses. Only bottom-up fatigue cracking requires unit conversion. Bottom-up cracking is defined as alligator cracking in the wheel-path. The MDOT PMS records alligator cracking in miles within wheel-paths. The conversion from miles to percent lane area is shown in Equation (12):

$$\%AC_{bottom-up} = \frac{\text{Length of cracking (miles)} \times \text{width of wheelpaths (feet)}}{\text{Length of project (miles)} \times \text{Lane width (feet)}} \times 100 \quad (12)$$

The widths of each wheel-path and lane were assumed to be 3 and 12 feet, respectively, as recommended by the LTPP distress identification manual (Miller and Bellinger 2003). It should be noted that the bottom-up and top-down fatigue cracking in the wheel-paths (6 feet) were

Haider, Brink, and Buch

combined due to the difficulty in determining the crack initiation at the surface (top or bottom) for the observed fatigue cracking.

The rutting model in the Pavement-ME predicts rut depth for the HMA, base, and subgrade layers in the pavement structure. The prediction models are different for each layer and the sum of the rut depth in three layers represents the total surface rutting. Since the rutting models are different for bound (HMA) and unbound layers (base, subbase, and subgrade), the measured rutting for each layer is needed for calibration. The MDOT PMS database only reports total surface rutting. Alternatively, due to lack of available trench data, the transverse profiles for the selected pavement sections were utilized to estimate the rutting in the HMA, base and subgrade layers. Transverse profile data were obtained for the selected pavement sections to estimate the surface rutting layer contributions. Analyses of transverse profiles assist in estimating the seat of rutting and the layer contributions to the total surface rutting. The width and depth of the measured rut channel can be used to determine the seat of rutting. A study sponsored by the National Cooperative Highway Research Program (NCHRP) developed a procedure to determine the seat of rutting from the transverse profiles (White et al. 2002). The magnitude of the individual layer rutting was determined based on the contribution percentage and the measured rut depth. The results indicate that on average; about 82% of the transverse profiles indicated seat of rutting in the HMA layer, 12% on the base/subbase and 6% in the subgrade layers. It should be noted that destructive testing such as trenching were not performed to verify the results of the transverse profile analysis.

Adequate Number of Pavement Sections

The next step consists of determining the adequate number of pavement sections for local calibration based on the statistical requirements. The local calibration guide (NCHRP Project 1-

Haider, Brink, and Buch

40B 2009) suggests a method to determine the minimum number of sections for each performance measure. The minimum number of sections is calculated using Equation (13) and the results are summarized in Table 2 for the flexible pavement performance measure.

$$n = \left(\frac{Z_{\alpha/2} \times \sigma}{e_t} \right)^2 \quad (13)$$

where:

$Z_{\alpha/2}$ = The two tail z-value from a standard normal distribution, $Z_{0.1/2} = 1.64$

n = Minimum number of pavement sections

σ = Performance threshold

e_t = Tolerable bias = $Z_{\alpha/2} \times \text{global model}_{\text{SEE}}$

Pavement Sections Identification

The pavement sections were identified to represent the current local construction practices in Michigan. Both reconstruct and rehabilitation pavement sections were considered for the local calibration. The flexible pavement types include new hot-mix asphalt (HMA) reconstruct, HMA overlay over existing HMA, HMA overlay over rubblized concrete pavements, HMA crush and shape, and HMA over intact concrete pavements. The pavement sections were selected based on the following experiment design for each pavement construction type. The factors and levels were selected to represent state-wide construction practices:

- *Site factors*: The site factors addressed seven regions, climate and subgrade soil types in the state.

Haider, Brink, and Buch

- *Traffic*: Three traffic categories were selected; < 1000 , 1000 to 3000 , and > 3000 AADTT. The three levels were selected based on pavement class, trunk routes, US routes and Interstate routes.
- *Surface Layer Thicknesses*: Three levels of HMA thickness levels were considered; < 3 , 3 to 7 , and > 7 inches. These ranges covered new flexible pavement and overlay construction practices in Michigan.
- *Pavement Age*: Three pavement age levels were selected; < 10 , 10 to 15 , and > 15 years.

Based on the experiment design, 169 flexible reconstruct and rehabilitation pavement sections were identified for the local calibration. The measured performance was extracted from the PMS database and analysed for consistency and to identify any abnormal performing pavement section.

Selection of Pavement Sections

The measured performance for each pavement section was compared to the FHWA expected pavement performance criteria (Rauhut et al. 1999). For a robust local calibration, the distress magnitude should cover a range above and below the threshold limits. There is a cause for concern if a pavement section exceeds the expected performance at an early age. The Pavement-ME will not capture such premature failure and this causes significant differences between the measured and predicted distress. Figure 1 shows the measured alligator cracking, rutting and IRI for the selected pavement sections. As shown in the figure, several pavement sections exceeded the expected alligator cracking performance at an early age. The rutting and IRI performance followed expected trends except for a few pavement sections. The pavement sections which

Haider, Brink, and Buch

exceeded the expected behavior at an early age (0-5 years) were excluded if the cause for such behavior was unknown. Such performance behavior can be attributed to many factors such as construction and measurement errors. However, the pavement sections which exhibited high levels of distresses in medium-term and cause of such behaviors were not available; those sections were kept in the calibration dataset.

Inputs Data Collection

The next step after pavement section selection consists of collecting the input data required to characterize the pavement sections in the Pavement-ME software. The accuracy of the input data directly impacts the performance predictions. The Pavement-ME uses a large number of inputs to characterize a pavement. Furthermore, the hierarchical structure of the Pavement-ME provides three levels for the important inputs. While the input data collection efforts were very time consuming, priority was given to the most sensitive inputs which affect the pavement performance predictions to make the process more efficient. The sensitive inputs which affect the performance prediction for new and rehabilitation design of flexible pavements were identified from previous studies (Brink et al. 2013, Buch et al. 2008, Buch et al. 2013, Harsini et al. 2013, Schwartz et al. 2011). Additionally, the best available hierarchical input level was used to characterize the selected pavement sections. All input data reflect Michigan construction practices and were obtained from either, construction records, or results from previous research studies. Table 3 summarizes the inputs, corresponding input levels and sources for the available data used.

Haider, Brink, and Buch

Local Calibration of the Performance Models

The adequacy of the global model predictions is determined through performing three hypothesis tests. The hypothesis tests provide an indication if the models are biased. Bias is defined as the consistent under- or over-prediction of distress or IRI. The bias between measured and predicted distress/IRI is determined by performing linear regression, hypothesis tests and a paired *t*-test using a significance level of 0.05. Figure 2 shows a representation of model bias and standard error for various conditions. The model bias and standard error were determined by using Equations (14) and (15):

$$Bias = \frac{\sum (\text{Predicted} - \text{Measured})}{n} \quad (14)$$

$$SEE = \sqrt{\frac{\sum (\text{Predicted} - \text{Measured})^2}{n - 2}} \quad (15)$$

Where *n* represents the total number of time series distress measurements. The next section presents a brief description for various sampling techniques.

Traditional Approach

The NCHRP Project 1-40B recommended split-sampling for local calibration of the Pavement-ME performance models (NCHRP Project 1-40B 2009). The approach is termed as the traditional approach in this paper. This two-step approach includes splitting the data into calibration and validation subsets. Typically, 70% and 30% splits are used for calibration and validation, respectively. This calibration-validation approach is dependent on the number of pavement sections in calibration dataset. A large enough sample size (i.e., > 30 pavement sections) is needed to have more confidence on the calibration and validation results. The model validation procedure is used to demonstrate that the calibrated model can produce accurate

Haider, Brink, and Buch

predictions of pavement distress for sections other than the ones used for calibration. The success of the validation process can be determined based on the model bias in the predicted values and the model standard error. Statistical hypotheses tests should be performed to determine if a significant difference between the calibrated model and the model validation exists (NCHRP Project 1-40B 2007, 2009).

Bootstrapping

At its simplest, for a dataset with a sample size of N , B "bootstrap" samples of size N are randomly selected with replacement from the original dataset. The bootstrap can be used to estimate variances, confidence intervals and other statistical properties of population from a sample. However, these properties are obtained by drawing samples from a sample. Each bootstrap sample typically omits several observations and has multiple copies of others. This procedure selects the new bootstrap samples "with replacement" and hence, there is an equal chance of selecting each sample multiple times. Validity of bootstrap variance estimates requires that a key analogy holds for the statistics of interest. The resampling properties of the bootstrap must be similar to the sampling properties of the population. The bootstrapped resampling can be performed using different methods: (a) resampling randomly or (b) resampling based on the residuals. The appropriate choice of resampling type for bootstrapping depends on the data structure. For example, if fixed regressors are needed for an experiment design, the bootstrapping can be performed using the residuals (error-terms). On the other hand, if the random effects of regressors are to be estimated in a regression model, random resampling can be employed in bootstrapping. Bootstrapping based on resampling the observations is an approach that is applied when the regression models are built from data that have random regressors and responses (Fox 2002, Sahinler and Topuz 2007).

Haider, Brink, and Buch

Jackknifing

Jackknifing is a resampling procedure for refining and confirming the calibration coefficients of a regression model. The model validation statistics are developed independently of the data used for calibration. Multiple jackknifing is used to assess the sensitivity of the validation goodness-of-fit statistics. To develop jackknife statistics from a sample of N sets of pavement section, the data are divided into two groups, one part for calibration and the other for validation. For an $N-1$ jackknife validation the procedure begins by removing one pavement section from the data and calibrating the model with the remaining $N-1$ pavement sections. The pavement section that was withheld is then used to predict the pavement performance, from which the error is computed as the difference between the predicted and measured values of the performance measure. A second pavement section is removed while replacing the first one and the new $N-1$ pavement sections are used to calibrate the new model and calculate error for the withheld pavement section. The procedure is repeated until all N pavement sections have been used for validation. This yields N values of the error, from which the jackknifing goodness-of-fit statistic can be computed. The jackknifed errors are computed from the pavement sections that were not used in calibrating the model coefficients. Thus, the jackknifing goodness-of-fit statistics are considered to be independent measures of model accuracy (Quintus et al. 2003).

When comparing jackknife with bootstrap procedures, jackknife is less demanding computationally than the bootstrap and relies on dividing the sample observations into disjoint subsets, each having the same number of observations. In addition, the jackknife takes a fundamentally different view of the possible replicates of the statistic; it treats them as a finite collection whereas bootstrap resampling assumes that the replicates are a sample from the population of infinite size. When bootstrap is used to estimate the standard error of a statistic, it

Haider, Brink, and Buch

gives slightly different results when repeated on the same data, whereas the jackknife gives exactly the same result each time. The bootstrap is a more general technique and preferred to the jackknife.

In summary, bootstrap and jackknife are nonparametric and robust resampling techniques for estimating standard errors and confidence intervals of a population parameter such as mean, median, proportion, odds ratio, or regression coefficients. The main advantage of these techniques, especially bootstrapping, is that estimation of parameters are possible without making distribution assumptions. In addition, the approach is valid when such assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the standard errors estimation. On the other hand, there are several limitations of bootstrap method, especially when (a) using the methods for small data sets with outliers and (b) adopting bootstrap for time series data when the independent assumption is violated. Therefore, in order to address the limitations, the pavement sections were resampled instead of the individual distress measurements to maintain the time series data structure. It should be noted that the estimate of parameters by using bootstrapping is robust since it can be used to estimate the confidence interval (in this case 95% CI) for each coefficient, SEE and bias of the models. Therefore, the estimates are much better than just one value. However, there are caveats and conditions for obtaining the reliable and accurate results by using the resampling techniques like bootstrapping. If the same size is limited and doesn't represent the population, the bootstrap estimates are not accurate. In addition, presence of outliers and high variability in the data can't be addressed by these techniques like any other statistical analysis methodology.

LOCAL CALIBRATION RESULTS AND DISCUSSIONS

The calibration process adopted for Michigan includes several data subsets (options) and re-

Haider, Brink, and Buch

sampling methods. This paper presents the local calibration results for only combined reconstruct and rehabilitation pavement sections.

The performance prediction models were locally calibrated by minimizing the sum of squared errors (SSE) between the measured and predicted distresses by using the following sampling methods (Haider et al. 2014):

1. No sampling (includes all data i.e., all sections were used to estimate calibration coeffs.)
2. Traditional split sampling (random selection of 70% sections for calibration and 30% for validation)
3. Repeated split sampling (1000 random resamples for method 2)
4. Bootstrapping (1000 random resamples with replacement)
5. Jackknifing ($n-1$ samples for calibration and 1 sample for validation, resample n times)

The different resampling techniques were used to sample data and then determine the best estimates of the local calibration coefficients along with the associated model SEE and bias.

These techniques were considered to: (a) address data limitations, especially due to limited sample size, and (b) employ more robust ways of quantifying model standard error and bias.

Comparisons of the SEE and bias between all the resampling techniques are presented in Figure

3 for the alligator cracking, rutting, and IRI models, respectively. The results show that the

bootstrapping provided the lowest or somewhat similar SEE compared to all other resampling

techniques. Additionally, it is expected that bootstrap resampling provides the lowest SEE

because it can account for the variability within the dataset (random sampling with replacement).

Since, the calibration was performed on a large enough sample size; the bootstrapping technique

did not improve the SEE and bias among different sampling approaches as expected. However, it

is anticipated that there will be significant difference in SEE and bias among resampling

Haider, Brink, and Buch

approaches for a smaller sample size. Such results were observed for local calibration of the rigid pavement sections in Michigan (Haider et al. 2014).

As an example, the bootstrap and jackknifing model calibration coefficients and statistics distributions are shown for the rutting model in Figure 4. The results show the variability of the SEE, bias and calibration coefficients as well as the 95% confidence interval (CI). As discussed previously, the bootstrapping technique considers multiple random selections (with replacement) of the pavement sections. However, if multiple samples are taken, the SEE, bias, β_{rI} , β_{sI} , and β_{sgI} values will vary for each realization. Traditional split sampling partitions the dataset only once for calibration and validation. Therefore, the results of a single split sample may not indicate an accurate representation of all the sections on average, especially when the sample size is limited. In order to address this limitation, bootstrap resampling was performed 1000 times to study the variability of the calibration parameters. Figure 4(a) shows that the SEE ranged between 0.06 and 0.1 inches, while the bias ranged from -0.02 to -0.005 inches. The β_{rI} calibration coefficient ranged between 0.8 and 1.1, β_{sI} ranged between 0.05 and 0.16 and, β_{sgI} varied between 0.01 and 0.06. Figure 4(b) illustrates the parameter distributions for the jackknifing. For this technique, the SEE varied from 0.0857 and 0.0864 and the bias ranged between -0.0155 and -0.01528. The β_{rI} calibration coefficient varied between 0.939 and 0.944, β_{sI} varied between 0.0972 and 0.0979 and, β_{sgI} varied between 0.0365 and 0.0367. It should be noted that jackknifing only considers $(N-1)$ samples which limits the total number of realizations to the overall sample size. The jackknife distributions do not follow a normal distribution and shows a skewed distribution for SEE and bias. This difference occurs because of the fundamental differences between the resampling needs for bootstrapping and jackknifing.

Haider, Brink, and Buch

The locally calibrated performance models should be able to predict pavement performance for the projects which were not included in the calibration. This process is necessary to validate the robustness, accuracy and adequacy of the models predictions. Therefore, considering the robustness and efficiency of bootstrapping, the original dataset was divided into two subsets. While 80% of pavement sections were used to calibrate the models, 20% were later utilized for validation. A total of 1000 bootstrap samples were obtained for 80% of the sections to estimate expected model coefficients and then those coefficients were used to validate the predicted performance for the remaining 20% of the sections.

Figure 5 shows the goodness-of-fit for locally calibrated model validation and verifies that the model coefficients obtained from bootstrap resampling are able to predict pavement performance using pavement sections which were not included in the local calibration. The model validation SEE and bias are similar to those of the local model calibration. The results indicate that the local model coefficients can accurately predict cracking, rutting and IRI for pavement sections which were not included in the local calibration.

In order to test the hypothesis that the difference between predicted vs. measured is significant, paired t -tests ($H_o : \mu_p - \mu_m = 0 ; H_a : \mu_p - \mu_m \neq 0$) were performed for all resampling techniques within each model. If the null hypothesis is rejected for the hypothesis tests, it indicates that the model may still have some bias and predicted and measured performance is significantly different from each other. On the other hand, if the null hypothesis cannot be rejected (i.e., $p - value > 0.05$), it shows that there is no significant difference between predicted and measure performance. Table 4 summarizes the hypothesis test results for the bootstrap validation. The results show that there is a significant difference between the measured and predicted distress for the global model. The local model validation results show that there is no

Haider, Brink, and Buch

significant difference between the measured and predicted distress for rutting and IRI. However, there is still a significant difference for alligator cracking. Even though the hypothesis tests are rejected, the SEE and bias is lower as compared to the global model and gives a better prediction using local design practices. Table 5 summarizes the recommended local calibration coefficients for the flexible pavement performance prediction models in Michigan. The reliability of each model was updated to reflect the changes due to local calibration. Figure 6 shows an example of residual plots for global and local rutting models. The residuals can be used to validate the assumptions (e.g. linearity and constant variance) of regression models. The global rutting model residuals clearly show the violation of regression assumptions and indicate a significant bias between predicted and measured rutting. On the other hand, as shown in Figure 6(b) the local calibration of the rutting model significantly reduced bias and distributes the residual in a random fashion. Similar results were observed for IRI model. However, the residual distribution for the local alligator cracking was a little skewed but much better than that for the global model.

The reliability of the trial design is dependent on the model prediction error (standard error) of the distress prediction equations. Therefore, the reliability equations (i.e., the relationship between the mean predicted distress and the standard deviation of the measured distress) will need modifications to incorporate the impact of the local calibration. There are two main methods to determine the design reliability for various performance models flexible pavements:

- Method 1: Reliability determined based on the relationship between the mean predicted distress and standard deviation of the measured distress.
- Method 2: Reliability based on the standard error by using the variance approach.

Haider, Brink, and Buch

The alligator cracking and rutting model reliability was determined based on method 1 while method 2 was adopted for the IRI model. Figure 7 presents the differences in standard errors between global and local models for alligator cracking and rutting models for reliability determination. The comparison between the standard errors for alligator cracking model show a gradual change in the function. From the design point of view, the local model will predict higher cracking as compared to the global model, especially at lower and higher levels of fatigue damage and thus will be more conservative for design. These results are expected based on the variability of measured alligator cracking on Michigan pavements. On the other hand, the global model will predict much higher rutting than local model and thus will result in more conservative design based on rutting reliability. Much lower rutting levels were observed; therefore, the local reliability equations are a better representation for the pavement sections in Michigan. These results of local calibrations show that most of the flexible pavement designs will be controlled by alligator cracking in Michigan.

CONCLUSIONS

The evaluation of the Pavement-ME global performance models showed that there is a need for local calibration for Michigan pavements. This paper presents the use of resampling techniques to locally calibrate the performance models for flexible pavements. The locally calibrated Pavement-ME models improved performance prediction accuracy. The results of the local calibration show that SEE and bias obtained from bootstrapping were somewhat similar among resampling techniques. However, the bootstrap validation model statistics were similar to that of the calibration, which indicates robustness of the local model coefficients. Since, bootstrap estimates of the model coefficients are more robust, the design reliability will also be more representative for the local conditions provided the sample size is adequate.

Haider, Brink, and Buch

From the design point of view, the local alligator cracking model will predict higher cracking as compared to the global model and thus will be more conservative. The local rutting model will predict lower amount of rutting as compared to the global model based on the observed rutting performance in Michigan. Therefore, the local reliability equations are a better representation for the pavement sections based on observed pavement performance in Michigan. These results of local calibrations show that most of the flexible pavement designs will be controlled by alligator cracking in Michigan.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Michigan Department of Transportation (MDOT) for funding the study. The authors also want to acknowledge Mr. Leslie Titus-Glover of ARA for providing insight related to the IRI model.

REFERENCES

- AASHTO. 2008. Mechanistic-Empirical Pavement Design Guide: A Manual of Practice: Interim Edition. American Association of State Highway and Transportation Officials.
- Banerjee, A., Aguiar-Moya, J.P., and Prozzi, J.A. 2009. Texas Experience using LTPP for Calibration of the MEPDG Permanent Deformation Models. Transportation Research Record.
- Brink, W., Harsini, I., Haider, S.W., Chatti, K., Buch, N., Baladi, G., and Kutay, E. 2013. Sensitivity of Input Variables for Rigid Pavement Rehabilitation Strategies in the MEPDG. ASCE T&DI, Congress Airfield and Highway Pavements: 528-538.
- Buch, N., Chatti, K., Haider, S.W., and Manik, A. 2008. Evaluation of the 1-37A Design Process for New and Rehabilitated JPCP and HMA Pavements, Final Report Research Report RC-1516. Michigan Department of Transportation, Construction and Technology Division, P.O. Box 30049, Lansing, MI 48909, Lansing, .

Haider, Brink, and Buch

Buch, N., Haider, S.W., Chatti, K., Baladi, G.Y., Brink, W., and Harsini, I. 2013. Preparation for Implementation of the Mechanistic-Empirical Pavement Design Guide in Michigan - Part 2: Rehabilitation Evaluation. Michigan Department of Transportation.

Fox, J. 2002. Bootstrapping regression models. An R and S-PLUS Companion to Applied Regression: A Web Appendix to the Book. Sage, Thousand Oaks, CA. URL <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-bootstrapping.pdf>.

Glover, L.T., and Mallela, J. 2009. Guidelines for Implementing NCHRP 1-37A M-E Design Procedures in Ohio: Volume 4 - MEPDG Models Validation & Recalibration. Ohio Department of Transportation.

Haider, S.W., Buch, N., Brink, W., Chatti, K., and Baladi, G.Y. 2014. Preparation for Implementation of the Mechanistic-Empirical Pavement Design Guide in Michigan - Part 3: Local Calibration and Validation of the Pavement-ME Performance Models Final Report RC-1595. Michigan Department of Transportation.

Hall, K.D., Xiao, D.X., and Wang, K.C.P. 2010. Calibration of the MEPDG for Flexible Pavement Design in Arkansas. Transportation Research Record.

Harsini, I., Brink, W., Haider, S.W., Chatti, K., Buch, N., Baladi, G., and Kutay, E. 2013. Sensitivity of Input Variables for Flexible Pavement Rehabilitation Strategies in the M-E PDG, Los Angeles, California, pp. 539-550.

Kim, S., Ceylan, H., Ma, D., and Gopalakrishnan, K. 2014. Calibration of Pavement ME Design and Mechanistic-Empirical Pavement Design Guide Performance Prediction Models for Iowa Pavement Systems. Journal of Transportation Engineering **140**(10): 04014052.

Li, J., Pierce, L.M., and Uhlmeier, J. 2008. Calibration of the Flexible Pavement Portion of the Mechanistic-Empirical Pavement Design Guide for the Washington State Department of Transportation. Transportation Research Record.

Mallela, J., Titus-Glover, L., Quintus, H.V., Darter, M.I., Stanley, M., and Rao, C. 2009a. Implementing the AASHTO Mechanistic-Empirical Pavement Design Guide in Missouri Volume II: MEPDG Model Validation and Calibration. Missouri Department of Transportation.

Haider, Brink, and Buch

Mallela, J., Titus-Glover, L., Quintus, H.V., Darter, M.I., Stanley, M., Rao, C., and Sadasivam, S. 2009b. Implementing the AASHTO Mechanistic-Empirical Pavement Design Guide in Missouri Volume I: Study Findings, Conclusions, and Recommendations. Missouri Department of Transportation.

Mallela, J., Titus-Glover, L., Sadasivam, S., Bhattacharya, B., Darter, M.I., and Quintus, H.L.V. 2013. Implementation of the AASHTO Mechanistic-Empirical Pavement Design Guide for Colorado. Colorado Department of Transportation.

Miller, J.S., and Bellinger, W.Y. 2003. Distress Identification Manual. FHWA-RD-03-031.

NCHRP Project 1-40B. 2007. User Manual and Local Calibration Guide for the Mechanistic-Empirical Pavement Design Guide and Software In Progress. NCHRP.

NCHRP Project 1-40B. 2009. Local Calibration Guidance for the Recommended Guide for Mechanistic-Empirical Pavement Design of New and Rehabilitated Pavement Structures. Final NCHRP Report.

Pierce, L.M., and McGovern, G. 2014. Implementation of the AASHTO Mechanistic-Empirical Pavement Design Guide and Software NCHRP Synthesis 457 NCHRP Synthesis 457

Quintus, H.V., and Moulthrop, J.S. 2007. Mechanistic-Empirical Pavement Design Guide Flexible Pavement Performance Prediction Models for Montana: Volume I Executive Research Summary FHWA/MT-07-008/8158-1. The State of Montana Department of Transportation,.

Quintus, H.V., Schwartz, C., McCuen, R.H., and Andrei, D. 2003. Jackknife Testing - An Experiment Approach to Refine Model Calibration and Validation Research Results Digest 283. NCHRP.

Rauhut, J.B., Eltahan, A., and Simpson, A.L. 1999. Common Characteristics of Good and Poorly Performing AC Pavements FHWA-RD-99-193. Federal Highway Administration,.

Sahinler, S., and Topuz, D. 2007. Bootstrap and jackknife resampling algorithms for estimation of regression parameters. Journal of Applied Quantitative Methods 2(2): 188-199.

Haider, Brink, and Buch

Schwartz, C.W., Li, R., Kim, S., Ceylan, H., and Gopalakrishnan, R. 2011. Sensitivity Evaluation of MEPDG Performance Prediction NCHRP 1-47 Final Report.

Velasquez, R., Hoegh, K., Yut, I., Funk, N., Cochran, G., Marasteanu, M., and Khazanovich, L. 2009. Implementation of the MEPDG for New and Rehabilitated Pavement Structures for Design of Concrete and Asphalt Pavements in Minnesota. Minnesota Department of Transportation.

White, T.D., John, E.H., Adam, J.T.H., and Hongbing, F. 2002. Contribution of Pavement Structural Layers to Rutting of Hot Mix Asphalt Pavements Report 468. NCHRP, Washington, DC.

Draft

Haider, Brink, and Buch

Table 1 Local calibration model coefficients for different states

Model	Coeff.	Global	AZ	CO	MO	OR	AR	NM	WA	MT	NC	OH	TX	IA
Bottom-up Cracking	C_1	1	1	0.07	1	0.56	0.688	0.625	1.071	1	1	1	1	1
	C_2	1	4.5	2.35	1	0.225	0.294	0.25	1	1	1	1	1	1
	C_4	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000
Rutting HMA	β_{r1}	1	0.69	1.34	1	1.48	1.2	1.1	1.05	1	13.1	0.51	2.39	1
	β_{r2}	1	1	1	1	1	1	1.1	1.109	1	0.4	1	1	1.15
	β_{r3}	1	1	1	1	0.9	0.8	0.8	1.1	1	1.4	1	0.856	1
Rutting base	β_{s1}	1	0.37	0.84	0.437	1	1	0.8	-	1	0.303	0.32	1	0.001
Rutting subgrade	β_{sg1}	1	0.14	0.4	0.01	1	0.5	1.2	0	1	1.102	0.33	0.5	0.001
IRI	C_1	40	1.228	35	17.7	40	40	0.015	40	40	40	17.6	40	40
	C_2	0.4	0.117	0.3	0.975	0.4	0.4	0.4	0.4	0.4	0.4	0.01	0.4	0.4
	C_3	0.008	0.008	0.02	0.008	0.008	0.008	0.008	0.008	0.008	0.008	1.37	0.008	0.008
	C_4	0.015	0.028	0.019	0.01	0.015	0.015	0.015	0.015	0.015	0.015	0.066	0.015	0.015

Haider, Brink, and Buch

Table 2 Minimum number of sections for local calibration in Michigan

Performance Model	Nationally calibrated SEE	$Z_{\alpha/2}$	Threshold ³	N (required number of sections)	Number of sections used	Number of distress measurements
Alligator cracking (%)	5.01	1.64	20	16	121	344
Rutting (in)	0.107		0.5	22	162 ¹	577
IRI (in/mile)	18.9		172	83	167 ²	468

¹129 reconstruct and 33 rehabilitated pavement sections, ²127 reconstruct and 40 rehabilitated pavement sections

³ The performance threshold value is dependent on distress type and local practices of a highway agency.

N = minimum number of samples required at 90% confidence level

Draft

Haider, Brink, and Buch

Table 3 Summary of input levels and data source

Input			Input level	Input source
Traffic	AADTT		1	MDOT Historical Traffic counts
	TTC		2	Cluster analysis
	ALS Tandem		2	Cluster analysis
	HDF		2	Cluster analysis
	MDF		3	MDOT traffic characterization study
	AGPV		3	MDOT traffic characterization study
	ALS single, tridem, quad		3	MDOT traffic characterization study
Cross-section (new and existing)	HMA thickness		1	Project specific HMA thicknesses based on design drawings
	Base thickness		1	Project specific base thicknesses based on design drawings
	Subbase thickness		1	Project specific subbase thicknesses based on design drawings
Construction materials	HMA	Binder type	3	Project specific binder and mixture gradation data obtained from data collection
		HMA mixture aggregate gradation	3	Project specific binder and mixture gradation data obtained from data collection
		Binder type	1	Pseudo Level 1- MDOT HMA mixture characterization study
		HMA mixture aggregate gradation	1	Pseudo Level 1- MDOT HMA mixture characterization study
	Base/subbase	MR	2	Recommendations from MDOT unbound material study
	Subgrade	MR	2	Soil specific MR values - MDOT subgrade soil study
		Soil type	1	Location based soil type - MDOT subgrade soil study
Climate			1	Closest available climate station

Note:

1. Level 1 is project specific data, pseudo level 1 means that the inputs are not project specific but the material properties (lab measured) corresponds to similar materials used in the project
2. Level 2 inputs are based on regional averages in Michigan
3. Level 3 inputs are based on statewide averages in Michigan

Haider, Brink, and Buch

Table 4 Bootstrap validation t-test results

Model	Global p -value	Local p -value	Validation p -value
Alligator cracking	0.00	0.05 ¹	0.02
Rutting	0.00	0.05 ²	0.78
IRI	0.00	0.05 ³	0.96

Note: ¹58% of the iterations showed p -values > 0.05
²80% of the iterations showed p -values > 0.05
³97% of the iterations showed p -values > 0.05

Draft

Haider, Brink, and Buch

Table 5 Summary of locally calibrated flexible pavement performance models in Michigan

Performance prediction model		Reliability	Local coefficient
Fatigue cracking		$S_{e(Alligator)} = 0.7874 + \frac{17.817}{1 + e^{0.0699 - 0.4559 \times \log(D)}}$	$C_1 = 0.50$ $C_2 = 0.56$
Rutting	HMA	$S_{e(HMA)} = 0.1126(Rut_{HMA})^{0.2352}$	$\beta_{1r} = 0.948$ $\beta_{2r} = 1.3$ $\beta_{3r} = 0.7$
	Base/subgrade	$S_{e(base)} = 0.1145(Rut_{base})^{0.3907}$ $S_{e(subgrade)} = 3.6118(Rut_{subgrade})^{1.0951}$	$\beta_{s1} = 0.094$ $\beta_{sg1} = 0.037$
IRI		Internally estimated by the software	$C_1 = 32.300, C_2 = 0.404$ $C_3 = 0.006, C_4 = 0.016$

Haider, Brink, and Buch

List of Figures

Fig. 1 Comparison between the measured and expected pavement performance

Fig. 2 Schematic of bias and standard error for model calibration

Fig. 3 Local calibration SEE and bias results for various sampling techniques

Fig. 4 Bootstrapping and jackknifing distributions for rutting model

Fig. 5 Measured versus predicted model validation

Fig. 6 Residuals comparison between global and locally calibrated models

Fig. 7 Comparison of standard error between global and local model for reliability

Draft

Fig. 1 Comparison between the measured and expected pavement performance

Fig. 2 Schematic of bias and standard error for model calibration

Fig. 3 Local calibration SEE and bias results for various sampling techniques

Fig. 4 Bootstrapping and jackknifing distributions for rutting model

(a) Alligator cracking

(b) Rutting

(c) IRI

Fig. 5 Measured versus predicted model validation

Fig. 6 Residuals comparison between global and locally calibrated models

Fig. 7 Comparison of standard error between global and local model for reliability