01 Jan 2006

Improving Prediction Accuracy in Mechanistic-Empirical Pavement Design Guide

Scott Schram

Magdy Abdelrahman
Missouri University of Science and Technology, abdelrahmanm@mst.edu

Recommended Citation

S. Schram and M. Abdelrahman, "Improving Prediction Accuracy in Mechanistic-Empirical Pavement Design Guide," *Transportation Research Record*, no. 1947, pp. 59-68, National Research Council (U.S.), Jan 2006.
The definitive version is available at https://doi.org/10.3141/1947-06

Improving Prediction Accuracy in Mechanistic–Empirical Pavement Design Guide

Scott Schram and Magdy Abdelrahman

a b a a a a a a
a a a a a
a a b a a b a
a ( ) a aba a a
a a a a a a b
a b a a a a a
a b a a a a
a b b a a a
a a a a a a a b a
a a
a a a a b a a a a a a a
a a b a a
b a a a b a
a a a a a b a
a a b a

Researchers have invested a great deal of time and effort in the default calibration of over 20 performance prediction models in the mechanistic–empirical design guide. Data sets from the Long-Term Pavement Performance (LTPP) database used in the calibrations not only spanned the climatic conditions of several states, including some Canadian provinces, but also encompassed wide ranges of traffic, materials, and surface layer thicknesses (*1*). Project-level conditions boast unique characteristics that may be diluted in a network-level calibration and allow the prediction accuracy to suffer during local design and analysis.

Prediction accuracy weighs heavily on the model's functional form and the data occupied during its calibration. Data sets containing pavement design characteristics and performance ratings that best represent the structure being designed will yield the calibrated model providing the maximum prediction accuracy. This study isolates traffic levels and surface layer thicknesses to form focused data subsets from which to perform local calibrations. The primary objective of this paper was to explore the effects of focus calibration on distress prediction accuracy. For the purpose of this study, focus calibration was defined as calibration performed using subsets within a complete data set categorized by a focusing parameter. The focusing parameters in this study are average daily truck traffic (ADTT) and thickness. The models for jointed plain concrete pavement (JPCP) smoothness prediction and hot-mix asphalt (HMA) overlay of rigid pavements smoothness prediction acted as test models.

This study calibrates the test models to default, local network-level, and local project-level conditions and compares the resulting prediction accuracies. Local data were focused by categorizing surface layer thickness within ADTT levels.

MODEL CALIBRATION IN DESIGN GUIDE

As previously mentioned, more than 20 models were calibrated with performance data from the LTPP database (*1*). Once the models' functional forms were finalized, LTPP pavement sections were chosen to be included in the calibration process. An iterative calibration procedure developed default calibration coefficients for the mechanistic–empirical models in the design guide. The LTPP database provided as-built cross-section geometrics, material data, distress data, traffic design conditions, and climatic data for each pavement section used in the calibration. Each section was essentially redesigned with the known design inputs entered into the design guide software for analysis. On completion of each design run, predicted distress values were gathered. Once predicted distresses were obtained for each pavement section, the process was repeated, with calibration coefficients altered between iterations. The set of coefficients yielding the minimum error between observed and predicted distress values served as the default parameters used in the design guide (*2*).

Because the design guide software is highly complex, computer run times can reach 90 s per design year for flexible pavements and 30 s per design year for rigid pavements. To display the immense time investment involved in model calibration for the design guide, more than 1,600 h of computer run time was needed to calibrate the flexible permanent deformation model with 88 new flexible pavement sections (*3*).

TEST MODELS USED IN CALIBRATION

Considering the high demands of time and effort involved in model calibration, new and rehabilitated pavement performance models were each calibrated once with data sets that enveloped all ranges of traffic and thicknesses. To assess the effects of focus calibration in model prediction accuracy, the aforementioned calibration effort would have to be repeated several times. To complete this prediction accuracy study, the models for JPCP smoothness and HMA overlays of rigid pavement smoothness were selected as test models. The main advantage of choosing such models lies in their purely empirical makeup.

Department of Civil Engineering, North Dakota State University, CIE 201F, Fargo, ND 58105.

Transportation Research Record: Journal of the Transportation Research Board, No. 1947, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 59–68.

International Roughness Index

AASHTO has used pavement serviceability as the sole performance criterion in pavement design since its inception following the AASHO Road Test (*4*). In 1982 the international roughness index (IRI) was developed to provide a universal way of measuring roughness in pavements (*5*). Since then it has been adopted as a performance-monitoring standard by FHWA and has replaced the present serviceability index in the new design guide (*1*).

JPCP IRI Model

Key distresses recognized as influencing pavement roughness were used to develop the JPCP IRI model in the design guide. Each of the key distresses (cracking, spalling, and faulting) is predicted by using mechanistic–empirical models. To relate the distresses to the climatic conditions and subgrade material, a site factor was introduced. The site factor considers the freezing index and the percent of material passing the No. 200 sieve. Also influencing pavement smoothness over time is the initial as-constructed IRI. The model chosen for use in the mechanistic–empirical pavement design guide is shown in its calibrated form (*6*):

$$\mathrm{IRI} = \mathrm{IRI}_I + C1 * \mathrm{crk} + C2 * \mathrm{spall} + C3 * \mathrm{tfault} + C4 * \mathrm{sf} \qquad (1)$$

where

IRI = predicted IRI (in./mi.),
IRI_I = initial smoothness measured as IRI (in./mi.),
crk = percent slabs with transverse cracks (all severities),
spall = percentage of joints with spalling (medium and high severities),
tfault = total joint faulting cumulated per mile (in.),
$C1$–$C4$ = calibration coefficients,
$C1$ = 0.8203 (default),
$C2$ = 0.4417 (default),
$C3$ = 1.4929 (default),
$C4$ = 25.24 (default), and
sf = site factor.

$$\mathrm{sf} = \mathrm{age}(1 + 0.5556 * \mathrm{fi})(1 + P_{200}) * 10^{-6}$$

where

age = pavement age (years),
fi = freezing index (°F-days), and
P_{200} = percent subgrade material passing No. 200 sieve.

IRI Model for HMA Overlays of Rigid Pavements

A similar concept to that of the JPCP IRI model was used to develop the smoothness model for HMA overlays of rigid pavements. The IRI prediction model is considered as a function of distresses that contribute to the reduction of pavement smoothness. Age as well as initial IRI, average rut depth, and the average spacing of medium- and high-severity transverse cracks are among the distress variables used to predict the IRI of HMA overlays of rigid pavements. The model is (*7*)

$$\mathrm{IRI} = \mathrm{IRI}_I + C1 * \mathrm{age} + C2 * \mathrm{rd} + C3 * \frac{1}{(TC_S)_{\mathrm{MH}}} \qquad (2)$$

where

IRI = predicted IRI (m/km),
IRI_I = initial smoothness measured as IRI (m/km),
rd = average rut depth (mm),
$(\mathrm{tc}_S)_{\mathrm{mh}}$ = average spacing of medium- and high-severity transverse cracks (m),
$C1$ = 0.0082627 (default),
$C2$ = 0.0221832 (default), and
$C3$ = 1.33041 (default).

LOCAL STATEWIDE CALIBRATION OF TEST MODELS FOR NEBRASKA

The data used in the default calibration of the test models in the design guide included available JPCP sections and HMA overlay sections in the LTPP database. The data did not target specific traffic levels or thicknesses but rather represented a broad range of design elements. This concept of general calibration was applied in performing a statewide local calibration of the IRI model to Nebraska pavements.

Because the test models are purely empirical, the calibration procedure differs from that of other calibrations in the design guide. The design guide software was not needed to predict values for cracking, rutting, spalling, and faulting. Instead, distress values were taken directly from the pavement management system (PMS) database, and a nonlinear optimization algorithm was employed by a computer processor to determine the set of calibration coefficients that corresponded to the minimum prediction error. Statewide calibration of the test models employed the following process:

1. Identify the calibration data set for the network,
2. Filter the calibration data set to ensure valid data, and
3. Select the coefficients that minimize the sum of squared differences between predicted IRI and measured IRI through an unconstrained, simply bound optimization algorithm.

The calibration procedure is discussed in further detail in the following sections.

Calibration Data

The design guide suggests that local agencies consult the LTPP database as an initial data resource when local calibrations are performed (*1*). Figure 1 shows the availability of LTPP data in selected central states. Although the LTPP database provides superior detail of pavement performance history, it does not offer sufficient calibration data volumes to every agency (*8*). The local agency's PMS database will prove invaluable in the design guide implementation. Proper utilization of the data will yield effective calibrations.

Identification of Calibration Data Set

A major underlying assumption in model calibration is that a model cannot be assumed to be any more accurate than the data occupied in

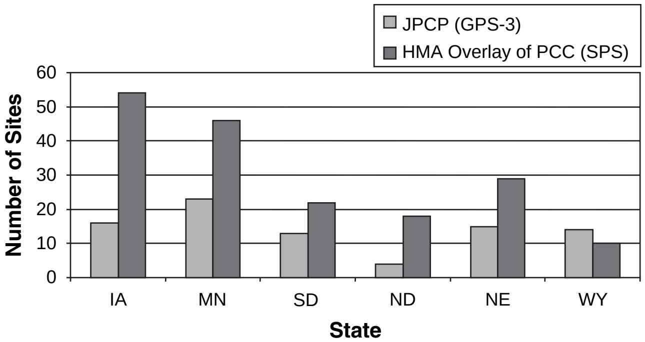

FIGURE 1 Number of LTPP sites available by state.

its calibration (*9*). A careful evaluation of the calibration data set was made to ensure that only applicable, valid data were used. The Nebraska Department of Roads (NDOR) PMS conducts annual pavement condition surveys on the entire network. An electronic database exists containing the distress ratings from these surveys dating back to 1997. Several categories including surface type, highway number, district, county, functional class, traffic, and thickness are used to catalog the data. More than 5,400 rigid pavement sections exist in the NDOR database ranging from 0.01 to 12.8 mi in length (*10*).

Detailed distress ratings are gathered on the first 10 joints and panels of each rigid pavement section. A windshield survey determines if the ratings are representative of the entire segment by recording the extent and severity of several functional and structural distresses. The database also contains more than 2,000 HMA overlay sections ranging from 0.01 to 11.8 mi in length. The first 200 ft of each flexible pavement section is closely monitored, and the remainder is rated via the windshield survey. A road profiler is used to measure rut depth, faulting amount, and smoothness (*11*). Transverse cracking severity is rated according to the same standards used in the design guide. Rutting and faulting are measured in millimeters. Faulting had to be converted to cumulated inches per mile by dividing the faulting amount by the joint spacing and converting to the proper units of inches per mile.

Data Filtering

Initial IRI

The NDOR pavement management software shows the distress ratings of each pavement section in a time series. Absent from the database is the initial smoothness IRI. Titus-Glover and Darter used a reasonable method for estimating the initial IRI (*6*). A majority of the sections contained IRI values within 12 months of construction. Considering IRI as a function of age, the slope and intercept values were found by performing a linear regression on the time series. The functional form of the best-fit line is

$$\mathrm{IRI} = \alpha \mathrm{age} + \beta \tag{3}$$

where

α = slope,
age = pavement age (years), and
β = intercept (initial IRI).

Although the method proved successful, poor correlation between IRI and age was observed on some sections, resulting in large IRI_I confidence intervals. A decrease in IRI was observed on some sections without any documentation of maintenance. Using such pavement sections in the calibration procedure introduced large unexplained variations in the model. Therefore, a filtering process was implemented in the data collection before calibration. A coefficient of determination, R^2, of 0.64 was used as a criterion for filtering pavement sections to be used for calibration. This value corresponds to a correlation coefficient of 0.8, classified as strong (*12*). Reasonable initial IRI values were observed after filtering.

Maintenance Considerations

While the calibration data sets were being built, consideration was given to maintenance and rehabilitation (M&R) activities. Because M&R activities such as diamond grinding, patching, and milling result in improved pavement condition, sections undergoing any activity affecting the smoothness measurements were only partially used. All data before an M&R activity were kept. The age of the section was then reset to zero.

Minimum Number of Time-Series Observations

Following recommendations in the design guide, priority was given to pavement sections containing at least 3 years of data (*13*).

Optimization

PaveCARE (Pavement Condition Assessment and Rehabilitation Effectiveness) is the pavement management software that uses NDOR historical pavement condition ratings to assess the condition of the pavement network as well as to analyze the effectiveness of rehabilitation strategies applied throughout the state (*10*). Because the IRI model is purely empirical, the design guide software is not needed to determine the parameters in the model. Instead, actual field measurements of cracking, spalling, and faulting can be used directly (*6*).

In recognition of this aspect and the data capabilities of PaveCARE, the NDOR software was retrofit with a nonlinear simple-bound optimization algorithm to minimize the sum of squared difference, ssd, with *n* observations (*14*). The optimization function for the JPCP IRI model is

$$\mathrm{ssd} = \sum_{x=1}^{n} \{\mathrm{IRI}(x) - [\mathrm{IRI}_I(x) + C1 * \mathrm{crk}(x) + C2 * \mathrm{spall}(x) + C3 * \mathrm{tfault}(x) + C4 * \mathrm{sf}(x)]\}^2 \tag{4}$$

Bounds

The design guide places acceptable ranges on each of the calibration coefficients in the IRI models. The lower bound for the JPCP IRI model is zero for each of *C*1, *C*2, *C*3, and *C*4, and the upper bounds are 10, 10, 10, and 100, respectively. Lower and upper bounds for the HMA overlay model coefficients are −1000 and 1000, respectively (*1*).

Calibration Results

Utilizing rutting, cracking, spalling, faulting, and site data in the pavement management database, the PaveCARE software found the set of coefficients that minimized the error between measured IRI and predicted IRI for both test models. Because the design guide software uses models to predict values for rutting, cracking, spalling, and faulting, a source of variability is introduced on the basis of the individual calibration goodness-of-fit statistics of each model. The procedure used in this study uses measured values for each distress and does not consider the error resulting from their individual predictions.

Results of each statewide calibration and the associated statistics are shown in Table 1. The network-level calibrations using local data are compared with the results from the design guide, which use LTPP data.

VALIDATION OF STATEWIDE CALIBRATION

Validation is needed to demonstrate the ability of the models to predict smoothness values for pavement sections not used in the calibration effort. No consensus exists on acceptable model performance criteria. However, basic truths of model calibration and validation are commonly accepted among professionals in the field of modeling. Models cannot precisely predict natural systems but are rather approximations of reality. To quantify these approximations, several descriptive statistics and statistical tests are used to show validity; no single statistical test exists to determine validation. The process of verifying the validity of a model should include a combination of statistical evidence and graphical comparisons (*9*).

Validation Data

The same process of elimination used to build calibration data sets needed to be repeated for validation. Pavement sections used in calibration could not be used to test the validity of the model. The LTPP database was considered as an alternative data source for validation sections. Commonly, pavement performance databases do not offer the required data volumes needed to complete validation, calibration, or both. Such was the case with the LTPP database. Only three Nebraska pavement sections, each averaging four observations, were available in the LTPP database for JPCP IRI model validation (*6*).

Von Quintus et al. explored an experimental approach to model validation (*15*). A statistical procedure known as jackknifing provided the validation of a permanent deformation model calibrated with MnROAD test sections. Von Quintus et al. found that $n - 1$ jackknifing provides a more realistic assessment of model accuracy for small sample sizes. To overcome limited data availability, this procedure was used to validate the test models in the current study.

Jackknifing

Jackknifing provides a more accurate assessment of model performance than calibration statistics because its goodness-of-fit statistics are based on predictions rather than on data used in calibration. The procedure involves calibrating a model with the use of a sample size $n - 1$, where n is the size of the calibration data set. In the first iteration, one observation is withheld. A calibration using the remaining $n - 1$ observations yields a calibrated model. The withheld observation is then used to test the model and generate the first error term, e_1, calculated as the difference between the predicted value, Y_p, and the measured value, Y_m. Once the e_1 is calculated, the observation is returned to the data set, and the next observation is withheld. The process is repeated until n error terms have been calculated. Goodness-of-fit statistics can be calculated upon completion of the jackknifing procedure. This process provides an independent measure of model accuracy.

Statistical Validation

Tools used for validation included error statistics (e.g., the standard error of the estimate), correlation statistics (e.g., the coefficient of determination), and hypothesis testing. The validation statistics were comparable with those obtained through calibration.

Hypothesis Testing

A two-sample t-test was used to test the null hypothesis, H_0, which stated that no difference exists between the predicted mean and the measured mean against the alternative. Failure to reject H_0 indicates that no significant difference exists between the predicted mean IRI resulting from calibration and the measured IRI (*12*). The equation used to calculate the t-statistic is

$$t = \frac{(\mu_{\text{predicted}} - \mu_{\text{measured}}) - 0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \tag{5}$$

where

$\mu_{\text{predicted}}$ = mean predicted IRI,
μ_{measured} = mean measured IRI, and

$$S_p^2 = \frac{SS_{\text{predicted}} + SS_{\text{measured}}}{(n_1 - 2) + (n_2 - 2)}$$

TABLE 1 Statewide Calibration Results Compared Design Guide Defaults

Calibration	$C1$	$C2$	$C3$	$C4$	N	R^2	SEE
JPCP IRI default	0.820	0.442	1.493	25.240	183	0.600	27.3 (in./mi)
Nebraska JPCP IRI	0.000	0.000	1.563	87.160	1270	0.662	25.031 (in./mi)
HMA overlay IRI default	0.008	0.022	1.330	n/a	367	0.543	0.197 (m/km)
Nebraska HMA overlay IRI	0.087	0.031	0.100	n/a	670	0.621	0.485 (m/km)

SSE = standard error of estimates.

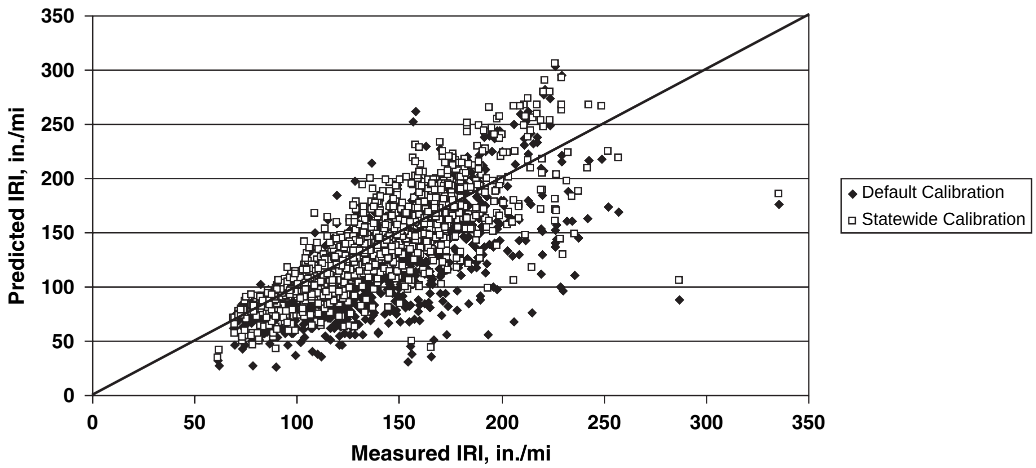

FIGURE 2 Measured versus predicted IRI scatter plot: JPCP IRI model.

where SS is the sum of squares and n_1, n_2 is the sample size.

With an α-level of 0.05, t-tests for both models showed no significant difference between measured and predicted IRI values.

Graphical Validation

Graphical evidence was used to validate the test models. Scatter plots of the predicted versus measured IRI values are shown in Figures 2 and 3. Included are comparisons with the respective design guide default models.

The line of equity is shown in both plots. A scatter that is concentrated along this line is desirable. A slight reduction in scatter was observed in Figure 2 after local calibration. Figure 3 shows a significant improvement in prediction accuracy after local calibration. As shown, the default HMA IRI overlay model severely underpredicts IRI on Nebraska composite pavements since the data points fall well below the line of equity. A better fit of the line was observed in the scatter of the local calibration.

The validation statistics shown in this section represent the entire population of JPCP and HMA overlay pavement sections in the state. Although the improvements in prediction accuracy observed in the statewide calibration–validation process are recognized, this study is aimed at achieving even greater accuracy at the project level.

FOCUS CALIBRATION

Introduction

The process set forth in this study for calibrating the test models parallels the basic process and methods found in the design guide. The general process is outlined in the following steps (*16*):

1. Compile a calibration data set;
2. Compare predicted and measured values to determine if there is a need for calibration (e.g., hypothesis testing, measured versus predicted scatter plot);
3. If calibration is needed, alter the calibration coefficients until the error between measured and predicted values is minimized; and
4. Validate the calibration with statistical and graphical methods.

The design guide does not attempt to quantify the benefits of focusing the calibration to a set of specific conditions. In most cases, only one set of default calibration coefficients exists to represent all design scenarios. In the remainder of this discussion it will be shown that a single statewide calibration proves insufficient for isolated thicknesses and ADTT levels when compared with project-level calibrations. Focus calibrations proved effective in increasing the prediction accuracy of the test models.

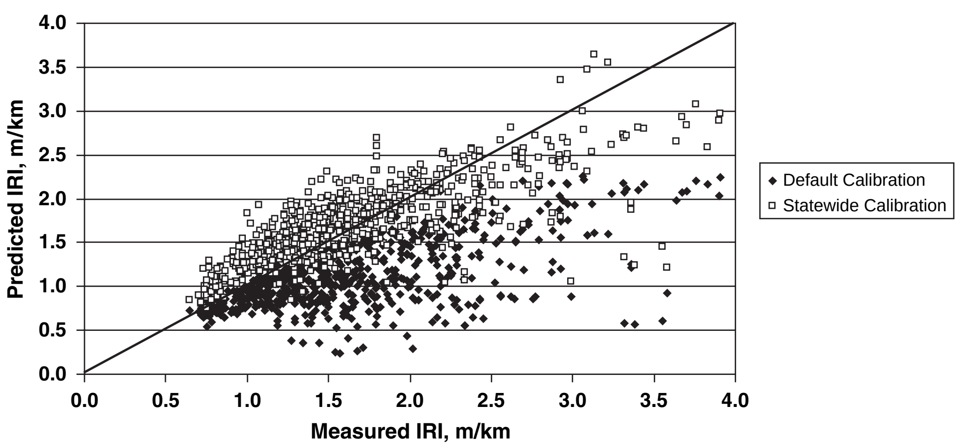

FIGURE 3 Measured versus predicted IRI scatter plot: HMA overlay IRI model.

Focus Calibration of Test Models

Focus calibration of the test models employed the following process:

1. Identify the focus parameters,
2. Identify the calibration data set,
3. Filter the calibration data set to ensure valid data, and
4. Select the coefficients minimizing the sum of squared differences between predicted IRI and measured IRI through an unconstrained, simply-bound optimization algorithm.

The calibration procedure is discussed in further detail in the next sections.

Focus Parameters

Focus parameters are the categories used to narrow the data set into subsets. The focus parameters used in the focus calibration of the test models were ADTT and surface layer thickness. Three categories of ADTT were considered: low, medium, and high. The categories as defined by NDOR were as follows:

1. Low, 0 to 200 trucks/day;
2. Medium, 201 to 500 trucks/day; and
3. High, more than 500 trucks/day.

The surface layer thicknesses considered ranged from 6 to 14 in. for JPC pavements and 0 to 8 in. for HMA layers.

Calibration Data

The same data available for the statewide calibration were used in the focus calibrations.

Identification of Calibration Data Set

The PaveCARE software and the established focus parameters were used to form the calibration data sets. The number of data subsets expanded as the focus was narrowed. Figure 4 shows the data tree used to complete the focus calibration of the JPCP IRI model. During the focus calibration, the complete data set used for the network-level calibration was divided into three smaller samples corresponding to low, medium, and high ADTT. After the calibration and validation were completed for each ADTT level, the three data sets were divided further into thicknesses. For the JPCP model, nine thickness ranges (6 to 14 in.) were used for each level of ADTT, as shown in Figure 4. Each thickness range within each ADTT level was calibrated. A total of 30 calibrations and validations were completed beyond the initial statewide calibration. The calibration and validation of the thickness ranges for each specific ADTT level are considered the local project-level calibration.

Data Filtering

The aforementioned filtering criteria were used in constructing the final focused data subsets:

- Initial IRI. The same backcasting concept utilized in the statewide calibration was used for the focus calibrations.
- M&R. The same M&R considerations applied during focus calibration.
- Minimum number of time-series observations. Priority was given to pavement sections containing at least 3 years of data. Sections having fewer than three time-series observations were excluded from all calibration data sets (*13*). Pavement sections more than 30 years old were limited in the database. Sections exceeding 30 years in age were not considered eligible for calibration.

Optimization

As in the statewide calibration, PaveCARE was used in the optimization process to achieve the set of calibration coefficients corresponding to the minimum sum of squared residuals (*14*).

Bounds

The bounds remained unchanged for the focus calibration.

Calibration Results

Table 2 shows the calibration coefficients deemed appropriate for predicting smoothness in JPC pavements at the project level for Nebraska. Likewise, Table 3 contains coefficients for the smoothness model of HMA overlays of rigid pavements. These calibration values should be used with the calibration statistics in mind. Small sample sizes were observed for some focus categories. The

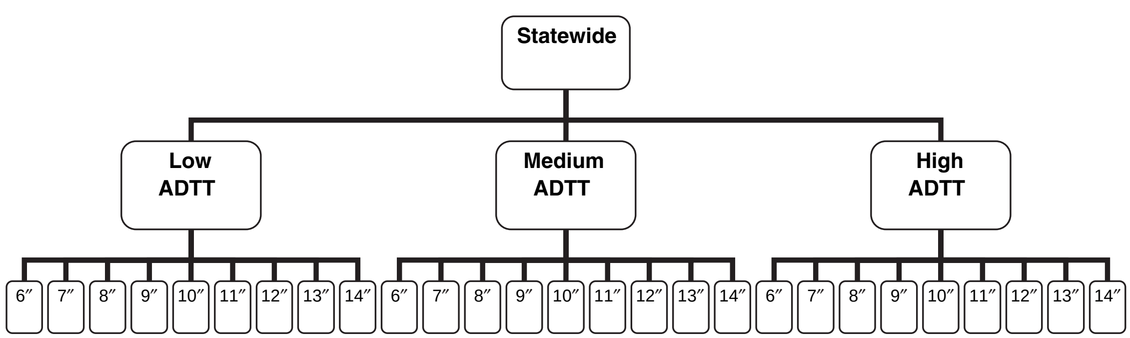

FIGURE 4 JPCP IRI calibration data tree.

TABLE 2 JPCP IRI Calibration Coefficients for Surface Layer Thickness Within ADTT

ADTT	Thickness	*C*1	*C*2	*C*3	*C*4	*N*	R^2	SEE (in./mi)
Low	6″–7″	0.0000	0.0000	1.0621	74.8461	33	0.434	26.885
	7″–8″	0.0000	0.0000	1.9923	46.9256	37	0.961	8.235
	8″–9″	0.8274	0.0000	0.0000	86.9721	39	0.904	14.465
	9″–10″	0.3458	0.0000	1.5983	64.3453	110	0.537	26.230
	10″–11″	0.0300	0.0000	3.4462	10.7893	37	0.893	17.280
	11″–12″	—	—	—	—	—	—	—
	12″–13″	—	—	—	—	—	—	—
	13″–14″	—	—	—	—	—	—	—
	14″–15″	—	—	—	—	—	—	—
Medium	6″–7″	0.0000	0.0000	4.1422	0.0000	3	0.966	5.094
	7″–8″	0.0000	1.5628	0.0000	71.9009	22	0.968	9.952
	8″–9″	0.0000	0.0000	1.7162	53.0179	122	0.291	40.537
	9″–10″	0.1910	0.0000	0.9644	89.3990	609	0.686	24.945
	10″–11″	0.0000	0.0000	2.0945	73.1246	314	0.812	18.535
	11″–12″	0.0000	0.0090	1.3617	100.0000	27	0.792	10.166
	12″–13″	—	—	—	—	—	—	—
	13″–14″	0.0000	0.0100	2.2226	24.9354	4	0.924	3.948
	14″–15″	—	—	—	—	—	—	—
High	6″–7″	—	—	—	—	—	—	—
	7″–8″	—	—	—	—	—	—	—
	8″–9″	0.0000	0.1376	0.4352	79.5526	46	0.151	48.576
	9″–10″	0.1561	0.0000	1.1024	62.9556	81	0.333	31.255
	10″–11″	0.0000	0.0000	1.6344	100.0000	228	0.653	22.295
	11″–12″	0.1125	1.8207	1.1678	100.0000	29	0.739	13.366
	12″–13″	0.0000	0.0000	1.5331	100.0000	151	0.719	17.724
	13″–14″	0.0100	0.0100	0.5184	0.0000	4	0.623	1.728
	14″–15″	0.1904	0.0000	2.1387	51.4053	146	0.838	9.018

TABLE 3 HMA Overlay IRI Calibration Coefficients for Surface Layer Thickness Within ADTT

ADTT	Thickness	*C*1	*C*2	*C*3	*N*	R^2	SEE (m/km)
Low	2″–3″	0.1318	0.0018	0.3971	3	0.994	0.02
	4″–5″	0.0704	−0.0048	−2.8771	16	0.813	0.11
	5″–6″	−0.0038	0.2409	−4.6360	5	0.039	1.15
Medium	2″–3″	0.0639	0.1337	−0.7896	21	0.612	0.5
	3″–4″	0.0733	0.0282	1.4725	65	0.532	0.36
	4″–5″	0.0781	−0.0032	1.1116	82	0.546	0.31
	5″–6″	0.0649	0.0169	3.5543	84	0.535	0.31
	6″–7″	0.0794	−0.0312	4.3652	31	0.888	0.17
	7″–8″	0.0674	−0.0164	1.7122	19	0.674	0.13
	8″–9″	0.0683	0.0192	−3.6231	13	0.936	0.1
High	0″–1″	0.2019	0.1158	−10.0646	27	0.392	0.45
	2″–3″	0.1866	0.0498	−16.7082	19	0.565	0.6
	3″–4″	0.1835	−0.0579	8.1863	32	0.010	0.9
	4″–5″	0.1170	−0.0100	1.4057	101	0.299	0.51
	5″–6″	0.2422	0.0371	−23.4448	62	0.713	0.85
	6″–7″	0.0756	0.0127	0.9250	64	0.597	0.22
	7″–8″	0.0604	0.0574	−2.4936	7	0.624	0.2
	8″–9″	0.0578	0.0706	−10.9179	28	0.103	0.25
	9″–10″	0.1005	−0.0001	−0.5216	8	0.845	0.13

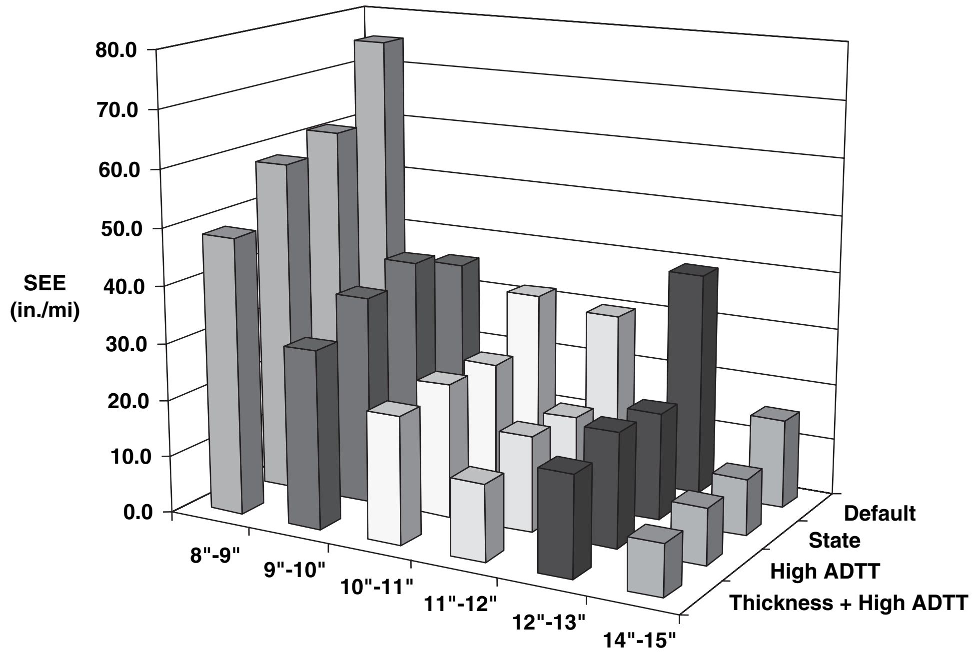

FIGURE 5 JPCP IRI model: standard error of the estimate (SEE) for high ADTT level.

comparisons of these values with those of default and statewide calibrations can be seen in Figures 5 through 7. In general, as the focus increases, the prediction error decreases.

VALIDATION

The validation process used for the focus calibrations is identical to the process used for validating the statewide calibration. The (n − 1) jackknifing method was used to construct a data set of pavement sections that were not used during calibration. The validation process shows how local data can be used in focus calibrations to provide greater prediction accuracy than that of a statewide calibration.

Statistical Validation

Tools available for statistical validation include error statistics (e.g., the standard error of the estimate), correlation statistics (e.g., the coefficient of determination), and hypothesis testing. As stated earlier, the data sets are narrowed from broad samples to those that are focused

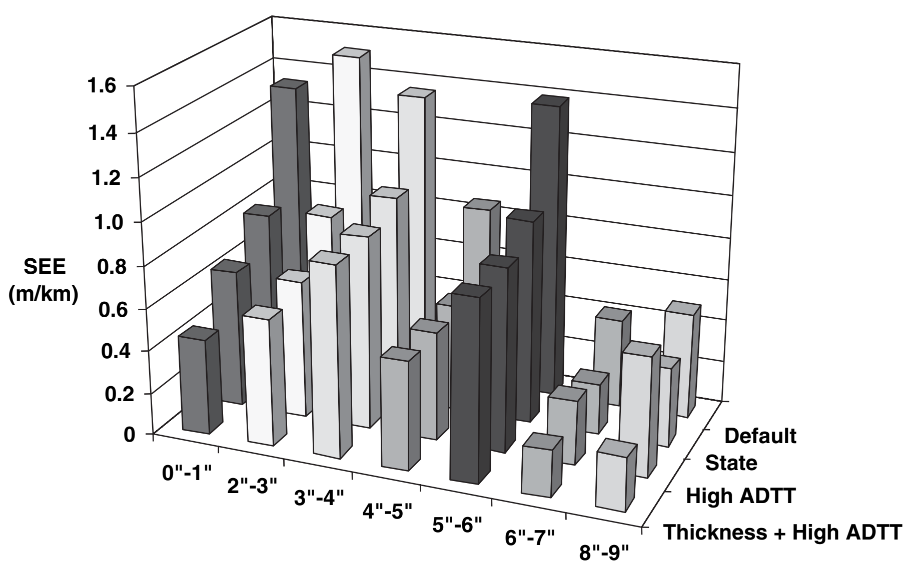

FIGURE 6 HMA overlay IRI model: SEE for high ADTT level.

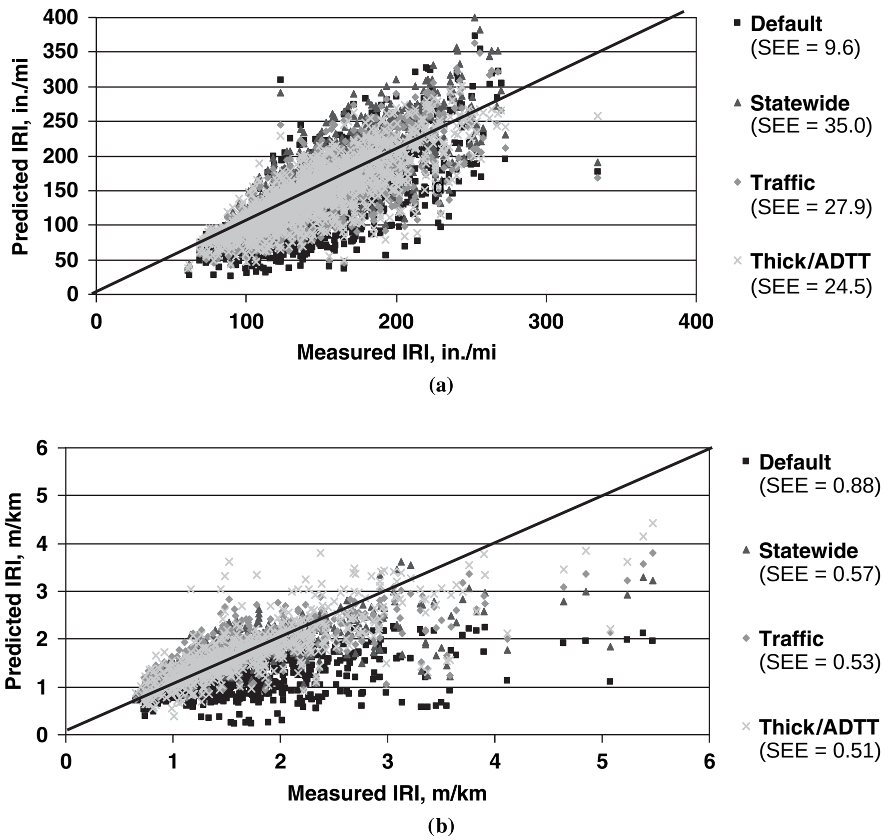

FIGURE 7 Scatter plot of measured IRI versus predicted IRI.

to surface layer thickness within traffic levels. With these jackknifed data sets, the four sets of focus calibration coefficients (default, statewide, ADTT, and thickness within ADTT) were used to predict IRI. Figures 5 and 6 show the standard error of the estimates and focus level for each thickness range in the high traffic level for the test models. As the focus increased, an overall decrease in error was observed for the test models.

For example, by using Figure 5, an engineer designing an 8-in. JPCP section with high ADTT could use the high ADTT coefficients. However, although these coefficients were obtained from a calibration data set that included only high ADTT sections, they also included all thicknesses. Using the thickness + high ADTT coefficients would reveal an even lower error because they were obtained through a calibration data set that focused only on ADTT sections 8 to 9 in. high. These project-level coefficients revealed the highest accuracy.

All calibrations yielded no significant difference between measured and predicted IRI.

Graphical Validation

Graphical evidence was used to validate the test models. Figure 7 shows scatter plots of measured versus predicted IRI for each model. A reduction in scatter was observed as the focus progressed from a network-level to a project-level calibration. It can be seen that the focus calibrations yield the smallest scatter in both test models.

CONCLUSION

The mechanistic–empirical pavement design guide uses LTPP data encasing the full spectrum of traffic, material, climatic, and geometric conditions with which to calibrate the mechanistic–empirical models it employs. Through calibration, the maximum agreement between predicted and measured distress values was achieved for general default conditions. However, each project boasts unique, local elements that cannot be tamed with a systematic model. Local calibration attempts to harness the range of area conditions and use them to improve prediction accuracy. A comparison between local and default IRI predictions showed such an improvement.

Although statewide calibrations showed accuracy improvements from the design guide defaults, they still represented the entire PMS network. Focusing the data sets to reflect project-level conditions proved effective in reducing prediction error, offering a window into the improvement potential that focus calibration can bring to the models used in the design guide. This study focused the data by the selected design elements: thickness and ADTT. Comparable accuracy improvements are anticipated with other design elements such as subbase type and dowel bar use.

Statewide calibration of the JPCP IRI model resulted in a 15% reduction in standard error of the estimates from that of the default model. The reduction nearly doubled with focus calibrations to 29%. Focus calibration of the IRI model for HMA overlays of rigid pavements reduced the standard error of the estimates from 37% at the network level to 43% at the project level.

The test models used in this study were empirical. Nevertheless, the same accuracy improvements can be anticipated for mechanistic–empirical models. The challenge of applying focus calibrations to M-E models lies in the amount of time required to complete the calibrations. The time effort put forth in the default calibration was significant and would have to be multiplied by the number of focus parameters involved. However, the improvement in prediction accuracy may justify the effort.

Acceptance of the new design guide weighs heavily on its ability to predict reasonable values of local distress. Only through local calibration will model error be minimized. Its importance has been exposed by comparing calibrated predictions with default predictions of the test models. Calibration effectiveness is anchored in LTPP and local PMS data. Harnessing the local data to develop project-level calibrations has proved a valuable vehicle to achieving improved accuracy beyond that at the network level.

ACKNOWLEDGMENT

The authors acknowledge NDOR for providing the data.

REFERENCES

1. *Guide for Mechanistic–Empirical Design of New and Rehabilitated Pavement Structures.* Final Report, NCHRP Project 1-37A. TRB, National Research Council, Washington, D.C., March 2004. http://trb.org/mepdg/guide.htm.
2. Von Quintus, H. L., C. Schwartz, R. H. McCuen, and D. Andrei. *NCHRP Report 284: Refining the Calibration and Validation of Hot Mix Asphalt Performance Models: An Experimental Plan and Database.* TRB, National Research Council, Washington, D.C., 2003.
3. Witczak, M. W., and M. M. El-Basyouny. Appendix GG-1: Calibration of Permanent Deformation Models for Flexible Pavements. In *Guide for Mechanistic–Empirical Design of New and Rehabilitated Pavement Structures,* TRB, National Research Council, Washington, D.C., March 2004. http://trb.org/mepdg/guide.htm. Accessed Sept. 2004.
4. *Special Report 61-E: Pavement Research.* HRB, National Research Council, Washington, D.C., 1962.
5. Sayers, M. W., T. D. Gillespie, and W. D. Paterson. *Guidelines for Conducting and Calibrating Road Roughness Measurements.* Technical Paper 46. World Bank, Washington, D.C., 1986.
6. Titus-Glover, L., and M. I. Darter. Appendix PP: Smoothness Prediction for Rigid Pavements. In *Guide for Mechanistic–Empirical Design of New and Rehabilitated Pavement Structures,* TRB, National Research Council, Washington, D.C., March 2004. http://trb.org/mepdg/guide.htm. Accessed Sept. 2004.
7. Von Quintus, L., and A. Yau. Appendix OO-2: Revised Smoothness Prediction Models for AC. In *Guide for Mechanistic–Empirical Design of New and Rehabilitated Pavement Structures.* TRB, National Research Council, Washington, D.C., March 2004. http://trb.org/mepdg/guide.htm. Accessed Sept. 2004.
8. *DataPave 3.0 Long-Term Pavement Performance* (CD-ROM). FHWA, U.S. Department of Transportation, 2001.
9. Donigian, A. S., Jr. Watershed Model Calibration and Validation: The HSPF Experience. Presented at Water Environment Federation 2002 Specialty Conference, Phoenix, Ariz., 2002.
10. *PaveCARE User's Guide 2.0: Pavement Condition Assessment and Rehabilitation Effectiveness.* Nebraska Department of Roads, Lincoln, 2005.
11. *Pavement Condition Survey Manual.* Nebraska Department of Roads, 1985, Jan. 2003.
12. Myers, J. L., and A. D. Well. *Research Design and Statistical Analysis.* Lawrence Erlbaum Associates, Inc., Mahwah, N.J., 2003.
13. Witczak, M. W., and M. M. El-Basyouny, Appendix II-1: Calibration of Fatigue Cracking Models for Flexible Pavements. In *Guide for Mechanistic–Empirical Design of New and Rehabilitated Pavement Structures.* TRB, National Research Council, Washington, D.C., March 2004. http://trb.org/mepdg/guide.htm. Accessed Sept. 2004.
14. Nash, S. G. Newton-Type Minimization via the Lanczos Method. *SIAM Journal of Numerical Analysis,* Vol. 21, 1984, pp. 770–778.
15. Von Quintus, H. L., C. Schwartz, R. H. McCuen, and D. Andrei. *NCHRP Report 283: Jackknife Testing: An Experimental Approach to Refine Model Calibration and Validation.* TRB, National Research Council, Washington, D.C., 2003.
16. Hoerner, T. E., M. I. Darter, L. Khazanovich, L. Titus-Glover, and K. L. Smith. *Improved Prediction Models for PCC Pavement Performance-Related Specifications, Volume I: Final Report.* Publication FHWA-RD-00-130. FHWA, U.S. Department of Transportation, 2000.

The Rigid Pavement Design Committee sponsored publication of this paper.