ONTARIO'S LOCAL CALIBRATION OF THE MEPDG DISTRESS AND PERFORMANCE MODELS FOR FLEXIBLE ROADS: A SUMMARY

Xian-Xun Yuan

Department of Civil Engineering, Ryerson University
350 Victoria Street, Toronto, ON, Canada

Warren Lee, Ningyuan Li

Pavement and Foundation Section, Ministry of Transportation of Ontario,
1201 Wilson Avenue, Toronto, ON, Canada

Paper prepared for presentation
at the *Innovations in Pavement, Management, Engineering and Technology* Session

of the 2017 Conference of the
Transportation Association of Canada
St. John's, NL

ABSTRACT

This paper summarizes the research outcomes from the multiple research projects devoted to local calibration of the distress and performance models of the AASHTO Mechanistic-Empirical Pavement Design Guide (MEPDG) for Ontario's flexible highway pavements. The study started with development of a local calibration database, which was later enhanced with a focus on Superpave sections. The permanent deformation or rutting models, fatigue cracking models, thermal cracking models, reflective cracks models, and finally the IRI models were all studied, and calibrated if every possible. The following main results are highlighted: (1) After several attempts and innovation on calibration database and calibration method development, the rutting models have been well calibrated. (2) Among the several types of cracking models, only the bottom-up fatigue cracking model has been successfully calibrated, whereas the top-down cracking, thermal cracking and reflective cracking models are still facing major challenges, reason being either a lack of reliable observation data or continuous updating status of the global models. (3) Since the thermal and reflective cracking models are subject to further development and global and local calibrations, the IRI model has been partially calibrated for its rutting and fatigue cracking terms. However, the full local calibration of the IRI model can be readily done after all cracking models are calibrated. The paper is concluded with a reflection of the work, which serves a good guide for other transportation agencies, either American or Canadian, for their local calibration study.

ACKNOWLEDGEMENTS

The financial supports from the Ministry of Transportation Ontario through its Highway Infrastructure Innovation Funding Program in FY2010-11, 2013-15, and 2015-17 are gratefully acknowledged. A number of graduate students have worked on the local calibration studies. They are Gulfam Jannat, Afzal Waseem, Gyam Gautam, Sifat Ahmed, Maryam Amir, and Iliya Nemtsov. Their research provided important inputs for the work presented in this paper. Mr. Joseph Ponniah, a former senior pavement research engineer of MTO provided technical assistance on the local calibration database development in the early stage of the research program. Thanks are due also to Dr. Medhat Shehata of Ryerson University for his generous discussions and suggestions.

1. INTRODUCTION

Developed under multiple NCHRP projects including 1-37A[1], 1-40[2] and 9-30A [3] over the past 15 years, the AASHTO Mechanistic-Empirical Pavement Design Guide (MEPDG) is emerging as a mainstream pavement design method throughout North America. The method established a direct tie between pavement distresses and various design inputs including material properties, pavement structures, traffic loadings, climate, soil conditions, construction quality, and so on. The design method has been packaged in a user-friendly working platform now called the AASHTOWare Pavement ME software (originally DARWin-ME). As one of the leading transportation agencies in Canada, the Ministry of Transportation of Ontario (MTO) has been mandated to implement the MEPDG for future pavement design.

Preliminary studies for Ontario's conditions have shown that the global (or default) distress models in the MEPDG do not accurately predict the pavement distresses and performance for Ontario roads. Rutting has been found to be drastically over-predicted, whereas fatigue cracking is often under-predicted [4]. The need for local calibration was thus obvious. To perform the local calibration, three research projects have since 2010 been commissioned to Ryerson University, with the last one jointly to the University of Waterloo as well, under the support of the MTO Highway Infrastructure Innovation Funding Program (HIIFP). The pavement performance data from the MTO's second-generation pavement management system (PMS-2) would be used for the local calibration. The first HIIFP project focused on the development of a local calibration database that included a number of typical pavement sections with accurate design input data as well as high-quality performance and distress data. The second project was tasked mainly to perform local calibration for the rutting models. The third and last project continued on the local calibration for the cracking models and international roughness index (IRI) model by using more accurate performance data collected by the new ARAN 9000 system [5]. While the first two projects were targeted exclusively to flexible pavements, the last project is expanded to include rigid pavements, though a much smaller portion in Ontario's provincial highway system.

The challenges faced in the database development were discussed in [4] and [6]. Some intermediate local calibration results for the rutting, fatigue cracking and IRI models were reported in [7-12]. This paper presents a brief summary of the major efforts and the key findings from these studies.

The paper is organized as follows. First of all, the distress and performance models used in MEPDG are briefly reviewed in Section 2. The review is important because the notations used in the original MEPDG documents [13] and AASHTO Manual of Practice [14, 15] are confusing and we are trying to provide a clearer presentation of those models, with the local calibration coefficients highlighted. The notations are unified as much as reasonably allowed. Second, the need for local calibration is reiterated in Section 3. Section 4 presents a summary of the major work. The local calibration methodology used in the Ontario study is discussed in Section 5. Several innovations proposed during our studies are underlined. Section 6 presents the major results and discussions. The paper is concluded in Section 7 with a reflection of the work, hoping to serve a guide for other agencies in their future local calibration study.

2. DISTRESS AND PERFORMANCE MODELS IN MEPDG

2.1. The Rutting Models

In contrast to several previous design methods in which rutting is thought to be a result of settlement of subgrade soil only, the MEPDG relates rut depth to the vertical permanent deformation of different structural layers. The mechanistic analysis starts with first calculating the resilient strain in each analysis layer based on elastic layer theory. After the resilient strains are obtained, the plastic strain of each analysis layer is then calculated by using one of the three empirical rutting models, namely, the asphalt concrete (AC) or hot mix asphalt (HMA) model, the granular base or subbase model, and the fine-grained subgrade model. The latter two models, collectively called the unbound granular material model, have the same model structure but different global and local calibration coefficients.

The empirical AC rutting model is expressed as [14]

$$\frac{\varepsilon_{p,AC}}{\varepsilon_{r,AC}} = k_z \beta_{AC} 10^{-3.3541} T^{1.5606\beta_T} N^{0.4791\beta_N} \quad (1)$$

where $\varepsilon_{r,AC}$ denotes the resilient strain of AC at the mid-depth of a given analysis layer under a specific traffic load; $\varepsilon_{p,AC}$ the corresponding *accumulated* plastic strain; k_z the depth confinement factor as a function of total asphalt layer thickness and depth to computational point; T the temperature at the given analysis layer in Fahrenheit degree; N the number of load repetitions; and finally, $\beta_{AC}, \beta_T, \beta_N$ the local calibration factors, which all equal 1.0 by default. Note that β_{AC} is also called the AC-scale factor, and β_T and β_N are called the temperature and traffic exponents, respectively.

The rutting models for the unbound granular materials and fine-grained soil have the same functional structure except for a different scale factor. Since the notations used in the MEPDG documents for these two rutting models are inconsistent, for the sake of local calibration, the transfer function can be rearranged and expressed as the following:

$$\frac{\varepsilon_{p,i}}{\varepsilon_{r,i}} = k_s \beta \phi(N, \alpha) \quad (2)$$

where k_s represents the global calibration factor, and β the local calibration factor. In this paper, β_{GB} and β_{SG} are used for the granular and fine-grained materials, respectively, to differentiate the two models. The function $\phi(N, \alpha)$ lumps the effect of repetitive traffic loading N and soil moisture with a transformed parameter α describing moisture content (W_c) in the soil. The global calibration factor for granular materials was set to 1.673 originally and changed to 2.03 in the new 2015 MEPDG Manual of Practice. It equals 1.35 for fine-grained materials.

Note that the calculation of the plastic strains in the AC layers also involves a so-called 'strain-hardening procedure'; for details, refer to [1]. Once the plastic strains are obtained, the total rut depth at age t, denoted by $\mathrm{RD}(t)$, is then calculated from the following summation expression:

$$\mathrm{RD}(t) = \sum_{i=1}^{M} \varepsilon_{p,i}(t)\, h_i \quad (3)$$

where h_i is the thickness of the i-th analysis layer, and M the total number of analysis layers. The above summation process is repeated for each traffic loading level, sub-season, and month of the analysis period. For the detailed analysis procedure used to predict permanent deformation for flexible pavements, refer to [1].

Therefore, the MEPDG contains three rutting models, one for each type of pavement material (AC, granular material, and subgrade soil). In total, there are five calibration coefficients that are subject to adjustment during local calibration: three in the HMA model ($\beta_{AC}, \beta_T, \beta_N$), one in the unbound granular materials (β_{GB}), and one in the fine-grained materials (β_{SG}). Note that $\beta_{AC}, \beta_{GB}, \beta_{SG}$ serve as a scaling factor that changes proportionally the permanent deformation in each layer along the whole life. In contrast, β_N is an exponent parameter associated with N and it changes the overall time profile of the permanent deformation curve of the AC layer. It is clear from the mathematical form that as β_N increases, the absolute value of rut depth will increase. Meanwhile, the other exponent parameter β_T associated with temperature T serves only a localized adjustment of the overall performance curve because of the seasonal variation of temperature. Therefore, the effect of β_T on the overall trend of rutting along time is harder to assess. As for the rutting model of unbound materials, although the moisture content would also change the permanent deformation rate, the coefficients in the function $\phi(N, \alpha)$ are not open for local calibration in the MEPDG. These observations are important because otherwise one would not know which local parameter(s) should be adjusted in the local calibration. Although the local calibration guide of AASHTO [16] suggested one to perform sensitivity analyses before local calibration, the understanding of the mathematical structure of the empirical model is more useful than scattered numerical results. A weakness of the local calibration guide is that too much emphasis is placed on the overall bias and residual sum of squares (RSS) while the growth trend of distresses along time and traffic are not explicitly addressed. For this reason, one must be careful when selecting the proper coefficients to calibrate.

2.2. The Fatigue Cracking Models

MEPDG predicts two types of fatigue cracks: the bottom-up cracking due to repetitive tensile strain at the bottom of the AC layer, and the top-down cracking due to repetitive shear strain on the surface of a pavement along the edge of a travelling tire. A matured bottom-up crack often appears to be an alligator crack, whereas a top-down crack is often longitudinal and meandering along travel paths. The computation of fatigue cracking in MEPDG is based on Miner's cumulative damage concept. According to this concept, fatigue damage $D(t)$ at any given time period t is expressed as

$$D(t) = \sum_{i=1}^{\mathrm{TC}} \frac{n_i(t)}{N_{f,i}} \tag{4}$$

Where i represents the i-th category of traffic loading; TC the total number of traffic loading categories; $n_i(t)$ the accumulative traffic of the i-th category up to time t; and $N_{f,i}$ the fatigue life under the i-th traffic loading. Whereas $n_i(t)$ can be obtained directly from the traffic inputs for any given time period, the fatigue life N_f of a certain traffic loading is a material characterization that is a function of several factors. In MEPDG, it is expressed as [14]

$$N_f = C_V C_H k_f \beta_f \varepsilon_t{}^{-k_\varepsilon \beta_\varepsilon} E^{-k_E \beta_E} \tag{5}$$

where ε_t denotes the tensile strain at the critical location; E the dynamic modulus of AC; k_f, k_ε, k_E the global calibration coefficients and $\beta_f, \beta_\varepsilon, \beta_E$ the corresponding local calibration coefficients; and C_V and C_H two adjusting factors for asphalt volume content and AC layer thickness, respectively. The global coefficients take the following value: $k_f = 0.007566, k_\varepsilon = 3.9492,$ and $k_E = 1.281$. It should be pointed out that the 2nd edition of the Manual of Practice [14] made a wrong revision for the sign of the three coefficients. It is unclear to the authors if the computational codes in the AASHTOware Pavement ME are correct – hopefully yes!

Another noteworthy point is that although the locations of critical strain for bottom-up and top-down cracking are different and thus the calculated N_f for the two cracking damages will be different, the model structure of the fatigue life for the two cracking modes are the same. This means that the sets of global or local calibration coefficients in model (5) must be applicable for both top-down and bottom-up cracking.

The Miner's rule stipulates that once the fatigue damage $D(t)$ reaches one, the material suffers from a 'fatigue failure.' However, the exact failure point is not well defined in practice. When the Miner's rule was initially proposed, it was mainly focused on metals and their end of life was often signified in testing by the rupture of specimens. For asphalt concrete and Portland cement concrete, however, the fatigue failure is a gradual process that is accompanied with stiffness deterioration and crack propagation. Although test standards (AASHTO T321 and ASTM D7460) may specify their own data processing methods to find the exact failure point, the fatigue damage calculated from eq. (5) cannot be directly used for pavement design. To fill this missing gap, the MEPDG developed another layer of transfer functions to translate the unobservable fatigue damage $D(t)$ into field observable cracking damage. In practice, cracking damage is measured by two quantities: extent and severity. The extent quantifies how extensive the cracking damage is in a specific pavement lane area, whereas the severity measures the crack width and sometimes depth as well. The MEPDG uses two transfer functions to predict the extent of the cracking damage only.

For the bottom-up cracking, the transfer model can be simplified as

$$\mathrm{FC}_{\mathrm{bt}}(t) = \frac{1}{1 + \exp(C_1 C_1' - C_2 C_2' \log_{10} 100 D(t))} \times 100\% \tag{6}$$

The fatigue crack is expressed in percentage of the total lane area. C_1' and C_2' are global calibration coefficients, and C_1 and C_2 are local calibration coefficients. Note that the global calibration coefficient $C_2' = 2.40874 + 39.748(1 + H_{AC})^{-2.856}$ obviously is not a constant. In addition, the two coefficients C_1' and C_2' are forced to obey a functional relationship, i.e., $C_1' = 2C_2'$. This relation is forced for the unvalidated assumption the MEPDG adopts that the fatigue cracking is 50% when the calculated

cumulative fatigue damage D equals 1, i.e., the AC material reaches its calculated fatigue life N_f. However, the local calibration guide published by the AASHTO does not clearly specify if this assumption should be honored in local calibration. Previous local calibration studies for the DOTs in the US do not comply to this assumption.

The transfer function for the top-down fatigue cracking was developed in a similar manner, except that the amount of top-down cracking is expressed in terms of relative length, i.e., ft/mile or m/km. This selection was understandable because the longitudinal top-down crack is a linear crack, whereas the alligator, bottom-up crack an areal one. The transfer function is expressed as (note that 1 mile is 5280ft and that there are two wheel paths subject to top-down cracking)

$$\mathrm{FC}_{\mathrm{top}}(t) = \frac{1}{1+\exp(C_3' - C_4' \log_{10} 100D(t))} \times (10560\mathrm{ft/mile\ or\ } 2000\mathrm{m/km}) \quad (7)$$

where $C_3' = 7.0$ and $C_4' = 3.5$ are the global calibration coefficients. Observe that $C_3' = 2C_4'$ for the same assumption of 50% cracking damage at $D = 1$. Note that the current MEPDG does not include any local calibration coefficients in the transfer function. This was possibly because the top-down model is still subject to further development. Anecdotes indicate that drastic change (e.g. a shift to fracture mechanics-based method) will be soon introduced.

2.3. The Thermal Cracking Models

Thermal cracking is often transverse to the travel direction of pavement. It results from repetititve thermal strains due to annual thermal cycling. The MEPDG applies crack propagation theory in fracture mechanics, or the Paris-Erdogan law in particular, to predict the crack propagation, and then uses a transfer function to establish the empirical relation between the mechanistically predicted crack depth ΔC and the observable crack length in ft/mile. The Paris-Erdogan model is expressed as

$$\Delta C = A\,(\Delta K)^n \quad (8)$$

where ΔC denotes the increment of crack depth, ΔK the incremental stress intensity factor, and A and n two fracture parameters for the HMA mixture. Whereas n is often readily obtained from the compliance curve, the parameter A involves a calibration coefficient that, according to the AASHTO documents, is subject to change depending upon the input accuracy level, and thus is subject to local calibration. The parameter A is expressed as

$$A = k_t \beta_{t1} 10^{4.389 - 2.52 \log_{10}(E_{AC}\sigma_m n)} \quad (9)$$

where E_{AC} is the HMA indirect tensile modulus, σ_m the mixture tensile strength, both in psi; and k_t and β_t are the global and local calibration coefficients.

In the prediction of the observable transverse, thermal crack, MEPDG uses a log-normal distribution-based transfer function multiplying with an empirical upper limit β_{t2} shown as the following:

$$\mathrm{TC} = \beta_{t2} \Phi\left[\frac{1}{\sigma_d} \log\left(\frac{C_d}{h_{AC}}\right)\right] \quad (10)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function; β_{t2} in the global model equals 400; $\sigma_d = 0.769$ represents the standard deviation of the log crack depth; and h_{AC} the AC layer thickness. The thermal crack is predicted in terms of ft/mile. Note again that $\beta_{t2} = 400$ represents the upper limit of the predicted thermal cracking amount. However, it is not clear from the AASHTO Local Calibration Guide if β_{t2} is subject to local calibration. In the original model of the NCHRP 1-37A report (Appendix HH, pp.79 [1]), the maximum thermal cracking was assumed to be 200ft in the standard LTPP section of 500ft, corresponding to a crack frequency of 1 transverse crack per 30ft. From Eq. (10) this maximum is reached when the cumulative crack length C_d reaches the total AC layer thickness, i.e., a through crack is formed.

However, the transfer function (10) contains a hidden small but significant mistake. That is, the output of the function is the amount of transverse crack for a standard LTPP section of 500ft long.

Therefore, the unit of TC must be ft/500ft. If the calculated TC is output in terms of ft/mile, as suggested in the Manual of Practice, then the whole function should multiply with a unit conversion factor of 5280/500 or 10.56. If the SI unit (i.e. m/km) is used, the whole function should multiply with 2.

2.4. The Reflection Cracking Models

For overlay and other types of rehabilitated structures, the MEPDG also predicts the amount of reflection cracks along time. It predicts the total AC fatigue cracking that includes both the new bottom-up fatigue crack and the reflection cracks in percentage lane area. It also predicts the total AC transverse cracking including new and reflection cracks in ft/mile. The reflection cracking models in the MEPDG have already gone through a thorough modification since its first version. Whereas the original model is largely empirical and based solely on the bottom-up fatigue cracking of the underlain layers, AASHTOWare 2.2 and later versions consider both fatigue and thermal cracks [17]. In addition, it takes the Paris-Erdogan crack propagation theory as the basic modeling skeleton, rather than the Miner's cumulative damage theory as used in the fatigue cracking. Because all types of crack in the old underlain layers may be reflected and propagated to the upper surface layer, the total incremental crack depth calculated from the Paris-Erdogan law is a sum of components due to the alligator, longitudinal, and thermal cracks. For this reason, the total cumulative critical response parameter (D_T) is estimated as:

$$D_T = C_1 \Delta C_{bend} + C_2 \Delta C_{shear} + C_3 \Delta C_{thermal} \tag{11}$$

where $\Delta C_{bend} = k_1 \sum AK_b^n \, dN$, $\Delta C_{shear} = k_2 \sum AK_s^n \, dN$ and $\Delta C_{thermal} = k_3 \sum AK_t^n \, dN$; C_1, C_2 and C_3 are global calibration coefficients. With the calculated damage D_T, the percentage reflection cracking is estimated by the following transfer function

$$RC = \frac{\alpha}{C_4 + \exp(C_5 \log_{10} D_T)} \times 100\% \tag{12}$$

The software has made all the three k's and five C's available for adjustment and thus local calibration for reflective fatigue cracking and reflective thermal cracking. Therefore, there seems to be in total 16 coefficients that can be adjusted. However, note that C_i and k_i for $i = 1, \ldots, 3$ are duplicating to each other. Local calibration is only required for either one of them and another one could be kept at its global value.

2.5. The IRI Models

IRI measures the overall riding smoothness of the pavement surface. In MEPDG, it is predicted as a function of several factors including the initial IRI after construction IRI_0, the rut depth (RD), the total fatigue cracking as a sum of alligator, longitudinal, and refection cracking in percentage lane area (FC), the total transverse cracking including both new and reflection thermal cracking (TC), and a site factor (SF). The prediction model is expressed as

$$\text{IRI} = \text{IRI}_0 + C_1 \times \text{RD} + C_2 \times \text{FC} + C_3 \times \text{TC} + C_4 \times \text{SF} \tag{13}$$

Since the IRI is a composite performance index, the IRI model can be calibrated only after all of the previous distress models have been adequately calibrated. As far as the calibration coefficients are concerned, there has been some debates whether IRI_0 should be calibrated, and if yes, how it should be determined.

2.6. Summary and Remarks

Table 1 summarizes the empirical functions and the associated coefficients that are subject to local calibration. Except for the IRI model, the other distress models include more than one empirical model to be calibrated. In addition, each model involves several coefficients to be adjusted. Therefore, selection of the proper model and coefficients for the local calibration is a crucial step. Although the AASHTO has published a guide for local calibration in 2010, the document should be selectively

followed for at least two reasons. First, the empirical models are in a state of continuous updating. Some of the revision has rendered the guide partly outdated. Second, local calibration is a very delicate exercise that relies on comprehensive understanding of the interaction between the mechanistic analysis and empirical models, and the interaction between different empirical models. Although the local calibration guide was developed based on many global and local calibration studies as well as complementary sensitivity analyses, the guide itself is also open for further enhancement.

Table 1: Summary of the empirical functions and local calibration coefficients

Distress/Performance	Number of Empirical Functions	LC Coefficients
Rutting	3 (one for each material layer)	$\beta_{AC}, \beta_{GB}, \beta_{SG}, \beta_N, \beta_T$
Fatigue cracking	3 (N_f, FC_{bt}, FC_{top})	$\beta_f, \beta_\varepsilon, \beta_E, C_1, \dots, C_4$
Thermal cracking	2 (A, TC)	β_{t1}, β_{t2}
Reflection cracking	2 (D_T, RC)	$k_1, k_2, k_3, C_1, \dots, C_5$
IRI	1	IRI_0, $C_1, \dots, C_4$

3. NEED FOR LOCAL CALIBRATION

The need for local calibration for Ontario's flexible roads was established before the first HIIFP project was initiated. According to MTO's study back in 2008 and even earlier, it was found that the MEPDG permanent deformation model overestimates rutting for Ontario's roads. After the first comprehensive local calibration database was developed, a need of local calibration for almost all distress and performance models in MEPDG was confirmed [6, 8].

Figure 1 compares the measured distresses with the predicted ones by the default models for rutting, IRI, and fatigue cracking. While rutting is greatly over-predicted, the alligator and longitudinal fatigue cracking both are overly under-predicted. Amazingly, the IRI predictions on average has very small bias; however, its trend with the measured IRI is not clear. These have further confirmed the need for local calibration.

4. SUMMARY OF MAJOR EFFORTS

Since 2010 the Ryerson research team has been commissioned to conduct the local calibration. In the first project, Jannat (2012) developed a database [4] that later turned out to cover predominantly the Marshall type of mixes, which are no longer used by the MTO. Nevertheless, using the database Jannat re-confirmed the dire need for local calibration of the rutting models and identified the major challenges that would be faced if the cracking models were to be calibrated. She further confirmed the need for calibration of the IRI model, although the extent of bias and residuals in this model was less than that found in the other distress models. Based upon her study, the research group, upon discussion with the MTO, established the local calibration priority in the following order: rutting models first, followed by cracking models (fatigue, thermal and reflection), and concluding with IRI.

Using the same calibration database, Waseem (2013) performed a sophisticated rutting model calibration [10], in which he identified the major challenge of rutting model calibration to be the unknown layer contributions among asphalt concrete (AC), granular base and subbase, and subgrade soil. Using the observed rutting series from PMS database, Waseem also experimented with a section-by-section longitudinal calibration method, which was the first attempt at a local calibration study in North America [9]. He also developed a semi-automated calibration procedure, which could drastically improve the efficiency of the calibration exercise. Details of the semi-automated calibration procedure can be found in [10].

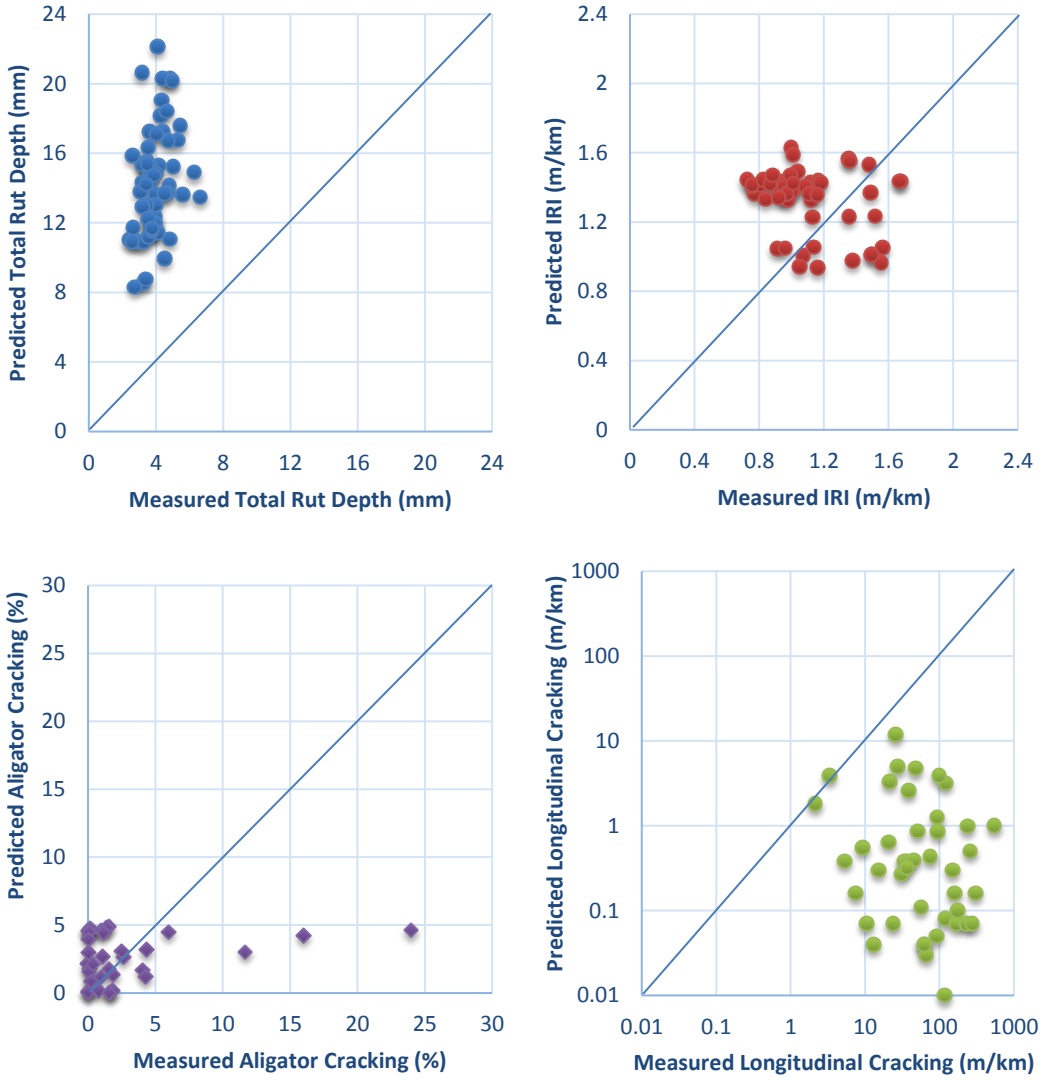

Figure 1: Comparison of the observed and predicted distresses (default models)

Later, the research group realized the defect of the local calibration database. In subsequent studies, the group developed a new calibration database for Superpave and SMA mixes, which have been used in Ontario since 2001. Based upon the new database, Gautam (2015) recalibrated the rutting model and compared the results of Marshall and Superpave materials. In addition, he made a great effort to enhance the local calibration method developed by Waseem earlier on by pre-fixing the two exponent coefficients β_T and β_N in the AC rutting model based on the results of the most recent rutting study, NCHRP project 9-30A [7]. In doing so, the indeterminacy issue due to lack of layer contribution was largely reduced, if not completely resolved.

Using the new Superpave database, Ahmed (2017) performed local calibration for the fatigue and thermal cracking models, which turned out to be a more demanding task. Details of the study were reported in [11]. The first challenge was the processing of the cracking data collected by the ARAN 9000 system and convert them to the observed crack data compatible to the LTPP data. This step is important to obtain reliable local calibration results that are consistent to the global calibration. In the calibration of the fatigue cracking models, Ahmed tried to use a transferred linear model (c.f. Eqs. (6) and (7)) to obtain the optimized local calibration coefficients of the fatigue models (i.e., $C_1, \ldots, C_4$), but the results were not very promising. The reason for this attempt at all was that the two sets of local calibration coefficients could be easily found by using routine linear regression, rather than a constrained optimization. More importantly, with the transformation she tried to establish relatively homoscedastic residuals, which is a neglected assumption in global calibration. For the thermal cracking model, she experimented several values of β_{t1} (or k_t) in the A model shown in Eq. (9) and found that

use of calibration factor of 7 showed the minimum average bias. However, the residual standard deviation did not show significant decrease with different values of β_{t1}.

Gautam's calibrated rutting models include zero or close-to-zero coefficients in the subgrade or base/subbase models, which disagree to engineering experience. This motivated a further enhancement of the rutting local calibration. To do this Amir (2017) did a final examination of the Superpave database. Using the same local calibration method Gautam developed, she obtained a new set of rutting calibration coefficients. In addition, Amir also performed a local calibration of the IRI model. Because the cracking models were not fully calibrated, this IRI calibration should be considered preliminary only. An interesting point of the IRI calibration was the determination of the initial IRI. Several approaches were explored. Details of the enhanced rutting model calibration and the preliminary IRI calibration can be found in [12].

5. THE LOCAL CALIBRATION METHODOLOGY

5.1. General Principles and Procedure

The general principles and steps suggested by the AASHTO Local Calibration Guide [16] was followed. It starts with the need analysis of local calibration, which is shown above, and then development of a local calibration database, which is discussed in next subsection. After that, the local calibration coefficients of each distress model can be determined by minimizing the residual sum of squares (RSS) while maintaining the average bias to be zero. An important note is that although the two objectives of RSS minimization and bias elimination are completely compatible in a full linear model (that is, the RSS minimization will automatically render the average bias to zero, and vice versa), this is not the case in nonlinear models or an incomplete linear model (e.g., a linear regression model with the intercept prefixed to zero). In this case, a constrained optimization can be formulated, i.e., minimizing RSS with an additional constraint to ensure the average bias is zero. Note that an absolute zero bias is meaningless, because another important factor in local calibration is the validation process. Therefore, if the zero-bias constraint has made the optimization very difficult to converge, one may consider to relax this constrain to a certain range and this does not deteriorate the overall quality of the calibration when both the calibration and validation datasets are considered.

5.2. Development of Local Calibration Database

A local calibration database includes three key parts: (1) pavement sections that are selected for calibration and validation, (2) MEPDG input data of each section for distress and performance prediction, and (3) observed performances data.

A guiding principle for the selection of pavement sections is efficiency of local calibration, which means a balance between the efforts of data collection and the statistical power in accepting or rejecting the null hypothesis that a new local calibration model is necessary. From the statistical power point of view, the sample size (i.e., the number of pavement sections) is obviously dependent upon the standard deviation of the residuals of the distress and performance models in the global calibration model. The greater is the residual standard deviation, the more sections will be needed. The AASHTO local calibration guide [16] recommends two statistical equations to determine the sample size. Based on the variability of the LTPP data, the guide suggests minimum numbers of test sections for each distress, and the greatest minimum number is 30 sections. The guide also recommends the use of fractional factorial design to select test sections. This recommendation is followed in this project.

Ontario has more than 16,500 lane-km pavements under the jurisdiction of MTO. These pavements are divided into about 1,800 pavement sections. Majority of these sections are flexible pavement. Early pavements were mainly built with Marshall mixes. A lot of hot mix types, including DFC, HL 1, HL 2, HDBC, etc. (refer to OPSS 1150 [18] for details) were used in different functional layers. MTO started to introduce Superpave mix including SMA into pavement construction as the top asphalt layers in 2001. The most frequently used Superpave mixes include SP 12.5 (and its varieties), 19.0 and 25.0 (refer to OPSS 1151 [19] for details). SMA includes 9.5, 12.5 and 19.0 three mixes. By 2006 all new and reconstructed asphalt pavements and overlaid pavements in provincial roads had been Superpave

specified. As mentioned in Section 4, two local calibration databases were developed. The first one was mainly Marshall types of section, and the second Superpave. Since the key results of the local calibration are for the Superpave sections, the subsequent discussions are all based on the Superpave database.

For the Superpave database, a total of 87 Superpave projects was selected, which contains 140 pavement sections in the pavement management system. These projects have different pavement structures and highway types, spreading over all five climatic zones of Ontario. After a series of data cleansing, only 63 sections are qualified for calibration because the rest lack reliable input information. The 63 pavement sections are further divided into a calibration set of 46 sections and a validation set of 17 sections.

In terms of input data, the best available information was used. Contract documents were retrieved to find the exact project, site, material and structural data. Existing pavement conditions were mainly Level 3 accuracy. The provincial default parameters [20] were largely used for various material characterizations. Level 1 to Level 3 data were used for climate and traffic inputs. Traffic growth rate was calibrated to the historical AADTT data.

Unlike the global calibration, many local calibration studies reported in literature had to rely on pavement performance data collected for pavement management system (PMS). Ontario is no exception. One challenge faced in the study was the changing data collection technologies. Although MTO is very well-known for its long history of the PMS, a vast majority of the long-term pavement performance data collected since 1980s could not be directly used for the local calibration study since the surface crack data before 2011 were manually collected by windshield method. Cracks were evaluated in terms of severity and extent, each scored 1 to 3. Therefore, the cracking data were of subjective and categorical nature. MTO started in 2013 to use the modern laser scanning and image processing technology (the patented LCMS technology in particular) to measure the severity and extent of various cracks at different location of a pavement surface. Accuracy of IRI and rut measurement has also been drastically improved since they were first measured in early 2000s.

The rutting data of the selected pavement sections were retrieved from MTO's PMS-2 system. Although MTO has started to collect rut depth data since 2002, the data collection technology has undergone a significant change in 2012. Starting in 2012, the rut depth was collected using Fugro Roadware's ARAN 9000 automated pavement data collection system. ARAN 9000 contains a Laser XVP that uses two synchronized, laser-based devices to measure the transverse profile of a 4.1m (13.5ft) lane width, with a lateral resolution of approximately 1,280 points. The rut depth measure accuracy is reported to be 1mm [5], which represents a significant improvement from the previous rut depth measurement system. For this reason, only the rutting data of Year 2012 were used in the local calibration. This type of local calibration is called cross-sectional calibration because only one-year of data are used. This is in contrast with the so-called longitudinal calibration in which the multiple-year longitudinal histories of rutting are used to track the prediction trend. Although the rutting data used in the study were collected only in 2012, the data covers a wide range of pavement age from 1 year to 11 years. Therefore, it can be stated that the rutting data have a good life-cycle representation.

Observed cracking data is a key input of the calibration study. Due to complexity, the issue is discussed separately in the next subsection. For all the selected sections, observed distress data and related section name, route name, route direction, station beginning mile, station end mile, facility type, functional class, AADTT, sub-grade modulus, axle configuration, vehicle class, materials properties etc. were collected and compiled to the proper format for the use of local calibration. The latitude, longitude and elevation for specific section are collected from google map and complied accordingly in to the data file. Finally, the integrated database was used for pavement analysis.

5.3. Preprocessing Protocol for Observed Cracking Data

As mentioned earlier, the field measured data provided by MTO using the ARAN-9000 LCMS technology cannot be directly used in calibration for two reasons: First, there is no direct match between the cracking morphology detected from the LCMS images of ARAN9000 vehicle and the cracking

mechanisms evaluated in MEPDG. The morphologically based cracking report as the cracking extent and severity is sufficient for pavement management because it provides adequate information for asset management decision-making. However, the same reporting without identifying the root causes of the cracking cannot be used for local calibration. Unfortunately, this cannot be easily resolved unless more accurate sensing technology has been applied in pavement evaluation.

Second, pavement sections in MTO's current PMS, and thus the sections in the local calibration databased developed above, are not defined the same way as the sections in LTPP were defined. In the local calibration database, the length of pavement sections varies from 0.7km to 30km. However, the LTPP database specifies 500ft (152.4m) as the standard length of a pavement section [1]. The development and global calibration of the cracking models, as well as the local calibration in the USA, were all hinged upon the standard dimension of the pavement sections. The local calibration guide [16] also suggests the use of the same dimension of the pavement section as in LTPP database. If not properly done, this length difference may result in data inconsistency, which may cause unnecessary model bias after local calibration. It may also result in invalid design thresholds for cracking distresses. While analyzing the raw data, it was found that the crack damage is more in shorter length and is decreasing with increase in section length. So, converting the PMS-2 performance data into MEPDG format is one of the major constraints in this study.

With the data given as they are, the best one can do is to follow some mutually agreed data process protocol so that other people in local calibration, if they so wish, can repeat the work and obtain the same results. Fortunately, MTO maintains the raw crack data recorded for each section at every 50m interval. These data can be used to derive a consistent observed cracking damage value that is comparable to the cracking damage defined in LTPP and MEPDG. To keep consistency with the global calibration of the MEPDG format, the pavement sections used in this study was decomposed into 150m long segments to match with MEPDG format. This decomposition results in a number of 150m segments for each pavement sections. It is not hard to imagine that each of the 150m segments has different cracking damage. Therefore, how to aggregate the 150m-segment cracking data into one cracking damage data becomes an issue. To do this, two aggregation methods were tried: The first one is maximization that takes the maximum value of the 150m-segment crack data, and the second is the averaging method that reports the average value. For alligator damage, only areas of wheel-path and mid-lane were counted. Crack data are divided by section area (since data are recorded as square-meter). For longitudinal damage, crack data are divided by section length and multiplied by 1000 to get the data in meter/kilometer. Thermal damage calculation is same as longitudinal damage.

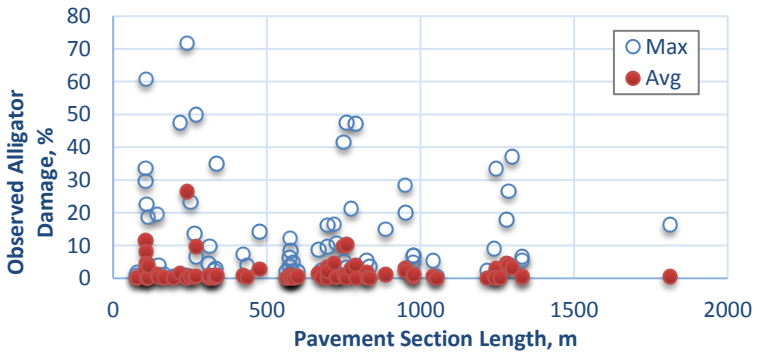

Figure 2: Comparison of the maximum and averaged alligator cracking

For illustration purpose, Figure 2 depicts the large difference between the two aggregation methods for alligator cracking. The mean and standard deviation of the maxima are 13.5% and 15.9%, respectively, whereas those of the averages of the same sections are only 1.97% and 3.83%. Unsurprisingly, the maximization method offers much greater mean and standard deviation. This large different results from the relative large spatial variation of cracking damage along the traffic direction. The decision between the two aggregation methods is not easy. The maximization method provides more conservative local calibration model, whereas the averaging method provides more stabilized model due to its smaller standard deviation. On the other hand, this comparison also brings up the issue

whether the PMS data are reliable enough to support local calibration. These are open questions for further study.

For thermal cracking, a weighted average has been used in the data input, after processing the observed data. Weighted value for the observed cracking is considered by using the following formula:

$$\text{Transverse Damage} = \frac{1}{9} \times (\text{Low} + 3 \times \text{Medium} + 5 \times \text{High}) \quad (14)$$

5.4. The Calibration Method for Rutting Models

The key challenge of the rutting model calibration is the identifiability issue due to the lack of layer rutting information. Different layer contributions to rutting would result in different sets of calibration coefficients. Unfortunately, all one has is the total surface rut depth. As a result, the scale coefficients cannot be uniquely determined. This identifiability issue is not new. It appeared in the global calibration, but has not been properly addressed; for more details, refer to [9]. We proposed a novel approach by pre-fixing the traffic and temperature exponents in the AC rutting model. After these two exponent coefficients are fixed, the other three scale coefficients are then determined by optimization.

To determine the two exponent calibration factors, a series of secondary analyses were performed based upon results from the recalibration study of NCHRP 9-30A project. For the traffic exponent $m = 0.4791\beta_N$, the section- or specimen-specific traffic exponents presented in NCHRP Report 719 of Project 9-30A [3] were studied. With a range from 0.15 to 0.55, the 88 data points of m have a mean value of 0.30 and standard deviation 0.08. Detailed statistical tests concluded that

- The new constructed sections and the overlay sections have the same traffic exponent;
- The field-derived traffic exponent can be the same as the laboratory test-derived traffic exponent, provided that the test is performed under the constant-height shear testing protocol;
- The traffic exponents obtained from the triaxial test are significant different than the field-derived traffic exponents and those from the shear tests.

Note that in the global calibration model, the traffic exponent equals 0.4791, which is greater than the exponent of 84/88 cases, or 95 per cent of the cases. This has partly explained why the default rutting model always over predicts rut depth. In addition, Report 719 also compared other three AC rutting models: the Asphalt Institute (AI) model, the modified Leahy model, and the Verstraeten model, which all include a traffic term in exponential form. The traffic exponent values of the three models are 0.4354, 0.25, and 0.25, respectively. Furthermore, Waseem and Yuan [9] performed a longitudinal calibration of the rutting models based on the Marshall-mix sections and found the traffic exponent varying from 0.11 to 0.57. Finally, a sensitivity study was carried out to check the overall impact of the traffic exponent (m) on rutting prediction. Under different m values ranging from 0.17 to 0.57, fourteen Superpave sections from the MTO database were analyzed. It was found that the predicted rut depths do not change significantly when $m \leq 0.35$. With these considerations, it was concluded that the traffic exponent be set at 0.30 (or $\beta_N = 0.30/0.4791 = 0.6262$) for the subsequent local calibration study.

Regarding the temperature exponent β_T, Report 719 found that the laboratory-derived temperature exponent is largely dependent upon the material testing methods. The histograms of the triaxial and shear loading test data reported in Tables 24 and 25 of the Report indicate that the triaxial results are systematically smaller than those from the shear tests. The mean value of the triaxial test results is 2.665, which is close to the exponent value set in the AI model, whereas the mean of the shear test is 7.720. This large variation in the temperature exponent blurs the issue. It is our belief that an intensive study of the temperature effect is required to settle down this issue. With consideration that many past local calibration studies simply left the temperature exponent at the default value, this study also chose to fix the temperature exponent at its global value of 1.5606. In other words, the local calibration factor β_T is preset to 1.0 in this local calibration study.

Once the two exponent coefficients are determined, the local calibration of the three scaling coefficients, namely β_{AC}, β_{GB} and β_{SG}, can be readily estimated by Excel Solver or any equivalent

optimizer. More importantly, no AASHTOware computational iterations are needed. All one needs to do is to compute the rut depths of the different structural layers (AC, GB and SG layers) with $\beta_N = 0.6262$ and $\beta_T = \beta_{AC} = \beta_{GB} = \beta_{SG} = 1.0$. After that, the rut depth of each layer at different scaling coefficient values can be obtained by multiplying the so calculated rut depth with the corresponding scaling coefficient. The conventional RSS minimization with zero bias constraint is then used to find the optimal solutions for the three scaling coefficients.

It has to be reiterated that this method does not completely resolve the identifiability issue; it only lessens the severity. For a certain calibration set of pavement sections, it is still possible that the results are converged to zero or negative values, which is nonsensical; refer to [7] for examples of this phenomenon. Trench analysis can completely solve the problem, but it is expensive.

5.5. The Local Calibration Method for the Fatigue Cracking Models

As discussed in Section 2, the MEPDG contains three empirical models for fatigue cracking. However, the N_f model relies heavily on material characterization. Without detailed Level 1 material inputs, particularly fatigue life data, calibrating the N_f model would lack solid theoretical support and have little practical significance. The local calibration is thus focused on the calibration of $C_1 \ldots, C_4$ used in Eqs. (6) and (7). The two transfer functions both are nonlinear function of $D(t)$, a mechanistically computed quantity. To calibrate the C's, there are two alternatives. The conventional method is by comparing the observed and predicted cracks. Alternatively, one can also take a logarithmic transformation of the transfer functions to obtain a linear function, and then the C's can be obtained by simple linear regression. For example, the transfer function of the bottom-up cracking can be modified as:

$$\ln\left(\frac{100}{FC_{bt}} - 1\right) = C_1 C_1' - C_2 C_2' \log_{10} 100 D(t) \quad (15)$$

Treating the left hand side as the prediction and comparing the observed equivalence results in a typical linear regression formulation that can readily solve the unknown C_1 and C_2. Similar transformation is done for the top-down cracking model.

The alternative method is advantageous to the conventional one in two counts. First, the RSS minimization and bias elimination become compatible—accomplishment of one objective leads automatically the accomplishment of the other. Second, the residuals can be more conveniently examined in the linear models. Since the conventional method is widely used in practice, the results of the conventional method is presented in the paper. For results of the other method, refer to [11].

5.6. The Local Calibration Method for the IRI Model

The main challenge of calibrating the IRI model was the determination of the initial IRI, or IRI_0 in eq. (13). IRI_0 refers to the roughness of pavement within six months after construction completion. According to the software manual [21], "the initial IRI value provided must be what is typically attained in the field." The Ontario's construction quality assurance practice is, according to Special Provision SSP 103F31 [22], when the post construction IRI measurement is between 0.650 to 1.000m/km, the payment adjustment factor is 1.0, i.e., no payment adjustment. In other words, the MTO expects that the initial IRI of a readily acceptable new pavement will be in that range. The default value used in AASHTOware is 1 m/km [20]. The Superpave calibration database includes the IRI measurements in the first year of the pavement, which range from 0.72 to 1.97m/km. But this measurement, strictly speaking, is not the initial IRI that should be measured at the time when the payment adjustment factor was determined. However, the real IRI_0 is unknown. Therefore, a debate arose during the local calibration what IRI_0 value should be used in the local calibration and future design.

For ease of explanation, the debate is divided into two related parts. First, what IRI_0 value should be used in the local calibration? Second, can we use the IRI_0 used in or determined by the local calibration in future design?

Regarding the first question, three schemes have been suggested. The first suggestion was to use the actual initial IRI values as a Level 1 input for the calibration. This would result in R^2 value in the local calibration. However, some argue that this method is not recommended, because one should not forget that IRI_0 is part of design specification and it is similar to design reliability, failure thresholds and design life that is not the value that is actually achieved, but one that is expected. Some other people defend that they use the actual IRI_0 only in local calibration to determine the local calibration coefficients C_1 to C_4. In future design, the IRI_0 value should be a unified, rather than section-specific, value that reflects the owner or authorization's expectation on the initial IRI. However, there is a doubt if so-calibrated IRI model would underestimate the variability of the IRI residuals and thus over-predict the pavement reliability.

The second suggestion was to treat the IRI_0 as a local calibration coefficient and let the optimization (a simple linear regression) determine the optimal one. However, it was soon realized that sometimes the so-calculated IRI_0 could be sensitive to the local calibration data, particularly when only one year (either 2014 or 2015) IRI data are used for the local calibration. It is possible in theory that the calculated IRI0 will be a negative value. Therefore, the result was not robust. The people supporting this suggestion defends that this method at least captures the growth trend of IRI along time, which is the main purpose of the local calibration. They further suggested that in future design the IRI_0 value should be determined separately.

The third and last suggestion was to use a default value that is determined by the agency in both local calibration and future design. This suggestion treats the initial IRI as an exogenous variable of local calibration. But a major problem is that this method often results in a very poor calibration.

Regarding the second question, it is generally agreed that the IRI_0 value for future design should be determined separately. To determine this value, the post-construction IRI measurements can be used to perform the statistical analysis. The mean, standard deviation and probability distribution of the initial IRI can be determined. If necessary, regression models that relate it to other explanatory variables such as road functional class, AC type, geographical zone, season of construction, etc. can be developed. After these statistical results are obtained, the determination of the initial IRI as a policy number requires several considerations. First, one needs to consider the current smoothness-related payment adjustment practice. This does not suggest that we simply follow the current practice and adopt the current 100% pay line. Rather, the statistical analyses provide a great review of the practice and impact analysis can be done if a change is made. Second, the impact of the design reliability needs to be assessed. Note that the current MEPDG uses exclusively a residual-based method for the reliability analysis. For IRI model, only the model residual of the global calibration is considered. Now that the IRI_0 is treated as an exogenous variable, which is also a random variable, the variability should be included in the standard deviation function. If the suggested IRI_0 value is not the mean value, then the calculation of the reliability needs to be adjusted. Refer to [23] for more detailed discussion of the reliability models of the MEPDG.

6. RESULTS & DISCUSSIONS

6.1. The Rutting Model

Using the local calibration database and method presented above, the three scaling local calibration coefficients are determined as follows: $\beta_{AC} = 1.7692$, $\beta_{GB} = 0.0968$, $\beta_{SG} = 0.2787$ with the residual standard deviation reduced to 1mm and average bias to 0. This represents a significant improvement compared to previous studies. The summary statistics of the calibration and validation sets are presented in Table 2 and compared with the global calibration result. Figure 3 shows the scatterplot of the observed vs. predicted rut depth for both calibration and validation sets. Both have shown satisfactory results.

The layer contributions to total rutting are a very important indicator to the quality of the model, as this can be directly benchmarked with engineering experience. These results are presented in Table 3 below. As it can be seen, the layer contributions of the local calibration model are very close to those of the global model. That being said, however, note that the global calibration model was originally

calibrated to the early 1960s AASHO Road Test results in terms of the layer contributions [1, 9]. Therefore, the small gap to the global model itself does not validate that the local calibration model is benchmarked to the actual situations in field. Several studies have been done in order to determine the realistic layer contributions, but majority of them were based on software calculations. But one needs to remember that rutting is a permanent deformation resulting from complicated nonlinear viscoelastoplatic behaviors of asphalt concrete and granular materials. Using software to validate the layer contributions suffers from circular reasoning. Only can extensive trench analyses determine the actual distribution of total rutting among different layers.

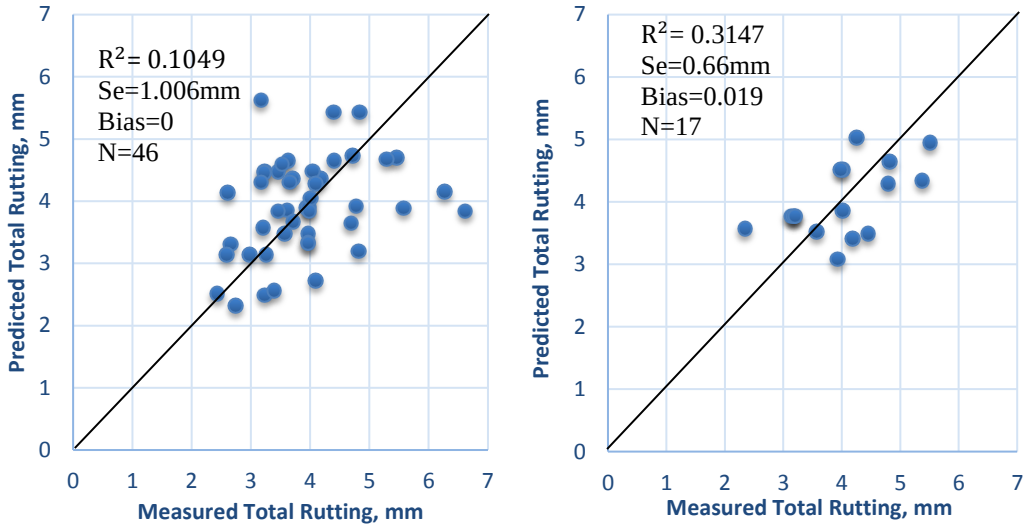

Figure 3: The predicted vs. measured total rut depth after local calibration (left – calibration sections; right – validation sections)

Table 2. Summary statistics of the rutting model calibration

Statistical Parameters	Calibration	Validation	Global Calibration
R^2	0.1049	0.3147	0.577
Residual standard deviation (mm)	1.006	0.66	2.717
Number of data points	46	17	334

Table 3. Layer contributions to total rutting

Layers	Global model	Calibrated Model
AC	23%	19%
GB	11%	9%
SG	66%	72%

The current AASHTOware allows different rutting models for different AC courses. For example, the surface course, binding course and existing overlaid AC layers can have different calibration coefficients. This can be an interesting question because the new courses are now Superpave mix whereas the existing AC layers are still mainly Marshall mix. We did not proceed further because this would just worsen the abovementioned identifiability issue. However, given that we have developed two different datasets, it would be interesting to re-evaluate the models when separate Marshall and Superpave models are used for different AC layers.

A weakness of our rutting calibration is the relative small range of the observed rut depth data. The greatest rut depth in the database is less than 7mm. However, the rutting failure threshold in Ontario is

12mm for freeways. Therefore, more data are required to further calibrate and validate the rutting model. Another weakness is that the current local calibration, and in fact, majority of local calibration studies using PMS data, does not consider the longitudinal trend of the calibration. In order to do so, longer time series of observed rutting data are needed. Waseem and Yuan (2012) did a longitudinal local calibration using the Marshall section database [9]. Their calibration method can be used when more rutting time series are available.

6.2. The Bottom-Up Fatigue Cracking Model

A total of 44 pavement sections was selected for the calibration of the cracking model. These sections have a well coverage in terms of pavement age and the amount of alligator cracks within the pavement sections. The resulting standard deviation of residuals (Se) obtained from the local calibration 6.14%, very close to the Se of 5.01% in the global calibration.

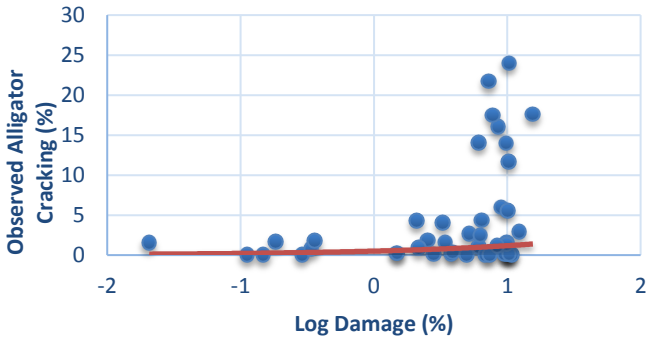

Figure 4: Calibration results of the alligator cracking model

Although splitting the results into new and overlay sections would result in smaller Se for those overlay sections, it is a bias to choice that only good overlay sections are used for the calibration. Those overlay sections have gone through a strict selection process to ensure that the sections with propagation of reflection cracks have been eliminated.

With a default N_f model, the calibration conclude that $C_1 = 0.5236$ and $C_2 = 0.1404$. Figure 4 shows the plot that is very similar to that of the global calibration. Significant heteroscedasticity is observed in the residual plots and this is why the fatigue cracking model also includes a nonlinear standard deviation model. This standard deviation model is not calibrated in this study.

6.3. The Other Cracking Models

Our study indicates that the other cracking models, namely the top-down fatigue cracking model, the thermal cracking model, and the reflection cracking model cannot be properly calibrated at this moment. The reasons are multifold, but the major one is that these models are subject to substantial improvement from the original model developers. Once the models are finalized, they can be calibrated using the same database we have developed.

6.4. The IRI Model

Only 20 new construction or reconstruction sections were considered for the IRI calibration. The overlay sections were not included because the reflection crack models have not been calibrated and yet the amount of reflection cracks observed in the overlaying sections were very high in magnitude (above 30%). As a result, only the first two calibration coefficients associated with rutting and fatigue cracking were calibrated while keeping C_3 and C_4 unchanged as in the global model. Using the actual initial IRI of each section, it was found that $C_1 = 55.0960$ and $C_2 = 1.0883$. Figure 5 shows the predicted versus measured IRI after local calibration.

It has to be pointed out that the residual standard deviation did not change before and after local calibration. It remained to be 0.30 m/km. However, the average bias was reduced from 0.11 m/km to

zero. In addition, the second method introduced in Section 5.6 that treating IRI_0 as a local calibration coefficient was also tried, but it resulted in a negative C_1.

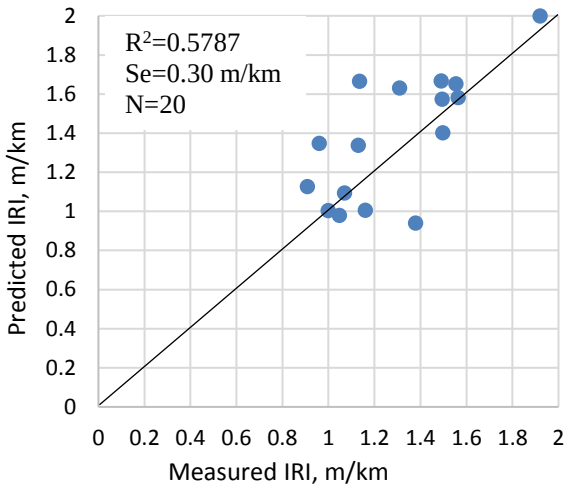

Figure 5: The predicted vs. measured IRI after local calibration.

7. CONCLUSIONS

The Ontario's local calibration has gone a long way since 2010. Over the past seven years, a substantial amount of time has been spent on the development and refinement of the local calibration database. Among them, a lot of work has been devoted to the identification of proper pavement sections. Equal amount of time has also been spent on the retrieval and post-processing of the historical performance data, particularly in the case of cracking data. Extensive experience has accumulated during this process. The best results have been achieved on the calibration of the rutting models and the bottom-up fatigue cracking model. Great efforts have been made to try to calibrate the other models, with limited success. It was originally hoped that with the new ARAN data the cracking models can be reasonably calibrated to the local conditions and practice. However, due to the pending development of the top-down and thermal cracking global models, the local calibration work for these models is held off.

Obviously, there are many outstanding issues to be addressed in order to fully calibrate the MEPDG distress and performance models, for example, the determination of initial IRI, sensitivity analysis of different protocols to preprocess the observed cracking data, calibration of the other three cracking models, collecting and including longer distress time-series into the calibration database, verification of the time trend of the empirical models through longitudinal calibration, calibration of the standard deviation functions, and impact analysis on pavement design and rehabilitation.

Nevertheless, the rutting and alligator cracking models have been well calibrated. With prefixed exponent coefficients for the AC rutting model, the other three scaling coefficients have been accurately calibrated and validated. Similarly, the alligator cracking models were calibrated and validated with a prefixed N_f model. The IRI model has also been calibrated to the best in the sense that only the coefficients associated with the rutting and alligator cracking models were determined. Note that the default IRI model has little bias anyway, the significance of the IRI model is less important than the other distress models.

As reviewed in Section 2, the MEPDG is a very complicated pavement analysis system. Different analysis modules are at different maturity stage. Some models (e.g., rutting, bottom-up fatigue cracking) use theories that are generally agreed in the community, but some distress phenomena such as the top-down fatigue cracking and reflection cracking are still subject to further development and verification. Meanwhile, new materials and new construction and rehabilitation technologies are emerging every year, if not every day. All these force the local calibration to be a dynamic process. The agencies should also be ready to invest on collecting more Level 1 inputs particularly in traffic, AC material characterizations, soil characterizations, and existing conditions.

8. REFERENCES

1. NCHRP, *Guide for Mechanistic Empirical Design of New and Rehabilitated Pavement Structures*. 2004: ARA, Inc., ERES Division 505 West University Avenue Champaign, Illinois 61820.
2. Brown, S.F., M.R. Thompson, and E.J. Barenberg, *Research Results Digest 307*, in *Research Results Digest, National Cooperative Highway Research Program*. 2006.
3. Von Quintus, H.L., et al., *NCHRP REPORT 719 - Calibration of Rutting Models for Structural and Mix Design*. 2012, AASHTO: Washington, D.C.
4. Jannat, G.E., *Database Development for Ontario's Local Calibration of Mechanistic-Empirical Pavement Design Guide (MEPDG) Distress Models*, in *Department of Civil Engineering*. 2012, Ryerson University.
5. Fugro Roadware, *Laser XVP*. 2013: http://www.roadware.com/related/Laser-XVP_2014.pdf.
6. Jannat, G.E., et al., *Database Development for Ontario's Local Calibration of the MEPDG Distress Models for Flexible Pavements* in *The 9th International Transportation Specialty Conference, CSCE*. 2012: Edmonton, AB.
7. Gautam, G.P., et al., *Local calibration of the MEPDG rutting models for Ontario's flexible roads: recent findings*, in *The 2016 Annual Meeting Transportation Research Board*. 2016: Washington DC. p. poster.
8. Jannat, G.E., X.-X. Yuan, and M. Shehata, *Development of regression equations for local calibration of rutting and IRI as predicted by the MEPDG models for flexible pavements using Ontario's long-term PMS data.* International Journal of Pavement Engineering, 2015. **17**(2): p. 166-175.
9. Waseem, A. and X.-X. Yuan, *Longitudinal local calibration of MEPDG permanent deformation models for reconstructed flexible pavements using PMS data.* International Journal of Pavement Research and Technology, 2013. **6**(4): p. 304-312.
10. Waseem, A., *Methodology Development and Local Calibration of MEPDG Permanent Deformation Models for Ontarios Flexible Pavements* in *Department of Civil Engineering* 2013, Ryerson University.
11. Ahmed, C.J.S.E., *Local Calibration of Cracking Models of MEPDG for Ontario's Flexible Pavements*, in *Department of Civil Engineering*. 2017, Ryerson Univeristy: Toronto, ON, Canada.
12. Amir, M., *Local Calibration of MEPDG for Superpave Pavements in Ontaro: Enhanced Rutting Calibration and Preliminary Study on IRI Model*, in *Department of Civil Engineering*. 2017, Ryerson Univeristy: Toronto, ON, Canada.
13. ARA, *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures: Appendix C*. 2004: Transportation Research Board.
14. AASHTO, *Mechanistic-Empirical Pavement Design Guide: A Manual of Practice*. 2nd ed. 2015, Washington DC: American Association of State Highway and Transportation Officials.
15. AASHTO, *Mechanistic-Empirical Pavement Design Guide: A Manual of Practice, Interim Edition*. 2008, Washington DC: American Association of State Highway and Transportation Officials.
16. AASHTO, *Guide for the Local Calibration of the Mechanical-Empirical Pavement Design Guide*, ed. Joint Technical Committee on Pavements. 2010, Washington DC: American Association of State Highway and Transportation Officials.
17. Lytton, R.L., et al., *Models for Predicting Reflection Cracking of Hot-Mix Asphalt Overlays*, in *NCHRP Report 669*. 2010, National Cooperative Highway Research Program: Washington, D. C.
18. Ontario Provincial Standard Specification, *Material Specification for Hot Mix Asphalt*, in *OPSS 1150*. 2008, Ontario Provincial Standard Specification: www.raqsb.mto.gov.on.ca.
19. Ontario Provincial Standard Specification, *Material Specification for Superpave and Stone Mastic Asphalt Mixtures*, in *OPSS 1151*. 2006, Ontario Provincial Standard Specification: www.raqsb.mto.gov.on.ca.

20. MTO, *Ontario's Default Parameters for AASHTOWare Pavement ME Design, Interim Report*. 2014, Downsview, ON: Ontario Ministry of Transportation, Pavement and Foundations Section, Materials Engineering and Research Office.
21. AASHTO, *AASHTOware Pavement M-E Design v2.0 Help Manual*. 2014, AASHTOWare.
22. Ontario Provincial Standard Specification, *Asphaltic Concrete Payment Adjustment for Surface Smoothness Based on Quality Assurance Measurements Taken by an Inertial Profiler, Special Provision No. 103F31*. 2014.
23. Yuan, X.-X., M. Shehata, and G. Jannat. *Evaluation of the Reliability Method Used in the AASHTO MEPDG*. in *International Symposium on Reliability Engineering and Risk Management (ISRERM'2012)*. 2012. Kanagawa University, Yakohama, Japan.