Exploring the Need for Having Region-Specific Calibration Factors

Srinivas Reddy Geedipally*, Ph.D., P.E.
Associate Research Engineer
Texas A&M Transportation Institute, Arlington, TX 76013, United States
Tel. (817) 462-0519
Email: srinivas-g@tti.tamu.edu

Mohammadali Shirazi
Ph.D. Candidate, Zachry Department of Civil Engineering
Texas A&M University, College Station, TX 77843, United States
Email: alishirazi@tamu.edu

Dominique Lord, Ph.D.
Professor, Zachry Department of Civil Engineering
Texas A&M University, College Station, TX 77843, United States
Tel. (979) 458-3949
Email: d-lord@tamu.edu

Paper submitted to the 96th Annual Meeting of the Transportation Research Board

Word Count: 4,593+ 6*250 (4 tables + 2 figures) = 6,093 words

April 4, 2017

*Corresponding Author
ABSTRACT

States, or even large urban cities, may experience different crash rates in different regions or parts of the city. This can be attributed to differences in terrain, population, weather, and other unobserved characteristics. Hence, it can impact the calibration procedure and consequently the calibration factor when it is used for a very large area. This study first investigated whether or not having region-specific calibration factors are required and justified for large states, such as Texas and Michigan. Next, a procedure was proposed to determine if a specific region needs a calibration factor that is different from the one developed for the whole state. If it is determined that a separate factor is needed, then the agency should derive a region-specific calibration factor using data that are collected within the region. Otherwise, the common statewide factor can be used. The proposed procedure is based on the general characteristics of data at the network level: (1) the total number of crashes, (2) the mean value of traffic flow (ADT/AADT), and (3) the total segment length (or the number of intersections). The procedure was validated for the two states listed above and showed that the calculated calibration factor matched the factor proposed in the recommended decision making procedure.
1. INTRODUCTION

Safety Performance function (SPF) models play an important role in various stages of systematic safety analysis and the improvement in safety. These models are used to estimate the frequency of crashes and conduct different types of safety evaluation or analysis. The Highway Safety Manual (HSM) (1) has started documenting pre-fitted SPF models for different facility types, such as rural two-lane roads, rural multilane highways, and urban and suburban arterials. However, since these SPFs were fitted and validated with data from a few selected number of states in the United States, they are needed to be calibrated to the local conditions when applied to a new jurisdiction. Calibration is a tool to account the variation in characteristics of crash data in different jurisdictions or differences in factors that cannot or were not included in the SPF models, such as weather patterns or driver behavior.

The HSM provides a methodology to calibrate the model through a scalar calibration factor (C-factor). The C-factor is estimated after collecting detailed data for an adequate sample-size (Note: see Bahar (2) and Shirazi et al. (3) for guidelines on required sample size) as follows (Eq. 1):

\[
C = \frac{\sum_{i=1}^{n} N_{\text{obs},i}}{\sum_{i=1}^{n} N_{\text{pre},i}} = \frac{\sum_{i=1}^{n} N_{\text{obs},i}}{\sum_{i=1}^{n} \prod_{j=1}^{d} N_{\text{base},i} \times CMF_{ji}}
\]  

where,

- \(C\) = calibration factor;
- \(N_{\text{obs},i}\) = the observed number of crashes at site \(i\);
- \(N_{\text{pre},i}\) = the unadjusted predicted number of crashes at site \(i\);
- \(N_{\text{base},i}\) = the predicted number of crashes at site \(i\) at the base conditions;
- \(CMF_{ji}\) = the \(j\)-th crash modification factor (CMF) at site \(i\);
- \(n\) = sample-size; and,
- \(d\) = number of CMFs in the model.

Since the release of HSM in 2010, state-wide C-factors (SWCFs) have been estimated for different facilities in the United States, including but not limited to Oregon (4), Illinois (5), and Alabama (6). In almost all these cases, it was assumed that the crash pattern and the effect of unobserved variables in the SPF is uniform and spread homogeneously across the state. Hence, data across the whole state were used to estimate a single state-wide C-factor. However, this assumption may not hold true when large states or even large cities are considered. In fact, the C-factor might be different within a region of a large state because attributes within that region are not uniform across the entire area. The HSM also noted this issue and recommended finding separate C-factors for large jurisdictions that are characterized by different topographical or weather conditions (1). Unfortunately, the HSM does not provide guidelines for determining the detailed conditions when separate C-factors are warranted or justified. Without guidelines, it is not possible to know whether the region-specific calibration factors are needed or not.

Recently, Bahar (2) studied the conditions needed for having separate C-factors across the states and provided a practical approach to find a separate C-factor. The core points of the study can be reviewed as follows:
First, it was hypothesized that there is no need to be more precise with the C-factor than the errors involved in the base model or the product of CMFs. Based on this hypothesis, a conservative guideline was provided, “...the coefficient variation of the C-factor does not need to be less than, say, half of the coefficient of variation of the product of the CMFs.” If this error is met, there is no need to find a separate C-factor for the model; otherwise, having a separate C-factor is advised. However, the document did not provide clear guidelines on what the typical CV of the product of the CMFs should be and calculation of the CV of the product of CMFs was left to the user (see Lord (7) for additional discussion about calculating the variance for the product of CMFs).

Second, it was suggested to collect data across the state; then, group data based on different variables and conditions, such as AADT, segment length or crash severities. If a major difference across the groups was observed, a separate C-factor is suggested. This method is not only useful to consider a separate C-factor for a region or terrain but also can be used to consider separate C-factors for sites that seemed to have similar patterns based on collected data. For example, different C-factors can be recommended for different ranges of AADTs.

The method proposed by Bahar (2) is based on the availability of detailed data that are used for the calibration process. More specifically, it was assumed that the analyst first collects all data that are required for the calibration, and then he or she goes through grouping the variables to determine whether or not a separate C-factor is needed, for example, for different regions. This method may not be efficient since the analyst may need to know if a separate C-factor is desired for different regions in advance before the calibration procedure begins in order to collect enough data for the required sample size.

Based on the information provided above, a data-driven procedure is needed for an analyst to make a decision on the need of region-specific calibration factors (RSCFs). The objective of this study is therefore to provide simple and convenient procedure to determine when the analyst is advised to develop region-specific calibration factors before collecting detailed data. The proposed procedure is based on general characteristics of data at the network level. In order to use the procedure, the analyst needs to secure only three sources of information for the regions or areas of interest. These sources are: (1) the total number of crashes (2) the traffic flow average and (3) the total segment length (or the total number of intersections). Once the information is secured, the procedure can simply be used to make a decision on the development of region-specific calibration factors.

2. REGION-SPECIFIC CALIBRATION FACTORS

This section further explores the idea of having region-specific calibration factors in large states by providing two examples. The empirical data from two facility types, one in Texas and the other one in Michigan, are used to accomplish this objective.
First, the region-specific C-factors are estimated for Texas urban four-lane divided arterials. The dataset was collected in three-year frequency from 2007 to 2014. This means, for example, if the traffic and geometric data in 2009 are considered, then the crash data in this dataset will include crashes occurred in a three year period from 2007 to 2009. Texas is divided into four regions: North, South, East, and West. The division of the state is based on administrative boundaries used by the Texas Department of Transportation (TxDOT). Table 1 summarizes the C-factors calculated for each region and the state. To develop the C-factors, the network was first divided into homogenous segments and SPFs from the HSM for the base conditions are applied. For segments that deviate from the base conditions, CMFs are developed and multiplied with the base SPFs to obtain the predicted crashes for each site. All variables in the dataset met the base conditions except for the median width. Thus, the model includes one CMF for the median width. As indicated in Table 1, the difference between the C-factors calculated in different regions is significant. For instance, the analysis with the 2007-2009 data show that the difference between the C-factor in the North region is about 35 percent when compared to the West region. The large difference between the value of RSCFs and SWCF would justify having a separate C-factor for at least some of the regions in Texas. Figure 1 shows the region-specific C-factors for the urban 4-lane divided arterials in different regions of Texas using 2012 to 2014 data. A similar analysis with one-year frequency data can be found in Lord et al. (8).

Table 1. Region Specific C-Factors for Texas Urban Four-Lane Divided Arterials

<table>
<thead>
<tr>
<th>Region</th>
<th>District Numbers</th>
<th>Number of Sites</th>
<th>Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>07-09</td>
</tr>
<tr>
<td>North</td>
<td>1,2,3,9,10,18,19,23</td>
<td>1163</td>
<td>0.947</td>
</tr>
<tr>
<td>South</td>
<td>13,14,15,16,21,22</td>
<td>1549</td>
<td>0.950</td>
</tr>
<tr>
<td>East</td>
<td>11,12,17,20</td>
<td>868</td>
<td>1.081</td>
</tr>
<tr>
<td>West</td>
<td>4,5,6,7,8,24,25</td>
<td>684</td>
<td>1.284</td>
</tr>
<tr>
<td>Statewide</td>
<td>All</td>
<td>4264</td>
<td>1.011</td>
</tr>
</tbody>
</table>
As a second example, the region-specific C-factors were estimated for Michigan four-legged signalized intersections. This dataset consisted of one-year crash frequency. The state of Michigan is divided into seven regions, which is also based on administrative boundaries. The regions were indicated by numbers from one to seven in Table 2. For this dataset, the SPF included four CMFs: left turn lane, right turn lane, right turn on red and lighting. As shown in Table 2, the calibration factors are significantly dissimilar for different regions. For instance, in 2008, the difference between the C-factor in superior region (region 1) is almost 2.5 times larger than the C-factor in Metro region (region 7). Given that there is a large difference between the RSCFs and SWCF in this state, it is justified and recommended to derive RSCFs for Michigan four-legged signalized intersections. Figure 2 shows the region-specific C-factors for the four-legged signalized intersections in different regions of Michigan using 2012 data.

Table 2. Region Specific C-Factors for Michigan four-Legged Signalized Intersections.

<table>
<thead>
<tr>
<th>Region</th>
<th>Region Name</th>
<th>Number of Sites</th>
<th>Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Superior</td>
<td>46</td>
<td>2.62</td>
</tr>
<tr>
<td>2</td>
<td>North</td>
<td>50</td>
<td>2.44</td>
</tr>
<tr>
<td>3</td>
<td>Grand</td>
<td>51</td>
<td>1.66</td>
</tr>
<tr>
<td>4</td>
<td>Bay</td>
<td>50</td>
<td>2.09</td>
</tr>
<tr>
<td>5</td>
<td>Southwest</td>
<td>52</td>
<td>2.63</td>
</tr>
<tr>
<td>6</td>
<td>University</td>
<td>50</td>
<td>1.99</td>
</tr>
<tr>
<td>7</td>
<td>Metro</td>
<td>50</td>
<td>1.14</td>
</tr>
<tr>
<td>Statewide</td>
<td></td>
<td>349</td>
<td>1.83</td>
</tr>
</tbody>
</table>
It is worth pointing out that, for each year, we had about 50 sites in each region of Michigan that can be used to find the region-specific C-factors. This sample size satisfies the HSM recommendation of having 30-50 sites for calibration. However, the CV of crash data for different regions in Michigan 4-legged signalized intersections data varies between 0.8 and 1.0. Based on sample size guidelines in Shirazi et al. (3), using 50 sites provide only 70 percent confidence that the C-factor lies within the 10 percent of the true factor; that is, the C-factor derived for different regions might still provide a biased or erroneous estimate. However, the effect of the sample size bias on the observed trend is rather negligible considering the huge difference in C-factors derived, especially since the same trend repeated for all five years analyzed. In fact, to minimize the effect of the sample size bias, Table 2 also includes the five years average of the calibration factors. This table shows that the five year average C-factors for different regions are also significantly different, so the region-specific C-factors are advised.

Figure 2. Region Specific C-Factors for Michigan Four-Legged Signalized Intersections in 2012.
In summary, based on these observations, it is justified to find and use region-specific C-factors instead of a state-wide C-factor for large states, such as Texas or Michigan. However, important questions should first be asked before segmenting a state or large region into sub-regions. When are the region-factors required? Can we get any intuition about its requirement before developing the region-specific factor? For instance, the analyst may want to know if the sample should be collected in different regions or at the statewide level. The next section addresses this issue by providing recommendations to determine the need for region-specific c-factors. The recommendations are based on the general characteristics of the data at hand at the network level and can be used to inform the agency (or the analyst) whether or not having the region-specific C-factor is required.

3. RECOMMENDATIONS

This section provides a procedure to determine the need for having region-specific C-factors based on the general data in the network. This section is divided into two parts. The first part describes the characteristics of the proposed method. The second part covers the procedure to use the recommendations for segment and intersection models.

3.1 Methodology

It is hypothesized that different regions may experience different number of crashes that can be attributed to two factors: (1) experience different traffic flows and/or having different lengths of facilities (or number of locations), or (2) difference in crash risk (or difference in safety). Then, it is argued that if the difference in the number of crashes in different regions is pertained to comparable difference in traffic flow or length of facility (or number of locations), the analyst can use the statewide calibration factor. On the other hand, if the difference in number of crashes in different regions is attributed to a particular factor that induces a safety issue (or safety improvement) ceteris paribus, a region-specific is recommended.

It is then important to estimate the number of crashes caused in the state and the region as a function of traffic flow and segment length. Fortunately, base models [Note: the flow only models with all other variables at the base conditions are referred to as base models] documented in HSM can assist for this matter. A typical HSM base model to estimate the number of crashes on roadway segments is shown in Eq. 2:

\[ N_{\text{base},i} = e^{b_0 + b_1 \times \ln(F_i) + \ln(L_i)} \]

where \( F_i \) denote the traffic flow (AADT or ADT) and \( L_i \) the segment length at the site “i”. The parameters \( b_0 \) and \( b_1 \) denote the intercept and the traffic flow coefficient, respectively. To simplify the procedure and since we seek an approximate procedure to estimate the number of crashes, the value of the traffic flow at each individual site is replaced by the mean value of the traffic flow in the state or the region (\( \bar{F} \)). Hence Eq. 2 is revised as.

\[ \sum_i N_{\text{base},i} \approx \sum_i e^{b_0 + b_1 \times \ln(\bar{F}) + \ln(L_i)} \]
Eq. 3 can be transformed as:

$$\sum_i N_{\text{base},i} = e^{b_0 + b_1 \times \ln(F)} \times \sum_i L_i$$

(4)

In other words,

$$\sum_i N_{\text{base},i} = e^{b_0 + b_1 \times \ln(F)} \times L^T$$

(5)

where $L^T$ denote the total segment length in the state or the region.

Next, the parameter $\tilde{C}$ (which is referred to as the C-proxy in this study) is calculated for the region and the state (the notations ‘R’ and ‘S’ are used to denote the values collected or estimated for the region and the state respectively) as follows:

$$\tilde{C}_R = \left( \frac{N_{\text{obs}}^T}{e^{b_0 + b_1 \times \ln(F)} \times L^T} \right)_R$$

(6)

$$\tilde{C}_S = \left( \frac{N_{\text{obs}}^T}{e^{b_0 + b_1 \times \ln(F)} \times L^T} \right)_S$$

(7)

where,

$\tilde{C}$ = C-proxy;

$N_{\text{obs}}^T$ = the total number of crashes on the network;

$F$ = the mean value of the traffic flow (AADT or ADT); and,

$L^T$ = the total combined length of all sites.

Note that only three sources of information, all at the network level, are required to calculate the variables $\tilde{C}_R$ and $\tilde{C}_S$. These sources are (1) the total number of crashes ($N_{\text{obs}}^T$), (2) the mean value of the traffic flow ($F$), and (3) the total segment length ($L^T$).

Now, if the relative difference between the statewide C-proxy ($\tilde{C}_S$) and a particular region’s C-proxy ($\tilde{C}_R$) is less than a certain threshold, the analyst can ignore developing a region-specific factor and can use the statewide C-factor for the region as well; however, if they are significantly different, the analyst is recommended to develop a C-factor that is specific to that region [Note: similar recommendations can be used to determine when the SPF model should be recalibrated. For further information, the interested readers are referred to the reference (8) and (9)]. In this study, we set the threshold to 10 percent. However, it should be pointed out that the procedure works for any threshold. The threshold is in fact a function of or a balance between the
3.2 Instructions to Use the Procedure

This section documents a procedure for determining whether or not a region-specific calibration factor is required for a specific region. First, the analyst should compute the value of the C-proxy in the region and the state ($\tilde{C}_R$ & $\tilde{C}_S$). Next, the value of the $\tilde{C}_R$ is compared with the value of the $\tilde{C}_S$. If the relative difference ($\bar{e}_R$) is more than a certain threshold, the agency is recommended to calculate and use the region-specific C-factor. Otherwise, the statewide C-factor can be used. It is recommended to set the threshold to a value between 10 to 20 percent (we used a 10 percent threshold in this study). The proposed procedure can be used for both segment and intersection models.

**Segment Models**

The detailed steps to compute the $\tilde{C}_R$ and $\tilde{C}_S$, and determine if a region specific C-factor is needed for segment roadway models are as follows:

**Step 1.** Find the total number of crashes and the total segment length in the state ($N_{S,obs}^T$ and $L_S^T$) and in the region ($N_{R,obs}^T$ and $L_R^T$).

**Step 2.** Find the mean value of the traffic flow (i.e., ADT or AADT) in the state ($F_S$) and in the region ($F_R$).

**Step 3.** Take the base SPF model from the HSM. Let $b_0$ and $b_1$, respectively, denote the intercept and the coefficient of traffic flow. Find the variables $N_{S,pre}$ and $N_{R,pre}$ as follows:

\[
N_{S,pre} = e^{b_0 + b_1 \times \ln(F_S)} \quad (8)
\]

\[
N_{R,pre} = e^{b_0 + b_1 \times \ln(F_R)} \quad (9)
\]

**Step 4.** Find the parameter $\tilde{C}$ in the state ($\tilde{C}_S$) and in the region ($\tilde{C}_R$) as follows:

\[
\tilde{C}_S = \frac{N_{S,obs}^T}{N_{S,pre} \times L_S^T} \quad (10)
\]

\[
\tilde{C}_R = \frac{N_{R,obs}^T}{N_{R,pre} \times L_R^T} \quad (11)
\]

**Step 5.** Find $\bar{e}_R$ as follows:

\[
\bar{e}_R = \frac{|\tilde{C}_R - \tilde{C}_S|}{\tilde{C}_S} \times 100 \quad (12)
\]
Step 6. If $\tilde{e}_R \geq 10\%$, estimate and use a region-specific C-factor. Otherwise, use the statewide C-factor.

Intersection Models

A similar procedure with some modifications can be used for intersection models. The detailed steps to calculate the $\tilde{C}_R$ and $\tilde{C}_S$, and how to use the procedure to determine if a region-specific C-factor is needed for intersection models are as follows:

Step 1. Find the total number of crashes and the total number of intersections in the state ($N_{S,obs}^T$ and $N_S$) and in the region ($N_{R,obs}^T$ and $N_R$).

Step 2. Find the average traffic flow on the state major ($\bar{F}_{S,major}$) and minor ($\bar{F}_{S,minor}$) streets and on the region major ($\bar{F}_{R,major}$) and minor ($\bar{F}_{R,minor}$) streets.

Step 3. Take the base SPF model from the HSM. Let $b_0$, $b_1$, and $b_2$, respectively, denote the intercept, the coefficient of traffic flow on the major street, and the coefficient of the traffic flow on the minor street. Find the variables $\bar{N}_{S,pre}$ and $\bar{N}_{R,pre}$ as:

$$\bar{N}_{S,pre} = e^{b_0+b_1 \times \ln(\bar{F}_{S,major})+b_2 \times \ln(\bar{F}_{S,minor})}$$

(13)

$$\bar{N}_{R,pre} = e^{b_0+b_1 \times \ln(\bar{F}_{R,major})+b_2 \times \ln(\bar{F}_{R,minor})}$$

(14)

Step 4. Find the parameter $\tilde{C}$ for the state ($\tilde{C}_S$) and the region ($\tilde{C}_R$) using Eq. 10 and Eq. 11, respectively. Note: In Eq. 10 and Eq. 11, the variables $L_S^T$ and $L_R^T$ should be replaced by the number of intersections in the state ($N_S$) and the region ($N_R$), respectively.

Step 5. Find $\tilde{e}_R$ using Eq. 12.

Step 6. If $\tilde{e}_R \geq 10\%$, estimate and use a region-specific C-factor. Otherwise, use the statewide C-factor.

4. VALIDATING THE PROCEDURE WITH OBSERVED DATASETS

In this section, the method is validated using two observed datasets. First, the recommendations are examined with Texas urban four-lane divided arterials using data collected in three-year frequency from 2012 to 2014. Next, the recommendations are evaluated with Michigan four-legged signalized intersections dataset collected in 2008. The proposed methodology is validated by calculating C-factors in the region and the state. Recall that our proposed methodology needs only three sources of information at the network level (as shown in Eq. 6 and Eq. 7). However, to estimate the C-factors, we need the detailed data (crash, exposure and characteristics) at individual sites as shown in Eq.1.
It is worth noting that the proposed procedure can be used for any boundaries. However, since the calibration procedure is not only needed to account for differences in topographical or climate changes, but also any other unobserved heterogeneity, it is advised to use the administration boundaries to divide the state. The administration regions can be useful in defining areas with more homogenous characteristics and consistent variable collection procedures. In addition, data within the administration regions could be easier to collect.

4.1 Texas Urban Four-Lane Divided Arterials

Table 3 shows the application of the proposed methodology to predict the need for having region-specific C-factor to the Texas urban four-lane arterials dataset. The data collected in 2014 are used for this analysis. For the purpose of analysis the threshold was set to 10 percent. As indicated in the table, for the North region, the difference between $\hat{C}_{\text{North}}$ and $\hat{C}_s$ is about 15 percent, so a region-specific calibration factor for the North region is recommended. As shown in the table, the difference between the North region and statewide C-factors is about 17 percent, which validates the decision based on the proposed methodology. The method, however, does not recommend considering a region-specific C-factor for the South region ($\hat{C}_{\text{South}}$=2.9 percent). The difference between South region and Statewide C-factors is also negligible ($\hat{C}_{\text{South}}$= 2.4 percent). Hence, the methodology is verified here as well. Next, the region-specific C-factor is also recommended for the East region ($\hat{C}_{\text{East}}$=23.6 percent). Here too the actual difference between C-factors is significant ($\hat{C}_{\text{East}}$= 26.1 percent), which confirms the decision based on the guidelines. For the West region, the $\hat{C}_{\text{West}}$ is equal to 13.8 percent, which is greater than 10 percent, so the region-specific factor is recommended. Again, the difference between estimated C-factors is also significant which validate the methodology.

<table>
<thead>
<tr>
<th>Region</th>
<th>North</th>
<th>South</th>
<th>East</th>
<th>West</th>
<th>Statewide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Crashes</td>
<td>3214</td>
<td>2851</td>
<td>2902</td>
<td>1105</td>
<td>10072</td>
</tr>
<tr>
<td>Average Traffic Flow</td>
<td>18599.3</td>
<td>18757.7</td>
<td>18198.7</td>
<td>11025.7</td>
<td>17346.1</td>
</tr>
<tr>
<td>$\hat{N}_{\text{pre}}$</td>
<td>8.40</td>
<td>8.50</td>
<td>8.16</td>
<td>4.13</td>
<td>7.64</td>
</tr>
<tr>
<td>Total Length (Mile)</td>
<td>376.7</td>
<td>289.8</td>
<td>241.3</td>
<td>197.4</td>
<td>1105.2</td>
</tr>
<tr>
<td>C-proxy ($\hat{C}$)</td>
<td>1.015</td>
<td>1.157</td>
<td>1.474</td>
<td>1.356</td>
<td>1.192</td>
</tr>
<tr>
<td>Difference in $\hat{C}$ (%)</td>
<td>14.9†</td>
<td>2.9</td>
<td>23.6</td>
<td>13.8</td>
<td>-</td>
</tr>
<tr>
<td>C-factor ($C$)</td>
<td>0.928</td>
<td>1.089</td>
<td>1.407</td>
<td>1.247</td>
<td>1.116</td>
</tr>
<tr>
<td>Difference in C (%)</td>
<td>16.9</td>
<td>2.4</td>
<td>26.1</td>
<td>11.7</td>
<td>-</td>
</tr>
</tbody>
</table>

† Underlined values: $\hat{C} \geq 10\%$.

4.2 Michigan four-Legged Signalized Intersections

Table 4 indicates the results for application of the region-specific guidelines for Michigan four-legged signalized intersections dataset. The data collected in 2008 are used for this analysis. As
shown in this table, the region-specific factor was recommended for all regions, except for region 6. For all these regions, the difference between the region and statewide C-factors is also very large which validates the proposed methodology. For region 6, the region-specific factor is not recommended. The estimated C-factor for this region also affirms the decision based on our proposed method.

Table 4. Region Specific Guidelines - Michigan four-Legged Signalized Intersections

<table>
<thead>
<tr>
<th>Region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Statewide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Crashes</td>
<td>250</td>
<td>372</td>
<td>454</td>
<td>344</td>
<td>483</td>
<td>374</td>
<td>578</td>
<td>2855</td>
</tr>
<tr>
<td>Major Avg. Traffic Flow</td>
<td>12324.4</td>
<td>17065.6</td>
<td>24908.9</td>
<td>16454.5</td>
<td>17575.9</td>
<td>18082.4</td>
<td>18082.4</td>
<td>39281.0</td>
</tr>
<tr>
<td>Minor Avg. Traffic Flow</td>
<td>4617.9</td>
<td>7865.4</td>
<td>10677.5</td>
<td>8099.9</td>
<td>7890.3</td>
<td>8305.3</td>
<td>13677.0</td>
<td>8781.2</td>
</tr>
<tr>
<td>N_{pre}</td>
<td>3.02</td>
<td>4.28</td>
<td>7.61</td>
<td>4.65</td>
<td>4.95</td>
<td>5.16</td>
<td>13.06</td>
<td>6.08</td>
</tr>
<tr>
<td>No. of Intersections</td>
<td>46</td>
<td>50</td>
<td>51</td>
<td>50</td>
<td>52</td>
<td>50</td>
<td>50</td>
<td>349</td>
</tr>
<tr>
<td>C-proxy (C)</td>
<td>1.80</td>
<td>1.74</td>
<td>1.17</td>
<td>1.48</td>
<td>1.88</td>
<td>1.45</td>
<td>0.89</td>
<td>1.35</td>
</tr>
<tr>
<td>Difference in C (%)</td>
<td>33.9†</td>
<td>29.2</td>
<td>13.2</td>
<td>10.0</td>
<td>39.4</td>
<td>7.71</td>
<td>34.3</td>
<td></td>
</tr>
<tr>
<td>C-factor (C)</td>
<td>2.62</td>
<td>2.44</td>
<td>1.66</td>
<td>2.09</td>
<td>2.63</td>
<td>1.99</td>
<td>1.14</td>
<td>1.83</td>
</tr>
<tr>
<td>Difference in C (%)</td>
<td>43.2</td>
<td>33.3</td>
<td>9.3</td>
<td>14.2</td>
<td>43.7</td>
<td>8.74</td>
<td>37.7</td>
<td>-</td>
</tr>
</tbody>
</table>

† Underlined values: \(\triangle > 10\%\).

As discussed above, we used a threshold of 10 percent for our analysis. In cases when the safety analyst would prefer to have a more accurate C-factor, a lower threshold could be used. However, the model would then need to be calibrated for almost all regions. On the other hand, by choosing a higher threshold a less accurate C-factor will be used, but the analyst will place less effort on calibrating models for the state or a large region. In the end, the analyst needs to balance the recalibration process between accuracy and the effort or resources needed for recalibrating the models.

5. SUMMARY AND CONCLUSIONS

The HSM recommends finding separate C-factors for large jurisdictions that are characterized by different topographical or weather conditions. Unfortunately, the HSM does not provide guidelines for determining the detailed conditions when separate C-factors are warranted or justified. Without guidelines, it is not possible to know whether the region-specific calibration factors are needed or not. In this research, first, the region-specific calibration factors were developed for two facility types, one in Texas and the other one in Michigan. The results suggested that the region-specific C-factors are significantly different from one region to the other in both states. Then, this study documented a data-driven procedure to determine when having region-specific C-factors are recommended. The proposed procedure was successfully validated by applying them to several observed datasets in Texas and Michigan. The procedure is straightforward and can be used to determine the need for region-specific C-factors, irrespective of the type of facility, such as intersections and roadway segments. The only sources of information needed for segment models are: (1) the total number of crashes, (2) the mean value of traffic flow (ADT or AADT) and (3) the total segment length. For intersection models, the
only needed information are (1) the total number of crash data, (2) the mean value of traffic flow on major and minor streets and (3) the number of intersections. The agency is required to secure these values at the network level in each region and the state and follow the procedure that were documented in the study to determine if a region-specific C-factor should be derived for the facility.

Although the study recommendations are applicable to any threshold, further research is needed to determine the appropriate or statistically-oriented threshold for each facility being analyzed. In this case, the errors involved in both the base model and CMFs can be used to decide on the accuracy needed for the C-factor (2, 7). In addition, if the SPF includes the variables that account for different topographical or weather conditions, the region-specific calibration may not be needed. Future work should include explore the need of region-specific C-factors when the SPFs include such variables or when the SPF is developed using methods that specifically consider unobserved heterogeneity in data (10, 11, 12). Further research is also needed to investigate the need for having region-specific C-factors when calibrating the severity distribution function (SDF) models (13).

ACKNOWLEDGMENTS

The authors would like to thank the ATLAS Center for funding this project. The views expressed by the authors of this paper do not necessarily reflect the views of the Center.

REFERENCES


