Renewal and Update of MTO IDF Curves: Defining the Uncertainty

Eric D. Soulis, Daniel Princz and John Wong

University of Waterloo, Waterloo, Ontario.

Received 2014 05 14, accepted 2015 01 29, published 2015 02 20.

Abstract

The University of Waterloo has been commissioned by the Design and Contract Standards Office of the Ministry of Transportation of Ontario (MTO) to update the intensity–duration–frequency (IDF) curves that are used to estimate design storms for drainage infrastructure. Environment Canada provides analyzed rainfall data for weather stations in Ontario. Project engineers determine the design storm parameters using a variety of methods. Thus, there is a need for a consistent interpolation procedure that can deal with the sparseness and unevenness of the Ontario station network. This paper describes the procedures that form the basis for the MTO IDF Curve Lookup system. Confidence limits assist in the design process and allow comparison with other results. These were used to assess the spatial bias in the error fields. There appears to be none, but there is room for improvement where there are sharp changes in geography, such as the highlands west of Thunder Bay. Validation curves are presented for several Canadian stations and one American station.

A single set of ten regression equations is sufficient for the whole province, eliminating the need for regional IDF curves and simplifying the design process.

Keywords: IDF curve, extreme rainfall, interpolate IDF, IDF lookup, MSC dataset.

1 Introduction

Good estimates of peak rainfall intensity are essential when designing highway drainage infrastructure such as culverts, bridges, sewer systems and roadside ditches. Quality rainfall data enables designers to make calculations that meet drainage capacity design standards and avoid the over- or under-design of drainage elements. Both can be expensive: over-design may waste resources and under-design can result in additional maintenance or repair costs.

Intensity–duration–frequency (IDF) curves are used to characterize the statistics of extreme rainfall for the design of flood protection infrastructure on small watersheds. Environment Canada (EC) periodically publishes curve parameters, in the form of $AB$ coefficients, at selected meteorological station locations in Ontario and elsewhere.

Engineers are faced with the problem that data are only available for certain spots in the province. Data for other locations must be interpolated for every project. The Ontario Ministry of Transportation (MTO) IDF Curve Lookup tool uses the latest EC data available from 147 Meteorological Service of Canada (MSC) stations across Ontario to determine rainfall intensities for any location in the province. The goal of the tool is to provide a consistent and objective method.

The U. S. Geological Survey (USGS) GTOPO30 digital elevation dataset (USGS 1997) is used to derive physiographic characteristics that become independent variables in a stepwise multiple linear regression analysis with MSC station statistics. The regression produces a set of generating equations for the parameters used to produce IDF curves. The technique is modified for this study and weights station data by length of record to ensure that more reliable data have greater influence on the interpolation.

MTO began revising its IDF curves in the late 1990s due to administrative changes, the availability of new data sets from MSC, and the concern for the impact that climate change could have on infrastructure design. Historically, MTO provided IDF curves for each MTO district in a hard copy document developed with EC data (Hogg, 1989). Ontario has the problem of varying gauge density throughout its geography. It has a densely gauged south, and a sparsely gauged north. The IDF Curve Lookup Tool uses the regional and local topography of the area of interest to interpolate the climate parameters that are relevant for a design.
The variables of interest in this study are the statistics that define the precipitation frequency.

1.1 IDF Curves

IDF curves summarize the annual probability of exceedance of storm intensity given a storm duration, \( P(R, \tau) \) of a volume of rainfall \( R \) in a single event of duration \( \tau \) (h) and intensity \( RR \) (mm/h). It is common practice to express \( (R, \tau) \) by return period \( T \):

\[
T = \frac{1}{P(R, \tau)} \tag{1}
\]

In this document, rainfall intensity and volume are expressed by:

\[
RR = At^B \tag{2}
\]

where:

\[
RR = \frac{R}{\tau} \tag{3}
\]

\( A, B = AB \) coefficients that vary with location and return period,

\( RR = \) rainfall rate in mm/h,

\( R = \) rainfall volume in mm,

\( \tau = \) event duration in hours, and

\( T = \) return period in years.

1.2 MSC Data Set

The National Climate Data Information Archive of Environment Canada (MSC 2012) provided the rainfall dataset used in this study. The values for the intensity of rainfall, \( R \), and the \( AB \) parameters used in the modeling of IDF curves for the 147 Ontario stations are for return periods of 2 y, 5 y, 10 y, 25 y, 50 y and 100 y. One of the 147 stations was removed from the analysis as an outlier. Figure 1 shows the station locations. Record lengths varied from 9 y to 76 y with an average record length of 28 y.

1.3 Sample station data analysis

EC publishes \( AB \) coefficients with the data for the stations. EC uses the classical Gumbel extreme value distribution method, and finds the relevant quantiles by least squares fit to the histogram. The \( AB \) coefficients are found by method of moments for each return period of interest (Hogg et al. 1989).

Figure 2 shows a typical curve on a log–log plot. \( \log A \) determines the 1 h intercept and \( B \) determines the slope. Figure 3 shows sample station data with the standard error line highlighted. The standard deviation of the mean for this sample station ranges from 7.9 mm for the 2 y return period to 19.3 mm for the 100 y return period.

![Figure 1](image1.png)

Figure 1  Map of Ontario showing all weather stations used in this study.

![Figure 2](image2.png)

Figure 2  Sample IDF curve (MSC 2012).

![Figure 3](image3.png)

Figure 3  Sample station statistics (MSC 2012), standard error line highlighted.
1.4 Spatial Interpolation

The IDF Curve Lookup tool uses the Waterloo Multiple Physiographic Parameter Regression (WATMAPPR) model (Segleniêks 2009), which is based on the square grid technique (Solomon et al. 1968), to estimate the curve parameters. The square grid technique uses UTM 10 km grid squares as elementary sub-catchments. The premise of the tool is that local and regional topography strongly influenced local climate. Thus, topographic parameters are useful interpolators of surface fields of interest, such as temperature, runoff and, in this case, IDF curve AB parameters.

The A coefficients were determined using stepwise-linear regression. The independent variables were station latitude and longitude, station elevation, barrier height, local slope of X and Y, and the station’s distance to water. The dependent variables are log transformed Gumbel quantiles. A coefficients depend on both return period and station location. B typically does not show a dependence on station properties and is adequately represented by a single number. This is equivalent to having parallel IDF curves as shown in Figure 2 above.

The interpolation method uses a large domain digital elevation model (DEM) to derive a comprehensive set of secondary physiographic characteristics. The characteristics used for each grid square are east–west slope, north–south slope, distance to major water bodies, and barrier height from the west. These are combined with basic station characteristics—latitude, longitude and elevation—to generate a powerful set of independent variables that have been used in a number of climatological studies.

In the stepwise regression, each independent variable was put into the regression individually, and the resulting value of the objective function was determined. The independent variable that resulted in the lowest value of the objective function was the first variable to enter the regression. The process was repeated with the first variable and each of the remaining variables entered individually into the regression. The variable that gave the lowest value of the objective function was the second variable to enter the regression. Other variables were added to the regression until the resulting objective function no longer decreased by a certain threshold amount.

The results of this process were equations for each return period which related the value of the maximum precipitation to the physiographic parameters:

\[ \hat{Y} = X\beta + \varepsilon \]  

(4)

where:

- \( \hat{Y} \) = the vector of observations (A or B),
- \( X \) = the matrix of independent variables,
- \( \beta \) = the vector of parameters to be estimated, and
- \( \varepsilon \) = a vector of errors.

These equations were used to calculate the expected value of the maximum precipitation at all points in the study area based on the values of the physiographic parameters at those points (see Table 1). The WATMAPPR technique is a non-exact interpolation, since the interpolated value at a measurement station may not necessarily be the measured value at that station.

Table 1  Typical partial regression coefficients for determining A parameters, standard error highlighted.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>RMS Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.4200</td>
<td>0.1145</td>
<td>3.5506</td>
<td>0.0005</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>-6.1635</td>
<td>0.7841</td>
<td>-8.6066</td>
<td>1.01E-12</td>
</tr>
<tr>
<td>X Variable 2</td>
<td>1.7353</td>
<td>0.4887</td>
<td>3.5306</td>
<td>0.0005</td>
</tr>
<tr>
<td>X Variable 3</td>
<td>7.25E-05</td>
<td>4.14E-05</td>
<td>1.7501</td>
<td>0.0823</td>
</tr>
<tr>
<td>X Variable 4</td>
<td>0.0020</td>
<td>0.0034</td>
<td>0.5866</td>
<td>0.5585</td>
</tr>
<tr>
<td>X Variable 5</td>
<td>-0.0003</td>
<td>0.0002</td>
<td>-1.3035</td>
<td>0.1946</td>
</tr>
<tr>
<td>X Variable 6</td>
<td>0.0010</td>
<td>0.0020</td>
<td>-0.5027</td>
<td>0.6160</td>
</tr>
<tr>
<td>X Variable 7</td>
<td>0.0570</td>
<td>0.0353</td>
<td>1.6141</td>
<td>0.1088</td>
</tr>
<tr>
<td>X Variable 8</td>
<td>0.4599</td>
<td>0.0169</td>
<td>27.2628</td>
<td>3.38E-57</td>
</tr>
</tbody>
</table>

The Cartesian standard error, \( E \), is found using

\[ E = A \cdot (10^R - 1) \]  

(5)

The objective function for the regression is the expected value of \( E \). The multiplication of each observation by its number of years of record recovers the total square error for each station dataset. The objective function of the regression therefore includes spatial errors, temporal errors, and model errors for the entire dataset. The results for this are under the column RMS for A Weighted Regression of Table 2.

Table 2  MSC station standard errors.

<table>
<thead>
<tr>
<th>Return Period</th>
<th>MSC Mean Station Value</th>
<th>MSC Standard Deviation</th>
<th>Weighted Station Mean</th>
<th>RMS for A Weighted Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 y</td>
<td>20.20</td>
<td>2.32</td>
<td>20.35</td>
<td>0.70</td>
</tr>
<tr>
<td>5 y</td>
<td>27.34</td>
<td>3.36</td>
<td>27.49</td>
<td>1.43</td>
</tr>
<tr>
<td>10 y</td>
<td>32.04</td>
<td>4.16</td>
<td>32.20</td>
<td>2.27</td>
</tr>
<tr>
<td>25 y</td>
<td>37.90</td>
<td>5.30</td>
<td>38.08</td>
<td>3.64</td>
</tr>
<tr>
<td>50 y</td>
<td>42.38</td>
<td>6.06</td>
<td>42.56</td>
<td>5.00</td>
</tr>
<tr>
<td>100 y</td>
<td>46.75</td>
<td>6.89</td>
<td>46.93</td>
<td>6.60</td>
</tr>
</tbody>
</table>

\( R \) was used indirectly in the development of the MTO confidence intervals. \( R \), an aggregate of the mean intensity across durations, has a standard deviation as large as the value itself. Used directly, \( R \) proved little help for regionalization. Instead of \( R \), the AB parameters were used directly, and each return period was analyzed separately. While this process removes most of the uncertainty from \( R \), some of it survives, and is carried forward in the standard error of each IDF curve.
The standard error is used to calculate confidence limits in Equation 6:

\[ \hat{Y}_i \pm t_{0.975, se} \sqrt{X_i^T C^{-1} X_i} \]  \hspace{1cm} (6)

where:

- \( \hat{Y}_i \) = the expected value of rainfall at point \( i \),
- \( t_{0.975} \) = the student \( t \)-distribution right tail probability,
- \( se \) = the standard error of the regression,
- \( X_i \) = a \( 1 \times 7 \) column vector that contains the difference between each independent variable and its mean for the station \( i \), and
- \( C \) = the covariance matrix (Draper and Smith 1966).

The square root term describes the curvature in the confidence limits.

In the development of confidence limits, \( A \) and \( B \) are treated as stationary in time and space. The confidence limits, therefore, reflect both temporal and spatial uncertainty. While Figure 4 and Figure 5 show that spatial trends do not exist in the residuals, there are likely time trends. The 95% confidence limits allow for these potential time trends. Their potential identification and quantification is part of the next phase of research. The investigation may allow the further reduction of the standard error.

2 Results

2.1 Examining the Residuals

The residual shown in Figures 4 and 5 overleaf indicates station \( R \)-positive errors (white dots) and station \( R \)-negative (black dots). Those stations that are between these extremes are indicated with varying degrees of shading. There is no trend in the vicinity of major topographical elements such as the Niagara Escarpment or the area west of Thunder Bay.

2.2 Comparison to Station Data

There are no identifiable spatial trends in the standard error, even in unusual geographic areas. The errors occur in specific geographic areas including the northern shore of Lake Superior and the area below the escarpment on Lake Ontario. These are well-known climate anomalies in the province.

Figures 6 and 7 overleaf show the upper 95% confidence limits and the expected values for a 100 y 1 h storm in the Ontario dataset. In both of the figures, the reddest colors represent the lowest value of the plus or minus term, and the bluest colors represent the highest value of the plus or minus term.

2.3 Comparison with Canadian Data

Typical results from selected MSC weather stations are in Figures 8 through 13 below. The points are the MTO database interpolation for the station. For the comparison of station data for the model validation, the original MSC data was excluded. When the data was excluded, the regression equation changed little due to the individual length of station records within the overall size of the dataset.

The A coefficient for the 1 h duration curve is presented, for ease of interpretation, in mm/h. This is possible because, in this case, where \( \tau \) = 1 h, \( R R = A \), as seen in Equation 2. The figures above illustrate how, at station locations, the MTO data is comparable to the MSC weather stations. The Kenora values show the least similarity. This is due to the orographic effect between Lake Winnipeg and Kenora. Further research will address this issue. In the meantime, the extent of the confidence limits compensates for this. The MTO dataset, with 4073 station-years, has the advantage of having values based on the entire MSC dataset and consistency between stations.

2.4 Comparison with NOAA Data

The U.S. National Oceanic and Atmospheric Administration (NOAA) uses a three parameter \( ABC \) IDF model that plots as a curve in a log–log space rather than a straight line (Perica et al. 2013). Data is converted so it could be compared with the Canadian \( AB \) data. Figures 14 to 17 below show an example of an informal comparison between the two datasets. Confidence levels are at 95% for Canadian curves, and 90% for the American curves.

The NOAA curve is determined by fitting the three-parameter \( ABC \) model to the data. The Canadian model uses the two-parameter \( AB \) model. There are advantages and disadvantages to each but in a data-sparse environment, such as Canada, having one less parameter to estimate is often helpful. The linear best fit to the American data points is also shown and it matches the Canadian \( AB \) curve. Based on preliminary studies, with visual comparisons, the difference between the two curves is insignificant at a 95% confidence level for individual observations.

3 Results and Implementation

The new IDF curves are project location based. The limits of a highway project are identified on the user interface by two representative locations. These locations are defined either by entering their latitude and longitude, or by selecting the locations from the Google Maps interface.

The system determines the IDF curve parameters representative of these areas covered by a project. The display shows the percentage error at the two identified points by assigning a representative colour to the icons. If the percentage error is greater than the acceptable value, the highway section can be divided into two sections by adding a third intermediate point. This will result in dividing the highway project into two areas each with a different IDF curve. The produced IDF curves have similar uncertainty to the curves provided at individual MSC stations. The similarity exists because the introduced spatial error offsets the eliminated temporal error. The obvious advantage is that, with the site-specific temporal error removed, the MTO model can generate IDF curves for anywhere in Ontario, not just at MSC stations. With the interpolated MTO curves, designers have calibrated rainfall extremes for projects, such as highways, that are too far from MSC stations for the MSC curves to be reliable.
Figure 4  The locations of positive and negative residuals in Ontario.

Figure 5  Detail of Figure 4 showing the locations of positive and negative residuals in Southern Ontario and the Toronto–Hamilton area.

Figure 6  Upper 95% confidence limit for 100 y 1 h storm.

Figure 7  Expected values for 100 y 1 h storm.
Figure 8  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for Chatham WPCP (6131415) and MTO values with 95% confidence limits.

Figure 9  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for Toronto City (6158355) and MTO values with 95% confidence limits.

Figure 10  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for Ottawa CDA RCS (6105978) and MTO values with 95% confidence limits.

Figure 11  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for North Bay A (6085700) and MTO values with 95% confidence limits.

Figure 12  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for Moosonee RCS (6075435) and MTO values with 95% confidence limits.

Figure 13  Comparison of MSC $A$ coefficient for 2 y to 100 y 1 h storm for Kenora A (6034075) and MTO values with 95% confidence limits.
4 Conclusions

The MTO IDF Look-Up Tool is a reliable interpolator of IDF curves in Ontario, and provides a consistent estimate of $AB$ parameters. The key to its successful interpolation of extreme rainfall statistics is the integration of physiographic characteristics. One equation for each return period represents the entire province, with only one value for $B$. Despite this, the errors show no regional bias, and the standard error is 10%, which is the standard error of the individual stations. Time trends are a potential source of error, and are the focus of further study. In the meantime, the confidence limits are broad enough to accommodate any error that time trends introduce. The interpolator was validated with the use of select NOAA and MSC stations. The MTO model results match the data and RMS error from these stations.

Future work will include the addition of data from other sources from within Ontario, along with neighbouring states and provinces. It will also investigate time trends in the data for a more refined interpolation of extreme rainfall estimates.

With the MTO Look-Up Tool, engineers have a consistent set of interpolated IDF curves for the entire province. In urban areas, where existing MSC IDF curves exist, the MTO tool provides results that match the RMS error for the MSC stations. In rural setting, the tool is of particular use. There, it provides reliable IDF curves where curves do not otherwise exist.

The interface tool is available at http://www.mto.gov.on.ca/IDF_Curves/.

References


