Water Resources Research Report

Guidelines for Flood Frequency Estimation under Climate Change

By:

Samiran Das
and
Slobodan P. Simonovic

Report No: 082
Date: October 2012

ISSN: (print) 1913-3200; (online) 1913-3219;
ISBN: (print) 978-0-7714-2973-6; (online) 978-0-7714-2974-3;
Guidelines for Flood Frequency Estimation under Climate Change

Samiran Das*

Slobodan P. Simonovic

Department of Civil and Environmental Engineering

The University of Western Ontario

London, Ontario, Canada

October 2012

*Current affiliation: Asian University for Women, 20/G M.M. Ali Road, Chittagong-4000, Bangladesh.

E-mail: samirandas@gmail.com (S. Das), simonovic@uwo.ca (S. P. Simonovic)
Abstract

The assessment of climate change impacts on frequency of floods is important for management of flood disasters. It is recognized that methods for the assessment are subject to various sources of uncertainty (choice of climate model and emission scenario, course spatial and temporal scales, etc.). This report presents the guidelines and associated procedures for assessing climate change impact on the frequency of flood flows. These guidelines are prepared for the implementation of the methodology developed by Das and Simonovic (2012).
# Table of Contents

Abstract........................................................................................................................................... ii

Table of Contents ............................................................................................................................... iii

List of Tables ....................................................................................................................................... iv

List of Figures ...................................................................................................................................... iv

1 Introduction ..................................................................................................................................... 1

2 Preparing Datasets for Frequency Analysis .................................................................................. 3

3 Frequency Analysis ....................................................................................................................... 6

4 Uncertainty Assessment ................................................................................................................. 7

5 Analyses and Results of a Case Study ............................................................................................ 8

6 Conclusion ...................................................................................................................................... 13

Acknowledgements .......................................................................................................................... 14

References ......................................................................................................................................... 14

Appendix A: Study Area and Data ..................................................................................................... 19

Appendix B: Climate Model and Scenarios ...................................................................................... 22

Appendix C: Weather Generator ...................................................................................................... 24

Appendix D: Hydrological Model ...................................................................................................... 25

Appendix E: Selection of Peaks ........................................................................................................ 29

Appendix F: Common Statistical Distributions and Parameter Estimation Procedures ................. 30

Appendix G: Previous Reports in the Series .................................................................................... 35
List of Tables

Table A.1 Location of stations in the Upper Thames River basin .............................................. 21
Table A.2 List of AOGCM models and emission scenarios used .................................................. 22
Table A.3 Calculation of L-moments and their dimensionless ratios, modified after (Das, 2010) ............................................................ 30

List of Figures

Figure 5.1 Simulated flood frequency results for climate data. Each plot relates to different time horizons. Data from 375 scenarios (15 climate model scenarios x 25 model runs each) are used for future climate. Each line with a specific color represents a different AOGCM. The upper and lower bound frequency curves are indicated in the plots. The flood frequency curve derived from historic data is also shown for information only. Data from 25 runs are used to produce a range of results for baseline. ........................................................................................................... 11

Figure 5.2 Probability density functions (PDFs) of 100-year return period flood for all AOGCMs at the time horizon 2050. Each PDF is constructed using 25 model runs. ............................. 11

Figure 5.3 Probability density plots for flood magnitudes at return period, T = 10, 100 and 250 at the four time horizons ........................................................................................................... 12

Figure 5.4 Cumulative distribution functions for flood magnitudes at return period, T = 10, 100 at the four time horizons ........................................................................................................... 12

Figure A.1 Map of the Upper Thames River basin ......................................................................... 20

Figure A.2 Continuous HEC-HMS hydrologic model structure ..................................................... 26

Figure A.3 Soil moisture accounting losses module ........................................................................ 27
1 Introduction

The estimation of flood for various return periods, referred to as frequency curve, is needed when analysing flood risk. Statistical investigation of peak flows extracted from daily time series is required to determine the flood frequency curve, i.e., the flood magnitude (Q) return period (T) relationship. With the effect of global climate change the hydrologic variables such as precipitation patterns are changing, and in many places they are drastically increasing. Changes in the frequency of flooding events are thus expected under climate change. The use of frequency curves based on historical flood data might, therefore, underestimate the risk associated with the design of water resources infrastructure systems. Flood estimation under climate change requires flow series to be produced under climate change using the climate data obtained from Global Climate Models (GCMs). While most hydrologists are greatly influenced by the result of a single GCM data when making decision about climate change impact assessment on flood risk, it must be borne in mind that the utilization of a single GCM may only represent a single realization out of a multiplicity of possible realizations and therefore cannot be representative of future. Considerations that should also be borne in mind when estimating design flood magnitude under climate change including

1. Aware that a good number of climate models (GCMs) and scenarios are available. Criteria such as consistency with global projections, physical plausibility, applicability in impact assessments, representative, and accessibility should be met by climate scenarios if they are to be useful for impact researchers and policy makers suggested by IPCC (2007).

2. Understand that the challenges using GCMs such as climate data generated at course resolution relative to the scale of exposure units in most impact assessments. Also, it is extremely unlikely that these models properly reproduce highly variable fields, such as precipitation, on a regional scale.
3. Understand that the use of downscaling in climate impact studies. The pros and cons of the selected downscaling technique should be known beforehand. Such as, a widely used downscaling technique, weather generator, did not take into account daily extreme changes in climate variability from climate models.

4. Understand that the selection of a hydrological model in the study region concerned. For a small and medium sized basin lumped models are often preferred for climate impact assessment studies. Semi distributed models can be opted if the basin size is relatively big. Complex and distributed models are often not preferred by hydrologists in climate impact studies.

5. Understanding the difference between the two primary flood frequency models annual maxima (AM) and peak over threshold (POT).

6. Understand that the theoretical distribution based on historical observations might be different for the future climate conditions in the production of flood frequency curves.

7. Understand that there is a great deal of uncertainty involved in the estimation of future extreme floods using any climate impact assessment procedure. Also, understand that the choice of GCM used in producing the extreme floods is the major source of uncertainty.

8. Understand that probabilistic assessment is a preferred approach to encompass the uncertainty linked with climate data. This can be done by incorporating many GCMs and carefully chosen emission scenarios.

9. Understand that it is impossible to predict the future floods accurately. The goal of an impact assessment is to include the uncertainty associated with future design floods into engineering and management practices.
Keeping in mind the considerations, the methodologies that should be used in the study for the assessment of climate change impacts on flood frequency (Das and Simonovic, 2012) are

1. A good number of climate models (GCMs) and emission scenarios should be used to obtain a range of climate data
2. A downscaling technique is employed to downscale climate data from GCMs.
3. A continuous hydrological model is used for the hydrological simulations
4. A flood frequency approach is used to assess peak flows extracted from the simulated flow series and
5. Finally a probabilistic model should be employed to encompass uncertainty linked with climate data for a design flood

The aim of this report is to provide guidelines for the assessment of climate change impact on design flood for engineering practice. The outline of the report is organised as follows. Chapter 2 presents climate data preparation for frequency analysis. Chapter 3 describes methodologies for frequency analysis. Chapter 4 describes procedures for uncertainty assessment. Chapter 5 presents and discuss the results of a river basin namely The Upper Thames River basin (UTRb) as a case study. Finally the study is concluded in Chapter 6. It is hoped that this manual can contribute to a better understanding of climate change impacts on flood frequency, and is expected to help water practitioners to assess flood risk in a changing climate.

2 Preparing Datasets for Frequency Analysis

Preparing datasets is the first step in the climate change impact assessment. This section outlines the procedures to be employed to prepare data sets for frequency analysis under climate change. The following steps are implemented to produce peak flow series for a stream gauge under climate change. The Byron stream gauge, located in South-East of the Upper Thames River basin, is selected. The description of the Upper Thames Basin is provided in Appendix A.
1. **Obtain climate data at grid points:** Climate data for selected Atmosphere-Ocean Global Climate Models (AOGCM’s) scenarios is to be collected from the nearest grid points surrounding the study area. The Canadian Climate Change Scenarios Network (CCCSN) provides access to those AOGCM models and emissions scenarios. Data can be obtained for several time slices: such as for baseline period 1961-1990, and for future period such as 2011-2040, 2041-2050 and 2071-2100. Seven variables such as minimum temperature, maximum temperature, precipitation, and specific humidity, northward wind component, southward wind component and mean sea level pressure are need to be collected. Appendix B provides a brief description of climate model and scenarios. A list of several climate models including their origin and associated scenarios used in this report is also provided in the appendix.

2. **Interpolation of climate data:** Climate variables from the nearest grid points are need to be interpolated to provide a data set for each of the stations of interest located around the study area. The inverse distance weighting (IDW) method is often used for this purpose. The four nearest grid points to the station of interest are chosen, and the distance, d from the station to each point is computed. Each gridded value is then assigned a weight, \( w \) using Eq. 2.1. The weighted average of the variable, \( p \), for the station is then computed using Equation 2.2 (where the subscript j represents the jth grid point and the subscript i represents the station).

\[
w = \frac{1/d_i^2}{1/d_1^2 + 1/d_2^2 + 1/d_3^2 + 1/d_4^2}
\]

\[
p_i(t) = \sum_{j=1}^{4} w_j P_j(t)
\]

3. **Calculation of Change Factors from AOGCM Outputs:** Calculation of change factors for future climate is to be performed for the purpose of downscaling using weather generator. Using the AOGCM datasets for each station, monthly averages are to be computed for each variable for both the baseline (1960-1990) and the future time slices.
(2011-2040, 2041-2070 or 2071-2100). For maximum temperature, minimum temperature, northward wind speed, eastward wind speed and mean sea level pressure, the monthly change factors are computed as the difference between the baseline and the future averages. For precipitation and humidity, the change factors are taken as the percent change between the baseline and the future averages. The change factors are used to modify the historic datasets for each station. The historical daily data for humidity and precipitation are multiplied by the monthly change factors. For the rest of the variables, the change factors are added to modify the historical data. Appendix A lists the historic datasets used in this study. This is done because daily climate data was not available. It is understood that the way historic data was modified, it does not allow for more complex changes in daily extreme climate data.

4. **Producing synthetic data series using weather generator**: Modified historic data sets, are then used as input into the weather generator model (knnCADV3) to produce synthetic data series for future climate. Appendix C provides a detailed description of weather generator, KnnCADV3. Synthetic data series can be produced for the stations in question. This report uses data from 22 stations for the period of 1979-2005 (N=27) to simulate data series. Another scenario “baseline” is developed by perturbing historical data to produce data series at baseline. Inter climate variability of a climate model should be performed. At least 25 model runs are to be performed to account for inter climate variability. Modified historic data sets, are used as input into the WG-PCA to produce a synthetic dataset.

5. **Use of hydrological model to generate flow series**: A hydrological model is required to be setup, calibrated and validated for the study basin. This study uses HEC-HMS to do hydrologic simulations. HEC-HMS is widely used in North American river basins and often recommended by hydrologists. Appendix D provides a detailed description of HEC-HMS model. The locations of 22 stations used for obtaining climate data do not correspond to the locations of the sub-basins of UTWb. The synthetic data series derived
from the weather generator is therefore spatially interpolated in order to be used by the hydrologic model. The inverse distance weighting (IDW) method can be used for interpolation. The interpolated synthetic data series of precipitation, maximum and minimum air temperature are fed into the calibrated hydrological model to get the simulated flow series.

6. **Extraction of Flood series data:** Peaks are extracted from the daily flow series. A highest peak value is extracted for each year for annual maximum flood frequency analysis. An average rate of three per year (i.e. the peak threshold is implicit), using the set of rules outlined by Willems (2008) is extracted for POT flood frequency analysis.

## 3 Frequency Analysis

In flood frequency analysis (FFA), a relationship between a flood magnitude Q and its return period T is developed by statistical modelling of a time series of peak flows. Two types of flood peak series, namely annual maximum (AM) series and peaks over threshold (POT) are primarily used. The AM series consists of one value, the maximum peak flow, from each year of record while the POT series consists of all well-defined peaks above a specified threshold value.

The \( Q_T - T \) relationship in an AM model has the form,

\[
1 - F(Q_T) = \frac{1}{T}
\]  
(3.1)

\( F(\quad) \) is the cumulative frequency distribution of flood magnitude \( Q \)

In a POT model, a series of well-defined flood peaks above a specified threshold \( q_0 \) is fitted with a continuous probability distribution. The flood events are modelled by a discrete probability distribution, such as Poisson distribution, and the model is of the form:
\[1 - F(Q_T | Q_T > q_0) = \frac{1}{\lambda T}\] (3.2)

where \(F(\cdot)\) is the cumulative frequency distribution of flood magnitude, \(Q > q_0\). \(\lambda\) is the number of peaks per year included in the POT series. According to Cunnane (1989), the POT model is statistically more efficient than the Annual Maximum (AM) model when \(\lambda > 1.65\). The criteria for selecting peaks for POT analysis are described in Appendix E.

The Generalized Extreme Value (GEV) distribution, of which EV1 is a special case, is the most popular model for annual maximum (AM) series analysis. The Generalized Pareto Distribution (GPD), of which the exponential distribution is a special case, with Poisson arrival rate is the most popular model for POT series analysis. It has been shown that under the assumption of peak independence, the generalized Pareto distribution with Poisson arrival rate leads to the generalized extreme value (GEV) distribution for the AM series (Wang, 1991).

The fitting of a distribution to a flood series is often carried out using the method of L-moments (Hosking and Wallis, 1987). The theoretical formula for GEV and GPD distributions and the formula for estimating parameters using L-moments are described in Appendix F. Formula for quantile estimation are also provided in the appendix.

4 Uncertainty Assessment

The approach of assessing uncertainty of design flood under climate change is subject to various sources of uncertainty. Climate data (GCM structure, future emission scenarios, climate variability, course spatial and temporal scales) and simulated hydrologic regimes (future land use scenarios, hydrological model structure, model parameters) are the main source of uncertainty (Prudhomme et al., 2003; Bates et al., 2008; Jung et al., 2011). Considering most of the above
uncertainty sources in their study, Jung et al. (2011) found that changes in flood frequency are more sensitive to climate change and the uncertainty caused by climate models is higher than due to other sources. The recommended approach of quantifying uncertainty linked with climate data is to use climate projections obtained from the combinations of several GCMs and carefully chosen emission scenarios. Flood peak series derived from different AOGCMs are used to estimate flood magnitude - return period relationships (Q-T curves). A good number of design values obtained for different scenarios for a particular return period can be assumed to be a good representation of variability under climate change and these can be used to establish an uncertainty measure. The consideration of large number of climate models and scenarios also permits a probabilistic assessment of future flood flow uncertainty. The probabilistic treatment of climate related uncertainty was performed in many recent studies. Other than simple normal assumption, non-parametric assumption (e.g., Ghosh and Mujumdar, 2007; Solaiman and Simonovic, 2011) or Bayesian approach (Tebaldi et al., 2004) can be employed to estimate an uncertainty bound. Non parametric based approach, normal kernel function (Bowman and Azzalini, 1997), is used in this study to construct probability density functions (PDF). The kernel density estimator ‘ksdensity’ function provided by MATLAB is used in this study to calculate the probability density function.

5 Analyses and Results of a Case Study

The methodologies described in the previous chapters are applied to the Upper Thames River basin (UTRb). This chapter presents the results of the statistical procedures applied to the POT data series produced for Byron stream gauging station located at the Upper Thames River basin under changing climate conditions.

Figure 5.1 displays the flood frequency results (Q-T curve) for all the time horizons (2020, 2050 and 2080) and for the baseline (1979-2005). The flood frequency curves for all climate model scenarios are produced using the GPD (Generalized Pareto Distribution) model. Peaks are extracted from the flow series for the Byron station at an average rate of three per year (i.e. the peak threshold is implicit), using the set of rules outlined by Willems (2008). Thus 81 most
extreme floods are selected for each of the flow series. A total of 375 POT series (15 AOGCMs x 25 model runs each) are derived for future climate projections (2020, 2050 and 2080). For baseline, 25 POT series are obtained by perturbing historical data 25 times using the weather generator. The flood frequency curve derived from historic data is also shown for information only. The AOGCMs for the highest and the lowest frequency curves at T= 250 are also shown in Figure 5.1. For example, the highest and lowest frequency curves (i.e. upper and lower boundary) for 2020 are obtained from a model run derived from CGCM3T63-A2 and MICRO3MEDRES-A1B, respectively. It is found that the corresponding magnitudes for 100 (250)-year floods are respectively 1175(1626) and 391(478) m$^3$/s. It is to be mentioned that the corresponding 100 and 250-year floods based on historic data are 955 and 1107 m$^3$/s, respectively. These indicate how the return values vary with the application of different AOGCMs, due to the assumptions made in each model. It is to be noted that the upper and lower boundary for the future time horizons are provided by different climate models. In terms of emission scenarios, A2 accommodates the upper boundary most often, which is expected. A1B provides the lower boundary in 2 out of 3 cases, while B1 provides the remaining one. That suggests uncertainty linked with climate data should be better-quantified by incorporating climate projections from available GCMs and carefully chosen emission scenarios.

A large number of flow values (15 AOGCMs X 25 model runs each = 375 data series) obtained for different model scenarios for a particular return period can be assumed to be a good representation of flow variability under climate change and these can be used to derive an uncertainty measure. Non parametric based approach, normal kernel function, is used in this study to construct probability density functions (PDF). Figure 5.2 shows, for example, the PDFs of 100-year return period flood for each AOGCM at time horizon 2050. The PDFs are constructed for each AOGCM using 25 model runs. The PDFs are different for different AOGCMs. A greater variance is observed for the climate models, CGCM3T63-A2 and MICRO3MEDRES-B1. Probability density plots are also constructed by incorporating data from all the 15 climate projections. They are displayed in Figure 5.3 for floods at return period, T = 10, 100 and 250 at the four time horizons for comparison. The plots show that variability increases with time, as PDFs become wider. With wider PDFs, a greater variance of design floods is noticed. It is found that the average percentage changes of the 100-year flood
magnitude between the future climate (2020, 2050 and 2080) and the baseline (1979-2005) are respectively 8, 12 and 12.3%. The corresponding percentage changes for the 250-year flood are respectively 19, 32 and 32.5%. In case of 10-year flood, very little changes have been observed between future and baseline climate projections. The information from Figure 5.3, are converted to cumulative distribution functions (CDFs). They are displayed in Figure 5.4. The CDFs allow the uncertainty of design flood to be quantified with high level of confidence.

From the results of flood magnitude – return period relationship (Q-T) it can be said that there is significant variation between climate models and this variation contributes more when an effort is carried out to predict floods for a far distant future. Hence, climate change impact studies based on only one AOGCM and/or emission scenario should be considered with a great care. The use of two carefully chosen climate projections (dry and wet projections, for example) may be more appropriate than using single model and this has been done in several recent climate change impact studies (e.g. Prodanovic and Simonovic, 2006).

One of the limitations to the approach presented in this case study is linked with hydrologic model calibration. In this case study no attempt has been made to recalibrate the model. The approach implicitly assumes that the calibration is equally acceptable for the baseline and the future conditions. It is recommended that a hydrological model should be calibrated before applying to a river basin. Another limitation is linked to downscaling approach based on monthly change factor. The approach did not take into account daily extreme changes in climate variability from climate models. This could affect the results and might lead to an underestimation of changes in future flood frequency. For incorporating inter climate variability of a climate model, 25 model runs are performed in this study. A greater number of model runs should provide a better variability incorporated by a climate model.
Figure 5.1 Simulated flood frequency results for climate data. Each plot relates to different time horizons. Data from 375 scenarios (15 climate model scenarios x 25 model runs each) are used for future climate. Each line with a specific color represents a different AOGCM. The upper and lower bound frequency curves are indicated in the plots. The flood frequency curve derived from historic data is also shown for information only. Data from 25 runs are used to produce a range of results for baseline.

Figure 5.2 Probability density functions (PDFs) of 100-year return period flood for all AOGCMs at the time horizon 2050. Each PDF is constructed using 25 model runs.
Figure 5.3 Probability density plots for flood magnitudes at return period, $T = 10, 100$ and 250 at the four time horizons

Figure 5.4 Cumulative distribution functions for flood magnitudes at return period, $T = 10, 100$ at the four time horizons.
6 Conclusion

This study presents the guidelines and associated procedures that should be used to assess design flood under climate change. This report emphasis the use of multi model multi scenario approach to estimate uncertainty associated with climate data. The recommended procedures are applied to the Upper Thames River basin (UTRb) and the associated results are presented. In addition to traditional guidelines (e.g. downscaling climate data from climate model, application of a hydrological model, frequency analysis) the following key guidelines should be borne in mind when assessing climate change impact on design future floods.

1. Climate change impact studies based on only one AOGCM and/or emission scenario should be considered with a great care. The use of two carefully chosen climate projections may be more appropriate than using single model. Multi-model multi-scenario approach is preferred.

2. A large uncertainty exists in all the projected future design floods.

3. The application of a wide range of climate models and scenarios allows performing probabilistic approach to better outline the uncertainty linked with climate data.

4. For engineering and management practices it is recommended to include the uncertainty associated with future design floods.
Acknowledgements

The authors would like to acknowledge the financial support made available by the Natural Sciences and Engineering Research Council of Canada.

References

Beard, L.R., 1974. Flood Flow Frequency Techniques, Tech. Report CRWR119, Center for Research in Water Resources, University of Texas at Austin, Austin, TX, USA.


Appendix A: Study Area and Data

The Upper Thames River basin (UTRb) located in London Ontario, Canada is used a case study. The Upper Thames River basin has an area of 3,842 km² located between Lake Huron and Lake Erie in Southwestern Ontario. Majority of the river basin is covered with agricultural lands (80%), with forest cover and urban uses taking about 10% each. London is the major urban centre with a population of around 366,151 inhabitants, many of whom experience the effects of flooding as the Thames River runs directly through the City. The Thames River with a total length of 273 km has an average annual discharge of 35.9 m³/s. The UTRb receives approximately 1,000 mm of annual precipitation; however 60% of this is lost due to evaporation and evapotranspiration. Figure A.1 shows a schematic map of the Upper Thames River basin. Several weather stations are located throughout the basin to provide point measurements of climatic variables. Stations chosen for this study are listed in Table A.1. Daily weather data (precipitation, maximum temperature and minimum temperature) for the period of 1979-2005 was obtained from Environment Canada (http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html) for each of the stations listed in Table A.1. Stations were chosen based on the completeness and length of the observed data. The historic daily flow data for the Byron gauging station was obtained from Environment Canada (http://www.wateroffice.ec.gc.ca).
Figure A.1 Map of the Upper Thames River basin
Table A.1 Location of stations in the Upper Thames River basin

<table>
<thead>
<tr>
<th>Station</th>
<th>Latitude(deg N)</th>
<th>Longitude(deg W)</th>
<th>Elevation(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blyth</td>
<td>43.72</td>
<td>81.38</td>
<td>350.5</td>
</tr>
<tr>
<td>Brantford MOE</td>
<td>43.13</td>
<td>80.23</td>
<td>196</td>
</tr>
<tr>
<td>Chatham</td>
<td>42.38</td>
<td>82.2</td>
<td>198</td>
</tr>
<tr>
<td>Delhi CS</td>
<td>42.87</td>
<td>80.55</td>
<td>255.1</td>
</tr>
<tr>
<td>Dorchester</td>
<td>43</td>
<td>81.03</td>
<td>271.3</td>
</tr>
<tr>
<td>Embro</td>
<td>43.25</td>
<td>80.93</td>
<td>358.1</td>
</tr>
<tr>
<td>Exeter</td>
<td>43.35</td>
<td>81.5</td>
<td>262.1</td>
</tr>
<tr>
<td>Fergus</td>
<td>43.73</td>
<td>80.33</td>
<td>410</td>
</tr>
<tr>
<td>Foldens</td>
<td>43.02</td>
<td>80.78</td>
<td>328</td>
</tr>
<tr>
<td>Glen Allan</td>
<td>43.68</td>
<td>80.71</td>
<td>404</td>
</tr>
<tr>
<td>Hamilton A</td>
<td>43.17</td>
<td>79.93</td>
<td>238</td>
</tr>
<tr>
<td>Ilderton</td>
<td>43.05</td>
<td>81.43</td>
<td>266.7</td>
</tr>
<tr>
<td>London A</td>
<td>43.03</td>
<td>81.16</td>
<td>278</td>
</tr>
<tr>
<td>Petrolia Town</td>
<td>42.86</td>
<td>82.17</td>
<td>201.2</td>
</tr>
<tr>
<td>Ridgetown</td>
<td>42.45</td>
<td>81.88</td>
<td>210.3</td>
</tr>
<tr>
<td>Sarnia</td>
<td>43</td>
<td>82.32</td>
<td>191</td>
</tr>
<tr>
<td>Stratford</td>
<td>43.37</td>
<td>81</td>
<td>354</td>
</tr>
<tr>
<td>St. Thomas WPCP</td>
<td>42.78</td>
<td>81.21</td>
<td>209</td>
</tr>
<tr>
<td>Tillsonburg</td>
<td>42.86</td>
<td>80.72</td>
<td>270</td>
</tr>
<tr>
<td>Waterloo Wellington</td>
<td>43.46</td>
<td>80.38</td>
<td>317</td>
</tr>
<tr>
<td>Woodstock</td>
<td>43.14</td>
<td>80.77</td>
<td>282</td>
</tr>
<tr>
<td>Wroxeter</td>
<td>43.86</td>
<td>81.15</td>
<td>355</td>
</tr>
</tbody>
</table>
Appendix B: Climate Model and Scenarios

Global circulation models namely, coupled Atmosphere-Ocean Global Climate Models (AOGCMs) are current state of the art in climate impact research. AOGCMs are the most viable tools for simulating physical processes in the atmosphere, ocean, cryosphere and land surface that determine global climate (IPCC, 2007). They are based on various assumptions about the effects of the concentration of greenhouse gases in the atmosphere coupled with projections of CO₂ emission rates (Smith et al., 2009). AOGCMs are associated with model structure developed by various countries, and the emission scenarios. Three emission scenarios “A1B”, “B1” and “A2” out of the family of emission scenarios (Nakićenović and Swart, 2000) are most commonly used in climate impact studies. These represent respectively “the productive world with rapid economic expansion and abundance of energy sources”, “the sustainable world with clean technologies” and “the world of independent nations with increasing population and slower technological advancements”. In terms of greenhouse gas emission, these represent respectively medium, low and high emission scenarios for the 21st century (Randall et al., 2007). A total of 15 climate projections from 6 AOGCMs, each with two to three emission scenarios are selected for investigation. A list of these models including their origin and associated scenarios is provided in Table A.2.

Table A.2 List of AOGCM models and emission scenarios used

<table>
<thead>
<tr>
<th>GCM models</th>
<th>Sponsors, Country</th>
<th>Emission Scenarios</th>
<th>Atmospheric Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGCM3T47, 2005</td>
<td>Canadian Centre for Climate Modelling and Analysis, Canada</td>
<td>A1B, B1, A2</td>
<td>3.75° 3.75°</td>
</tr>
<tr>
<td>CGCM3T63, 2005</td>
<td></td>
<td>A1B, B1, A2</td>
<td>2.81° 2.81°</td>
</tr>
<tr>
<td>CSIROMK3.5, 2001</td>
<td>Commonwealth Scientific and Industrial Research Organization (CISRO) Atmospheric Research, Australia</td>
<td>B1, A2</td>
<td>1.875° 1.875°</td>
</tr>
<tr>
<td>Model</td>
<td>Institution</td>
<td>A1B, B1</td>
<td>Resolution</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>GISSAOM, 2004</td>
<td>National Aeronautics and Space Administration (NASA)/Goddard Institute for Space Studies (GISS), USA</td>
<td>A1B, B1</td>
<td>3°</td>
</tr>
<tr>
<td>MIROC3.2HIRES, 2004</td>
<td>Centre for Climate System Research (University of Tokyo), National Institute for Environmental Studies, and Frontier Research Centre for Global Change (JAMSTEC), Japan</td>
<td>A1B, B1</td>
<td>1.125°</td>
</tr>
<tr>
<td>MIROC3.2MEDRES, 2004</td>
<td>Research Centre for Global Change (JAMSTEC), Japan</td>
<td>A1B, B1, A2</td>
<td>2.8°</td>
</tr>
</tbody>
</table>
Appendix C: Weather Generator

Downscaling is employed to address the deficiencies (i.e. coarse spatial and temporal resolution) of global climate models for use at local scales. Downscaling based on weather generator is widely used in climate impact studies (Rajagopalan and Lall, 1999; Yates et al., 2003; Cunderlik and Simonovic, 2005; Solaiman and Simonovic, 2011). Weather generator stochastically simulates climate information for an area by combining both, local and global weather data. The local data are used to address the fine spatial and temporal scale issues needed for impact studies by including historically observed data obtained from stations in and around the study area. The global data provide the general direction of change of the climate within the region of interest by including outputs obtained from global climate models.

The principle component analysis integrated stochastic weather generator (KnnCADV3) (Eum et al., 2009) is used in this study to produce synthetic data sets. The model based on K-Nearest Neighbour (K-NN) algorithm is an extended version of the weather generator model developed by Sharif and Burn (2006). The model operates by generating weather for a new day for a station of interest. This has been done by extracting all days with similar characteristics, known as nearest neighbours, from the historic record from which a single value is selected according to a defined set of rules. The set of rules are described in detail by Eum et al. (2009). The addition of principal component analysis provides reduction in computational requirements and allows more variables to be included for an improved selection of nearest neighbours. The model also includes a perturbation mechanism which allows newly generated values to be outside of the observed range. The data sets produced in this way take into account natural variability when predicting the effects of climate change.
Appendix D: Hydrological Model

This study uses HEC-HMS model developed by the US Army Corps of Engineers (USACE) to carry out hydrological simulations in a continuous mode. The model has been widely applied in many geographical locations for solving a variety of hydrological problems (Yu et al., 2002; Fleming and Neary, 2004; Cunderlik and Simonovic, 2004; Grillakisa et al., 2011) and used by the local water authority (the Upper Thames River Conservation Authority-UTRCA), in everyday practice. Precipitation, air temperature, and estimated potential evapotranspiration are used as input data for HEC-HMS model. Additionally, soil information and land use data are required for estimating initial parameter sets for the model.

The overall structure of the continuous simulation version of the HEC-HMS model used in this study is presented in Figure A.2. The model includes four components. Each component represents a module that mathematically represents a physical processes functioning in the river basin. Snow module takes precipitation and air temperature (maximum and minimum) data obtained from the weather generator as inputs to separate solid and liquid form of precipitation. The algorithm of the snow module is based on a degree-day method (Cunderlik and Simonovic, 2004). The output of the snow module is adjusted precipitation, used for computation of losses.

The losses module integrated with HMS is soil-moisture accounting (SMA). The module is used to estimate and subtract the losses (interception, infiltration and evapotranspiration) from adjusted precipitation. The 5-layer SMA module shown in Figure A.3 is based on Precipitation-Runoff modeling System, PRMS (Leavesly and Stannard, 1995) designed to compute runoff discharge on a continuous time basis. The SMA uses four types of conceptual storage: canopy-interception, surface-interception, soil profile, and a number of ground water storage. The inflow and outflow rates (i.e. evapotranspiration, infiltration, percolation, surface runoff and ground water flow) between the reservoirs regulate the amount of water stored in each conceptual reservoir.

Surface excess, groundwater flow and ground water recharge are outputs from the losses module. The Clark unit hydrograph (USACE, 2006) is used to convert surface excess to direct runoff.
The groundwater flow is transformed into baseflow by a series of linear reservoir model. Both direct runoff and baseflow enter the river. The translation and attenuation of flow in river reach is simulated by the modified pulse method (USACE, 2006). The ground water recharge enters deep aquifers and does not return to the stream.

Figure A.2 Continuous HEC-HMS hydrologic model structure
Figure A.3 Soil moisture accounting losses module
The hydrologic model used in this study was originally developed and applied to the Upper Thames River in the work by Cunderlik and Simonovic (2004; 2005). The model consists of thirty three sub-basins, twenty one river reaches, and three reservoirs namely Wildwood, Fanshawe and Pittock (Prodanovic and Simonovic, 2006; Fig 9). Each sub-basin is provided with interpolated precipitation and maximum and minimum temperature data. The outputs of each sub basin are flow hydrographs joined by junctions where the flows are added together. River reaches represent the major rivers in the basin connected between two junctions. The routing module (i.e. modified puls) is applied to each river reach, and thus acts as a passage of a flood wave as it moves through the river system. The same routing rules are also applied to reservoirs.

The model was calibrated and verified with extensive sensitivity analyses in the work by Cunderlik and Simonovic (2004; 2005). The model is seasonal in nature with different parameters referring to the summer and winter seasons. The parameter sets for the summer and winter seasons are presented by Cunderlik and Simonovic (2004) and Prodanovic and Simonovic (2007). The current program runs in Java program (Prodanovic and Simonovic, 2006). The necessary codes for developing the model can be downloaded from the FIDs website at (http://www.eng.uwo.ca/research/iclr/fids/publications/products/ContinuousModelReport2. pdf, retrieved on 5/17/2012).
Appendix E: Selection of Peaks

In an annual maximum (AM) model, the largest peak discharge of a year is selected. An annual maximum flood series is a sequence of the largest peak of each year of record.

In a POT model, flood peaks can be obtained using different methods from a time series. The number of floods (M) generally will be different to the number of years of record (N), and will depend on the selected threshold discharge. The US Geological Survey (Dalrymple, 1960) recommended that M should equal 3N. The UK Flood Studies Report (NERC, 1975) recommended that M should equal 3N to 5N. A criterion for independence of successive peaks must also be applied in selecting events. Beard (1974) used a criterion that flood peaks should be separated by five days plus the natural logarithm of the square miles of drainage area, with the additional requirement that intermediate flows must drop to below 75% of the lower of the two separate flood peaks. The UK Flood Studies Report (NERC, 1975) used a criterion that flood peaks should be separated by three times the time to peak and that the flow should decrease between peaks to two thirds of the first peak. An excellent review on the selection of flood peaks is presented by Lang et al. (1999). The method proposed by Willems (2003, 2008) is used in this report where two adjacent peaks are considered independent if:

(i) the time between the two peaks is longer than the recession constant of the quick flow runoff components for the given basin;

(ii) the minimum discharge between the two peaks is smaller than 37% of the peak discharge.

The POT values were extracted by applying the criteria mentioned above using the WETSPRO software, which has been developed by the Hydraulics Laboratory of K.U. Leuven in Belgium (Willems, 2003; 2008).
Appendix F: Common Statistical Distributions and Parameter Estimation Procedures

L-Moments

The method of L-moments is often used and is recommended by many research hydrologists for estimating distribution parameters. L-moments are analogous to conventional moments defined as linear combinations of the probability weighted moments (PWMs). Theoretical formula in terms of the basic population quantities can be obtained from Hosking and Wallis (1997). Formulae for the calculation of these statistics from the sample data are provided in Table A.3.

Table A.3 Calculation of L-moments and their dimensionless ratios, modified after (Das, 2010)

<table>
<thead>
<tr>
<th>Definitions based on Probability Weighted Moments (PWMs)</th>
<th>Sample estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population quantities</td>
<td>Sample estimates</td>
</tr>
<tr>
<td>M(_{100}) = 1(^{st}) PWM</td>
<td>(\hat{M}_{100} = \bar{Q}) = sample mean</td>
</tr>
<tr>
<td>M(_{110}) = 2(^{nd}) PWM</td>
<td>(\hat{M}<em>{110} = \frac{1}{N} \sum</em>{i=1}^{N} \left(\frac{i-1}{N-1}\right) Q_{(i)})</td>
</tr>
<tr>
<td>M(_{120}) = 3(^{rd}) PWM</td>
<td>(\hat{M}<em>{120} = \frac{1}{N} \sum</em>{i=1}^{N} \left(\frac{(i-1)(i-2)}{(N-1)(N-2)}\right) Q_{(i)})</td>
</tr>
<tr>
<td>M(_{130}) = 4(^{th}) PWM</td>
<td>(\hat{M}<em>{130} = \frac{1}{N} \sum</em>{i=1}^{N} \left(\frac{(i-1)(i-2)(i-3)}{(N-1)(N-2)(N-3)}\right) Q_{(i)})</td>
</tr>
<tr>
<td>(\lambda_1) = 1(^{st}) L-Moment</td>
<td>L1 = M(_{100})</td>
</tr>
<tr>
<td>(\lambda_2) = 2(^{nd}) L-Moment</td>
<td>L2 = 2M(<em>{110}) - M(</em>{100})</td>
</tr>
<tr>
<td>(\lambda_3) = 3(^{rd}) L-Moment</td>
<td>L3 = 6M(<em>{120}) - 6M(</em>{110}) + M(_{100})</td>
</tr>
</tbody>
</table>

where Q\(_{(i)}\) is the \(i^{th}\) smallest value.
\[ \lambda_4 = 4^{th} \text{L-Moment} \]
\[ L4 = 20M_{130} - 30M_{120} + 12M_{110} - M_{100} \]

\[
\begin{array}{|c|c|}
\hline
\tau_2 & L-CV \\
\hline
\tau_3 & L-Skewness \\
\hline
\tau_4 & L-Kurtosis \\
\hline
\end{array}
\]

\[ t_2 = \frac{L2}{L1} \]

\[ t_3 = \frac{L3}{L2} \]

\[ t_4 = \frac{L4}{L2} \]

**Generalized Pareto Distribution (GPD)**

This is a three parameter distribution and the distribution function is

\[
F(q) = P(Q < q / q > q_0) = 1 - \left[ 1 - \frac{k}{\beta} (q - q_0) \right]^\frac{1}{k}
\]

(A.1)

where \( q_0 \) is the threshold \( \beta \) is a scale parameter and \( k \) is a shape parameter.

When \( k =0 \), this is reduced to exponential distribution of the form

\[
F(q) = 1 - \exp \left[ - \frac{1}{\beta} (q - q_0) \right]
\]

(A.2)

The inverse form of the GPD is

\[
q(F) = q_0 + \frac{\beta}{k} \left[ 1 - (1 - F)^k \right] \quad , k \neq 0
\]

(A.3)

\[
q(F) = q_0 - \beta \ln [1 - F] \quad , k = 0
\]

(A.4)

The estimation of the parameters can be done in either of two distinct ways from a record of \( N \) years.
a) Fix \( q_0 \) a priori and abstract from the record of flows every peak value exceeding \( q_0 \). Let there be M of them.

b) An alternative to fix \( \lambda \) a priori. This determines \( M = \lambda N \) the required sample size. The largest M peaks are then extracted from the record and both \( q_0 \) and \( \beta \) and \( k \) are estimated from a sample of data.

The estimation of the parameters using L-moments are as follows:

For case (a) the two parameters \( \beta \) and \( k \) are given by (Hosking and Wallis, 1997)

\[
k = \left( l_1 - q_0 \right) / l_2 - 2 \tag{A.5}
\]

\[
\beta = (1 + k)(l_1 - q_0) \tag{A.6}
\]

For case (b), the three parameters are given by (Hosking and Wallis, 1997)

\[
k = \left( l_1 - 3t_3 \right) / \left( 1 + t_3 \right) \tag{A.7}
\]

\[
\beta = (1 + k)(2 + k)l_2 \tag{A.8}
\]

\[
q_0 = l_1 - (2 + k)l_2 \tag{A.9}
\]

where \( l_1 \) is 1\(^{st}\) L-moment, \( l_2 \) is 2\(^{nd}\) L-moment and \( t_3 \) is L-skewness

**Generalized Extreme Value distribution (GEV)**

The cumulative distributional form of GEV is (Hosking and Wallis 1997):

\[
F(Q) = \exp \left\{ - \left( 1 - k(P - \xi) / \alpha \right)^{1/k} \right\} \quad \text{for} \quad k \neq 0 \tag{A.10}
\]
\[ F(Q) = \exp \left[ -\exp \left( -\frac{P - \xi}{\alpha} \right) \right] \quad \text{for} \quad k = 0 \]  

(A.11)

where \( \xi \) is a location parameter, \( \alpha \) is a scale parameter; and \( k \) is a shape parameter. The shape parameter controls the shape and size of the tails of each of the three distributions. When the shape parameter, \( k = 0, k < 0 \) and \( k > 0 \), this is respectively the EV1 (Gumbel), EV2 (Frechet), and EV3 (Weibull) distribution.

The parameters are estimated using L-moments as follows (Hosking and Wallis 1997):

\[ k = 7.8590c + 2.9554c^2, \quad \text{in which} \quad c = \frac{2}{3 + \tau_3} - \frac{\ln 2}{\ln 3} \]  

(A.12)

\[ \alpha = \frac{\lambda_1 k}{(1 - 2^{-k}) \Gamma(1 + k)} \]  

(A.13)

\[ \xi = \lambda_1 - \alpha \{1 - \Gamma(1 + k)\} / k \]  

(A.14)

where \( \lambda_1 \) and \( \lambda_2 \) are the first two L-moments and \( \tau_3 \) is the L-skewness. \( \Gamma \) denotes the complete gamma function.

Using the estimated parameters, the T-year return period events, \( P_T \), are estimated for the GEV as follows:

\[ P_T = \xi + (\alpha / k) \left[ 1 - \left( -\log \left( \frac{T - 1}{T} \right) \right) \right] \]  

(A.15)

The parameters of an EV1 distribution are estimated by the first two L-moments \( \lambda_1 \) and \( \lambda_2 \) as follows: (Hosking and Wallis 1997)

\[ \lambda_1 = \xi + \alpha \epsilon \]  

(A.16)
\[ \lambda_2 = \alpha \ln 2 \quad \text{(A.17)} \]

where \( \varepsilon \) is the Euler’s constant = 0.5772.

The T-year event for the EV1 distribution is estimated as

\[ P_T = \xi + \alpha y_T \quad \text{(A.18)} \]

where \( y_T = -\ln(-\ln(1 - 1/T)) \) is the EV1 reduced variate for a T-year return period.
Appendix G: Previous Reports in the Series

ISSN: (print) 1913-3200; (online) 1913-3219


