**THE UNIVERSITY OF WESTERN ONTARIO DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING**

**Water Resources Research Report**

**Quantifying Uncertainties in the Modelled Estimates of Extreme Precipitation Events at the Upper Thames River Basin**

> **By: Tarana A. Solaiman and Slobodan P. Simonovic**



**Report No: 070 Date: February 2011**

**ISSN: (print) 1913-3200; (online) 1913-3219; ISBN: (print) 978-0-7714-2878-4; (online) 978-0-7714-2880-7;**

## **Quantifying Uncertainties in the Modelled Estimates of Extreme**

## **Precipitation Events at the Upper Thames River Basin**

**By: Tarana A. Solaiman And Slobodan P. Simonovic**



**Department of Civil and Environmental Engineering The University of Western Ontario London, Ontario, Canada**

**February 2011**

### **Executive Summary**

<span id="page-2-0"></span>Assessment of climate change impact on hydrology at watershed scale incorporates downscaling of global scale climatic variables into local scale hydrologic variables and computations of risk of hydrologic extremes in future for water resources planning and management. Atmosphere-Ocean General Circulation (AOGCM) models are designed to simulate time series of future climate responses accounting for enthropogenically induced green house gas emissions. The climatological inputs obtained from several AOGCMs suffer the limitations due to incomplete knowledge arising from the inherent physical, chemical processes and the parameterization of the model structure. This study explores the methods available for quantifying uncertainties from the AOGCM outputs by considering fixed weights from different climate model means for the overall data lengths and provides an extensive investigation of the variable weight nonparametric kernel estimator based on the choice of bandwidths for investigating the severity of extreme precipitation events over the next century. The results of this study indicate that the variable width method is better equipped to provide more useful information of the uncertainties associated with different AOGCM scenarios. This study further indicates an increase of probabilities for higher intensities and frequencies of events. The applied methodology is flexible and can be adapted to any uncertainty estimation studies with unknown densities.

# **Table of Contents**

<span id="page-3-0"></span>



# **List of Tables**

<span id="page-5-0"></span>

# **List of Figures**

<span id="page-6-0"></span>

# **Introduction**

### <span id="page-7-1"></span><span id="page-7-0"></span>**1.1 Problem Definition**

Hydrologic research and modeling is largely dependent on climatological inputs due to the inextricable link of water with climate. Water is the most vulnerable resource to climate change (Minville et al., 2008; Srikanthan and McMohan, 2001; Xu and Singh, 2004) resulting in an increased evaporation due to higher temperatures, changes in amount, variability, and frequency of regional precipitation. Studies related to the impact of climate change on water resources have shown a significant impact on the mean annual discharge with changes in the intensity and frequency of precipitation (Whitfield and Cannon, 2000; Muzik, 2001), larger changes in reservoir storage because of a modest change in the natural inflow or even a changed effect in the energy production and flood control measure due to any effect in the hydrologic cycle (Xu and Singh, 2004).

Climate modeling studies involving anthropogenic increase in the concentration of greenhouse gases have suggested an increase in the frequency and intensity of climatic extremes in a warmer world (Cubasch et al., 2001). The evidence of an altered climate has already become noticeable. Recent studies related to the Canadian climate have indicated a 12% increase of precipitation in southern Canada during the twentieth century (Zhang et al., 2000; Vincent and Mekis, 2006). This provides the justification of over a decade long effort to determine the impacts of anthropogenic climate change in water resources. However, most efforts have focused on studying the changes of means, although extremes usually have the greatest and most direct impact on our everyday lives,

community and environment. Study on the detection of changes in extremes is limited and hence require further investigation.

Assessment of climate change impacts on hydrology incorporates projection of climate variables into a global scale, downscaling of global scale climatic variables into local scale hydrologic variables and computations of risk of future hydrologic extremes for purposes of water resources planning and management. Global scale climate variables are commonly projected by Coupled Atmosphere-Ocean Global Climate Models (AOGCMs), which provide a numerical representation of climate systems based on the physical, chemical and biological properties of their components and feedback interactions between these (IPCC, 2007). These models are currently the most reliable tools available for obtaining the physics and chemistry of the atmosphere and oceans and for deriving projections of meteorological variables (temperature, precipitation, wind speed, solar radiation, humidity, pressure, etc). They are based on various assumptions about the effects of the concentration of greenhouse gases in the atmosphere coupled with projections of  $CO<sub>2</sub>$  emission rates (Smith et al., 2009).

There is a high level of confidence that AOGCMs are able to capture large scale circulation patterns and correctly model smoothly varying fields, such as surface pressure, especially at continental or larger scales. However, it is extremely unlikely that these models properly reproduce highly variable fields, such as precipitation (Hughes and Guttorp, 1994), on a regional scale, let alone for small to medium watershed. Although confidence has increased in the ability of AOGCMs to simulate extreme events, such as hot and cold spells, the frequency and the amount of precipitation during intense events are still underestimated.

2

### <span id="page-9-0"></span>**1.2 Definition and Types of Uncertainty**

A proper understanding of the uncertainties resulting from human induced climate change will help decision makers to interpret different projected hydrologic impacts with confidence. Three broad areas of uncertainties have been identified by Colglazier (1991):

- Predicting future climate
- Predicting future impacts
- Assessing costs and benefits of policy responses

The first two areas, related to the present research, are described here.

Predictions of the timing and magnitude of any future global warming are associated with uncertainties in estimating future anthropogenic emissions of greenhouse gases; understanding the resulting changes in the carbon cycle, especially the uptake of carbon in the oceans; understanding the dynamic climatic response with all the relevant feedback mechanisms, such as those from clouds and Ocean currents; projecting regional variations; and estimating the frequency of severe events such as hurricanes and droughts (Colglazier, 1991). Although the basic theory of the enhanced green house gas effect is now well established, and the rise in carbon dioxide concentrations since the industrial revolution has also been well documented, there is still much debate regarding the timing and quantity of warming. For decades AOGCMs have been used to predict these values; however there is continued uncertainty even with the improvements of the resolution of GCMs.

The interpretation of uncertainties from climate models can be described from five sources. 'Forcing uncertainty' consists of using the future elements/aspects that are not a part of the climate system, but have the potential to affect it. One possible form of forcing uncertainty arises from using climate model simulations based on different scenarios of future concentrations of atmospheric GHGs, which depend entirely on the actions taken to control the GHG emissions (Cubasch et al., 2001).

'Initial condition uncertainty' involves uncertainty arising from an initial state or ensemble of states (Stainforth et al., 2007) applied to the climate models. It can be 'macroscopic' and found in state variables with relatively large slowly mixing scales, such that the predicted distribution is effected by the imprecise knowledge of the current state of the system. 'Microscopic' uncertainty, on the other hand, has no significant effect on the targeted climate distribution; the effects are only identified during weather forecast.

'Model imperfection' describes the uncertainty that results from a limited understanding and ability to simulate the Earth's climate. It is sub-divided into two types: 'uncertainty' and 'inadequacy'. 'Model uncertainty' describes uncertainties in the most relevant parameter values to be used in the model (Murphy et al., 2004). It characterizes the impact of known uncertainties and can be large at regional scales. Climate models, in this respect, are considered rather complicated. Extending this from parameter values to parameterizations enables an improved representation of various processes within the model and makes model uncertainty an extended form of the 'parameter uncertainty'(Kennedy and O'Hagan, 2001). 'Model inadequacy' results from the limited ability of the climate models to represent natural systems. These models provide no information on important processes related to climate change on decadal to centennial time scales, such as the carbon cycle, atmospheric and oceanic chemistry and stratospheric circulation. They further suffer from limited spatial resolution, inadequate representation of hurricanes, the diurnal cycle of tropical precipitation, characteristics of El Nino Southern Oscillation (ENSO) and the inter tropical convergence zone (Trenberth et al., 2003).

### <span id="page-11-0"></span>**1.3 Organization of the Report**

The report is organized as follows: literature relevant to the development of methods for assessing climate change impacts are presented in chapter 2. The development of proposed uncertainty estimation methodology is presented step by step in Chapter 3, including the model and parameter set up. Finally, the performances of the variable weight and fixed weight methods for quantification of AOGCM uncertainties in extreme precipitation events are presented. Finally, the findings of the results are discussed in Chapter 4.

### **2. Literature Review**

<span id="page-12-0"></span>In recent years, quantifying uncertainties from AOGCM choice and scenario selections used for impact assessments has been identified as critical for climate change and adaptation research. Climate change impact studies derived from AOGCM outputs are associated with uncertainties due to "incomplete" knowledge originating from insufficient information or understanding of the relevant biophysical processes, or a lack of analytical resources. Examples of uncertainty include the simplification of complex processes involved in atmospheric and oceanographic transfers, inaccurate assumptions about climatic processes, limited spatial and temporal resolution resulting in a disagreement between AOGCMs over regional climate change, etc. Uncertainties also emerge due to "unknowable" knowledge, which arises from the inherent complexity of the Earth system and from our inability to forecast future socio-economic and human behavioral patterns in a deterministic manner (New and Hulme, 2000; Allan and Ingram, 2002; Proudhomme et al., 2003; Wilby and Harris, 2006; Stainforth et al., 2007; IPCC, 2007, Buytaert et al, 2009). Selection of the most appropriate AOGCM for the realization of future climate depends on user's ability to assess the model's strengths and weaknesses, the inability of which is recognized as one of the major sources of uncertainty (Wilby and Harris, 2006, Ghosh and Mujumdar, 2007; Tebaldi and Smith, 2010). It has been established that the accuracy of AOGCMs decrease at finer spatial and temporal scales; a typical resolution of AOGCMs ranges from 250 km to 600 km, but the need for impact studies conversely increases at finer scales. The representation of regional precipitation is distorted due to this coarse resolution and thus it cannot capture the subgrid-scale processes required for the formation of site-specific precipitation

conditions. While some models are parameterized, details of the land-water distribution or topography in others are not represented at all (Widmann et al., 2003). Studies have found that the models failed to predict the high variability in daily precipitation and could not accurately simulate present-day monthly precipitation amounts (Trigo and Palutikof, 2001; Brissette et al., 2006).

### <span id="page-13-0"></span>**2.1 Downscaling of AOGCM Outputs**

In order to estimate uncertainties at smaller spatial scales, output from climate models are scaled down to a suitable level. Statistical and dynamic downscaling represents two common branches for the development of climate scenarios depending on regions, time periods and the variables of interest. The approaches for dynamic downscaling involve (i) running a regional scale limited area model with coarse GCM data as geographical or spectral boundary conditions, (ii) performing global-scale experiments with high resolution Atmosphere-GCM (AGCM), with coarse GCM data as initial (as partially and boundary) conditions, and (iii) the use of a variable-resolution global model with the highest resolution over the area of interest (Rummukainen, 1997). The most common technique for dynamic downscaling involves utilizing Regional Climate Models (RCMs), which perform at a much higher resolution and can simulate climatic variables more accurately for any region (Brissette et al., 2006). AOGCM output variables are used as boundary inputs for the RCMs, and provide a more accurate representation of the local climate than the coarsely gridded AOGCM data alone. The works of Vidal and Wade (2008), Wood et al. (2004) and Schmidli et al. (2006) compared dynamic downscaling to other methods. A limitation of the dynamic approach is the scale of RCM's (approximately 40 km x 40 km according to Brissette et al., 2006), which is still too coarse for application to smaller basins. The computational effort required for the dynamic approach makes it impractical where several AOGCMs and emissions scenarios are used (Maurer, 2007). Furthermore, RCMs have only been produced for selected areas; moving to a slightly different region requires repeating the experiment (Kay and Davies, 2008).

The second approach, namely statistical downscaling, is more popular in climate change impact assessments due to its computational ease and its ability to produce synthetic datasets of any desired length. In statistical downscaling statistical relationships are developed to transfer large-scale features of the predictors (AOGCM) to regional scale predictands (variables). Hewitson and Crane (1992) pointed out three underlying assumptions related to statistical downscaling: (i) the predictors are variables of relevance and are realistically modeled by the host AOGCM; (ii) the empirical relationship is also valid under altered climate conditions; and (iii) the predictors employed fully represent the climate change signal.

Several methods of statistical downscaling can be broadly divided into three categories: Transfer function, weather typing and weather generator. Transfer functions rely on the direct quantitative relationship between the global large scale and local small scale variables obtained from different choices of mathematical transfer functions, predictors or statistical fitting processes. Applications of neural networks, regression based methods, least square methods, support vector machines, empirical orthogonal functions (Zorita and von Storch, 1999), etc., fall in this category.

Weather typing involves grouping local meteorological variables with respect to different classes of atmospheric circulation. Future regional climate scenarios are constructed either by resampling from the observed variable distribution or by first generating synthetic sequences of weather patterns using Monte Carlo techniques and resampling from the generated data. The relative frequencies of the weather classes are weighted to derive the mean or frequency distribution of the local climate. Climate change is then determined from the changes of the frequency of the weather classes.

Stochastic weather generators simulate weather data to assist in the formulation of water resource management policies. They are essentially complex random number generators, which can be used to produce a synthetic series of data. This allows the researcher to account for natural variability when predicting the effects of climate change. Weather generators have an advantage over other downscaling methods because by producing long duration rainfall series, it is possible to examine rare events and extremes in the river basin (Brissette et al., 2007; Diaz-Nieto and Wilby, 2005; Wilks and Wilby, 1999). The underlying assumption of weather generator is that the past (control experiment) would be a representative of the future. It is, however, difficult to guarantee that the statistical relationship derived from current climate will remain same for future in the presence of climate change (Hewitson and Crane, 1996; Schulze, 1997; Joubert and Hewitson, 1997).

Parametric, empirical or semi-parametric, and non-parametric (Brissette et al., 2007) weather generators are commonly used by the scientific community. In most parametric weather generators, a Markov chain is used to determine the probability of a wet or dry day and a probability distribution is assumed to determine the amount of precipitation

(Kuchar, 2004; Hanson and Johnson, 1998). Most of the parametric weather generators are extensions of Richardson's WGEN, which was developed in 1981 (Richardson, 1981). Some examples of the parametric weather generators successfully employed using the Richardson approach are CLIGEN, WGENK, GEM, WXGEN, and SIMMENTO (Kuchar, 2004; Schoof et al., 2005; Hanson and Johnson, 1998; Soltani and Hoogenboom, 2003). Hanson and Johnson (1998) compared outputs from GEM to historical data using the means and standard deviations. Results from that study showed that simulated total precipitation values were significantly underestimated for some months, and annual precipitation values were considerably less than the historical record (Hanson and Johnson, 1998). A study employing the SIMMENTO weather generator found that the variability (standard deviations) of wet fractions and amounts were significantly overestimated by the synthetic historical series (Elshamy et al., 2006). A major drawback of the parametric approach is that the Markov chain does not take into account the previous days' weather. As a result of this, the rare events, such as droughts or wet spells are not adequately produced (Sharif and Burn, 2007; Semenov and Barrow, 1997; Dibike and Coulibaly, 2005). Another limitation of the parametric weather generators is that an assumption must be made about the probability distribution of precipitation amounts, and different distributions do not give similar results (Sharif and Burn, 2007). Furthermore, the weather generators cannot be easily transferred to other basins as their underlying probability assumptions would change (Sharif and Burn, 2006). The computational effort is also significantly higher than other methods since many parameters must be estimated and statistically verified (Mehrotra et al., 2006). Parametric

weather generators are less easily applied to multiple sites as simulations occur independently and thus spatial correlations would have to be assumed.

Semi-Parametric or Empirical weather generators include LARS-WG and the Wilks model, SDSM (Semenov and Barrow, 1997; Wilks and Wilby, 1999). LARS-WG differs from the parametric approaches described above because it employs a series-approach in which the wet and dry spells are determined by taking into account the observed values and assuming mixed-exponential distributions for dry/wet series as well as precipitation amounts (Semenov and Barrow, 1997). The wet/dry day status is first chosen, and then the amount is chosen conditional on the status. As such, the LARS-WG was able to satisfactorily reproduce wet and dry spells, unlike the parametric weather generators (Dibike and Coulibaly, 2005). Wilks (1998) improved on the parametric models of Richardson (1981) by introducing Markov-chains of higher order that have a better "memory" of the preceding weather. The Richardson (1981) model was further extended for multi-site applications by using a collection of single site models in which a conditional probability distribution is specified and thus spatially correlated random numbers can be generated (Mehrotra, 2006; Wilks, 1998). A drawback to these empirical approaches is that there is still a subjective assumption about the type of probability distribution for precipitation amounts and spell lengths, and the spatial correlation structure is empirically estimated for use with multiple sites.

Non-parametric weather generators are computationally simple and do not require any statistical assumptions to be made. They work by using a nearest-neighbor resampling procedure known as the K-NN approach (Sharif and Burn, 2007; Brandsma and Buishand, 1998; Beersma et al., 2002; Yates et al., 2003). The nearest neighbor algorithm works by searching the days in the historical record that have similar characteristics to those of the previously simulated day, and then randomly selecting one of these as the simulated value for the next day (Beersma et al., 2002). This approach is easily used in multi-site studies because the values are simulated concurrently, thus spatial correlation is preserved (Mehrotra et al. 2006). The K-NN algorithm has been successfully used for multi-site hydrological impact assessments in the Rhine Basin, accurately preserving spatial correlation and climatic variability (Beersma et al., 2002; Brandsma and Buishand, 1998). Apipattanavis et al. (2007) compared a K-NN to a semiparametric weather generator. Box plots of wet-spell lengths showed that for some months the semi-parametric model could not reproduce maximum wet spell lengths, and average spell lengths were underestimated for the traditional K-NN model. A major limitation to the K-NN approach is that the values are merely reshuffled, thus no new values are produced (Sharif and Burn, 2007). Climatic extremes are essential in predicting flooding events in response to climate change, thus Sharif and Burn (2007) modified the K-NN algorithm to produce unprecedented precipitation amounts by introducing a perturbation component in which a random component is added to the resampled data points (Sharif and Burn, 2007). Monthly total precipitation and total monthly wet day box plots were used to evaluate the performance of the Modified K-NN algorithm. The algorithm was able to satisfactorily reproduce the statistics of the original dataset while adding variability, which is crucial in hydrologic impact assessments (Sharif and Burn, 2007). Prodanovic and Simonovic (2006) altered the Modified K-NN algorithm of Sharif and Burn (2007) to account for the leap year. In order to allow for more variables for an improved selection of nearest neighbor, principal components are

added in the weather generator (WG-PCA). With the inclusion of more variables and perturbations, the updated model is expected to more accurately define both present day climate conditions and also to produce estimates of future climate scenarios.

However, studies have indicated that the task of downscaling can sometimes become challenging due to the absence of proper station measurements. Gridded databases, such as the National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR) Global Reanalysis – NNGR (Kalnay et al., 1996) and the North American Regional Reanalysis – NARR (Mesinger et al., 2006) can be viable alternatives for alleviating these limitations of missing data and spatial bias resulting from uneven and unrepresentative spatial modelling (Robeson and Ensor, 2006; Ensor and Robeson, 2008). The reanalysis data are advantageous in impact studies because they are based on the AOGCMs with a fixed dynamic core, physical parameterizations and data assimilation systems (Castro et al., 2007).

Global (NNGR) and regional (NARR) reanalysis databases are also gaining use in uncertainty assessment studies. In many of their applications, however, the NNGR resolution (250 km  $\times$  250 km) is not satisfactory, especially in regions with a complex topography (Choi et al., 2009; Tolika et al, 2006; Rusticucci and Kousky, 2002; Haberlandt and Kite, 1998; Castro et al., 2007). The NARR dataset (Mesinger et al., 2006) is a major improvement upon the global reanalysis datasets in both resolution and accuracy. Literature related to an inter-comparison between the global and regional datasets (Nigam and Ruiz-Barradas, 2006; Woo and Thorne, 2006; Castro et al., 2007; Choi et al., 2007 and 2009) shows better agreement of NARR data. More recently, Solaiman and Simonovic (2010a) conducted a rigorous assessment of the NARR and NNGR database for application in the Upper Thames river basin (Ontario, Canada), for hydrological modeling and/or climate change impact studies.

### <span id="page-20-0"></span>**2.2 Multi-Model Ensembles for Uncertainty Research**

In most of the climate change impact assessment studies, single AOGCMs have been used for predicting future climate. It is well understood that in the current context of huge uncertainties, the utilization of a single AOGCM may only represent a single realization out of a multiplicity of possible realizations, and therefore cannot be representative of the future. So, for a comprehensive assessment of future changes in climate conditions, it is important to use collective information by utilizing all available models and by synthesizing the projections and uncertainties in a probabilistic manner.

Studies that used multiple climate model information are, however, cannot be found in abundance. Of the literatures available, one of the common approaches is the use of reliability estimates to multi-model ensembles. The earliest research, to our knowledge, to consider a multimodel ensemble approach was conducted by Raisanen and Palmer (2001), who treated the ensemble members as equally probable realizations and determined probabilities of climate change by computing the fraction of ensemble members in which the differential properties of models, such as bias and rate of convergence, were disregarded.

Giorgi and Mearns (2003) confronted the approach undertaken in Raisanen and Palmer by introducing the "Reliability Ensemble Averaging (REA)" technique, which considered the reliability-based likelihood of realization by models to calculate the

probability of regional temperature and precipitation change. They found this technique to be more flexible in the assessment of risk and cost in regional climate change studies.

Tebaldi et al. (2004; 2005) used Bayesian statistics to estimate a distribution of future climates through the combination of past observational data and the corresponding AOGCM simulated climates. This technique was motivated by the assumption that an AOGCM ensemble represents a "sample of the full potential climate model space compatible with the observed climate using probability distributions (PDFs)" at a regional scale.

Recently, Smith et al. (2009) extended the work of Tebaldi et al. by introducing the univariate approach to consider one region at a time. They are still using a multivariate approach, including cross validation, to confirm the resemblance of the Bayesian predictive distributions. Other literature on Bayesian methods in multi-model ensembles includes work from Allan et al. [2000], Benestad [2004], Stone and Allan [2005], and Jackson et al. [2004].

Another class of new but promising uncertainty estimation methods incorporates the downscaling of AOGCM scenarios and quantifying uncertainties by separately weighting outputs from different AOGCMs in each time step based on their performances. The results can be presented in a probabilistic framework. Wilby and Harris [2006] developed a probabilistic framework to combine information from four AOGCMs, two greenhouse scenarios where the AOGCMs were weighted to an index of reliability for downscaled effective rainfall. A Monte Carlo approach was adopted to explore components of uncertainty affecting projections for the river Thames for 2080s. The resulting cumulative

distribution functions appeared to be most sensitive to uncertainty in (i) the selection of climate change scenarios, and (ii) the downscaling of different AOGCMs.

Ghosh and Mujumdar (2007) used NNGR to develop a methodology capable of assessing AOGCM uncertainty due to different AOGCMs by considering different probability density functions for each time step. They used the information on uncertainty in examining future drought scenarios in a nonparametric manner. Samples of drought indicators were generated with results from downscaled precipitation using a statistical regression approach from available AOGCMs and scenarios. The severity of droughts was presented in a nonparametric kernel estimation and orthonormal approach.

The implications of uncertainties in estimating the severity of extreme precipitation events is an area of research not yet fully explored. The present study aims to compare the uncertainties of precipitation change on a watershed scale by using two very different methods: the Bayesian Reliability Ensemble Average (BA-REA) by Tebaldi et al. (2004; 2005) and the nonparametric kernel estimator. A classification scheme for investigating the severity level of extreme precipitation indices is addressed. Finally, the nonparametric data driven kernel density estimation methods are investigated to quantify uncertainties associated with AOGCM and scenario outputs for extreme precipitation events.

# **3. Methodology**

### <span id="page-23-1"></span><span id="page-23-0"></span>**3.1 Study Area**

The Upper Thames River (UTR) basin (Figure 1)  $(42^035'24''N, 81^08'24''W)$ , located in Southwestern Ontario, Canada, is a  $3,500 \text{ km}^2$  area nested between the Great Lakes of Huron and Erie. The basin often experiences major hydrologic hazards, such as floods and droughts. The basin has a well documented history of flooding events dating back to the 1700s (Prodanovic and Simonovic 2006).



<span id="page-23-2"></span>Figure 1: The Upper Thames River Basin

High flows mostly take place in early March after snowmelt, and again in July and August as a result of summer storms. Khaliq et al (2008) reported that in the Canadian regime, low flow conditions follow a seasonal behavioral pattern: summer low flow occurs between June to November and winter low flow during the December and May periods. The UTR basin experiences frequent low flow conditions between June and September (Prodanovic and Simonovic 2006).

The population of the basin is approximately 450,000 (2006), of which 350,000 are the residents of the City of London. The Thames river basin consists of two majors tributaries of the river Thames: the North Branch  $(1,750 \text{ km}^2)$ , flowing southward through Mitchell and St. Mary's, and eventually into London, and the South Branch  $(1,360 \text{ km}^2)$ , flowing through Woodstock, Ingersoll, and east London. The Upper Thames River basin receives about 1,000 mm of annual precipitation, 60% of which is lost through evaporation and/or evapotranspiration, stored in ponds and wetlands, or recharged as groundwater (Prodanovic and Simonovic 2006).

### <span id="page-24-0"></span>**3.2 Database**

### <span id="page-24-1"></span>**3.2.1 Selection of Predictors**

Daily precipitation and temperature are the most important atmospheric forcing parameters required for any hydrologic impact study for a larger river basin (Salathe Jr., 2003). However, climate models do not resolve important mesoscale and surface features that control precipitation in an area. The choice of appropriate predictors or characteristics from the large-scale atmospheric circulation is one of the most important steps in downscaling. Rainfall can be related to air mass transport and thus related to

atmospheric circulation, which is a consequence of pressure differences and anomalies (Bardossy, 1997). Mean sea level pressure is the basis of derived variables such as surface vorticity, airflow strength, meridional and zonal flow components and divergence (Wilby and Wigley, 2000). Specific humidity is recognized as significant for AOGCM precipitation schemes (Hennessy et al., 1997). Considering all the above factors, predictor variables mentioned in Table 1 are initially chosen to generate precipitation in this study.

<span id="page-25-0"></span>

<b>Predictors</b>	<b>Abbreviations</b>
Precipitation (mm/day)	Precip
Maximum temperature $({}^{0}C)$	Tmax
Minimum temperature $(^{0}C)$	Tmin
Mean sea level pressure (Pa)	<b>MSL</b>
Specific humidity $(Kg/Kg)$	<b>SPFH</b>
Zonal (eastward) wind velocity component $(m/s)$ at 10 m	UGRD
Meridional (northward) wind velocity component $(m/s)$ at 10 m	VGRD

Table 1: Definition of Predictor Variables

Daily observed precipitation (precip), maximum and minimum temperature (Tmax and Tmin) data from 22 stations covering the UTR basin for the period of 1979-2005 is collected from Environment Canada [\(http://www.climate.weatheroffice.gc.ca/climateData/canada\\_e.html\)](http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html) (Table 2).

The rest of the atmospheric variables are collected from the NARR reanalysis dataset for the period of 1979 – 2005. Precipitation values less than 0.5 mm day<sup>-1</sup> are considered zero as suggested by Reid et al. (2001) and Choi et al. (2007). NARR data for this study has been made available through the Data Access Integration of the Canadian Climate Change Scenarios Network of Environment Canada.

<span id="page-26-0"></span>

Table 2: Weather Stations used for Uncertainty Estimation

Data source: National Climate Data and Information Archive of Environment Canada [\(http://climate.weatheroffice.ec.gc.ca/climateData/canada\\_e.html,](http://climate.weatheroffice.ec.gc.ca/climateData/canada_e.html) Retrieved 14/11/2007)

### <span id="page-27-0"></span>**3.2.2 Future Climate Change Scenarios**

The Canadian Climate Change Scenarios Network (CCCSN) provides access to several AOGCM models and emissions scenarios. The website allows the user to specify the range of geographical co-ordinates required, as well as the climatic variable and time period of interest. For the purpose of this study, the time slices collected were 1960- 1990 (baseline), 2011-2040 (2020's) and 2071-2100 (2080s). Seven variables were chosen: minimum temperature, maximum temperature, precipitation, specific humidity, northward wind component, southward wind component and mean sea level pressure. Six AOGCM models were collected, each with two to three emissions scenarios, as specified by the IPCC's Special Report on Emissions Scenarios (Nakicenovic et al, 2000). Full descriptions of the emissions scenarios can be found in Appendix A. Table 3 lists the AOGCM's along with the emissions scenarios available and their origin. Appendix B provides descriptions of each AOGCM.

Both NARR and the AOGCM datasets are processed to conform to the station's grid points.

#### <span id="page-27-1"></span>**3.3 Multi-Model Uncertainty Estimation Methods**

Two approaches based on fundamentally different assumptions are applied to estimate uncertainty in climate model projections of future precipitation under different forcing scenarios. First, a Bayesian statistics approach is applied to estimate a distribution of future climates from the combination of past observed and corresponding AOGCMsimulated data. Next, a methodology combining statistical downscaling using a PCAbased weather generator approach and nonparametric kernel density estimation technique is developed to quantify the uncertainties from AOGCMs. The difference between these

two approaches lies in the fact that the BA-REA method combines uncertainties from different AOGCMs based on its mean bias, so a single weight for different models is present; whereas the nonparametric kernel estimator is capable of providing weights for any point of interest.

			Atmospheric	
<b>GCM Models</b>	<b>Sponsors, Country</b>	<b>SRES</b> <b>Scenarios</b>	<b>Resolution</b>	
				Long
CGCM3T47, 2005	<b>Canadian Centre for Climate</b>	A1B, A2, B1	$3.75^\circ$	$3.75^\circ$
CGCM3T63, 2005	Modelling and Analysis, Canada	A1B, A2, B1	$2.81^\circ$	$2.81^\circ$
	Commonwealth Scientific and			
CSIROMK3.5,	<b>Industrial Research Organization</b>	A2, B1	$1.875^\circ$	$1.875^\circ$
2001	(CISRO) Atmospheric Research,			
	Australia			
	National Aeronautics and Space			
	Administration (NASA)/ Goddard		$3^\circ$	$4^\circ$
GISSAOM, 2004	Institute for Space Studies (GISS),	A1B, B1		
	<b>USA</b>			
MIROC3.2HIRES,	Centre for Climate System	A1B, B1	$1.125^\circ$	$1.125^{\circ}$
2004	Research (University of Tokyo),			
	National Institute for			
MIROC3.2MEDR	<b>Environmental Studies, and Frontier</b>		$2.8^\circ$	$2.8^\circ$
ES, 2004	Research Centre for Global Change	A1B, A2, B1		
	(JAMSTEC), Japan			

<span id="page-28-0"></span>Table 3: AOGCM Models and Emission Scenarios used for Uncertainty Estimation

Data source: Canadian Climate Change Scenario Network Website, [\(http://cccsn.ca/?page=dd](http://cccsn.ca/?page=dd-gcm)[gcm,](http://cccsn.ca/?page=dd-gcm) Retrieved 9/20/2008)

#### <span id="page-29-0"></span>**3.3.1 The Bayesian Reliability Ensemble Average (BA-REA) Method**

The methodology developed by Tebaldi et al. (2004; 2005) consists of a formal Bayesian implementation and extension of the reliability ensemble averaging (REA) approach of Giorgi and Mearns (2002; 2003). It combines data from observations and a multi-model ensemble of AOGCMs to compute PDFs of future temperature and precipitation change over large regions under different forcing scenarios. Three components constitute the model structure: prior, likelihood, and posterior. The assumption is that the variability of present and future climate from different AOGCMs are random quantities and have different variances which are priori unknown. Although uninformative prior distribution has been chosen, both model-generated and observational data are applied for calculating meaningful posterior distributions.

The choice of an uninformative prior distribution has the advantage of selecting parameter estimates similar to non-Bayesian approaches, such as maximum likelihood. In cases where there is no sufficient agreement between experts to determine a specific prior and no data from previous studies could be incorporated, ( a situation similar to wide range of future climate scenarios), selection of an uninformative prior is justified.

The choice of the likelihood or distribution of the data as a function of any random parameters constitutes the second parameter. The AOGCM responses are assumed to have a symmetric distribution whose center is the 'true value' of the variable of interest, but maintains an individual variability to be a measure of how well each AOGCM depicts the natural variability.

The prior and posterior distributions are combined into a joint posterior distribution using the Bayes' theorem. The empirical estimate of the posterior distribution is obtained using the Markov Chain Monte Carlo (MCMC) simulation by simulating samples from the posterior distribution.

#### *Likelihoods*

The likelihoods for the observations of current mean precipitation  $(X_0)$ , simulations of present  $(X_i)$  and future  $(Y_i)$  mean precipitation by the  $i<sup>th</sup>$  model can be written as:

 $X_0 \sim N[\mu, (\lambda_0)^{-1}]$ , the likelihood of the observations of current climate

Or

alternately,

$$
X_o = \mu + \epsilon_o
$$
  

$$
X_i \sim N[\mu, (\lambda_i)^{-1}]
$$
 (3.1)

Or alternately,

 $X_i = \mu + \eta_i$  (assuming a common Gaussian distribution for the error terms)

$$
Y_i \sim N[\nu + (\theta \lambda_i)^{-1}]
$$

Or alternately,

 $Y_i = v + \beta_x (X_i - \mu) + \frac{\xi}{2}$  $\sqrt{\theta}$  $\frac{d}{d\sigma}$  (assuming a common Gaussian distribution for the error

terms)

Where,  $\mu$  and  $\nu$  are random variables presenting the (unknown) true present and future mean precipitations respectively.  $(\lambda_0)^{-1}$  and  $(\lambda_i)^{-1}$  can be considered as a measure of *i*<sup>th</sup> AOGCM precision, and the estimates of natural observed variability, which depends on the season, region and time average of the observation. The parameter  $\lambda_0$  is fixed as the reciprocal of the squared value of the standard deviation of the observations. Random variable  $\theta$  allows for the possibility of the future and present precipitations having different variances by a multiplicative factor and is common to all AOGCMs.

The alternate forms of equation 3.1 links  $X_i$  and  $Y_i$  through a linear regression equation equivalent to assuming that  $(X_i, Y_i)$  are jointly normal when parameter values are given and the correlation coefficient is relaxed to vary between  $-1$  and  $+1$ . For  $\beta_x \neq 0$ , the modified equation for  $Y_i$  will create a direct (if positive) or inverse (if negative) relation between  $X_i - \mu$  and  $Y_i - \nu$ . The value of  $\beta_x$  is also significant for representing the correlation: a value of 1 denotes the conditional independence of the signal of precipitation change produced by any AOGCM and  $X_i - \mu$ , the model bias for current precipitation. Values greater or smaller than 1 imply a positive or negative correlation between them.

### *Prior Distribution*

The prior distributions are chosen for the following precision parameters:

 $\lambda_i = 1,2,3,...$  6 have Gamma prior densities (Ga(a,b)):

$$
\frac{b^a}{\Gamma(a)} \lambda_i^{a-1} exp^{-bx} \tag{3.2}
$$

Where, a and b are known. Similarly for  $\theta$ , c, d are assumed to be known. For the model,  $a = b = c = d = 0.001$  are chosen.

The true climate means μ and ν for present and future precipitation have uniform prior densities so that even in case of improper priors (do not integrate to one) they are assumed to have a proper posterior density function.

### *Posterior Distribution*

Bayes' theorem is applied to the likelihood and propers. The resulting joint posterior distribution is given by:

$$
\Pi_{i=1}^{15} \left[ \lambda_i^{a-1} e^{-b\lambda_i} \lambda_i \theta^{1/2} exp\left\{ -\frac{\lambda_i}{2} (X_i - \mu)^2 + \theta (Y_i - \nu)^2 \right\} \right] \cdot \theta^{c-1} e^{-d\theta} \cdot exp\left\{ -\frac{\lambda_0}{2} (X_0 - \mu)^2 \right\}
$$
\n
$$
(3.3)
$$

The above distribution does not represent any specific known parameter family. The posterior distribution fixes the parameters and considers a conditional posterior for others to synthesize the data and the prior assumptions. For example, the distribution of  $\mu$  for fixing all other parameters is Gaussian with

Mean:

$$
\tilde{\mu} = \frac{\left(\sum_{i=0}^{15} \lambda_i X_i\right)}{\left(\sum_{i=0}^{15} \lambda_i\right)} \tag{3.4}
$$

Variance:

$$
\left(\sum_{i=0}^{15} \lambda_i\right)^{-1} \tag{3.5}
$$

Similarly, the conditional distribution of  $\nu$  is Gaussian with

Mean:

$$
\tilde{\nu} = \frac{\left(\sum_{i=0}^{15} \lambda_i Y_i\right)}{\left(\sum_{i=0}^{15} \lambda_i\right)} \tag{3.6}
$$

Variance:

$$
\left(\theta \sum_{i=0}^{15} \lambda_i\right)^{-1} \tag{3.7}
$$

Equations 3.4 and 3.6 are comparable to the REA results as the weighted means of the 15 different AOGCMs with their scenarios and the observation with weights  $\lambda_1$ , ... ...  $\lambda_{15}$ ,  $\lambda_0$ , respectively. These weights are derived by assuming parameters with random quantities and hence can be used for uncertainty estimation. This uncertainty will inflate the width of the posterior distributions of  $\nu$ ,  $\mu$  and also the precipitation change,  $\Delta P$ .

The mean of the posterior distribution of the  $\lambda_i s$  for  $i = 1, 2, \dots 15$  is approximated as:

$$
E(\lambda_i | \{X_0, \dots \dots, X_{15}, Y_1, \dots, Y_{15}\}) \approx \frac{a+1}{b + \frac{1}{2}((X_i - \tilde{\mu})^2 + \theta(Y_i - \tilde{\nu})^2)}
$$
(3.8)

 Equation 3.8 expresses how the bias and convergence criteria are built into the model implicitly since the precision parameter or the weights  $\lambda_i$  for each AOGCM are large provided the bias  $|X_i - \mu|$  and convergence  $|Y_i - \nu|$  or the distance of the *i*<sup>th</sup> model future response from the overall average response are small. So the results are strictly constrained by their convergence into future projections determined by the weighted

ensemble of mean. For this study,  $a=b=0.001$  is chosen as per Tebaldi et al. (2004, 2005) to ensure that the contribution of the prior assumption to equation 3.8 is negligible.

Using the approximation similar to equation 3.1 the posterior mean can be written as:

$$
E(\lambda_i | \{X_0, \dots, X_{15}, Y_1, \dots, Y_{15}\})
$$
  

$$
\approx \frac{a+1}{b + \frac{1}{2} ((X_i - \tilde{\mu})^2 + \theta (Y_i - \tilde{v} - \beta_x (X_i - \tilde{\mu}))^2)}
$$
(3.9)

 Next, the marginal posterior distribution is derived using the MCMC approach. A large number of sample values are generated by applying the Gibbs Sampler using equation 3.3 for all parameters.

#### *MCMC Approach: The Gibbs Sampler*

The joint posterior distribution derived from assuming different distributions such as Gaussian, Uniform and Gamma in different stages, does not represent any known parametric family of distributions. Because they are conjugate, they allow for a closedform deviation of all full conditional distributions.

Auxiliary randomization parameters  $s_i$  and  $t_i$ ,  $i = 1, 2, 3, \dots, 15$  are used to ensure an efficient simulation from student's t distribution within the Gibbs sampler. Fixing  $s_i = t_i = 1, \beta_x = 0$ , returns the full conditionals to the prior parameters.

$$
\lambda_i | \dots | \dots | \dots \in Ga\left(a+1, b+\frac{s_i}{2}(X_i-\mu)^2+\frac{\theta t_i}{2}\{Y_i-\nu-\beta_x(X_i-\mu)\}^2\right) \tag{3.10}
$$

$$
s_i | \dots | \dots | \dots | \dots | \cdot Ga\left(\frac{\phi + 1}{2}, \frac{\phi + \lambda_i (X_i - \mu)^2}{2}\right)
$$
 (3.11)

$$
t_i | \dots | \dots | \dots | \dots | \cdot Ga\left(\frac{\phi + 1}{2}, \frac{\phi + \theta \lambda_i \{Y_i - \nu - \beta_x (X_i - \mu)\}^2}{2}\right)
$$
 (3.12)

$$
\mu | \dots | \dots | \dots | \dots | \cdot N \left( \tilde{\mu}, \left( \sum s_i \lambda_i + \theta \beta_x^2 \sum t_i \lambda_i + \lambda_0 \right)^{-1} \right) \tag{3.13}
$$

$$
\nu | \dots | \dots | \dots | \dots | \sim N \left( \tilde{\nu}, \left( \theta \sum t_i \lambda_i \right)^{-1} \right) \tag{3.14}
$$

$$
\beta_x | \dots | \dots | \dots | \dots | \cdot N(\widetilde{\beta_x}, (\theta \sum t_i \lambda_i + (X_i - \mu)^2)^{-1}) \tag{3.15}
$$

$$
\theta | \dots | \dots | \dots | \dots | \dots \in Ga(c + \frac{N}{2}, d + \frac{1}{2} \sum t_i \lambda_i \{Y_i - \nu - \beta_x (X_i - \mu)\}^2)
$$
(3.16)

Simplifying,

$$
\tilde{\mu} = \frac{\sum s_i \lambda_i X_i - \theta \beta_x \sum \lambda_i t_i (Y_i - \nu - \beta_x X_i) + \lambda_0 X_0}{\sum s_i \lambda_i + \theta \beta_x^2 \sum \lambda_i t_i + \lambda_0}
$$
\n(3.17)

$$
\tilde{v} = \frac{\sum t_i \lambda_i \{Y_i - \beta_x (X_i - \mu)\}}{\sum t_i \lambda_i}
$$
\n(3.18)
$$
\widetilde{\beta_x} = \frac{\sum t_i \lambda_i \{ (Y_i - \nu) (X_i - \mu) \}}{\sum t_i \lambda_i (X_i - \mu)^2}
$$
\n(3.19)

From this sequence of full conditional distributions, the Gibbs sampler is coded to simulate iteratively. After a series of iterations, the MCMC process ignores the arbitrary set of initial values for parameters. Values sampled at each iteration represents a draw from the joint posterior distribution of interest, and any summary statistic can be computed to a degree of approximation that is a direct function of the number of the sampled values available, and an inverse function of the correlation between successive samples.

The reliability of any AOGCM is measured by two criteria to form the shape of the posterior distribution as a consequence of assumptions formulated in the statistical model: mean bias of present climate and rate of convergence of the future climate models to weighted ensemble mean.

#### *Model parameters*

For this study, the area averaged precipitation response from all 15 AOGCMs and scenarios, averaged for the London station, is considered to compare with the PDFs generated by the methodology presented in Section 3.3.1.

# *Data and model setup*

To generate PDF of precipitation affected by climate change, simulated present (1961- 1990) and future (2041-2070) precipitation  $(X_i, Y_i)$  are considered for the winter (December-January-February) and summer (June-July-August) seasons. The outputs

from 15 different sets of experiments and from 6 AOGCMs for the two time slices are extracted for the 22 stations and averaged for the London station using the nearest neighbor approach. The natural variability is expressed as the inverse of the variance of observed precipitation for 1961-1990  $(X_0)$ . It is calculated as the inter-annual variance on the basis of the observed record  $(X_0)$ . The computer codes used in this study can be downloaded from the website of the National Centre for Atmospheric Research (*http://www.image.ucar.edu/~nychka/REA/*).

#### **3.3.2 Nonparametric Kernel Estimators**

#### *Downscaling*

Stochastic weather generators simulate weather data to assist in the formulation of water resource management policies. The basic assumption for producing synthetic sequences is that the past will be representative of the future. These sequences are essentially complex random number generators, which can be used to produce a synthetic series of data. This allows the researcher to account for natural variability when predicting the effects of climate change.

In order to reduce multi-dimensionality and collinearity associated with the large number of input variables, a principal component analysis has been integrated within the weather generator. The process requires selecting the appropriate principal components (PCs) that will adequately represent most of the information of the original dataset.

The WG-PCA algorithm with  $p$  variables and  $q$  stations works through the following steps:

1) Regional means of  $p$  variables for all  $q$  stations are calculated for each day of the

observed data:

$$
\bar{X}_t = [\bar{x}_{1,t}, \bar{x}_{2,t}, \dots, \bar{x}_{p,t}] \qquad \forall t = \{1, 2, \dots, T\}
$$
\n(3.20)

Where,

$$
\bar{x}_{i,t} = \frac{1}{q} \sum_{j=1}^{q} x_{i,t}^j \qquad \forall t = \{1, 2, \dots, p\}
$$
\n(3.21)

2) The user-set parameters are as follows: potential neighbors,  $L$  days long where  $L = (w + 1) \times (N - 1)$  for each of p individual variable with N years of historic record, and a temporal window of size  $w$ . The days within the given window are all potential neighbors to the feature vector.  $N$  data which correspond to the current day are deleted from the potential neighbors so the value of the current day is not repeated.

3) Regional means of the potential neighbors are calculated for each day at all  $q$ stations.

4) A covariance matrix,  $C_t$  of size  $L \times p$  is computed for day t.

5) The first time step value is randomly selected for each of  $p$  variables from all current day values in the historic record.

6) Next, using the variance explained by the first principal component, Mahalanobis distance is calculated with equation 3.33.

$$
d_k = \sqrt{\frac{(PC_t - PC_k)^2}{Var(PC)}} \qquad \forall k = \{1, 2, ..., K\}
$$
 (3.22)

where,

 $PC<sub>t</sub>$  is the value of the current day;

 $PC_k$  is the nearest neighbor transferred by the Eigen vector.

 $Var(PC)$  is the variance of the first principle component is for all Knearest neighbors.

7) The selection of the number of nearest neighbors,  $K$ , out of  $L$  potential values using  $K = \sqrt{L}$ .

8) The Mahalanobis distance  $d_k$  is put in order of smallest to largest, and the first  $K$ neighbors in the sorted list are selected (the *K* Nearest Neighbors). A discrete probability distribution is used that weights closer neighbors highest in order to resample out of the set of K neighbors. Using equations  $3.34$  and  $3.35$ , the weights, w, are calculated for each  $k$  neighbor.

$$
w_k = \frac{\frac{1}{k}}{\sum_{i=1}^k \frac{1}{i}}
$$
  $\forall k = \{1, 2, \dots, K\}$  (3.23)

Cumulative probabilities,  $p_i$ , are given by:

$$
p_j = \sum_{i=1}^j w_i
$$
 (3.24)

9) A random number  $u(0,1)$  is generated and compared to the cumulative probability calculated above in order to select the current day's nearest neighbor. If  $p_1 < u < p_k$ , the day *j* for which *u* is closest to  $p_i$  is selected. However, if  $p_1 > u$ , then the day that corresponds to  $d_1$  is chosen. For  $u = p_k$ , the day that corresponds to day  $d_k$  is selected. Upon selecting the nearest neighbor, the K-NN algorithm chooses the weather of the selected day for all stations in order to preserve spatial correlation in the data (Eum et al,

2009).

10) In order to generate values outside the observed range, perturbation is used. A conditional standard deviation  $\sigma$  for K nearest neighbors is estimated. For choosing the optimal bandwidth of a Gaussian distribution function that minimizes the asymptotic mean integrated square error (AMISE), Sharma et al. (1997) reduced Silverman's (Silverman 1986, pp. 86-87) equation of optimal bandwidth into the following form for a univariate case:

$$
\lambda = 1.06 \sigma K^{-\frac{1}{5}} \tag{3.25}
$$

Using the mean value of the weather variable  $x_{i,t}^j$  obtained in step 9 and variance  $(\lambda \sigma_i^j)^2$ , a new value  $y_i^j$  can be achieved through perturbation (Sharma et al. 1997).

$$
y_{i,t}^j = x_{i,t}^j + \lambda \sigma_i^j Z_t \tag{3.26}
$$

where,

 $z_t$  is a random variable, distributed normally (zero mean, unit variance) for day t. Negative values are prevented from being produced for precipitation by employing a largest acceptable bandwidth (Sharma and O'Neil, 2002):

$$
\lambda_a = \frac{x_{*,t}^j}{1.55\sigma_*^j} \tag{3.27}
$$

where,

\* refers to precipitation.

If again a negative value is returned, a new value for  $z_t$  is generated (Sharif and Burn, 2006).

### *Kernel based Nonparametric Uncertainty Estimator*

A practical approach to deal with AOGCM and scenario uncertainties initiating from inadequate information and incomplete knowledge should: (1) be robust with respect to model choice; (2) be statistically consistent in a uniform application across different area scales such as global, regional or local/watershed scales; (3) be flexible enough to deal with the variety of data; (4) obtain the maximum information from the sample; and (5) lead to consistent results. Most parametric methods do not meet all these requirements.

The Probability Density Function (PDF) is commonly used to describe the nature of data. In applications an estimate of the unknown  $PDF = f()$  based on random sample  $x_1, x_2, \dots, x_n$  from  $f()$  is calculated in the form of  $\widehat{PDF} = \widehat{f}(.)$ . Probability distribution functions estimated by any nonparametric method without prior assumptions are suitable for quantifying AOGCM and scenario uncertainties. Several approaches, such as kernel methods, orthogonal series methods, penalized-likelihood methods, k-nearest neighbor methods, Bayesian-spline methods, and maximum-likelihood or histogram like methods, are used throughout the the relevant literature (Adamowski, 1985).

A Kernel density estimation method has been widely used as a viable and flexible alternative to parametric methods in hydrology (Sharma et al., 1997; Lall, 1995), flood frequency analysis (Lall et al., 1993; Adamowski, 1985), and precipitation resampling (Lall et al., 1996) for estimating a probability density function.

A kernel density estimate is formed through the convolution of kernels or weight functions centered at the empirical frequency distribution of the data. A kernel density estimator involves the use of the kernel function  $(K(x))$  defined by:

$$
\int_{-\infty}^{\infty} K(x)dx = 1
$$
\n(3.28)

A PDF can thus be used as a kernel function. The Parzen-Rosenbalt kernel density estimate  $f_n(x)$  at x, from a sample of  $\{x_1, \ldots, x_i, \ldots, x_n\}$  of sample size *n* is given by:

$$
\widehat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h\left(\frac{x - x_i}{h}\right)
$$
\n(3.29)

Where  $t = \left(\frac{x}{x}\right)$  $\left(\frac{-x_i}{h}\right)$  and  $k_h(t)$  is a weight or kernel function required to satisfy criteria such as symmetry, finite variance, and integrates to unity. Successful application of any kernel density estimation depends more on the choice of the smoothing parameter or bandwidth *(h)* than it does on the type of kernel function *K(.)*, to a lesser extent.

The bandwidth for kernel estimation may be evaluated by minimizing the deviation of the estimated PDF from the actual one. Assuming a normal distribution for the bandwidth estimation, the optimal bandwidth for a normal kernel can be given by (Polansky and Baker, 2000):

$$
h_n = (1.587)\hat{\sigma}n^{-\frac{1}{3}}\tag{3.30}
$$

Where  $\hat{\sigma}$  is the sample standard deviation measured by Silverman [1986]:

$$
\hat{\sigma} = \min \left\{ S, \frac{IQR}{1.349} \right\} \tag{3.31}
$$

Where *S* is the sample standard deviation and *IQR* is the interquartile range.

This methodology is applied to derive the PDF of the mean monthly precipitation at different time steps.

### *Data preprocessing and experimental setup*

A schematic of estimating the PDFs combining uncertainties using downscaling technique is presented in Figure 2. For this study, daily input variables from NARR, as indicated in Table 2, are collected at the nearest grid points and spatially interpolated to the stations (Table 1) surrounding the Upper Thames River basin.

While the direct downscaling of minimum and maximum temperature has produced good results, precipitation values are not well reproduced directly from AOGCM data



Figure 2: Flow Chart of Uncertainty Estimation using Nonparametric Method

(Brissette et al., 2006). For selection of appropriate conditioning variables, several combinations of predictors, which are listed in Table 3, are used to generate synthetic versions of the historic dataset. A multi-objective Compromise Programming tool is then used to find an optimal set of predictors. Assessment of trade-offs between different combinations of variables (considered as alternatives) is done according to four variability measures (considered as criteria): mean, standard deviation, maximum and minimum values for each month. The rank of each combination is measured by the compromise programming distance metric, which is calculated as the distance from the ideal solution for each alternative. Table 4 presents the ranks obtained for each combination of predictors. It is clear that a combination of all seven predictors is closest to the ideal solution in most months, and is therefore selected for the purposes of further analysis.

<b>Cases</b>		<b>Months</b>										
		$1\quad 2$		$3\quad 4$		$5\,6$		78		9 10	11	12
P, Tmax, Tmin, PRMSL		76		5 6		3 <sub>1</sub>		$1\quad5$		74	6	
P, Tmax, Tmin, PRMSL, SPFH		41		75		4 2		77		6 1	3	7
P, Tmax, Tmin, PRMSL, SPFH,		$5\,2$		2 4 2 4 3 3 1 6							$\boldsymbol{4}$	$\mathbf{2}$
<b>UGRD, VGRD</b>												
P,Tmax,Tmin,PRMSL,UGRD,VGRD		6 4		1 7				75 21 42			5	6
P, Tmax, Tmin, SPFH		3 7		4 1				5 3 6 2	5 3		7	$\overline{4}$
P, Tmax, Tmin, SPFH, UGRD, VGRD		2 3		6 3		$1\quad7$		56		3 5	$\mathcal{D}_{\mathcal{L}}$	5
P, Tmax, Tmin, UGRD, VGRD		1 5		3 2		66		44		2.7		$\mathcal{R}$

Table 4: Rank Table of Different Combinations of Predictors

\* P: Precipitation, Tmax: Maximum temperature, Tmin: Minimum temperature, PRMSL: Mean sea level pressure, SPFH: Specific humidity, UGRD: Eastward wind component, VGRD: Northward wind component

Next, the monthly information from each of the AOGCM emission scenarios (Table 3) is collected for four time slices: 1961-1990, 2011-2040, 2041-2070 and 2071-2100. Because of the limited quality and unavailability of daily inputs from many AOGCMs, monthly inputs should be used. Climate variables taken from the nearest grid points are interpolated to provide a dataset for each of the stations of interest in the same way as the NARRs. In order to generate future climate data, the difference between the base climate and the AOGCM outputs (2041-2070 or 2050s) are computed for all predictors. The change factors are then used to modify the historic dataset collected for each station, thereby creating a future dataset. The differences between current and future climate are used to calculate the monthly change factor and then added to the predictors to generate a modified time series. This modified dataset is used as input into the weather generator to produce synthetic datasets of any length for the time period of interest.

In order to reduce multi-dimensionality and collinearity associated with the large number of input variables, a principal component analysis is integrated with the weather generator. The process requires the selection of the appropriate principal components (PCs) that will adequately represent most information in the original dataset. It is found that the first PC is able to explain over 95% of the variations associated with the inputs. Hence, only the first PC is considered for the weather generator.

The daily future data, downscaled using WG-PCA, are averaged to a monthly value in order to draw a PDF for comparison with the BA-REA approach. The average monthly total values for winter (DJF) and summer (JJA) for each scenario are considered. Values from each AOGCM for any specific year are considered as an independent set of realization and are used to draw PDFs.

# **3.4 Indexing Extreme Precipitation Events**

Simulation of extreme precipitation is dependent on resolution, parameterization and the selected thresholds. Sun et al. (2006) found that most AOGCM models tend to produce light precipitation  $\left($ <10mm day<sup>-1</sup> $\right)$  more often than observed, too few heavy precipitation events and much less precipitation during heavy events  $(>10$  mm day<sup>-1</sup>) (Randall et al., 2007). The situation gets worse in the absence of any extreme precipitation indices. In the IPCC (2007), several indices explaining extreme temperature and precipitation are proposed but most literature reports investigations of percent change in the occurrences of such indices without any acceptable definition of their severity level.

Three precipitation indices have been used for comparing the performance of the AOGCMs in generating extreme precipitation amounts. These indices describe precipitation frequency, intensity and extremes. The highest 5 day precipitation, number of very wet days and the number of heavy precipitation days express extreme features of precipitation. For very wet days, the  $95<sup>th</sup>$  percentile reference value has been obtained from all non-zero total precipitation events for the base climate. Heavy precipitation days are those days that experience more than 10 mm of precipitation.

For Canada, due to large variation of precipitation intensities in various regions, a fixed threshold may not be good to assess the severity level (Vincent and Mekis, 2006). Accordingly, in this study an attempt has been made to classify the severity level of these indices based on percentile values. The percentile method has several advantages. It is simple and computationally inexpensive. It is completely data driven and does not follow any specific distribution, so can be used at any location with different precipitation

patterns. Table 5 presents the classification scheme used for the summer and winter season. They can be easily used to assess the impact of climate change on extreme precipitation events.

Table 5: Classification of Extreme Precipitation Indices based on Percentile Approach

Serial	<b>Description</b>
	$\epsilon$ 25th percentile of 1961-1990 observed precipitation
2	$25th - 50th$ percentile of 1961-1990 observed precipitation
3	50th –75th percentile of 1961-1990 observed precipitation
4	75th – 95th percentile of 1961-1990 observed precipitation
5	>95th percentile of 1961-1990 observed precipitation

# **3.5 Extended Kernel Estimators**

Nonparametric estimators are erroneously considered to be less accurate with small sample sizes (Lall et al., 1993). With the increase in sample size, the choice of estimator selection (parametric or nonparametric) can only be more accurately identified. Nonparamtric kernel estimators based on (i) normal kernel estimator (Silverman, 1986), and (ii) the Orthonormal method (Efromovich, 1999) have been applied by Ghosh and Mujumder (2007) for assessing AOGCM and scenario uncertainties of future droughts. In the present study, the application of a normal kernel estimator is extended with the commonly used bandwidth selection methods for estimating densities and addressing model choice and scenario choice uncertainties.

### **3.5.1 Definition**

The nonparametric kernel density estimation described in section 3.3.2 is based on the conventional method of assuming a normal distribution function for unknown PDFs. Because of an uncertain future climate, it is not justifiable to assume a normal distribution of the PDFs. Allowing an extension for the kernel estimator by replacing the normal bandwidth for a data-driven procedure can better quantify the inherent uncertainties arising from different AOGCMs.

The behavior of the estimator (equation 3.28) may be analyzed mathematically under the assumption that the data sets represent independent realizations from a probability density  $f(x)$ . The basic methodology of the theoretical treatment aims to discuss the closeness of estimator  $\hat{f}$  to the true density, f. Successful application of the estimator depends mostly on the choice of a kernel and a smoothing parameter or bandwidth. the relevant literature shows that the choice of bandwidth is more critical. A change in kernel bandwidth can dramatically change the shape of the kernel estimate (Efromovich, 1999). For each x,  $\hat{f}(x)$  can be thought as a random variable because of its dependence on  $X_1, X_2, \ldots, X_n$ . Except otherwise stated,  $\Sigma$  will refer to a sum for  $i = 1$  to n and  $\int$  to an integral over the range  $(-\infty, \infty)$ .

The discrepancy of the density estimator  $\hat{f}$  from its true density f can be measured by mean square error (MSE):

$$
MSE_x(\hat{f}) = E[(\hat{f}(x) - f(x))^2]
$$
\n(3.32)

By standard elementary properties of mean and variance,

$$
MSE_x(\hat{f}) = \{E[(\hat{f}(x) - f(x)]^2\} + var \,\hat{f}(x) \tag{3.33}
$$

43

The sum of the squared bias and the variance at  $x$ . In many applications a trade-off is applied between the bias and the variance in equation 3.32; the bias can be reduced by increasing the variance and vice versa by adjusting the degree of smoothing. It can be obtained by minimizing the mean integrated squared error (MISE), a widely used measure of global accuracy of  $\hat{f}$  as an estimator of f (Rosenblatt, 1956; Adamowski, 1985; Scott et al., 1981, Jones et al., 1996) and defined as:

$$
MISE\left(\hat{f}\right) = E\int [(\hat{f}(x) - f(x))]^2 dx\tag{3.34}
$$

Or in alternative forms,

$$
MISE(\hat{f}) = \int MSE_x(\hat{f})dx
$$
  
=  $E \int [(\hat{f}(x) - f(x))^2 dx + \int var(\hat{f}) dx$  (3.35)

which gives the  $MISE$  as the sum of the integrated square bias and the integrated variance.

Asymptotic analysis provides a simple way of quantifying how the bandwidth *h* works as a smoothing parameter. Under standard assumptions, MISE is approximated by the asymptotic mean integrated squared error (AIMSE) (Jones et al., 1996):

$$
AMISE(h) = n^{-1}h^{-1}R(K) + h^4R(f^{\prime\prime})\left(\int x^2 K/2\right)^2\tag{3.36}
$$

Where  $R(\varphi) = \int \varphi^2(x) dx$  and  $\int x^2 K = \int x^2 K(x) dx$ , *n* is sample size, *h* is bandwidth. The first term (integrated variance) is large when *h* is too small, and the second term (integrated squared bias) is large when *h* is too large.

The minimizer of  $AIMSE(h)$  is easily calculated as:

$$
AMISE(h) = \left[\frac{R(k)}{nR(f^{''})(\int x^2 K)^2}\right]^{1/5}
$$
\n(3.37)

#### **3.5.2 Methods for Bandwidth Selection**

Data driven estimation methods are broadly classified as first generation and second generation methods by Jones et al (1996).

#### *First Generation Methods*

First generation methods used for the selection of smoothing parameter include those proposed before 1990. These include the rule of thumb, least square cross validation and biased cross validation methods.

The most basic method is the 'rule of thumb' used by Silverman (1986). The idea involves replacing the unknown part of  $h_{AMISE}$ ,  $R(f^{\prime})$ , in equation 3.34 with an estimated value based on a parametric family such as a normal distribution  $N(0, \sigma^2)$ . However, this method is known to provide an over-smoothed function (Terrell and Scott, 1985; Terrell, 1990) and has been proven to be unrealistic in many applications. In the present study,  $h_{ROT}$  is used to denote the bandwidth based on the standard deviation in Silverman (1986).

The idea of 'least squared cross validation', first used by Bowman (1984) and Rudemo (1982), incorporates integrated squared error (ISE) as

$$
ISE(h) = \int (\widehat{f}_h - f)^2 = \int \widehat{f}_h^2 - 2 \int \widehat{f}_h f + \int f^2 \tag{3.38}
$$

 The minimizer of the ISE is the same as the minimizer of the first two terms of the final form. The first term is known while the second term can be estimated by  $-2n^{-1}\sum_{i=1}^n \hat{f}_h(X_i)$ , where  $\hat{f}_i$  is the leave-out kernel density estimator with  $X_i$  removed. The largest minimizer is denoted by  $h_{CV}$  Hall and Marron (1991).

The biased cross validation (BLCV) proposed by Scott and Terrell (1987) seeks to directly minimize the AMISE by estimating the unknown  $R(f^{\prime})$  in equation 3.34. It proceeds by selecting another bandwidth treated as the dummy variable of minimization. The smallest local minimizer of

$$
BLCV(h) = n^{-1}h^{-1}R(K) + h^4 \times \left[ R\left(\hat{f}_a^{\ \ ^{n}} - \frac{R(K^{\ ^{n}})}{mh}\right) \right] \left(\int x^2 K/2\right)^2 \tag{3.39}
$$

is denoted by  $h_{BCV}$ .

#### *Second Generation Method*

 Second generation methods comply with those developed after 1990, such as the solve-the-equation-plug-in approach, the smoothed bootstrap approach, etc. In this study, only the solve-the-equation-plug-in approach is used, and hence is described below.

The main thought behind the 'solve the equation plug in' approach is to plug an estimate of the unknown  $R(f^{\prime})$  in the equation 3.40. The major challenge is to estimate a pilot bandwidth. The 'solve the equation' approach proposed by Hall (1980), Sheather (1983, 1986) and later refined by Sheather and Jones (1991) is used in this study. The smallest bandwidth, *hSJPI,* is considered as the solution of the fixed point equation

$$
h = \left[ \frac{R(K)}{nR(\hat{f}_{g(h)})(\int x^2 K)^2} \right]^{\frac{1}{5}} \tag{3.40}
$$

The major difference between the *BLCV* and *SJPI* approaches lies in the expression of the form  $g(h)$ , which provides a better representation of  $R(f^{\prime})$ . It is done by estimating an analogue of  $h_{AMISE}$  for estimating  $R(f^{\prime\prime})$  by  $R(\hat{f}^{\prime\prime})$ .

The minimizer of the asymptotic mean squared error (AMSE) is expressed as:

$$
g_{AMSE} = C_1 \{ R(f^{'''}) \} C_2(K) n^{\frac{-1}{7}}
$$
\n(3.41)

for suitable functional  $C_1$  and  $C_2$ . The expression of g in terms of h comes from solving the representation of  $h_{AIMSE}$  for n and substituting to get

$$
g(h) = C_3\{R(f''), R(f''')\}C_4(K)h^{\frac{5}{7}}
$$
\n(3.42)

For appropriate functionals  $C_3$ ,  $C_4$ . The unknowns  $R(f'')$  and  $R(f''')$  are estimated by  $R(\hat{f}^{\prime\prime})$  and  $R(\hat{f}^{\prime\prime\prime})$ , with bandwidths chosen by reference to a parametric family, as for  $h_{ROT}.$ 

While many variations have been tested for the treatment of  $R(\hat{f}^{\prime\prime})$  and  $R(\hat{f}^{\prime\prime\prime})$ , the major contribution has been to try to reduce the influence of the normal parametric family even further by using pilot kernel estimates instead of normal interference (Jones et al., 1996). Park and Marron (1992) has shown the improvements in terms of the asymptotic rate of convergence up to a certain point.

# **4. Results and Discussion**

The performances of all methods and comparison results described in Chapter 3 are presented here. First, the BA-REA method and non-parametric weather generator are evaluated for assessing AOGCM uncertainties. The indices for estimating the severity of extreme precipitation events are developed and compared for future climate. Finally, the probabilities of extreme precipitation events are assessed with associated AOGCM and scenario uncertainties.

# **4.1 Fixed Weight (BA-REA) Method**

The performance of the Bayesian reliability method can be assessed by model bias and convergence. Table 6 presents the values of the bias from six different AOGCMs. Bias is calculated as the difference between each AOGCM's response to the present climate and the present climate as generated by the model.

Table 6: Biases from AOGCM Responses to Present Climate (1961-1990) in London

	Model Bias $(\% )$								
<b>Season</b>		CGCM3 CGCM3 CSIRO		GISS <sup></sup>	<b>MIROC</b>	<b>MIROC</b>			
	T47	T63	MK3.5		AOM 3.2HIRES	3.2MEDRES			
Summer	22.50	$-2.12$	6.50		$12.07 - 14.92$	$-14.10$			
Winter	2.18	$-1.68$	11.46	$-0.04$	$-26.24$	$-5.64$			

Figure 3 presents posterior distributions of precipitation change  $\Delta P$  for London during the winter and summer seasons. For purposes of reference, the response of 15 models and the scenarios' individual responses  $Y_i - X_i$ , for  $i = 1, 2, ..., 15$ , are plotted



Figure 3: Posterior Distributions of **DP= ν – μ** in London for Winter and Summer

along the x axis (dots) together with the REA estimate of mean change (triangles). A measure of convergence can be assessed using the relative position of the individual responses. The relative position is used in the identification of the outlier models and the models that reinforce each other. The comparison of densities in Figure 3 and the bias measure in Table 7 identify the models that provide higher biases (Table 7) and act as outliers (Figure 3). Models with smaller biases receive larger weights. The cases that respect both criteria are the ones where the probability density is concentrated.

<b>Models/Scenarios</b>	<b>DJF</b>	<b>JJA</b>
CGCMT47_A1B	2.22	4.07
CGCMT47_A2	1.11	1.09
CGCMT47_B1	7.76	2.80
CGCMT63_A1B	4.30	31.83
CGCMT63_A2	11.06	36.56
CGCMT63_B1	1.32	0.41
CSIROMK35_B1	2.46	1.10
CSIROMK35_A2	3.37	2.77
GISSAOM_A1B	18.21	2.66
<b>GISSAOM B1</b>	24.25	4.10
MIROC32HIRES_A1B	0.07	4.28
MIROC32HIRES B1	0.09	4.51
MIROC32MEDRES A1B	8.75	1.26
MIROC32MEDRES B1	8.44	0.69
MIROC32MEDRES A2	6.57	1.86

Table 7: Relative Weighting of the 15 AOGCM Scenarios (2050s) for London

Figure 4 summarizes the posterior distributions for the precision parameters  $\lambda_j$ . It is considered as a random variable. The scoring of the AOGCM scenarios should be evaluated through the relative position of the boxplots, rather than by comparing point estimates. Comparison of their distributions across the models for a single region and any specific season provides the ordered performances of those scenarios in simulating future climate. Large  $\lambda_i$  values indicate that the distributions of the AOGCM responses are more concentrated to the true climate response, i.e. the posterior distributions which are shifted towards right indicate AOGCM's better performances than those shifted to the left. However, large overlaps among these distributions are evident indicating substantial uncertainty in the relative weighting of the models.



Figure 4: Posterior Distribution of  $\lambda_j$ , the Precision Parameter for Winter (Top) and Summer (Bottom)

So the posterior mean of each  $\lambda_j$  is calculated and the results are presented as percentages in Tables 6 and 7. The tables clearly indicate the varying result of the model

performances for different seasons, thereby suggesting a differential skill in reproducing present day climate and a different level of agreement among the models for different signals of precipitation change. Table 7 presents an overall measure of reliability for the AOGCMs by summing up the weights from each model through relative weighting. The results are ranked based on performances for summer (JJA) and winter (DJF) seasons separately. A difference in the relative weighting of the AOGCMs and scenarios can be seen.

Next, the posterior distribution of the inflation/deflation parameter θ is presented in Figure 5 to compare the simulations of the present day to future climate scenarios. A value below one represents a deterioration of the degree of the precision of the model performances.



Figure 5: Posterior Distribution of θ, the Inflation/Deflation Parameter

From Figure 5, it is seen that for summer and winter, the models and scenarios show improved performances, however with varying degree; the agreements are better represented during summer than winter. The figure presents an overall degree of performance for the REA method by considering a common value for all AOGCMs, which can limit the real representation of future climate.

# **4.2 Variable Weight (Kernel Estimator) Method**

The variable weight method involves downscaling the AOGCM responses for future climate scenarios and estimating the uncertainties using nonparametric density estimator by considering different weights at each time interval.

This study uses 22 stations for the period of 1979-2005 ( $N=27$ ) to simulate precipitation scenarios using seven meteorological variables. Employing the temporal window of 14 days ( $w=14$ ) and 27 years of historic data ( $N=27$ ), 404 days are considered as potential neighbors  $(L=(w+1) \times N-1=404)$  for each variable. 12 different runs, each comprising 27 years of daily precipitation are generated. Errors in the estimates of mean and variance of generated precipitation are evaluated using a statistical hypothesis test at 95% confidence level.

The performance of WG in representing the present climate is tested by using the nonparametric Wilcoxon-rank test and Levene's test (Levene, 1980). Table 8 presents the statistical significance test results (*p* values) in the estimate of daily precipitation for summer (JJA) and winter (DJF) for 1979-2005 in London. The *p* values at 95% confidence level for all runs are above the threshold (*0.05*), which clearly indicates that there is no evidence of different means between the observed and generated precipitations. The results of the Levene's test for the equality of variances of observed and simulated precipitation at 95% confidence level are presented in Table 8. The *p* values appear above 0.05 thresholds, indicating equal variability of the simulated

precipitation with the observed precipitation. So, the observed and the simulated precipitation can be assumed to have equal variances.

<b>Runs</b>		<b>Wilcoxon Rank Test</b>	<b>Levene's Test</b>			
	Summer	Winter	Summer	Winter		
$\mathbf{1}$	0.46	0.48	0.61	0.55		
$\overline{2}$	0.76	0.61	0.72	0.58		
3	0.64	0.67	0.56	0.99		
4	0.93	0.37	0.98	0.18		
5	0.60	0.98	0.87	0.59		
6	0.59	0.53	0.96	0.99		
7	0.91	0.95	0.64	0.20		
8	0.91	0.95	0.64	0.20		
9	0.76	0.67	0.98	0.84		
10	0.48	0.63	0.91	0.19		
11	0.77	0.80	0.41	0.66		
12	0.76	0.29	0.76	0.30		

Table 8: Test Results (*p* values) of the Wilcoxon Rank Test and Levene's Test

Frequency distributions of wet-spell lengths for winter and summer months are plotted in Figure 6. A comparison of observed and simulated values for wet-spell lengths shows very close agreement between the frequency distributions. The frequency of wetspell lengths in the simulated data for summer is almost identical to the observed values, except for the one day lengths where the simulated data show a slight overestimation. The same is the case for the winter months. The performance of the weather generator in reproducing wet-spell lengths is very good.

Using the synthetic data set created from the change factors from several AOGCMs, 324 years of data set is generated for each case. In order to investigate the intensity of wet





Figure 6: Frequency Plots of Wet Spell Lengths for Summer (Top) and Winter (Bottom)

spells for future climate, bar charts are made showing the percent change in wet spell intensity from the historical values to the future values. Intensities are calculated using the total amount of rain that fell during the spell over the length of the spell. The percent changes in wet spell intensities are determined for 3, 5 and 7 day wet spells. The plots are made for summer (June, July, August) and winter (December, January February) in both time periods. Figures 7 (a), (b) and (c) show the bar charts for the summer and winter months of the 2050s, respectively.



Figure 7 (a): Change in 3-Day-Spell Intensities for Summer, 2041-2070



Figure 7 (b): Change in 5-Day-Spell Intensities for Summer, 2041-2070

For summer wet spells, all models, except the MIROC3.2HIRES A1B, project an increase in 3-day intensities. The most significant increase in intensity is predicted by MIROC3MEDRES A2 (100%) and CSIROMK3.5 A2 (47%). For 5-day wet spells, all models predict an increase, with CSIROMK3.5 A2 and MIROC3MEDRES B1 predicting the highest intensities over 100%. The smallest increase is predicted by MIROC3.2MEDRES A1B and MIROC3HIRES A1B with below 20% of changes. The average change from all the models and scenarios is approximately between 35-70%.



Figure 7 (c): Change in 7-Day-Spell Intensities for Summer, 2041-2070

Except for CGCM3T63 B1, CGCM3T47 B1 and A2, and MIROC3MEDRES B1, most models predict a decrease in intensity for 7-day spells. Increases of 6-25% are predicted by these models. The remaining models predict a decrease in intensity of 7 to 38% for the 2050s, with the highest being generated by GISSAOM B1. Overall, the general trend for summer in the future as predicted by several AOGCM's is that shorter wet-spell intensities will increase as longer wet-spell intensities decrease.

In the next section, a comparison between both uncertainty estimation methods is presented. The mean precipitation obtained from each AOGCM and scenario is assumed to be an independent realization of future. Using this concept, climate density curves are generated by combining the information from all AOGCMs during the 2050s, the results of which are presented in section 4.3.

# **4.3 Fixed vs. Variable Weight Method**

This section presents a comparison of uncertainty estimation methods explained in sections 3.3.1 and 3.3.2 using density estimators. Figures 8 (a) and (b) present density estimates of precipitation change for the winter and summer seasons with the results obtained from the WG combined kernel density estimates and the BA-REA method for London station using 2050s (2041-2070) time slice.



Figure 8 (a): Density Estimate of the Mean Precipitation Change in London using BA-REA Method for Winter and Summer

The density estimate of the posterior distribution of the precipitation change during summer using BA-REA method is under-smoothed, many spurious bumps especially at the tails for both winter and summer can be seen which makes it harder to understand the structure of the data. The estimates calculated by the kernel estimator show evidence of a smoothed structure.

The extended benefit of kernel estimators is that unlike BA-REA, the generated outputs can be modified into indices of interest and the probabilities can be calculated for any frequency of data, monthly, daily, or yearly, while the BA-REA method only provides the mean change by combining the AOGCM scenarios. Moreover, the BA-REA method does not provide a single relative weight applicable to the overall data length. Instead, the weight/kernel function  $(K(.)$  in equation 3.29 can be calculated at any points of interest within the range of data.



Figure 8 (b): Density Estimate of the Mean Precipitation Change using Kernel Estimator for Winter (Top) and Summer (Bottom)

# **4.4 Uncertainty Estimation of Extreme Precipitation Events**

# **4.4.1 Changes in Future Extreme Precipitation Events**

Changes in the precipitation indices compared to the historic observed 1979-2005 values are computed from the downscaled precipitation for three time slices (2020s, 2050s, and 2080s) and presented in Table 9. Both summer and winter show different changing patterns. For summer, half of the scenarios show a decrease in number of heavy precipitation and very wet days for all three time slices, while most models show an increase in 5 day maximum precipitation amount.

<b>Models/Scenarios</b>	<b>Heavy Precip Days</b>				<b>Very Wet Days</b>		5 Day Precip		
	2020	2050	2080	2020	2050	2080	2020	2050	2080
<b>Summer</b>									
CGCM3T47 A1B	3.89	2.86	1.37	8.19	5.96	1.87	5.64	0.75	2.49
CGCM3T47 A2	1.87	3.45	$-2.51$	5.34	5.07	$-1.33$	5.88	3.03	$-0.55$
CGCM3T47 B1	7.26	$-2.93$	$-1.56$	13.97	0.44	5.87	8.27	1.96	2.45
CGCM3T63 A1B	$-2.38$	$-6.56$	$-1.18$	6.23	$-3.91$	2.05	12.17	$-2.54$	1.16
CGCM3T63 A2	$-10.78$	3.30	$-1.56$	$-11.57$	9.70	$-0.71$	$-4.73$	6.70	5.25
CGCM3T63 B1	$-7.51$	$-6.85$	$-7.50$	$-5.60$	$-6.85$	$-7.12$	$-2.80$	2.83	$-4.70$
CSIROMK3.5 A2	18.44	29.73	26.84	39.68	57.92	52.05	26.98	37.51	35.18
CSIROMK3.5 B1	5.61	19.57	16.37	14.77	39.68	29.00	9.81	30.38	18.45
<b>GISSAOM A1B</b>	1.03	6.38	15.91	3.56	14.59	32.38	3.64	9.59	20.15
<b>GISSAOM B1</b>	5.06	5.57	8.15	11.92	16.01	22.42	8.76	9.19	16.53
MIROC3HIRES A1B	$-25.84$	$-24.38$	$-26.72$	$-35.32$	$-38.26$	$-39.59$	$-19.41$	$-23.08$	$-26.93$
MIROC3HIRES B1	$-14.55$	$-25.70$	$-16.82$	$-18.68$	$-31.94$	$-24.82$	$-11.64$	$-19.30$	$-15.91$
MIROC3MEDRES A1B	$-13.31$	$-23.24$	$-33.12$	$-16.28$	$-31.58$	$-41.28$	$-12.08$	$-20.45$	$-27.70$
MIROC3MEDRES_A2	$-13.09$	$-12.50$	$-40.01$	$-15.75$	$-16.81$	$-56.41$	$-14.01$	$-9.18$	$-38.89$
MIROC3MEDRES B1	$-14.85$	$-20.38$	$-15.57$	$-17.53$	$-27.05$	$-20.82$	$-10.23$	$-17.58$	$-13.58$
				Winter					
CGCM3T47 A1B	26.15	38.60	47.13	40.00	59.88	76.66	19.40	27.09	29.02
CGCM3T47 A2	28.88	32.80	60.11	43.08	48.13	91.86	23.07	20.54	38.96
CGCM3T47 B1	25.31	48.85	45.16	33.38	73.36	66.94	21.49	27.87	25.75
CGCM3T63 A1B	19.57	23.31	35.05	22.77	26.50	54.55	10.54	9.52	20.76
CGCM3T63 A2	10.07	26.04	33.66	12.77	40.45	47.66	9.82	15.14	20.14
CGCM3T63 B1	20.32	7.21	19.52	16.62	5.65	29.00	11.21	1.30	10.26
CSIROMK3.5 A2	22.44	31.12	38.55	30.00	40.45	62.55	12.24	24.71	26.66
CSIROMK3.5 B1	20.04	39.51	21.30	23.54	60.51	21.17	12.65	25.41	12.75
<b>GISSAOM A1B</b>	6.87	10.70	27.38	6.31	4.08	41.54	5.80	$-0.94$	15.45
<b>GISSAOM B1</b>	17.03	11.66	18.71	19.23	16.62	23.05	12.55	5.30	6.26
MIROC3HIRES A1B	$-4.80$	6.73	6.10	$-9.38$	7.22	11.76	$-0.15$	2.02	$-2.13$
MIROC3HIRES B1	$-4.09$	$-2.91$	18.66	$-18.92$	$-5.64$	18.97	$-7.27$	$-7.47$	2.61
MIROC3MEDRES_A1B	$-7.67$	0.64	$-0.41$	$-14.77$	$-11.12$	$-2.35$	$-9.68$	$-7.52$	$-1.02$
MIROC3MEDRES A2	$-6.26$	$-1.61$	5.58	$-12.31$	$-7.67$	10.35	$-8.40$	$-5.01$	$-0.84$
MIROC3MEDRES_B1	$-9.64$	$-2.95$	6.63	$-16.92$	$-11.91$	$-0.15$	$-5.64$	$-10.02$	$-3.02$

Table 9: Percent Changes in Extreme Precipitation Events for 2020s, 2050s and 2080s

This clearly indicates a higher intensity of precipitation during extreme precipitation events. However, ranges of change are very high, indicating higher uncertainties in model projections during summer. For winter, most of the models are in agreement over the increasing trend of extreme precipitation indices for three time slices. In this case also, the uncertainty range is higher.

### **4.4.2 Distribution Fitting**

In the presence of uncertainties in AOGCM models, there is still concern over the choice of a unique distribution for the future climate responses. The comparison of the optimal distribution of different AOGCM data based on probability plots and goodness of fit test provides an insight into the level of inherent uncertainties. The performances of different distributions during summer and winter are evaluated using three goodness-of – fit-tests: Kolmogorov-Smirnov test, Anderson-Darling estimate, and Chi-Squared test.

### *Kolmogorov-Smirnov Test*

The Kolmogorov-Smirnov test is used to decide whether the sample comes from a hypothesized continuous distribution. The samples  $x_1, x_2, \dots, x_n$  are assumed to be random, originating from some distribution with Cumulative Distribution Function (CDF)  $F(x)$ . The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical CDF:

$$
D = \max_{1 \le i \le n} \left( F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \tag{4.1}
$$

### *Anderson-Darling Estimate*

 The Anderson-Darling procedure compares the fit of an observed CDF to an expected CDF. The method provides greater weight to the tail distribution than the Kolmogorov-Smirnov test. The Anderson-Darling statistic  $A^2$  is expressed as:

$$
A^{2} = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \cdot [lnF(X_{i}) + ln(1 - F(X_{n-i+1}))]
$$
\n(4.2)

### *Chi Squared Test*

The Chi –squared test is used to determine if a sample comes from a specific distribution. The test statistic is expressed as:

$$
\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \tag{4.3}
$$

where,

 $O_i$  is the observed frequency;

 $E_i$  is the expected frequency calculated by:

$$
E_i = F(x_2) - F(x_1)
$$

Where,

 $F$  is the CDF of the probability distribution being tested; and

 $x_1$  and  $x_2$  are the limits of the *i*<sup>th</sup> bin.

In terms of hypothesis tests, the distributional form is rejected at the chosen significance level  $\alpha$  if the test statistic is greater than the critical value defined as:  $\chi^2$  $1-\alpha, k-1$ , representing the Chi-squared inverse CDF with  $k-1$  degrees of freedom and a significant level of  $\alpha$ .

The performance of any specific distribution is ranked based on the goodness of fit values. The optimum parameters for the best fitted distribution function are summarized in Appendix C. From the tabulated results it can be observed that for extreme precipitation events, most models are fitted with the Generalized Extreme Value distribution with varying value of the shape  $(k)$ , location  $(\mu)$  and scale  $(\sigma)$  parameters.

However, the distribution of wet days with  $>95<sup>th</sup>$  percentile precipitation during the winter season fits a well defined Frechet distribution, indicating a distribution different than the historical perturbed/no change scenario.

The GEV distribution unites the type I, type II and type III extreme value distributions into a single family, thereby allowing a continuous range of possible shapes. For  $k < 0$ , the GEV is equivalent to the type III extreme value (Weibull). For  $k > 0$ , the GEV is equivalent to type II distribution (Frechet). As k approaches 0, the GEV becomes the type I (Gumbel). Although most of the models and scenarios show the best fit with extreme value distributions, to be more precise, with the Type II (Frechet) and Type III (Weibull) distributions with shape parameters greater and smaller than 0 respectively, the shape parameter values (k) appear close to 0. However, the differences in the k values show extent of the variations among the distributions for each index. The tables further point out the limitations of the parametric methods for quantification of uncertainties assuming any specific distribution and parameter values.

### **4.4.3 Comparison of Extended Kernel Estimators**

### *Selection of bandwidth*

To measure how well the bandwidth selection methods perform, this section proceeds with the comparison of various bandwidth selectors by applying them in the assessment of extreme precipitation indices. Figure 9 presents kernel density estimates with statistics constructed using several bandwidth selectors: (i) the rule of thumb (ROT; by Silverman, 1986) as explained in section 3.3.2), (ii) likelihood cross validation (LCV), which searches for bandwidth based on likelihood (by Terrell and Hall, 1990, as explained in section 3.5) and (iii) the plug in estimator that selects the bandwidth using the pilot estimator of the derivatives refined by Sheather and Jones, 1991 (SJPI; named after Sheather-Jones plug in estimator (section 3.5)).

The choice of kernel is strictly limited to examining two of the most widely used types: Gaussian and Epanetchnikov kernels, the functions of which are expressed as:

Gaussian: 
$$
K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}
$$

$$
Epanechnikov: \qquad K(u) = \frac{3}{4}(1 - u^2)
$$

The 'original' estimate is created by mixing the inputs, and 1000 samples are generated from the mixtures without any estimation of bandwidth. It is created for assessing how different techniques respond to the original data type. By comparing the generated estimators, it can be seen that the density estimate using ROT is highly oversmoothed, which may have missed important features of the generated data. For both kernel types, it failed to capture the multimodality. In the case of LCVs, there are suggestions of multiple modes in the density curve. However, it is still severely undersmoothed; the small bumps occurring from the uncertainties of different AOGCM types make it harder to understand the structure of real data. The bandwidth by SJPI seems to be in a better agreement with the 'original' estimate and provides a strong indication of multimodal distribution. From Figure 9, it is also evident that the choice of kernel merely plays a role in the estimation of density. So, for the present study the Gaussian kernel with Sheather-Jones plug in estimator was used to calculate the bandwidth for estimating density of the extreme precipitation indices.



Figure 9: Comparison of Various Bandwidths of Extreme Precipitation Indices
#### *Uncertainty estimation*

To examine uncertainties in future extreme precipitation events, the yearly values of the indices from each AOGCMs and scenarios are taken as a set of independent realizations. This set is then used at each time step to establish a PDF by applying the bandwidth values. The CDF values at the upper and lower ranges of each severity class are calculated by numerical integration. The difference between the upper and lower value can thus be considered as the probability of that specific class of extreme precipitation indices for future. Figures 10 through 12 present the probability of heavy precipitation days, very wet days, and 5 day precipitation for three time slices. Both indices show somewhat similar results for the summer and winter seasons. For  $< 25<sup>th</sup>$  percentile values, heavy precipitation days show an increase in probability for the later part of the century. For the  $25^{th}$ -50<sup>th</sup> and 50-75<sup>th</sup> percentile ranges, probabilities decrease slightly while approaching 2100. However, the higher probability of precipitation days over the time is observed for  $>75<sup>th</sup>$  and  $>90<sup>th</sup>$  percentile range. This trend is supported by the probabilities of very wet day and 5 day precipitation for the summer season. In summary, the increased probability of the high end extreme precipitation events indicates larger chance of high intensity events during the later part of the century. The method explained in this section can be seen as a major improvement over the 'normal' kernel (Silverman, 1986) method applied in other AOGCM and scenario uncertainty studies. The SJPI based kernel estimation method proposed here overcomes the limitations associated with the assumptions of normality in the case of unknown densities/distributions. It is completely data driven; hence, not only is it more robust, flexible, and independent, but and the methodology has been extensively revised by statisticians.



Figure 10 (a): Probability of Heavy Precipitation Days during Summer



Figure 10 (b): Probability of Heavy Precipitation Days during Winter



Figure 11 (a): Probability of Very Wet Days during Summer



Figure 11 (b): Probability of Very Wet Days during Winter



Figure 12 (a): Probability of 5 Day Precipitation during Summer



Figure 12 (b): Probability of 5 Day Precipitation during Winter

The Orthomornal method (Efromovich, 1999) proposed by Ghosh and Mujumder (2007) to estimate uncertainties of future droughts provides another important segment of the nonparametric uncertainty estimation technique. However, one major limitation of the orthonormal method is the use of a subset of the Fourier series, which consists of cosine functions without proper justification. The additional benefit of kernel density estimators for estimating AOGCM and scenario uncertainties derives from the fact that the scientific community is now highly confident that the trends in precipitation over future periods are not going to follow the same distribution as in the past. However, it is true for any statistical method that larger samples provide better estimates of any data distribution. It is our expectation that with the advance of more sophisticated global climate models, the kernel method will be applied with more confidence for uncertainty estimation problems.

## **5. Conclusions**

This study deals with the approaches for quantifying AOGCM and scenario uncertainties from the modeled outputs of extreme precipitation events for London, Ontario, Canada. This work is strictly limited to the uncertainties of the outputs from several AOGCMs and scenarios and does not consider the uncertainties due to parameterization or structure of the models.

Two very different multi-model ensemble methods namely, the Bayesian reliability ensemble average (BA-REA) and the downscaling based kernel density estimator are used for uncertainty estimation. A comparison of these two methods reveals that while the BA-REA method can be a good alternative for predicting mean changes in precipitation in any region, it cannot be used in estimating uncertainties of different extreme events occurring at a daily time scale. The capability of the BA-REA method to analyze the climate responses is fairly limited; whereas the downscaled outputs can be obtained in any frequency according to the need of the user. The data-driven kernel estimator is capable of assuming data values at each time step as an independent realization, instead of calculating weights based on the means. It has a significant implication for estimating uncertainties of extreme precipitation events; calculating weights based on the mean can ignore the higher or lower values which may cause an unrealistic representation of climate extremes, such as floods, droughts, etc. However, the kernel estimator has it's limitations too, from the extended chance of over or undersmoothing resulting from wrong selection of bandwidth. The comparison of the best fit curves for different AOGCM scenarios for extreme precipitation indices shows varying agreement and thereby the limited benefits of parametric distribution approach.

The choice of an appropriate bandwidth selection method is a significant step for kernel estimation. The shape of the distribution function is important in determining the performance of the bandwidth. The comparative results of different bandwidth selectors show that the rule of thumb (ROT) method assuming normal kernel suffers from oversmoothing for both indices while the least square cross validation (BLCV) method results in under-smoothed distributions. The SJPI estimator offered a useful compromise between the ROT and the BLCV methods. This trade-off between the distributions of the bandwidths seems to be an intrinsic criterion for assessing the performance of data-driven bandwidth selectors. Using the bandwidths calculated by the SJPI method, the CDFs for different severity classes are calculated for the extreme precipitation indices. This is estimated by the assumptions that the outputs from different AOGCMs are independent realizations; hence, indices have a different PDF at each time step and are not limited to any specific type of distribution. The nonparametric methods can be seen as a major improvement over the parametric methods, which otherwise assume specific distributions for estimating uncertainties. Considering the probabilities obtained, it can be said that the probability of severe and extreme events are going to increase for both summer and winter due to the changes in climate over the next century.

The future scope of the study includes generating probabilistic intensity-durationfrequency (IDF) curves for future extreme precipitation events by incorporating associated uncertainties from AOGCM and scenario outputs for decision making.

# **Acknowledgement**

The authors wish to gratefully thank the Canadian Foundation for Climate and Atmospheric Sciences for providing financial assistance. The constructive comments given by Dr. Claudia Tebaldi of the National Centre of Atmospheric Research for improving the BA-REA methodology are greatly acknowledged. The constructive comments of the reviewers of Water Resources Research Journal for improving the manuscript (under review) from the present work are gratefully acknowledged.

## **References**

- Adamowski, K. (1985). Nonparametric kernel estimation of flood frequencies, *Water Resources Research, 21 (11),* 1585-1590.
- Allan, M. R., and W. J. Ingram (2002). Constraints on future changes in climate and the hydrologic cycle, *Nature, 419,* 224– 232.
- Apipattanavis, S., G. Podesta, P. Rajagopalan, R. W. Katz (2007). A semiparametric multivariate and multisite weather generator. *Water Resources Research* 43, W11401.
- Bardossy, A. (1997). Downscaling from GCM to local climate through stochastic linkages, *Journal of Environmental Management, 49,* 7-17.
- Benestad, R. E. (2004). Tentative probabilistic temperature scenarios for northern Europe, *Tellus, Ser. A, 56,* 89– 101.
- Bowman, A. W. (1984). An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, *Biometrika, 71,* 353-360.
- Brandsma, T., T. A. Buishand, (1998). Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrology and Earth System Sciences* 2(2-3): 195-209.
- Brissette, F., R. Leconte, M. Minville, R. Roy (2006). Can we adequately quantify the increase/decrease of flooding due to climate change? EIC Climate Change Technology, IEEE. Doi: [10.1109/EICCCC.2006.277254.](http://dx.doi.org/10.1109/EICCCC.2006.277254)
- Brissette, F., R. Leconte, M. Khalili (2007). Efficient stochastic generation of multi-site synthetic precipitation data. *Journal of Hydrology*, vol 345, n3-4, 2007, p 121-133.
- Beersma, J. J., T. A. Buishand, R. Wojcik, (2001). Rainfall generator for the Rhine basin: multi-site simulation of daily weather variables by nearest-neighbour resampling. In:

Generation of Hydrometeorologicalreferenceconditionsfor the assessment of flood hazard in largeriverbasins, P. Krahe and D. Herpertz (Eds.), CHR-Report No. I-20, Lelystad, p. 69-77.

- Brown, M. B. and A. B. Forsythe (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association, 69,* 364-367.
- Buytaert, W., R. Ce´lleri, and L. Timbe (2009). Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty, *Geophysical Research Letters,36*, L07406, doi:10.1029/2008GL037048.
- Castro, C. L., R. A. Pielke Sr. and J. O. Adegoke (2007). Investigation of the summer climate of the contiguous United States and Mexico using the Regional Atmospheric modeling System (RAMS). Part I: Model Climatology (1950 - 2002). *Journal of Climate, 20,* 3844–3864.
- Choi, W., A. Moore, and P. F. Rasmussen (2007). Evaluation of temperature and precipitation data from NCEP-NCAR Global and Regional Reanalyses for hydrological modeling in Manitoba. *In* Proceedings of the CSCE 18<sup>th</sup> Hydrotechnical Conference on Challenges for Water Resources Engineering in a Changing World, Winnipeg, Manitoba. pp. 1-10.
- Choi, W., S. T. Jim, P.F. Rasmussen, and A. R. Moore (2009). Use of the North American Regional Reanalysis for Hydrologic Modelling in Manitoba, Canadian *Water Resources Journal, 34 (1),* 17-36.
- Colglazier, E. (1991). Scientific Uncertainties, Public Policy, and Global Warming: How Sure on Sure Enough, *Policy Studies Journal*, Vol. 19 (2), pp.: 61-72.
- Cubasch, U., G. A. Meehl, G. J. Boer, R. J. Stouffer, M. Dix, A. Noda, C. A. Senior, S. Raper, and K. S. Yap (2001), The Scientific Basic. Contribution of working group 1 to the Third Assessment Report of the Intergovernmental Panel of Climate Change, Cambridge University Press, Cambridge, UK and New York, USA, 881 pp.
- Diaz-Neito, J., R. L. Wilby (2005). A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom. *Climatic Change* 69: 245-268.
- Dibike Y. B. and P. Coulibaly (2005). Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, 307(1-4), 145-163.
- [Efromovich,](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V1D-48B0H10-3&_user=940030&_coverDate=07%2F01%2F2003&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1649253493&_rerunOrigin=google&_acct=C000048763&_version=1&_urlVersion=0&_userid=940030&md5=e1afa0adcf98e8b6702bf2c747a6e6b9&searchtype=a#bbib5) S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications*, Springer, New York.
- Elshamy, M. E., H.S. Wheater, N. Gedney and C. Huntingford, (2006). Evaluation of the rainfall component of a weather generator for climate impact studies, *J. Hydrol.* **326**: 1– 24.
- Ensor, L. A. and S. M. Robeson (2008). Statistical Characteristics of daily precipitation: comparisons of gridded and point datasets, *Journal of Applied Meteorology and Climatology, 47(9),* 2468- 2476.
- Eum, H-I., V. Arunachalam, S. P. Simonovic, (2009). Integrated Reservoir Management System for Adaptation to Climate Change Impacts in the Upper Thames River Basin. *Water Resources Research Report 62*, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada.

#### Furrer, E. M, R. W. Katz (2008). Improving the simulation of extreme precipitation

events by stochastic weather generators. *Water Resources Research* 44, W12439, doi: 10.1029/2008WR007316.

- Ghosh, S., and P. P. Majumdar (2007). Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment, *Water Resources Research, 43,* W07495, 19pp., doi:10.1029/2006WR005351.
- Giorgi, F., and L. O. Mearns, (2003), Probability of regional climate change calculated using the reliability ensemble averaging (REA) method, *Geophysical Research Letters, 30(12)*, 1629, doi: 10.1029/2003GL017130.
- Haberlandt, U., and G. W. Kite (1998), Estimation of daily space-time precipitation series for macroscale hydrologic modeling, *Hydrological Processes, 12,* 1419-1432.
- Hall, P., and J. S. Marron (1991), Lower Bounds for Bandwidth Selection in Density Estimation, *Probability Theory and Related Fields, 90,* 149-173.
- Hanson, C. L., and G. L. Johnson (1998). "GEM (Generation of weather Elements for Multiple applications): its application in areas of complex terrain". *Hydrology, Water Resources and Ecology in Headwaters IAHS* 248: 27-32.
- Hennessy, K. J., and J. F. B. Mitchell (1997), Changes in daily precipitation under enhanced greenhouse conditions, *Climate Dynamics, 13,* 667-680.
- Hewitson B. C, and R, G. Crane. (1992). Regional-scale climate prediction from the GISS GCM. Palaeogeogr Palaeoclimatol Palaeoecol (Global Planet Change Sec) 97: 249-267.
- Hughes J. P. (1993). A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomena. Ph.D. thesis, University of Washington, US.
- Hughes J.P, P. Guttorp, and S. P. Charles (1999). A non-homogeneous hidden markov model for precipitation occurrence. *Appl. Statist., 48 (1),* 15-30.
- Hughes, J. P., and P. Guttorp (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resources Research, 30(5),* 1535–1546.
- Intergovernmental Panel on Climate Change (IPCC) (2007). Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by M. Parry et al., Cambridge University Press, UK.
- IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of working group 1 to the Fourth Assessment Report of the Intergovernmental Panel of Climate Change, Annexes [Baede, A.P.M. (ed.)]. Cambridge University Press, Cambridge UK and New York, USA, 48pp.
- Jackson, C., M. K. Sen, and P. L. Stoffa (2004). An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions, *J. Climate, 17,* 2828–2841.
- Jones, M.C., J. S. Marron, and S. J. Sheather (1996). A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association, 91(433),* 401–407.
- Joubert A. M, and B. C. Hewitson (1997). Simulating present and future climates of southern Africa using general circulation models. *Prog Phys Geog 21*: 51–78.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woolen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W Higgins, J.

Janowiak, C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph (1996). The NCEP-NCAR Reanalysis Project, *Bulletin of the American Meteorological Society, 77 (3),* 437–471.

- Kay, A. L., Davies, H. N. (2008). Calculating potential evaporation from climate model data: A source of uncertainty for hydrological climate change impacts. *Journal of Hydrology 358*: 221-239.
- Kennedy, M., and A. O'Hagan (2001). Bayesian Calibration of Computer Models, *J. R. tat. Soc. Ser. B—Stat. Methodol 63,* 425–450.
- Kilsby C.G., P. D. Jones, A. Burton, A. C. Ford, H. J. Fowler, C. Harpham, P. James, A. Smith, R. L. Wilby (2007). A daily weather generator for use in climate change studies, *Environmental Modelling and Software, 22,* 1705-1719.
- Koutsoyiannis D. (2004). Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrological Sciences Journal, 49 (4),* 575–590.
- Kuchar, L. (2004). Using WGENK to generate synthetic daily weather data for modelling of agricultural processes. *Mathematics and Computers in Simulation* 65: 69–75.
- Lall, U. (1995). Nonparametric Function Estimation: Recent Hydrologic Contributions. Reviews of Geophysics, Contributions in Hydrology, U.S. National Report to the IUGG 1991-1994, 1093-1099.
- Lall, U., B. Rajagopalan and D. G. Tarboton (1996). A Nonparametric Wet/Dry Spell Model for Daily Precipitation, *Water Resources Research, 32(9),* 2803-2823.
- Levene, H. (1980). *Contributions to Probability and Statistics*. Stanford University Press, USA.
- Maurer, E.P., 2007, Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California under two emissions scenarios, *Climatic Change*, Vol. 82, No. 3-4, 309-325, doi: 10.1007/s10584-006-9180-9 [r](http://www.engr.scu.edu/~emaurer/papers/reprints/maurer_uncertainties_ca_cc_2007.pdf)
- Mehrotra, R., Srikanthan, R., Sharma, A. (2006). A comparison of three stochastic multisite precipitation occurrence generators. *Journal of Hydrology* 331: 280-292.
- Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovic, J. Wollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Ki, Y. Lin, G. Mankin, D. Parrish, and W. Shi (2006), North American Regional Reanalysis, *Bulletin of the American Meteorological Society, 87(3),* 343-360.
- Minville, M., Brissette, F., and Leconte, R. (2008). Uncertainty of the impact of climate change on the hydrology of a nordic watershed. *Journal of Hydrology, 358,* 70-83.
- Murphy, J. M. et al. (2004). Quantifying uncertainties in climate change from a large ensemble of general circulation model predictions. *Nature 430*, 768–772. (doi:10.1038/nature02771).
- Muzik, I. (2001). Sensitivity of hydrologic systems to climate change. *Canadian Water Resources Journal 26 (2),* 233–253.
- Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., and co-authors. (2000). IPCC Special Report on Emissions Scenarios. *UNEP/GRID-Ardenal Publications*.
- New, M., and M. Hulme, (2000), Representing uncertainty in climate change scenarios: a Monte-Carlo approach, *Integrated Assessment, 1,* 203-213.
- Nigam, S. And A. Ruiz-Barradas (2006). Seasonal hydroclimate variability over North American Global and Regional Reanalyses and AMIP simulations: varied representation, *Journal of Climate, 19 (5),* 815-837.
- Polansky, A. M. and E. R. Baker (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation and Simulation*, *65*, 63-80.
- Prodanovic, P. and S. P. Simonovic, (2006). [Inverse flood risk modelling of The Upper](http://www.eng.uwo.ca/research/iclr/fids/publications/products/EventModelReport2.pdf)  [Thames River Watershed.](http://www.eng.uwo.ca/research/iclr/fids/publications/products/EventModelReport2.pdf) Water Resources Research Report no. 052, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 163 pages. ISBN: (print) 978-0-7714-2634-6; (online) 978-0- 7714-2635-3.
- Prudhomme, C., D. Jakob, and C. Svensson (2003). Uncertainty and climate change impact on the flood regime of small UK catchments, *Journal of Hydrology, 277,* 1 – 23.
- Raisanen, J., and T. N. Palmer (2001). A probability and decision-model analysis of a multimodel ensemble of climate change simulations, *J. Climate, 14,* 3212– 3226.
- Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor,. (2007). Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Reid, P. A., P. D. Jones, O. Brown, C. M. Goodess, and T. D. Davies (2001). Assessments of the reliability of NCEP circulation data and relationships with surface climate by direct comparisons with station based data, *Climate Research, 17,* 247–261.
- Richardson, C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research 17*: 182–90.
- Richardson, C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* 17: 182–90.
- Robeson, S. M. and L. A. Ensor (2006). Comments on "Daily precipitation grids for South America", *Bulletin of American Meteorological Society, 87,* 1095- 1096.
- Rosenblatt (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27: 832-837.
- Rudemo, M. (1982). Empirical Choice of Histograms and Kernel Density Estimators, *Scandinavian Journal of Statistics, 9,* 65-78.
- Rummukainen, M. (1997). Methods for statistical downscaling of GCM simulations, *RMK No. 80*, SMHI, Norrkping.
- Rusticucci, M. M. and V. E. Kousky (2002). A comparative study of maximum and minimum temperatures over Argentina: NCEP-NCAR Reanalysis versus station data, *Journal of Climate, 15 (15),* 2089-2101.
- Salathe Jr., E. P. (2003). Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin, *International Journal of Climatology, 23,* 887-901.
- Scott, D. W., and G. R. Terrell, (1987). Biased and Unbiased Cross- Validation in Density Estimation, *Journal of the American Statistical Association, 82,* 1131-1146.
- Schoof, J. T., Arguez, A., Brolley, J., O'Brien, J. J. (2005). A new weather generator based on spectral properties of surface air temperatures. *Agricultural and Forest Meteorology 135*, 241–251.
- Schmidli, J., C. Frei, and P. L. Vidale (2006). Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods, *Int. J. Climatol. 26*, 679– 689.
- Schulze, R. E. (1997). Impacts of global climate change in a hydrologically vulnerable region: challenges to South African hydrologists. *Progress in Physical Geography 21*, 113-136
- Semenov, M. A, E. M Barrow (1997). Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change 35,* 397-414.
- Sharif, M., and D. H. Burn (2006). Simulating climate change scenarios using an improved K-nearest neighbor model, *Journal of hydrology, 325,* 179-196.
- Sharma, A., D. G. Tarbaton, and U. Lall, (1997). Streamflow simulation a nonparametric approach, *Water Resources Research, 33 (2),* 291-308.
- Sheather, S. J. (1983). A Data-Based Algorithm for Choosing the Window Width When Estimating the Density at a Point, *Computational Statistics and Data Analysis, 1,* 229- 238.
- Sheather, S. J. (1986). An Improved Data-Based Algorithm for Choosing the Window Width When Estimating the Density at a Point, *Computational Statistics and Data Analysis, 4, 61-65.*
- Sheather, S. J., and M. C. Jones (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation, *Journal of the Royal Statistical Society, Ser. B, 53,* 683-690.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability*, Chapman & Hall/ CRC., Washington, D.C.
- Smith, R. L., C. Tebaldi, D. Nychka, and L. Mearns (2009). Bayesian modeling of uncertainty in ensembles of climate models, *Journal of the American Statistical Association, 104 (485),* 97-116., doi:10.1198/jasa.2009.0007.
- Solaiman, T.A., S. P. Simonovic (2010*a*). [National Centers for Environmental Predicton -](http://www.eng.uwo.ca/research/iclr/fids/publications/cfcas-quantifying_uncertainty/Papers/final_published_paper.pdf) [National Center for Atmospheric Research \(NCEP-NCAR\) Reanalyses Data for](http://www.eng.uwo.ca/research/iclr/fids/publications/cfcas-quantifying_uncertainty/Papers/final_published_paper.pdf)  [Hydrologic Modelling on a Basin Scale,](http://www.eng.uwo.ca/research/iclr/fids/publications/cfcas-quantifying_uncertainty/Papers/final_published_paper.pdf) *Canadian Journal of Civil Engineering, 37(4),* 611-623.
- Solaiman, T. A., L. M. King, and S. P. Simonovic (2010*b*). Extreme precipitation vulnerability in the Upper Thames river basin: uncertainty in climate model projections, *International Journal of Climatology*, (DOI: 10.1002/joc.2244).
- Soltani, A., Hoogenboom, G. (2003). A statistical comparison of the stochastic weather generators WGEN and SIMMETEO. *Climate Research* 24: 215-230.
- Srikanthan, R., and T. A. McMahon (2001). Stochastic generation of annual, monthly and daily climate data: a review. *Hydrology and Earth Systems Sciences 5 (4),* 653–670.
- Stainforth, D. A., T. E. Downing, R. W. A. Lopez, and M. New (2007). Issues in the interpretation of climate model ensembles to inform decisions, *Philos. Trans. R. Soc., Ser. A, 365,* 2163–2177.
- Stone, D. A., and M. R. Allan (2005). The end-to-end attribution problem: From emissions to impacts, *Climate. Change, 77,* 303– 318.
- Sun, Y., S. Solomon, A. Dai, and R. Portmann (2006), How often does it rain? *Journal of Climate, 19*, 916–934.
- Tebaldi,C., L. O. Mearns, D. Nychka, and R. L. Smith (2004). [Regional probabilities of](http://www.agu.org/journals/gl/gl0424/2004GL021276/)  [precipitation change: A Bayesian analysis of multimodel simulations,](http://www.agu.org/journals/gl/gl0424/2004GL021276/) *Geophysical Research Letters, 31*.
- Tebaldi, C., R. L. Smith, D. Nychka and L.O. Mearns (2005). Quantifying uncertainty in Projections of Regional Climate Change: a Bayesian Approach to the Analysis of Multimodel Ensembles, *Journal of Climate, 18 (10),* 1524-1540.
- Tebaldi, C., and R. L. Smith (2010), Characterizing uncertainty of climate change projections using hierarchical models, In The Oxford Handbook of Applied Bayesian Analysis [eds. O'Hagan, T. and West, M.], Oxford University Press, UK, 896 pp.
- Terrell, G. R. (1990). The Maximal Smoothing Principle in Density Estimation, *Journal of the American Statistical Association, 85,* 470-477.
- Terrell, G. R., and D. W. Scott (1985). Oversmoothed Nonparametric Density Estimates, *Journal of the American Statistical Association, 80,* 209-214.
- Tolika, K., P. Maheras, H. A. Flocas, and A-P. Imitriou (2006). An evaluation of a General Circulation Model (GCM) and the NCEP-NCAR Reanalysis data for winter precipitation in Greece, *International Journal of Climatology, 26,* 935–955.
- Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. (2003). The changing character of precipitation*. Bull. Am. Meteorol. Soc. 84*, 1205–1217. (doi:10.1175/BAMS-84-9-1205)
- Trigo, R.M. and J.P. Palutikof (2001). Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. J. Climate 14, 4422- 4446.
- Vidal J.-P. and Wade S. D. (2008). A framework for developing high-resolution multimodel climate projections: 21st century scenarios for the UK. International Journal of Climatology, 28(7): 843-858.
- Vincent, L.A., and E. Mekis (2006). Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmosphere-Ocean, 44(2),* 177- 193.
- Whitfield, Paul H., Cannon, Alex J., (2000). Recent variation in climate and hydrology in Canada. *Canadian Water Resources Journal 25 (1),* 19–65.
- Widmann M, Bretherton CS, Salathe-Jr EP. (2003). Statistical precipitation downscaling over the North-western United States using numerically simulated precipitation as a predictor. *Journal of Climate 16(5),* 799-816.
- Wilby, R. L., and I. Harris (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resources Research, 42,* W02419, doi: 10.1029/2005WR004065.
- Wilby, R. L., and T. M. L. Wigley (2000). Precipitation predictors for downscaling: observed and general circulation model relationships, *International Journal of Climatology, 20,* 641-661.
- Wilks D. S and R. L. Wilby (1999). The weather generation game: a review of stochastic weather models, *Progress in Physical Geography 23(3),* 329-357.
- Wilks, D.S. (1998). Multi-site generalization of a daily stochastic precipitation model. *Journal of Hydrology 210*, 178–191.
- Woo, M-K, and R. Thorne (2006). Snowmelt contribution to discharge from a large mountainous catchment in subarctic Canada, *Hydrologic Processes, 20,* 2129-2139.
- Wood, A. W., Leung, L. 5 R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs*,* Climatic Change*, 62*, 189–216, 2004.
- Xu, Chong-Yu, Singh, V.P. (2004). Review on regional water resources assessment models under stationary and changing climate. *Water Resources Management 18,* 591– 612.
- Yates D, S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003). A technique for generating regional climate scenarios using a nearest-neighbour algorithm, *Water Resources Research 39(7),* 1199-1213.
- Zhang, X., L. A Vincent, W. D. Hogg, and A. Niitsoo (2000). Temperature and precipitation trends in Canada during the 20<sup>th</sup> century, *Atmosphere-Ocean* 38, 395-429.
- Zorita, E. and H. von Storch, 1999: The analog method [a simple statistical downscaling](http://coast.gkss.de/staff/storch/pdf/zorita_storch_1999.pdf)  [technique: comparison with more complicated methods.](http://coast.gkss.de/staff/storch/pdf/zorita_storch_1999.pdf) *J. Climate 12*, 2474-2489.

## **APPENDIX A: SRES Emission Scenarios**



#### **SRES** Scenarios

Figure A1: SRES Emission Scenarios (Nakicenovic et al, 2000)

- **A1B:** In scenario A1B, the storyline includes rapid economic expansion and globalization, a population peaking at 9 billion in 2050, and a balanced emphasis on a wide range of energy sources (Nakicenovic et al, 2000).
- **B1**: The storyline for the B1 scenario is much like A1B in terms of population and globalization; however there are changes toward a service and information economy with more resource efficient and clean technologies. Emphasis is put on finding global solutions for sustainability (Nakicenovic et al, 2000).
- **A2**: For scenario A2, the storyline consists of a world of independently operating nations with a constantly increasing population and economic development on a regional

level. Technological advances in this storyline occur more slowly due to the divisions between nations (Nakicenovic et al, 2000).

# **APPENDIX B: Comparison of Different Distributions of AOGCM Models and Scenarios for Extreme Precipitation Events**

Heavy Precipitation Days during Summer

## Historical Perturbed



#### CGCM3T47 A1B



## CGCM3T47 A2





#### CGCM3T47 B1





### CGCM3T63 A1B





## CGCM3T63 A2





#### CGCM3T63 B1





## CSIROMK3.5 A2











#### GISSAOM A1B




#### GISSAOM B1





## MIROC3HIRES A1B











### MIROC3MEDRES A1B





## MIROC3MEDRES A2





### MIROC3MEDRES B1





# Very Wet Days for Summer

## Historical Perturbed



# CGCM3T47 A1B





## CGCM3T47 A2

#### CGCM3T47 B1





## CGCM3T63 A1B

# CGCM3T63 A2





# GISSAOM A1B

#### MIROC3HIRES A1B







#### MIROC3MEDRES A2



# Maximum 5 Day Precipitation for Summer

## Historical Perturbed



#### CGCM3T47 A1B





## CGCM3T47 A2

#### CGCM3T63 A1B





# CGCM3T63 A2

### CSIROMK3.5 A2





#### GISSAOM A1B

## MIROC3HIRES A1B







# MIROC3MEDRES A2



# Heavy Precipitation Days for Winter

## Historical Perturbed



#### CGCM3T47 A1B





## CGCM3T47 A2

#### CGCM3T47 B1





ſ



# CGCM3T63 A2

Т





## CGCM3T63 B1

#### CSIROMK3.5 A2





### CSIROMK3.5 B1

### GISSAOM A1B





## GISSAOM B1

#### MIROC3HIRES A1B



# MIROC3HIRES B1



#### MIROC3MEDRES A1B



#### MIROC3MEDRES A2



## MIROC3MEDRES B1

#	<b>Distribution</b>	Kolmogorov <b>Smirnov</b>		Anderson <b>Darling</b>		<b>Chi-Squared</b>	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
$\mathbf{1}$	Frechet	0.1623	3	14.583	з	189.85	з
2	Frechet	0.1693	4	13.048	2	120.37	2
з	Gamma	0.3509	9	58.291	7	512.2	8
4	Gamma	0.31607	7	43.556	5	368.44	6
5	Gen. Extreme Value	0.16047	2	10.954	$\mathbf{1}$	50.514	1
6	Gen. Pareto	0.15909	1	60.891	9	N/A	
7	<b>Gumbel Max</b>	0.33105	8	59.366	8	1052.6	9
8	Normal	0.39562	10	72.323	10	1201.7	10
9	Rayleigh	0.47772	11	186.69	11	346.87	4
10	Rayleigh	0.66892	12	347.16	12.	1305.8	11
11	Weibull	0.29614	6	45.488	6	352.15	5
12	Weibull	0.28474	5	39.12	4	464.04	7
13	Pareto	No fit.					

Very Wet Days for Winter

### Historical Perturbed



#### CGCM3T47 A1B



#### CGCM3T47 A2



# CGCM3T47 B1



#### CGCM3T63 A1B



## CGCM3T63 A2



## CGCM3T63 B1



# CSIROMK3.5 A2



#### CSIROMK3.5 B1



#### GISSAOM A1B



# GISSAOM B1



#### MIROC3HIRES A1B



#### MIROC3HIRES B1



#### MIROC3MEDRES A1B



## MIROC3MEDRES A2



#### MIROC3MEDRES B1



5 day Maximum Precipitation for Winter 2050s





#### CGCM3T47 A1B



## CGCM3T47 A2



# CGCM3T47 B1



### CGCM3T63 A1B



### CGCM3T63 A2



# CGCM3T63 B1



#### CSIROMK3.5 A2



## CSIROMK3.5 B1


#### GISSAOM A1B



#### GISSAOM B1



#### MIROC3HIRES A1B



#### MIROC3HIRES B1



#### MIROC3MEDRES A1B



## MIROC3MEDRES A2



### MIROC3MEDRES B1



# **APPENDIX C: Distribution Fit of Extreme Precipitation Indices**



Table G-1: Heavy Precipitation Days for 2050s Summer





<b>AOGCM Models/Scenarios</b> Distribution Fit		<b>Parameters</b>					
		$\alpha$	B	$\bf k$	$\sigma$	$\mu$	
<b>Historical Perturbed</b>	<b>GEV III</b>			0.06	1.24	1.40	
CGCM3T47 A1B	Frechet	0.99	2.57				
CGCM3T47 A2	Frechet	0.92	2.14	$\overline{a}$		-	
CGCM3T47 B1	Frechet	0.99	2.74	$\blacksquare$	-	$\overline{\phantom{0}}$	
CGCM3T63 A1B	Frechet	0.89	1.80	$\overline{\phantom{a}}$		-	
CGCM3T63 A2	Frechet	0.90	2.02	$\overline{a}$			
CGCM3T63 B1	Frechet	0.82	1.43				
CSIROMK3.5 A2	Frechet	0.83	1.98				
CSIROMK3.5 B1	Frechet	0.90	2.45				
<b>GISSAOM A1B</b>	Frechet	0.81	1.43				
<b>GISSAOM B1</b>	Frechet	0.84	1.66	$\overline{a}$		-	
<b>MIROC3HIRES A1B</b>	Gen. Pareto	$\overline{\phantom{0}}$		0.63	3.52	$-0.15$	
<b>MIROC3HIRES B1</b>	Gen. Pareto			0.68	3.41	$-0.29$	
<b>MIROC3MEDRES A1B</b>	Gen. Pareto			0.64	3.33	$-0.33$	
MIROC3MEDRES A2	Frechet	0.83	1.29			-	
MIROC3MEDRES B1	Gen. Pateto			0.65	0.85	0.25	

Table G-3: Very Wet Days for Summer

<b>AOGCM Models/Scenarios</b> Distribution Fit		<b>Parameters</b>				
		K	$\sigma$	$\mu$	$\alpha$	β
<b>Historical Perturbed</b>	<b>GEV II</b>	$-0.070$	1.52	2.59		۰
CGCM3T47 A1B	<b>Gumbel Max</b>		2.17	3.40		
CGCM3T47 A2	<b>GEV II</b>	0.10	2.04	3.34		
CGCM3T47 B1	Gamma			$\overline{\phantom{a}}$	2.61	1.81
CGCM3T63 A1B	<b>GEV III</b>	$-0.14$	1.58	2.61		
CGCM3T63 A2	<b>GEV III</b>	0.14	1.72	2.99		
CGCM3T63 B1	<b>GEV III</b>	0.15	1.53	2.55		
CSIROMK3.5 A2	Weibull				2.57	6.09
CSIROMK3.5 B1	<b>GEV III</b>	$-0.16$	1.98	3.98		
<b>GISSAOM A1B</b>	<b>GEV III</b>	$-0.17$	1.77	3.20		
<b>GISSAOM B1</b>	<b>GEV III</b>	$-0.22$	1.82	3.32		
<b>MIROC3HIRES A1B</b>	<b>GEV III</b>	$-0.03$	1.18	1.5		
MIROC3HIRES B1	<b>Gumbel Max</b>		1.16	1.69		
MIROC3MEDRES A1B	Gumbel Max		1.15	1.68	$\overline{\phantom{a}}$	
MIROC3MEDRES A2	<b>GEV III</b>	$-0.20$	1.46	2.30	$\overline{a}$	
<b>MIROC3MEDRES B1</b>	<b>GEV III</b>	$-0.09$	1.30	1.88		

Table G-4: Very Wet Days for Winter





<b>AOGCM Models/Scenarios</b>	<b>Distribution Fit</b>	<b>Parameters</b>					
		K	$\sigma$	$\mu$	$\alpha$	β	γ
<b>Historical Perturbed</b>	<b>GEV II</b>	0.07	15.85	42.45	$\overline{\phantom{a}}$	$\overline{\phantom{0}}$	$\blacksquare$
CGCM3T47 A1B	<b>GEV II</b>	0.07	18.93	54.03	$\overline{a}$		$\overline{\phantom{a}}$
CGCM3T47 A2	<b>GEV II</b>	0.08	18.64	50.48	$\blacksquare$		$\overline{\phantom{0}}$
CGCM3T47 B1	<b>GEV II</b>	0.09	18.14	54.38	$\overline{a}$		
CGCM3T63 A1B	<b>GEV II</b>	0.04	17.46	46.26	$\overline{\phantom{a}}$		
CGCM3T63 A2	<b>GEV II</b>	0.05	18.58	48.35			
CGCM3T63 B1	<b>GEV II</b>	0.05	15.65	43.05			
CSIROMK3.5 A2	<b>GEVII</b>	0.098	21.17	50.56			
CSIROMK3.5 B1	<b>GEV II</b>	0.13	18.57	52.01			
<b>GISSAOM A1B</b>	<b>GEV II</b>	0.07	15.24	41.70			
<b>GISSAOM B1</b>	Frechet 3P				6.25	97.63	$-54.53$
<b>MIROC3HIRES A1B</b>	Gamma 3P				3.01	13.28	13.23
MIROC3HIRES B1	<b>GEV II</b>	0.1	13.83	38.85			$\overline{\phantom{a}}$
<b>MIROC3MEDRES A1B</b>	Gamma 3P				2.27	13.45	17.68
<b>MIROC3MEDRES A2</b>	Gamma 3P				3.26	10.843	14.15
<b>MIROC3MEDRES B1</b>	<b>Gumbel Max</b>		14.12	38.78			

Table G-6: Maximum 5 Day Precipitation for Winter

# **APPENDIX D: Previous Reports in Series**

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

(1) Slobodan P. Simonovic (2001). Assessment of the Impact of Climate Variability and Change on the Reliability, Resiliency and Vulnerability of Complex Flood Protection Systems. Water Resources Research Report no. 038, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 91 pages. ISBN: (print) 978-0-7714-2606-3; (online) 978-0-7714-2607-0.

(2) Predrag Prodanovic (2001). Fuzzy Set Ranking Methods and Multiple Expert Decision Making. Water Resources Research Report no. 039, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 68 pages. ISBN: (print) 978-0-7714-2608-7; (online) 978-0-7714-2609-4.

(3) Nirupama and Slobodan P. Simonovic (2002). Role of Remote Sensing in Disaster Management. Water Resources Research Report no. 040, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 107 pages. ISBN: (print) 978-0-7714-2610-0; (online) 978-0-7714- 2611-7.

(4) Taslima Akter and Slobodan P. Simonovic (2002). A General Overview of Multiobjective Multiple-Participant Decision Making for Flood Management. Water Resources Research Report no. 041, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 65 pages. ISBN: (print) 978-0-7714-2612-4; (online) 978-0-7714-2613-1.

(5) Nirupama and Slobodan P. Simonovic (2002). A Spatial Fuzzy Compromise Approach for Flood Disaster Management. Water Resources Research Report no. 042, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 138 pages. ISBN: (print) 978-0-7714-2614-8; (online) 978-0-7714-2615-5.

(6) K. D. W. Nandalal and Slobodan P. Simonovic (2002). State-of-the-Art Report on Systems Analysis Methods for Resolution of Conflicts in Water Resources Management. Water Resources Research Report no. 043, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 216 pages. ISBN: (print) 978-0-7714-2616-2; (online) 978-0-7714-2617-9.

(7) K. D. W. Nandalal and Slobodan P. Simonovic (2003). Conflict Resolution Support System – A Software for the Resolution of Conflicts in Water Resource Management. Water Resources Research Report no. 044, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 144 pages. ISBN: (print) 978-0-7714- 2618-6; (online) 978-0-7714-2619-3.

(8) Ibrahim El-Baroudy and Slobodan P. Simonovic (2003). New Fuzzy Performance Indices for Reliability Analysis of Water Supply Systems. Water Resources Research Report no. 045, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 90 pages. ISBN: (print) 978-0- 7714-2620-9; (online) 978-0-7714- 2621-6.

(9) Juraj Cunderlik (2003). Hydrologic Model Selection for the CFCAS Project: Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions. Water Resources Research Report no. 046, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 40 pages. ISBN: (print) 978-0-7714- 2622-3; (online) 978-0-7714- 2623-0.

(10) Juraj Cunderlik and Slobodan P. Simonovic (2004). Selection of Calibration and Verification Data for the HEC-HMS Hydrologic Model. Water Resources Research Report no. 047, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 29 pages. ISBN: (print) 978-0- 7714-2624-7; (online) 978-0-7714-2625-4.

(11) Juraj Cunderlik and Slobodan P. Simonovic (2004). Calibration, Verification and Sensitivity Analysis of the HEC-HMS Hydrologic Model. Water Resources Research Report no. 048, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 113 pages. ISBN: (print) 978- 0- 7714-2626-1; (online) 978-0-7714- 2627-8.

(12) Predrag Prodanovic and Slobodan P. Simonovic (2004). Generation of Synthetic Design Storms for the Upper Thames River basin. Water Resources Research Report no. 049, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 20 pages. ISBN: (print) 978- 0-7714-2628-5; (online) 978-0-7714-2629-2.

(13) Ibrahim El-Baroudy and Slobodan P. Simonovic (2005). Application of the Fuzzy Performance Indices to the City of London Water Supply System. Water Resources Research Report no. 050, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 137 pages. ISBN: (print) 978- 0-7714-2630-8; (online) 978-0-7714-2631-5.

(14) Ibrahim El-Baroudy and Slobodan P. Simonovic (2006). A Decision Support System for Integrated Risk Management. Water Resources Research Report no. 051, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 146 pages. ISBN: (print) 978-0-7714-2632-2; (online) 978-0- 7714-2633-9.

(15) Predrag Prodanovic and Slobodan P. Simonovic (2006). Inverse Flood Risk Modelling of The Upper Thames River Basin. Water Resources Research Report no. 052, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 163 pages. ISBN: (print) 978-0-7714-2634-6; (online) 978-0-7714-2635-3.

(16) Predrag Prodanovic and Slobodan P. Simonovic (2006). Inverse Drought Risk Modelling of The Upper Thames River Basin. Water Resources Research Report no. 053, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 252 pages. ISBN: (print) 978-0-7714-2636-0; (online) 978-0-7714-2637-7.

(17) Predrag Prodanovic and Slobodan P. Simonovic (2007). Dynamic Feedback Coupling of Continuous Hydrologic and Socio-Economic Model Components of the Upper Thames River Basin. Water Resources Research Report no. 054, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 437 pages. ISBN: (print) 978-0-7714-2638-4; (online) 978-0- 7714-2639-1.

(18) Subhankar Karmakar and Slobodan P. Simonovic (2007). Flood Frequency Analysis Using Copula with Mixed Marginal Distributions. Water Resources Research Report no. 055, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 144 pages. ISBN: (print) 978-0-7714-2658-2; (online) 978-0-7714-2659-9.

(19) Jordan Black, Subhankar Karmakar and Slobodan P. Simonovic (2007). A Web-Based Flood Information System. Water Resources Research Report no. 056, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 133 pages. ISBN: (print) 978-0-7714-2660-5; (online) 978-0- 7714-2661-2.

(20) Angela Peck, Subhankar Karmakar and Slobodan P. Simonovic (2007). Physical, Economical, Infrastructural and Social Flood Risk – Vulnerability Analyses in GIS. Water Resources Research Report no. 057, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 80 pages. ISBN: (print) 978-0- 7714-2662-9; (online) 978-0-7714-2663-6.

(21) Predrag Prodanovic and Slobodan P. Simonovic (2007). Development of Rainfall Intensity Duration Frequency Curves for the City of London Under the Changing Climate. Water Resources Research Report no. 058, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 51 pages. ISBN: (print) 978-0- 7714-2667-4; (online) 978-0-7714-2668-1.

(22) Evan G. R. Davies and Slobodan P. Simonovic (2008). An integrated system dynamics model for analyzing behaviour of the social-economic-climatic system: Model description and model use guide. Water Resources Research Report no. 059, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 233 pages. ISBN: (print) 978-0-7714-2679-7; (online) 978-0- 7714-2680-3.

(23) Vasan Arunachalam (2008). Optimization Using Differential Evolution. Water Resources Research Report no. 060, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 42 pages. ISBN: (print) 978-0-7714- 2689-6; (online) 978-0-7714-2690-2.

(24) Rajesh Shrestha and Slobodan P. Simonovic (2009). A Fuzzy Set Theory Based Methodology for Analysis of Uncertainties in Stage-Discharge Measurements and Rating Curve. Water Resources Research Report no. 061, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 104 pages. ISBN: (print) 978-0-7714-2707-7; (online) 978-0-7714-2708-4.

(25) Hyung-Il Eum, Vasan Arunachalam and Slobodan P. Simonovic (2009). Integrated Reservoir Management System for Adaptation to Climate Change Impacts in the Upper Thames River Basin. Water Resources Research Report no. 062, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 81 pages. ISBN: (print) 978-0-7714-2710-7; (online) 978-0-7714-2711-4.

(26) Evan G. R. Davies and Slobodan P. Simonovic (2009). Energy Sector for the Integrated System Dynamics Model for Analyzing Behaviour of the Social- Economic-Climatic Model. Water Resources Research Report no. 063. Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada. 191 pages. ISBN: (print) 978-0-7714-2712-1; (online) 978-0-7714-2713-8.

(27) Leanna King, Tarana Solaiman, and Slobodan P. Simonovic (2009). Assessment of Climatic Vulnerability in the Upper Thames River Basin. Water Resources Research Report no. 064, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 61pages. ISBN: (print) 978-0- 7714-2816-6; (online) 978-0-7714- 2817-3.

(28) Slobodan P. Simonovic and Angela Peck (2009). Updated Rainfall Intensity Duration Frequency Curves for the City of London under Changing Climate. Water Resources Research Report no. 065, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 64pages. ISBN: (print) 978-0-7714-2819-7; (online) 987-0-7714-2820-3.

(29) Leanna King, Tarana Solaiman, and Slobodan P. Simonovic (2010). Assessment of Climatic Vulnerability in the Upper Thames River Basin: Part 2. Water Resources Research Report no. 066, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 72pages. ISBN: (print) 978-0- 7714-2834-0; (online) 978-0-7714-2835-7.

(30) Christopher J. Popovich, Slobodan P. Simonovic and Gordon A. McBean (2010).Use of an Integrated System Dynamics Model for Analyzing Behaviour of the Social-Economic-Climatic System in Policy Development. Water Resources Research Report no. 067, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 37 pages. ISBN: (print) 978-0- 7714-2838-8; (online) 978-0-7714-2839-5.

(31) Hyung-Il Eum and Slobodan P. Simonovic (2009). City of London: Vulnerability of Infrastructure to Climate Change; Background Report 1 – Climate and Hydrologic Modeling. Water Resources Research Report no. 068, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 102pages. ISBN: (print) 978-0-7714-2844-9; (online) 978-0-7714-2845-6.

(32) Dragan Sredojevic and Slobodan P. Simonovic (2009). City of London: Vulnerability of Infrastructure to Climate Change; Background Report 2 – Hydraulic Modeling and Floodplain Mapping. Water Resources Research Report no. 069, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 147 pages. ISBN: (print) 978-0-7714-2846-3; (online) 978-0- 7714-2847-0.