Elsevier Editorial System(tm) for Journal of Hydrology

Manuscript Draft

Manuscript Number: HYDROL26154R2

Title: Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate

Article Type: Research paper

Keywords: extreme streamflow; variance structure

Corresponding Author: Dr. James Forrest Fox, PhD

Corresponding Author's Institution: University of Kentucky

First Author: Nabil Al Aamery

Order of Authors: Nabil Al Aamery; James Forrest Fox, PhD; Mark Snyder; Chandra Chandramouli

Abstract: Coupling global climate models, hydrologic models and extreme value analysis provides a method to forecast streamflow maxima, however the elusive variance structure of the results hinders confidence in application. Directly correcting the bias of forecasts using the relative change between forecast and control simulations has been shown to marginalize hydrologic uncertainty, reduce model bias, and remove systematic variance when predicting mean monthly and mean annual streamflow, prompting our investigation for maxima streamflow. We assess the variance structure of streamflow maxima using realizations of emission scenario, global climate model type and project phase, downscaling methods, bias correction, extreme value methods, and hydrologic model inputs and parameterization. Results show that the relative change of streamflow maxima was not dependent on systematic variance from the annual maxima versus peak over threshold method applied, albeit we stress that researchers strictly adhere to rules from extreme value theory when applying the peak over threshold method. Regardless of which method is applied, extreme value model fitting does add variance to the projection, and the variance is an increasing function of the return period. Unlike the relative change of mean streamflow, results show that the variance of the maxima's relative change was dependent on all climate model factors tested as well as hydrologic model inputs and calibration. Ensemble projections forecast an increase of streamflow maxima for 2050 with pronounced forecast standard error, including an increase of +30(±21), +38(±34) and +51(±85)% for 2, 20 and 100 year streamflow events for the wet temperate region studied. The variance of maxima projections was dominated by climate model factors and extreme value analyses.

March 9, 2018

Dear Editor McVicar:

Please find the second revised submittal to *Journal of Hydrology* entitled "Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate" by Nabil Al Aamery, Jimmy Fox, Mark Snyder, and Chandra Chandramouli.

We thank the reviewers and editorial board for their hard work on reviewing our paper, and our appreciation is extended in the Acknowledgements section. We have addressed all remaining minor comments.

My contact information is:

James F. Fox,
Raymond-Blythe Professor of Civil Engineering
Professor
Civil Engineering Department
University of Kentucky
161 O. H. Raymond Bldg. - Rm 354G
Lexington, KY 40506-0281
Phone: 859-257-8668
Fax: 859-257-4404
Email: james.fox@uky.edu

Please feel free to contact me for any information that might assist in the review of this manuscript.

Best regards,

Jimmy Fox

Revision Notes

Editor and Reviewer comments are Blue.

Author's Responses are Black.

****
COMMENTS FROM EDITORS AND REVIEWERS
****
Editor comments:

Hello again Jimmy, Nabil, Mark and Chandra,

thanks for submitting your revised manuscript (MS) back to JoH where it was assessed by the same two reviewers as your original submission. Also thanks for your comprehensive response letter, noting please use black-and-blue interleaved text as requested in my comment 18 of your original submission.

EC1) Most abstracts do not include separate paragraphs. Please see if it works OK if you remove these.

No problem, and Done.

EC2) RE your response to comment 11 the Associate Editor is not anonymous, it is Yongqiang Zhang - please see the names at the bottom of the previous and this version of the official correspondence from JoH to you.

We now include Associate Editor Yongqiang Zhang in the Acknowledgements.

EC3) L203, what are the remaining (17%) land uses? It would be good to get the total provided to be more than 95% as opposed to the current 83%.

We have added:
"The land use is dominated by agricultural equal to 72%. The remaining land uses are urban/suburban equal to 13%, forest equal to 14%, and open water and wetlands equal to 1%."

EC4) L442, this is very good addition to your MS. Its great to see the GCM hydroclimatic community assessing all meteorological variables that govern the evaporative process. I hope all others follow your lead, and this is something you can recommend when reviewing future submissions for JoH and elsewhere that will surely come your way once this MS is published in JoH.

Thank you for the comment and we agree.

EC5) Good luck; and I look forward to seeing the next revision (which I hope is 'near-final', if not 'final') ASAP.

Thank you.

****
Associate Editor comments:

AEC1) The authors addressed most of comments from the two reviewers. I would like to recommend its publication after some minor issues are solved. Congratulations!

Thank you and we have addressed the minor comments mentioned.

****

Reviewer #1:

R1C1) I am very satisfied with the careful revision. All my comments have been well addressed.

Thank you.

****

Reviewer #2:

Dear Authors,

I think the manuscript has been substantially improved. I like the inclusions and the new structure.
Indeed the numbered objectives matching the conclusions were a good choice.
As it was a very good job, I have just some minor suggestions:

R1C1) Lines 165 - 178: Almost all sentences in the paragraph starts with "The relative change approach", that does not allow a pleasurable reading.

Thank you, and we have edited the lines to read more nicely.

R2C2) Lines 495 - 496: I do not think that the sentence "simulated streamflow well" is clear enough for scientific communication;

We removed the phrase, and used a conjunction to combine with the next sentence, which describes what we were inferring regarding "well".

R2C3) Conclusions: Numbered conclusions are much better. But I would still use a small sentence before the numbers ("The main conclusions or our work are described following...")
[EiC: agree, break the ice a little.]

Yes, we have added this small sentence. Thank you.

Congratulations

Thank you.

****

***Highlights (3 to 5 bullet points (maximum 85 characters including spaces per bullet point)**

HIGHLIGHTS

- Variance of forecasted streamflow maxima is not dependent on the extremal method
- Variance of forecasted maxima increases with return period from extreme model fitting
- The variance of the maxima's relative change was dependent on all climate model factors
- Variance from climate and extreme model factors dominates over hydrologic model factors

Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate

By Nabil Al Aamery, James F. Fox, Mark Snyder and Chandra V. Chandramouli

Nabil Al Aamery, Ph.D. Candidate of Civil Engineering at the University of Kentucky, nabil.hussain@uky.edu.

James F. Fox, Professor of Civil Engineering at the University of Kentucky, james.fox@uky.edu, +1(859)257-8668.

Mark Snyder, Project Scientist of Earth and Planetary Sciences Department, University of California Santa Cruz, masnyder@ucsc.edu.

Chandra V Chandramouli, Associate Professor of Civil Engineering, Mechanical and Civil Engineering Department at Purdue University Northwest, cviswana@pnw.edu.

Corresponding Author: James F. Fox, james.fox@uky.edu, +1(859)257-8668, 161 Raymond Bldg, Lexington KY 40506.

Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate

Abstract:

Coupling global climate models, hydrologic models and extreme value analysis provides a method to forecast streamflow maxima, however the elusive variance structure of the results hinders confidence in application. Directly correcting the bias of forecasts using the relative change between forecast and control simulations has been shown to marginalize hydrologic uncertainty, reduce model bias, and remove systematic variance when predicting mean monthly and mean annual streamflow, prompting our investigation for maxima streamflow. We assess the variance structure of streamflow maxima using realizations of emission scenario, global climate model type and project phase, downscaling methods, bias correction, extreme value methods, and hydrologic model inputs and parameterization. Results show that the relative change of streamflow maxima was not dependent on systematic variance from the annual maxima versus peak over threshold method applied, albeit we stress that researchers strictly adhere to rules from extreme value theory when applying the peak over threshold method. Regardless of which method is applied, extreme value model fitting does add variance to the projection, and the variance is an increasing function of the return period. Unlike the relative change of mean streamflow, results show that the variance of the maxima's relative change was dependent on all climate model factors tested as well as hydrologic model inputs and calibration. Ensemble projections forecast an increase of streamflow maxima for 2050 with pronounced forecast standard error, including an increase of +30(±21), +38(±34) and +51(±85)% for 2, 20 and 100 year streamflow events for the wet temperate region studied. The variance of maxima projections was dominated by climate model factors and extreme value analyses.

1 INTRODUCTION

Streamflow maxima is one of the most sought after response variables within hydrologic research and application (Coles, 2001, Begueria & Vicente-Serrano, 2006, Reiss &Thomas, 2007). Streamflow extreme maxima re-contours the morphology of the fluvial system (Leopold et al., 2012), partially controls the stream biogeochemical function (Ford and Fox, 2015), can destroy human infrastructure (Melillo et al., 2014), and resupplies human water stores for consumption, food production and energy generation (Rosenzweig et al., 2001, Mirza, 2003, McMichael et al., 2007). The complex earth system for which streamflow maxima responds is no less encompassing of hydrology than streamflow itself and includes components such as the climates ability to produce precipitation and weather patterns, the watershed's physiogeographic configuration and ability to respond to precipitation, and human's influence on both the watershed and climate. Despite hydrologists' long historical emphasis upon study of streamflow extreme maxima, current disparity is prevalent in terms of both streamflow maxima's current estimations and its gradient as we forecast into the future (Khaliq et al., 2006). Scientific gaps associated with estimating and forecasting current and future streamflow maxima is qualitatively attributed to scientific uncertainty surrounding human's economic behavior and influence on the earth system, representation of the climate and its changes, hydrologic representation of streamflow, and scalar coupling of a changing climate within a hydrologic representation of the earth (Madsen et al., 2014, IPCC, 2013). The difficulty of streamflow extreme maxima estimation and forecasting in a non-stationary earth system has challenged hydrologists to consider the potential use of new methodologies for investigating and forecasting streamflow.

One methodology for which streamflow maxima investigation and forecasting has received some recent attention is through the use of non-stationary projection with global climate models that can be used to drive hydrologic and statistical forecasting (Prudhomme et al., 2003; Dankers and Feyen, 2008; Mantua et al., 2010; Lawrence &Hisdal, 2011; Zhang et al., 2014). This method involves application of the non-stationarity form of long-term climate change projected using global climate models as a means to provide a physics-based guideline for extrapolation (Lima et al., 2015, Shamir et al., 2015). The global climate model results are post-processed for scalar considerations and then propagated through hydrologic models for predicting multi-year streamflow time series. Thereafter, the extreme value theorem is adopted to study streamflow extremes because the theory provides a mathematical basis for the definition of extremes and has been used to prove that the distribution of extremes follow similarity at their limit (e.g., Coles, 2001). Somewhat analogous to the central limit theorem, the extreme value theorem focuses on the statistical distribution and behavior of maxima that may arise from an unknown distribution for a population of a sequence of values measured over many time units. In this manner, hydrologists can statistically investigate current and forecast extreme value extreme maxima such as 2-, 20- and 100-year events via time series generated from the mentioned hydrologic modeling.

The coupling of global climate and hydrologic models for forecasting streamflow extreme maxima has been recently criticized for water infrastructure planning in some engineering and management circuits (Moradkhani, 2017), and we tend to agree that use of the methodology in a infrastructure design capacity is a bit preliminary given that published applications and results of the method is still in its infancy. Yet, we argue that the time is ripe for elucidating the variance structure of streamflow maxima forecasted with global climate models. We offer several reasons for this contention. First, highlighting the variance structure of forecasted streamflow maxima provides hydrologic and climate researchers with knowledge of highly sensitive factors and parameters of streamflow forecasting that systemically increase the size of the solution space, so that researchers might focus their attention towards improving model structure and parameterization. Second, the variance structure of forecasted streamflow maxima allows researchers to see what extent the previous results of forecasted mean streamflow might be adopted and extrapolated for forecasting extremes. There is a plethora of studies that forecast mean streamflow with global climate models (Chen et al., 2011, Al Aamery et al., 2016, Fatichi et al., 2014) and there is a question as to what extent results from these mean-focused studies might be relevant to the study of extreme streamflow, especially in light of the extra level of uncertainty that is introduced to forecasting maxima during application of the extreme value theorem. Third, a reason for investigating the variance structure of forecasted streamflow maxima is to help provide balanced forecasts that can be compared with a meta-analysis of trends in existing literature results such as in wet temperate regions.

While variance analysis of streamflow extreme maxima is sparse in the literature, global climate model research and forecasting of mean annual and mean monthly streamflow tends to suggest that climate models are different in their structures and parametrizations (Randall et al. 2007), downscaling methods are distinct in their stucture and results to re-scale global results

(Wilby and Dawson, 2007, Warner, 2010, Mearns et al., 2013), emission scenarios address the uncertainty of future economic and CO_2 conditions (IPCC, 2007, IPCC, 2013), the version of climate model projects are different in their structures, and the newer version (CMIP5) is distinct realtive older versions of models (CMIP3) (Brekke et al., 2013), and the bias implementation of climate results is a source for variance presence in hydrologic models (Teutschbein and Seibert, 2012, Al Aamery et al., 2016); and therefore all such components could have the potential to impact the variance structure of forecasted streamflow maxima. The few studies that have forecasted streamflow maxima with global climate models (see Table 1) have results that tend to corroborate some findings from mean-focused studies and suggest the type of global climate model applied caused differences in streamflow extreme forecasts (Prudhomme et al., 2003; Lawrence and Hisdal, 2011), and emission scenario can also shift extreme predictions (Dankers and Feyen, 2008; Mantua et al., 2010; Zhang et al., 2014). Applications of extreme value theory suggest the statistical analysis associated with the choice of extreme value analysis method has the potential to impact the variance of forecasted streamflow maxima; and the annual maxima method is criticized for its neglect of multiple extremes per annum while the peak over threshold method has been criticized for subjectivity of threshold selection (Svensson et al., 2005; Scarrott and MacDonald, 2012; Bezak et al., 2014; Fischer and Schumann, 2016).

<Table 1 here please>

Beyond uncertainty surrounding the global climate model projections and extreme value methods, there are additional uncertainty considerations with respect to the future hydrologic balance and its simulation when forecasting streamflow maxima. As one example, future changes in the hydrologic cycle, and in turn streamflow, are primarily driven by changes in precipitation and evapotranspiration. Studies that forecast streamflow maxima with global climate models have focused on precipitation and temperature differences within simulation of future periods (Mantua et al., 2010; Zhang et al., 2014), however it is now recognized that evaporative demand is physically controlled by net radiation, vapor pressure and wind speed as well as air temperature (Donohue et al 2010). Future projections of these additional variables suggest decreases in wind speed and net radiation and an increase in relative humidity for some regions (Willet et al., 2008; Wild, 2009; McVicar et al., 2012). The directions of the projected shifts would decrease evapotranspiration and in turn could potentially increase variability when forecasting streamflow maxima. As a second example, hydrologic model fit and modeling uncertainty when simulating the water balance has the potential to increase the variability of projected streamflow maxima (Al Aamery et al., 2016). Recent study has tended to marginalize the importance of hydrologic model calibration and uncertainty for mean streamflow projections when considering relative future changes (Niraula et al., 2015), however streamflow maxima has not yet been tested in this context, to our knowledge.

Our objectives were to: (1) perform coupled climate, hydrologic and statistical model simulation and evaluation to build realizations of streamflow maxima; (2) perform variance analysis to test for systematic uncertainty from climate and extreme modeling factors potentially controlling streamflow maxima forecasting; (3) perform uncertainty analysis to quantify variance from hydrologic modeling; and (4) forecast streamflow maxima for the wet temperate region studied herein and provide literature comparison. These objectives provide the structural sub-headings used in the following Methods, Results and Discussion sections.

2 THEORETICAL BACKGROUND

The variance structure of forecasted streamflow maxima can be decomposed as a function of potentially controlling modeling factors. The conceptual model of factors that have the potential impact forecasted streamflow maxima variance is shown in Figure 1. As can be seen in the figure, modeling factors that may impact the variance structure can be grouped into those associated with climate modeling (CMFs in Figure 1) including global climate model (GCM) type, hydrologic modeling (HMFs) and uncertainty in inputs and parameterization, and statistical modeling of extremes (SMFs) associated with different fitting methods and distributions. Land use and management modeling is not shown in the figure and was treated as static in this study, but it is also recognized to potentially control future streamflow.

<Figure 1 here please>

The response variables are streamflow maxima associated with different return periods, including 2, 20 and 100 year return periods, so the distribution of extremes can be quantified (Lawrence and Hisdal, 2011). We consider response of the future relative change in streamflow equal to the percent difference of GCM-forecasted streamflow maxima relative to GCM-hindcast maxima ($\Delta Q_{F\text{-}H\,x}$, where x indicates the return period). The future streamflow maxima can be related to the 'real' streamflow maxima by using the relative changes derived from the forecast projections and hindcast-control projections coupled with the observations, such as using the delta method directly applied to streamflow model results.

The relative change approach has become rather popular in climate change studies that emphasize GCM-forecasted streamflow (Chien et al., 2012; Harding et al., 2012; Fitichi et al., 2014; Niraula et al., 2015; Al Aamery et al., 2016). The approach has been suggested to remove seasonal, spatial, and/or inter-annual biases of GCMs or statistical artifacts from the downscaling method that are not accounted for in bias correction methods (Harding et al., 2012). In addition, application of the relative change has recently shown no significant dependence upon calibrated versus un-calibrated hydrologic model simulation, thus suggesting the response variable does not require model calibration to see the projected direction of future streamflow (Niraula et al., 2015). The approach has also shown less dependence upon climate modelling factors (i.e., CMFs in Fig 1) as compared to the absolute forecasted streamflow suggesting that biases specific to a model structure could be accounted (Al Aamery et al., 2016). While the relative change approach has shown potential in past studies, these studies have tended to focus on the mean forecasting of streamflow. In the present study, we consider the method for streamflow maxima, which is one contribution of this paper.

Realizations of the relative change in streamflow maxima can be simulated as a function of climate, hydrologic, and statistical modeling factors within a variance analysis ensemble (Al Aamery et al., 2016). In the present study, we included permutations using seven emission scenarios (i.e., emission factor, CMF1) propagated through eight different GCMs associated with phase three and four climate projects, i.e., CMIP3, CMIP5, (i.e., GCM type and version factors, CMF2, 3) that were downscaled using two statistical downscaling methods and four dynamical downscaling methods (i.e., downscaling factor, CMF4). Further, our post-processing and hydrologic analyses of downscaled hindcast (1983-2000) and forecast (2048-2065) climate model results considered bias correction (i.e., bias factor, SMF1) propagated through a continuous simulation hydrologic model. We performed both annual maxima and peak over threshold extreme value analyses (i.e., extreme value factor, SMF1,2) of hydrologic model results given recent debate in the literature over the best method. We also investigated additional uncertainty considerations with respect to additional hydrologic inputs and hydrologic uncertainty (HMF 1, 3).

3 STUDY SITE AND MATERIALS:

The study site was South Elkhorn Watershed in Lexington, Kentucky USA (see Figure 2). This watershed is within a wet and temperate region where a future change in climate, including an increase in precipitation and temperature, is projected (Melillo et al., 2014). According to Melillo et al. (2014), a 20 to 30% increase in annual maximum precipitation is projected under RCP 8.5 emission scenario for the end of the century. Additionally, at least, 80% of the models used in Melillo et al. (2014) are in agreement for this region. The watershed covers an area of 478.6 km^2 with surface elevations ranging between 197 to 325 m asl. The land use is dominated by agricultural equal to 72%. The remaining land uses are urban/suburban equal to 13%, forest equal to 14%, and open water and wetlands equal to 1%.

<Figure 2 here please>

The results of eight GCMs were implemented in this analysis. The GCM models reflected four different GCM model types and two versions of each model, inculding a version from CMIP3 and the newer version from CMIP5 (Brekke et al., 2013; Al Aamery et al., 2016). The GCMs included the Canadian Global Climate Model including CGCM3 from CMIP3 and CanESM2 from CMIP5 (Flato, 2005); the National Center for Atmospheric Research Community Climate Model including CCSM3 from CMIP3 and CCSM4 from CMIP5 (Collins et al., 2006); the Geophysical Fluid Dynamics Laboratory including GFDL CM2.1 from CMIP3 and CM3 from CMIP5 (Delworth et al., 2006); and the United Kingdom Hadley Centre Climate Model including HadCM3 from CMIP3 and HadGEM2-ES from CMIP5 (Gordon et al., 2000). These GCMs were chosen for their representation in different climate projects, including CMIP3, CMIP5, and NARCCAP projects, and their available archives of climate results for the current and future periods focused on in this study (Brekke et al., 2013; Mearns et al., 2013; Al Aamery et al., 2016).

Statistical downscaling and dynamical downscaling results were included in this analysis. The statistical downscaling results were used from the Coupled Model Inter-comparison Project phase three (CMIP3) and phase five (CMIP5) (Brekke et al., 2013). The dynamical downscaling results were used from the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2013). These downscaling methods represent two distinct approaches for downscaling GCM results from their coarse scale to a finer watershed scale. The statistical downscaling method is statistically based and adopts empirical-statistical relationships to estimate the small-scale climate variables based on the large-scale atmospheric variables (Wilby and Dawson, 2007). The statistical downscaling method implemented in CMIP3 and CMIP5 projects adopt two schemes including bias correction and spatial disaggregation (BCSD) and bias-correction and constructed analogs (BCCA) (Brekke et al., 2013). The dynamical downscaling method is physically-based and uses regional climate models (RCMs) whose boundary conditions are forced by the results of the parent GCM to simulate the atmospheric physical processes on a regional scale (Warner, 2010). Six regional climate models were implemented through the NARCCAP project including the Canadian Regional Climate Model (CRCM) (Plummer et al., 2006), the Experimental Climate Prediction Center (ECPC) model (Juang et al., 1997), the Hadley Regional Model 3 (HRM3) (Jones et al., 2003), the MM5-PSU/NCAR mesoscale model (MM5I) (Chen and Dudhia, 2001), the Reginal Climate Model version 3 (RCM3) (Giorgi et al., 1993), and the Weather Research and Forecasting model (WRFP) (Skamarock et al., 2005).

4 METHODS

4.1 Modeling simulations and evaluation

The Soil and Water Assessment Tool (SWAT; the version was ArcSWAT 2012.10.1.13) model was applied to simulate the hydrology of South Elkhorn Watershed. This model is physically based and was applied successfully in this region and many other regions around the world (Palanisamy and Workman, 2014; Gassman et al., 2007; Arnold et al., 1998). The model was evaluated over 1981-2000 using the observed climate and streamflow data and applied for the hindcast period (1981-2000) and forecast period (2046-2065) using the GCMs results of daily precipitation and maximum and minimum temperature (see Figure 3 in Al Aamery et al., 2016 for evaluation methods of SWAT). We obtained all the data required by SWAT including topography, soil, and landuse data from publically available databases. The topography and streamlines data were obtained from the National Map website (http://viewer.nationalmap.gov/viewer/), and the soil data was obtained from the Data Gateway website (http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx). The landuse data of 1992 was used from the USGS website (http://www.mrlc.gov/nlcd2011.php). The observed climate data of daily precipitation and maximum and minimum temperature were obtained from Bluegrass Airport meteorological gage station. The model was evaluated using four USGS streamflow gage stations within the watershed including South Elkhorn Creek at Fort Spring, USGS03289000, Town Branch at Yarnallton Road, USGS03289200, South Elkhorn Creek near Midway, USGS03289300, and Elkhorn Creek near Frankfort, USGS03289500. Results for the South Elkhorn Creek near Midway station were analyzed for the relative change in streamflow maxima. Model evaluation including calibration, validation, and sensitivity analysis was performed semi-automatically via SWAT-CUP software for Sequential Uncertainty Fitting SUFI2 for the four gage stations in the watershed (Abbaspour et al., 2007). The first two years was left as a spin-up period for SWAT (Arnold et al., 2010).

<Figure 3 here please>

As input to the hydrologic modeling, the scaling of precipitation and temperature method of Lenderink et al. (2007) was applied to correct the bias in the climate data. The method operates with monthly correction values based on the difference between observed and current period simulated values as:

$$P^*_{sc}(d) = P_{sc}(d)\,\frac{\mu_m(P_{obs}(d))}{\mu_m(P_{sc}(d))}, \tag{1}$$

$$P^*_{sf}(d) = P_{sf}(d)\,\frac{\mu_m(P_{obs}(d))}{\mu_m(P_{sc}(d))}, \tag{2}$$

$$T^*_{sc}(d) = T_{sc}(d) + \mu_m\big(T_{obs}(d)\big) - \mu_m\big(T_{sc}(d)\big)\text{ , and} \tag{3}$$

$$T^*_{sf}(d) = T_{sf}(d) + \mu_m\big(T_{obs}(d)\big) - \mu_m\big(T_{sc}(d)\big)\,, \tag{4}$$

where $P^*_{sc}(d)$ and $T^*_{sc}(d)$ are the corrected daily precipitation and temperature for the simulated current period, $P^*_{sf}(d)$ and $T^*_{sf}(d)$ are the corrected daily precipitation and temperature for the simulated future period, $P_{sc}(d)$ and $T_{sc}(d)$ are the uncorrected daily precipitation and temperature for the simulated current period, $P_{sf}(d)$ and $T_{sf}(d)$ are the uncorrected daily precipitation and temperature for the simulated future period, $\mu_m(P_{obs}(d))$ is the average of observed daily precipitation values for a given month, $\mu_m(P_{sc}(d))$ is the average daily precipitation for the current simulated values, $\mu_m\big(T_{obs}(d)\big)$ is the average of observed daily temperature values, $\mu_m\big(T_{sc}(d)\big)$ is the average of daily temperature for the current simulated period, and *m* stands for "within monthly time step".

The annual maxima (AM) and peak over threshold (POT) methods were carried out for each realization from the hydrologic modeling results in order to analyze the extremes (see Figure 3). The AM series is constructed by selecting one value per a specific time over the sample size. In streamflow studies such as herein, this value is the maximum water discharge value selected over one year from the daily time series data (Khaliq et al., 2006, Haan, 2002). Thereby, the AM series replaces the flow series ($q_1, q_2, \ldots.., q_{365}$) of a year ($j$) by the largest flood value q_m^j (where $1 \le m \le$ total number of days in year j, $1 \le j \le n$, and n is the number of years). According to the extreme theorem, the probability of the rescaled M_n is approaching the General Extreme Value (GEV) family when $n \to \infty$. The GEV family distribution is expressed as follows:

$$G(q) = \exp\left\{ -\left[1 + \xi\left(\frac{q-\mu}{\sigma}\right) \right]^{-1/\xi} \right\}, \tag{5}$$

where $\left\{ q : 1 + \xi\left(\frac{q-\mu}{\sigma}\right) > 0 \right\}$, the location parameter $-\infty < \mu < \infty$, the scale parameter $\sigma > 0$, and the shape parameter $-\infty < \xi < \infty$. Depending on the value of the shape parameter ξ, the GEV family has three distinct probability distributions. The light tail Gumbel type when $\xi = 0$, the heavy tail Fréchet type when $\xi > 0$, and the bounded upper tail Weibull type when $\xi < 0$. The extreme quantiles of the return level T when $\xi \ne 0$ are then calculated as follows:

$$q_T = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-\log\left(1 - \frac{1}{T}\right)\right\}^{-\xi}\right] \tag{6}$$

and when $\xi = 0$

$$q_T = \mu - \sigma \log\left\{-\log\left(1 - \frac{1}{T}\right)\right\}. \tag{7}$$

The POT series was constructed by selecting all independent and identically distributed values ($q_1, q_2, \ldots..$) that are higher than a specific, and carefully chosen, value called threshold point (q_o) (see example in Figure 4). According to extreme value theory, for large enough q_o the distribution function of $y = (q - q_o)$ conditioned by $q > q_o$ is approximated by the Generalized Pareto (GP) family as follows (Coles, 2001):

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{8}$$

where $\left\{ y : y > 0 \ \mathit{and} \ \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0 \right\}$, $\mathit{and} \ \tilde{\sigma} = \cancel{\sigma} \quad \xi(q_o - \mu)$, q is a specific value in the sequence ($q_1, q_2, \ldots..$), and $\sigma, \mu,$ and ξ are the scale, location, and shape parameters. Depending on the value of the shape factor, the GP family consists of three probability distribution functions as follows: the heavy tail Pareto type when $\xi > 0$; the light tail Exponential type when $\xi = 0$; and bounded upper tail Beta type when $\xi < 0$. To calculate the extreme quantile (q_T) of the return period *T*, the probability $\zeta_{q_o} = \Pr\ (q > q_o)$ is calculated first and then the return period when $\xi \neq 0$ is given by:

$$q_T = q_o + \left(\frac{\tilde{\sigma}}{\xi}\right)\left[(\zeta_{q_o} T)^{\xi} - 1\right] \qquad (9)$$

When $\xi = 0$, the return level of a period *T* is given by:

$$q_T = q_o + \tilde{\sigma} \log(T \zeta_{q_o}) \qquad (10)$$

<Figure 4 here please>

Threshold Point Choice: We adopted the parameter stabilization method explained by Coles (2001) to choose threshold points used within the POT method. The method is based on fitting the General Pareto distribution across a range of different threshold points. When fitting, the model parameters including the shape parameter (ξ) and scale parameter $(\tilde{\sigma})$ were estimated for each point across the range. The shape parameter should be approximately constant, and the scale parameter should be linear in *q* when the GP distribution is valid above the q_o (Coles, 2001). Figure 4 shows an example of fitting the GP model using the maximum likelihood method over a range of 1 to 40 for the threshold point. As observed, the shape and the reparametrized scale parameters are nearly stable until reaching the point 21. We, therefore, specified the point 21 cms as the threshold point for the POT series of the observed daily streamflow series in this example; and the method was repeated for each model hydrologic realization performed in our study. In order to support our choice of the threshold, we compared our final results of threshold selection using the parameter stabilization method with three Rules of Thumb presented by Scarrott and MacDonald (2012). Using general order statistics convergence properties, methods including the upper 10% rule, square root rule $k_1 = \sqrt{n}$, and $k_2 = n^{2/3} / \log[\log(n)]$ rule were developed (see Scarrott and MacDonald, 2012). Figure 4 shows that our choices compared well to the three methods.

Temporal Independency in the POT Series: The values of the POT series, in the sense of extreme theorem, should admit to the temporal independence condition. By only selecting all values that are higher than the threshold point, we will obviously violate this condition within a streamflow time series. Therefore, to identify and remove the time dependency in the POT series values, de-clustering of the POT series was adopted. The de-clustering was performed by calculating the Extremal Index (θ) as follows (Coles, 2001):

$$\theta = (\text{limiting mean clustering size})^{-1}, \qquad (11)$$

where θ equal to one indicates an independent series. Therefore, the objective was to minimize the size of the clusters until θ reaches one. Our approach was to make manual iteration for each POT series to select the number of threshold deficits, *r*, used to define a cluster. Moreover, to support our independent choices of POT series, we performed the auto-tail dependence function plots for the data series (Reiss and Thomas, 2007) to test the dependency of the events in the series.

Trend Analysis: We analyzed the POT and the AM series with respect to the non-stationarity explained by trend analysis. We used the Mann-Kendall nonparametric test to identify the presence of trends in each independent POT and AM series (Haan, 2002). If the trend was present, we removed the trend from the series, although as will be discussed in the results, very few series exhibited a significant mean trend.

Likelihood Ratio Test: The likelihood ratio test was used to test the null hypothesis of the shape factor (ξ) to be zero. This test is used in statistics to test the goodness of fit of two distributions when one of them is a special case of the other, i.e., nested models (Hogg et al., 2014, Coles, 2001). In our case, the Gumbel distribution is nested within the GEV distribution, and the Exponential distribution is nested within the GP distribution.

Currently, the AM and POT series are the only two types of flood peak series that can be used for flood frequency analysis, and further discussion of a comprehensive comparison between the two series is provided in the literature in Bezak et al. (2014) and Madsen et al. (1997). To perform all the methods described in the extreme analysis methods section and shown in Figure 3, we have applied the R package extRemes version 2.0 described in Gilleland & Katz (2016).

4.2 Uncertainty from climate and extreme modeling factors

Our results from the coupled climate, hydrologic, and extreme modeling methods produced 226 realizations of model runs available for variance analysis based on a factorial design that considered emission type, GCM type and version, downscaling type, bias correction, and extreme value method type. Each factor was divided within variance decomposition as follows: the GCM type factor was divided into four levels for the four parent models mentioned previously; the GCM version factor was divided into two levels indicating CMIP3 and CMIP5 project phases of the models; the downscaling factor was divided into two levels for statistical and dynamical methods; the emission factor was divided into seven levels including the SRES type used in CMIP3 (A1B, A2, and B1) and the RCPs type used in CMIP5 (RCP2.6, RCP4.5, RCP6.0, and RCP8.5); the bias factor was divided into two levels indicating inclusion of methods in Equations (1-4) or lack thereof; and the extreme value factor was divided into two levels for AM and POT methods. Further details of the factorial levels for each of the 226 realizations are provided in the Supplementary On-line Table. We simulated variance analysis following both more traditional linear methods and more recently published nonlinear methods in order to maintain robustness of the analyses.

Linear Analysis of Variance (ANOVA): We performed statistical analysis through fitting the linear analysis of variance model (ANOVA) to the results of the maxima extreme analysis. ANOVA was applied separately for each streamflow maxima quantile. The extreme quantiles represent the response variables of 2-year, 20-year, and 100-year return periods ($\Delta Q_{F\text{-}H(2\text{-}year\text{-}ME)}$, $\Delta Q_{F\text{-}H(20\text{-}year\text{-}ME)}$, and $\Delta Q_{F\text{-}H(100\text{-}year\text{-}ME)}$ respectively) *via* the general linear model-univariate procedure in *SPSS 22* software (Pallant, 2013). ANOVA explores the effect of different factors on the variance of the response using the p-value of the statistical test and ranking factor importance by using the F-value. The F-value of each factor was divided by the summation of F-values in a single model to determine how much variance that factor explains from the total predictable variance. Several considerations were determined when applying ANOVA methods. First, because the datasets were not represented in the climate factors equally, we applied four separate models that balanced a set of factors. The reason for the multiple models is attributed to our climate datasets where the CMIP3 project has both statistically and dynamically downscaled results while the CMIP5 project has only statistically downscaled results. We, also, have different emission scenarios between the two projects. CMIP3 has SRES emission scenarios; and CMIP5 has RCPs emission scenarios. We built therefore four-way ANOVA models as the highest possible order to constrain the balanced and nested models. Figure 5 shows four possible 4-way ANOVA models that we built from our factorial design. Second, we analyzed the factors across the models using the highest possible order, however, if a factor was found to be unimportant, we omitted the factor to maximize repetitions.

<Figure 5 here please>

ANOVA assumes that the population is normally distributed, although the violation of this assumption should not cause major problems when the sample size is greater than 30 (Pallant, 2005, Gravetter&Wallnau, 2000, Stevens, 1996). In our factorial design, the least sample size was recorded in ANOVA model 3, where the sample size was 56. Therefore, our concern about the normality assumption is limited. The homogeneity of variance assumption was treated by using the Levene test for the equality of variance (Pallant, 2005). If the data failed in this test, the significant level by which we compare the variances of the different groups in the ANOVA models was 0.01, which overcomes the violation of this assumption (Pallant, 2005).

Nonlinear Artificial Neural Network (ANN): ANN models, on the other hand, were considered in this study to reinforce our robustness of the variance analysis. ANNs provides a model framework based on a set of multivariate nonlinear functions, and therefore could account for nonlinearity between factors controlling variance and the streamflow response variable, if it exists. In this manner, ANNs could overcome the underlying multivariate linear model limitation that ANOVA is based on. We used the ANN model to examine the climate factors importance on streamflow maxima projections through *SPSS 22* software (IBM , 2012, Tufféry, 2011). The input layer represented the climate and the statistical factors with nominal variables, and the output layer represented the relative change in streamflow maxima. We used one hidden layer with a randomly generated number of neurons. We used supervised training with multilayer perceptron and feedforward architecture. All values of the input and output layers were normalized so that all values ranged between 0 and 1. The hyperbolic tangent activation function was considered in the hidden layer. We used the same four models proposed in the ANOVA analysis to perform the ANN analysis. The dataset partitioning was performed with *SPSS-ANN* to divide the data into training and testing datasets. However, through generation of random numbers within *SPSS-ANN*, the partitioning values of training and testing will swing around the 70% and 30% marks for each run of many runs performed for each model. The values of training partitioning ranged between 60% and 80% affecting the testing portion and providing a new relative error value for both training and testing parts. Accordingly, the smallest relative error provides the best results for the ANN model (IBM, 2012). Therefore, our approach was to use an initial 70% of the dataset for training and the rest for testing, and then rerun the model until obtaining the minimum possible relative errors across the training and testing data.

4.3 Uncertainty from hydrologic modeling

Additional uncertainty from the future hydrologic balance and its simulation were also quantified as part of our study. Future projections of net radiation, vapor pressure and wind speed were tested in simulation for the study region with the premise that decreases in wind speed and net radiation and an increase in relative humidity could decrease future evapotranspiration and in turn increase streamflow maxima while at the same time increase uncertainty of forecasts. Future projections that consider hydrologic model fit and hydrologic parameter uncertainty were also tested to assess the potential to increase the variability of projected streamflow maxima.

Future climate change of wind speed, net radiation, and relative humidity were tested within hydrologic simulation by considering projected shifts reported in the literature. The average monthly wind speed in the study site ranges between 3 and 5 m/s. According to McVicar et al. (2012), the possible stilling in the middle of the current century is approximately 0.5 m/s for the study site region when assuming a linear trend of their observations reported therein. In turn, the percent climate change of wind speed is between -10% and -17% for the future period in the study region. Wild (2009) indicates that the surface solar radiation has a decadal variation and that the absolute trend was observed as -6 W m^{-2} per decade and 8 W m^{-2} per decade for the periods of 1961-1990 and 1995-2007, respectively, over the United States. We recognized that increasing radiation would offset decreasing wind speed when estimating evapotranspiration, and therefore we considered the decreasing trend of -6 W m^{-2} per decade for the future period, in order to test its sensitivity. The mean daily solar radiation ranges throughout the year between 81 W m^{-2} (1.9 kW h $m^{-2}d^{-1}$) and 300 W m^{-2} (7.2 kW h $m^{-2}d^{-1}$). Considering the mentioned net decrease produces a change in the solar radiation reaching the surface to be between -4% and -15%. Regarding the relative humidity, Willett et al. (2008) shows data that suggests an increase in the relative humidity for the northern hemisphere. The net increase shown was 0.07% for the 10 year period of investigation. We assumed the same change for the future period, which resulted in a range between +0.4% and +0.5% for the study region. Donohue et al. (2010) showed that the Penman equation produced the most reasonable estimation of evaporation demand, and this method is included within the hydrologic model used in the present study. Therefore, we considered a number of scenarios in hydrologic modeling that test the mentioned ranges of wind speed, net radiation, and relative humidity concurrently to see their added impact on streamflow maxima. We also tested the variables independently to see their individual sensitivity upon the streamflow maxima.

Future projections that consider hydrologic model fit and hydrologic modeling uncertainty were also tested with the hydrologic model to investigate their impact on forecasted streamflow maxima. Recent literature results have marginalized the importance of model fit when forecasting the relative change in future mean streamflow (Niraula et al., 2015), and we tested this concept for future streamflow maxima. The future streamflow maxima produced from the calibrated hydrologic model simulation for a set of GCM realizations was compared against the future streamflow maxima produced using the un-calibrated (i.e., default) parameterization of the hydrologic model for the same climate realizations. Additionally, the impact of hydrologic model uncertainty was considered by carrying forward uncertainty projections from the hydrologic model parameterization to the extreme value methods and thereafter to compute the relative change in future streamflow. The SWAT-CUP software provides parameter sets and solutions used to create uncertainty bounds during the model simulation. Realizations of all parameter sets that meet the objective function criteria were chosen and extreme value methods were performed for hindcast and forecast global climate pairs to compute the relative change in streamflow maxima.

4.4 Forecast of streamflow maxima for wet temperate regions

After quantifying the climate, hydrologic, and extreme modeling factors controlling variability of the projections, an ensemble was created to forecast the relative change in the streamflow maxima for the wet temperate study region (Al Aamery et al., 2016). The extreme forecasts for this study calculated the net effect on the mean and variance of the balanced ensemble from variation of climate modeling factors and extreme modeling factors, the added uncertainty from hydrologic model parameterization, and the added mean shift and its variance from climate change shifts in net radiation, vapor pressure and wind speed. Results were compared with other studies reported in the literature of streamflow maxima (see Table 1) that fell within wet temperate regions.

5 RESULTS AND DISCUSSION

5.1 Modeling simulations and evaluation

Results from the model evaluation showed that the hydrologic model performed within an acceptable range, and the simulated and observed daily streamflow signals showed close agreement (see Figure 6). The four quantitative matrics including coefficient of determination (R^2), percent bias (*PBIAS%*), Nash-Sutcliff Efficiency (*NS*), and the ratio of the root mean square error to the standard deviation of measured data (*RSR*) showed results within the acceptable range (Moriasi et al., 2007, Donigan, 2002, Gassman et al., 2007) in both calibration and validation periods for the majority of the four observation sites for which the model was compared against (see compiled metrics in Table 2), although one of the four sites showed values just below or equal to the acceptable range boundary during validation. Overall, 53 out of the 56 metrics that compared observations with model results were above the acceptable range showing that the model simulated streamflow well. According to Moriasi et al. (2007) the monthly time step model performance is considered satisfactory if the $NS>0.5$, $RSR<0.7$, and $PBIAS <\pm25\%$. The model performance on finer time steps (e.g. daily) is usually poorer than the coarser time steps model (e.g. monthly) in terms of the statistical matrices (e.g. *NS*, *RST*, *PBIAS*) (Moriasi et al., 2007, Engel et al., 2007). For instance, while the monthly *NS* was 0.656 for the calibration period in Fernandez et al. (2005), the daily one was 0.395. Moreover, Moriasi et al. (2007) indicated that when reviewing previous studies, *NS* and *PBIAS* were "as expected" lower in the validation period than the calibration period for streamflow. Note that the Midway station was our primary calibration, since all of the model's streamflow forecasts occurred from this location. The model we established for South Elkhorn watershed showed results for the Midway gage station to have *NS* values equal to 0.9 and 0.66 for the monthly and daily time steps, respectively, for the calibration period; and *NS* values equal to 0.88 and 0.46 for the monthly and daily time steps, respectively, for the validation period. In summary, the metrics showed adequate performance considering the above information and results.

<Figure 6 here please>

<Table 2 here please>

Results from fitting both the AM and POT extreme series methods to the streamflow results showed that in general the extreme series results had little mean trend and were dominated by the two parameter probability distributions (see Supplementary On-line Table). The Mann-Kendall test results showed that only 2% of the AM series included a mean trend that required removal and only 4% of POT series results had a mean trend that required removal. A regression approach was also carried out and provided identical results as the Mann-Kendall tests. The results highlight that although non-stationarity is exhibited when comparing extremes from the hindcast to the forecast periods, little significant non-stationarity is exhibited within the simulation periods. Statistical results showed that 91% of the AM series best followed the two-parameter Gumbel distribution while 85% of the POT series best followed the exponential distribution. The results tend to agree with the results of Dankers and Feyen (2008) who also found that a two parameter distribution was most adequate when fitting distributions from extreme value theory to streamflow results derived from global climate modeling. Additional results from the extreme value analyses is also compiled in the Supplemental On-line Table and includes: threshold selections, the value of the extremal index θ before de-clustering, the value of r required to make the extremal index θ equal to unity, the p-value of Mann-Kendall non-parametric test, and the resultant sample size (n).

We found less than 10% difference between observed and simulated maxima for all return periods (i.e., 2, 20 and 100 year return periods) for both AM and POT methods. Both observed and simulated maxima followed exponential distributions for the POT method; and both followed the Gumbel distribution for the AM method. Donigan (2002) indicates that an absolute hydrologic model calibration/validation target of less than 10% difference between the simulated and the observed hydrology flow is considered a very good target; and that the range of such target should be applied on the mean and the individual events may show larger differences while still acceptable. With this criteria in mind, our SWAT evaluation results for the extremes were deemed adequate.

Extreme quantiles for 2-, 20-, and 100-year maxima streamflow levels showed that forecast results were in general greater than hindcast results for simulation pairs with the same climate modeling factors, highlighting the non-stationarity of extremes mentioned previously. Figure 7 illustrates hindcast simulations corresponding to POT extreme series method, and all simulation results are shown in the Supplementary On-line Table. Statistical downscaling of the hindcast GCM realizations in general under-predicted hydrologic model results analyzed with the extreme series method; and the under-prediction was especially true for streamflow levels from the 100-year return period. Results from the dynamical downscaling hindcast realizations better bound the observed extremes. The result supports the idea that regional climate models can capture small-scale climate features, e.g., strong fronts, and realistically simulate extreme events (Fowler et al., 2007, Warner, 2010), which would suggest a better choice for extreme streamflow forecasting. Fowler et al. (2007) pointed out that the statistical downscaling methods poorly represent the extreme events and underestimate variance, which reflects the fact that both BCSD and BCCA methods use the distribution of precipitation from historical climate records to create the future distributions. Warner (2010) compared the statistical and dynamical downscaling with respect to their advantages and disadvantages, and he indicated that dynamical downscaling methods could better capture extreme events and variance. Sunyer et al. (2015) shows that the RCM-GCM projections are the main source of variability in their results, and between 30-50% of the total variance is explained by statistical downscaling in several catchments in their study. Trayhorn and DeGaetano (2001) compared several different downscaling methods for rainfall extremes over the Northeastern United States; and their results suggest that regional climate models overestimate the observed extremes. Aside from the Trayhorn and DeGaetano (2001) results, literature results and this study generally support the idea that hindcast extremes from dynamic downscaling agree better with observed extremes as compared to statistical downscaling results.

<Figure 7 here please>

We also examined specific results of individual climate models and downscaling methods in order to provide insight on how climate model structure may be impacting forecasted streamflow maxima. The four GCMs from CMIP3 all illustrate differences when comparing across the 2, 20 and 100 year return periods (Figure 7). The result was not surprising given that GCM has been found as a significant factor in studies of forecasted mean streamflow and precipitation, and climate scientists highlight variability of GCMs due to the differences in the models' structures and parameterizations (Randall et al., 2007; Weart, 2010; Mearns et al., 2013; Melillo et al., 2014; Al Aamery et al., 2016). Given the many differences between the four GCMs, it is difficult to discern specific processes represented within the climate models that might be controlling the extreme streamflow forecasts, however, direct comparison of CMIP3 and CMIP5 model versions provided some discussion.

Figure 7 reveals that CCSM has a pronounced difference between CMIP3 and CMIP5 forecasted streamflow maxima while the other GCMs (Had, GFDL and CGCM) do not show differences between model versions for our analyses. The reason is perhaps attributed to the newer version CCSM4 that produces El Nino-Southern Oscillation (ENSO) variability in a more realistic frequency distribution than CCSM3 by changing the deep convection scheme.

The Had, GFDL and CGCM models also made changes from CMIP3 to CMIP5 but these tend to have little differences in terms of streamflow extremes (Figure 7). The HadGEM2 of CMIP5 improved the performance of ENSO, northern continent land-surface temperature biases, SSTs, and wind stress compared to the previous models; however, Collins et al. (2008) suggests that the power spectrum of El Nino was not a substantial improvement. GFDL version 3 (CM3) used in CMIP5 made minimal changes to the ocean and sea ice models compared to those used in CM2.1 version of CMIP3; however, the newer version is extensively developed the atmosphere and land model components (Griffies et al., 2011). CanESM2 of CMIP5 combines the fourth generation atmospheric general circulation model (CanCM4) with terrestrial carbon cycle model (CTEM). Compared to the third generation of CanCM3 that was used in CGCM3.1 of CMIP3, CanCM4 is different in many aspects such as the finer resolution and the addition of new schemes such as shallow convection scheme (see Chylek et al., 2011).

Taken together, of all the changes to the four different GCMs between CMIP3 and CMIP5, only augmenting ENSO within the GCM seems to have a substantial impact on forecasted streamflow maxima. The suggestion is reasonable given that ENSO has been suggested to show significant impacts on precipitation in this region of North America (Gabler et al., 2009). Results suggest that the El Nino-Southern Oscillation and its representation within climate modeling may exhibit a substantial control on forecasting streamflow maxima for the wet temperate study region; and additional emphasis upon oscillations when forecasting streamflow maxima in wet temperate regions may be fruitful.

5.2 Uncertainty from climate and extreme modeling factors

Variance analysis results determined *via* ANOVA showed that the variance structure of forecasted streamflow maxima exhibits some dependence on all of the climate modeling considered factors but does not exhibit dependence upon the extreme value method applied (see Figure 8). The results are interesting due the fact that previous variance analysis of mean streamflow forecasted from GCMs only showed dependence on a subset of the climate modeling factors while debate in the literature suggests that AM and POT methods would give different results (Scarrott and MacDonald, 2012; Bezak et al., 2014; Al Aamery et al., 2016). Specifically, results of the ANOVA (Figure 8) show that variance of the 2 year and 20 year streamflow maxima are significantly dependent upon GCM type, downscaling method, emission scenario, GCM project phase, and bias implementation; and variance of the 100 year streamflow maxima is significantly dependent upon GCM type, GCM project phase, and bias implementation. For reference, results of forecasted mean streamflow are included in Figure 8 and show dependence on GCM type and phase and downscaling.

<Figure 8 here please>

The climate modeling factors that significantly influenced the forecasted streamflow maxima variances were ranked using the weighted F-value according to their variance contribution (see Figure 8) as GCM type, downscaling method, bias implementation, GCM version associated with the climate project phase, and the emission scenario input to the GCM. Results of the ANN non-linear variance analysis compared well with linear analysis *via* ANOVA (see comparisons in Figure 9) providing further confidence in our ranking results.

<Figure 9 here please>

In addition to the variance breakdown, the total variance of the forecasted extremes also displays pertinent information. The total variance of streamflow extremes increased substantially with return period—a result most easily observed with the standard error bars in Figure 10. In addition, the proportion of the variance that was predictable with the climate modeling factors tended to decrease with return period. The result suggests a propagation of unexplainable variance throughout the analysis that becomes more pronounced with the higher order extremes associated with higher return periods.

<Figure 10 here please>

We at least partially attribute the pronounced growth of uncertainty with return period to fitting the extreme value distributions to the hydrologic results. The 100 year return period falls at the tail end of the GEV and GP distributions (i.e., f=0.99) and therefore uncertainty introduced in fitting the distributions will be most pronounced for the highest return periods. To illustrate the point, we performed sensitivity of the extreme value parameterization method by assuming a known parent Gumbel distribution for M_n, drawing sets of realizations consistent with the years of data in our analyses, and fitting the extreme value distribution consistent with the maximum likelihood method of our analyses as well as typically performed by others (e.g., Gilleland and Katz, 2006). Results from the sensitivity show that the variance associated with the 100 year streamflow is about five times greater than that of the 2 year streamflow event (see Table 3). The result highlights one reason for pronounced increases in unexplainable variance within forecasted streamflow maxima.

<Table 3 here please>

On the other hand, factorial comparison between the AM and POT series fitted by the General Extreme Value (GEV) and General Pareto (GP) distributions did not show significance within the analysis of variance results. The result is surprising given recent debate and critique of each method, e.g., AM is criticized for its neglect of multiple extremes per annum while POT has been criticized for subjectivity of threshold selection (Svensson et al., 2005; Scarrott and MacDonald, 2012; Bezak et al., 2014; Fischer and Schumann, 2016). However, further investigation of the literature suggests that the variance analysis result is consistent with fundamental theory and that the methods might be used interchangeably, as needed, so long as care is taken in their application. Fundamentally, Coles (2001) shows that the GEV distribution provides the base that can be used to derive the GP distribution so long as the threshold point is sufficiently large and the events are independent and random. In this manner, we recommend that future coupled hydrologic and climate research studies that apply the POT method should strive for relatively high threshold values that fall within the Rules of Thumb outlined by Scarrott and MacDonald (2012) and ensure that the extremal index is not less than one (see Figure 3).

One noteworthy comparison of the present study's results with previously published results is that the variance of forecasted streamflow maxima is even more sensitive to climate modeling factors as compared to the variance of mean forecasted streamflow. Specifically, the variance of streamflow maxima showed significant dependence upon the choice of emission scenario and bias correction approach (see Figure 8) while the variance of mean streamflow did not exhibit significant dependence upon emission and bias (see Al Aamery et al., 2016 and results summarized in Figure 8). The streamflow maxima's dependence upon emission scenario is worthy of mentioning given that the mean atmospheric CO_2 concentration projected for the emission scenarios varies by just ±50 ppm for 2050 (IPCC, 2001; Meinshausen et al., 2011).

Further, the mean annual temperature has a total range of just 1.5°C for 2050 across emission scenarios projected within the GCMs applied in this study and the mean streamflow study of Al Aamery et al. (2016). The subtle mean changes in CO_2 and MAT for 2050 appear to mask temporal anomalies captured within the GCMs. The potential of emissions to help control streamflow maxima is somewhat corroborated by the work of Mantua et al. (2010) where they show streamflow maxima differences among two emission scenarios. Significance of emission scenario within variance analysis of forecasted streamflow maxima suggests that hydrologic and climate research is needed that examines how models might be coupled at a higher temporal resolution, rather than the more prevalent emphasis on mean coupling (e.g., see review Table 1 in Al Aamery et al., 2016). Similarly, the significance of bias correction upon the variance of forecasted streamflow maxima reflects the boundary between climate and hydrologic models that has emphasized mean coupling and thus linear shifts in rainfall and temperature data to show agreement with observations (Lenderink et al., 2007). More sophisticated bias correction methods are available (Teutschbein and Seibert, 2012) but typically come with the added conundrum of forcing functional constraints on climate model results that are sought after due to their non-stationarity. Surely, research might consider higher resolution model coupling to understand anomalies that control maxima streamflow.

5.3 Uncertainty from hydrologic modeling

Future climate change of wind speed, net radiation, and relative humidity were tested within hydrologic simulation by considering projected shifts reported in the literature. Results suggest that the net impact of wind speed, net radiation, and relative humidity could provide an additional 1 to 5% increase in streamflow maxima for 2, 20 and 100 year return periods for the wet temperate study region and future period considered (see Table 4). Average daily change in evapotranspiration ranged from 0.5 to 5% decreases. Streamflow maxima increases and standard error associated with the wind speed, radiation and relative humidity shifts were +3.2(±1.7), +2.2(±1.6) and +1.9(±1.6)% for 2, 20 and 100 year events. Relative to the increases of +27(±21), +36(±34) and +49(±85)% for streamflow maxima associated with GCM-projection of precipitation and temperature from ensemble analysis (see Figure 10), the effect of wind speed, net radiation and relative humidity were small for this region. Nevertheless, the effect is non-zero; and the variables may be more substantial in other regions or for forecasting to 2100.

<Table 4 here please>

Future projections that considered hydrologic model fit and hydrologic modeling uncertainty were also tested to investigate their impact on the relative change of streamflow maxima. The future streamflow maxima produced from the calibrated hydrologic model simulation was compared against the future streamflow maxima produced using the un-calibrated (i.e., default) parameterization of the hydrologic model for the realization pairs for the AM extreme value analysis method (*n*=74). Results for the uncalibrated hydrologic analysis of the relative change in streamflow maxima were +19(±28), +20(±35) and +24(±59)% for 2, 20 and 100 year events in comparison to the calibrated model results equal to +27(±23), +35(30) and +49(±92)% for 2, 20 and 100 year events. Results show that the uncalibrated model gives a much lower increase in future streamflow maxima compared to the calibrated model results, especially for the 100 year extreme. Note that the default model simulations tended to under-predict streamflow during peak events. The simulation bias is carried forward to the extreme modeling results and is not removed when considering the relative change. In this manner, the variance of the streamflow maxima was dependent on hydrologic model parameterization. These results contrast the work of Niraula et al. (2015) where we showed that the relative change in mean forecasted streamflow was not dependent on parameter selection during calibration. The results further highlight the variance structure's sensitivity when forecasting streamflow extremes.

Given the dependence on hydrologic calibration, the hydrologic uncertainty realizations were also performed. Results suggest that hydrologic model parameter sets generated during uncertainty analysis also impart variance upon relative changes in streamflow maxima. We calculated the error associated with the relative change in streamflow maxima using the parameter sets within SWAT-CUP that met model objective function criteria. Standard error was 3.1, 3.3 and 3.6% for the relative change of 2, 20 and 100 year events. Standard error is small in comparison to the error produced from climate and extreme modeling factors. Nevertheless the error is nonzero and may be larger for other regions. We also calculated the standard error from absolute forecasted streamflow maxima and found values of 11, 21, and 27 cms for 2, 20 and 100 year events. We compared these values with the standard error from direct bias-correction of the streamflow maxima via the relative change approach, and the standard error was 3, 6 and 9 cms for 2, 20 and 100 year events. The results highlight that the delta method applied to the direct observed streamflow via the relative change does reduce hydrologic uncertainty relative to the absolute forecasts.

5.4 Forecast of streamflow maxima for wet temperate regions

One corollary of variance analysis is inclusion of significant factors impacting prediction and thus forecasting of future streamflow. The relative change in streamflow maxima were increases of +30(±21), +38(±34) and +51(±85)% for the study region for 2, 20 and 100 year events. The increases are substantially larger than the 11% increases found for mean streamflow and mean precipitation for the study region (Al Aamery et al., 2016). Additionally, streamflow maxima increases as a function of return period. The variability of the projections is pronounced, and the uncertainty from climate and extreme model factors dominates the variance (see Table 5).

<Table 5 here please>

The forecasted results of increased maxima streamflow in 2050 for the wet temperate region of North America (1120 mm y^{-1}) is in agreement with scientific sentiment and forecasting that wet regions will get wetter and wet time periods will be wetter (Melillo et al., 2014). We performed analysis of published maxima streamflow forecasts in wet regions of Europe and their comparison corroborated the finding that maxima streamflow increases as a function of return period. Analysis of the results from Lawrence and Hisdal (2011) show an increase of maxima streamflow as a function of return period for Norway (760-2250 mm y^{-1}). Also, analysis of the results from Dankers and Feyen (2008) show an increase of maxima streamflow as a function of return period for their European sites studied where the mean annual precipitation was greater than 500 mm per year and is projected to be less in the end of this century.

The finding that forecasted maxima streamflow may show further increases as a function of return period further supports general scientific agreement that the most extreme flooding events will get even more extreme for wet temperate climates (Melillo et al., 2014). This concept is reflected in the timing of streamflow increases and extremities in the present study, and Table 6 shows that the months of the year with the highest future changes in mean precipitation and streamflow tend to also account for the majority of forecasted streamflow maxima events during the study period. The results also reflect the fundamental scientific consequences of climate change. That is, increased precipitation in wet regions is expected due to higher amounts of moisture in the atmosphere due to warmer atmospheric temperatures and expansion of the high Sub-tropical Belt as the air temperature increases and moist air is transported to higher and lower latitudes (Gabler et al., 2009; Melillo et al., 2014). In turn, climate change in wet temperate region may increase precipitation, temperature, and relative humidity while decreasing wind speed and net radiation, and the net effect both individually and cumulatively of all these shifts is an increase in streamflow maxima.

<Table 6 here please>

6 CONCLUSION

The main conclusions of our work are described as follows:

(1) Model simulation and evaluation results from comparison of different global climate model downscaling methods suggests that dynamic downscaling results more closely align with observations, presumably due to the explicit simulation of small-scale features such as strong fronts. Comparison of streamflow maxima forecasted with paired climate models from CMIP3 versus CMIP5 projects suggest that the El Nino-Southern Oscillation representation within modeling exhibits a control on forecasting streamflow maxima for the wet temperate region studied.

(2) Uncertainty from climate and extreme modeling factors was evaluated and showed that the relative change of streamflow maxima was not dependent on systematic variance from the annual maxima versus peak over threshold method applied. We find that the variance of streamflow maxima is an increasing function of the return period, which is at least partly attributed to fitting the extreme value distributions to the hydrologic model results. The variance of the relative change in streamflow maxima is dependent upon global climate model, emission scenario, project phase, downscaling, and bias correction.

(3) Uncertainty from hydrologic modelling was analyzed and unlike results from previous research focused on the relative change of mean streamflow, the relative change of streamflow maxima was dependent on hydrologic model fit and modeling uncertainty. The streamflow maxima also showed some dependence on climate projections of wind speed, net radiation and relative humidity.

(4) Ensemble projections forecast an increase of streamflow maxima for 2050 with pronounced forecast standard error, including +30(±21), +38(±34) and +51(±85)% for 2, 20 and 100 year

events for the wet temperate region studied. The variance of maxima projections was dominated by climate model factors and extreme value analyses with lesser control from hydrologic inputs and hydrologic model parameterization.

ACKNOWLEDGEMENTS:

We thank Editor McVicar, Associate Editor Yongqiang Zhang, and anonymous reviewers for their excellent input and comments, which in turn helped improve the quality of this paper. We acknowledge partial support from NSF Award Number 1632888 to assist with study of the watershed and river system.

REFERENCES:

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., & Srinivasan, R. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of hydrology*, *333*(2), 413-430.

Al Aamery, N., Fox, J. F., & Snyder, M. (2016). Evaluation of climate modeling factors impacting the variance of streamflow. *Journal of Hydrology*, *542*, 125-142.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73-89.

Arnold, J. G., Allen, P. M., Volk, M., Williams, J. R., & Bosch, D. D. (2010). Assessment of different representations of spatial variability on SWAT model performance. *Transactions of the ASABE*, *53*(5), 1433-1443.

Beguería, S., & Vicente-Serrano, S. M. (2006). Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of applied meteorology and climatology*, *45*(1), 108-124.

Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, *59*(5), 959-977.

Brekke, L., Thrasher, B. L., Maurer, E. P., & Pruitt, T. (2013). Downscaled CMIP3 and CMIP5 climate projections. *US Dep. Inter. Bur. Reclamation, Tech. Serv. Cent*.

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part II: Preliminary model validation. *Monthly Weather Review*, *129*(4), 587-604.

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, *401*(3), 190-202.

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian earth system model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22893-22907.

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). London: Springer.

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Coney, S.C., Hack, J.J., Henderson, T.B. and Kiehl, J. T. (2006). The community climate system model version 3 (CCSM3). *Journal of Climate*, *19*(11), 2122-2143.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Hinton, T., Jones, C. D., ... & Senior, C. (2008). Evaluation of the HadGEM2 model. *Hadley Cent. Tech. Note*, *74*.

Dankers, R., & Feyen, L. (2008). Climate change impact on flood hazard in Europe: An assessment based on high- resolution climate simulations. *Journal of Geophysical Research: Atmospheres*, *113*(D19).

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., ... & Durachta, J. W. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643-674.

Donigan, A. S. (2002). Watershed model calibration and validation: The HSPF experience. *Proceedings of the Water Environment Federation*, *2002*(8), 44-73.

Donohue, R. J., McVicar, T. R., & Roderick, M. L. (2010). Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *Journal of Hydrology*, 386(1-4), 186-197.

Engel, B., Storm, D., White, M., Arnold, J., & Arabi, M. (2007). A hydrologic/water quality model applicati1. *Journal of the American Water Resources Association*, 43(5), 1223-1236.

Fatichi, S., Rimkus, S., Burlando, P., & Bordoy, R. (2014). Does internal climate variability overwhelm climate change signals in streamflow? The upper Po and Rhone basin case studies. *Science of the Total Environment*, *493*, 1171-1182.

Fernandez, G. P., Chescheir, G. M., Skaggs, R. W., & Amatya, D. M. (2005). Development and testing of watershed-scale models for poorly drained soils. *Transactions of the ASAE*, 48(2), 639-652.

Fischer, S., & Schumann, A. (2016). Robust flood statistics: comparison of peak over threshold approaches based on monthly maxima and TL-moments. *Hydrological Sciences Journal*, *61*(3), 457-470.

Flato, G. M. (2005). The third generation coupled global climate model (CGCM3). *Available on line at http://www. cccma. bc. ec. gc. ca/models/cgcm3. shtml*.

Ford, W. I., Fox, J. F., & Rowe, H. (2015). Impact of extreme hydrologic disturbance upon the sediment carbon quality in agriculturally impacted temperate streams. *Ecohydrology*, *8*(3), 438-449.

Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International journal of climatology*, *27*(12), 1547-1578.

Gabler, R. E., Peter, J. F., Trapson, M., & Sack, D. (2009). Physical Geography: Brooks/Cole. *Belmont, USA*.

Gassman, P. W., Reyes, M. R., Green, C. H., & Arnold, J. G. (2007). The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASABE*, *50*(4), 1211-1250.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., ... & Worley, P. H. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973-4991.

Gilleland, E., & Katz, R. W. (2016). Extremes 2.0: an extreme value analysis package in r. *Journal of Statistical Software*, *72*(8), 1-39.

Giorgi, F., Marinucci, M.R., Bates, G.T. (1993). Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. *Mon. Weather Rev. 121*, 2794–2813.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., ... & Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate dynamics*, *16*(2), 147-168.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th edition). Belmont, CA: Wadsworth.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., ... & Palter, J. B. (2011). The GFDL CM3 coupled climate model: characteristics of the ocean and sea ice simulations. *Journal of Climate*, *24*(13), 3520-3544.

Haan, C. T. (2002). *Statistical methods in hydrology*. The Iowa State University Press.

Harding, B. L., Wood, A. W., & Prairie, J. R. (2012). The implications of climate change scenario selection for future streamflow projection in the Upper Colorado River Basin. *Hydrology and Earth System Sciences*, *16*(11), 3989.

Hogg, R. V., Tanis, E., & Zimmerman, D. (2014). *Probability and statistical inference*. Pearson Higher Ed.

IBM.(2012). IBM SPSS Neural Networks 21. Retrieved from http://www.sussex.ac.uk/its/pdfs/SPSS_Neural_Network_21

IPCC, 2001. Climate Change 2001: The Scientific Basis. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. (Eds.), Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 881 pp.

IPCC, 2007. Climate Change 2007: The Physical Science Basis. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996 pp.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 1535 pp.

Juang, H., Hong, S., Kanamitsu, M. (1997). The NMC nested regional spectral model: an update. *Bull. Am. Meteorol. Soc. 78*, 2125–2143.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of hydrology*, *329*(3), 534-552.

Knutti, R., & Sedláček, J. (2013). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, *3*(4), 369-373.

Lawrence, D., & Hisdal, H. (2011). Hydrological projections for floods in Norway under a future climate. *NVE Report*, *5*.

Lenderink, G., Buishand, A., & Deursen, W. V. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, *11*(3), 1145-1159.

Leopold, L. B., Wolman, M. G., & Miller, J. P. (2012). *Fluvial processes in geomorphology*. Courier Corporation.

Lima, C. H., Lall, U., Troy, T. J., & Devineni, N. (2015). A climate informed model for nonstationary flood risk prediction: Application to Negro River at Manaus, Amazonia. *Journal of Hydrology*, *522*, 594-602.

Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At- site modeling. *Water resources research*, *33*(4), 747-757.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, *519*, 3634-3650.

Mantua, N., Tohver, I., & Hamlet, A. (2010). Climate change impacts on streamflow extremes and summertime stream temperature and their possible consequences for freshwater salmon habitat in Washington State. *Climatic Change*, *102*(1), 187-223.

McMichael, A. J., Powles, J. W., Butler, C. D., & Uauy, R. (2007). Food, livestock production, energy, climate change, and health. *The lancet*, *370*(9594), 1253-1263.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., ... & Mescherskaya, A. V. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, 416, 182-205.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., ... & Thomson, A. G. J. M. V. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change*, *109*(1-2), 213.

Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe, Eds., 2014: *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.

Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., ... & Snyder, M. (2013). Climate change projections of the North American regional climate change assessment program (NARCCAP). *Climatic Change*, *120*(4), 965-975.

Mirza, M. M. Q. (2003). Climate change and extreme weather events: can developing countries adapt?. *Climate policy*, *3*(3), 233-248.

Moradkhani, H. (2017, May). *State-of-the-art Uncertainty Analysis in Hydroclimate Modeling Panel Discussion.* Presentation session presented at the meeting of the World Environmental& Water Resources Congress, Sacramento, CA.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

Niraula, R., Meixner, T., & Norman, L. M. (2015). Determining the importance of model calibration for forecasting absolute/relative changes in streamflow from LULC and climate changes. *Journal of Hydrology*, 522, 439-451.

Palanisamy, B., & Workman, S. R. (2014). Hydrologic modeling of flow through sinkholes located in streambeds of Cane Run Stream, Kentucky. *Journal of Hydrologic Engineering*, *20*(5), 04014066.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).

Plummer, D. A., Caya, D., Frigon, A., Côté, H., Giguère, M., Paquin, D., ... & De Elia, R. (2006). Climate and climate change over North America as simulated by the Canadian RCM. *Journal of Climate*, *19*(13), 3112-3132.

Prudhomme, C., Jakob, D., & Svensson, C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. *Journal of hydrology*, *277*(1), 1-23.

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: Cilmate 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.

Reiss, R. D., Thomas, M., (2007). *Statistical analysis of extreme values* (Vol. 2). Basel: Birkhäuser.

Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2001). Climate change and extreme weather events; implications for food production, plant diseases, and pests. *Global change & human health*, *2*(2), 90-104.

Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal*, *10*(1), 33-60.

Shamir, E., Megdal, S. B., Carrillo, C., Castro, C. L., Chang, H. I., Chief, K., ... & Prietto, J. (2015). Climate change and water resources management in the Upper Santa Cruz River, Arizona. *Journal of Hydrology*, *521*, 18-33.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J. G. (2005). A description of the advanced research WRF version 2. *NCAR Tech. Note NCAR/TN-468+STR*, 88 pp.

Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences* (3rd edition). Mahway, NJ: Lawrence Erlbaum.

Sunyer Pinya, M. A., Hundecha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., ... & Loukas, A. (2015). Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrology and Earth System Sciences*, *19*(4), 1827-1847.

Svensson, C., Kundzewicz, W. Z., & Maurer, T. (2005). Trend detection in river flow series: 2. Flood and low-flow index series/Détection de tendance dans des séries de débit fluvial: 2. Séries d'indices de crue et d'étiage. *Hydrological Sciences Journal*, *50*(5).

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456*, 12-29.

Tryhorn, L., & DeGaetano, A. (2011). A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. *International Journal of Climatology*, *31*(13), 1975-1989.

Tufféry, S. (2011). *Data mining and statistics for decision making* (Vol. 2). Chichester: Wiley.

Warner, T. T. (2010). *Numerical weather and climate prediction*. Cambridge University Press.

Weart, S. (2010). The development of general circulation models of climate. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *41*(3), 208-217.

Wilby, R. L., & Dawson, C. W. (2007). SDSM 4.2-A decision support tool for the assessment of regional climate change impacts. *User Manual. London, UK*.

Wild, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres*, 114(D10).

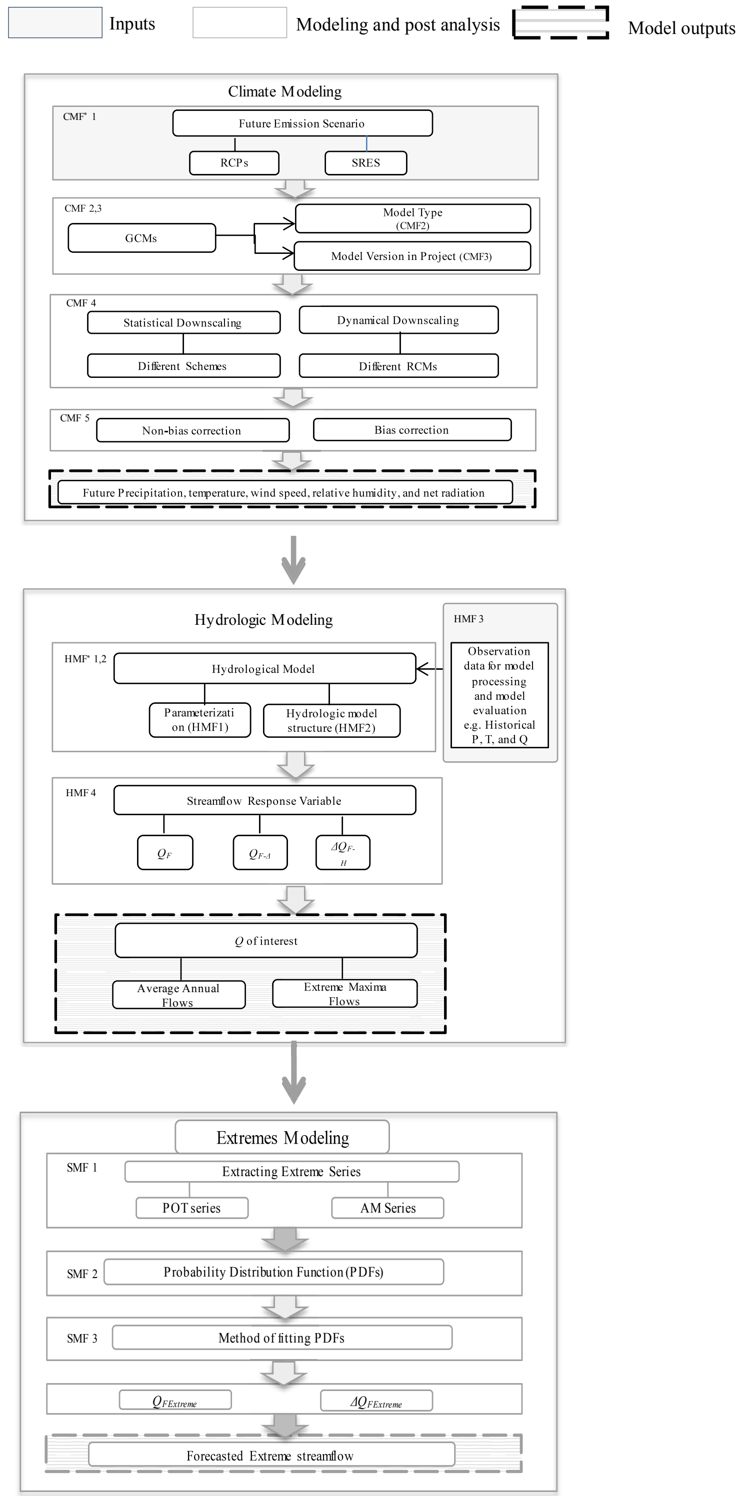
Willett, K. M., Jones, P. D., Gillett, N. P., & Thorne, P. W. (2008). Recent changes in surface humidity: Development of the HadCRUH dataset. *Journal of Climate*, 21(20), 5364-5383.

Yuan, X., & Wood, E. F. (2012). Downscaling precipitation or bias- correcting streamflow? Some implications for coupled general circulation model (CGCM)- based ensemble seasonal hydrologic forecast. *Water Resources Research*, *48*(12).

Zhang, X., Xu, Y. P., & Fu, G. (2014). Uncertainties in SWAT extreme flow simulation under climate change. *Journal of hydrology*, *515*, 205-222.

Figure 1

Figure 1. Conceptual model of variance structure for forecasted streamflow maxima.



*CMF : Climate Modeling Factor *HMF: Hydrology Modeling Factor *SMF: Statistical Modeling Factor

Figure 2

Figure 2. Study area of the South Elkhorn Watershed, Kentucky USA (adopted from Figure 2 in Al Amery et al., 2016)

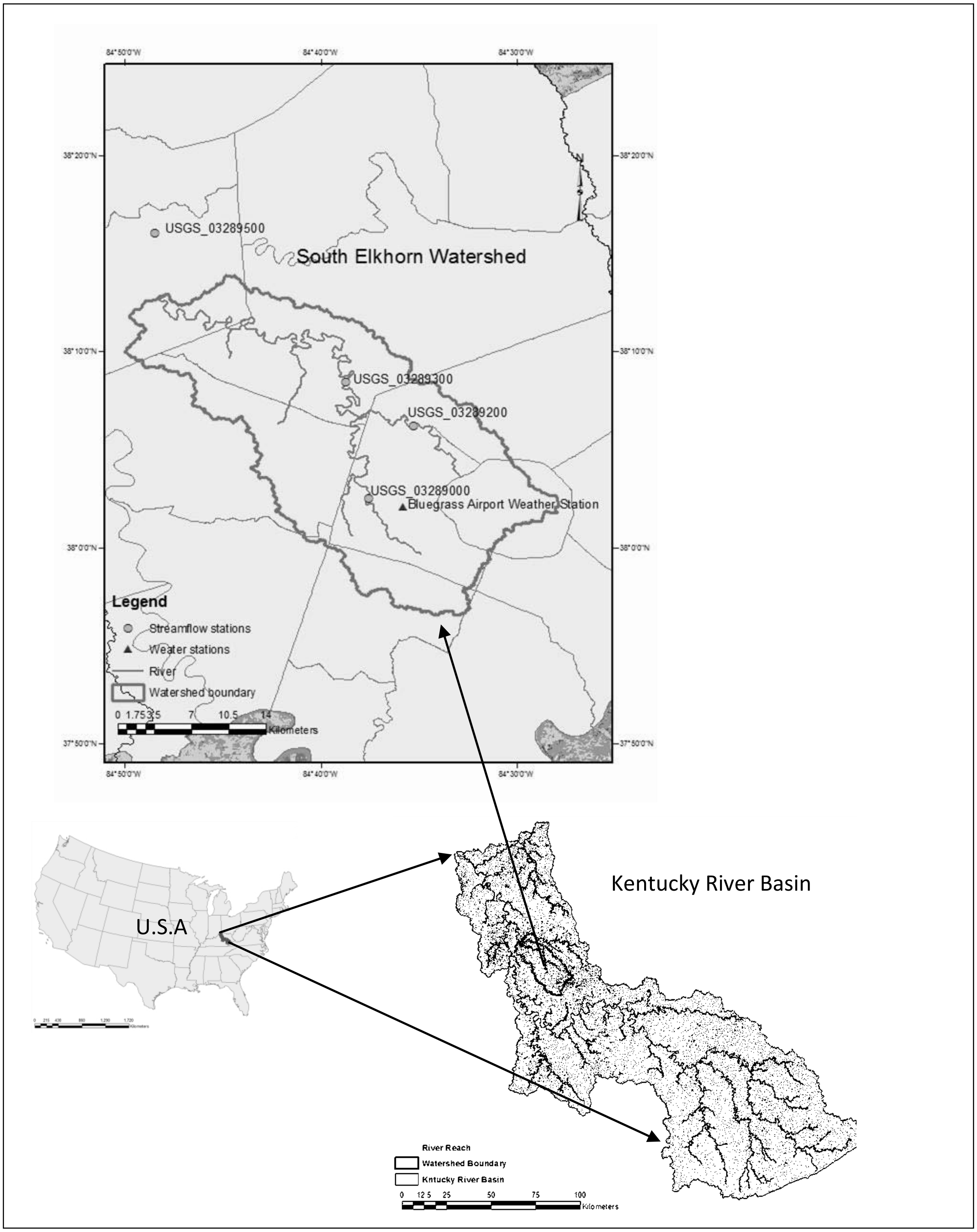


Figure 3

Figure 3. Methodology of extreme value analysis.

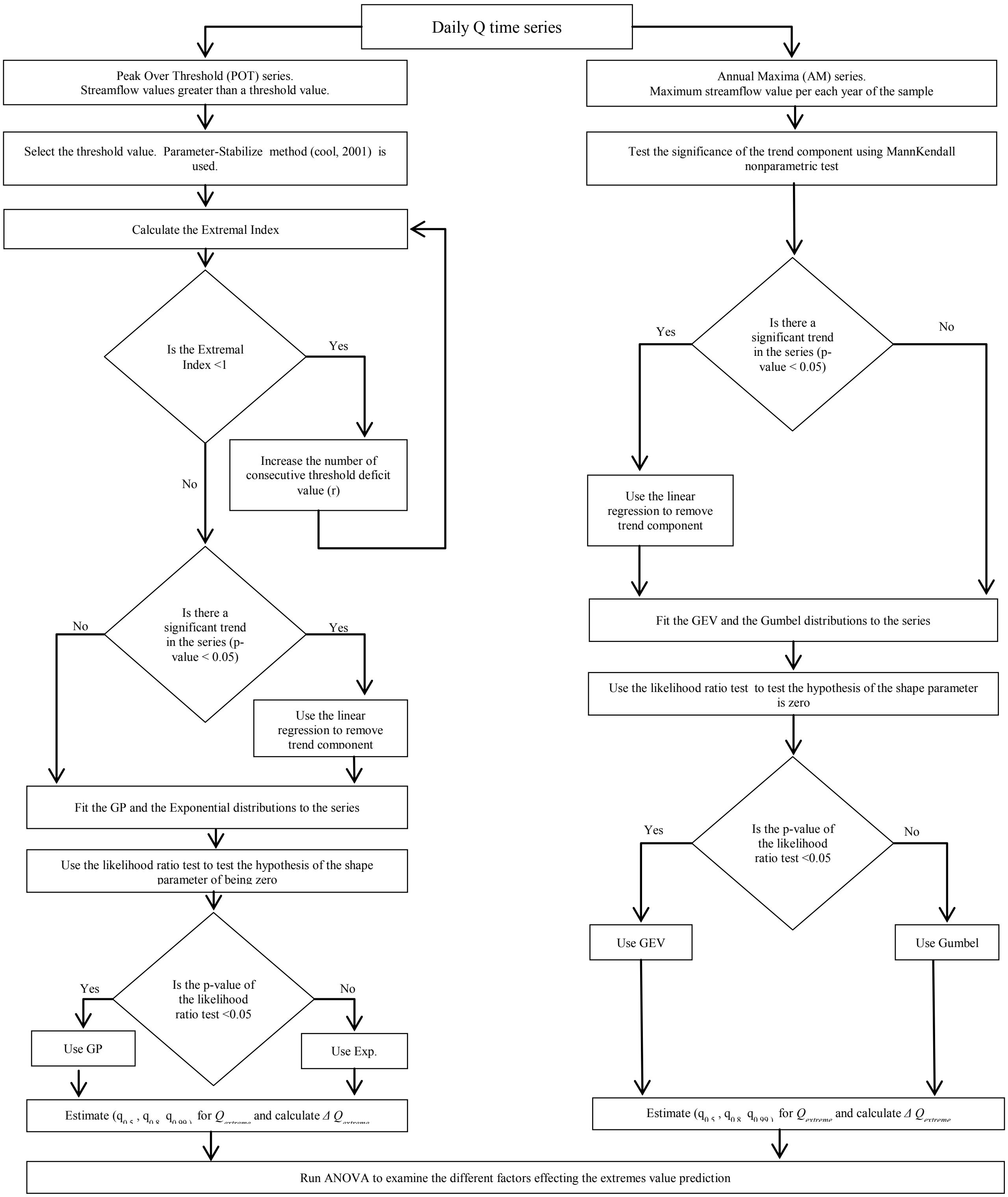


Figure 4

Figure 4. (a) AM series example, (b) POT series example, (c) Parameter stabilization drawing for observed *Q*, and (d) Threshold choices, comparison of three Rule of Thumb methods and parameter stabilization method.

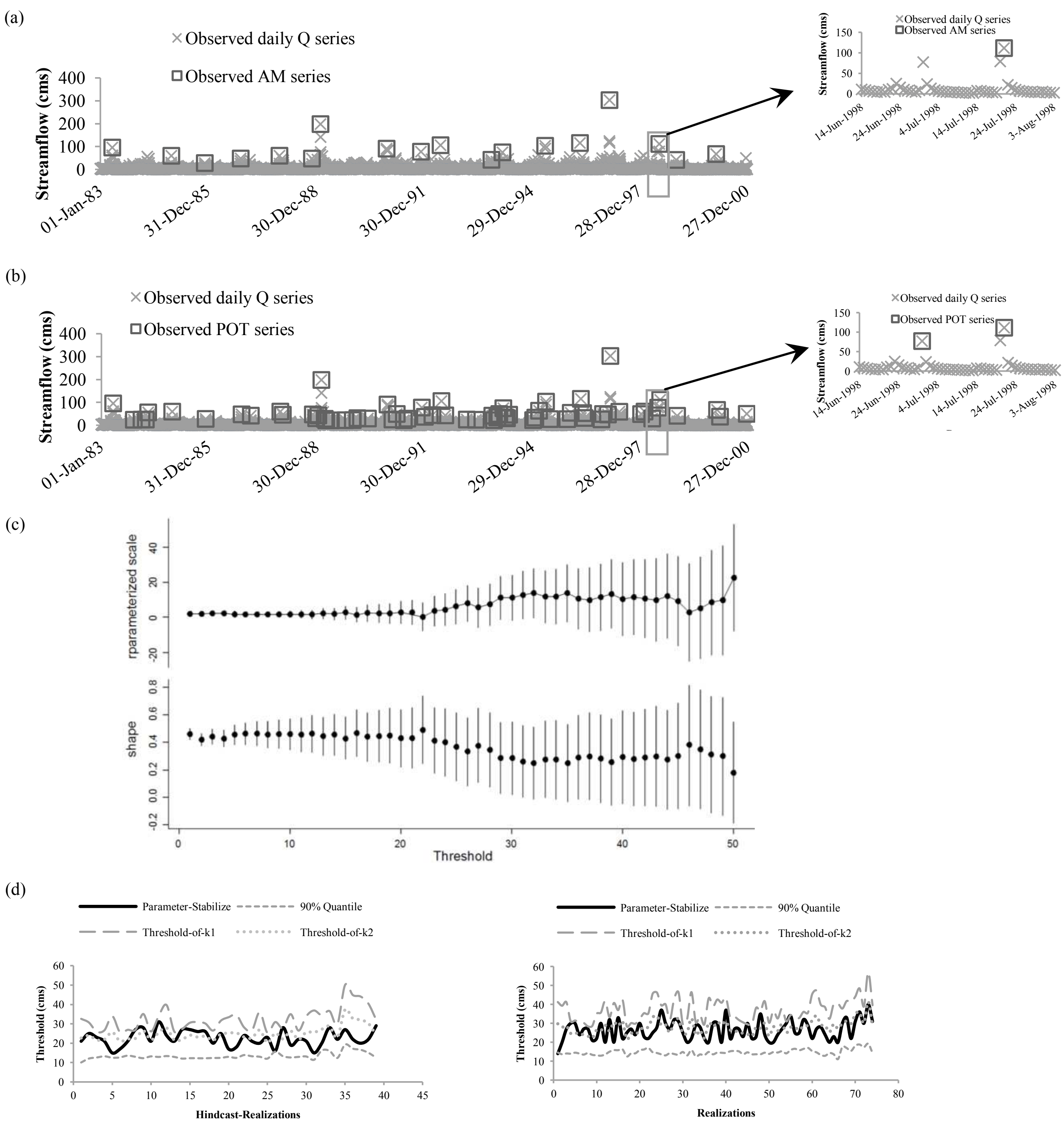


Figure 5

Figure 5. Factorial design for the four analysis of variance models.

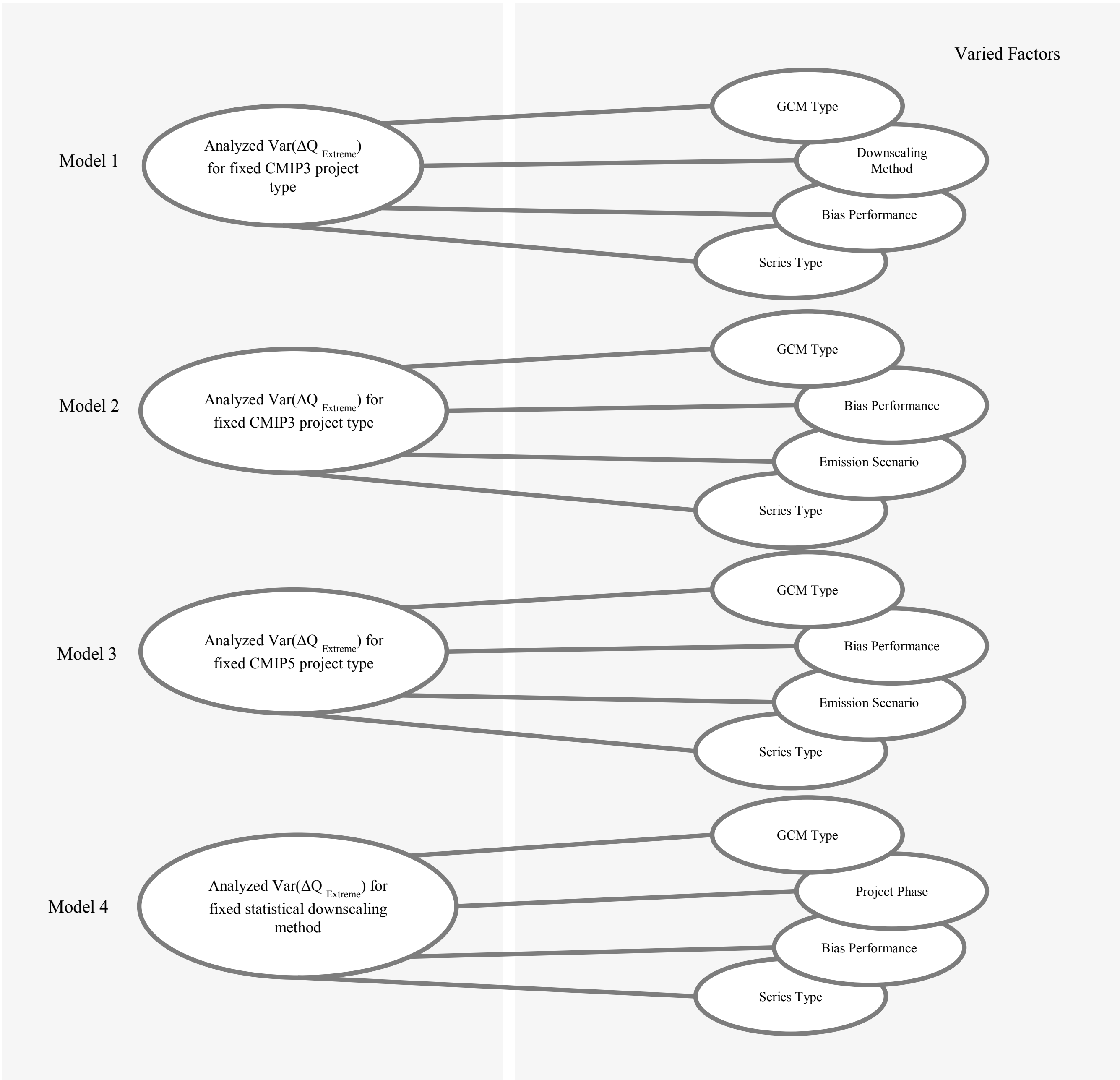


Figure 6

Figure 6. Observed and simulated streamflow by using SWAT.

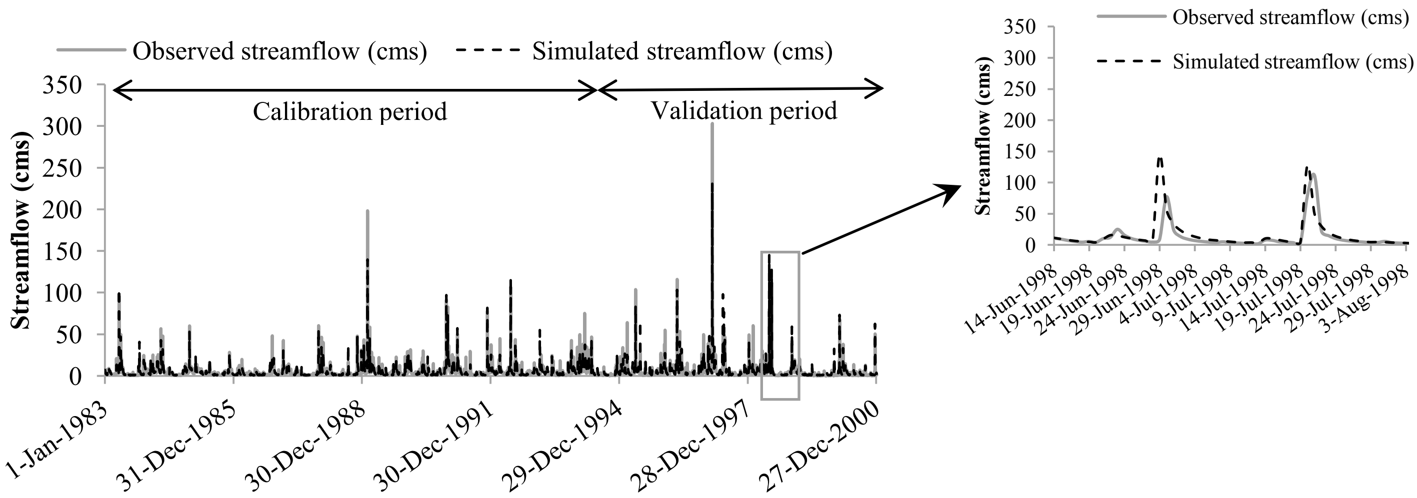


Figure 7

Figure 7. Results of the extreme quantiles fir (a) hindcast simulation, (b) different GCM results, and (c) CMIP3 versus CMIP5 results.

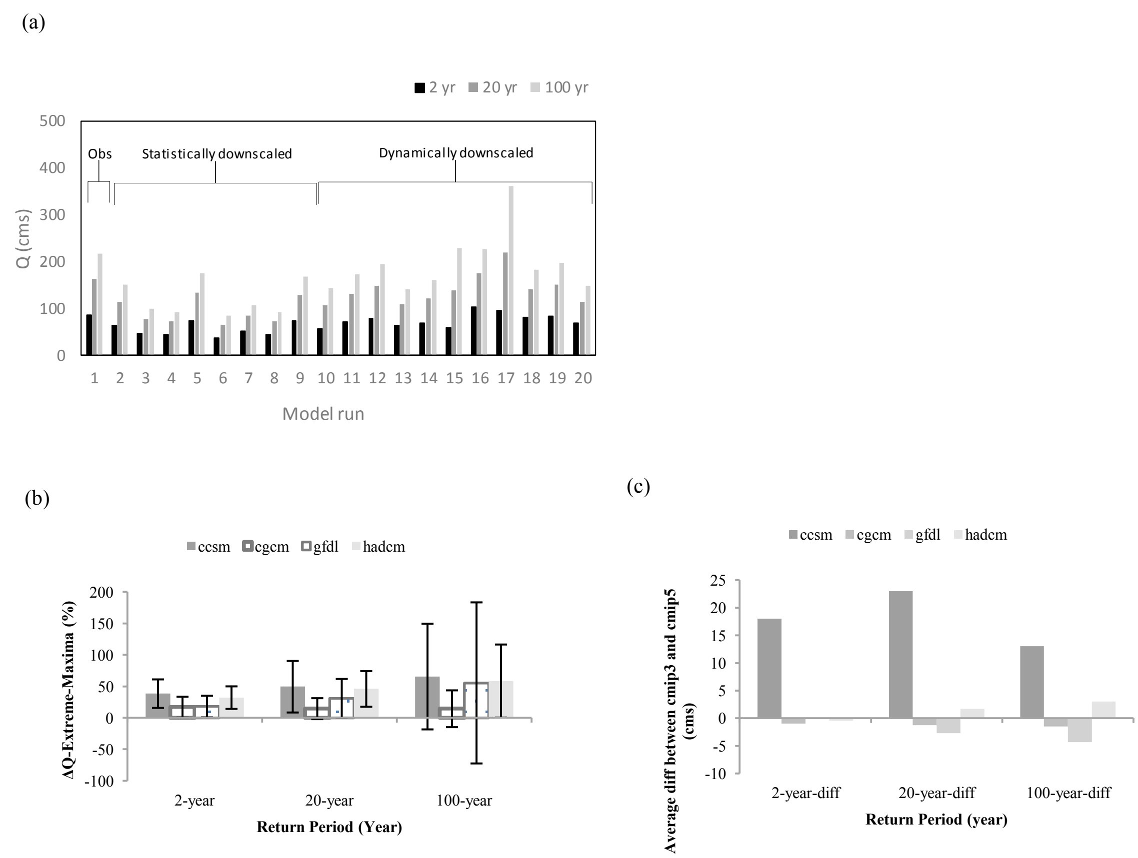
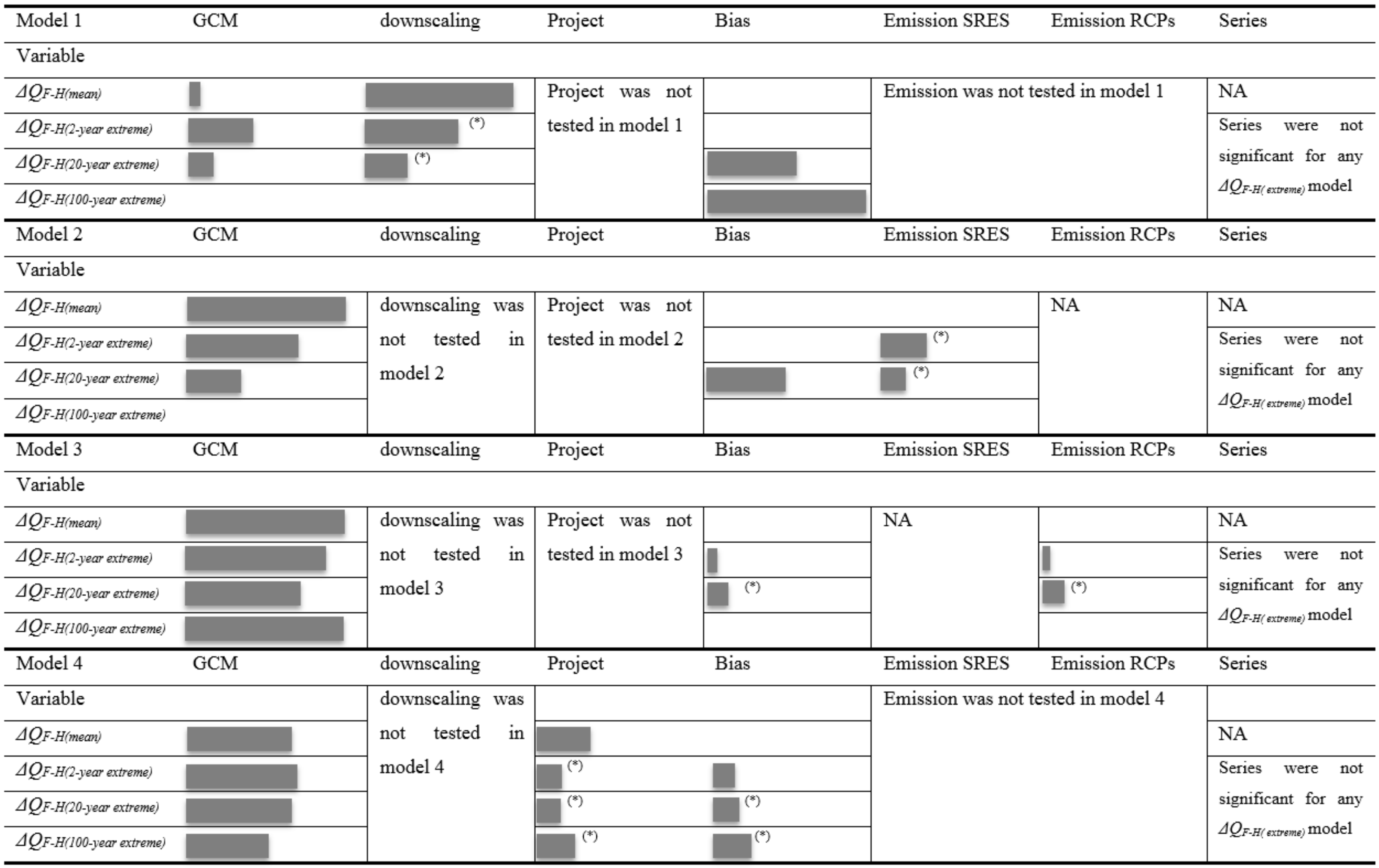


Figure 8

Figure 8. ANOVA results that present a comparison between the factors with respect to their significance and ranking. The horizontal dark grey bars represent the fraction of variance explained by the factor by using ANOVA method.



$^{(*)}$ Indicates that F-value was selected from the interaction effect. NA indicates not applicable simulation.

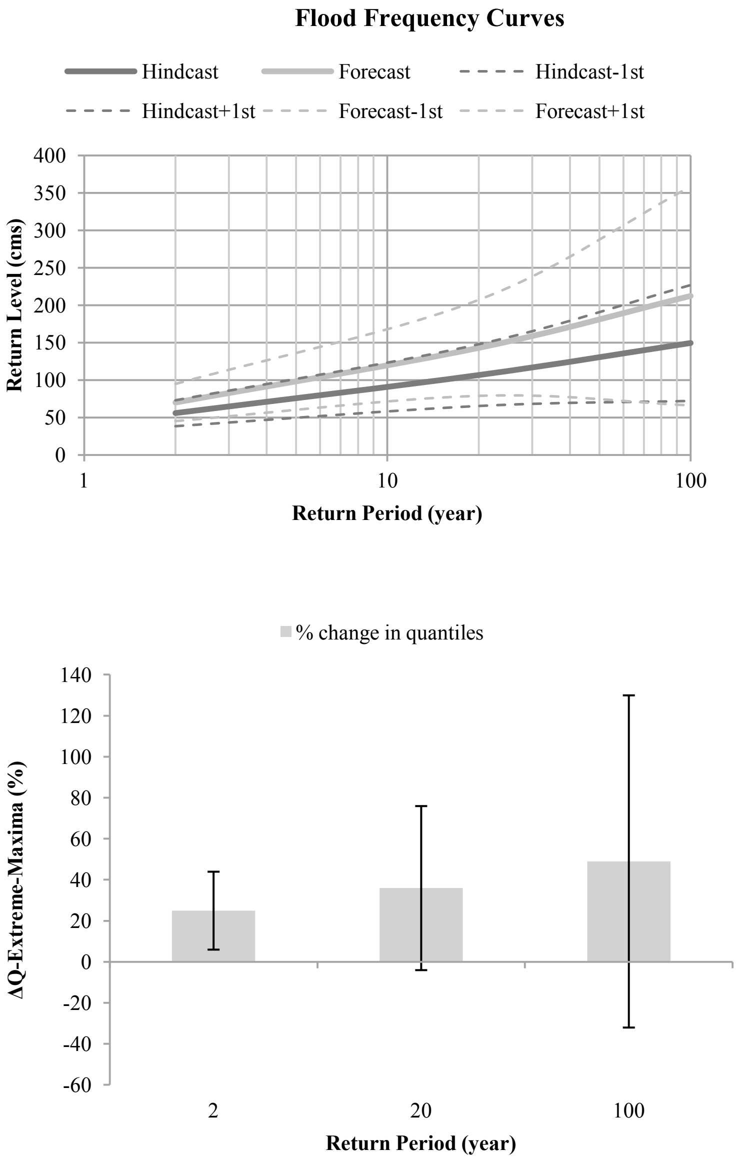
Figure 9

Figure 9. ANN and ANOVA results; comparison of variance decomposition. The horizontal grey bars represent the fraction of variance explained by the factor.

Model #	Return Period (Year)	ANOVA (R^2)	ANN (RE)		Method	Weights						
			Train	Test		GCM	Downscaling	Project	Bias	Emission (SRES)	Emission (RCPs)	Series
Model 1	2	0.54	0.43	0.43	ANN			Not tested		Not tested	Not tested	
					ANOVA				0			0
	20	0.46	0.64	0.66	ANN							
					ANOVA							0
	100	0.08	0.8	0.83	ANN							
					ANOVA	0	0					0
						GCM	Downscaling	Project	Bias	Emission (SRES)	Emission (RCPs)	Series
Model 2	2	0.43	0.74	0.72	ANN		Not tested	Not tested			Not tested	
					ANOVA				0			0
	20	0.49	0.75	0.76	ANN							
					ANOVA							0
	100	0.25	0.87	0.88	ANN							
					ANOVA	0			0	0		0
						GCM	Downscaling	Project	Bias	Emission (SRES)	Emission (RCPs)	Series
Model 3	2	0.96	0.37	0.35	ANN		Not tested	Not tested		Not tested		
					ANOVA							0
	20	0.83	0.5	0.5	ANN							
					ANOVA							0
	100	0.23	0.77	0.79	ANN							
					ANOVA				0		0	0
						GCM	Downscaling	Project	Bias	Emission (SRES)	Emission (RCPs)	Series
Model 4	2	0.71	0.3	0.28	ANN		Not tested			Not tested	Not tested	
					ANOVA							0
	20	0.55	0.5	0.51	ANN							
					ANOVA							0
	100	0.37	0.93	0.92	ANN							
					ANOVA							0

The horizontal grey bars represent the fraction of variance explained by the factor. The representation is by bar length where maximum and minimum lengths are 1 and 0 respectively and calculated from F-value. Adding the lengths for each extreme level will equal 1. The length of dark grey bars were calculated by using ANOVA method while the length of light grey bars were calculated by using ANN method.

Figure 10



N 38º, W 85º.

Pr= 1010 mm/year.

Jan: High pressure system.

July : High pressure system

Figure 10. Flood frequency curves and percent change in extreme quantiles found from ensemble analyses of climate modeling and statistical modeling analyses.

Table 1
Click here to download Table: Table1.docx

Table 1. Previous studies of streamflow maxima forecasted with global climate models.

Study/Location	**Number of GCMs**	**Downscaling method**	**Project Phase**	**Number of emission scenarios**	**Bias implementation for climate data**	**Extreme series type**	**Response variable**
Zhang et al., 2014 (China)	1	Dynamical	$CMIP3_{eq.}$	3	No	AM	$Q_{F(extreme)}$
Lawrence & Hisdal, 2011 (Norway)	8	Dynamical	$CMIP3_{eq.}$	3	Yes	AM	$\Delta Q_{F(extreme)}$
Mantua et al., 2010 (Washington state, USA)	10	Statistical	$CMIP3_{eq.}$	2	Yes	AM	$Q_{F(extreme)}$
Dankers and Feyen, 2008 (Europe)	1	Dynamical	$CMIP3_{eq.}$	2	No	AM	$\Delta Q_{F(extreme)}$
Prudhomme et al., 2003 (Great Britain)	7	Statistical	$CMIP3_{eq.}$	4	Yes	POT	$Q_{F(extreme)}$

GCM: Global Circulation Model, AM: Annual Maxima extreme series, POT: Peak Over Threshold extreme series

$CMIP3_{eq.}$: Equivalent to CMIP3 project where either the model version or the downscaling method is different but with the same forces (emission scenarios)

Table 2
Click here to download Table: Table2.docx

Table 2. Calibration and validation results for the SWAT model.

Optimization Gage	Time Step	Sub-basins Number	Soil Data Type	Landuse /landcover dataset	Total Flow Calibration (For the period 1/1/1983-12/31/1993)				Total Flow Validation (For the period 1/1/1994-12/31/2000)			
					R^2	PBIAS%	NS	RSR	R^2	PBIAS	NS	RSR
Near Midway	Daily	12	STAT-SGO	1992	0.67	-0.7	0.66	0.58	0.52	-14.97	0.46	0.74
	Monthly	12	STAT-SGO	1992	0.91	-1.08	0.9	0.31	0.91	-15.11	0.88	0.34
Near Frankfort	Daily	12	STAT-SGO	1992	0.6	0.14	0.58	0.65	0.52	-16.65	0.4	0.77
	Monthly	12	STAT-SGO	1992	0.84	-0.35	0.84	0.4	0.9	-16.7	0.84	0.4
					Total Flow Calibration (For the period 1/1/1983-9/30/1992)				Total Flow Validation (For the period 10/1/1997-12/31/2000)			
Yarnalton	Daily	12	STAT-SGO	1992	NA	NA	NA	NA	0.77	13.56	0.75	0.5
	Monthly	12	STAT-SGO	1992	NA	NA	NA	NA	0.84	13.5	0.78	0.46
Fort Spring	Daily	12	STAT-SGO	1992	0.7	5.9	0.68	0.58	0.53	-11.6	-0.01	1.00
	Monthly	12	STAT-SGO	1992	0.89	5.6	0.88	0.34	0.88	-11.62	0.73	0.51

NA indicates that there are no observation data were available.

Table 3. Sensitivity analysis of Extreme value modeling

Extreme event	Standard deviation of ΔQ	Variance of ΔQ
2 year event	±43%	1850
20 year event	±54%	2920
100 year event	±93%	8650

Table 4

Click here to download Table: Table4.docx

Table 4. Wind speed, net radiation and relative humidity impacts on streamflow maxima.

Model run	Δ Wind speed (%)	Δ net Radiation (%)	Δ rel humidity (%)	Q_2 (cms)	Q_{20} (cms)	Q_{100} (cms)	ΔQ_2 (%)	ΔQ_{20} (%)	ΔQ_{100} (%)
Control	NA	NA	NA	78	161	215	NA	NA	NA
Scenario 1	-10	-4	0.4	79	161	216	0.8	0.6	0.5
Scenario 2	-13.5	-9.5	0.45	81	165	220	3.6	2.7	2.5
Scenario 3	-17	-15	0.5	82	166	222	4.7	3.5	3.2
Scenario 4	-17	0	0	79	161	215	1.4	0.4	0.1
Scenario 5	0	-15	0	81	165	220	3.3	2.5	2.3
Scenario 6	0	0	0.5	78	160	214	-0.1	-0.2	-0.2

Table 5

Click here to download Table: Table5.docx

Table 5. Variance of the relative change in streamflow maxima from climate, extreme and hydrologic modeling components.

Modeling component	**Var[$\Delta Q_{2\ yr}$]**	**Var[$\Delta Q_{20\ yr}$]**	**Var[$\Delta Q_{100\ yr}$]**
1. Climate and extreme modeling factors	441	1122	7225
2. Additional climate shifts input to hydrologic model	3	3	3
3. Hydrologic model parameterization	9	11	13

Table 6. Monthly distribution of precipitation changes, streamflow changes and number of extremes.

Month	**ΔP_{mean}**	**ΔQ_{mean}**	**ΔQ_x number of extreme per 20 year period**
October	-9%	-2%	1
November	4%	0%	5
December	26%	34%	11
January	19%	27%	10
February	32%	21%	7
March	14%	15%	11
April	20%	25%	3
May	-2%	-2%	7
June	-3%	-11%	8
July	5%	2%	3
August	4%	10%	0
September	6%	17%	0
	$\Delta P_{yearlyr}$=10%	$\Delta Q_{yearlyr}$=11%	$\Delta Q_{2\text{-}year}$=25% $\Delta Q_{20\text{-}year}$=35% $\Delta Q_{100\text{-}year}$=49%

Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate

By Nabil Al Aamery, James F. Fox, Mark Snyder and Chandra V. Chandramouli

Nabil Al Aamery, Ph.D. Candidate of Civil Engineering at the University of Kentucky, nabil.hussain@uky.edu.

James F. Fox, Professor of Civil Engineering at the University of Kentucky, james.fox@uky.edu, +1(859)257-8668.

Mark Snyder, Project Scientist of Earth and Planetary Sciences Department, University of California Santa Cruz, masnyder@ucsc.edu.

Chandra V Chandramouli, Associate Professor of Civil Engineering, Mechanical and Civil Engineering Department at Purdue University Northwest, cviswana@pnw.edu.

Corresponding Author: James F. Fox, james.fox@uky.edu, +1(859)257-8668, 161 Raymond Bldg, Lexington KY 40506.

Variance Analysis of Forecasted Streamflow Maxima in a Wet Temperate Climate

Abstract:

Coupling global climate models, hydrologic models and extreme value analysis provides a method to forecast streamflow maxima, however the elusive variance structure of the results hinders confidence in application. Directly correcting the bias of forecasts using the relative change between forecast and control simulations has been shown to marginalize hydrologic uncertainty, reduce model bias, and remove systematic variance when predicting mean monthly and mean annual streamflow, prompting our investigation for maxima streamflow. We assess the variance structure of streamflow maxima using realizations of emission scenario, global climate model type and project phase, downscaling methods, bias correction, extreme value methods, and hydrologic model inputs and parameterization.

~~The~~ Results show that the relative change of streamflow maxima was not dependent on systematic variance from the annual maxima versus peak over threshold method applied, albeit we stress that researchers strictly adhere to rules from extreme value theory when applying the peak over threshold method. Regardless of which method is applied, extreme value model fitting does add variance to the projection, and the variance is an increasing function of the return period.

Unlike the relative change of mean streamflow, results show that the variance of the maxima's relative change was dependent on all climate model factors tested as well as hydrologic model inputs and calibration. Ensemble projections forecast an increase of streamflow maxima for 2050 with pronounced forecast standard error, including an increase of +30(±21), +38(±34) and +51(±85)% for 2, 20 and 100 year streamflow events for the wet temperate region studied. The variance of maxima projections was dominated by climate model factors and extreme value analyses.

Formatted: Normal

1 INTRODUCTION

Streamflow maxima is one of the most sought after response variables within hydrologic research and application (Coles, 2001, Begueria & Vicente-Serrano, 2006, Reiss &Thomas, 2007). Streamflow extreme maxima re-contours the morphology of the fluvial system (Leopold et al., 2012), partially controls the stream biogeochemical function (Ford and Fox, 2015), can destroy human infrastructure (Melillo et al., 2014), and resupplies human water stores for consumption, food production and energy generation (Rosenzweig et al., 2001, Mirza, 2003, McMichael et al., 2007). The complex earth system for which streamflow maxima responds is no less encompassing of hydrology than streamflow itself and includes components such as the climates ability to produce precipitation and weather patterns, the watershed's physiogeographic configuration and ability to respond to precipitation, and human's influence on both the watershed and climate. Despite hydrologists' long historical emphasis upon study of streamflow extreme maxima, current disparity is prevalent in terms of both streamflow maxima's current estimations and its gradient as we forecast into the future (Khaliq et al., 2006). Scientific gaps associated with estimating and forecasting current and future streamflow maxima is qualitatively attributed to scientific uncertainty surrounding human's economic behavior and influence on the earth system, representation of the climate and its changes, hydrologic representation of streamflow, and scalar coupling of a changing climate within a hydrologic representation of the earth (Madsen et al., 2014, IPCC, 2013). The difficulty of streamflow extreme maxima estimation and forecasting in a non-stationary earth system has challenged hydrologists to consider the potential use of new methodologies for investigating and forecasting streamflow.

One methodology for which streamflow maxima investigation and forecasting has received some recent attention is through the use of non-stationary projection with global climate models that can be used to drive hydrologic and statistical forecasting (Prudhomme et al., 2003; Dankers and Feyen, 2008; Mantua et al., 2010; Lawrence &Hisdal, 2011; Zhang et al., 2014). This method involves application of the non-stationarity form of long-term climate change projected using global climate models as a means to provide a physics-based guideline for extrapolation (Lima et al., 2015, Shamir et al., 2015). The global climate model results are post-processed for scalar considerations and then propagated through hydrologic models for predicting multi-year streamflow time series. Thereafter, the extreme value theorem is adopted to study streamflow extremes because the theory provides a mathematical basis for the definition of extremes and has been used to prove that the distribution of extremes follow similarity at their limit (e.g., Coles, 2001). Somewhat analogous to the central limit theorem, the extreme value theorem focuses on the statistical distribution and behavior of maxima that may arise from an unknown distribution for a population of a sequence of values measured over many time units. In this manner, hydrologists can statistically investigate current and forecast extreme value extreme maxima such as 2-, 20- and 100-year events via time series generated from the mentioned hydrologic modeling.

The coupling of global climate and hydrologic models for forecasting streamflow extreme maxima has been recently criticized for water infrastructure planning in some engineering and management circuits (Moradkhani, 2017), and we tend to agree that use of the methodology in a infrastructure design capacity is a bit preliminary given that published applications and results of the method is still in its infancy. Yet, we argue that the time is ripe for elucidating the variance structure of streamflow maxima forecasted with global climate models. We offer several reasons for this contention. First, highlighting the variance structure of forecasted streamflow maxima provides hydrologic and climate researchers with knowledge of highly sensitive factors and parameters of streamflow forecasting that systemically increase the size of the solution space, so that researchers might focus their attention towards improving model structure and parameterization. Second, the variance structure of forecasted streamflow maxima allows researchers to see what extent the previous results of forecasted mean streamflow might be adopted and extrapolated for forecasting extremes. There is a plethora of studies that forecast mean streamflow with global climate models (Chen et al., 2011, Al Aamery et al., 2016, Fatichi et al., 2014) and there is a question as to what extent results from these mean-focused studies might be relevant to the study of extreme streamflow, especially in light of the extra level of uncertainty that is introduced to forecasting maxima during application of the extreme value theorem. Third, a reason for investigating the variance structure of forecasted streamflow maxima is to help provide balanced forecasts that can be compared with a meta-analysis of trends in existing literature results such as in wet temperate regions.

While variance analysis of streamflow extreme maxima is sparse in the literature, global climate model research and forecasting of mean annual and mean monthly streamflow tends to suggest that climate models are different in their structures and parametrizations (Randall et al. 2007), downscaling methods are distinct in their stucture and results to re-scale global results

(Wilby and Dawson, 2007, Warner, 2010, Mearns et al., 2013), emission scenarios address the uncertainty of future economic and CO_2 conditions (IPCC, 2007, IPCC, 2013), the version of climate model projects are different in their structures, and the newer version (CMIP5) is distinct realtive older versions of models (CMIP3) (Brekke et al., 2013), and the bias implementation of climate results is a source for variance presence in hydrologic models (Teutschbein and Seibert, 2012, Al Aamery et al., 2016); and therefore all such components could have the potential to impact the variance structure of forecasted streamflow maxima. The few studies that have forecasted streamflow maxima with global climate models (see Table 1) have results that tend to corroborate some findings from mean-focused studies and suggest the type of global climate model applied caused differences in streamflow extreme forecasts (Prudhomme et al., 2003; Lawrence and Hisdal, 2011), and emission scenario can also shift extreme predictions (Dankers and Feyen, 2008; Mantua et al., 2010; Zhang et al., 2014). Applications of extreme value theory suggest the statistical analysis associated with the choice of extreme value analysis method has the potential to impact the variance of forecasted streamflow maxima; and the annual maxima method is criticized for its neglect of multiple extremes per annum while the peak over threshold method has been criticized for subjectivity of threshold selection (Svensson et al., 2005; Scarrott and MacDonald, 2012; Bezak et al., 2014; Fischer and Schumann, 2016).

<Table 1 here please>

Beyond uncertainty surrounding the global climate model projections and extreme value methods, there are additional uncertainty considerations with respect to the future hydrologic balance and its simulation when forecasting streamflow maxima. As one example, future changes in the hydrologic cycle, and in turn streamflow, are primarily driven by changes in precipitation and evapotranspiration. Studies that forecast streamflow maxima with global climate models have focused on precipitation and temperature differences within simulation of future periods (Mantua et al., 2010; Zhang et al., 2014), however it is now recognized that evaporative demand is physically controlled by net radiation, vapor pressure and wind speed as well as air temperature (Donohue et al 2010). Future projections of these additional variables suggest decreases in wind speed and net radiation and an increase in relative humidity for some regions (Willet et al., 2008; Wild, 2009; McVicar et al., 2012). The directions of the projected shifts would decrease evapotranspiration and in turn could potentially increase variability when forecasting streamflow maxima. As a second example, hydrologic model fit and modeling uncertainty when simulating the water balance has the potential to increase the variability of projected streamflow maxima (Al Aamery et al., 2016). Recent study has tended to marginalize the importance of hydrologic model calibration and uncertainty for mean streamflow projections when considering relative future changes (Niraula et al., 2015), however streamflow maxima has not yet been tested in this context, to our knowledge.

Our objectives were to: (1) perform coupled climate, hydrologic and statistical model simulation and evaluation to build realizations of streamflow maxima; (2) perform variance analysis to test for systematic uncertainty from climate and extreme modeling factors potentially controlling streamflow maxima forecasting; (3) perform uncertainty analysis to quantify variance from hydrologic modeling; and (4) forecast streamflow maxima for the wet temperate region studied herein and provide literature comparison. These objectives provide the structural sub-headings used in the following Methods, Results and Discussion sections.

2 THEORETICAL BACKGROUND

The variance structure of forecasted streamflow maxima can be decomposed as a function of potentially controlling modeling factors. The conceptual model of factors that have the potential impact forecasted streamflow maxima variance is shown in Figure 1. As can be seen in the figure, modeling factors that may impact the variance structure can be grouped into those associated with climate modeling (CMFs in Figure 1) including global climate model (GCM) type, hydrologic modeling (HMFs) and uncertainty in inputs and parameterization, and statistical modeling of extremes (SMFs) associated with different fitting methods and distributions. Land use and management modeling is not shown in the figure and was treated as static in this study, but it is also recognized to potentially control future streamflow.

<Figure 1 here please>

The response variables are streamflow maxima associated with different return periods, including 2, 20 and 100 year return periods, so the distribution of extremes can be quantified (Lawrence and Hisdal, 2011). We consider response of the future relative change in streamflow equal to the percent difference of GCM-forecasted streamflow maxima relative to GCM-hindcast maxima ($\Delta Q_{F\text{-}H\,x}$, where x indicates the return period). The future streamflow maxima can be related to the 'real' streamflow maxima by using the relative changes derived from the forecast projections and hindcast-control projections coupled with the observations, such as using the delta method directly applied to streamflow model results.

The relative change approach has become rather popular in climate change studies that emphasize GCM-forecasted streamflow (Chien et al., 2012; Harding et al., 2012; Fitichi et al., 2014; Niraula et al., 2015; Al Aamery et al., 2016). The~~ relative change~~ approach has been suggested to remove seasonal, spatial, and/or inter-annual biases of GCMs or statistical artifacts from the downscaling method that are not accounted for in bias correction methods (Harding et al., 2012). ~~The~~In addition, application of the relative change ~~also~~has recently ~~showed~~shown no significant dependence upon calibrated versus un-calibrated hydrologic model simulation, thus suggesting the response variable does not require model calibration to see the projected direction of future streamflow (Niraula et al., 2015). The ~~relative change~~approach has also shown less dependence upon climate modelling factors (i.e., CMFs in Fig 1) as compared to the absolute forecasted streamflow suggesting that biases specific to a model structure could be accounted (Al Aamery et al., 2016). ~~The~~While the relative change approach has shown potential in past studies, ~~yet~~ these studies have tended to focus on the mean forecasting of streamflow. In the present study, we consider the ~~relative change is considered~~method for streamflow maxima, which is one contribution of this paper.

Realizations of the relative change in streamflow maxima can be simulated as a function of climate, hydrologic, and statistical modeling factors within a variance analysis ensemble (Al Aamery et al., 2016). In the present study, we included permutations using seven emission scenarios (i.e., emission factor, CMF1) propagated through eight different GCMs associated with phase three and four climate projects, i.e., CMIP3, CMIP5, (i.e., GCM type and version factors, CMF2, 3) that were downscaled using two statistical downscaling methods and four dynamical downscaling methods (i.e., downscaling factor, CMF4). Further, our post-processing and hydrologic analyses of downscaled hindcast (1983-2000) and forecast (2048-2065) climate model results considered bias correction (i.e., bias factor, SMF1) propagated through a continuous simulation hydrologic model. We performed both annual maxima and peak over threshold extreme value analyses (i.e., extreme value factor, SMF1,2) of hydrologic model results given recent debate in the literature over the best method. We also investigated additional uncertainty considerations with respect to additional hydrologic inputs and hydrologic uncertainty (HMF 1, 3).

3 STUDY SITE AND MATERIALS:

The study site was South Elkhorn Watershed in Lexington, Kentucky USA (see Figure 2). This watershed is within a wet and temperate region where a future change in climate, including an increase in precipitation and temperature, is projected (Melillo et al., 2014). According to Melillo et al. (2014), a 20 to 30% increase in annual maximum precipitation is projected under RCP 8.5 emission scenario for the end of the century. Additionally, at least, 80% of the models used in Melillo et al. (2014) are in agreement for this region. The watershed covers an area of 478.6 km^2 with surface elevations ranging between 197 to 325 m asl. The land use is dominated by agricultural ~~lands with about 71~~equal to 72%. The remaining land uses are urban/suburban equal to 13%, forest equal to 14%, and ~~urbanization lands that spread over 12% of the watershed.~~ open water and wetlands equal to 1%.

<Figure 2 here please>

The results of eight GCMs were implemented in this analysis. The GCM models reflected four different GCM model types and two versions of each model, inculding a version from CMIP3 and the newer version from CMIP5 (Brekke et al., 2013; Al Aamery et al., 2016). The GCMs included the Canadian Global Climate Model including CGCM3 from CMIP3 and CanESM2 from CMIP5 (Flato, 2005); the National Center for Atmospheric Research Community Climate Model including CCSM3 from CMIP3 and CCSM4 from CMIP5 (Collins et al., 2006); the Geophysical Fluid Dynamics Laboratory including GFDL CM2.1 from CMIP3 and CM3 from CMIP5 (Delworth et al., 2006); and the United Kingdom Hadley Centre Climate Model including HadCM3 from CMIP3 and HadGEM2-ES from CMIP5 (Gordon et al., 2000). These GCMs were chosen for their representation in different climate projects, including CMIP3, CMIP5, and NARCCAP projects, and their available archives of climate results for the current and future periods focused on in this study (Brekke et al., 2013; Mearns et al., 2013; Al Aamery et al., 2016).

Statistical downscaling and dynamical downscaling results were included in this analysis. The statistical downscaling results were used from the Coupled Model Inter-comparison Project phase three (CMIP3) and phase five (CMIP5) (Brekke et al., 2013). The dynamical downscaling results were used from the North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2013). These downscaling methods represent two distinct approaches for downscaling GCM results from their coarse scale to a finer watershed scale. The statistical downscaling method is statistically based and adopts empirical-statistical relationships to estimate the small-scale climate variables based on the large-scale atmospheric variables (Wilby and Dawson, 2007). The statistical downscaling method implemented in CMIP3 and CMIP5 projects adopt two schemes including bias correction and spatial disaggregation (BCSD) and bias-correction and constructed analogs (BCCA) (Brekke et al., 2013). The dynamical downscaling method is physically-based and uses regional climate models (RCMs) whose boundary conditions are forced by the results of the parent GCM to simulate the atmospheric physical processes on a regional scale (Warner, 2010). Six regional climate models were implemented through the NARCCAP project including the Canadian Regional Climate Model (CRCM) (Plummer et al., 2006), the Experimental Climate Prediction Center (ECPC) model (Juang et al., 1997), the Hadley Regional Model 3 (HRM3) (Jones et al., 2003), the MM5-PSU/NCAR mesoscale model (MM5I) (Chen and Dudhia, 2001), the Reginal Climate Model version 3 (RCM3) (Giorgi et al., 1993), and the Weather Research and Forecasting model (WRFP) (Skamarock et al., 2005).

4 METHODS

4.1 Modeling simulations and evaluation

The Soil and Water Assessment Tool (SWAT; the version was ArcSWAT 2012.10.1.13) model was applied to simulate the hydrology of South Elkhorn Watershed. This model is physically based and was applied successfully in this region and many other regions around the world (Palanisamy and Workman, 2014; Gassman et al., 2007; Arnold et al., 1998). The model was evaluated over 1981-2000 using the observed climate and streamflow data and applied for the hindcast period (1981-2000) and forecast period (2046-2065) using the GCMs results of daily precipitation and maximum and minimum temperature (see Figure 3 in Al Aamery et al., 2016 for evaluation methods of SWAT). We obtained all the data required by SWAT including topography, soil, and landuse data from publically available databases. The topography and streamlines data were obtained from the National Map website (http://viewer.nationalmap.gov/viewer/), and the soil data was obtained from the Data Gateway website (http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx). The landuse data of 1992 was used from the USGS website (http://www.mrlc.gov/nlcd2011.php). The observed climate data of daily precipitation and maximum and minimum temperature were obtained from Bluegrass Airport meteorological gage station. The model was evaluated using four USGS streamflow gage stations within the watershed including South Elkhorn Creek at Fort Spring, USGS03289000, Town Branch at Yarnallton Road, USGS03289200, South Elkhorn Creek near Midway, USGS03289300, and Elkhorn Creek near Frankfort, USGS03289500. Results for the South Elkhorn Creek near Midway station were analyzed for the relative change in streamflow maxima. Model evaluation including calibration, validation, and sensitivity analysis was performed semi-automatically via SWAT-CUP software for Sequential Uncertainty Fitting SUFI2 for the four gage stations in the watershed (Abbaspour et al., 2007). The first two years was left as a spin-up period for SWAT (Arnold et al., 2010).

<Figure 3 here please>

As input to the hydrologic modeling, the scaling of precipitation and temperature method of Lenderink et al. (2007) was applied to correct the bias in the climate data. The method operates with monthly correction values based on the difference between observed and current period simulated values as:

$$P^*_{sc}(d) = P_{sc}(d)\ \frac{\mu_m(P_{obs}(d))}{\mu_m(P_{sc}(d))}, \tag{1}$$

$$P^*_{sf}(d) = P_{sf}(d)\ \frac{\mu_m(P_{obs}(d))}{\mu_m(P_{sc}(d))}, \tag{2}$$

$$T^*_{sc}(d) = T_{sc}(d) + \mu_m\big(T_{obs}(d)\big) - \mu_m\big(T_{sc}(d)\big)\text{ , and} \tag{3}$$

$$T^*_{sf}(d) = T_{sf}(d) + \mu_m\big(T_{obs}(d)\big) - \mu_m\big(T_{sc}(d)\big)\text{ ,} \tag{4}$$

where $P^*_{sc}(d)$ and $T^*_{sc}(d)$ are the corrected daily precipitation and temperature for the simulated current period, $P^*_{sf}(d)$ and $T^*_{sf}(d)$ are the corrected daily precipitation and temperature for the simulated future period, $P_{sc}(d)$ and $T_{sc}(d)$ are the uncorrected daily precipitation and temperature for the simulated current period, $P_{sf}(d)$ and $T_{sf}(d)$ are the uncorrected daily precipitation and temperature for the simulated future period, $\mu_m(P_{obs}(d))$ is the average of observed daily precipitation values for a given month, $\mu_m(P_{sc}(d))$ is the average daily precipitation for the current simulated values, $\mu_m\big(T_{obs}(d)\big)$ is the average of observed daily temperature values, $\mu_m\big(T_{sc}(d)\big)$ is the average of daily temperature for the current simulated period, and *m* stands for “within monthly time step”.

The annual maxima (AM) and peak over threshold (POT) methods were carried out for each realization from the hydrologic modeling results in order to analyze the extremes (see Figure 3). The AM series is constructed by selecting one value per a specific time over the sample size. In streamflow studies such as herein, this value is the maximum water discharge value selected over one year from the daily time series data (Khaliq et al., 2006, Haan, 2002). Thereby, the AM series replaces the flow series ($q_1, q_2, \ldots\ldots, q_{365}$) of a year ($j$) by the largest flood value q_m^j (where $1 \le m \le$ total number of days in year j, $1 \le j \le n$, and n is the number of years). According to the extreme theorem, the probability of the rescaled M_n is approaching the General Extreme Value (GEV) family when $n \to \infty$. The GEV family distribution is expressed as follows:

$$G(q) = \exp\left\{-\left[1 + \xi\left(\frac{q-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{5}$$

where $\left\{q : 1 + \xi\left(\frac{q-\mu}{\sigma}\right) > 0\right\}$, the location parameter $-\infty < \mu < \infty$, the scale parameter $\sigma > 0$, and the shape parameter $-\infty < \xi < \infty$. Depending on the value of the shape parameter ξ, the GEV family has three distinct probability distributions. The light tail Gumbel type when $\xi = 0$, the heavy tail Fréchet type when $\xi > 0$, and the bounded upper tail Weibull type when $\xi < 0$. The extreme quantiles of the return level T when $\xi \neq 0$ are then calculated as follows:

$$q_T = \mu - \frac{\sigma}{\xi}[1 - \{-\log(1 - \frac{1}{T})\}^{-\xi}] \tag{6}$$

and when $\xi = 0$

$$q_T = \mu - \sigma \log\{-\log(1 - \frac{1}{T})\}. \tag{7}$$

The POT series was constructed by selecting all independent and identically distributed values ($q_1, q_2, \ldots\ldots$) that are higher than a specific, and carefully chosen, value called threshold point (q_o) (see example in Figure 4). According to extreme value theory, for large enough q_o the distribution function of $y = (q - q_o)$ conditioned by $q > q_o$ is approximated by the Generalized Pareto (GP) family as follows (Coles, 2001):

$$H(y) = 1 - (1 + \frac{\xi y}{\tilde{\sigma}})^{-1/\xi} \tag{8}$$

where $\left\{ y : y > 0 \ and \ (1+\frac{\xi y}{\tilde{\sigma}}) > 0 \right\}$, $and \ \tilde{\sigma} = \cancel{\sigma} \ \ \xi(q_o - \mu)$, q is a specific value in the sequence (q_1, q_2, ...), and $\sigma, \mu,$ and ξ are the scale, location, and shape parameters. Depending on the value of the shape factor, the GP family consists of three probability distribution functions as follows: the heavy tail Pareto type when $\xi > 0$; the light tail Exponential type when $\xi = 0$; and bounded upper tail Beta type when $\xi < 0$. To calculate the extreme quantile (q_T) of the return period *T*, the probability $\zeta_{q_o} = \Pr \ (q > q_o)$ is calculated first and then the return period when $\xi \neq 0$ is given by:

$$q_T = q_o + \left(\frac{\tilde{\sigma}}{\xi}\right)\left[(\zeta_{q_o} T)^{\xi} - 1\right] \tag{9}$$

When $\xi = 0$, the return level of a period *T* is given by:

$$q_T = q_o + \tilde{\sigma} \log(T\zeta_{q_o}) \tag{10}$$

<Figure 4 here please>

Threshold Point Choice: We adopted the parameter stabilization method explained by Coles (2001) to choose threshold points used within the POT method. The method is based on fitting the General Pareto distribution across a range of different threshold points. When fitting, the model parameters including the shape parameter (ξ) and scale parameter $(\tilde{\sigma})$ were estimated for each point across the range. The shape parameter should be approximately constant, and the scale parameter should be linear in q when the GP distribution is valid above the q_o (Coles, 2001). Figure 4 shows an example of fitting the GP model using the maximum likelihood method over a range of 1 to 40 for the threshold point. As observed, the shape and the reparametrized scale parameters are nearly stable until reaching the point 21. We, therefore, specified the point 21 cms as the threshold point for the POT series of the observed daily streamflow series in this example; and the method was repeated for each model hydrologic realization performed in our study. In order to support our choice of the threshold, we compared our final results of threshold selection using the parameter stabilization method with three Rules of Thumb presented by Scarrott and MacDonald (2012). Using general order statistics convergence properties, methods including the upper 10% rule, square root rule $k_1 = \sqrt{n}$, and

$k_2 = n^{2/3}/\log[\log(n)]$ rule were developed (see Scarrott and MacDonald, 2012). Figure 4 shows that our choices compared well to the three methods.

Temporal Independency in the POT Series: The values of the POT series, in the sense of extreme theorem, should admit to the temporal independence condition. By only selecting all values that are higher than the threshold point, we will obviously violate this condition within a streamflow time series. Therefore, to identify and remove the time dependency in the POT series values, de-clustering of the POT series was adopted. The de-clustering was performed by calculating the Extremal Index (θ) as follows (Coles, 2001):

$$\theta = (\text{limiting mean clustering size})^{-1},$$

(11)

where θ equal to one indicates an independent series. Therefore, the objective was to minimize the size of the clusters until θ reaches one. Our approach was to make manual iteration for each POT series to select the number of threshold deficits, *r*, used to define a cluster. Moreover, to support our independent choices of POT series, we performed the auto-tail dependence function plots for the data series (Reiss and Thomas, 2007) to test the dependency of the events in the series.

Trend Analysis: We analyzed the POT and the AM series with respect to the non-stationarity explained by trend analysis. We used the Mann-Kendall nonparametric test to identify the presence of trends in each independent POT and AM series (Haan, 2002). If the trend was present, we removed the trend from the series, although as will be discussed in the results, very few series exhibited a significant mean trend.

Likelihood Ratio Test: The likelihood ratio test was used to test the null hypothesis of the shape factor (ξ) to be zero. This test is used in statistics to test the goodness of fit of two distributions when one of them is a special case of the other, i.e., nested models (Hogg et al., 2014, Coles, 2001). In our case, the Gumbel distribution is nested within the GEV distribution, and the Exponential distribution is nested within the GP distribution.

Currently, the AM and POT series are the only two types of flood peak series that can be used for flood frequency analysis, and further discussion of a comprehensive comparison between the two series is provided in the literature in Bezak et al. (2014) and Madsen et al. (1997). To perform all the methods described in the extreme analysis methods section and shown in Figure 3, we have applied the R package extRemes version 2.0 described in Gilleland & Katz (2016).

4.2 Uncertainty from climate and extreme modeling factors

Our results from the coupled climate, hydrologic, and extreme modeling methods produced 226 realizations of model runs available for variance analysis based on a factorial design that considered emission type, GCM type and version, downscaling type, bias correction, and extreme value method type. Each factor was divided within variance decomposition as follows: the GCM type factor was divided into four levels for the four parent models mentioned previously; the GCM version factor was divided into two levels indicating CMIP3 and CMIP5 project phases of the models; the downscaling factor was divided into two levels for statistical and dynamical methods; the emission factor was divided into seven levels including the SRES type used in CMIP3 (A1B, A2, and B1) and the RCPs type used in CMIP5 (RCP2.6, RCP4.5, RCP6.0, and RCP8.5); the bias factor was divided into two levels indicating inclusion of methods in Equations (1-4) or lack thereof; and the extreme value factor was divided into two levels for AM and POT methods. Further details of the factorial levels for each of the 226 realizations are provided in the Supplementary On-line Table. We simulated variance analysis following both more traditional linear methods and more recently published nonlinear methods in order to maintain robustness of the analyses.

Linear Analysis of Variance (ANOVA): We performed statistical analysis through fitting the linear analysis of variance model (ANOVA) to the results of the maxima extreme analysis. ANOVA was applied separately for each streamflow maxima quantile. The extreme quantiles represent the response variables of 2-year, 20-year, and 100-year return periods ($\Delta Q_{F\text{-}H(2\text{-}year\text{-}ME)}$, $\Delta Q_{F\text{-}H(20\text{-}year\text{-}ME)}$, and $\Delta Q_{F\text{-}H(100\text{-}year\text{-}ME)}$ respectively) *via* the general linear model-univariate procedure in *SPSS 22* software (Pallant, 2013). ANOVA explores the effect of different factors on the variance of the response using the p-value of the statistical test and ranking factor importance by using the F-value. The F-value of each factor was divided by the summation of F-values in a single model to determine how much variance that factor explains from the total predictable variance. Several considerations were determined when applying ANOVA methods. First, because the datasets were not represented in the climate factors equally, we applied four separate models that balanced a set of factors. The reason for the multiple models is attributed to our climate datasets where the CMIP3 project has both statistically and dynamically downscaled results while the CMIP5 project has only statistically downscaled results. We, also, have different emission scenarios between the two projects. CMIP3 has SRES emission scenarios; and CMIP5 has RCPs emission scenarios. We built therefore four-way ANOVA models as the highest possible order to constrain the balanced and nested models. Figure 5 shows four possible 4-way ANOVA models that we built from our factorial design. Second, we analyzed the factors across the models using the highest possible order, however, if a factor was found to be unimportant, we omitted the factor to maximize repetitions.

<Figure 5 here please>

ANOVA assumes that the population is normally distributed, although the violation of this assumption should not cause major problems when the sample size is greater than 30 (Pallant, 2005, Gravetter&Wallnau, 2000, Stevens, 1996). In our factorial design, the least sample size was recorded in ANOVA model 3, where the sample size was 56. Therefore, our concern about the normality assumption is limited. The homogeneity of variance assumption was treated by using the Levene test for the equality of variance (Pallant, 2005). If the data failed in this test, the significant level by which we compare the variances of the different groups in the ANOVA models was 0.01, which overcomes the violation of this assumption (Pallant, 2005).

Nonlinear Artificial Neural Network (ANN): ANN models, on the other hand, were considered in this study to reinforce our robustness of the variance analysis. ANNs provides a model framework based on a set of multivariate nonlinear functions, and therefore could account for nonlinearity between factors controlling variance and the streamflow response variable, if it exists. In this manner, ANNs could overcome the underlying multivariate linear model limitation that ANOVA is based on. We used the ANN model to examine the climate factors importance on streamflow maxima projections through *SPSS 22* software (IBM , 2012, Tufféry, 2011). The input layer represented the climate and the statistical factors with nominal variables, and the output layer represented the relative change in streamflow maxima. We used one hidden layer with a randomly generated number of neurons. We used supervised training with multilayer perceptron and feedforward architecture. All values of the input and output layers were normalized so that all values ranged between 0 and 1. The hyperbolic tangent activation function was considered in the hidden layer. We used the same four models proposed in the ANOVA analysis to perform the ANN analysis. The dataset partitioning was performed with *SPSS-ANN* to divide the data into training and testing datasets. However, through generation of random numbers within *SPSS-ANN*, the partitioning values of training and testing will swing around the 70% and 30% marks for each run of many runs performed for each model. The values of training partitioning ranged between 60% and 80% affecting the testing portion and providing a new relative error value for both training and testing parts. Accordingly, the smallest relative error provides the best results for the ANN model (IBM, 2012). Therefore, our approach was to use an initial 70% of the dataset for training and the rest for testing, and then rerun the model until obtaining the minimum possible relative errors across the training and testing data.

4.3 Uncertainty from hydrologic modeling

Additional uncertainty from the future hydrologic balance and its simulation were also quantified as part of our study. Future projections of net radiation, vapor pressure and wind speed were tested in simulation for the study region with the premise that decreases in wind speed and net radiation and an increase in relative humidity could decrease future evapotranspiration and in turn increase streamflow maxima while at the same time increase uncertainty of forecasts. Future projections that consider hydrologic model fit and hydrologic parameter uncertainty were also tested to assess the potential to increase the variability of projected streamflow maxima.

Future climate change of wind speed, net radiation, and relative humidity were tested within hydrologic simulation by considering projected shifts reported in the literature. The average monthly wind speed in the study site ranges between 3 and 5 m/s. According to McVicar et al. (2012), the possible stilling in the middle of the current century is approximately 0.5 m/s for the study site region when assuming a linear trend of their observations reported therein. In turn, the percent climate change of wind speed is between -10% and -17% for the future period in the study region. Wild (2009) indicates that the surface solar radiation has a decadal variation and that the absolute trend was observed as -6 W m^{-2} per decade and 8 W m^{-2} per decade for the periods of 1961-1990 and 1995-2007, respectively, over the United States. We recognized that increasing radiation would offset decreasing wind speed when estimating evapotranspiration, and therefore we considered the decreasing trend of -6 W m^{-2} per decade for the future period, in order to test its sensitivity. The mean daily solar radiation ranges throughout the year between 81 W m^{-2} (1.9 kW h $m^{-2}d^{-1}$) and 300 W m^{-2} (7.2 kW h $m^{-2}d^{-1}$). Considering the mentioned net decrease produces a change in the solar radiation reaching the surface to be between -4% and -15%. Regarding the relative humidity, Willett et al. (2008) shows data that suggests an increase in the relative humidity for the northern hemisphere. The net increase shown was 0.07% for the 10 year period of investigation. We assumed the same change for the future period, which resulted in a range between +0.4% and +0.5% for the study region. Donohue et al. (2010) showed that the Penman equation produced the most reasonable estimation of evaporation demand, and this method is included within the hydrologic model used in the present study. Therefore, we considered a number of scenarios in hydrologic modeling that test the mentioned ranges of wind speed, net radiation, and relative humidity concurrently to see their added impact on streamflow maxima. We also tested the variables independently to see their individual sensitivity upon the streamflow maxima.

Future projections that consider hydrologic model fit and hydrologic modeling uncertainty were also tested with the hydrologic model to investigate their impact on forecasted streamflow maxima. Recent literature results have marginalized the importance of model fit when forecasting the relative change in future mean streamflow (Niraula et al., 2015), and we tested this concept for future streamflow maxima. The future streamflow maxima produced from the calibrated hydrologic model simulation for a set of GCM realizations was compared against the future streamflow maxima produced using the un-calibrated (i.e., default) parameterization of the hydrologic model for the same climate realizations. Additionally, the impact of hydrologic model uncertainty was considered by carrying forward uncertainty projections from the hydrologic model parameterization to the extreme value methods and thereafter to compute the relative change in future streamflow. The SWAT-CUP software provides parameter sets and solutions used to create uncertainty bounds during the model simulation. Realizations of all parameter sets that meet the objective function criteria were chosen and extreme value methods were performed for hindcast and forecast global climate pairs to compute the relative change in streamflow maxima.

4.4 Forecast of streamflow maxima for wet temperate regions

After quantifying the climate, hydrologic, and extreme modeling factors controlling variability of the projections, an ensemble was created to forecast the relative change in the streamflow maxima for the wet temperate study region (Al Aamery et al., 2016). The extreme forecasts for this study calculated the net effect on the mean and variance of the balanced ensemble from variation of climate modeling factors and extreme modeling factors, the added uncertainty from hydrologic model parameterization, and the added mean shift and its variance from climate change shifts in net radiation, vapor pressure and wind speed. Results were compared with other studies reported in the literature of streamflow maxima (see Table 1) that fell within wet temperate regions.

5 RESULTS AND DISCUSSION

5.1 Modeling simulations and evaluation

Results from the model evaluation showed that the hydrologic model performed within an acceptable range, and ~~simulated streamflow well. The~~the simulated and observed daily streamflow signals showed close agreement (see Figure 6). The four quantitative matrics including coefficient of determination (R^2), percent bias (*PBIAS%*), Nash-Sutcliff Efficiency (*NS*), and the ratio of the root mean square error to the standard deviation of measured data (*RSR*) showed results within the acceptable range (Moriasi et al., 2007, Donigan, 2002, Gassman et al., 2007) in both calibration and validation periods for the majority of the four observation sites for which the model was compared against (see compiled metrics in Table 2), although one of the four sites showed values just below or equal to the acceptable range boundary during validation. Overall, 53 out of the 56 metrics that compared observations with model results were above the acceptable range showing that the model simulated streamflow well. According to Moriasi et al. (2007) the monthly time step model performance is considered satisfactory if the NS>0.5, RSR<0.7, and $PBIAS$ <±25%. The model performance on finer time steps (e.g. daily) is usually poorer than the coarser time steps model (e.g. monthly) in terms of the statistical matrices (e.g. *NS*, *RST*, *PBIAS*) (Moriasi et al., 2007, Engel et al., 2007). For instance, while the monthly *NS* was 0.656 for the calibration period in Fernandez et al. (2005), the daily one was 0.395. Moreover, Moriasi et al. (2007) indicated that when reviewing previous studies, *NS* and *PBIAS* were “as expected” lower in the validation period than the calibration period for streamflow. Note that the Midway station was our primary calibration, since all of the model’s streamflow forecasts occurred from this location. The model we established for South Elkhorn watershed showed results for the Midway gage station to have *NS* values equal to 0.9 and 0.66 for the monthly and daily time steps, respectively, for the calibration period; and *NS* values equal to 0.88 and 0.46 for the monthly and daily time steps, respectively, for the validation period. In summary, the metrics showed adequate performance considering the above information and results.

<Figure 6 here please>

<Table 2 here please>

Results from fitting both the AM and POT extreme series methods to the streamflow results showed that in general the extreme series results had little mean trend and were dominated by the two parameter probability distributions (see Supplementary On-line Table). The Mann-Kendall test results showed that only 2% of the AM series included a mean trend that required removal and only 4% of POT series results had a mean trend that required removal. A regression approach was also carried out and provided identical results as the Mann-Kendall tests. The results highlight that although non-stationarity is exhibited when comparing extremes from the hindcast to the forecast periods, little significant non-stationarity is exhibited within the simulation periods. Statistical results showed that 91% of the AM series best followed the two-parameter Gumbel distribution while 85% of the POT series best followed the exponential distribution. The results tend to agree with the results of Dankers and Feyen (2008) who also found that a two parameter distribution was most adequate when fitting distributions from extreme value theory to streamflow results derived from global climate modeling. Additional results from the extreme value analyses is also compiled in the Supplemental On-line Table and includes: threshold selections, the value of the extremal index θ before de-clustering, the value of r required to make the extremal index θ equal to unity, the p-value of Mann-Kendall non-parametric test, and the resultant sample size (n).

We found less than 10% difference between observed and simulated maxima for all return periods (i.e., 2, 20 and 100 year return periods) for both AM and POT methods. Both observed and simulated maxima followed exponential distributions for the POT method; and both followed the Gumbel distribution for the AM method. Donigan (2002) indicates that an absolute hydrologic model calibration/validation target of less than 10% difference between the simulated and the observed hydrology flow is considered a very good target; and that the range of such target should be applied on the mean and the individual events may show larger differences while still acceptable. With this criteria in mind, our SWAT evaluation results for the extremes were deemed adequate.

Extreme quantiles for 2-, 20-, and 100-year maxima streamflow levels showed that forecast results were in general greater than hindcast results for simulation pairs with the same climate modeling factors, highlighting the non-stationarity of extremes mentioned previously. Figure 7 illustrates hindcast simulations corresponding to POT extreme series method, and all simulation results are shown in the Supplementary On-line Table. Statistical downscaling of the hindcast GCM realizations in general under-predicted hydrologic model results analyzed with the extreme series method; and the under-prediction was especially true for streamflow levels from the 100-year return period. Results from the dynamical downscaling hindcast realizations better bound the observed extremes. The result supports the idea that regional climate models can capture small-scale climate features, e.g., strong fronts, and realistically simulate extreme events (Fowler et al., 2007, Warner, 2010), which would suggest a better choice for extreme streamflow forecasting. Fowler et al. (2007) pointed out that the statistical downscaling methods poorly represent the extreme events and underestimate variance, which reflects the fact that both BCSD and BCCA methods use the distribution of precipitation from historical climate records to create the future distributions. Warner (2010) compared the statistical and dynamical downscaling with respect to their advantages and disadvantages, and he indicated that dynamical downscaling methods could better capture extreme events and variance. Sunyer et al. (2015) shows that the RCM-GCM projections are the main source of variability in their results, and between 30-50% of the total variance is explained by statistical downscaling in several catchments in their study. Trayhorn and DeGaetano (2001) compared several different downscaling methods for rainfall extremes over the Northeastern United States; and their results suggest that regional climate models overestimate the observed extremes. Aside from the Trayhorn and DeGaetano (2001) results, literature results and this study generally support the idea that hindcast extremes from dynamic downscaling agree better with observed extremes as compared to statistical downscaling results.

<Figure 7 here please>

We also examined specific results of individual climate models and downscaling methods in order to provide insight on how climate model structure may be impacting forecasted streamflow maxima. The four GCMs from CMIP3 all illustrate differences when comparing across the 2, 20 and 100 year return periods (Figure 7). The result was not surprising given that GCM has been found as a significant factor in studies of forecasted mean streamflow and precipitation, and climate scientists highlight variability of GCMs due to the differences in the models' structures and parameterizations (Randall et al., 2007; Weart, 2010; Mearns et al., 2013; Melillo et al., 2014; Al Aamery et al., 2016). Given the many differences between the four GCMs, it is difficult to discern specific processes represented within the climate models that might be controlling the extreme streamflow forecasts, however, direct comparison of CMIP3 and CMIP5 model versions provided some discussion.

Figure 7 reveals that CCSM has a pronounced difference between CMIP3 and CMIP5 forecasted streamflow maxima while the other GCMs (Had, GFDL and CGCM) do not show differences between model versions for our analyses. The reason is perhaps attributed to the newer version CCSM4 that produces El Nino-Southern Oscillation (ENSO) variability in a more realistic frequency distribution than CCSM3 by changing the deep convection scheme.

The Had, GFDL and CGCM models also made changes from CMIP3 to CMIP5 but these tend to have little differences in terms of streamflow extremes (Figure 7). The HadGEM2 of CMIP5 improved the performance of ENSO, northern continent land-surface temperature biases, SSTs, and wind stress compared to the previous models; however, Collins et al. (2008) suggests that the power spectrum of El Nino was not a substantial improvement. GFDL version 3 (CM3) used in CMIP5 made minimal changes to the ocean and sea ice models compared to those used in CM2.1 version of CMIP3; however, the newer version is extensively developed the atmosphere and land model components (Griffies et al., 2011). CanESM2 of CMIP5 combines the fourth generation atmospheric general circulation model (CanCM4) with terrestrial carbon cycle model (CTEM). Compared to the third generation of CanCM3 that was used in CGCM3.1 of CMIP3, CanCM4 is different in many aspects such as the finer resolution and the addition of new schemes such as shallow convection scheme (see Chylek et al., 2011).

Taken together, of all the changes to the four different GCMs between CMIP3 and CMIP5, only augmenting ENSO within the GCM seems to have a substantial impact on forecasted streamflow maxima. The suggestion is reasonable given that ENSO has been suggested to show significant impacts on precipitation in this region of North America (Gabler et al., 2009). Results suggest that the El Nino-Southern Oscillation and its representation within climate modeling may exhibit a substantial control on forecasting streamflow maxima for the wet temperate study region; and additional emphasis upon oscillations when forecasting streamflow maxima in wet temperate regions may be fruitful.

5.2 Uncertainty from climate and extreme modeling factors

Variance analysis results determined *via* ANOVA showed that the variance structure of forecasted streamflow maxima exhibits some dependence on all of the climate modeling considered factors but does not exhibit dependence upon the extreme value method applied (see Figure 8). The results are interesting due the fact that previous variance analysis of mean streamflow forecasted from GCMs only showed dependence on a subset of the climate modeling factors while debate in the literature suggests that AM and POT methods would give different results (Scarrott and MacDonald, 2012; Bezak et al., 2014; Al Aamery et al., 2016). Specifically, results of the ANOVA (Figure 8) show that variance of the 2 year and 20 year streamflow maxima are significantly dependent upon GCM type, downscaling method, emission scenario, GCM project phase, and bias implementation; and variance of the 100 year streamflow maxima is significantly dependent upon GCM type, GCM project phase, and bias implementation. For reference, results of forecasted mean streamflow are included in Figure 8 and show dependence on GCM type and phase and downscaling.

<Figure 8 here please>

The climate modeling factors that significantly influenced the forecasted streamflow maxima variances were ranked using the weighted F-value according to their variance contribution (see Figure 8) as GCM type, downscaling method, bias implementation, GCM version associated with the climate project phase, and the emission scenario input to the GCM. Results of the ANN non-linear variance analysis compared well with linear analysis *via* ANOVA (see comparisons in Figure 9) providing further confidence in our ranking results.

<Figure 9 here please>

In addition to the variance breakdown, the total variance of the forecasted extremes also displays pertinent information. The total variance of streamflow extremes increased substantially with return period—a result most easily observed with the standard error bars in Figure 10. In addition, the proportion of the variance that was predictable with the climate modeling factors tended to decrease with return period. The result suggests a propagation of unexplainable variance throughout the analysis that becomes more pronounced with the higher order extremes associated with higher return periods.

<Figure 10 here please>

We at least partially attribute the pronounced growth of uncertainty with return period to fitting the extreme value distributions to the hydrologic results. The 100 year return period falls at the tail end of the GEV and GP distributions (i.e., f=0.99) and therefore uncertainty introduced in fitting the distributions will be most pronounced for the highest return periods. To illustrate the point, we performed sensitivity of the extreme value parameterization method by assuming a known parent Gumbel distribution for M_n, drawing sets of realizations consistent with the years of data in our analyses, and fitting the extreme value distribution consistent with the maximum likelihood method of our analyses as well as typically performed by others (e.g., Gilleland and Katz, 2006). Results from the sensitivity show that the variance associated with the 100 year streamflow is about five times greater than that of the 2 year streamflow event (see Table 3). The result highlights one reason for pronounced increases in unexplainable variance within forecasted streamflow maxima.

<Table 3 here please>

On the other hand, factorial comparison between the AM and POT series fitted by the General Extreme Value (GEV) and General Pareto (GP) distributions did not show significance within the analysis of variance results. The result is surprising given recent debate and critique of each method, e.g., AM is criticized for its neglect of multiple extremes per annum while POT has been criticized for subjectivity of threshold selection (Svensson et al., 2005; Scarrott and MacDonald, 2012; Bezak et al., 2014; Fischer and Schumann, 2016). However, further investigation of the literature suggests that the variance analysis result is consistent with fundamental theory and that the methods might be used interchangeably, as needed, so long as care is taken in their application. Fundamentally, Coles (2001) shows that the GEV distribution provides the base that can be used to derive the GP distribution so long as the threshold point is sufficiently large and the events are independent and random. In this manner, we recommend that future coupled hydrologic and climate research studies that apply the POT method should strive for relatively high threshold values that fall within the Rules of Thumb outlined by Scarrott and MacDonald (2012) and ensure that the extremal index is not less than one (see Figure 3).

One noteworthy comparison of the present study's results with previously published results is that the variance of forecasted streamflow maxima is even more sensitive to climate modeling factors as compared to the variance of mean forecasted streamflow. Specifically, the variance of streamflow maxima showed significant dependence upon the choice of emission scenario and bias correction approach (see Figure 8) while the variance of mean streamflow did not exhibit significant dependence upon emission and bias (see Al Aamery et al., 2016 and results summarized in Figure 8). The streamflow maxima's dependence upon emission scenario is worthy of mentioning given that the mean atmospheric CO_2 concentration projected for the emission scenarios varies by just ±50 ppm for 2050 (IPCC, 2001; Meinshausen et al., 2011). Further, the mean annual temperature has a total range of just 1.5°C for 2050 across emission scenarios projected within the GCMs applied in this study and the mean streamflow study of Al Aamery et al. (2016). The subtle mean changes in CO_2 and MAT for 2050 appear to mask temporal anomalies captured within the GCMs. The potential of emissions to help control streamflow maxima is somewhat corroborated by the work of Mantua et al. (2010) where they show streamflow maxima differences among two emission scenarios. Significance of emission scenario within variance analysis of forecasted streamflow maxima suggests that hydrologic and climate research is needed that examines how models might be coupled at a higher temporal resolution, rather than the more prevalent emphasis on mean coupling (e.g., see review Table 1 in Al Aamery et al., 2016). Similarly, the significance of bias correction upon the variance of forecasted streamflow maxima reflects the boundary between climate and hydrologic models that has emphasized mean coupling and thus linear shifts in rainfall and temperature data to show agreement with observations (Lenderink et al., 2007). More sophisticated bias correction methods are available (Teutschbein and Seibert, 2012) but typically come with the added conundrum of forcing functional constraints on climate model results that are sought after due to their non-stationarity. Surely, research might consider higher resolution model coupling to understand anomalies that control maxima streamflow.

5.3 Uncertainty from hydrologic modeling

Future climate change of wind speed, net radiation, and relative humidity were tested within hydrologic simulation by considering projected shifts reported in the literature. Results suggest that the net impact of wind speed, net radiation, and relative humidity could provide an additional 1 to 5% increase in streamflow maxima for 2, 20 and 100 year return periods for the wet temperate study region and future period considered (see Table 4). Average daily change in evapotranspiration ranged from 0.5 to 5% decreases. Streamflow maxima increases and standard error associated with the wind speed, radiation and relative humidity shifts were +3.2(±1.7), +2.2(±1.6) and +1.9(±1.6)% for 2, 20 and 100 year events. Relative to the increases of +27(±21), +36(±34) and +49(±85)% for streamflow maxima associated with GCM-projection of precipitation and temperature from ensemble analysis (see Figure 10), the effect of wind speed, net radiation and relative humidity were small for this region. Nevertheless, the effect is non-zero; and the variables may be more substantial in other regions or for forecasting to 2100.

<Table 4 here please>

Future projections that considered hydrologic model fit and hydrologic modeling uncertainty were also tested to investigate their impact on the relative change of streamflow maxima. The future streamflow maxima produced from the calibrated hydrologic model simulation was compared against the future streamflow maxima produced using the un-calibrated (i.e., default) parameterization of the hydrologic model for the realization pairs for the AM extreme value analysis method (*n*=74). Results for the uncalibrated hydrologic analysis of the relative change in streamflow maxima were +19(±28), +20(±35) and +24(±59)% for 2, 20 and 100 year events in comparison to the calibrated model results equal to +27(±23), +35(30) and +49(±92)% for 2, 20 and 100 year events. Results show that the uncalibrated model gives a much lower increase in future streamflow maxima compared to the calibrated model results, especially for the 100 year extreme. Note that the default model simulations tended to under-predict streamflow during peak events. The simulation bias is carried forward to the extreme modeling results and is not removed when considering the relative change. In this manner, the variance of the streamflow maxima was dependent on hydrologic model parameterization. These results contrast the work of Niraula et al. (2015) where we showed that the relative change in mean forecasted streamflow was not dependent on parameter selection during calibration. The results further highlight the variance structure's sensitivity when forecasting streamflow extremes.

Given the dependence on hydrologic calibration, the hydrologic uncertainty realizations were also performed. Results suggest that hydrologic model parameter sets generated during uncertainty analysis also impart variance upon relative changes in streamflow maxima. We calculated the error associated with the relative change in streamflow maxima using the parameter sets within SWAT-CUP that met model objective function criteria. Standard error was 3.1, 3.3 and 3.6% for the relative change of 2, 20 and 100 year events. Standard error is small in comparison to the error produced from climate and extreme modeling factors. Nevertheless the error is nonzero and may be larger for other regions. We also calculated the standard error from absolute forecasted streamflow maxima and found values of 11, 21, and 27 cms for 2, 20 and 100 year events. We compared these values with the standard error from direct bias-correction of the streamflow maxima via the relative change approach, and the standard error was 3, 6 and 9 cms for 2, 20 and 100 year events. The results highlight that the delta method applied to the direct observed streamflow via the relative change does reduce hydrologic uncertainty relative to the absolute forecasts.

5.4 Forecast of streamflow maxima for wet temperate regions

One corollary of variance analysis is inclusion of significant factors impacting prediction and thus forecasting of future streamflow. The relative change in streamflow maxima were increases of +30(±21), +38(±34) and +51(±85)% for the study region for 2, 20 and 100 year events. The increases are substantially larger than the 11% increases found for mean streamflow and mean precipitation for the study region (Al Aamery et al., 2016). Additionally, streamflow maxima increases as a function of return period. The variability of the projections is pronounced, and the uncertainty from climate and extreme model factors dominates the variance (see Table 5).

<Table 5 here please>

The forecasted results of increased maxima streamflow in 2050 for the wet temperate region of North America (1120 mm y^{-1}) is in agreement with scientific sentiment and forecasting that wet regions will get wetter and wet time periods will be wetter (Melillo et al., 2014). We performed analysis of published maxima streamflow forecasts in wet regions of Europe and their comparison corroborated the finding that maxima streamflow increases as a function of return period. Analysis of the results from Lawrence and Hisdal (2011) show an increase of maxima streamflow as a function of return period for Norway (760-2250 mm y^{-1}). Also, analysis of the results from Dankers and Feyen (2008) show an increase of maxima streamflow as a function of return period for their European sites studied where the mean annual precipitation was greater than 500 mm per year and is projected to be less in the end of this century.

The finding that forecasted maxima streamflow may show further increases as a function of return period further supports general scientific agreement that the most extreme flooding events will get even more extreme for wet temperate climates (Melillo et al., 2014). This concept is reflected in the timing of streamflow increases and extremities in the present study, and Table 6 shows that the months of the year with the highest future changes in mean precipitation and streamflow tend to also account for the majority of forecasted streamflow maxima events during the study period. The results also reflect the fundamental scientific consequences of climate change. That is, increased precipitation in wet regions is expected due to higher amounts of moisture in the atmosphere due to warmer atmospheric temperatures and expansion of the high Sub-tropical Belt as the air temperature increases and moist air is transported to higher and lower latitudes (Gabler et al., 2009; Melillo et al., 2014). In turn, climate change in wet temperate region may increase precipitation, temperature, and relative humidity while decreasing wind speed and net radiation, and the net effect both individually and cumulatively of all these shifts is an increase in streamflow maxima.

<Table 6 here please>

6 CONCLUSION

Formatted: Font: Bold

The main conclusions of our work are described as follows:

(1) Model simulation and evaluation results from comparison of different global climate model downscaling methods suggests that dynamic downscaling results more closely align with observations, presumably due to the explicit simulation of small-scale features such as strong fronts. Comparison of streamflow maxima forecasted with paired climate models from CMIP3 versus CMIP5 projects suggest that the El Nino-Southern Oscillation representation within modeling exhibits a control on forecasting streamflow maxima for the wet temperate region studied.

(2) Uncertainty from climate and extreme modeling factors was evaluated and showed that the relative change of streamflow maxima was not dependent on systematic variance from the annual maxima versus peak over threshold method applied. We find that the variance of streamflow maxima is an increasing function of the return period, which is at least partly attributed to fitting the extreme value distributions to the hydrologic model results. The variance of the relative change in streamflow maxima is dependent upon global climate model, emission scenario, project phase, downscaling, and bias correction.

(3) Uncertainty from hydrologic modelling was analyzed and unlike results from previous research focused on the relative change of mean streamflow, the relative change of streamflow maxima was dependent on hydrologic model fit and modeling uncertainty. The streamflow maxima also showed some dependence on climate projections of wind speed, net radiation and relative humidity.

(4) Ensemble projections forecast an increase of streamflow maxima for 2050 with pronounced forecast standard error, including +30(±21), +38(±34) and +51(±85)% for 2, 20 and 100 year events for the wet temperate region studied. The variance of maxima projections was dominated by climate model factors and extreme value analyses with lesser control from hydrologic inputs and hydrologic model parameterization.

ACKNOWLEDGEMENTS:

We thank Editor McVicar, ~~the anonymous~~ Associate Editor Yongqiang Zhang, and anonymous reviewers for their excellent input and comments, which in turn helped improve the quality of this paper. We acknowledge partial support from NSF Award Number 1632888 to assist with study of the watershed and river system.

REFERENCES:

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., & Srinivasan, R. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of hydrology*, *333*(2), 413-430.

Al Aamery, N., Fox, J. F., & Snyder, M. (2016). Evaluation of climate modeling factors impacting the variance of streamflow. *Journal of Hydrology*, *542*, 125-142.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73-89.

Arnold, J. G., Allen, P. M., Volk, M., Williams, J. R., & Bosch, D. D. (2010). Assessment of different representations of spatial variability on SWAT model performance. *Transactions of the ASABE*, *53*(5), 1433-1443.

Beguería, S., & Vicente-Serrano, S. M. (2006). Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of applied meteorology and climatology*, *45*(1), 108-124.

Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, *59*(5), 959-977.

Brekke, L., Thrasher, B. L., Maurer, E. P., & Pruitt, T. (2013). Downscaled CMIP3 and CMIP5 climate projections. *US Dep. Inter. Bur. Reclamation, Tech. Serv. Cent*.

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part II: Preliminary model validation. *Monthly Weather Review*, *129*(4), 587-604.

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, *401*(3), 190-202.

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian earth system model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22893-22907.

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). London: Springer.

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Coney, S.C., Hack, J.J., Henderson, T.B. and Kiehl, J. T. (2006). The community climate system model version 3 (CCSM3). *Journal of Climate*, *19*(11), 2122-2143.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Hinton, T., Jones, C. D., ... & Senior, C. (2008). Evaluation of the HadGEM2 model. *Hadley Cent. Tech. Note*, *74*.

Dankers, R., & Feyen, L. (2008). Climate change impact on flood hazard in Europe: An assessment based on high- resolution climate simulations. *Journal of Geophysical Research: Atmospheres*, *113*(D19).

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., ... & Durachta, J. W. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643-674.

Donigan, A. S. (2002). Watershed model calibration and validation: The HSPF experience. *Proceedings of the Water Environment Federation*, *2002*(8), 44-73.

Donohue, R. J., McVicar, T. R., & Roderick, M. L. (2010). Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *Journal of Hydrology*, 386(1-4), 186-197.

Engel, B., Storm, D., White, M., Arnold, J., & Arabi, M. (2007). A hydrologic/water quality model applicati1. *Journal of the American Water Resources Association*, 43(5), 1223-1236.

Fatichi, S., Rimkus, S., Burlando, P., & Bordoy, R. (2014). Does internal climate variability overwhelm climate change signals in streamflow? The upper Po and Rhone basin case studies. *Science of the Total Environment*, *493*, 1171-1182.

Fernandez, G. P., Chescheir, G. M., Skaggs, R. W., & Amatya, D. M. (2005). Development and testing of watershed-scale models for poorly drained soils. *Transactions of the ASAE*, 48(2), 639-652.

Fischer, S., & Schumann, A. (2016). Robust flood statistics: comparison of peak over threshold approaches based on monthly maxima and TL-moments. *Hydrological Sciences Journal*, *61*(3), 457-470.

Flato, G. M. (2005). The third generation coupled global climate model (CGCM3). *Available on line at http://www. cccma. bc. ec. gc. ca/models/cgcm3. shtml*.

Ford, W. I., Fox, J. F., & Rowe, H. (2015). Impact of extreme hydrologic disturbance upon the sediment carbon quality in agriculturally impacted temperate streams. *Ecohydrology*, *8*(3), 438-449.

Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International journal of climatology*, *27*(12), 1547-1578.

Gabler, R. E., Peter, J. F., Trapson, M., & Sack, D. (2009). Physical Geography: Brooks/Cole. *Belmont, USA*.

Gassman, P. W., Reyes, M. R., Green, C. H., & Arnold, J. G. (2007). The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASABE*, *50*(4), 1211-1250.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., ... & Worley, P. H. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973-4991.

Gilleland, E., & Katz, R. W. (2016). Extremes 2.0: an extreme value analysis package in r. *Journal of Statistical Software*, *72*(8), 1-39.

Giorgi, F., Marinucci, M.R., Bates, G.T. (1993). Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. *Mon. Weather Rev. 121*, 2794–2813.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., ... & Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate dynamics*, *16*(2), 147-168.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th edition). Belmont, CA: Wadsworth.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., ... & Palter, J. B. (2011). The GFDL CM3 coupled climate model: characteristics of the ocean and sea ice simulations. *Journal of Climate*, *24*(13), 3520-3544.

Haan, C. T. (2002). *Statistical methods in hydrology*. The Iowa State University Press.

Harding, B. L., Wood, A. W., & Prairie, J. R. (2012). The implications of climate change scenario selection for future streamflow projection in the Upper Colorado River Basin. *Hydrology and Earth System Sciences*, *16*(11), 3989.

Hogg, R. V., Tanis, E., & Zimmerman, D. (2014). *Probability and statistical inference*. Pearson Higher Ed.

IBM.(2012). IBM SPSS Neural Networks 21. Retrieved from http://www.sussex.ac.uk/its/pdfs/SPSS_Neural_Network_21

IPCC, 2001. Climate Change 2001: The Scientific Basis. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. (Eds.), Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 881 pp.

IPCC, 2007. Climate Change 2007: The Physical Science Basis. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996 pp.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 1535 pp.

Juang, H., Hong, S., Kanamitsu, M. (1997). The NMC nested regional spectral model: an update. *Bull. Am. Meteorol. Soc. 78*, 2125–2143.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of hydrology*, *329*(3), 534-552.

Knutti, R., & Sedláček, J. (2013). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, *3*(4), 369-373.

Lawrence, D., & Hisdal, H. (2011). Hydrological projections for floods in Norway under a future climate. *NVE Report*, *5*.

Lenderink, G., Buishand, A., & Deursen, W. V. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, *11*(3), 1145-1159.

Leopold, L. B., Wolman, M. G., & Miller, J. P. (2012). *Fluvial processes in geomorphology*. Courier Corporation.

Lima, C. H., Lall, U., Troy, T. J., & Devineni, N. (2015). A climate informed model for nonstationary flood risk prediction: Application to Negro River at Manaus, Amazonia. *Journal of Hydrology*, *522*, 594-602.

Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At- site modeling. *Water resources research*, *33*(4), 747-757.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, *519*, 3634-3650.

Mantua, N., Tohver, I., & Hamlet, A. (2010). Climate change impacts on streamflow extremes and summertime stream temperature and their possible consequences for freshwater salmon habitat in Washington State. *Climatic Change*, *102*(1), 187-223.

McMichael, A. J., Powles, J. W., Butler, C. D., & Uauy, R. (2007). Food, livestock production, energy, climate change, and health. *The lancet*, *370*(9594), 1253-1263.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., ... & Mescherskaya, A. V. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, 416, 182-205.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., ... & Thomson, A. G. J. M. V. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change*, *109*(1-2), 213.

Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe, Eds., 2014: *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.

Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., ... & Snyder, M. (2013). Climate change projections of the North American regional climate change assessment program (NARCCAP). *Climatic Change*, *120*(4), 965-975.

Mirza, M. M. Q. (2003). Climate change and extreme weather events: can developing countries adapt?. *Climate policy*, *3*(3), 233-248.

Moradkhani, H. (2017, May). *State-of-the-art Uncertainty Analysis in Hydroclimate Modeling Panel Discussion.* Presentation session presented at the meeting of the World Environmental& Water Resources Congress, Sacramento, CA.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

Niraula, R., Meixner, T., & Norman, L. M. (2015). Determining the importance of model calibration for forecasting absolute/relative changes in streamflow from LULC and climate changes. *Journal of Hydrology*, 522, 439-451.

Palanisamy, B., & Workman, S. R. (2014). Hydrologic modeling of flow through sinkholes located in streambeds of Cane Run Stream, Kentucky. *Journal of Hydrologic Engineering*, *20*(5), 04014066.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).

Plummer, D. A., Caya, D., Frigon, A., Côté, H., Giguère, M., Paquin, D., ... & De Elia, R. (2006). Climate and climate change over North America as simulated by the Canadian RCM. *Journal of Climate*, *19*(13), 3112-3132.

Prudhomme, C., Jakob, D., & Svensson, C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. *Journal of hydrology*, *277*(1), 1-23.

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: Cilmate 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.

Reiss, R. D., Thomas, M., (2007). *Statistical analysis of extreme values* (Vol. 2). Basel: Birkhäuser.

Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2001). Climate change and extreme weather events; implications for food production, plant diseases, and pests. *Global change & human health*, *2*(2), 90-104.

Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal*, *10*(1), 33-60.

Shamir, E., Megdal, S. B., Carrillo, C., Castro, C. L., Chang, H. I., Chief, K., ... & Prietto, J. (2015). Climate change and water resources management in the Upper Santa Cruz River, Arizona. *Journal of Hydrology*, *521*, 18-33.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J. G. (2005). A description of the advanced research WRF version 2. *NCAR Tech. Note NCAR/TN-468+STR*, 88 pp.

Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences* (3rd edition). Mahway, NJ: Lawrence Erlbaum.

Sunyer Pinya, M. A., Hundecha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., ... & Loukas, A. (2015). Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrology and Earth System Sciences*, *19*(4), 1827-1847.

Svensson, C., Kundzewicz, W. Z., & Maurer, T. (2005). Trend detection in river flow series: 2. Flood and low-flow index series/Détection de tendance dans des séries de débit fluvial: 2. Séries d'indices de crue et d'étiage. *Hydrological Sciences Journal*, *50*(5).

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456*, 12-29.

Tryhorn, L., & DeGaetano, A. (2011). A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. *International Journal of Climatology*, *31*(13), 1975-1989.

Tufféry, S. (2011). *Data mining and statistics for decision making* (Vol. 2). Chichester: Wiley.

Warner, T. T. (2010). *Numerical weather and climate prediction*. Cambridge University Press.

Weart, S. (2010). The development of general circulation models of climate. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *41*(3), 208-217.

Wilby, R. L., & Dawson, C. W. (2007). SDSM 4.2-A decision support tool for the assessment of regional climate change impacts. *User Manual. London, UK*.

Wild, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres*, 114(D10).

Willett, K. M., Jones, P. D., Gillett, N. P., & Thorne, P. W. (2008). Recent changes in surface humidity: Development of the HadCRUH dataset. *Journal of Climate*, 21(20), 5364-5383.

Yuan, X., & Wood, E. F. (2012). Downscaling precipitation or bias- correcting streamflow? Some implications for coupled general circulation model (CGCM)- based ensemble seasonal hydrologic forecast. *Water Resources Research*, *48*(12).

Zhang, X., Xu, Y. P., & Fu, G. (2014). Uncertainties in SWAT extreme flow simulation under climate change. *Journal of hydrology*, *515*, 205-222.