HYDROLOGICAL PROCESSES
Hydrol. Process. (2012)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9376

Analysis of precipitation extremes with the assessment of regional climate models over the Willamette River Basin, USA

Andrew Halmstad, Mohammad Reza Najafi and Hamid Moradkhani*

Department of Civil and Environmental Engineering, Portland State University, Portland, OR, USA

Abstract:

An appropriate, rapid and effective response to extreme precipitation and any potential flood disaster is essential. Providing an accurate estimate of future changes to such extreme events due to climate change are crucial for responsible decision making in flood risk management given the predictive uncertainties. The objective of this article is to provide a comparison of dynamically downscaled climate models simulations from multiple model including 12 different combinations of General Circulation Model (GCM)–regional climate model (RCM), which offers an abundance of additional data sets. The three major aspects of this study include the bias correction of RCM scenarios, the application of a newly developed performance metric and the extreme value analysis of future precipitation. The dynamically downscaled data sets reveal a positive overall bias that is removed through quantile mapping bias correction method. The added value index was calculated to evaluate the models' simulations. Results from this metric reveal that not all of the RCMs outperform their host GCMs in terms of correlation skill. Extreme value theory was applied to both historic, 1980–1998, and future, 2038–2069, daily data sets to provide estimates of changes to 2- and 25-year return level precipitation events. The generalized Pareto distribution was used for this purpose. The Willamette River basin was selected as the study region for analysis because of its topographical variability and tendency for significant precipitation. The extreme value analysis results showed significant differences between model runs for both historical and future periods with considerable spatial variability in precipitation extremes. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS precipitation extreme; climate change; dynamic downscaling; bias correction; NARCCAP

Received 18 December 2011; Accepted 29 March 2012

INTRODUCTION

Among the potentially significant impacts of future climate change, the spatial and temporal variation of precipitation extremes, in terms of intensity and frequency, is of paramount importance to water resources engineers and decision makers. The primary approach for evaluating potential changes to hydroclimatic variables is through the use of climate models that simulate aspects of the global climate cycle. During the last three decades, the number and complexity of climate models have increased substantially, more physical processes are simulated and the coupling between individual sea, atmosphere and land-based processes has been improved (Xue *et al*., 2001; Wang *et al*., 2004; Diffenbaugh *et al*., 2005; Mearns, 2007; Solomon, 2007; Mearns *et al*., 2011; Yuan and Liang, 2011). Recent advancements in modelling spatial and temporal climate variables at finer scales allow for regional impact analysis studies. The ability to investigate the impact of climatic change at a regional scale has the potential to inform water resource managers and decision makers regarding the changes in the climate cycle that will influence extreme events, that is, floods and droughts. To provide valuable information regarding potential climatic changes, the results from multiple climate model simulations can be investigated and compared with reduce the overreliance on one model and quantify model uncertainty.

During the last two decades, numerous improvements in the field of climate change research have bolstered confidence in the predictive capability of climate models. Through increased international research efforts made possible by initiatives such as multimodel ensemble investigation projects (described in greater detail in the section on Multimodel Ensemble Projections), climate models have undergone extensive analysis by an increasing number of investigators at virtually all levels of research (Solomon, 2007). All major component phases (atmospheric, oceanic and terrestrial) have seen improvement in terms of model formulation (improved transport and dynamics schemes), increased resolution (vertical, horizontal and temporal) and represented processes (such as direct and indirect aerosol effects) as well as many other aspects (Solomon, 2007). Most notably for this study, the overall distribution of precipitation and the capability of models to simulate extreme events are noted by the Intergovernmental Panel on Climate Change-Fourth Assessment Report (IPCC-AR4) as areas, which have seen improvement.

Within the climate modelling community, it has long been speculated that increasing the resolution of climate models is necessary to improve the estimates of regional-scale phenomena, such as precipitation (e.g. Giorgi, 1990; McGregor, 1997; Murphy, 1999; Caldwell, 2010; Di Luca *et al*., 2011;). The process of downscaling outputs from GCMs has been established as the primary approach

*Correspondence to: Hamid Moradkhani, Department of Civil and Environmental Engineering, Portland State University, Portland, OR, USA.
E-mail: hamidm@cecs.pdx.edu

Copyright © 2012 John Wiley & Sons, Ltd.

for addressing the inadequacies of large-scale resolution models. There are two main classes of downscaling procedures: statistical and dynamical. Numerous studies during the last several decades have provided detailed comparisons and exploration of both downscaling types (e.g. Giorgi, 1990; McGregor, 1997; Murphy, 1999; Caldwell, 2010; Di Luca *et al.*, 2011; Najafi *et al.*, 2011a). Statistical approaches involve determining reliable statistical relationships between large-scale climate variables that are well represented by GCMs, such as pressure fields, and local-scale variables, such as temperature or precipitation (Najafi *et al.*, 2011a; Najafi *et al.*, 2011c). There is currently an extensive variety of statistically based approaches (for a more comprehensive review, see Fowler *et al.*, 2007; Salathe *et al.*, 2007). Dynamic downscaling approaches, on the other hand, are based on the same numerical integration of differential equations, as in GCMs, but over a smaller spatial and temporal domain. Furthermore, dynamical downscaling approaches can include modified physical schemes that have been demonstrated to better address topographical variations (Yuan and Liang, 2011). Given recent advancements in computational efficiency and resources, dynamic downscaling, via regional climate models (RCMs), has expanded to the point where numerous RCMs exist and the need for multimodel comparison is beginning to be addressed (Mitchell, 2009; Van der Linden and Mearns *et al.*, 2009; Kendon *et al.*, 2010).

Caldwell (2010) found that RCMs tend to overpredict precipitation estimates and that, contrary to expectations, 'improved resolution does not translate into improved simulation. . ..' For that study, the author investigated the performance of gridded observational data sets, RCMs and GCMs ranging in spatial resolution from 1/12th of a degree up to 4.5 F in terms of their ability to reproduce wintertime precipitation over the state of California. It was noted that the removal of bias from the climate models was critical.

Di Luca *et al.* (2011) showed that temporal scale is one aspect where RCMs do provide noticeable improvements compared with coarser resolution models. The central concern of that study was to identify a manner of objectively quantifying the amount of information gained from RCM efforts. As pointed out by the authors, although RCM simulations may not add substantive value across all aspects of climate change predictions, identifying areas where they do add significant information should be an area of greater concern and research. Instead of concentrating on whether RCMs improve desirable climate-related information at all spatial and temporal scales, it would be more beneficial for resources to be focused on identifying those aspects that are improved via RCM simulation, thereby resulting in more skilful impact and adaptation investigations. Results of that study reveal several aspects where the added value of RCMs is noticeable, including shorter temporal scale and warmer seasons and in regions of complex topology (Di Luca *et al.*, 2011).

With the previously mentioned findings, the development of a performance metric that describes the ability of RCMs to improve upon GCM simulations is a current field of research for those investigating regional climatology and the impact of climate variation. Kanamitsu and DeHaan (2011) provided a concise discussion of the merits and shortcomings of commonly used performance metrics concerning climate model simulations. The authors pointed out that deficiencies exist in several aspects of currently used performance metrics and propose a novel measure intended to accurately capture the geographic distribution of a model's skill, thereby yielding a quantitative and descriptive measure of the ability of high resolution models to add information in particular regions.

Model bias exists within climate models for multiple reasons. Commonly identified causes of bias are attributed to model structure and initial/forcing condition treatment. To more accurately compare historic and future climate model simulations, recent studies suggest the use of bias correction techniques such as the quantile mapping approach (e.g. Fowler *et al.*, 2007; Mote and Salathe, 2010; Shrestha *et al.*, 2011). The need for bias correction of climate model simulations over future periods is widely accepted for hydrologic impact studies (Wood *et al.*, 2004); however, the relative strengths and weaknesses of each individual correction technique are still a focus of research (Johnson and Sharma, 2011; Najafi *et al.*, 2011b).

Because of the climate change impact on hydroclimatic events, the study of the ability of climate models to capture and simulate extreme precipitation events is of paramount importance. Frei *et al.* (2006) and Villarini *et al.* (2011) studied the variability in both observed and climate model simulated extreme event occurrence. The potential for change in the occurrence of extreme events in the future in conjunction with projected climate variability is also prevalent in recent studies (Mote and Salathe, 2010; Tryhorn and DeGaetano, 2011). To evaluate the characteristics of extreme events, statistical extreme value theory (EVT) is commonly used in water resource and hydrology-related studies in recent decades (Katz *et al.*, 2002). EVT is a statistical method used for analyzing the tails of probability distributions of a random variable. The parameters of the extreme value distributions yield estimates of the intensity and frequency of extreme events. As such, the distributions can be used for estimating the magnitude of extreme event return values in the future and historical periods to investigate the climate change impact on hydroclimate extremes (Katz, 2010). EVT has been widely applied in studies of precipitation (Katz *et al.*, 2002; Rusticucci and Tencer, 2008 Acero *et al.*, 2010; Wehner *et al.*, 2010 Kharin *et al.*, 2010a; Kharin *et al.*, 2010b;), temperature (Kharin *et al.*, 2010b;), streamflow (Katz *et al.*, 2002; Hurkmans *et al.*, 2010; Lima and Lall, 2010) and wind speed (Hundecha *et al.*, 2008 Brabson and Palutikof, 2010; Caires and Sterl, 2010;), among others.

This article is organized as follows: a brief introduction illustrating the relevance and growth of regionally based climate model ensemble projects is provided in the section on Multimodel Ensemble Projections. The sections on Study Area and Data introduce the chosen study area as well as the data sets used, respectively. The section on Methodology provides a more detailed

Copyright © 2012 John Wiley & Sons, Ltd.

description of the methods used in this study. The results of the analysis are presented and discussed in the Results section, followed by conclusions in the Conclusions section.

MULTIMODEL ENSEMBLE PROJECTIONS

Because of the expansion of climate modelling efforts resulting in an abundance of distinct climate models, there is a need to evaluate how these models perform relative to one another. Multiple model intercomparison projects have been organized to meet this need. On a global scale, the Coupled Model Intercomparison Project (CMIP) and Atmospheric Model Intercomparison Project are the most notable collaborations undertaken with this goal in mind (Meehl *et al*., 2005 Meehl *et al*., 2007; Kreienkamp *et al*., 2011;). Beginning in the mid-1990s, the World Climate Research Programme (WCRP) committee, now known as the WCRP/Climate Variability and Predictability Working Group on Coupled Models, set about to organize one of the first generations of intercomparison projects (Meehl *et al*., 2007). Their efforts have since resulted in multiple CMIP generations, recently culminating in an open-access data set, the WCRP CMIP3 multimodel data set which represents 'an unprecedented, comprehensive coordinated set of global couple climate model experiments' (Meehl *et al*., 2007).

Several regional programs have been conducted in the last decade focused on addressing the need for appropriate scale level assessment of climate change impacts. In Europe, the Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects project described by Christensen and Christensen (2007) followed by the Ensembles-Based Predictions of Climate Changes and Their Impacts project (Van der Linden and Mitchell, 2009) provided an array of regional data sets for investigating future climate variation. The Statistical and Regional Dynamical Downscaling of Extremes for European regions project focused on the frequency and intensity of twenty-first century extreme events over Europe (http://www.cru.uea.ac.uk/projects/stardex/).

In North America, the North American Regional Climate Change Assessment Program (NARCCAP) provides data from multiple GCM–RCM-coupled simulations over most the continent (Mearns *et al*., 2009). The RCM data used in this study was provided by NARCCAP efforts. NARCCAP's goal is the production of climate simulations at a resolution that allows for regional-scale investigation of future climate variation. The products are intended to be useful in generating and studying impact scenarios across much of North America. The program consists of multiple RCMs driven by multiple AOGCMs. Simulations of both future (2041–2070) and historic (1971–2004) periods were produced by the NARCCAP modellers at a spatial resolution of 50 km and subdaily temporal resolution. Future scenarios were forced for the 21st century using the Special Report on Emissions Scenarios (SRES) A2 emissions scenario.

STUDY AREA

Oregon's Willamette River basin (WRB) (see Figure 1) covers a drainage area of 29,728 km^2 (11,478 sq mi), roughly 12% of the entire state, and intersects or contains 13 of the 36 counties in the state (Hulse *et al*., 2002). It is home for more than two thirds of Oregon's population and serves urban, agricultural, wildlife and recreational land use interests (Hulse *et al*., 2002 Chang and Jung, 2010;). The Willamette River, 13th largest in the continental USA, captures more runoff than its higher ranked counterparts, per

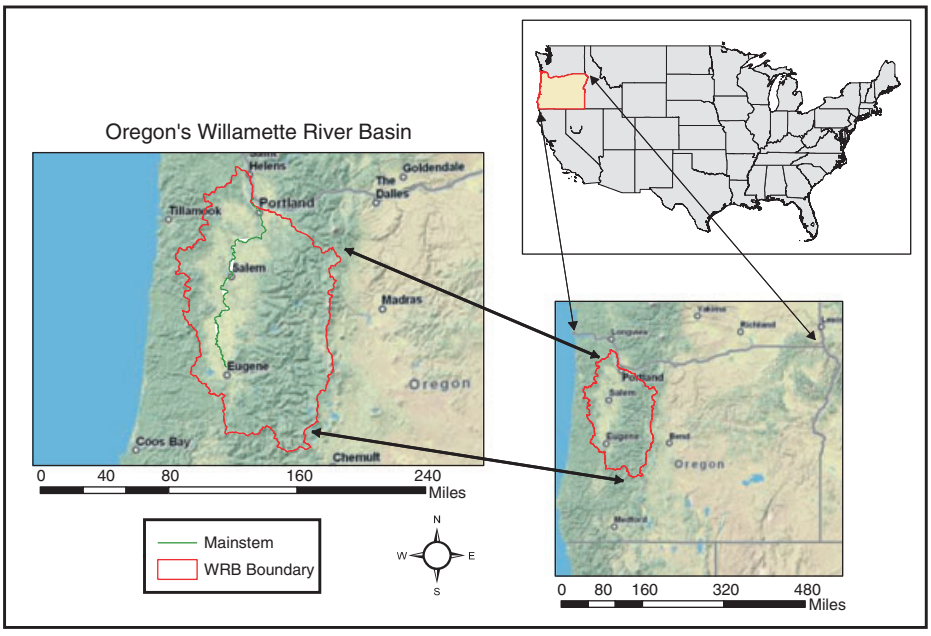

Figure 1. Study area: Willamette River basin, Oregon, USA

Copyright © 2012 John Wiley & Sons, Ltd.

unit of land area (Hulse *et al*., 2002; Chang and Jung, 2010; Jung *et al*., 2011a).

The temperate marine climate of the basin translates into cool wet winters, with 80% of annual precipitation occurring between October and May, and warmer mostly dry summers (Lee and Risley, 2002). The average annual temperatures in the region depend primarily on elevation and range from 40 °F to 65 °F. The annual mean precipitation also varies with elevation, from approximately 40 inches at the lowest elevations up to 175 inches at the highest elevations. Precipitation in the form of snow at the higher elevations within the basin is an influential component of the overall water cycle. Recent studies estimate that as much as 75% of precipitation falls as snow at or higher than 6500 ft (Chang and Jung, 2010). Higher than 4000 ft, 35% of precipitation falls as snow (Lee and Risley, 2002).

The WRB has been the subject of multiple recent climate change and precipitation-related studies (e.g. Chang and Jung, 2010; Moradkhani *et al*., 2010; Risley *et al*., 2011 Jung *et al*., 2011a; Najafi *et al*., 2011a,c;). Moradkhani *et al*. (2010) used a combination of LiDAR and GCM data to investigate the potential changes to floodplains and flood magnitude in the Tualatin River Basin, a subbasin of the WRB. Najafi *et al*. (2011a) used a multimodel approach to investigate the influence of a variety of statistically downscaled GCM outputs and hydrologic models also over the Tualatin River Basin. Najafi *et al*. (2011c) investigated the statistically downscaled GCM data sets over the upper WRB with a focus on estimating precipitation based on optimal predictor variables selected using an independent component analysis developed by Moradkhani and Meier (2010). Chang and Jung (2010) investigated spatial and temporal changes in runoff over the WRB using multiple GCMs, emission scenarios and a hydrologic model. In contrast to the previously mentioned studies, the current study investigates the entire WRB using dynamically downscaled climate model data sets with a focus on the ability of the RCMs to accurately simulate historic precipitation events.

Given the projected population growth and influence of precipitation on the WRB, understanding the effect of future climate variability on the region is of crucial importance for all stakeholders in the region. Exploratory analysis of climate models yields one approach for addressing this issue.

DATA

The NARCCAP project provides dynamically downscaled GCM outputs at a spatial resolution averaging 50 km. Data from six RCM simulations were available at the time of this study; these RCMs are listed in Table I. The RCMs were driven by boundary condition data sets provided by the AOGCMs. Four distinct AOGCMs were selected by NARCCAP to provide boundary conditions required as inputs to the RCMs. These AOGCMs are listed in Table II along with the group name, aliases and other distinct information regarding the model differences. The NARCCAP Web site catalogued the individual ensemble member data sets used to drive the RCMs for all AOGCMs, historic and future simulations, except for the Hadley Centre Coupled Model, version 3 (HADCM3) model.

Precipitation rate data [$kgm^{-2}\,s^{-1}$], at a temporal resolution of 3 h, was obtained over both a historical period (1979–2004 for the NCEP reanalysis driven data and 1976–2000 for the GCM driven data) as well as a future period (2038–2069). The spatial location of each RCM's grid points within the WRB is displayed in Figure 2. The number and location of grid points within the WRB varies between RCMs, owing to inherent design differences of each model. Although the amount of RCM grid points within the study region is rather sparse, it still represents an improvement upon the spatial resolution of GCMs.

To provide a comparison with observed precipitation over the WRB, the University of Washington (UW) gridded data set described by Maurer *et al*. (2002) was used. This data set covers the period 1950–2000 and provides surface level information regarding numerous climatic variables at three hourly time intervals. Specifically for this study, the UW data set provides values of total daily precipitation over the

Table I. RCM information

Model	Aliases	Modelling group	Full name	References
CRCM	MRCC	OURANOS/UQAM	Canadian Regional Climate Model	http://www.ouranos.ca/fr/programmation-scientifique/science-du-climat/simulations-climatiques/MRCC/eng/crcm.html#crcm42
ECP2	ECPC, RSM	UC San Diego/Scripps	Experimental Climate Prediction Center Regional Spectral Model	http://www.emc.ncep.noaa.gov/mmb/RSM/
HRM3	PRECIS,HadRM3	Hadley Centre	Hadley Regional Model 3	http://www.metoffice.gov.uk/precis/
MM5I	MM5, MM5P	Iowa State University	MM5-PSU/National Center for Atmospheric Research mesoscale model	http://www.mmm.ucar.edu/mm5/
RCM3	RegCM3	UC Santa Cruz	Regional Climate Model version 3	http://users.ictp.it/~pubregcm/RegCM3/
WRFG	WRFP, WRF	Pacific Northwest National Lab	Weather Research and Forecasting Model	http://www.wrf-model.org/index.php

Copyright © 2012 John Wiley & Sons, Ltd.

Table II. General circulation model (drivers for RCMs) information

Model	Full name	Modelling group	Ensemble member used	More information (references)
CCSM	Community Climate System Model	National Center for Atmospheric Research	b30.030e (ctl), b30.042e (fut)	http://www.cesm.ucar.edu/
CGCM3	Third Generation Coupled Global Climate Model	Canadian Centre for Climate Modeling and Analysis	CGCM #4	http://www.ec.gc.ca/ccmac-cccma/default.asp?lang=En&n=1299529F-1
GFDL	Geophysical Fluid Dynamics Laboratory GCM	GFDL/National Oceanic and Atmospheric Administration	20C3M, run2; sresa2, run1	
HADCM3	Hadley Centre Coupled Model, version 3	Hadley Centre for Climate Prediction and Research, Met Office, UK	Custom NARCCAP run	Gordon *et al*. (2000), Pope *et al*. (2000)

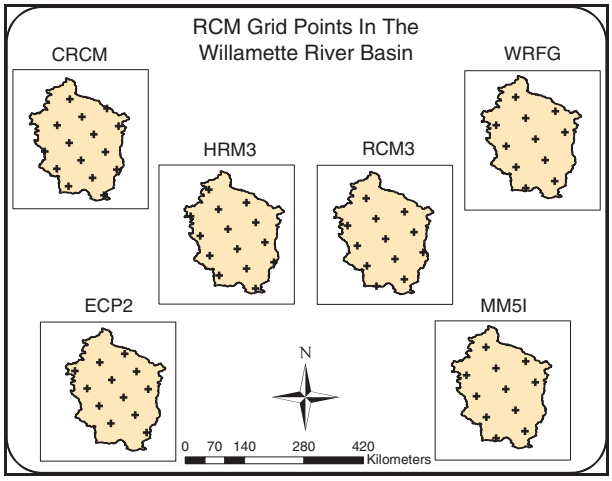

Figure 2. Location of RCM grid points within the study area

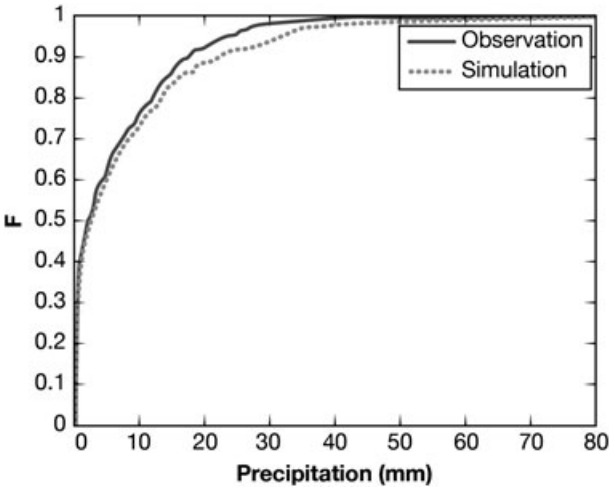

Figure 3. Example scenario depicting observed and simulated data CDFs over the historic period 1980–1998

continental USA obtained from the stations of the National Oceanic and Atmospheric Administration's Cooperative Observer (Maurer *et al*., 2002). The precipitation data over the WRB used in this study was obtained at 1/8th degree resolution and served as an observational benchmark upon which the dynamically downscaled NARCCAP data sets were compared.

METHODOLOGY

Bias correction

For this study, the quantile mapping approach was implemented on the NARCCAP data sets. In this procedure, the observed and simulated data sets are each characterized in terms of their full distribution of daily values for each month, a so-called *nonparametric approach* (Johnson and Sharma, 2011) because it does not rely on adjusting the mean, standard deviation or other standard statistical parameters. As in most bias correction approaches, a scaling factor is developed between the observed and the simulated data sets over a historic period. For both the observed and the simulated data sets, the cumulative distribution functions (CDFs) are computed on a monthly basis. Figure 3 shows the CDFs for an example case. After computing the CDFs, the scaling factor determined based on the respective quantile values during the observed period are then applied for the projected (future) period. The results, which will be discussed in the Results section, demonstrated that this approach is suitable for the given data set and suggest that such a bias correction step would be crucial when using the given data for impact analysis.

Performance of the GCMs versus RCMs

To evaluate the performance of RCM simulations, various performance metrics have been proposed in recent studies. In this study, the performance metrics focus on the ability of the model simulations to accurately capture the occurrence of precipitation events over the study area. This study used a newly developed performance metric focusing on the ability of RCMs to accurately simulate precipitation events. The added value index (AVI) is defined as 'the area beyond critical useful skill where the regional model skill is greater than that of the coarse-resolution model' (Kanamitsu and DeHaan, 2011). In fact AVI was designed to identify the ability of an RCM to improve on the simulated characteristics compared with its host GCM simulation.

Copyright © 2012 John Wiley & Sons, Ltd.

Computing the AVI begins by calculating the skill of the simulations compared with an observed data set. For this study, the skill metric computed is based on correlation values between observations and simulations at each grid point. We used Spearman's correlation coefficient (ρ). The observed (x) and simulated values (y), from the UW data set and each RCM–GCM pair, respectively, were ranked (i), and the correlation coefficient was computed for each month as follows:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (1)$$

The seasonal correlation values were considered as the combination of the monthly correlation values for that season at each grid point and for each model. This was performed for both the RCM–GCM as well as the original GCM forcing data set. The seasonal correlation values at all grid points were then fit to a normal distribution through the transformation:

The seasonal correlation values were then averaged at each grid point and for each model. This was performed for both the RCM–GCM as well as the original GCM forcing data set. The correlation values at each grid point were then fit to a normal distribution by the transformation:

$$P^* = \frac{P}{(1 - \mathrm{ABS}(P)^n)} \quad (2)$$

where P is the original averaged correlation value at each grid point, P^* is the transformed value and n is set to a value of 8 as suggested by (Kanamitsu and DeHaan, 2011). From these normalized correlation values, a probability density function (PDF) was established. In some simulations, the two PDFs would cross one another, the skill values at these cross points (XPs) provides useful information regarding the data sets. To aid understanding those PDF characteristics used in describing this metric, Figure 4 was generated containing two PDFs, and they illustrate the location of model means, XP and the region considered in calculating the AVI.

Rather than focusing on the entire skill distribution, it is useful to select a threshold skill above which the distributions will be investigated. This is important because investigating model simulation should focus on skill values that represent accurate portrayal of observed characteristics. For this study, the threshold skill value of 0.3 was selected. When computing the value of the AVI for each model combination, only the area between the PDFs above this threshold will be considered and computed. Furthermore, to investigate and portray situations where the RCM outperforms the GCM at critical high skill values, the skill values where the distributions cross one another are calculated. When the distributions cross one another above the previously mentioned threshold value, that location is termed the XP skill. Results of the AVI analysis on the NARCCAP simulation data sets is presented in the Results section.

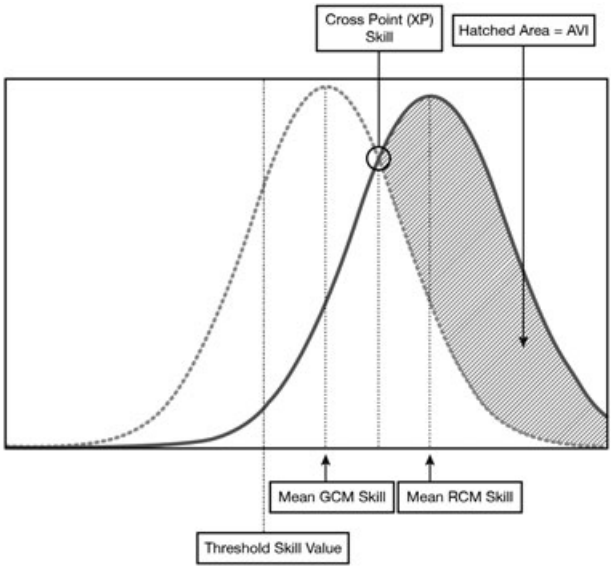

Figure 4. Calculation of the AVI. Two idealized skill distribution functions are shown, the GCM skill in dash line and the RCM skill in solid line. The XP is defined as the skill where the two distributions intersect. The Threshold Skill Value is defined as a user selected value below which the skill is not considered. The mean skill of each model (GCM or RCM) is defined as the skill value at peak of the distribution curve. The AVI is then calculated as the area between the distributions above the XP skill value, demonstrated by the hatched area in the figure

Extreme value analysis

Given a sufficiently large number of maximum observations in finite sized blocks, such as annual maximum, EVT leads to the generalized extreme value (GEV) distribution (Fisher and Tippett, 1928):

$$F(D_{ijt} \leq d) = \exp\left[-\left(1 + \xi_{ij}\frac{D - \mu_{ij}}{\sigma_{ij}}\right)^{-1/\xi_{ij}}\right] \quad (3)$$

where μ is termed the location parameter, σ is the scale parameter and ξ is the shape parameter (Kharin and Zwiers, 2000 Katz *et al.*, 2002;). i, j and t denote the latitude, longitude of each grid and the time, respectively. D represents independently and identically distributed (iid) data $D_1, D_2, \ldots, D_n$.

The GEV distribution commonly considers the maximum extreme in annual blocks (Cooley, 2009; Mínguez *et al.*, 2010; Towler *et al.*, 2010); therefore, it may provide an insufficient data series especially for short periods. It also disregards extreme data less than the annual maximum in each year (Cooley *et al.*, 2007). Instead, in the peaks-over-threshold approach, one considers all the extreme values, which exceed a large enough threshold (Naveau *et al.*, 2005 Cooley *et al.*, 2007; Acero *et al.*, 2010;).

The generalized Pareto distribution (GPD), from the EVT, models the probability distribution of exceedances over the threshold u (Coles, 2001). Similar to GEV, this

Copyright © 2012 John Wiley & Sons, Ltd.

method is based on asymptotic justification (i.e. sufficiently large data set). In this case, the tail of the distribution is represented by

$$F\left(D_{ijt} > d + u \middle| D_{ijt} > u\right) = \left(1 + \xi_{ij}\frac{d}{\sigma_{u_{ij}}}\right)_{+}^{-1/\xi_{ij}} \quad (4)$$

Assuming $a = \left(1 + \xi_{ij}\frac{d}{\sigma_{u_{ij}}}\right)$, then $a_+ = a$ if $a \geq 0$ and $a_+ = 0$ if $a < 0$. $\sigma_u(>0)$ and ξ are the scale and shape parameters, respectively. The scale parameter provides information about the variability of the exceedances, and ξ gives information about the tail of the distribution as discussed previously. The threshold u is chosen so that the previously mentioned data follow a GPD.

By differentiating the CDF of Equation 4, the following probability distribution is obtained:

$$p\left(D_{ijt}\right) = \frac{1}{\sigma_{u_{ij}}}\left(1 + \frac{\xi_{ij}d}{\sigma_{u_{ij}}}\right)^{-1/\xi_{ij}-1} \quad (5)$$

To remove the conditional form of the GPD, the following equation can be used instead based on the theory of total probability (Coles, 2001):

$$F\left(D_{ijt} > d + u\right) = \zeta_{u_{ij}}\left(1 + \xi_{ij}\frac{d}{\sigma_{u_{ij}}}\right)^{-1/\xi_{ij}} \quad \zeta_{u_{ij}} = F\left(D_{ijt} > u\right) \quad (6)$$

The EVT approach provides estimates of various return levels. The r-year return level (d_r) is obtained from the GPD parameters:

$$d_{r_{ij}} = u + \frac{\sigma_{u_{ij}}}{\xi_{ij}}\left[\left(rn_{y_{ij}}\zeta_{u_{ij}}\right)^{\xi_{ij}} - 1\right] \quad (7)$$

where n_y is the number of observation in a year.

Optimal threshold selection is an important factor in GPD approach. A high value of u results in very few exceedance points and high variance in the parameter estimates; however, a very low u violates the asymptotic assumption of the EVT approach. In this study, data were fit to the GPD distribution several times with different thresholds. The estimated GPD parameters were then evaluated to find a stable range of parameter values (Thompson *et al.*, 2009). Considering various RCMs and grid points the 95% quantile of the data provided reliable parameter estimates. Therefore, $u = Q_{95}$ was selected as the GPD threshold.

The independency assumption is another factor that may invalidate the EVT approach because the occurrence of precipitation extreme in one day may influence the probability for the one in the next day. According to 'runs declustering', scheme exceedances belong to the same cluster if they are separated by a fixed number of observations r (Acero *et al.*, 2010). In this study, $r = 1$ was chosen, meaning that if the distance between two extreme events was over a day then those extremes were considered as belonging to different clusters.

The final independent, declustered extreme NARCCAP precipitation data were fit to the GPD distribution for each grid cell. The maximum likelihood approach was then used to estimate the distribution parameters.

RESULTS

The results from the performance metric analysis are displayed in Figure 5 and in Table III. Figure 5 displays the PDFs of each RCM–GCM pair in terms of their correlation skill values. It is evident from Figure 5 that not all of the RCMs outperform their host GCMs in terms of correlation skill. In fact, the Community Climate System Model (CCSM) GCM outperforms all three of the RCMs (CRCM, MM5I and WRFG); it is coupled with in terms of both mean/average skill as well as the overall distribution of skill. Figure 5, row 1 column 2 (CRCM-CCSM), exhibits the scenario where the average skill of the GCM is greater than the skill of the RCM and the GCM outperforms the RCM even at high skill levels. In this case, the RCM can be viewed as having no added information compared with the host GCM. In Figure 5, however, the Experimental Climate Prediction-2(ECP2)–Geophysical Fluid Dynamics Laboratory (GFDL) combination reveals the opposite and most desirable situation. In this case, the average skill of the RCM is greater than that of the GCM, and the high correlation simulations are better represented in the RCM. RCM3–Third Generation Coupled Global Climate Model (CGCM3) shows the situation where both the RCM and GCM appear to perform with equal skill. Table III yields the relevant skill values as well as the computed AVI for all simulations except for the Hadley Regional Model 3 (HRM3)–HADCM3 combination [data for the host GCM (HADCM3) was not available at daily resolution so the metrics were not computed]. The mean skill of both RCM and GCM individually reveals that the overall skill of these simulations compared with observed data was low, with the highest mean skill being approximately 0.05. As mentioned earlier, in some cases, the distributions crossed one another over the threshold skill level of 0.3. The value of the AVI is given in the last column of Table III. Positive values indicate that the RCM improved upon the host GCM, and negative values indicate that the GCM simulation had higher correlation with observed data. AVI values with an 'x' indicate that the RCM outperformed the GCM above the XP skill value, meaning that the RCM demonstrated higher skill at high correlations.

Results from the EV analysis are displayed in Figures 6–8. The return level analysis in Figures 6 and 7 reveal noticeable differences between both RCM simulations as well as GCM boundary condition influence. Figure 6 displays the spatial distribution of the 2-year return level magnitudes over the WRB as simulated by each of the RCM–GCM data sets modelled via representative GPD distributions. The 2-year return level magnitudes vary substantially between the RCMs, with

Copyright © 2012 John Wiley & Sons, Ltd.

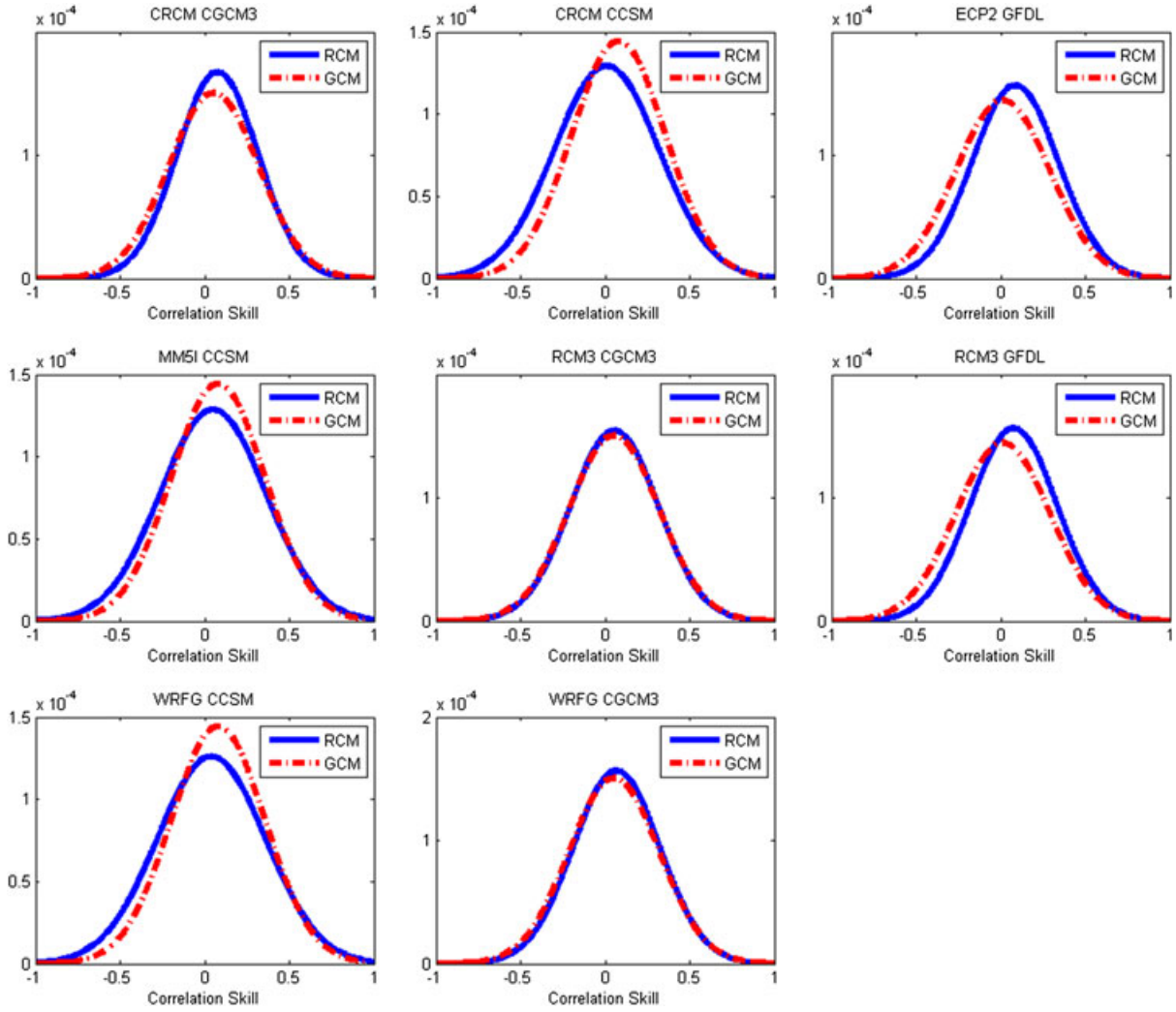

Figure 5. PDFs for all RCM and GCM combinations based on correlation between simulated and observed historic (1980–1998) precipitation data sets. The correlation between the simulation and the observation is calculated for each month of a year, and then the average value is computed over the period of analysis for each month. The PDFs are then generated for each season

values ranging from less than 75 mm/day to more than 135 mm/day. The upper limit of the 25-year return level magnitudes extends to values in excess of 150 mm/day. Some very clear differences between both RCMs and GCMs are evident in these figures. It is notable that only the HRM3–HADCM3 simulation data set contains magnitudes more than 110 mm/day for the 2-year return level. Most future simulations do not reveal a substantial increase in the magnitude of the 2-year return level. The 25-year return levels also do not change dramatically between the historic and the future periods. As with the 2-year return level estimates, the HRM3–HADCM3 simulations exhibit the highest magnitudes; however, other model combinations, such as WRFG–CGCM3&CCSM and RCM3–GFDL, also exhibit high magnitudes. The spatial distribution of these magnitudes is also informative. The topography of the WRB, high elevations on the Eastern and Western edges and low valley floor in between, should influence the distribution of precipitation, theoretically with more precipitation falling at higher elevations. However, in most of the model simulations, the topographical influence is not discernible.

Table III. Mean skill, XP skill, area difference between PDFs and AVI computed from dynamically downscaled RCM data sets over the WRB, historic period (1980–1998)

Model (RCM–GCM)	RCM mean skill	GCM mean skill	XP skill	AVI
WRFG–CGCM3	0.0351	0.0173	0.4740	−0.0032
WRFG–CCSM	0.0245	0.0353	0.5232	0.0140x
RCM3–GFDL	−0.0013	0.0155	0.8916	−6.949e.05
RCM3–CGCM3	0.0314	0.0173	0.2746	−0.0088
MM5I–CCSM	0.0305	0.0353	0.5148	0.0123x
ECP2–GFDL	0.0050	0.0155	No XP	0.0505
CRCM–CGCM3	0.0448	0.0173	0.4455	−0.0088
CRCM–CCSM	0.0163	0.0353	No XP	−0.0394

The shape parameter of the estimated GPD is displayed in Figure 8 across the WRB. In terms of the simulated precipitation events over the WRB, a positive shape parameter indicates the presence of a heavy upper tail (higher likelihood of extreme magnitude events) and a negative shape parameter indicates a bounded distribution (an identifiable upper limit to those extreme events), and when the shape parameter is equal to zero, the

Copyright © 2012 John Wiley & Sons, Ltd.

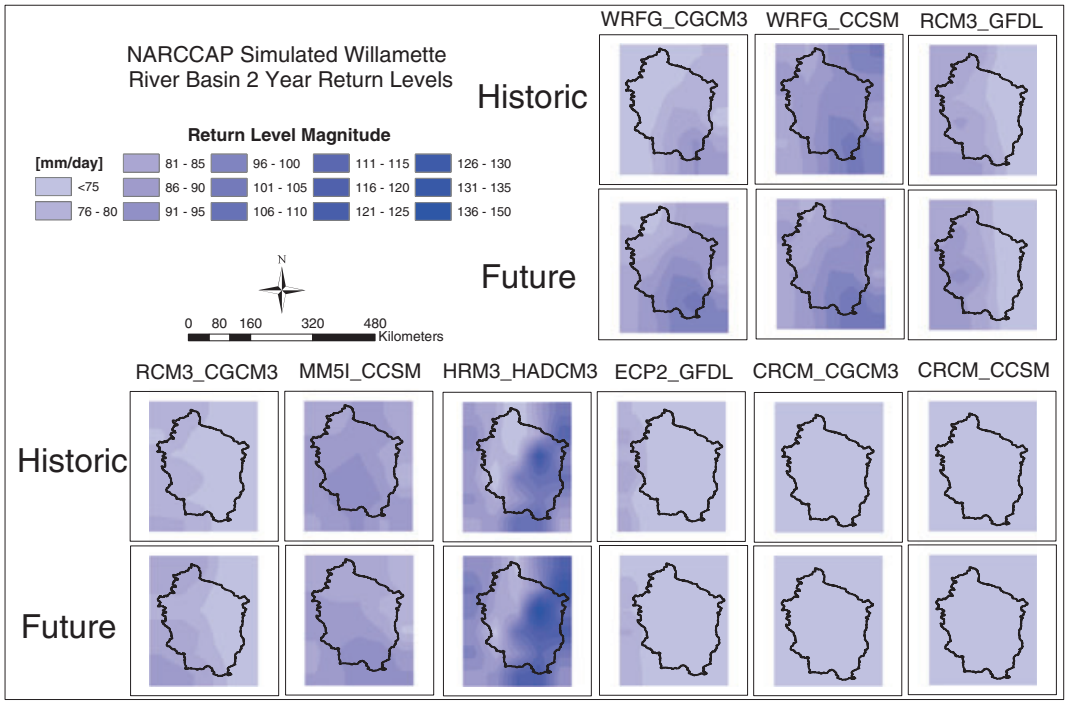

Figure 6. Simulated historic (1980–1998) and future (2040–2069) 2-year return levels for all RCMs

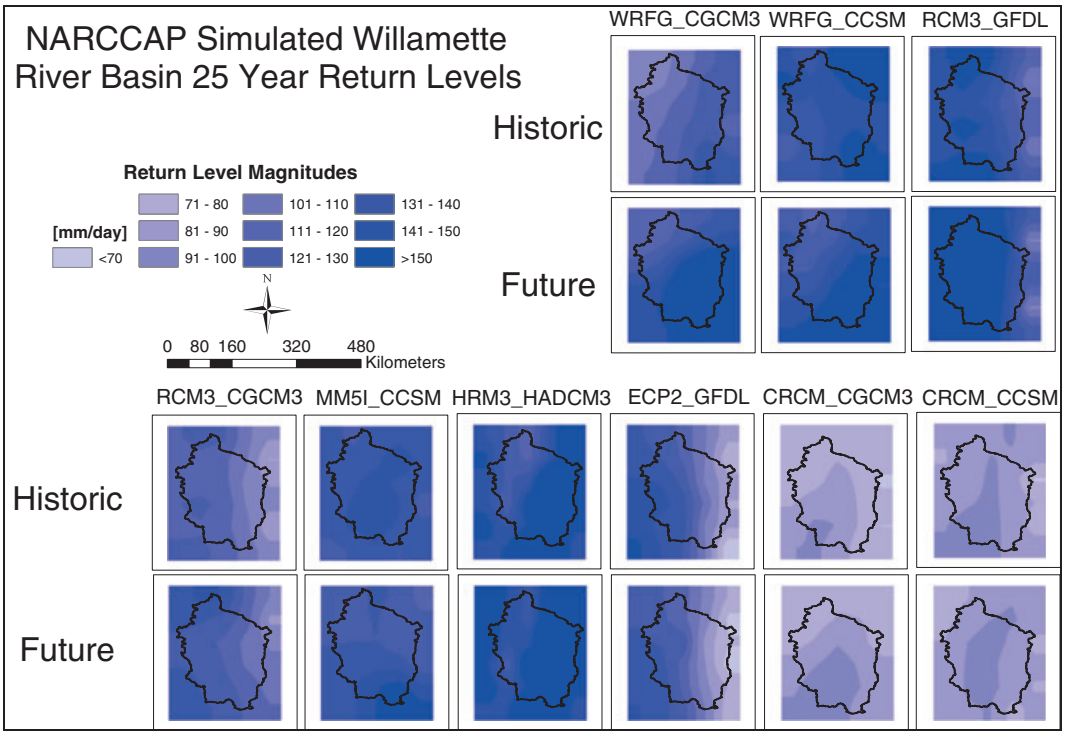

Figure 7. Simulated historic (1980–1998) and future (2040–2069) 25-year return level magnitudes for all RCMs

distribution is unbounded but has a thin upper tail. However, when the distribution is wider, or unbounded, the uncertainty in the distribution increases. Areas in green represent regions where the shape parameter is negative, yellow areas represent shape parameter values that are positive but nearly zero and red and pink areas are indicative of areas where the estimated distribution's shape parameter is higher than 0.1. The shape parameter results also reveal differences due to RCM characteristics as well as the influence of GCM driving conditions on the behaviour of the RCMs. From Figure 8, it is evident that the GFDL (a GCM) simulations are represented by unbounded heavy upper tails regardless of RCM and period.

Copyright © 2012 John Wiley & Sons, Ltd.

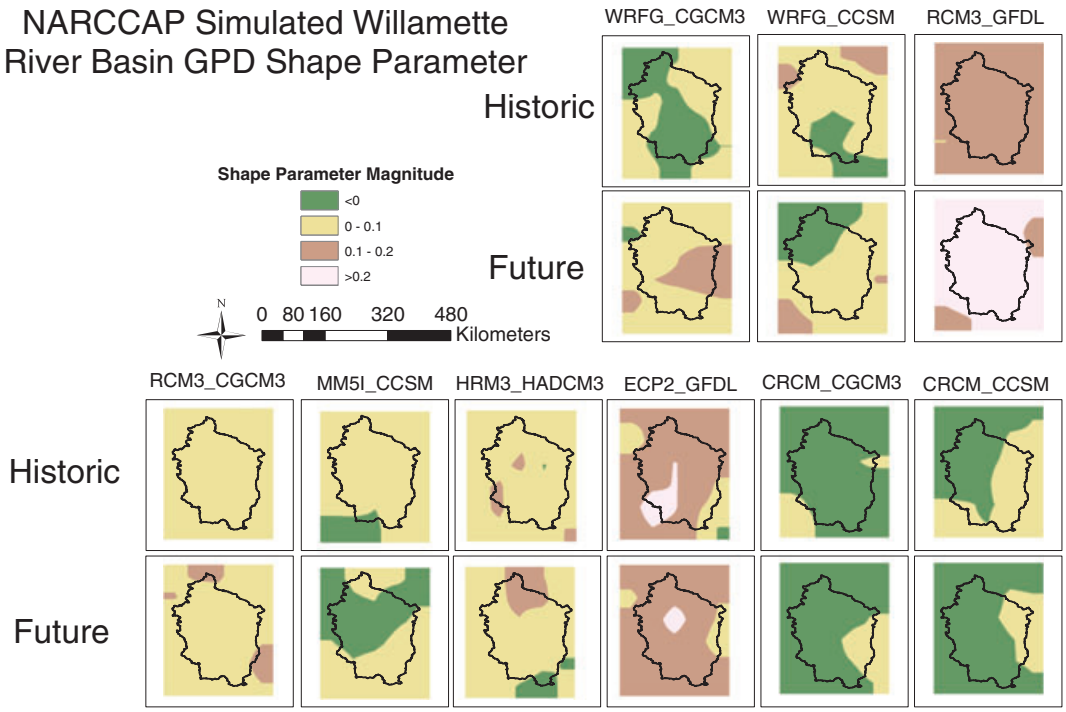

Figure 8. Representative GPD distribution shape parameter (ξ) for historic (1980–1998) and future (2040–2069) RCM simulations

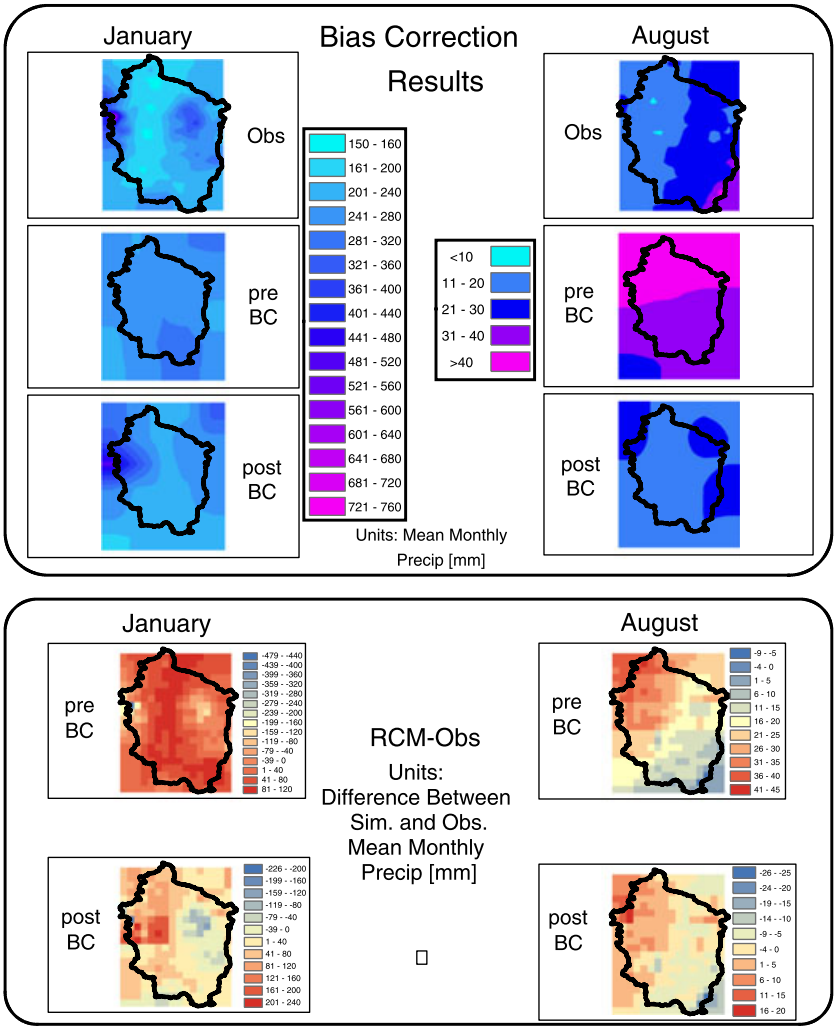

Figure 9. Results from bias correction procedure. Taking the values from the WRFG-CCSM simulation and the UW observation data sets for August and January, then calculating mean monthly precipitation values (mm) from 1980 to 1998

Copyright © 2012 John Wiley & Sons, Ltd.

Results of the quantile mapping bias correction technique are displayed in Figure 9, for the WRFG–CCSM simulation. Mean monthly precipitation values for January and August over the historical period, 1980–1998, are displayed over the WRB both before and after the bias correction procedure was applied. These two months were selected because they represent both heavy precipitation, January, and relatively low precipitation, August, months. The observed gridded data set used to bias correct the simulated data is also displayed for the selected months in Figure 9, row 1. Rows 2 and 3 of Figure 9 display the simulated data before and after, respectively, the application of the bias correction procedure. Rows 4 and 5 of Figure 9 show the difference (simulated − observed) between the data sets before and after bias correction. The bias is stronger in the month of January because of the higher magnitude of precipitation during the month, whereas August, a relatively dry month over the WRB, reveals a smaller magnitude in the bias. During the month of January, the simulation tends to over predict precipitation in the WRB. This positive bias is an attribute that has been documented before in dynamically downscaled data sets (Caldwell, 2010). The bias present in the simulation data set during the month of August is also slightly positive, demonstrated by the range of bias values present before bias correction. The bias correction procedure effectively corrects for this positive bias and is able to reduce the overall magnitude of bias as well.

CONCLUSIONS

In this study, precipitation data from the multiple RCMs and GCMs were investigated over the WRB. Multiple performance metrics, a bias correction scheme and extreme value analysis were applied to the data. Results of the various performance metrics and information from the extreme value analysis were described, including the AVI, 2- and 25-year return level magnitudes and the shape parameter of a representative GPD distribution. The results of this study demonstrate two key facts regarding the use of dynamically downscaled climate data sets. First, applying a bias correction scheme to any downscaled data set is a needed and important step yielding more accurate results. For the NARCCAP data sets, the quantile mapping procedure was implemented and successfully reduced the difference between observed and simulated precipitation over the WRB by correcting for the positive bias that is present in RCM data sets and reducing overall bias magnitude. Second, implementing fundamentals of EV theory to climate data sets provides estimates of changes to variable values, such as precipitation, because of climate change. Using the GPD distribution, this study obtained estimates of changes to 2- and 25-year extreme precipitation event magnitudes over the WRB. The results indicate that these return level magnitudes will increase in the future period 2038–2069 compared with simulations over the historical period 1980–1998. In addition, the shape parameter of the GPD distribution derived from the NARCCAP data sets indicates that the RCM models in the NARCCAP study provide different depictions of future changes over the WRB.

ACKNOWLEDGEMENTS

Partial financial support for this project was provided by the National Science Foundation, Water Sustainability and Climate (WSC) program (grant no. EAR-1038925).

The authors thank the NARCCAP for providing the data used in this article. NARCCAP is funded by the National Science Foundation, the US Department of Energy, the National Oceanic and Atmospheric Administration and the US Environmental Protection Agency Office of Research and Development.

REFERENCES

Acero FJ, García JA, Gallego MC. 2010. Peaks-over-Threshold Study of Trends in Extreme Rainfall over the Iberian Peninsula. *Journal of Climate* **24**: 1089–1105.

Brabson B, Palutikof J. 2010. Tests of the generalized Pareto distribution for predicting extreme wind speeds.

Caires S, Sterl A. 2010. 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data.

Caldwell P. 2010. California wintertime precipitation bias in regional and global climate models. *Journal of Applied Meteorology and Climatology* **49**: 2147–2158. DOI: http://dx.doi.org/10.1175/2010JAMC2388.1.

Chang H, Jung IW. 2010. Spatial and temporal changes in runoff caused by climate change in a complex large river basin in Oregon. *Journal of Hydrology* **388**: 186–207.

Christensen JH, Christensen OB. 2007. A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change* **81**: 7–30.

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag: London, UK.

Cooley D. 2009. Extreme value analysis and the study of climate change. *Climatic Change* **97**: 77–83.

Cooley D, Nychka D, Naveau P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**: 824–840.

Di Luca A, de Elía R, Laprise R. 2011. Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations. *Climate Dynamics*: 1–19. DOI 10.1007/s00382-011-1068-3.

Diffenbaugh NS, Pal JS, Trapp RJ, Giorgi F. 2005. Finescale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15,774–15,778. DOI: 10.1073/pnas.0506042102.

Fisher RA, Tippett LHC. 1928. *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*. Cambridge Univ Press; 180–190.

Fowler H, 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**: 1547–1578.

Frei C, Schöll R, Fukutome S, Schmidli J, Vidale PL. 2006. Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *Journal of Geophysical Research* **111**: D06105.

Giorgi F. 1990. Simulation of regional climate using a limited area model nested in a general circulation model. *Journal of Climate* **3**: 941–964.

Gordon C, Cooper C, Senior CA, Banks HT, Gregory JM, Johns TC, Mitchell JFB, Wood RA. 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**: 147–168.

Hulse D, Gregory S, Baker JP, Consortium PNER. 2002. *Willamette River Basin Planning Atlas: Trajectories of Environmental and Ecological Change*. Oregon State University Press Corvallis: Oregon, USA.

Hundecha Y, St-Hilaire A, Ouarda T, El Adlouni S, Gachon P. 2008. A nonstationary extreme value analysis for the assessment of changes in

Copyright © 2012 John Wiley & Sons, Ltd.

extreme annual wind speed over the Gulf of St. Lawrence, Canada. *Journal of Applied Meteorology and Climatology* **47**: 2745–2759.

Hurkmans R, Terink W, Uijlenhoet R, Torfs P, Jacob D, Troch PA. 2010. Changes in streamflow dynamics in the Rhine basin under three high-resolution regional climate scenarios. *Journal of Climate* **23**: 679–699.

Johnson F, Sharma A. 2011. Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Water Resources Research* **47**: W04508.

Jung I, Chang H, Moradkhani H. 2011a. Quantifying uncertainty in urban flooding analysis considering hydro-climatic projection and urban development effects. *Hydrology and Earth System Sciences* **15**: 617–633.

Kanamitsu M, DeHaan L. 2011. The Added Value Index: A new metric to quantify the added value of regional models. *Journal of Geophysical Research* **116**: D11106.

Katz RW. 2010. Statistics of extremes in climate change. *Climatic Change* **100**: 71–76.

Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. *Advances in water resources* **25**: 1287–1304.

Kendon EJ, Jones RG, Kjellström E, Murphy JM. 2010. Using and designing GCM–RCM ensemble regional climate projections. *Bulletin of the American Meteorological Society*.

Kharin VV, Zwiers FW. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *Journal of Climate* **13**: 3760–3788.

Kharin VV, Zwiers FW, Zhang X. 2010a. Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations.

Kharin VV, Zwiers FW, Zhang X, Hegerl GC. 2010b. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations.

Kreienkamp F, Baumgart S, Spekat A, Enke W. 2011. Climate Signals on the Regional Scale Derived with a Statistical Method: Relevance of the Driving Model's Resolution. *Atmosphere* **2**: 129–145.

Lee KK, Risley JC. 2002. *Estimates of Ground-Water Recharge, Base Flow, and Stream Reach Gains and Losses in the Willamette River Basin*. Citeseer: Oregon.

Lima CHR, Lall U. 2010. Spatial scaling in a changing climate: A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *Journal of Hydrology* **383**: 307–318.

Maurer E, Wood A, Adam J, Lettenmaier D, Nijssen B. 2002. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States*. *Journal of Climate* **15**: 3237–3251.

McGregor J. 1997. Regional climate modelling. *Meteorology and Atmospheric Physics* **63**: 105–117.

Mearns L. 2007. The North American Regional Climate Change Assessment Program dataset. *National Center for Atmospheric Research Earth System Grid Data Portal, Boulder, CO*.

Mearns L, Gutowski W, Jones R, Leung R, McGinnis S, Nunes A, Qian Y. 2009. A regional climate change assessment program for North America. *Eos Trans. AGU* **90**: 311.

Mearns L, Biner S, Caya D, Laprise R, Nunes A, Jones R, Moufouma-Okia W, Tucker S, Gutowski W, Arritt R, Flory D, Takle G, Zoellick C, Macintosh C, Snyder M, Sloan L, O'Brien T, Leung R, Correia J, Qian Y, Duffy P, Teng H, Strand G, Held I, Wyman B, McGinnis S, McDaniel L, Thompson J, Anitha A. 2011. The North American Regional Climate Change Assessment Program dataset, National Center for Atmospheric Research Earth System Grid data portal. *National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO. Data downloaded 2011-10-17*.

Meehl GA, Covey C, McAvaney B, Latif M, Stouffer RJ. 2005. Overview of the coupled model intercomparison project. *Bulletin of the American Meteorological Society* **86**: 89–93.

Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE. 2007. The WCRP CMIP3 multimodel dataset. *Bulletin of the American Meteorological Society* **88**: 1383–1394.

Mínguez R, Menéndez M, Méndez F, Losada I. 2010. Sensitivity analysis of time-dependent generalized extreme value models for ocean climate variables. *Advances in water resources* **33**: 833–845.

Moradkhani H, Meier M. 2010. Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis. *Journal of Hydrologic Engineering* **15**(10): 744–762. DOI: 10.1061/ASCEHE.1943-5584.0000246.

Moradkhani H, Baird RG, Wherry SA. 2010. Assessment of climate change impact on floodplain and hydrologic ecotones. *Journal of Hydrology* **395**: 264–278. DOI:10.1016/j.jhydrol.2010.10.038.

Mote PW, Salathe EP. 2010. Future climate in the Pacific Northwest. *Climatic Change* **102**: 29–50.

Murphy J. 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. *Journal of Climate* **12**: 2256–2284.

Najafi M, Moradkhani H, Jung I. 2011a. Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrological Processes* **25**: 2814–2826.

Najafi M, Moradkhani H, Piechota TC. 2011b. Ensemble Streamflow Prediction: Climate Signal Weighting vs. Climate Forecast System Reanalysis. *Journal of Hydrology* **442-443**: 105–116.

Najafi MR, Moradkhani H, Wherry SA. 2011c. Statistical Downscaling of Precipitation using Machine Learning with Optimal Predictor Selection. *Journal of Hydrologic Engineering* **16**: 650–644. DOI:10.1061/(ASCE)HE.1943-5584.0000355.

Naveau P, Nogaj M, Ammann C, Yiou P, Cooley D, Jomelli V. 2005. Statistical methods for the analysis of climate extremes. *Comptes Rendus Geosciences* **337**: 1013–1022.

Pope V, Gallani ML, Rowntree PR, Stratton RA. 2000. The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics* **16**: 123–146.

Risley J, Moradkhani H, Hay L, Markstrom S. 2011. Statistical Comparisons of Watershed Scale Response to Climate Change in Selected Basins across the United States. *Earth Interactions* **15**: 617–633. DOI: 10.1175/2010EI364.1.

Rusticucci M, Tencer B. 2008. Observed changes in return values of annual temperature extremes over Argentina. *Journal of Climate* **21**: 5455–5467.

Salathe Jr EP, Mote PW, Wiley MW. 2007. Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest. *International Journal of Climatology* **27**: 1611–1621.

Shrestha RR, Dibike YB, Prowse TD. 2011. Modelling of climate-induced hydrologic changes in Lake Winnipeg Watershed. *Journal of Great Lakes Research*. DOI:10.1016/j.jglr.2011.02.004.

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL. (Ed.) 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge.

Thompson P, Cai Y, Reeve D, Stander J. 2009. Automated threshold selection methods for extreme wave analysis. *Coastal Engineering* **56**: 1013–1021.

Towler E, Rajagopalan B, Gilleland E, Summers RS, Yates D, Katz RW. 2010. Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research* **46**: W11504.

Tryhorn L, DeGaetano A. 2011. A comparison of techniques for downscaling extreme precipitation over the Northeastern United States. *International Journal of Climatology* **31**: 1975–1989.

Van der Linden P, Mitchell J. 2009. ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. *Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3 PB, UK* **160**.

Villarini G, Smith JA, Serinaldi F, Ntelekos AA, Schwarz U. 2011. Analyses of extreme flooding in Austria over the period 1951–2006. *International Journal of Climatology*. DOI: 10.1002/joc.2331.

Wang YQ, Leung LR, McGregor JL, Lee DK, Wang WC, Ding YH, Kimura F. 2004. Regional climate modeling: progress challenges and prospects. *J. Meteor. Soc. Japan* **82**: 1599–1628.

Wehner MF, Smith RL, Bala G, Duffy P. 2010. The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dynamics* **34**: 241–247.

Wood AW, Leung LR, Sridhar V, Lettenmaier D. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**: 189–216.

Xue Y, Zeng F, Mitchell K, Janjic Z, Rogers E. 2001. The impact of land surface processes on simulations of the U.S. Hydrological cycle: a case study of the 1993 flood using the SSiB land surface model in the NCEP eta regional model. *Monthly Weather Review* **129**: 2833–2860.

Yuan X, Liang X-Z. 2011. Improving cold season precipitation prediction by the nested CWRF-CFS system. *Geophysical Research Letters* **38**: L02706. DOI: 10.1029/2010GL046104.

Copyright © 2012 John Wiley & Sons, Ltd.