Weather and Climate Extremes 12 (2016) 15–23

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Statistical downscaling of regional climate model output to achieve projections of precipitation extremes

Eric M. Laflamme [a,*], Ernst Linder [b], Yibin Pan [b]

[a] Department of Mathematics Plymouth State University, MSC 29, 17 High Street Plymouth, NH 03264, United States
[b] Department of Mathematics and Statistics University of New Hampshire, W383 Kingsbury Hall, 33 Academic Way Durham, NH 03824, United States

ARTICLE INFO

Article history:
Received 2 July 2015
Received in revised form
30 October 2015
Accepted 21 December 2015
Available online 22 December 2015

Keywords:
Statistical downscaling
Generalized Pareto distribution
Regional climate models
Bootstrapping
Bayesian analysis
Uncertainty quantification

ABSTRACT

In this work we perform a statistical downscaling by applying a CDF transformation function to local-level daily precipitation extremes (from NCDC station data) and corresponding NARCCAP regional climate model (RCM) output to derive local-scale projections. These high-resolution projections are essential in assessing the impacts of projected climate change. The downscaling method is performed on 58 locations throughout New England, and from the projected distribution of extreme precipitation local-level 25-year return levels are calculated. To obtain uncertainty estimates for return levels, three procedures are employed: a parametric bootstrapping with mean corrected confidence intervals, a non-parametric bootstrapping with BCa (bias corrected and acceleration) intervals, and a Bayesian model. In all cases, results are presented via distributions of differences in return levels between predicted and historical periods. Results from the three procedures show very few New England locations with significant increases in 25-year return levels from the historical to projected periods. This may indicate that projected trends in New England precipitation tend to be statistically less significant than suggested by many studies. For all three procedures, downscaled results are highly dependent on RCM and GCM model choice.

1. Introduction

There is great societal interest in assessing the impacts of projected climate change, and more specifically, there is an intense interest in the impact of change in variability and extreme events that could accompany global climate change predictions (Tebaldi et al., 2006). Increases in these extremes have already been observed as precipitation events, heat waves, and drought are occurring with greater intensity and frequency over the past few decades (U.S. Climate Change Science Program (USCCSP), 2008). Other analyses have provided additional evidence that precipitation extremes are becoming more extreme and will continue to do so in the future (e.g. Zwiers and Kharin, 1998; Groisman et al., 1999; Meehl et al., 2000; Tank and Konnen, 2003; Karl and Knight, 1998; Kharin and Zwiers, 2005). In their survey of recent projections of climate extremes provided by global circulation models (GCMs), Tebaldi et al. (2006) concluded that models agree with observations over the historical period and that there is a trend towards a world characterized by intensified precipitation, with a greater frequency of heavy-precipitation and extreme events.

Precipitation extremes are a primary concern as these events are typically more impactful than precipitation events alone and are responsible for a disproportionately large part of climate-related damages (Kunkel et al., 1999; Easterling et al., 2000; Meehl et al., 2000). Natural systems may also be affected by changes in precipitation extremes, as these events have been shown to cause shifts in ecosystem distributions, to trigger extinctions, and to alter species morphology and behavior (Parmesan et al., 2000). Furthermore, extreme rainfall often translates into extreme flooding and consequently great material and economic losses, erosion and damage to crops, collapse of lifeline infrastructure, the breakdown of public health services (Douglas and Fairbank, 2011), fatalities (Kunkel et al., 1999), and structural damage to dams, bridges, and coastal roads.

In this work, we outline a procedure to examine potential change in precipitation extremes in New England. In the following section (Section 2), we discuss methods of statistical/probabilistic downscaling as well as some elements of extreme value theory germane to our analysis. In Section 3, we describe our data: precipitation observations for locations throughout New England, historical climate model output, and projected series for future precipitation. Section 4 outlines our downscaling approach and discusses all details of our methods. Section 5 presents our

* Corresponding author.
E-mail addresses: emlaflamme@plymouth.edu (E.M. Laflamme), elinder@unh.edu (E. Linder), ybj3@unh.edu (Y. Pan).

http://dx.doi.org/10.1016/j.wace.2015.12.001

downscaling results. Lastly, in Section 6, a discussion and conclusion is presented.

2. Downscaling

Atmosphere-ocean general circulation models, or AOGCMs, are coupled atmosphere and ocean models that simulation weather at a global scale. AOGCMs are the main component of global climate models (GCMs) which are the primary tools used to quantify and assess climate change impacts (Wilby and Harris, 2006). However, because global weather simulation is so computationally expensive, these models provide predictions at an extremely coarse scale (250 KM by 250 KM, in most cases). The issue is that environmental impact models are sensitive to local climate characteristics, and the drivers of local climate variation are not captured at the coarse scales of GCMs (Maurer and Hidalgo, 2008). That is, GCMs do not provide an accurate description of local climate. To overcome this discrepancy, methods of 'downscaling' are applied to produce local-scale climate predictions based on corresponding GCM scenarios.

Downscaling appears in two forms: Dynamical and statistical downscaling (or empirical statistical downscaling, ESD). Dynamical downscaling is a computationally-intensive technique which makes use of the lateral boundary conditions combined with regional-scale forcings such as land-sea contrast, vegetation cover, etc., to produce regional climate models (RCMs) from a GCM. RCM outputs are typically produced over regular geographic grids with scales in the tens of kilometers.

Statistical downscaling (SD), on the other hand, is a computationally less demanding alternative that may be applied to achieve a variety of results. Essentially, statistical downscaling is a two-step process consisting of 1) the development of statistical relationships between local climate variables and large-scale predictors, and 2) the application of such relationships to the output of large-scale output to simulate local climate characteristics in the future (Hoar and Nychka, 2008). Statistical downscaling is a realistic approach to develop a specific, local-level climate prediction. Typically, SD methods are applied to GCM projections, but may also be applied to RCM output as these results may not be representative for the local climate (Skaugen et al., 2002; Engen-Skaugen, 2004). Furthermore, RCM output may simply have inadequate spatial resolution for some impact studies, and hence additional statistical downscaling must be applied to the dynamical model results (Benestad et al., 2007).

2.1. Probabilistic downscaling

This analysis focuses on a method of 'probabilistic downscaling' to project a single variable, extreme precipitation, into the future. While traditional ESD models the link between large- and local-scale variables, probabilistic downscaling is a type of statistical downscaling that models the relationship between large- and local-scale statistical entities. In this case, the statistical entities are the corresponding cumulative distribution functions (CDFs) of the large- and local-scale precipitation extremes. In this way, probabilistic downscaling techniques do not retain the chronology, or exact ordering, of events. However, accurate descriptions of future climate distributions are themselves sufficient predictions as we do not aim to predict weather, but rather the distribution of a weather variable (precipitation extremes).

When working exclusively with cumulative distribution functions, the simplest form of downscaling is what is referred to as 'quantile mapping' or 'quantile matching'. This non-parametric technique downscales a large-scale value x by selecting a local-scale value y based on the following:

$$F_Y(y)=F_X(x) \text{ with } y=F_Y^{-1}\left(F_X(x)\right)$$

where F is a CDF of a climate random variable. Once a mapping has been defined, it is then applied to large-scale dataset to create a local-scale prediction. The method does not take into account the information of the distribution of the future modeled dataset (Michelangeli et al., 2009). Furthermore, the method of quantile mapping cannot provide local-scale quantiles outside the range of the historical observations (Michelangeli et al., 2009). Proposed by Wood et al. (2004), the technique was applied to downscale monthly precipitation and temperature output from a GCM, and became known as bias-correction and spatial downscaling (BCSD).

To overcome the shortcomings of the quantile matching methodology, Michelangeli et al. (2009) proposed an extension to this mapping called the CDF-t. The CDF-t is similar to quantile mapping as it compares local- and large-scale distributions, but it accounts for changes in the large-scale CDF between historical and future periods. Let X denote a variable from climate model output and let X_C denote the series of the variable over the current, or calibration, period. Then, X_P denotes the variable projected into the future, the time series from runs of the climate model in the future. Similarly, let Y_C and Y_P denote the current and future series for the local-level station. We note that while Y_C is observed, Y_P will need to be predicted or downscaled. Finally, a transformation, $T(\bullet)$, is assumed to exist between the large- and local-scale variable such that $T(\bullet):[0,1]\rightarrow[0,1]$. We then have the relationship:

$$F_{Y_P}(x)=T\left(F_{X_P}(x)\right)=F_{Y_C}\left(F_{X_C}^{-1}\left(F_{X_P}(x)\right)\right) \quad (1)$$

where F_{Y_P} and $F_{X_P}$are the respective empirical CDFs for the local- and large-scale prediction, and F_{Y_C} and F_{X_C} are the respective CDFs of observed (historical) local-level data and observed large-scale, or regional data. For further details see Michelangeli et al. (2009). The improvement over quantile mapping is that the future, local-scale distribution is a function of both historical observations and large-scale information that may be distributed differently between calibration and projection periods.

However, for precipitation data, we are more concerned with the extreme events. In these cases, where the tails, which correspond to the extremes or high quantiles, are of primary interest, the non-parametric CDF-t is not ideal. Generally speaking, these rare, extreme values result in empirical CDFs for precipitation that are heavy-tailed. With few data at the extreme ends of the distribution, non-parametric quantile estimates in these tails have large variance and they may be strongly influenced by a single extreme event. Also, observations of historical changes, as well as future projections, confirm that changes in the distributional tails of precipitation (extremes) may not occur in proportion to changes in the mean and may not be symmetric in nature (Kharin and Zwiers, 2005; Robeson, 2004; Tank and Konnen, 2003; Easterling et al., 2000).

In light of this, Kallache et al. (2011) proposed the XCDF-t technique to downscale the distribution of extremes exclusively. The technique is analogous to the CDF-t technique of Michelangeli et al. (2009) in that is makes use of the same transformation function form (see Eq. (1)) to link large- and local-scale distributions of climate variables. Unlike the CDF-t method, however, the XCDF-t links estimated parametric distributions of large- and local-scale extremes only. To do this, 'exceedances over a threshold' based on extreme value theory (EVT) are used to fit appropriate distributions to extremes based on limiting properties of max-stable processes (See, for example, Coles, 2001). The framework of EVT allows for more precise estimation of the extreme portions of distributions.

For the XCDF-t, F_{X_P}, F_{Y_C}, and F_{X_C} are cumulative Generalized Pareto distribution (GPD) for the extremes of the modeled

predicted, local-level observed (historical), and modeled historical series, respectively. Each of the GPD CDFs have the following form:

$$F_u(z)=1-\left(1+\frac{\xi(z-u)}{\sigma}\right)^{\frac{-1}{\xi}},$$

where, for $z > u$ are the exceedances above some sufficiently large threshold, u, and where ξ and σ are shape and scale model parameters, respectively (Pickands, 1975; Balkema and de Haan, 1974). The qualitative behavior of the GPD is dominated by ξ, the shape parameter. For $\xi > 0$, the distribution of excesses is unbounded and has the traditional 'Pareto', heavy tail; for $\xi < 0$, the distribution has a finite upper bound and resembles an inverted Weibull-type distribution; and $\xi = 0$ corresponds to an unbounded, exponential-type distribution.

Lastly, we note that the XCDF-t method assumes that the relationship between large- and local-scale extreme remains constant between calibration and prediction period. That is, the models are based on historical data, and there is no guarantee that the past statistical relationships between different data fields will hold in the future. This so-called 'stationarity' assumption is made with any ESD procedure.

3. Data

Local-level precipitation data, Y_C, were obtained from the U.S. National Climatic Data Center (NCDC), a climate data archiving and retrieval system operated by the U.S. National Oceanographic and Atmospheric Administration (NOAA) Satellite and Information Service (http://www.ncdc.noaa.gov/). Hourly precipitation accumulations reported in hundredths of inches were originally obtained for 69 meteorological stations from all six New England states covering a period from 1948 to 2010. Because long series were required to observe extreme events, few stations with shorter precipitation series were omitted from our analysis. In the end, only stations with continuous measurements between 1970 and 2000 were retained, a total of 58 locations of the original 69 New England monitoring stations.

Regional climate model (RCM) output is used for large-scale precipitation data, both current and projected, or X_C and X_P from Eq. (1), respectively. Four regional model outputs used in downscaling were acquired from the North American Climate Change Assessment Program (NARCCAP), an international program which acts as a custodian for regional climate simulations to be used in impact assessment and research (Mearns et al., 2012). Each of the RCM outputs is driven by larger-scale atmospheric and oceanic boundary conditions provided by global circulation models, or GCMs. Thus, RCMs are highly dependent on their GCM 'drivers.' A total of eight different RCM/GCM combinations were used in conjunction with the NCDC station-level data. Information regarding the specific RCMs used is listed below (Table 1). We also note that all GCMs have been forced with the SRES A2 greenhouse gas emissions scenario for the 21st century (See, Nakicenovoic et al. (2000) for Special Report on Emissions Scenarios (SRES) commissioned by the IPCC).

For each of the eight cases, current RCM data, three-hourly values of 42 climate variables are available for the period 1970–1999. Similarly, three-hourly RCM projected model outputs are available for the period 2040–2069. Large-scale precipitation output was converted from instantaneous flux in units of kg/m^{2}s to three-hourly accumulations in inches. For both model outputs, current and future (projected), and the station data, respective three-hourly and hourly accumulations were aggregated into daily totals of precipitation measured in inches.

4. Methodological details

4.1. Data matching

To downscale the 58 New England locations, each station must be matched with a corresponding region from the eight RCMs to establish the required X_C, X_P, and Y_C series. Although all RCMs are comprised of equally sized 50 km by 50 km grids, these layouts are not identical. In the majority of cases, stations are matched to model output based simply on gridded boundaries. In few coastal locations, however, where nearest grids were predominately ocean-related, stations were matched to the nearest terrestrial gridpoints. Ultimately, each of the precipitation data from the 58 stations was matched to eight corresponding historical and future RCM outputs.

4.2. Generalized Pareto Distribution and CDF transfer function

Following the method of Kallache et al. (2011), a GPD is first fitted to the precipitation exceedances associated with the three available series, X_C, X_P, and Y_C. This requires selecting a threshold for the three series, X_C, X_P, and Y_C, to distinguish extreme values. The method of Karl et al. (1996) was employed where a fixed, high percentile was chosen as a threshold. In our case, the 98.5th percentile was chosen as a threshold for all series, at all stations, which equates to 5 observations per year, on average, or about 200 observations for the entire collection period. Also, we coerce the

Table 1
RCM/GCM combinations used for large-scale model output.

Combination	RCM Name	Modeling group	GCM name
CRCM-CGCM3	Canadian Regional Climate Model (Caya and Laprise, 1999)	OURANOS / UQAM	3rd Generation Coupled Global Climate Model (McFarlane et al., 2005)
CRCM-CCSM	Canadian Regional Climate Model (Caya and Laprise, 1999)	OURANOS / UQAM	Community Climate System Model (Kiehl and Gent, 2004)
HRM3-GFDL	Hadley Regional Model 3 / Providing Regional Climates for Impact Studies (Jones et al., 2003)	Hadley Centre	Geophysical Fluid Dynamics Laboratory GCM (Delworth et al., 2006)
HRM3-HADCM3	Hadley Regional Model 3 / Providing Regional Climates for Impact Studies (Jones et al., 2003)	Hadley Centre	Hadley Centre Coupled Model, version 3 (Johns et al., 2003)
RCM3-GFDL	Regional Climate Model, version 3 (Giorgi et al., 1993a, b; Pal et al., 2007)	University of California, Santa Cruz	Geophysical Fluid Dynamics Laboratory GCM (Delworth et al., 2006)
RCM3-CGCM3	Regional Climate Model, version 3 (Giorgi et al., 1993a, b; Pal et al., 2007)	University of California, Santa Cruz	3rd Generation Coupled Global Climate Model (McFarlane et al., 2005)
WRFP-CGCM3	Weather Research and Forecasting Model (Skamarock et al., 2005)	Pacific Northwest National Laboratory	3rd Generation Coupled Global Climate Model (McFarlane et al., 2005)
WRFP-CCSM	Weather Research and Forecasting Model (Skamarock et al., 2005)	Pacific Northwest National Laboratory	Community Climate System Model (Kiehl and Gent, 2004)

fitted GPDs, F_{X_C}, F_{X_P}, and F_{Y_C} from Eq. (1), to have positive or identically zero shape parameters corresponding to heavy and exponential tails, respectively. In most cases, the shape parameters are positive, but few cases yield negative parameter estimates which were subsequently set to zero. This assumption is supported by the widely held belief that precipitation data, particularly maxima and extremes, consistently appear as heavy-tailed (e.g. Katz et al., 2002; Smith, 2001). Lastly, to produce a distribution of future precipitation exceedances at the local-level, F_{Y_P}, the transfer function (1) may be applied to the three fitted GPDs corresponding to the X_C, X_P, and Y_C series. This process is done for each station and model combination.

4.3. Return levels

A return level is a high quantile of an extreme value distribution that will be exceeded with some known probability. Return levels are universally understood and relate directly to location-specific climate impact assessments such as flooding, potential erosion, etc. As stated in Coles (2001), it is usually more convenient to interpret extreme value models in terms of quantiles or return levels, rather than individual parameter values.

We extend the work of Kallache et al. (2011) to include climate projections in the form of return levels. In their downscaling of extreme precipitation, Kallache et al. (2011) focus on transforming the entire distribution of large-scale predicted values to unknown, smaller-scale predicted values, an approach similar to Michelangeli et al. (2009). Instead of predicting an entire distribution of exceedances, we have chosen to predict a specified return level via the quantile matching of the XCDF-t procedure.

When estimating the q-quantile of Y_P, we can directly apply the transformation (1) by noting that, if $F_{X_P}^{-1}(q)=x_P$, and

$$\text{ReturnLevel}\,(q)=\hat{y}_P\,(q)=F_{Y_C}^{-1}\left(F_{X_C}\left(F_{X_P}^{-1}(q)\right)\right) \tag{2}$$

since the two quantiles are matched by having the same cumulative value. In our case, for the r-year return level, the value q is given by:

$$q=1-\left[\frac{1}{r*n_y*\hat{\lambda}_u}\right]$$

where r is the return level period, n_y is the number of observations per year, and $\hat{\lambda}_u$ is the empirical threshold exceedance rate that estimates $\lambda_u=Pr\,(X>u)$. In our case, $r=25$, n_y=365.25, and $\hat{\lambda}_u$=. 015. Point estimates for local-level, predicted 25-year return levels are produced from the application of the transfer function (2) to fitted GPD parameters estimated from the three series, X_C, X_P, and Y_C.

4.4. Climate change uncertainty quantification

4.4.1. Parametric bootstrapping and confidence intervals

We further extend the work of Kallache et al. (2011) to include a parametric bootstrapping procedure to calculate measures of uncertainty associated with return levels. Our parametric bootstrapping procedure was devised and implemented for all 58 locations (local-level stations) and all 8 RCM/GCM combinations. For this procedure, GPD model forms were first fit to X_C, X_P, and $Y_C$series. Then, from each of the three pairs of parameter estimates (a shape and scale for each series), random exceedances were generated. The result was a simulated series of extremes corresponding to each of the X_C, X_P, and $Y_C$series. For each of these simulated series, GPD parameters were refit via maximum likelihood. Next, for these parameter estimates and a given q value associated with the return level, both a predicted and historical local-level 25-year return level is identified via application of Eq. (2) and the corresponding quantile function, respectively. Differences between historical and predicted values were observed. This process was repeated 200 times (bootstrap iterations) to yield a distribution of differences in 25-year return level estimates. Lastly, from this distribution, 'mean corrected' bootstrap confidence intervals (See, for example, Davison and Hinkley, 1997) were calculated.

4.4.2. Nonparametric bias correction and acceleration (BCa) confidence intervals

Just as some skew (asymmetry) is present in the individual distributions of 25-year return levels, it is likely that the distribution of differences in return levels between predicted and current periods is also skewed. Furthermore, it is known that bootstrap procedures may not be consistent for extreme value problems as there is a tendency for samples to generate shorter tails than the true sample distribution (Coles and Simiu, 2003). In light of this, a non-parametric bootstrapping procedure with BCa (bias-corrected and accelerated) adjustment (Efron and Tibshirani, 1993) was employed. In this procedure, BCa interval endpoints are given by percentiles of the bootstrap distribution, but adjusted to account for the skew and bias of the data. The actual percentiles used depend on the acceleration and bias correction, which, generally speaking, measure the rate of change of the standard error and median bias of our estimator, respectively. In practice, these values are estimated by repeatedly sampling the data, and resulting BCa intervals are simply calculated.

For this procedure, the three series, X_C, X_P, and Y_C, are resampled with replacement. GPD models are fit to each sample and, from the fitted parameter estimates and a given q value associated with the return level, both a predicted and historical local-level 25-year return level is identified via application of Eq. (2) and the corresponding quantile function, respectively. The difference between the historical and predicted return level is then calculated. This resampling, fitting, and calculation process is repeated 500 times to produce a distribution of differences in return levels between the predicted and historical periods. From this distribution in differences, the adjusted BCa interval is calculated as given by Efron and Tibshirani (1993, see pg. 185). We note that, compared to the previous bootstrapping procedure, the procedure including this bias correction is considerably more computer intensive.

4.4.3. Bayesian models and credibility intervals

We further extend the analysis of Kallache et al. (2011) by pursuing a Bayesian model approach. Like the non-parametric bootstrap with BCa approach, this approach is pursued to address skewness and produce more accurate and precise intervals than the bootstrap procedure. As with our other approaches, precipitation exceedances are assumed to follow a GPD which is asymptotically justified by extreme value theory. The Bayesian procedure will result in distributions for these extreme value distribution parameters that will serve as a basis for downscaling and ultimately return level estimation. That said, posterior estimates are based on assumed prior distributions and results may be sensitive to such specification.

In the Bayesian approach, posterior distributions are obtained using Bayes theorem. For ease of calculation, however, Markov Chain Monte Carlo (MCMC) techniques (see, for example, Gelman et al., 2014) are used to simulate the posterior draws of the unknown quantities and obtain simulation-based estimates of posterior parameters of interest. MCMC is often performed under a Gibbs sampling framework where a sequence of samples from conditional individual distributions is generated to approximate a joint distribution.

R statistical software (R Core Team, 2013) was used in

conjunction with the open-source software OpenBUGs (see http://www.openbugs.net) through the 'R2WinBUGs' R package to produce posterior distributions for the GPD parameters, ξ and σ. For each location, given the data from the three known series, X_C, X_P, and Y_C, posterior distributions for the respective parameters ξ_{X_C}, σ_{X_C}, ξ_{X_P}, σ_{X_P}, ξ_{Y_C}, and σ_{Y_C} were produced. For each series, a beta prior distribution was assumed for shape parameters. That is,

$$\xi_{X_C} \sim \text{Beta}\left(\alpha_{\xi_{X_C}}, \beta_{\xi_{X_C}}\right),$$

$$\xi_{X_P} \sim \text{Beta}\left(\alpha_{\xi_{X_P}}, \beta_{\xi_{X_P}}\right),$$

$$\xi_{Y_C} \sim \text{Beta}\left(\alpha_{\xi_{Y_C}}, \beta_{\xi_{Y_C}}\right),$$

where the beta parameters, alpha and beta ($\alpha_{\cdot}$, $\beta_{\cdot}$), are set to be 1.5 and 2.5, respectively, for each series. These distributions are semi-informative as they restrict estimates to values between 0 and 1 (thus, fairly non-informative). As stated earlier, precipitation extremes typically follow a heavy-tailed or exponential-tailed distribution, so it is commonplace to assume a constant zero or slightly positive shape parameter and thus restrict our model to only such cases, or values deemed 'sensible.'

For scale parameters, prior distributions were assumed to be uniform (Gelman, 2006), or

$$\sigma_{X_C} \sim \text{Unif.}\left(a_{\sigma_{X_C}}, b_{\sigma_{X_C}}\right),$$

$$\sigma_{X_P} \sim \text{Unif.}\left(a_{\sigma_{X_P}}, b_{\sigma_{X_P}}\right),$$

$$\sigma_{Y_C} \sim \text{Unif.}\left(a_{\sigma_{Y_C}}, b_{\sigma_{Y_C}}\right).$$

For these parameters, the uniform distribution is defined from 0 to 50 (a and b, respectively). Thus, the prior specification is non-informative as parameter estimates from traditional model-fitting were typically found to be between 0.2 and 1.2 for daily rainfall in inches.

The Gibbs sampler was initialized with starting values, $\xi^{(0)} = 0.48$ and $\sigma^{(0)} = 3$ for each series, values based on previous model-fitting. For each station/model combination, individual samplers, each based on 20,000 iterations, were run for the three series to produce distinct sets of shape and scale posterior distributions. MCMC sampled values require a 'burn-in' period before convergence to an ergodic Markov chain (before the Markov chain stabilizes), so 5000 initial iterations were discarded and not used for inference. For ease of calculations, the 15000 remaining iterations were further thinned to 5000. Thus, for each station/model combination, 5000 sets (ξ_{X_C}, σ_{X_C}, ξ_{X_P}, σ_{X_P}, ξ_{Y_C}, and σ_{Y_C}) of parameter estimates are produced. Convergence of the individual chains was then checked by diagnostic commands contained in the R CODA package.

For each station/model combinations, and for each of the 5000 sets of GPD parameters (ξ and σ) fit to the exceedances of the X_C, X_P, and Y_C series, a GPD cumulative or quantile function is calculated as required for application of Eq. (2). For a certain q-quantile of Y_P, we apply Eq. (2) to produce a single, predicted 25-year return level estimate. Additionally, for each of these sets of parameter estimates, a 25-year return level is calculated for the Y_C series, the station-level historical data. The difference in return levels between predicted and historical series are calculated and, after considering all 5000 sets of parameter estimates, a distribution of differences in return levels between the predicted and historical periods is produced for each station/model combination. Finally, from this distribution of differences, a 90% credibility (confidence) interval is extracted using the 5th and 95th percentiles.

5. Results

5.1. Return level point estimates

From local-level, predicted return levels and return levels produced from the local-level historical data (based on the fitted GPD parameters for the Y_C data), we produce a bubble plot of the differences (Fig. 1, below) at each of the 58 stations throughout New England. This figure represents the projected effect of climate change on extreme precipitation at the local-level via downscaling. In Fig. 1, red and blue circles indicate positive and negative changes, respectively, and the size of the circles indicates the magnitude of such change. We note that in this figure and all subsequent figures, dot sizes are on a common scale so that plots are comparable.

We observe that results are sensitive to the choice of GCM driver (GCM denoted by second part of combination title in lower case). Among common RCMs, different GCM drivers yield considerably different results (HRM3_gfdl versus HRM3_hadc, for example). Also, differences in station-level return levels are highly dependent on our choice of RCM, and this choice yields considerably different and occasionally contradictory results (RCM3_cgcm versus WRFG_cgcm, for example). Finally, the station results generally exhibit fewer decreases (negative values) than increases. The largest positive values are consistently observed in northern New Hampshire at the Pinkham Notch location. This is realistic as this area has the highest elevation among all 58 stations, is known for extreme weather, and is likely sensitive to climate change projections.

5.2. Parametric bootstrap confidence intervals

Fig. 2 below illustrates the variability attributable to downscaling by observing the 90% bootstrapped lower bound (the 5th percentile) of the distribution of differences in 25-year return levels between predicted and historical periods. Positive (>0) lower bounds are marked in red and identify locations with significant increases in 25-year return levels from historical to future periods. Locations marked in blue indicate differences in return levels that do not show a significant increase.

When comparing results from the model combinations, we notice that the choice of GCM has an influence on the resulting differences in return levels. For example, plots for WRFG_ccsm and WRFG_cgcm (which share a common RCM) show conflicting results at several locations in southern New England and northern New Hampshire/Vermont. This suggests that our choice of GCM does contribute to the variability observed in downscaling results. Also, lower bounds for station-level differences in return levels show significant dependence on our choice of RCM. This is illustrated in the significant disagreement between results obtained from RCM3_cgcm, CRCM_cgcm, and WRFG_cgcm combinations, for example. We conclude that there is little agreement between the 8 RCM/GCM combinations in terms of return level projections, and both the regional model and GCM driver choice contribute to such variability.

Generally speaking, observing lower bounds for differences in return levels, we find significant increases at relatively few stations, mostly sporadic increases. There may be some weak evidence of significant increases throughout locations in southern New England (Connecticut, Massachusetts, and Rhode Island). While subtle, this trend is most clearly seen from the WRFG_cgcm

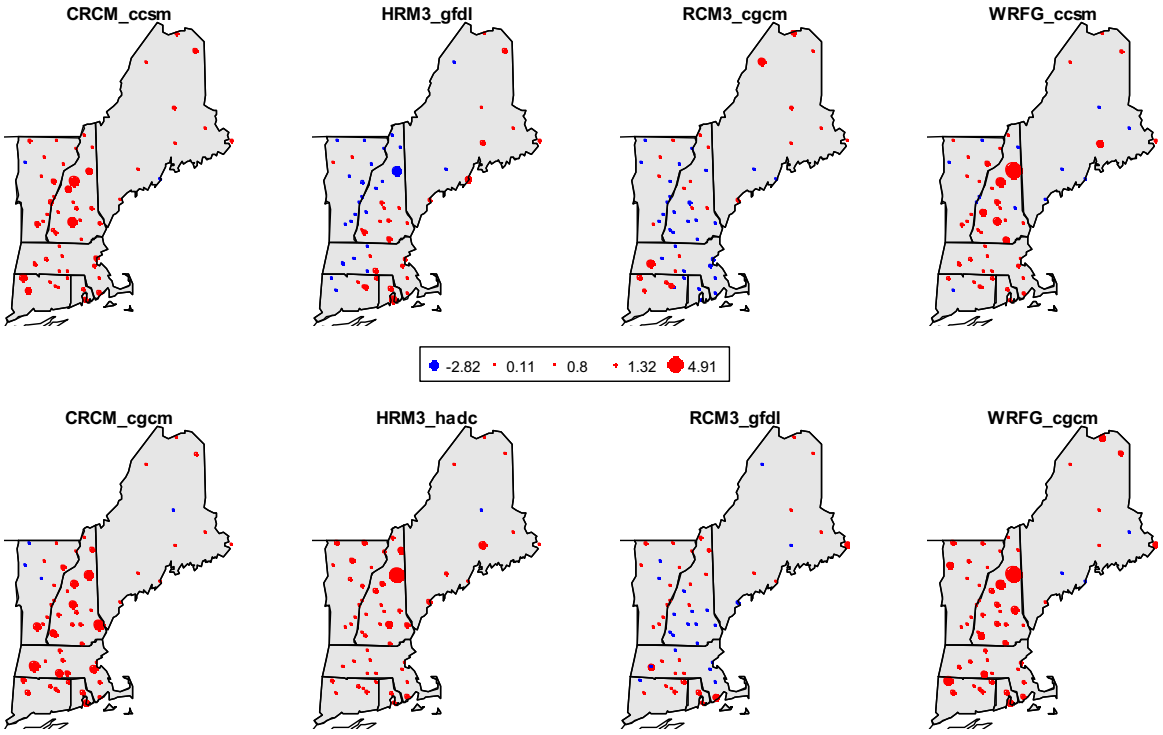

Fig. 1. Difference in 25-year return level estimates between predicted and historical periods (predicted minus historical) for the 8 climate model combinations.

and HRM3_gfdl model combinations. Overall, our results above (Fig. 2) are quite different from station results based on return level point estimates only, where numerous increases were observed (Fig. 1). These comparisons illustrate the variability attributable to the procedure, thus how downscaling adds uncertainty to our results.

5.3. Nonparametric bootstrap and BCa confidence intervals

Fig. 3 below compares the 90% lower bounds (5th percentile) of the BCa interval for the differences in 25-year return levels between predicted and historical periods. As before, positive lower bounds are marked in red and identify locations with significant increases in 25-year return levels from historical to future periods.

From the plot above, the majority of lower bounds are negative, but some significant increases are observed in northern New

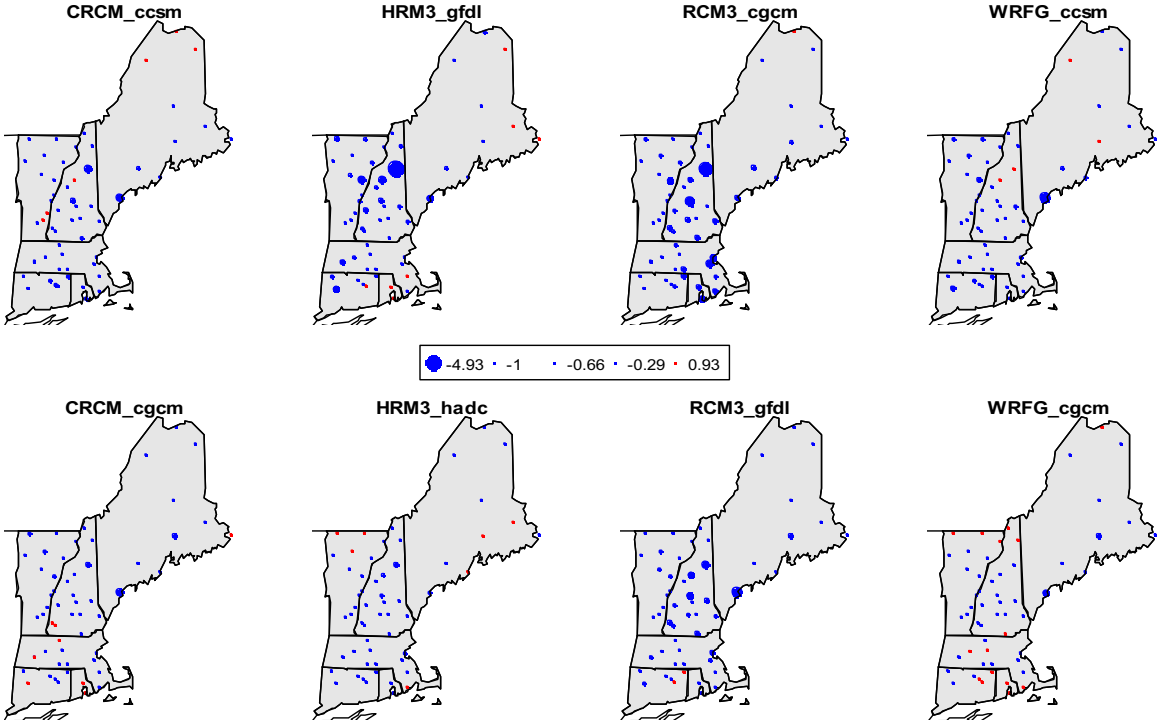

Fig. 2. 90% lower bound for distribution of differences in bootstrapped 25-year return levels between predicted and historical periods.

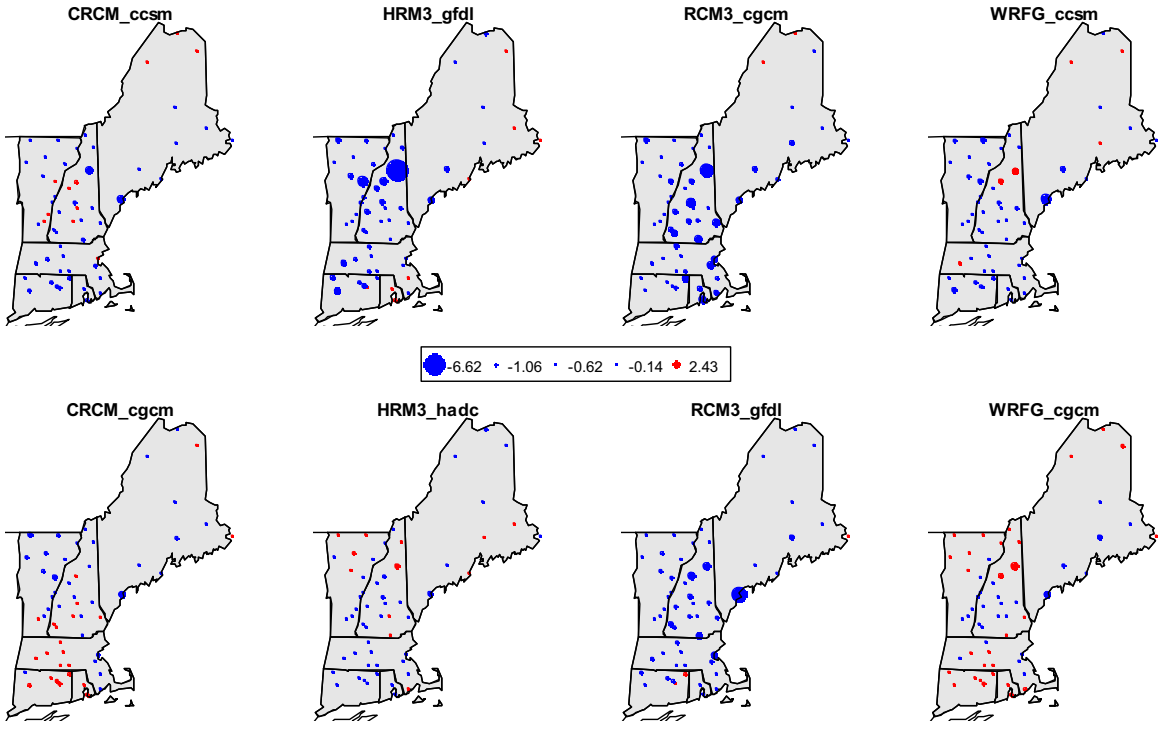

Fig. 3. 90% lower bound for distribution of differences in BCa 25-year return levels between predicted and historical periods.

Hampshire (WRFG models) and throughout southern New England (CRCM_cgcm and WRFG_cgcm combinations). Generally speaking, we find that lower bounds of the BCa intervals are comparable to those obtained without the adjustment (Fig. 2). That is, regardless of method (parametric bootstrap or non-parametric bootstrap with BCa), increases and decreases follow the same broad pattern across locations and model combinations. We note that the BCa adjustment, however, tends to inflate the positive differences. Also, in general, there are more positive differences associated with the BCa procedure. This is expected and consistent with our understanding of the BCa, how the procedure is designed to capture the right-skew of the distribution of differences, and how the procedure typically inflates lower confidence limits. Lastly, as before, there is little agreement between the 8 RCM/GCM combinations, and both the regional model and GCM driver choice contribute to such variability.

5.4. *Bayesian credibility intervals*

The intervals for differences in 25-year return levels produced under the Bayesian framework generally tend to resemble those produced via non-parametric bootstrapping with BCa adjustment. Below (Fig. 4) we compare our Bayesian results, the 90% credibility intervals for difference in return levels, to the estimates and confidence intervals obtained from both our parametric bootstrap and non-parametric/BCa procedures. For readability, this plot shows results for the six Connecticut locations and eight model combinations only.

Next, we compare the 90% lower bounds for differences in 25-year return levels between predicted and historical periods. These plots are directly comparable to those produced previously via parametric bootstrap and non-parametric/BCa methods. In Fig. 5 below, red and blue circles indicate positive and negative changes, respectively, and the size of the circle indicates the magnitude of change.

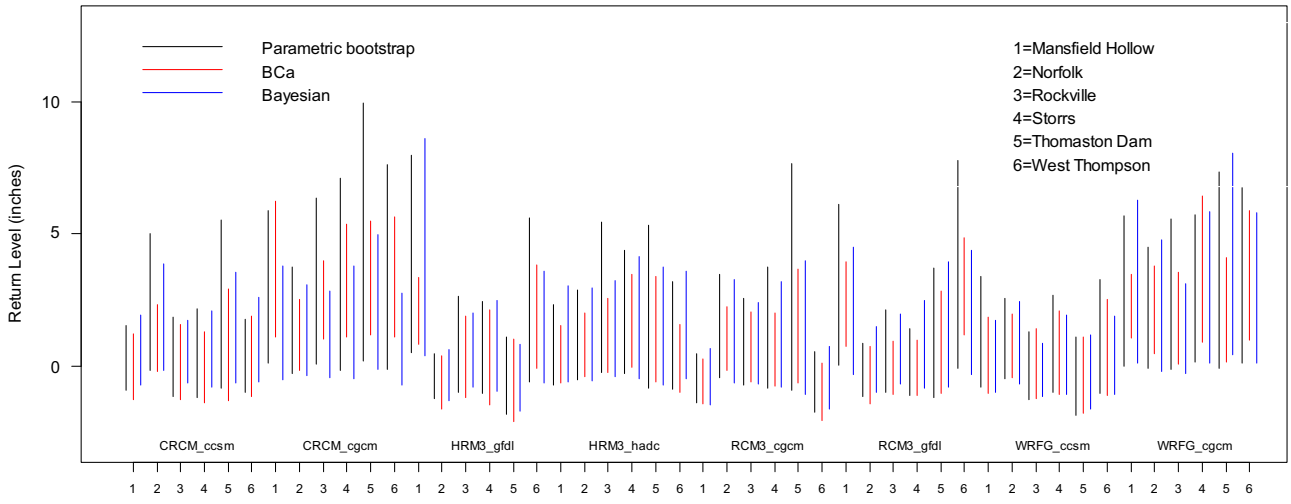

Fig. 4. Comparison of Bayesian intervals (blue) to parametric bootstrap (black) and BCa (red) intervals.

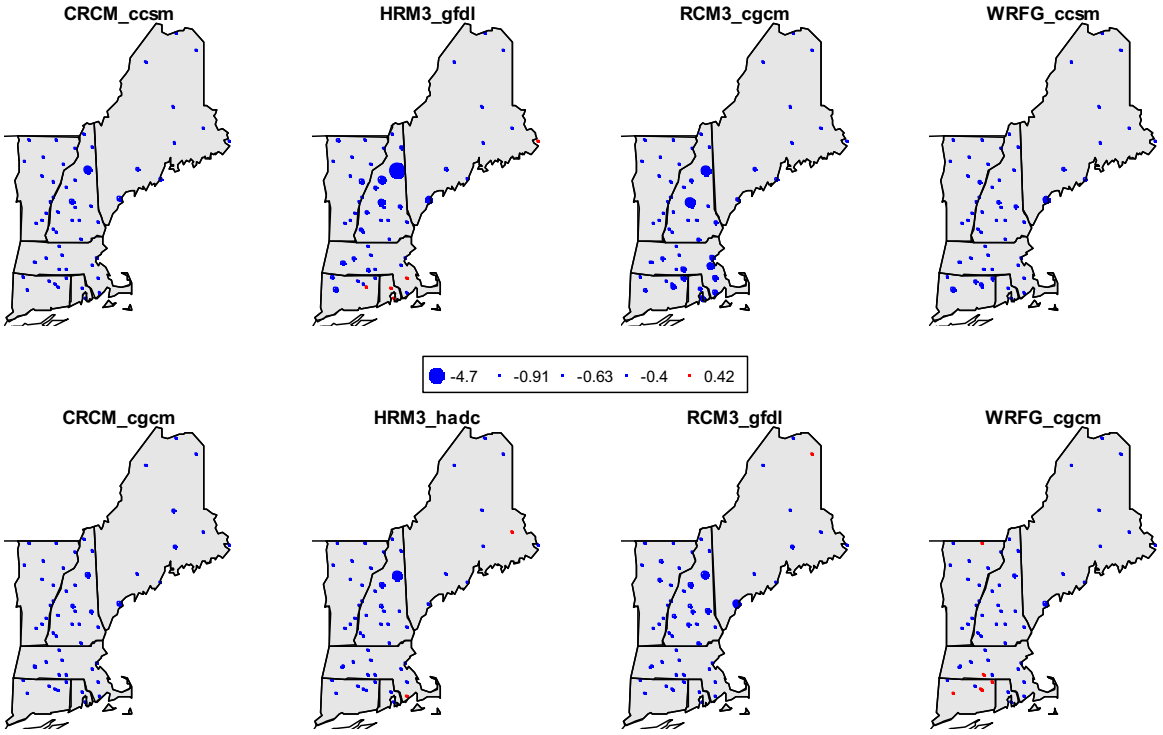

Fig. 5. 90% lower bound for distribution of differences in Bayesian 25-year return levels between predicted and historical periods.

From the plots above, the vast majority of locations have negative lower bounds (blue circles) which indicate differences in return levels that do not show a significant increase between historical and predicted periods. The few significant increases (red circles) observed are primarily found at locations in southeast New England (HRM3_gfdl, HRM3_hadc, and WRFG_cgcm) in Connecticut, Massachusetts, and Rhode Island. This may provide some weak evidence of a trend in increases in return levels, but, as seen from previous results (parametric and non-parametric bootstrap), these results are dependent on choice of both GCM and RCM.

6. Conclusion and discussion

Theoretically, this work is heavily influenced by the work of both Michelangeli et al. (2009) and Kallache et al. (2011) who respectively introduced the CDF-t and its parametric extension to extremes, the XCDF-t. While this work is made possible by these procedures, we extend the procedure in a variety of ways. First, we use the downscaling procedure to produce predicted, local-level return level estimates, widely interpretable estimates critical to climate change impact assessment. To produce measures of uncertainty associated with these return levels, a parametric bootstrapping procedure was employed. Next, we develop a non-parametric resampling scheme using a BCa (bias correction and acceleration) confidence interval to account for potential skewness/bias. Lastly, to further pursue measures of uncertainty associated with return levels, a Bayesian estimation framework is developed. In all cases, for all techniques, results are presented via comparison between predicted and historical return levels.

Our results are most comparable to those of Douglas and Fairbank (2011) who investigated trends in extreme precipitation for stations in northern New England. Unlike their work, however, our results, the return levels, consider both predicted model output as well as current data. In their work, the authors conclude that coastal areas of northern New England have significant increases in maximum precipitation records. Our results, however, do not support this claim. Differences in point estimates between local-level observations and downscaled future estimates show widespread increases across most of New England, although many of these significant increases are not observed when we consider measures of uncertainty. From the three techniques used to assess uncertainty in our estimates (the parametric bootstrap, the non-parametric bootstrap with BCa adjusted confidence interval, and the Bayesian/MCMC framework), few significant increases in 25-year return levels were observed, and only subtle evidence of significant increases throughout locations in southern New England was identified. These results, however, are highly dependent on model choice. Certainly, future work will entail developing techniques to decrease sampling variability and achieve more precise estimates.

The choice of RCM and GCM driver has been shown to impact the downscaling procedure substantially. While it was not the primary purpose of this analysis, we have indirectly illustrated the effect this choice can have on downscaling and climate predictions. Related to this, in their presentation of guidelines for downscaling climate variables, Wilby et al. (2004) note that it is increasingly recognized that any comprehensive impact study should be founded on multiple GCM (or large-scale) model outputs. Schliep et al. (2010) analyzed historical output from six RCMs via a spatial Bayesian hierarchical model and found different RCMs yielded substantially different 100-year return level estimates. Going forward, quantifying the effect of RCM and GCM on projections should be pursued. Such work may entail a functional analysis of variance or functional ANOVA. Related and recent work in this area has been undertaken by Kaufman and Sain (2010) and Sain et al. (2010), for example.

References

Balkema, A., de Haan, L., 1974. Residual life time at great age. Ann. Probab. 2, 792–804.
Benestad, R.E., Hanssen-Bauer, I., Chen, D., 2007. Empirical Statistical Downscaling.

World Scientific Publishing, Singapore.
Caya, D., Laprise, R., 1999. A semi-implicit semi-lagrangian regional climate model: The Canadian RCM. Mon. Wea. Rev. 127, 341–362.
Coles, S., 2001. An Introduction To Statistical Modeling of Extreme Values. Springer, London.
Coles, S., Simiu, E., 2003. Estimating uncertainty in the extreme value analysis of data generated by a hurricane simulation model. J. Eng. Mech. 129 (11), 1288–1294.
Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press, New York.
Delworth, T.L., et al., 2006. GFDL's CM2 global coupled climate models - Part 1: Formulation and simulation characteristics. J. Clim. 19 (5), 643–674.
Douglas, E.M., Fairbank, C.A., 2011. Is precipitation in northern new england becoming more extreme? statistical analysis of extreme rainfall in massachusetts, new hampshire, and maine and updated estimates of the 100-year storm. J. Hydrol. Eng. 16 (3), 1.
Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., Mearns, L.O., 2000. Is precipitation in northern new england becoming more extreme? statistical analysis of extreme rainfall in massachusetts, new hampshire, and maine and updated estimates of the 100-Year storm. Science 289 (5487), 2068–2074.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, New York.
Engen-Skaugen, T., 2004. Refinement of dynamically downscaled precipitation and temperature scenarios. DNMI Climate Report 15, 1–20.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. J. Stat. Anal. 1 (3), 515–533.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian Data Analysis 2. Chapman & Hall/CRC, London.
Giorgi, F., Marinucci, M.R., Bates, G.T., 1993a. Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. Mon. Weather Rev. 121, 2794–2813.
Giorgi, F., de Canio, G., Bates, G.T., 1993b. Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions. Mon. Wea. Rev. 121, 2814–2832.
Groisman, P. Ya, et al., 1999. Changes in the probability of heavy precipitation: important indicators of climatic change. Clim. Chang. 42, 243–283.
Hoar, T., Nychka, D., 2008. Statistical downscaling of the community climate system model (CCSM) monthly temperature and precipitation projections. White paper preprint, Institute for Mathematics Applied to Geosciences/National Center for Atmospheric Research, Boulder, CO 80307.
Johns, T.C., Gregory, J.M., Ingram, W.J., Johnson, C.E., Jones, A., Lowe, J.A., Mitchell, J. F.B., Roberts, D.L., Sexton, B.M.H., Stevenson, D.S., Tett, S.F.B., Woodage, M.J., 2003. Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. Clim. Dyn. 20, 583–612.
Jones, R.G., Hassell, D.C., Hudson, D., Wilson, S.S., Jenkins, G.J., Mitchell, J.F.B., 2003. Workbook on generating high resolution climate change scenarios using PRECIS, UNDP, p. 32.
Kallache, M., Vrac, M., Naveau, P., Michelangeli, P.-A., 2011. Non-stationary probabilistic downscaling of extreme precipitation. J. Geophys. Res. 116, D05113.
Karl, T.R., Knight, R.W., 1998. Secular trend of precipitation amount, frequency, and intensity in the United States. Bull. Am. Meteorol. Soc. 79, 231–242.
Karl, T.R., Knight, R.W., Easterling, D.R., Quayle, R.G., 1996. Indices of climate change for the United States. Bull. Am. Meteorol. Soc. 77, 279–292.
Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Resour. 25, 1287–1304.
Kaufman, C.G., Sain, S.R., 2010. Bayesian functional ANOVA modeling using gaussian process prior distributions. Bayesian Anal. 5, 123–150.
Kharin, V.V., Zwiers, F.W., 2005. Estimating extremes in transient climate change simulations. J. Clim. Am. Meteorol. Soc. 18, 1156–1173.
Kiehl, J.T., Gent, P.R., 2004. The community climate system model, version two. J. Clim. 17, 3666–3682.
Kunkel, K.E., Andsager, K., Easterling, D.R., 1999. Long-term trends in extreme precipitation events over the conterminous United States and Canada. J. Clim. Am. Meteorol. Soc. 12, 2515–2527.
Maurer, E.P., Hidalgo, H.G., 2008. Utility of daily vs. monthly large-scale climate data: an inter-comparison of two statistical downscaling methods. Hydrol. Earth Syst. Sci. 12, 551–563.
McFarlane, N.A., Scinocca, J.F., Lazare, M., Harvey, R., Verseghy, D., Li, J., 2005. The CCCma third generation atmospheric general circulation model CCCma Internal Report, p. 25.
Mearns, L.O., Arritt, R., Biner, S., Bukovsky, M.S., McGinnis, S., Sain, S., Caya, D., Correia Jr., J., Flory, D., Gutowski, W., Takle, E.S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A.M.B., Qian, Y., Roads, J., Sloan, L., Snyder, M., 2012. The north american regional climate change assessment program: overview of phase I results. Bull. Amer. Meteorol. Soc. 93, 1337–1362.
Meehl, G.A., Collins, W.D., Boville, B., Kiehl, J.T., Wigley, T.M.L., Arblaster, J.M., 2000. Response of the NCAR climate system model to increased CO_2 and the role of physical processes. J. Clim. Am. Meteorol. Soc. 13, 1879–1898.
Michelangeli, P.-A., Vrac, M., Loukos, H., 2009. Probabilistic downscaling approaches: application to wind cumulative distribution functions. Geophys. Res. Lett. 36, L11708.
Nakicenovoic, N. et al., 2000. Special report on emissions scenarios a special report of working group III of the intergovernmental panel on climate change, Cambridge: Cambridge University Press.
Pal, J.S., et al., 2007. Regional climate modeling for the developing world: the ICTP RegCM3 and RegCNET. Bull. Amer. Meteorol. Soc. 88, 1395–1409.
Parmesan, C., Root, T.L., Willig, M.R., 2000. Impacts of extreme weather and climate on terrestrial biota. Bull. Am. Meteorol. Soc. 81, 443–450.
Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Stat. 3, 119–131.
R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.
Robeson, S.M., 2004. Trends in time-varying percentiles of daily minimum and maximum temperature over North America. Geophys. Res. Lett. 31, L04203.
Sain, S.R., Nychka, D., Mearns, L., 2010. Functional ANOVA and regional climate experiments: A statistical analysis of dynamic downscaling. Unpublished manuscript. http://www.image.ucar.edu/~ssain/publications/narccap_gfdl.pdf.
Schliep, E.M., Cooley, D., Sain, S.R., Hoeting, J.A., 2010. A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. Extremes 13, 219–239.
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 version 2. NCAR Technical Note NCAR/TN-4681STR, p. 88.
Skaugen, T., Astrup, M., Roald, L.A., Skaugen, T.E., 2002. scenarios of extreme precipitation of duration 1 and 5 days for Norway caused by climate change. Technical report. Norgesvassdrags-ogenergidirektorat, NVE, Postboks 5091 Majorstua, 0301 Oslo, Norway.
Smith, R.L., 2001. Trends in Rainfall Extremes Technical report, Department of Statistics, University of North Carolina, Chapel Hill NC 27599-3260 USA.
Tank, A.M.G. Klein, Konnen, G.P., 2003. Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. J. Clim., Am. Meteorol. Soc. 16, 3665–3680.
Tebaldi, C., Hayhoe, K., Arblaster, J.M., Meehl, G.A., 2006. Going to the extremes: an inter-comparison model-simulated historical and future changes in extreme events. Clim. Chang. 79 (3–4), 185–211.
U.S. Climate Change Science Program (USCCSP), 2008. Weather and climate extremes in a changing climate, U.S. Climate Change Science, Program Synthesis Analysis Production 3.3 ⟨http://www.climatescience.gov/Library/sap/sap3-3/final-report/sap3-3-final-allpdf⟩.
Wilby, R.L., Harris, I., 2006. A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK. Water Resour. Res. 42, W02 419.
Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Published on-line, supporting material to the IPCC, pp. 27.
Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim. Chang. 62, 189–216.
Zwiers, F.W., Kharin, V.V., 1998. Changes in the Extremes of the Climate Simulated by CC GCM2 under CO2 Doubling. J. Clim. Am. Meteorol. Soc. 11, 2200–2222.