Journal of Hydrology 538 (2016) 49–62

Contents lists available at ScienceDirect

Journal of Hydrology

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

Use of beta regression for statistical downscaling of precipitation in the Campbell River basin, British Columbia, Canada

Sohom Mandal [a,*], Roshan K. Srivastav [b], Slobodan P. Simonovic [a]

[a] Department of Civil and Environmental Engineering, Western University, London, ON, Canada
[b] School of Civil and Chemical Engineering, VIT University, Vellore, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form 20 March 2016
Accepted 3 April 2016
Available online 9 April 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Uwe Haberlandt, Associate Editor

Keywords:
Climate change
CART
K-means clustering
Beta regression
Statistical downscaling

SUMMARY

Impacts of global climate change on water resources systems are assessed by downscaling coarse scale climate variables into regional scale hydro-climate variables. In this study, a new multisite statistical downscaling method based on beta regression (BR) is developed for generating synthetic precipitation series, which can preserve temporal and spatial dependence along with other historical statistics. The beta regression based downscaling method includes two main steps: (1) prediction of precipitation states for the study area using classification and regression trees, and (2) generation of precipitation at different stations in the study area conditioned on the precipitation states. Daily precipitation data for 53 years from the ANUSPLIN data set is used to predict precipitation states of the study area where predictor variables are extracted from the NCEP/NCAR reanalysis data set for the same interval. The proposed model is applied to downscaling daily precipitation at ten different stations in the Campbell River basin, British Columbia, Canada. Results show that the proposed downscaling model can capture spatial and temporal variability of local precipitation very well at various locations. The performance of the model is compared with a recently developed non-parametric kernel regression based downscaling model. The BR model performs better regarding extrapolation compared to the non-parametric kernel regression model. Future precipitation changes under different GHG (greenhouse gas) emission scenarios also projected with the developed downscaling model that reveals a significant amount of changes in future seasonal precipitation and number of wet days in the river basin.

1. Introduction

Impacts of the climate change on hydro-meteorology include earlier snowmelt, lower flows in summer and fall, and an increase in the frequency of flooding and drought in many regions of the world. It has been reported that due to climate change, the frequency of warmer days has increased where the frequency of cold nights has decreased over the Canadian landmass during 1950–2010 (Warren and Lemmen, 2014). A stronger warming trend has been found in the west, and the northern part of Canada compared to the east coast. Mekis and Vincent (2011) reported that in Canada especially on the west coast the total precipitation has increased in fall and spring while it has decreased in winter during the period of 1950–2009.

According to IPCC (2007) increase in global average surface temperature since the mid-20th century is very likely due to an increase of greenhouse gasses (GHGs) concentration in the atmosphere and changes in surface temperature affects atmospheric circulation pattern which influences precipitation. Moreover, the global climate will continue to change in the future as a result of continued emissions of GHGs into the atmosphere. Therefore, we can anticipate that there will be further changes to global and regional temperature and precipitation patterns in the future. Altered patterns of future precipitation will affect the regional hydrology and water resources. Due to significant changes in precipitation patterns in past decades, water resources managers and planners are expressing interest in future precipitation projection under changing climate condition. Generally, impacts of climate change on regional water resources are assessed for future climate scenarios obtained from Global Climate Model (GCM) simulations. GCMs represent the state of the art with respect to the simulation of global climate variables in response to emission scenarios of greenhouse gasses. GCMs can satisfactorily model smoothly varying fields such as mean sea level pressure, but often fail to capture

* Corresponding author at: Department of Civil and Environmental Engineering, Western University, London, ON N6A 3K7, Canada.
E-mail addresses: sohomiitb@gmail.com (S. Mandal), roshan@vit.ac.in (R.K. Srivastav).

http://dx.doi.org/10.1016/j.jhydrol.2016.04.009
0022-1694/

non-smooth fields such as precipitation (Hughes and Guttorp, 1994). In addition, the spatial scale of GCM output is very coarse (>100 km^2). Therefore on a regional scale, capturing the impacts of climate change on hydro-meteorological variables (e.g. temperature, precipitation, soil moisture) is more difficult and uncertain. At the catchment level (<50 km), downscaling of coarsely gridded GCM data is necessary for a better understanding and assessment of future hydrologic conditions in response to climate change.

Downscaling methods can be used to improve the spatial resolution of GCM output to overcome the challenge of assessing climate related impacts on water resources at the catchment scale. Downscaling methods are broadly classified as dynamic or statistical. Dynamic downscaling is based upon nesting a finer scale regional climate model (RCM) (up to $10\ km \times 10\ km$ horizontal resolution) within GCMs (Wood et al., 2004). The major drawbacks of dynamic downscaling are model complications, high computational requirements, their dependence on boundary conditions obtained from GCMs and lack of transferability to different regions. Statistical downscaling (SD) methods use parametric/nonparametric and/or linear/nonlinear relationships between predictor and predictand variables (Wilby and Wigley, 1997). Statistical downscaling is more adaptable, flexible and popular because of low computational requirements, simple modeling structure and easy modifications for use at various locations. SD methods developed so far are can be classified into three groups: (a) classification/weather typing methods (Hay et al., 1991; Hughes and Guttorp, 1994; Hughes et al., 1999; Mehrotra and Sharma, 2005); (b) regression/transfer function methods (Chen et al., 2014; Ghosh, 2010; Goyal and Ojha, 2010; Hashmi et al., 2009; Kannan and Ghosh, 2013; Von Storch et al., 1993; Wilby et al., 2002; Wilby et al., 1999) and (c) weather generators (WG) (Wilks, 1999; Wilks and Wilby, 1999; Sharif and Burn, 2006; Eum and Simonovic, 2012; Lee et al., 2012; King et al., 2014; Srivastav and Simonovic, 2014).

Weather typing approaches develop the relationship between local climate and atmospheric circulation variables based on a given weather classification scheme. The observed local climate variables are related to weather classes which include principal component analysis (PCA) (Schoof and Pryor, 2001; Wetterhall et al., 2005) fuzzy rules (Bardossy et al., 1995), canonical correlation analysis (CCA) (Gyalistras et al., 1994), analogues procedure (Martin et al., 1997) or other pattern recognition methods based on correlation (Wilby and Wigley, 1997). The major drawbacks of this approach are the stationary relationships between local climate variables and different types of atmospheric circulation and the additional effort for weather classification. Non-stationarities are inherent traits of the climate system and can be observed in different spatio-temporal scale (Hertig and Jacobeit, 2013). Hence, ignoring the nonstationary relationships between climate variables may mislead the downscaling process. Transfer function based models usually build a statistical relationship between GCM or RCM outputs (large scale predictor) and local-scale climate variables (predictand). Generally multivariate linear or nonlinear regression (Chen et al., 2014; Vrac et al., 2007), non-parametric regression (Kannan and Ghosh, 2013; Sharma and O'Neill, 2002) and support vector machine (SVM) approach (Ghosh, 2010; Tripathi et al., 2006) are used for deriving those relationships. These approaches are widely used and known as 'perfect-prog nosis' downscaling methods. Weather generators are statistical models that stochastically simulate random sequences of synthetic climate variables that preserve statistical properties of observed climate data (Dibike and Coulibaly, 2005; Hashmi et al., 2011; Khan et al., 2006; Sharif and Burn, 2006).

In spite of progress made in the development of downscaling models in the past, the challenges still exist in representing temporal and spatial variability in the generated sequences (Wilby et al., 2004), accurately generating extremes (maximum and minimum) (Pour et al., 2014) and generating multisite sequences with spatial dependence. Most of the models developed in the past have failed to capture spatial dependence in rainfall occurrence and they assume that the probability distributions of observed and future climate variables remain the same, which can be a limiting assumption. Raje and Mujumdar (2009) developed a conditional random field (CRF) downscaling method which does not require the assumption of independence for climate variables and their distribution. In this method, four surface flux variables (precipitation flux, surface temperature, maximum and minimum surface temperature) and four surface/pressure variables (specific humidity, sea level pressure, U wind and V wind) are needed to maintain spatial and temporal dependence which make this method computationally demanding. In addition, the CRF method moderately captures spatial correlation and also overestimates the mean value of the predictand (precipitation). Individually downscaling at multiple stations may be the reason for poor spatial correlations and discretization of historical rainfall data into different classes without confirming an exact number of rainfall classes using clusters validity test may produce bias toward over-prediction of mean precipitation values at different stations. For this reason non-parametric statistical methods like K-nearest neighbors (K-nn) (Brandsma and Buishand, 1998; Eum and Simonovic, 2012; King et al., 2014; Sharif and Burn, 2006; Young, 1994) or Kernel density estimator are referred in the literature as plausible approaches for the downscaling purposes (Kannan and Ghosh, 2013; Mehrotra and Sharma, 2010). Although non-parametric methods can successfully capture the spatial dependence of observed data, they often fail to capture extreme events in the case of precipitation. Markov based downscaling models (Hughes and Guttorp, 1994; Mehrotra and Sharma, 2005; Mehrotra and Sharma, 2007) perform satisfactory in capturing spatial variability of daily precipitation but they fail to reproduce the variability of a non-stationary climate as exogenous climate predictors are not considered.

In spite of considerable progress in development of downscaling methods, especially for simulation of precipitation, challenges still exist in accurately capturing extreme behavior in generated precipitation sequences, simulating multisite sequences with realistic spatial and temporal dependence (Raje and Mujumdar, 2009). Moreover, downscaling method should be efficient and computationally inexpensive to simulate the underlying processes present in the observed data. Recently, Kannan and Ghosh (2013) developed a multisite statistical downscaling model using a non-parametric kernel density function. Exogenous climate predictors were used in this method for generating multisite precipitation. This method encouraged us to develop a statistical downscaling model based on a new regression approach.

In this study, we propose a new statistical downscaling approach which considers the historical effect of exogenous climate variables for the generation of multisite precipitation amounts. Since the occurrence of precipitation is influenced by global circulation patterns, in this approach we derived precipitation states in the river catchment using a classification and regression tree (CART) (Breiman et al., 1984). Modeling spatial dependence is the biggest challenge in downscaling (Yang et al., 2005) and it is addressed here in an innovative way.

Precipitation states of the basin are obtained from large scale circulation patterns to capture the spatial patterns within the basin. We also use multivariate beta regression model to downscale multisite precipitation amounts conditioned on precipitation states of the catchment.

Based on the precipitation states, beta regression is used to generate precipitation at each individual location within the catchment. This regression method based on the beta distribution has proven to be very versatile and flexible to model exogenous variables (Ferrari and Cribari-Neto, 2004) and is novel in its application as a statistical downscaling technique. For model performance evaluation, the results obtained from the proposed method are compared with those obtained from a recently developed model based on Kernel density estimation (Kannan and Ghosh, 2013). The primary objective of the comparison is to analyze the advantages and disadvantages of the proposed beta regression based downscaling method.

The methodology is developed in two steps. First, precipitation states are predicted using the CART algorithm. Second, we generate time series of multisite daily precipitation by downscaling outputs of CanESM2 for a historical time period (1983–2005) and a future time period (2036–2065). The proposed downscaling method is applied to a case study in the Campbell River basin, British Columbia, Canada (Fig. 1). Information regarding case study area and datasets used for this study are given in the next section followed by methodology. Next model application is discussed followed by results and discussion. Summary and conclusions are presented in the last section of this report.

2. Case study area and data used

Campbell River is situated in between the dry east coast and wet west coast climate on the Vancouver Island, Canada. The total drainage area of this basin is approximately 1856 km^2 with a length of 33 km from the origin (Strathcona provincial park). The Campbell River basin consists of three lakes namely Buttle Lake and Upper Campbell Lake, Lower Campbell Lake and John Hart Lake. Campbell River system produces 2.5% of total BC Hydro's hydroelectric power which is equivalent to 11% of Vancouver Island's annual energy demand (BC Hydro Generation Resource Management, 2012). In this river basin, streamflow is a mixture of melting snow and rainfall. Generally, the streamflow is high during fall and spring and low during the summer season. The salient features (longitude, latitude, elevation) of the gauging stations in the basin are given in Table 1.

Historical daily precipitation data (0.1° latitude × 0.1° longitudes) for a 40 years span (1961–2013) have been obtained from the ANUSPLIN Data Set, Environment Canada (Hutchinson and Xu, 2013). ANUSPLIN data is developed using "thin plate smoothing splines" algorithm. This technique interpolates climate variables as a function of latitude, longitude, and elevation. For this study, the daily precipitation data is used at ten locations covering the entire Campbell River basin. Large-scale climate circulation patterns govern the regional climate. Therefore, selection of the predictors is necessary for the downscaling process (Wetterhall et al., 2005; Wilby et al., 2004). According to Wilby et al. (1999), predictors used for downscaling need to be: (a) easily available, (b) reliably simulated and (c) strongly correlated with response variable (precipitation in this case). Considering those conditions, daily maximum and minimum air surface temperature (Tmax and Tmin), mean sea level pressure (mslp), specific humidity (hus) at 500 hPa, zonal (u-wind) and meridional (v-wind) wind are used as predictors.

Table 1
Salient features of precipitation stations in the Campbell River basin, BC, Canada.

Station	Elevation (m)	Latitude (°N)	Longitude (°W)	Station abbreviation
Elk R ab Campbell Lk	270	49.85	125.8	ELK
Eric creek	280	49.6	125.3	ERC
Gold R below Ucona R	10	49.7	126.1	GLD
Heber river near gold river	215	49.82	125.98	HEB
John hart substation	15	50.05	125.31	JHT
Quinsam R at argonaut Br	280	49.93	125.51	QIN
Quinsam R nr Campbell R	15	50.03	125.3	QSM
Salmon R ab Campbell div	215	50.09	125.67	SAM
Strathcona dam	249	49.98	125.58	SCA
Wolf river upper	1490	49.68	125.74	WOL

Due to inadequate historical climate data for a longer period, predictor data is extracted from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis dataset (Kalnay et al., 1996) for 53 years spanning 1961–2013. NCEP/NCAR data set is a combination of physical process and model forecast gridded data at the 2.5° × 2.5° spatial resolution. In the context of GCM outputs downscaling, historical data from CanESM2 (1983–2013) is used for proposed model performance evaluation. CanESM2 (2.813° latitude × 2.79° longitude) is a second generation earth system model from the Coupled Model Inter-comparison Project (CMIP5) developed by the Canadian Centre for Climate Modeling and Analysis.

ANUSPLIN, NCEP/NCAR and GCM (CanESM2) data have a different spatial resolution. Therefore, all the data sets are spatially interpolated to a location of interest (gauging station) using inverse distance square method (Gaur and Simonovic, 2013). Six

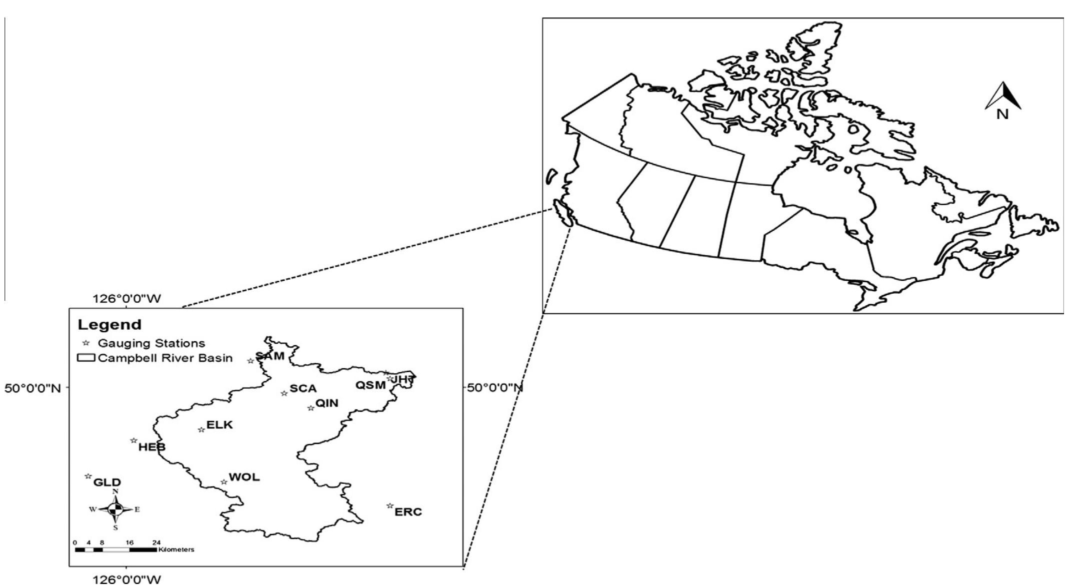

Fig. 1. The Campbell River basin with the location of gauging stations.

climate variables (Tmax, Tmin, mslp, hus, u-wind and v-wind) at ten locations in the basin are used as model predictors where precipitation is model predictand.

Standardization procedure (Wilby et al., 2004) is applied to the predictors data to reduce the systematic bias among the variable means and standard deviations. Standardization is carried out by subtracting the mean and dividing by standard deviation from all respective variables. A 30 years span (1961–1990) is considered as a model training period, where 23 years (1991–2013) daily data is used for model validation. Predictors for a particular station are expected to have a high correlation with other nearby stations which may lead to the multicollinearity problem (Ghosh, 2010). Multicollinearity is a statistical phenomenon which refers to highly correlated predictors in multiple regression analyses. It occurs when predictors are not only correlated with response variable but also to each other. Multicollinearity may lead to larger changes in the regression model estimation for small changes in the data. Therefore, it is necessary to remove multicollinearity from the predictor variables (Salvi et al., 2013). Apart from this, the model is expected to be computationally inexpensive for its multiple dimensions. Now if the dimensions are reduced without considering the internal variability and patterns of the data, it may lead us to an erroneous model result. Hence, to reduce the multicollinearity and dimensionality, the principal component analysis (PCA) is used. PCA is a powerful statistical tool which can identify patterns in multidimensional data. On the other hand, it can reduce dimensions without reducing the internal variability of the original data. There are no clear rules for choosing a number of principal components that explains the maximum percentage of variance. Srivastav and Simonovic (2014) investigated the performance of a multisite weather generator with different principal components and considered first principal component for their study. Kannan and Ghosh (2013) adopted Kasier's rule for selecting principal components that explain more than the average amount of total variance. In this study, we considered first five principal components that explain 97% variability of the original data (Fig. 2).

3. Methodology

The details of beta regression based statistical downscaling technique conditioned to the precipitation states are outlined in this section. The proposed modeling framework is shown in Fig. 3. This framework is divided into two parts. In the first part

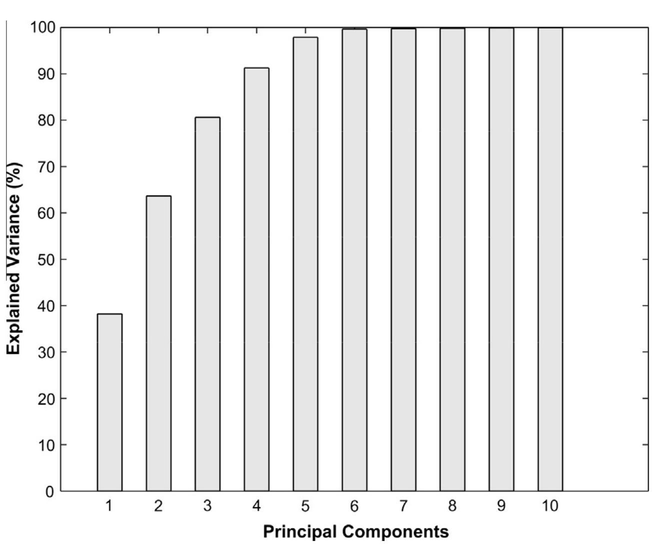

Fig. 2. Cumulative percent of variance explained by principal components.

(Fig. 3(a)), the daily precipitation states are generated using a supervised classification technique, namely CART (classification and regression trees) wherein the second part (Fig. 3(b)), the daily precipitation sequences are generated for a particular location using multivariate beta regression. CART classifies predictor variables or builds relationship in terms of explanatory power and variance using an "acyclic tree". The following subsections describe in details procedures for generation of precipitation states (part 1) and daily precipitation generation (part 2).

3.1. Generation of daily precipitation states

The daily precipitation state is a qualitative representation of precipitation status for a given day in a particular region where multiple sites of interest belong. For predicting daily precipitation states in the river basin, CART algorithm coupled with an unsupervised classification method (K-means clustering) is used following Kannan and Ghosh (2013). K-means clustering helps to identify daily precipitation states in the river basin. CART is a classification and regression algorithm based on 'if-then' logic. The advantages of using CART are: (1) it does not follow a prior statistical distribution of predictors; (2) it is flexible and efficient with high dimensional data; and (3) it can effectively deal with a mixture of categorical, discrete and continuous predictor variables (James et al., 2013). The procedure for daily precipitation states estimation is explained in Fig. 3(a). It includes few steps as follows:

Step-I: Use K-means clustering technique for identifying precipitation states from the observed ANUSPIN precipitation data (1961–1990). For an optimum number of clusters, we used cluster validity index e.g. Silhouette Index, Davis–Bouldin index, Dunn Index and Connectivity measures (Brock et al., 2008).
Step-II: Standardize the NCEP/NCAR predictor variables by subtracting mean and dividing the data by standard deviation. After standardization, PCA is used to reduce the dimension and remove multicollinearity from the standardized predictor variables. Preserve the principal component/scores and Eigen vectors/factors for the next step.
Step-III: Apply the standardization procedure and PCA to historical NCEP/NCAR predictor data and historical GCM predictor data (CanESM2) for a different time period.
Step-IV: Build the CART with the help of principal components obtained from NCEP/NCAR predictor data and precipitation states obtained from K-means.
Step-V: Apply the CART model to GCM historical data (1983–2005) and historical NCEP/NCAR data (1991–2013) to derive rainfall states for a different time period and compare statistics with observed historical data for the same time period. This two different historical time periods are used for validate the proposed downscaling model with GCMs and NCEP/NCAR data set.
Step-VI: For calculating future precipitation states under different climate change scenarios the CART model is applied to standardized future GCM (CanESM2) predictor data (2036–2065).

Preserving the spatial correlation and capturing the variability of predictand are the important aspects of the statistical downscaling. Hence, it will be more acceptable if the procedure provides for derivation of precipitation states first and then generate precipitation amount. Precipitation states of the river basin combined with data driven regression approach (beta regression) preserve the spatial dependence in the precipitation fields. This combined procedure retains the marginal and joint density structure of historical precipitation series which includes nonlinearity and state dependence.

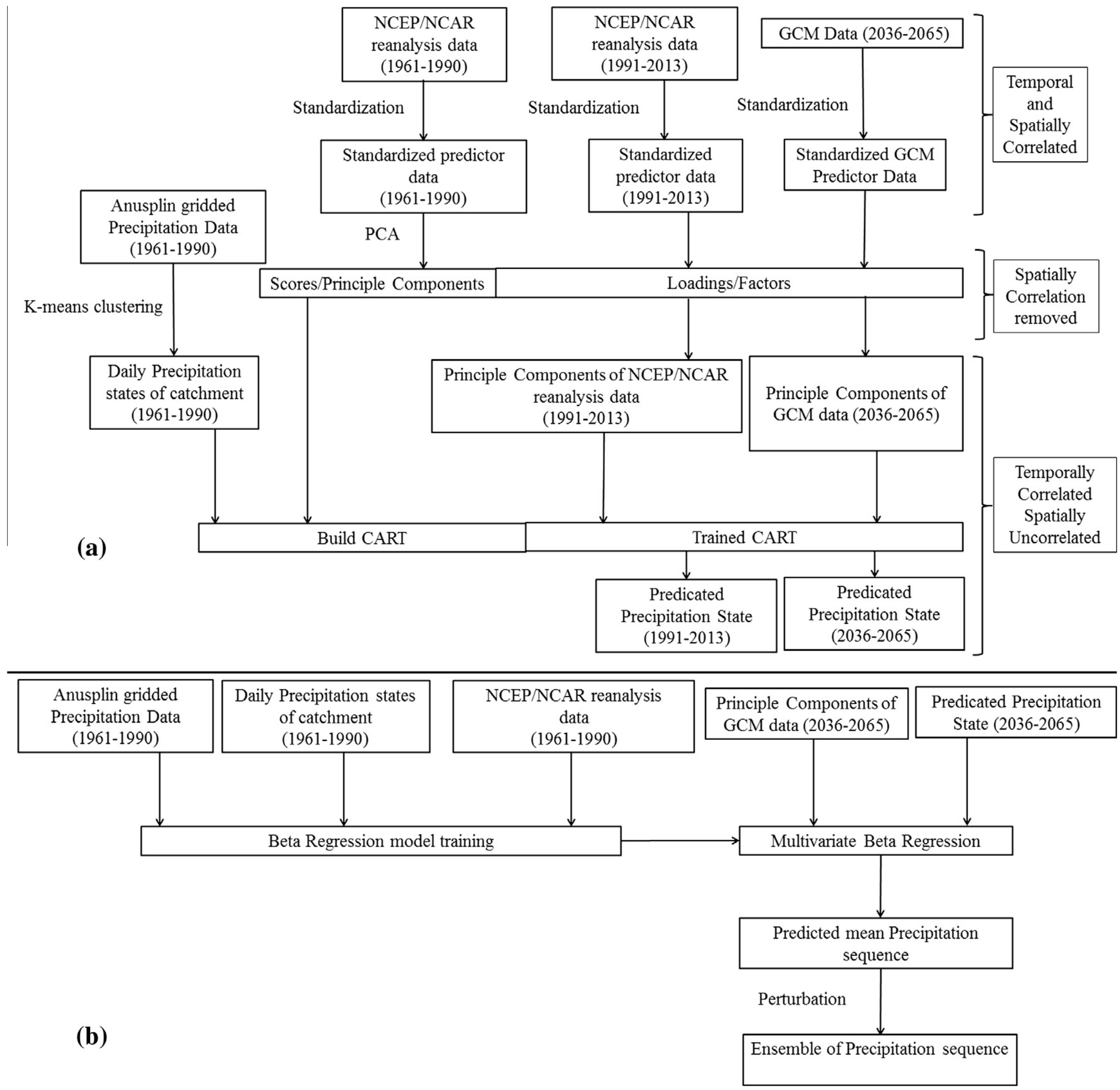

Fig. 3. The schematic of proposed downscaling framework. (a) Prediction of precipitation state using CART. (b) Multivariate beta regression model for synthetic precipitation generation.

3.2. Multisite precipitation generation

For multisite precipitation generation, a relationship between predictors and predictand climate variables has to be developed.

$$P_t = F_R(X_t/S_t) \tag{1}$$

The generalized relationship between predictors and predictand is described by Eq. (1) where P_t is the precipitation at a certain station at time t, X_t is predictor variables at time t and S_t is precipitation state of the river basin at time t.

Generally this kind of relationship is developed using regression (parametric/non-parametric) or probabilistic approach (Wilby and Harris, 2006; Mehrotra and Sharma, 2007; Srivastav and Simonovic, 2014). In this study, beta regression is applied to model the above mentioned relationship. The predictors used for build the regression model are current day principal components of reanalysis predictor data and current day precipitation states from CART where predictand is present day precipitation at different stations (generated separately).

3.2.1. Beta regression

Regression analysis builds a relationship between independent variables (x) and dependent variable (y). In this study, large-scale global climate variables are independent variables or predictors and precipitation is dependent variable or predictand. The relationship between them can be formulized as follows:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots n \tag{2}$$

where ε_i is a normally distributed non-zero error term. If the relationship is linear then the expression (2) is modified as follows:

$$y = x^T\beta + \varepsilon_i = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots x_d\beta_d + \varepsilon_i \tag{3}$$

where x is a vector of predictor variables with dimension d and β is a coefficient vector. The relationship in Eq. (3) is developed using beta regression (BR). This regression approach follows the beta distribution. The beta distribution is very flexible for modeling dependent variables since its density can assume a number of different shapes based on its parameters. Apart from this, the beta distribution is heteroskedastic and can successfully accommodate asymmetric data (Ferrari and Cribari-Neto, 2004; Schmid et al., 2013). Another advantage of using beta distribution is that it can model nonlinear relationship (Simas et al., 2010). The beta density function of the predictand can be written as:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},$$
$$0 < y < 1, 0 < \mu < 1, \phi > 0 \tag{4}$$

μ is the mean of predictand, ϕ is precision parameter, y is dependent variable and $\Gamma(\cdot)$ is gamma function. Beta distribution includes gamma function. In past, gamma function was successfully implemented to model precipitation (Groisman et al., 1999; Stern and Coe, 1984; Wilks and Wilby, 1999). The shape of beta density function can change depending on the values of μ and ϕ which help to estimate and model underlying structure of the data without assuming any functional form of estimators (Schmid et al., 2013). If $\mu = 1/2$ then the model is symmetric and if $\mu \neq 1/2$ then the model becomes asymmetric.

The proposed regression model assumed that the dependent variable or predictand is beta distributed and constrained to the unit interval (0, 1). Therefore any dependent variable bounded in an interval (a, b) where a and b are known scalar values ($a < b$) need to be scaled to (0, 1) interval. For this case y (predictand) is scaled into (0, 1) interval using the following two steps:

Step (i):

$$y' = (y-a)/(b-a) \tag{5}$$

Step (ii):

$$\Pr_{scaled} = (y'(n-1)+0.5)/n; \tag{6}$$

where y is precipitation data, n is sample size and Pr_{scaled} is scaled precipitation data into (0, 1).

To relate the conditional expectation function $E(y/x)$ for multivariate predictors, beta regression assumes a predictor-predictand relationship given by

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i \tag{7}$$

$$x_{ti} = (x_{t1}, \ldots, x_{tk}); \; t = 1, \ldots, n \tag{8}$$

$$\beta_i = (\beta_1, \ldots \beta_k)^T (\beta \in \mathbb{R}^k) \tag{9}$$

where β_i is a vector of unknown regression parameters and x_{ti} are observations of k covariates ($k < n$). $g(\cdot)$ is strictly monotonic and invertible link function that maps (0, 1) into $\mathbb{R}$. Many types of link functions are possible here (e.g. probit, logit, log–log). Logit transformation is used for this work following Ferrari and Cribari-Neto (2004). Maximum likelihood estimation (MLE) is used to estimate β.

One of the major challenges of downscaling methods is generation of precipitation data outside of the observed range. A perturbation technique is used with stochastic precipitation simulations and enhances the generation of extreme precipitation following King et al. (2015). The following equation is used for perturbation:

$$y^j_{ppt,t+i} = \lambda_{ppt} x^j_{ppt,t+i} + (1-\lambda_{ppt}) z_{t+i}; \; t = 1, 2, \ldots n \tag{10}$$

where $y^j_{ppt,t+i}$ is the perturbed precipitation value for $t + i$th day in jth location, $x^j_{ppt,t+i}$ is precipitation value for $t + i$th day in jth location and t is number of days and z_{t+i} comes from two parameter lognormal distribution (King et al., 2015). λ_{ppt} value varies in between 0 and 1 (0 means data series are totally perturbed and 1 means no perturbation in the results) and larger value of λ_{ppt} is reasonable to preserve spatial correlation. It has been found that $\lambda_{ppt} = 0.9$ can adequately preserve spatial correlation and other statistics (i.e. mean, variance) while it can still produce precipitation values outside of the observed range (King et al., 2015).

KNN algorithm is used to resample a block of days and ranks them. A cumulative probability distribution is calculated based on a day's rank. The next day precipitation is selected based on this probability distribution and a random number u (0, 1) which selects the closest day. For instance, precipitation of a day which is similar to present day precipitation has a higher probability of being selected and that helps to preserve temporal correlation of climate variables. After the resampling, perturbation is used to reshuffle the precipitation values. This process can be repeated several times for generating alternative precipitation realizations.

3.3. Model application

An unsupervised K-means clustering method is used to identify historical daily precipitation states (1961–1990) in the river basin. The optimum number of clusters or precipitation states are obtained from cluster internal validity tests e.g. Connectivity measure, Silhouette index and Dunn index and Davis–Bouldin index (Brock et al., 2008). Each validity index has different criteria for identification of an optimum number of clusters. For an optimum number of clusters, connectivity index value should be minimized where Silhouette index, Davis–Bouldin index and Dunn index value should be maximized. All four indices are tested for a number of clusters varying from 2 to 10 (Fig. 4). Apart from cluster validity index there is a hydrological aspect in selecting number of precipitation states or optimum number of clusters (Kannan and Ghosh, 2010). Supplementary Table-1 (ST1) shows cluster centroids calculated using k-means clustering technique for clusters varying from 2 to 4. It is found that the dry condition states (low cluster centroid value) are well separated from the other states in all clusters (ST1). To preserve the daily temporal correlational among predictor (large scale global climate variable) and predictand (precipitation) dry state condition need to be identified. Hence, the number of clusters exceeding 2 is considered following Kannan and Ghosh (2010). Cluster validity indices show that the optimal number of clusters should be greater than 2 where Davis–Bouldin index indicates 3 clusters as the optimal number. After the cluster validity measure analysis and consideration of other hydrologic issues, 3 clusters are selected to be used in this study. These clusters divide precipitation states into "almost dry", "medium" and "high," based on precipitation amount stored in the cluster centroid. Daily precipitation amount divided into different clusters provides more realistic prediction of precipitation states (Kannan and Ghosh, 2010).

CART model is constructed to predict precipitation states in the river basin using principal components derived from NCEP/NCAR predictor data and historical daily precipitation states obtained from the K-means clustering. CART prunes a classification tree conditioned to daily precipitation states. Principal components of NCEP/NCAR predictor variables for 30 years period (1961–1990) are used to prune the tree where the remaining 23 years of data (1991–2013) is used for validation of the model. It has been reported that the performance of classification tree was acceptable using NCEP/NCAR data with a lag-1 precipitation state (Kannan and Ghosh, 2010). The following relationship is used for building the CART model:

$$s_t = f\{p_t, p_{t-1}, s_{t-1}\} \tag{11}$$

where s_t is precipitation state, p_t is set of climate variables on tth day and p_{t-1}/s_{t-1} is precipitation state/set of climate variables on $(t-1)$th day.

Therefore, CART model build in this study used principal components of NCEP/NCAR predictor variables of the current day and the previous day with lag-1 precipitation state. BR model constructs a featured linear space based on identified daily precipitation states for the daily multisite precipitation generation. The linear space contains standardized and dimensionally reduced NCEP/NCAR predictors and corresponding daily observed

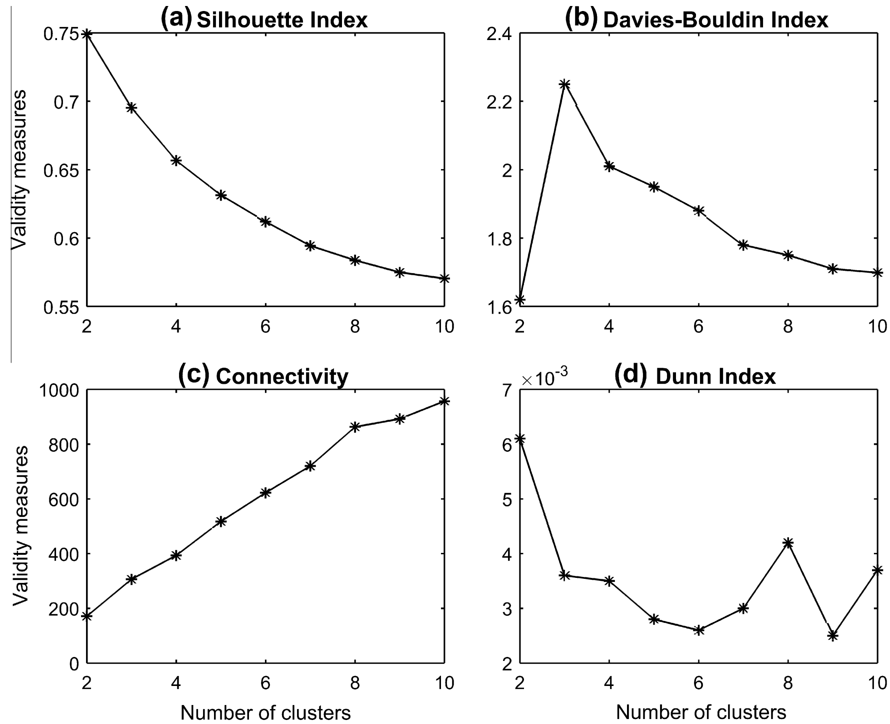

Fig. 4. Cluster validity measures.

precipitation data for 30 years period (1961–1990). For BR model validation, the remaining 23 years (1991–2013) of standardized and dimensionally reduced NCEP/NCAR predictors are used conditioned to the precipitation states. In the context of GCM data downscaling and model performance test using GCM outputs, standardized and dimensionally reduced historical (1983–2005) predictors from the CanESM2 are downscaled and compared with daily observed data for the same period.

The proposed BR model simulated output is compared with the multi-site non-parametric kernel regression (KR) model (Kannan and Ghosh, 2013). The kernel regression model has been used for generating multisite precipitation in the Mahanadi river basin, India. This model combines K-means, bias-correction, PCA, CART and kernel regression to generate synthetic precipitation. Simulation results from the BR model without rainfall state conditioning (BRWS) is also compared with the proposed BR model results in order to better understand the role of rainfall state in the downscaling process.

The comparison details and advantage of the BR model are discussed in the next section. A brief description of models used in this study with their acronyms is listed in Table 2.

4. Results and discussion

The objective of this study is to demonstrate the efficacy of the proposed multisite BR model. Using BR model, 30 independent realizations are generated for the validation period (1991–2013). The present downscaling method is also applied to GCM (CanESM2) simulated standardized predictor data for a future time (2036–2065) periods. Proposed BR model performance evaluation is based on the reproduction of historical statistics such as (1) temporal mean and standard deviation, (2) seasonal total precipitation (3) temporal and spatial cross correlation, and (4) preservation of quantiles. Results from different downscaling approaches such as BR, BRWS and KR are evaluated for spatial and temporal variation of precipitation over the river basin.

Table 2
Brief description of models used for comparison.

Acronym	Description
BR	Beta regression conditioned to precipitation states
BRWS	Beta regression not conditioned to precipitation states
KR	Kernel regression conditioned to precipitation states

4.1. Model evaluation over the validation period

4.1.1. Comparison of statistical characteristics

The statistical characteristics (such as mean and standard deviation) from BR, BRWS and KR model applications are shown in Table 3 and they are compared with observed precipitation for the validation period (1991–2013) at ten downscaling locations. Student *t*-test is conducted to check if the means of model simulated precipitation series at different stations similar to those of the observed data. The hypothesis is stated as "H_0: means of two series are the same" at 5% significance level. Results from the *t*-test are presented in Table 4. It can be seen that the BR model can generate precipitation time series similar to observed precipitation at different stations except two locations: GLD and WOL. The BRWS and KR model results show mixed outcomes at a 5% significance level.

4.1.2. Basin average annual and monthly precipitation

Streamflow of the Campbell River is affected by snowmelt and rain. Peak streamflows are observed in spring and fall while the low flows are usually experienced during the summer and winter (Zwiers et al., 2011). Hence, annual or seasonal changes in precipitation (snow/rain) will affect streamflow in the river. We compared the annual and monthly variability of basin average precipitation (50th percentile estimates) simulated from different models for the validation period Fig. 5(a) and (b) compares annual and mean monthly precipitation generated by BR, BRWS and KR

Table 3
Mean and standard deviation for observed and simulated precipitation (mm) series.

	Downscaling location									
	ELK	ERC	GLD	HEB	JHT	QIN	QSM	SAM	SCA	WOL
	Mean									
Observed	6.01	5.91	7.29	7	4.56	5.45	4.53	5.5	5.47	6.31
BR	5.4	7.2	7.37	6.65	3.95	3.49	3.94	4.27	4	7.16
BRWS	3.2	5.52	6.49	9.79	6.23	6.58	6.33	6.51	6.96	5.97
KR	9.00	9.08	10.50	9.77	6.02	7.71	6.00	7.69	7.49	9.61
	Standard deviation									
Observed	10.22	10.22	12.7	12.4	8.19	9.84	8.20	9.63	9.79	10.81
BR	10.58	11.65	10.5	12.8	7.89	9.62	7.63	10.02	8.8	9.8
BRWS	8.40	7.28	9.02	9.36	4.94	8.77	7.94	8.75	9.26	7.99
KR	12.9	13.02	15.76	14.93	9.26	11.67	9.25	11.47	11.45	13.87

Table 4
Hypothesis test results for testing mean of observed and simulated precipitation series.

Station	Student's *t* test result for acceptance/rejection of the null hypothesis at 5% confidence		
	KR	BRWS	BR
ELK	Reject	Do not reject	Do not reject
ERC	Reject	Do not reject	Do not reject
GLD	Reject	Do not reject	Reject
HEB	Do not reject	Do not reject	Do not reject
JHT	Do not reject	Do not reject	Do not reject
QIN	Do not reject	Reject	Do not reject
QSM	Do not reject	Reject	Do not reject
SAM	Do not reject	Reject	Do not reject
SCA	Do not reject	Reject	Do not reject
WOL	Reject	Do not reject	Reject

models in the river basin for a 23 year time period (1991–2013). Fig. 5(c) and (d) presents the correlation coefficient between basin average annual and monthly precipitation simulated by different downscaling models and observed precipitation for the validation time period. BR model simulated mean annual precipitation (basin average) series shows a high correlation (correlation coefficient of 0.88) with the observed precipitation, which means that the BR model can capture annual variability fairly well over the basin. For monthly basin average precipitation BR shows a satisfactory match with the observed series and obtained correlation coefficients is 0.83, where KR performs moderately well with a monthly correlation coefficient of 0.64. Overall beta regression based method outperforms all other models in terms of capturing annual and monthly variability.

Fig. 6(a)–(c) compares cumulative distribution function (CDF) of basin-average simulated precipitation series generated from different downscaling methods with those obtained using observed rainfall series. Compare to KR and BRWS, the CDF computed from BR model simulated data shows minimum deviation from the CDF obtained using observed precipitation. CDF of basin average precipitation obtained from BR model using historical CanESM2 GCM predictors (1983–2005) data is shown in Fig. 6(d) together with CDF of the observed precipitation. It seems BR model fairly well represents basin average precipitation using GCM (CanESM2) predictors data (Tmax, Tmin, mslp, hus, u-wind and v-wind). Another important observation is that the BR model is capable of capturing the percent of dry days (precipitation $\leqslant$ 1 mm/day) adequately. Using the BR, percent of dry days in the river basin calculated from simulated precipitation data is 42% for validation period, almost equal to actual observed dry day percent (Fig. 6 (a)). Although, KR performs well in capturing percent of dry day (42%) but it has an upward shift which refers a decreased frequency of precipitation events from actual. Percent of dry days calculated from the CanESM2 (48%) is also acceptable when compared

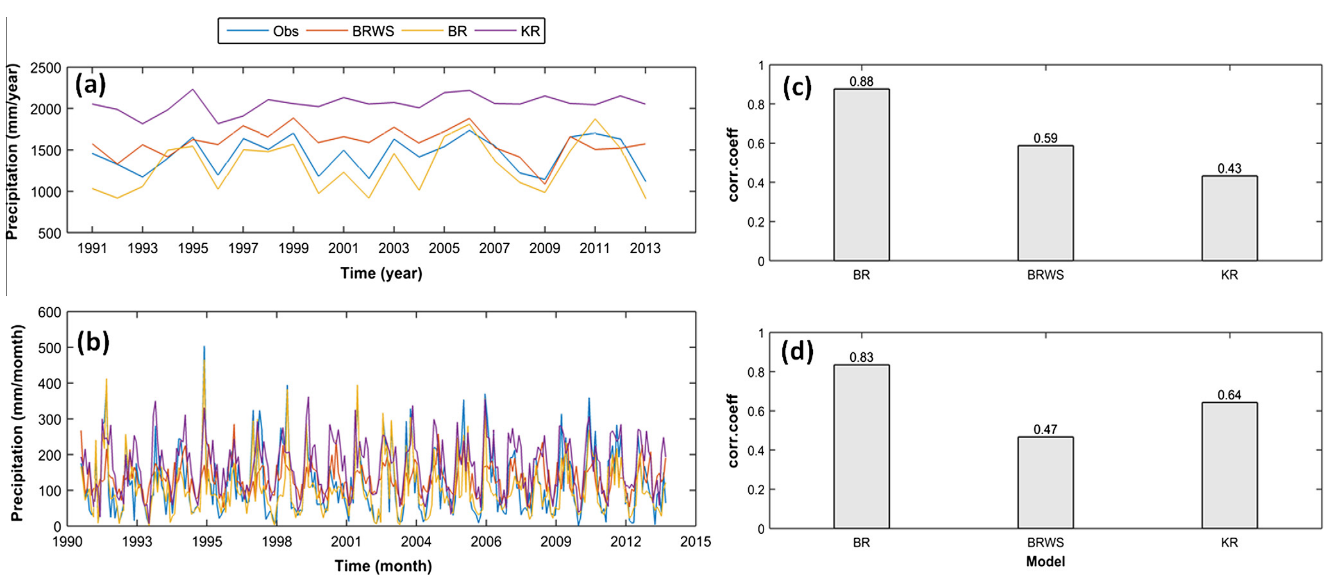

Fig. 5. (a) Annual and (b) monthly mean precipitation, spatially averaged over the Campbell River basin. The corresponding temporal correlation coefficients for different downscaling approaches are shown in (c and d).

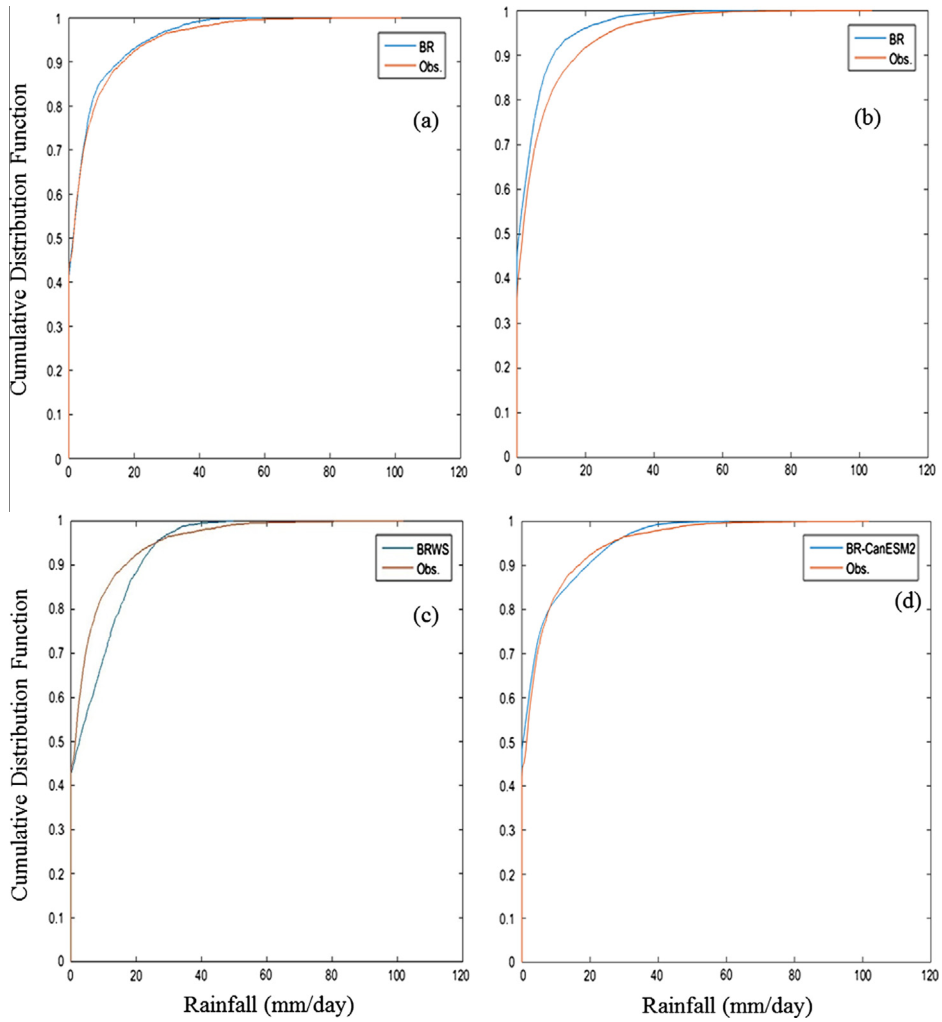

Fig. 6. (a–c) CDF of basin average precipitation obtained from different downscaling methods using reanalysis data (1991–2013). (d) CDF of basin average precipitation obtained from BR model using CanESM2 data (1983–2005).

to observed dry day percent (Fig. 6(d)). However, BR model ability to capture extreme precipitation (maximum) is very poor.

4.1.3. Basin average wet/dry spell length and seasonal precipitation amounts

Wet and dry spell lengths are very important in water resource planning and management especially for countries where reservoir needs a certain water storage level for hydropower generation. Hence, reproduction of wet/dry spell lengths along with seasonal precipitation is a very important aspect of the downscaling process. Although there are many definitions presented in the literature for wet/dry spell (WS/DS) length, we used the following definition for WS/DS from WCRP (2009). A WS (DS) defined as a maximum number of consecutive days with precipitation greater than (less or equal to) 1 mm.

Table 5 shows the annual average total seasonal precipitation and compares 5th, 50th and 95th percentile for both observed and downscaled precipitation. It is found that overall performance of BR model is better compared to KR in terms of capturing seasonal total precipitation. KR performs well in spring and summer period. However, KR simulates the high amount of precipitation in winter and fall compare to historical precipitation which is not acceptable in water resource planning and management. Table 6 describes annual average wet and dry spell lengths from simulated precipitation. BR and KR perform similar in reproducing dry spell length, but BR performs well in capturing wet spell length. It seems from both Table (Tables 5 and 6) BR model also performs satisfactorily in capturing seasonal precipitation (except fall) and WS/DS length using CanESM2 predictors data.

4.1.4. Temporal variability and spatial dependence

Assessment of temporal and spatial variability of precipitation is high importance for water resource management (municipal water supply, irrigation scheduling, hydropower generation, etc.). A better understanding of rainfall variability (temporal and spatial) is needed to better manage impacts of natural disasters (e.g. floods, droughts) in a changing climate. Therefore, the downscaling models should capture the temporal and spatial variability of precipitation accurately. We examined the performance of all downscaling methods in capturing temporal and spatial dependence of simulated precipitation series. Table 7 provides correlation coefficient between models simulated precipitation time series and observed precipitation at all ten downscaling locations for the validation period. From the results, it can be concluded that the overall performance of BR model conditioned to precipitation states is moderately better when compared to other methods. Fig. 7 shows the scatter plot of interstation correlation coefficients computed from model-simulated daily precipitation series and observed precipitation for all station pairs using different modeling approaches. From the plot in Fig. 7 it can be concluded that the BR model captures

Table 5
Observed and downscaled annual average seasonal total precipitation (5th, 50th (median) and 95th percentile) for testing period (1991–2013).

Season	Rainfall amount (mm)				
		Simulated percentile estimate			Percentage change in median value
	Observed	5th percentile	Median	95th Percentile	Rainfall amount
Model using Reanalysis data for 1991–2013					
Model: BR					
Winter	253.61	240.32	284.92	310.25	12.34
Spring	409.33	354.23	376.86	404.21	−7
Summer	263.10	224.5	267.06	289.36	1.5
Fall	188.39	180.5	217.37	266.52	15.3
Model: BRWS					
Winter	253.61	358.23	410.55	425.36	61.8
Spring	409.33	554.23	605.11	630.23	47.8
Summer	263.10	359.36	404.73	456.32	53.4
Fall	188.39	219.32	237.31	265.36	25.9
Model: KR					
Winter	253.61	290.32	355.63	420.32	40.22
Spring	409.33	410.35	494.11	550.35	20.71
Summer	263.10	265.36	290.90	310.85	10.56
Fall	188.39	289.36	308.43	340.25	63.72
Downscaled precipitation data using current climate data of GCM (CanESM2) for 1983–2005					
Model: BR					
Winter	204.65	178.36	195.62	230.23	−4.41
Spring	304.21	256.36	290.23	331.65	−4.59
Summer	257.99	240.36	278.36	339.36	7.89
Fall	129.99	155.36	160.36	225.36	23.35

Table 6
Annual averaged dry spell length and wet spell length of observed and downscaled precipitation.

Obs	BR	BRWS	KR
Model: using Reanalysis data for 1991–2013			
Dry spell length			
21	19	10	18
Wet spell length			
23	20	14	37
Obs			BR
Model: using current climate data of GCM (CanESM2) for 1983–2005			
Dry spell length			
26			23
Wet spell length			
28			26

Table 7
Correlation coefficients obtained for observed and simulated precipitation series at different stations in the Campbell River basin, BC, Canada (Validation period: 1991–2013).

Stations	Correlation coefficient from model generated precipitation series		
	BR	BRWS	KR
ELK	0.6999	0.6160	0.5697
ERC	0.6804	0.5984	0.4737
GLD	0.7086	0.6321	0.5389
HEB	0.6925	0.6301	0.5674
JHT	0.6312	0.5788	0.5282
QIN	0.6842	0.6058	0.5678
QSM	0.6299	0.5831	0.4685
SAM	0.6997	0.6155	0.6075
SCA	0.6858	0.6107	0.4480
WOL	0.6906	0.6131	0.3938

spatial correlation better than the KR. Artificial correlation has been added during the simulation by the conditioned rainfall states which can lead the BR model to overestimation of precipitation. Hence, the rainfall states should be used cautiously when the spatial correlation is of primary interest. The BRWS model fails to preserve the spatial correlation between data series.

4.1.5. Seasonal wet days characteristics

Changes in wet days precipitation may lead to extreme hydrological events such as floods and droughts. Investigation of wet days characteristics is important for water resource planning and management. Accurate reproduction of wet days is one of the important aspects of statistical downscaling. Although there are different criteria used in the literature to assess the wet days (WD), this work follows the definition of WD from Gaur and Simonovic (2013). According to Gaur and Simonovic (2013) if the amount of precipitation in a day is greater than 1 mm, then it will consider as a wet day. Fig. 8 represents 5th and 95th percentile estimates of downscaled monthly wet days for JHT station (considered as the only location to shorten the manuscript length). Fig. 8 (a) and (b) shows WD characteristics obtained from the simulated reanalysis monthly precipitation data where Fig. 8(c) and (d) shows WD characteristics of downscaled data obtained using the historical CanESM2 (1983–2005) monthly data. From Fig. 8 it can be observed that the BR model can generate values beyond extremes, but sometimes it underestimates extremes precipitation. This may be caused by scaling the response variable (precipitation) to (0, 1) interval.

4.1.6. Adding value to GCM (CanESM2) projections

Future climate change projections using GCMs simulation are very sensitive due to the existence of historical climate bias (Liang et al., 2008). If a GCM reasonably simulates present and historical climate then the credibility of future climate projection using the same GCM simulation will be higher. This can be possible if a downscaling model adds value to historical and present GCM climate variables. We conducted an experiment to explore whether BR model adds value to GCMs historical climate change following Racherla et al. (2012). The steps we followed for this experiment are given below:

Step-I: First we divided CanESM2 simulated historical precipitation data into two-time slices e.g. 1983–1994 (T1 hereafter) and 1995–2005 (T2 hereafter).

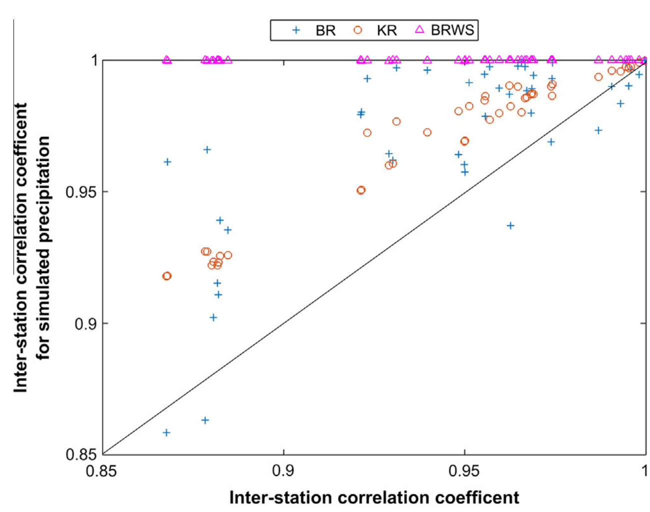

Fig. 7. Interstation correlation coefficients for different downscaling approaches.

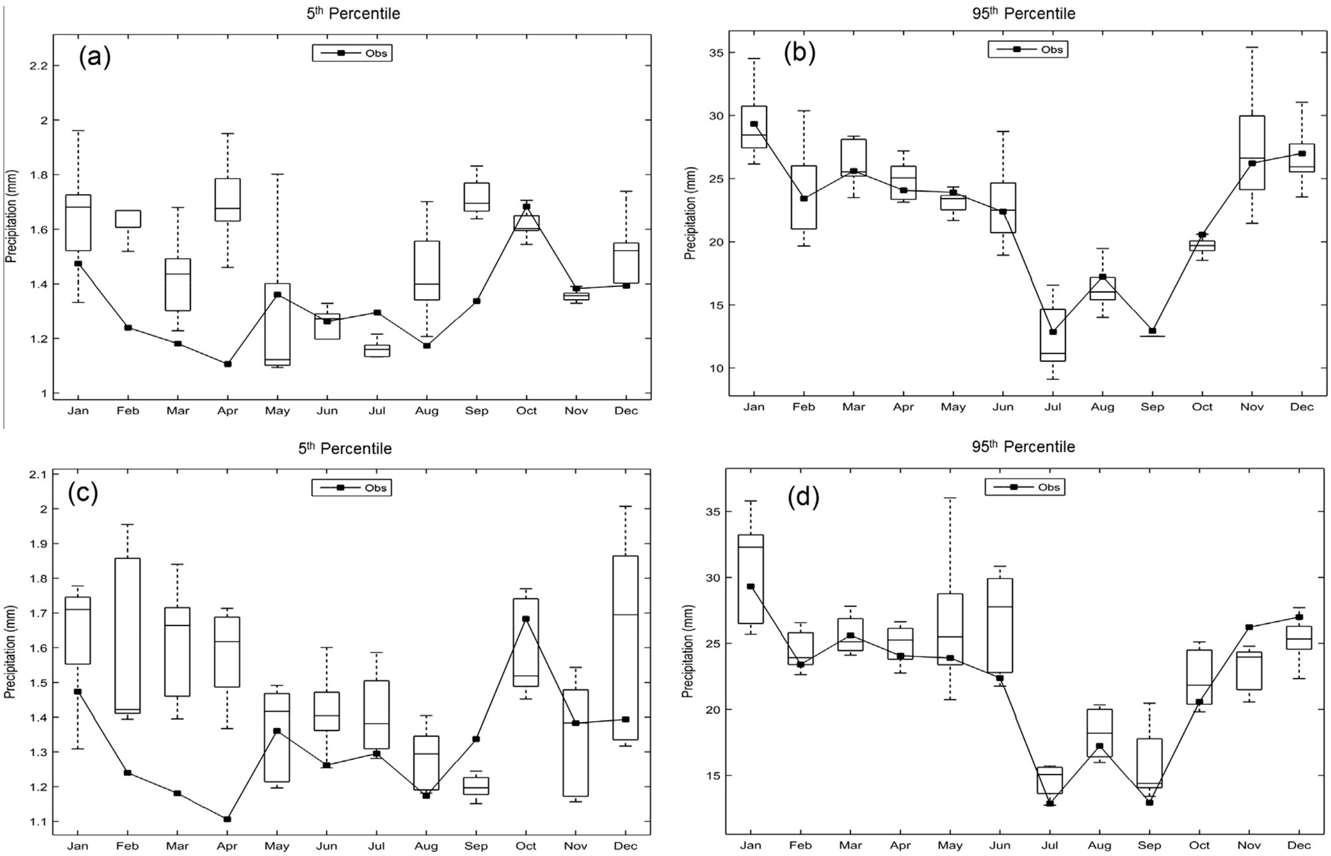

Fig. 8. Characteristics of monthly wet day extremes for observed and simulated precipitation at the JHT station. (a and b) Obtained from NCEP/NCAR (time period: 1991–2013). (c and d) Obtained from CanESM2 (time period: 1983–2005).

Step-II: ANUSPLIN and BR simulated precipitation data also divided into same time slices following step-I.

Step-III: Precipitation biases are calculated by subtracting daily historical precipitation data (ANUSPLIN) from CanESM2/BR simulated precipitation data. These biases are calculated for T1 and T2 time period and converted to seasonal mean precipitation biases shown in Supplementary Fig. 1 (SF1).

Step-IV Historical climate change (T2 minus T1) is calculated using above mentioned three data sets (ANUSPLIN, CanESM2 and BR simulated precipitation) and presented in Supplementary Fig. 2 (SF2).

It has been found that a wet bias is present in CanESM2 T1 time period (SF1 (a)) especially in fall and winter season where a dry bias has been found in T2 time period except spring season (SF1 (b)). The biases are reduced in BR simulated precipitation in both time periods except GLD station (SF1 (c) and (d)). Downscaled changes in seasonal precipitation between T1 and T2 time periods are presented in SF2. The most visible observed positive changes in precipitation have found in fall and winter, where a small decreasing trend found in summer time (SF2 (a)). Evidently observed changes are not reproduced well in CanESM2 except summer (SF2 (b)). However, BR model fairly reproduces observed changes except winter season (SF2 (c)). From this experiment, it can be concluded that GCMs historical bias can be reduced using BR model but not for all seasons and all stations. This limitation may be overcome if we consider different GCMs for the same experiment.

4.2. Future projection using GCM simulation

The BR model applied with standardized predictor data pertaining to RCP 2.6 RCP4.5 and RCP 8.5 scenarios of CanESM2 where RCP 2.6 represents low carbon emission scenario, RCP 4.5 referred as intermediate carbon emission scenario and RCP 8.5 is high emission scenario. To investigate the impact of future climate change on precipitation under different emission scenarios, a future time slice (2036–2065) is selected.

Fig. 9 represents CDF of daily precipitation at four downscaling locations. These four downscaling stations are selected based on their geographical location. JHT located near John hart dam where QIN and SCA are located near Strathcona dam. All of these three stations are located in downstream of the Campbell River basin where WOL is in upstream of the river.

The CDFs obtained for three scenarios are similar to each other and almost match with the CDF of observed precipitation (1991–2013). However, a downward shift pertaining to all three scenarios can be observed for JHT which indicates an increased frequency of high precipitation events during 2036–2065 compared to 1991–2013 (Fig. 9(a)). JHT station is located in downstream of the Campbell River and surrounded by forest. According to Sheil and Murdiyarso (2009) winds travel through forests can produce more than twice times precipitation compare to when they travel over the land which can be the reason of increased precipitation events at JHT. Although we only compared results obtained from a single GCM output using BR. More variation can be expected if the present analysis is performed with multiple GCMs (Werner, 2011).

4.2.1. Projected future seasonal precipitation changes during 2036–2065

Tables 8 and 9 provide information on estimated changes in number of wet days and seasonal precipitation amounts during 2036–2065. Percentage change in the median of any scenario is estimated with respect to observed data (1991–2013). For all three scenarios, summer precipitation amounts are going to decrease

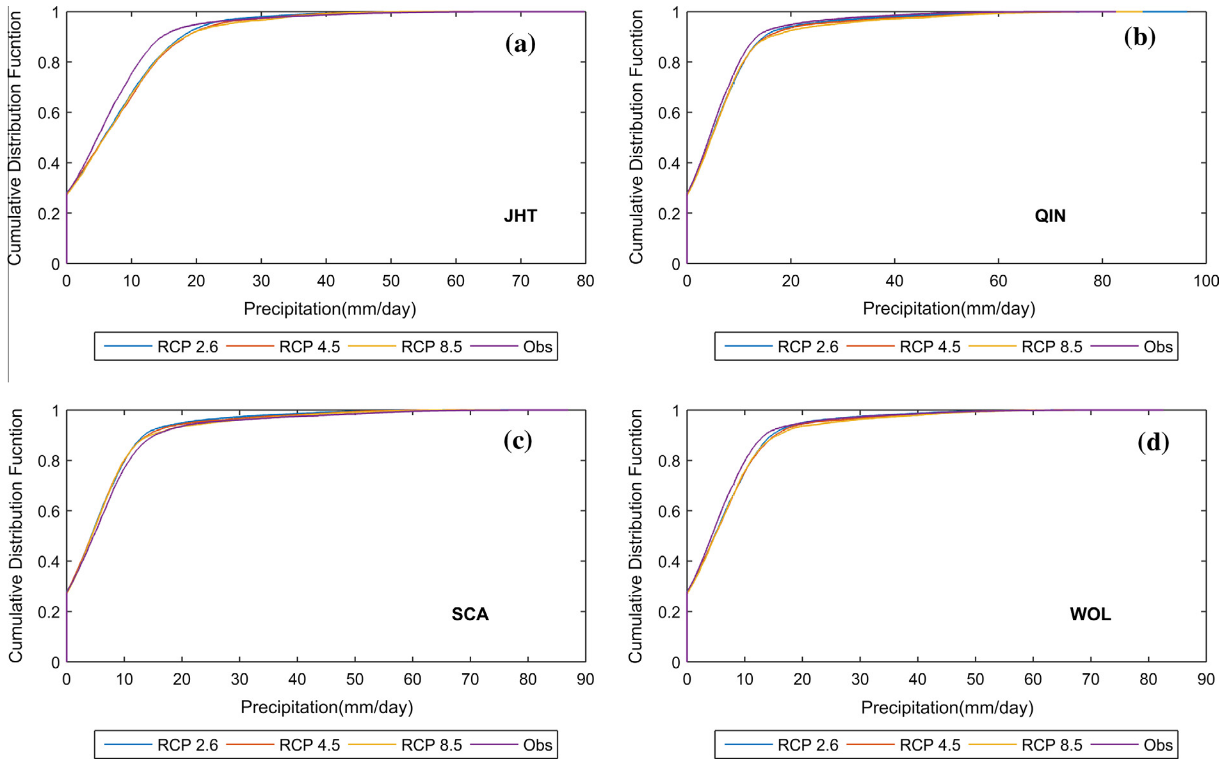

Fig. 9. CDF of simulated future (2036–2065) daily precipitation using CanESM2 predictor data under three emission scenarios (RCP 2.6, RCP 4.5 and RCP8.5) at different locations compare with observed precipitation (1991–2013).

Table 8
Seasonal changes in numbers of wet days during 2036–2065.

Season	Obs (1991–2013)	Scenario: RCP 2.6		RCP 4.5		RCP 8.5	
	Median estimate of number of wet days	Median estimate of number of wet days	Percentage change in median estimate	Median estimate of number of wet days	Percentage change in median estimate	Median estimate of number of wet days	Percentage change in median estimate
Winter	23	22	−4	26	13	26	13
Spring	19	24	26	23	21	24	26
Summer	16	15	−6	13	−18	15	−6
Fall	15	18	20	24	60	23	53

along with wet days. Maximum 58% can be decreased in summer precipitation amount where wet days can be reduced up to 18%. The changes detected for all three scenarios show precipitation amount in the fall is going to increase with wet days. For RCP 8.5, the increase is 22% in precipitation and 13% in wet days during winter time. For RCP 2.6 and RCP 4.5 both project a small amount of precipitation increase in spring time but the wet days increases 26% and 21% respectively.

5. Summary and conclusions

In this study, a new multisite statistical downscaling model is proposed for generating precipitation for a river basin using large scale climate variables conditioned to daily rainfall states. The proposed downscaling approach can reproduce spatiotemporal structure of the historical data at daily time scale, in addition to other statistics. The proposed downscaling method involves two main steps: (1) rainfall state generation using CART; and (2) generation of multisite precipitation amounts using multivariate BR model. To capture multicollinearity and reduce dimensionality we combine principal components analysis (PCA) with the BR. First five principal components are selected for this study which explains 97% variability of the original data.

CART constructs a classification tree based on the categorical and continuous predictors to generate precipitation state of the river basin. Lag-1 precipitation is used to prune the classification tree. The multisite precipitation sequences in the Campbell River basin (British Columbia, Canada) are generated using beta regression conditioned to precipitation states in the river basin. As BR model estimates mean precipitation values, perturbation method is added to the model for stochastic generation of precipitation outside the observed range following King et al. (2014). The model performs well in terms of preserving temporal and spatial dependence. Although BR overestimates spatial interstation cross-correlation.

Since there is no clear guidance for determining the optimal number of principal components, we considered a number of components which represent a large fraction of the variability (here 97%) contained in the original data. However, with availability of large data set obtained from the GCM simulated climate variables we may follow the step wise procedure described by Srivastav and

Table 9
Seasonal changes in precipitation amount during 2036–2065.

Season	Obs (1991–2013)	Scenario RCP 2.6		RCP 4.5		RCP 8.5	
	Median estimate of number of wet days	Median estimate of number of wet days	Percentage change in median estimate	Median estimate of number of wet days	Percentage change in median estimate	Median estimate of number of wet days	Percentage change in median estimate
Winter	253	260	2	289	13	311	22
Spring	409	425	3	437	6	485	18
Summer	263	116	−55	110	−58	108	−58
Fall	188	236	25	252	33	267	41

Simonovic (2014). BR method is a data driven method which builds a relationship between climate variables and daily precipitation. It is considered a stationary relationship among predictand and predictors variable (precipitation) which may not always be true. The basic relationships between climate variables controlled by conservation laws are not going to alter because of climate change. However, if the downscaling model is calibrated under stationary conditions and regional warming (e.g. El Nino-Southern Oscillation) influences the convective precipitation fraction then the stationary relationship in the downscaling process may indeed change. Salvi et al. (2015) observed that the kernel regression (KR) based statistical downscaling model failed to capture the changes in mean precipitation under non-stationary climate. They also identified that the assumption of stationarity was violated during the model testing period. It may be the reason for the changes in climate pattern occurring at large-scale or interference by some local factors e.g. urbanization. The urban areas have different climatology (Kishtawal et al., 2010; Shastri et al., 2015) and the effect of urbanization is not included in the BR model, therefore the same outcomes might be possible from BR if we test the BR model under non-stationary condition. Hence, identifying the exact reason of non-stationary behavior and validating the proposed model under non-stationary climate condition may be considered as a future scope of the present work.

Another important factor is link function in beta regression model. Several link functions are available such as logit, probit and log–log link. In this study we used only logit link function. Different outcomes may be expected if other link functions are used. In the present study, we used only one GCM output for downscaling. Future precipitation estimation may be different for use of other GCMs. Uncertainty modeling of downscaled precipitation from different GCMs is a potential research area under consideration.

The main advantage of using BR based downscaling model is multisite rainfall sequence generation which captures the temporal and spatial variability of the predictand at each downscaled location. The proposed model is computationally inexpensive and ideal for practical engineering application. It can use any number of predictor variables which may be considered the scope of future work.

Acknowledgement

This research work is supported by Discovery grant from Natural Sciences and Engineering Council (NSERC) of Canada to the third author.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhydrol.2016.04.009.

References

Bardossy, A., Duckstein, L., Bogardi, I., 1995. Fuzzy rule-based classification of atmospheric circulation patterns. Int. J. Climatol. 15, 1087–1097. http://dx.doi.org/10.1002/joc.3370151003.
BC Hydro Generation Resource Management, 2012. Campbell River System Water Use Plan. <http://www.bchydro.com/content/dam/hydro/medialib/internet/documents/planning_regulatory/wup/vancouver_island/2012q4/campbell_river_WUP_accept_2012_11_21.pdf>. 86p.
Brandsma, T., Buishand, T.A., 1998. Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. Hydrol. Earth Syst. Sci. 2 (2–3), 195–209. http://dx.doi.org/10.5194/hess-2-195-1998.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Brooks/Cole Publishing, Monterey.
Brock, G., Pihur, V., Datta, S., Datta, S., 2008. ClValid, an R package for cluster validation. J. Stat. Software 25 (4).
Chen, J., Brissette, F.P., Leconte, R., 2014. Assessing regression-based statistical approaches for downscaling precipitation over North America. Hydrol. Process. 28, 3482–3504. http://dx.doi.org/10.1002/hyp.9889.
Dibike, Y.B., Coulibaly, P., 2005. Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models. J. Hydrol. 307, 145–163. http://dx.doi.org/10.1016/j.jhydrol.2004.10.012.
Eum, H.-I.I., Simonovic, S.P., 2012. Assessment on variability of extreme climate events for the Upper Thames River basin in Canada. Hydrol. Process. 26, 485–499. http://dx.doi.org/10.1002/hyp.8145.
Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31, 799–815. http://dx.doi.org/10.1080/0266476042000214501.
Gaur, A., Simonovic, P.S., 2013. Climate change impact on flood hazard in the Grand river basin. Water Resources Research Report no. 084, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, London, Ontario, Canada, 92p. ISBN: (print) 978-0-7714-3063-3; (online) 978-0-7714-3064-0.
Ghosh, S., 2010. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. J. Geophys. Res. 115, D22102. http://dx.doi.org/10.1029/2009JD013548.
Goyal, M.K., Ojha, C.S.P., 2010. Evaluation of various linear regression methods for downscaling of mean monthly precipitation in arid Pichola watershed. Nat. Resour. 01, 11–18. http://dx.doi.org/10.4236/nr.2010.11002.
Groisman, P.Y., Karl, T.R., Easterling, D.R., Knight, R.W., Jamason, P.F., Hennessy, K.J., Suppiah, R., Page, C.M., Wibig, J., Fortuniak, K., Vyacheslav, R.N., Douglas, A., Førland, E., Zhai, P.M., 1999. Changes in the probability of heavy precipitation. Important indicators of climate change. Clim. Change 42, 243–283.
Gyalistras, D., von Storch, H., Fischlin, A., Beniston, M., 1994. Linking GCM-simulated climatic changes to ecosystem models: case studies of statistical downscaling in the Alps. Clim. Res. 4, 167–189.
Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2009. Statistical downscaling of precipitation: state-of-the-art and application of bayesian multi-model approach for uncertainty assessment. Hydrol. Earth Syst. Sci. 6, 6535–6579. http://dx.doi.org/10.5194/hessd-6-6535-2009.
Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stoch. Environ. Res. Risk Assess. 25, 475–484. http://dx.doi.org/10.1007/s00477-010-0416-x.
Hay, L.E., McCabe, G.J., Wolock, D.M., Ayers, M.A., 1991. Simulation of precipitation by weather type analysis. Water Resour. Res. 27 (4), 493–501. http://dx.doi.org/10.1029/90WR02650.
Hertig, E., Jacobeit, J., 2013. A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the mediterranean area. J. Geophys. Res. Atmos. 118, 520–533. http://dx.doi.org/10.1002/jgrd.50112.
Hughes, J.P., Guttorp, P., 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour. Res. 30 (5), 1535–1546. http://dx.doi.org/10.1029/93WR02983.
Hughes, J.P., Guttorp, P., Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. Appl. Stat. 48 (1), 15–30.
Hutchinson, M.F., Xu, T., 2013. Anusplin Version 4.4 User Guide. <http://fennerschool.anu.edu.au/files/anusplin44.pdf>. 52p.
IPCC, 2007. Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovermental

Panel on Climate Change. Cambridge University Press, Cambridge, UK & NY, USA, p. 996.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. http://dx.doi.org/10.1007/978-1-4614-7138-7.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kannan, S., Ghosh, S., 2013. A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin. Water Resour. Res. 49, 1360–1385. http://dx.doi.org/10.1002/wrcr.20118.

Kannan, S., Ghosh, S., 2010. Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stoch. Environ. Res. Risk Assess. 25, 457–474. http://dx.doi.org/10.1007/s00477-010-0415-y.

Khan, M.S., Coulibaly, P., Dibike, Y., 2006. Uncertainty analysis of statistical downscaling methods. J. Hydrol. 319, 357–382. http://dx.doi.org/10.1016/j.jhydrol.2005.06.035.

King, L.M., Mcleod, I.A., Simonovic, S.P., 2015. Improved weather generator algorithm for multisite simulation of precipitation and temperature. J. Am. Water Resour. Assoc. 7, 1–16. http://dx.doi.org/10.1111/1752-1688.12307.

King, L.M., Mcleod, A.I., Simonovic, S.P., 2014. Simulation of historical temperatures using a multi-site, multivariate block resampling algorithm with perturbation. Hydrol. Process. 28, 905–912. http://dx.doi.org/10.1002/hyp.9596.

Kishtawal, C.M., Niyogi, D., Tewari, M., Pielke, R.A., Shepherd, J.M., 2010. Urbanization signature in the observed heavy rainfall climatology over India. Int. J. Climatol. 30, 1908–1916. http://dx.doi.org/10.1002/joc.2044.

Lee, T., Ouarda, T.B.M.J., Jeong, C., 2012. Nonparametric multivariate weather generator and an extreme value theory for bandwidth selection. J. Hydrol. 452–453, 161–171. http://dx.doi.org/10.1016/j.jhydrol.2012.05.047.

Liang, X.Z., Kunkel, K.E., Meehl, G.A., Jones, R.G., Wang, J.X.L., 2008. Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophys. Res. Lett. 35, 1–5. http://dx.doi.org/10.1029/2007GL032849.

Martin, E., Timbal, B., Brun, E., 1997. Downscaling of general circulation model outputs: simulation of the snow climatology of the French Alps and sensitivity to climate change. Clim. Dyn. 13, 45–56. http://dx.doi.org/10.1007/s003820050152.

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resour. Res. 46, 1–17. http://dx.doi.org/10.1029/2009WR008423.

Mehrotra, R., Sharma, A., 2007. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J. Hydrol. 335, 180–193. http://dx.doi.org/10.1016/j.jhydrol.2006.11.011.

Mehrotra, R., Sharma, A., 2005. A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. J. Geophys. Res. 110, 1–13. http://dx.doi.org/10.1029/2004JD005677.

Mekis, É., Vincent, L.A., 2011. An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. Atmosphere–Ocean 49 (2), 163–177. http://dx.doi.org/10.1080/07055900.2011.583910.

Pour, S., Harun, S., Shahid, S., 2014. Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular Malaysia. Atmosphere (Basel) 5, 914–936. http://dx.doi.org/10.3390/atmos5040914.

Racherla, P.N., Shindell, D.T., Faluvegi, G.S., 2012. The added value to global model projections of climate change by dynamical downscaling: a case study over the continental U.S. using the GISS-ModelE2 and WRF models. J. Geophys. Res. Atmos. 117, D20118. http://dx.doi.org/10.1029/2012JD018091.

Raje, D., Mujumdar, P.P., 2009. A conditional random field-based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin. Water Resour. Res. 45, W10404. http://dx.doi.org/10.1029/2008WR007487.

Salvi, K., Kannan, S., Ghosh, S., 2013. High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment. J. Geophys. Res. Atmos. 118, 3557–3578. http://dx.doi.org/10.1002/jgrd.50280.

Salvi, K., Ghosh, S., Ganguly, A.R., 2015. Credibility of statistical downscaling under nonstationary climate. Clim. Dyn. 46, 1991–2023. http://dx.doi.org/10.1007/s00382-015-2688-9.

Shastri, H., Paul, S., Ghosh, S., Karmakar, S., 2015. Impacts of urbanization on Indian summer monsoon rainfall extremes. J. Geophys. Res. Atmos. 120, 495–516. http://dx.doi.org/10.1002/2014JD022061.

Schmid, M., Wickler, F., Maloney, K.O., Mitchell, R., Fenske, N., Mayr, A., 2013. Boosted beta regression. PLoS ONE 8 (4). http://dx.doi.org/10.1371/journal.pone.0061623.

Schoof, J.T., Pryor, S.C., 2001. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int. J. Climatol. 21, 773–790. http://dx.doi.org/10.1002/joc.655.

Sharif, M., Burn, D.H., 2006. Simulating climate change scenarios using an improved K-nearest neighbor model. J. Hydrol. 325, 179–196. http://dx.doi.org/10.1016/j.jhydrol.2005.10.015.

Sharma, A., O'Neill, R., 2002. A nonparametric approach for representing interannual dependence in monthly streamflow sequences. Water Resour. Res. 38 (7). http://dx.doi.org/10.1029/2001WR000953.

Sheil, D., Murdiyarso, D., 2009. How forests attract rain: an examination of a new hypothesis. Bio Sci. 59 (4), 341–347. http://dx.doi.org/10.1525/bio.2009.59.4.12.

Simas, A.B., Barreto-Souza, W., Rocha, A.V., 2010. Improved estimators for a general class of beta regression models. Comput. Stat. Data Anal. 54, 348–366. http://dx.doi.org/10.1016/j.csda.2009.08.017.

Srivastav, R.K., Simonovic, S.P., 2014. Multi-site, multivariate weather generator using maximum entropy bootstrap. Clim. Dyn. http://dx.doi.org/10.1007/s00382-014-2157.

Stern, R.D., Coe, R., 1984. A model fitting analysis of daily rainfall data. J. R. Stat. Soc. 147, 1–34.

Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J. Hydrol. 330, 621–640. http://dx.doi.org/10.1016/j.jhydrol.2006.04.030.

Von Storch, H., Zorita, E., Cubasch, U., 1993. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J. Clim. http://dx.doi.org/10.1175/1520-0442(1993)006<1161:DOGCCE>2.0.CO;2.

Vrac, M., Paillard, D., Naveau, P., 2007. Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. Clim. Past 3, 669–682. http://dx.doi.org/10.5194/cpd-3-899-2007.

Warren, F.J., Lemmen, D.S., 2014. Canada in a Changing Climate: Sector Perspectives on Impacts and Adaptation. Government of Canada, Ottawa, ON, p. p. 286.

Werner, A.T., 2011. BCSD Downscaled Transient Climate Projections for Eight Select GCMs over British Columbia, Canada. Hydrologic Modelling Project Final Report (Part I). p. 63.

World Climate Research Programme (WCRP), 2009. ETCCDI/CRD Climate Change Indices: Definitions of the 27 Core Indices. <http://etccdi.pacificclimate.org/list_27_indices.shtml>. (Accessed 4 September 2015).

Wetterhall, F., Halldin, S., Xu, C.Y., 2005. Statistical precipitation downscaling in central Sweden with the analogue method. J. Hydrol. 306, 174–190. http://dx.doi.org/10.1016/j.jhydrol.2004.09.008.

Wilby, R., Dawson, C., Barrow, E., 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. Environ. Model. Software 17, 147–159. http://dx.doi.org/10.1016/S1364-8152(01)00060-3.

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, pp. 1–27.

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. http://dx.doi.org/10.1029/2005WR004065.

Wilby, R.L., Hay, L.E., Leavesley, G.H., 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. J. Hydrol. 225, 67–91. http://dx.doi.org/10.1016/S0022-1694(99)00136-5.

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Progr. Phys. Geogr. 21 (4), 530–548. http://dx.doi.org/10.1177/030913339702100403.

Wilks, D.S., 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. Clim. Res. 11, 125–136. http://dx.doi.org/10.3354/cr011125.

Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic weather models. Progr. Phys. Geogr. 23 (3), 329–357. http://dx.doi.org/10.1191/030913399666525256.

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. Change 62, 189–216. http://dx.doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Yang, C., Chandler, R.E., Isham, V.S., Wheater, H.S., 2005. Spatial–temporal rainfall simulation using generalized linear models. Water Resour. Res. 41, 1–13. http://dx.doi.org/10.1029/2004WR003739.

Young, K.C., 1994. A multivariate chain model for simulating climatic parameters from daily data. J. Appl. Meteorol. http://dx.doi.org/10.1175/1520-0450(1994)033<0661:AMCMFS>2.0.CO;2.

Zwiers, F., Schnorbus, M.A., Maruszeczka, G.D., 2011. Hydrologic Impacts of Climate Change on BC Water Resources: Summary Report for the Campbell, Columbia and Peace River Watersheds, pp. 1–24.