CHAPTER 4

Introduction to Physical Scaling: A Model Aimed to Bridge the Gap Between Statistical and Dynamic Downscaling Approaches

Abhishek Gaur[1], Slobodan P. Simonovic[2]
[1]Construction Research Centre, National Research Council Canada, Ottawa, ON, Canada;
[2]Facility for Intelligent Decision Support, Western University, London, ON, Canada

Contents

Trends and Changes in Hydroclimatic Variables
ISBN 978-0-12-810985-4
https://doi.org/10.1016/B978-0-12-810985-4.00004-9

1. INTRODUCTION

In their pioneering work, Lowry (1977) identified "background climate," "effects of local landscape," and "effects of local urbanization" as three main components shaping climate at any location. Although first proposed 40 years ago, this fundamental and simple theoretical framework forms the basis of several detailed and complex climate models used currently like the global climate models (GCMs). GCMs simulate complex large-scale bio-geophysical and chemical processes, as well as their interactions with geophysical characteristics like land-use and topography to model global climate system. They are very useful mathematical tools that have been used extensively to study the impacts of historically observed and future projected greenhouse gas emissions on climate.

Very high computational demand is associated with GCM simulations even though they simulate climate at a very coarse spatial resolution of 3—5 degrees (~330—~550 km). To provide a rough estimate, at 3 degrees spatial resolution a GCM performs computations at 7200 grids across the globe each day. For simulating long-term climate projections for 250 years (as typically projections are made from 1850 to 2100), it will be performing the calculations ~700 million times to produce one set of climatic projections. Due to such high computational demand associated with GCM runs, climate simulations cannot be performed at very high resolutions, which will produce more realistic estimates of climate but will also be impractical. As such, at 3–5 degrees spatial resolution, GCM outputs are unsuitable for local and regional-scale risk assessment studies. Additionally, climatic effects of subgrid-scale geophysical features like land-cover and elevation are averaged out over a large area in the data associated with a GCM grid.

To estimate higher resolution climate projections from lower resolution GCM projections, a step referred to as downscaling is performed. Two broad classes of downscaling methods have been used in the past: statistical and dynamic. In the statistical downscaling methods, higher resolution climate data are estimated by developing statistical relationships between low resolution climatic or atmospheric data and locally observed data. The developed relationships are then used to estimate downscaled future climatic projections. A review of the state-of-the-art statistical downscaling techniques used in previous climate change impact assessment studies is provided in Section 2.2 of this chapter. In dynamic downscaling methods, local-scale climate is estimated by coupling a mesoscale higher resolution model or regional climate model (RCM) with a GCM. Boundary conditions generated from the GCMs are used as inputs into a RCM to estimate higher resolution climate taking into consideration biogeophysical interactions occurring at the land–atmosphere interface within the grid. Thus in dynamic downscaling methods, climate variability within a GCM grid is modeled taking into consideration geophysical variability present within a GCM grid-cell. On the other hand, in existing statistical downscaling methods same is modeled based on the location-specific differences in statistical relationships between low-resolution GCM and observed local climate.

The statistical relationships derived when performing statistical downscaling can be different for different locations within a GCM grid because of the differences in "background climate," "effects of local landscape," and "effects of local urbanization" as discussed in Lowry (1977). In previous studies, changes in "background climate" between different subgrid locations have been considered when downscaling GCM data. However, the contribution of land-cover and topography has largely been ignored. To model changes in statistical relationships between large-scale and local-scale climate at a location more accurately, land-cover and elevation characteristics of the location should be also considered when performing statistical downscaling of future GCM projections. Another important benefit of performing statistical downscaling, considering geophysical characteristics of a location, is that the developed relationships can then be used in geophysical environments that are drastically different from the present. Typically in statistical downscaling, relationships are derived for a location over the current time-period, and the relationships are used directly to model future projections at that location without considering possible future shifts in geophysical characteristics of the location which can change the nature of statistical relationships developed on the current time-period. When making decadal to century-scale predictions, considerable changes in geophysical characteristics at local to regional scales can be expected. It is therefore important to develop statistical downscaling methods that can model the effects of changes in geophysical characteristics in future climatic projections.

Physical scaling (SP) downscaling model is one of the first steps toward this direction. In this chapter we introduce an original model, perform validation of the model in current climate, compare the performance of the model with other state-of-the-art models, demonstrate an example application of the model, as well as explain R codes that can be used to apply the model at any location across the globe. The chapter is organized as follows. A review of literature related to SP model is provided in Section 2. This is followed by an introduction of SP method in Section 3. The advantages of SP method over other state-of-the-art statistical downscaling methods are presented in Section 4. Results from the validation of SP method are discussed in Section 5. A case study of SP method where it is used to quantify land-cover change–induced hydroclimatic changes in four catchments located in southern Saskatchewan region of Canada is presented in Section 6. The R code developed to formulate SP model and used to downscale future GCM projections is presented and discussed in Section 7. This is followed by a summary of conclusions and future work in Section 8.

2. LITERATURE REVIEW

In this section we provide theoretical background on concepts related to SP model. Most of the discussion provided in this section revolves around the downscaling of two important climate variables: precipitation and temperature, and how that can be performed more realistically using SP model.

First a brief introduction to the phenomenon of climate change and land-use land-cover (LULC) change is provided in Section 2.1. Then a review of statistical downscaling methods used in the past is presented in Section 2.2. This is followed by a discussion on the mechanisms and observation of the evidence of the effects of land-cover and elevation on local climate in Section 2.3. Finally a review of previous studies which have quantified the impact of future land-cover change on hydroclimatic variables is provided in Section 2.4.

2.1 Climate Change and LandUse Land-Cover Change

In this subsection, a brief introduction to climate change and its impacts on the past and future climates are presented. This is followed by a discussion of the historically observed and future projected LULC trends across the globe.

2.1.1 Climate Change

Climate change is referred to as the phenomenon of a persistent and detectable long-term change in climate conditions of a location. It is undeniable that the global climate has changed drastically over the past 60 years or so. Observational evidences of drastic shifts in climatic patterns have been observed across the globe. Global temperatures have been found to increase consistently with a mean increase of ~0.85°C rise noted over the period 1880–2012. Other hydroclimatic changes include an increase in global sea levels, decrease in snow-cover and sea-ice content, changes in precipitation and flow patterns, changes in circulation patterns, etc. (Stocker et al., 2013). Greenhouse gases emitted as a result of rapid human and industrial growth have been found to be the most prominent reason behind this change. Greenhouse gases emitted in the atmosphere hinder longwave radiations reflected by the earth's surface to travel back into the space thereby increasing temperatures near and around the earth's surface. Indeed, continual emission of greenhouse gases as a result of continuous industrial and human growth has resulted in considerable changes in the earth's radiation budget. It is estimated that earth is receiving more energy from the sun than it is releasing since at least about 1970 (Stocker et al., 2013). A change in the earth's radiation budget affects a range of atmospheric and climatic variables as evident in the observational findings discussed before.

Observed trends of increases in greenhouse gas concentration are expected to continue over the 21st century. To study the possible range of future emissions and their implications on climate, future greenhouse gas emission scenarios have been developed by different integrated assessment modeling (IAM) groups operating across the globe. Out of all the proposed future greenhouse gas trajectories, Van Vuuren et al. (2011) identified four representative concentration pathways (RCPs): RCP2.6, RCP4.5, RCP6, and RCP8.5 that correspond to an end of the 21st century radiative forcing of 2.6, 4.5, 6, and 8.5 W/m^2, respectively. A description of the four RCPs, their respective modeling groups, and related references are provided in Table 4.1. One or more of the selected RCPs are typically chosen to perform climate change impact assessment studies. In the Fifth Assessment Report (AR5) published by the Intergovernmental Panel for Climate Change (IPCC) (Stocker et al., 2013), future climate projections made by a range of GCMs, corresponding to the above specified four RCPs, have been presented and discussed.

Future climatic and hydrologic regimes are expected to be shaped by the greenhouse gas emission trajectory that is experienced. Temperature is expected to increase for all RCPs although the magnitude of change varies among different RCPs. The projected global mean increases in temperature between 1986–2005 and 2081–2100 time-periods as simulated by different GCMs are expected to be in the range of 0.3–1.7°C, 1.1–2.6°C, 1.4–3.1°C, and 2.6–4.8°C for RCP2.6, RCP4.5, RCP6, and RCP8.5,

Table 4.1 The Four Representative Concentration Pathways (RCPs) Chosen and Described in Van Vuuren et al. (2011)

RCP	Modelling Group	Description	Reference
RCP2.6	IMAGE	Rising radiative forcing pathway leading to 8.5 W/m^2 by 2100.	Van Vuuren et al. (2006, 2007)
RCP4.5	GCAM	Stabilization without overshoot pathway to 4.5 W/m^2 at stabilization after 2100.	Clarke et al. (2007), Smith and Wigley (2006) and Wise et al. (2009)
RCP6	AIM	Stabilization without overshoot pathway to 6 W/m^2 at stabilization after 2100.	Fujino et al. (2006) and Hijioka et al. (2008)
RCP8.5	MESSAGE	Peak in radiative forcing at ~3 W/m^2 before 2100 and then decline to 2.6 W/m^2 by 2100.	Riahi et al. (2007)

respectively (Collins et al., 2013). Future mean precipitation is expected to increase with rising temperatures. At a global scale mean precipitation is projected to increase at a rate of 1% $°C^{-1}$ to 3% $°C^{-1}$ for all RCPs except RCP2.6 in which case the rate of change in precipitation is associated with a considerable uncertainty among different GCMs (0.5% $°C^{-1}$ to 4% $°C^{-1}$). It should be noted that considerable regional and distributional variations are associated with the global temperature and precipitation changes outlined above (Stocker et al., 2013). Temperature and precipitation extremes are projected to change more drastically in the future than the means. For instance, it is projected that a one in 20-year maximum temperature event will at least become a one in 10-year event at most places across the globe; however, at many places it will become an annual or one in 2-year event. Similarly, for precipitation, a clear tendency for an increase in precipitation extremes with temperatures at a rate of roughly 5% $°C^{-1}$ to 10% $°C^{-1}$ is projected (Stocker et al., 2013; Collins et al., 2013).

2.1.2 LULC Change

Earth has experienced significant LULC change with human development and growth so much so that carbon emissions owing to land use change have been found comparable to that due to fossil fuel emissions over the period 1850–1990 (Houghton, 1999). Studies have found that during the last 300 years more than 50% of the land surface has been affected by land-use change. A rapid clearing of around 25% of the forested lands and a rapid growth of agricultural land to around 30% of the total land surface has also been observed (Turner et al., 1990; Hurtt et al., 2006).

With increasing population and agricultural and industrial growth, LULC change is projected to increase across the globe over the 21st century. The degree of change however is projected to vary depending on the manner in which future society growth occurs. Harmonized land-use data is a $0.5° \times 0.5°$ continuous annual estimate of land-use for the period 1500–2100 and is prepared by merging historical reconstructions of land-use with future estimates corresponding to different RCPs as provided by the modeling groups listed in Table 4.1. An analysis of harmonized land-use datasets produced by Hurtt et al. (2011) suggests that future LULC change is highly dependent on the emission scenario that future society follows while developing. An analysis of future projected land-use indicated significant change among all land-use classes. Most consistently the total area of secondary land (i.e., land which had been influenced by human activity in the past) is projected to increase under all RCPs. On the contrary, few land-use classes such as total agricultural land can change by +13% to −24% depending on the choice of RCP.

In summary, rapid changes in LULC have been observed in the past 300 years or so. It is expected that this trend will continue in the future; however, the magnitude of change will depend on the manner in which human development progresses in the future.

2.2 Statistical Downscaling

Statistical downscaling models aim to estimate regional to local-scale climate from low spatial resolution GCM climate data using statistical techniques. Here we provide an overview on some of the models that have been used in the past to perform statistical downscaling of GCM projections. A discussion on other relevant aspects of statistical downscaling such as inherent assumptions and commonly used predictors is also presented.

2.2.1 Assumptions

All statistical downscaling methods must adhere to following assumptions (Hewitson and Crane, 1996; Giorgi et al., 2001; Wilby et al., 2004; Benestad et al., 2008). Firstly, there must be a strong relationship between the predictor and predictant variables chosen in the method. In statistical terms, the predictor variable should have high degree of covariance and similar time structure to that of the predictant variable for it to be selected for prediction. Secondly, chosen predictor variables must be adequately simulated by the GCMs. This is very important because if important physical and statistical aspects of the predictor variable are not accurately simulated by the GCMs, then downscaling will only result in a more precise estimate of the erroneous predictor variables. Third assumption is that chosen predictors must be able to represent the climate change signal in the predictant variable adequately. For instance, sea level pressure is considered a very good predictor for observed temperature. However, if it is used alone to predict temperatures it can lead to the prediction of unrealistically low temperature estimates (Benestad et al., 2008). Therefore when performing downscaling, it is important to identify a set of predictor variables which can model the predictant variable with acceptable accuracy. Lastly and most importantly, the relationships derived are assumed to be time-invariant, that is, the relationships do not change over time. The validity of this assumption is very critical to obtaining reliable future estimates of climate in drastically changed future environments. Although this assumption cannot be tested explicitly, several studies have tested the robustness of the statistical relationships by calibrating and validating the methods on contrasting climatic conditions. For instance, Wilks (1999) tested the robustness of a statistical downscaling model towards downscaling precipitation by calibrating the method on wet days and validating it on dry days and vice versa. These tests do not guarantee superior method skill in future climates but models that perform well in such can be considered preferable for making downscaled future climatic projections than the ones that do not perform well.

2.2.2 Common Predictors

The predictor variables considered for downscaling in the past have been guided by the availability of long time-series of atmospheric variables in the calibration and prediction timelines. Atmospheric data for calibration is generally extracted from the reanalysis products and for the prediction period is obtained from the GCMs. The most relevant set of predictors are chosen by accessing the relationship between the predictor variables and observed climate data. Predictors that are found to be consistently correlated with the observed climate are selected for prediction.

The choice of variables has also been found to depend on the temporal resolution of the predictant variable. For instance, monthly mean predictant variables have been found to be linked to upper level atmospheric and circulation variables (like geopotential height and sea level pressure) while daily predictant variables, which are more variable and skewed, are likely also dependent on variables linked to horizontal flux and convergence of moisture (like specific humidity, winds, and vorticity) (Cavazos and Hewitson, 2005; Schoof, 2012). Another important factor that can influence the choice of predictor variables in statistical downscaling is the spatial, temporal, and distributional scale at which the relationship between predictors and observed climate are accessed. Gaur and Simonovic (2015) identified this as one of the three prominent sources of uncertainty in climate change impact assessment studies. They demonstrated that GCM-based climatic and atmospheric variables differ considerably with the spatial, temporal, and distributional scale of analysis. Masson and Knutti (2011) and Räisänen and Ylhäisi (2011) have discussed the variability of GCM skill over different spatial scales and have suggested an optimal spatial scale at which GCMs perform best. Evaluation of predictors at different spatial, temporal, and distributional scales is therefore very important and should be considered when selecting optimal predictor variables for statistical downscaling of GCM projections.

2.2.3 Statistical Downscaling Models

Statistical downscaling methods used in the past can be classified as being either PerfectProg or model output statistics (MOS) models (Fowler et al., 2007; Schoof, 2013). In PerfectProg models, the relationship between predictor and predictant variables are derived by using observed or reanalysis based predictors. On the other hand, in MOS techniques GCM data is used directly for downscaling. The raw GCM data is sometimes bias-corrected (Teutschbein and Seibert, 2012) before it is used for downscaling in the MOS models. The PerfectProg models can be further classified into three categories of models: (1) regression-based, (2) weather pattern-based, and (3) weather generator-based methods (Wilby and Wigley, 1997).

Several MOS methods have been used in the past to estimate higher resolution downscaled climatic projections from raw GCM projections. The most popular method under this category is the bias correction and spatial disaggregation (BCSD) technique (Salathé, 2003; Wood et al., 2004). In this method the raw GCM data are first bias-corrected with reference to gridded observed climate data. The bias-corrected data is then used to obtain local-scale spatial anomaly pattern which is then used to modify future bias-corrected data. The method has been used extensively to downscale GCM projections in many climate change impact assessment studies (Salathé, 2005; Widmann et al., 2003; Sharma et al., 2007 among others).

Among the PerfectProg models, regression-based statistical downscaling models are perhaps the most popular. In regression-based downscaling methods a regression relationship is developed between predictor and predictant variables. The predictor and predictant variables have previously been linked using functions such as standard ordinary least squares (Wilby et al., 2002), quantile regression (Friederichs and Hense, 2007), partial least squares regression (Bergant and Kajfez-Bogataj, 2005), generalized linear modeling (Fealy and Sweeney, 2007), etc. In some studies, techniques such as singular value decomposition and canonical correspondence analysis (CCA) have also been used to identify relationships between the predictor fields (Hertig and Jacobeit, 2008; Huth, 2002). Studies have also employed artificial neural networks (ANN) and machine learning algorithms such as genetic programming (Coulibaly, 2004), support vector machines (Tripathi et al., 2006), and relevance vector machines (Ghosh and Mujumdar, 2008) to link predictor and predictant variables.

One of the most popular regression-based downscaling techniques is statistical downScaling model (SDSM). Downscaling by SDSM method

(Wilby et al., 2002) is performed by analyzing a range of atmospheric variables available from the reanalysis and GCM databases. In SDSM a set of 25 candidate predictor variables have been suggested for initial screening (Wilby et al., 2002). Out of the candidate predictor variables, a set of most relevant atmospheric predictor variables which are highly correlated with locally observed climate variable of interest are chosen for prediction. This step is referred to as the initial screening process. Multiple linear regression relationship between the selected large-scale atmospheric variables (predictors) and locally observed climate (predictant) data is then formulated over the calibration period. The calibrated regression relationship is thereafter used to predict climate variable of interest over the validation period. Another regression-based precipitation downscaling model is generalized linear modeling (GLM)—based model. In GLM model, first relevant set of atmospheric predictors are identified for prediction using an approach similar to that used in the SDSM model. Precipitation occurrence is modeled from the chosen predictor variables using a logistic regression model and precipitation magnitudes on wet days are modeled using generalized additive model (GAM). The GLM model has been used extensively to perform precipitation downscaling across the globe (Fealy and Sweeney, 2007; Kigobe et al., 2011; Beecham et al., 2014; Lee et al., 2011; Manzanas et al., 2015).

Weather pattern—based downscaling methods are based on the principle that observed climatic variations at the local scale are largely shaped by synoptic climatic patterns. The most common method under this category is referred to as the analog method (Zorita and von Storch, 1999). In this method historical records are searched for synoptic patterns matching the GCMs. The observed climate corresponding to those patterns is then considered as the downscaled GCM data. Several studies have employed analog methods to downscale GCM data (Timbal and Jones, 2008; Timbal et al., 2008; Li and Smith, 2009; Chiew et al., 2010; Frost et al., 2011; Hope et al., 2010). Another popular method in this category is the nonhomogeneous hidden Markov model (Hughes and Guttorp, 1994), which has been widely used to downscale GCM precipitation data (Robertson et al., 2004, 2007; Bellone et al., 2000; Hughes et al., 1999; Samuels et al., 2009; Liu et al., 2011; Kwon et al., 2011). In this method precipitation occurrence probabilities and wet -day precipitation amounts at a location are linked to the classes of synoptic scale atmospheric fields such as geopotential height. The applicability of weather pattern—based downscaling methods is highly dependent on the availability of long observational climatic and atmospheric records so that patterns found in the GCM data can be associated with an observational synoptic scale pattern.

Weather generators are stochastic models that can be used to produce long synthetic replicates of observed climate time-series. A wide range of parametric, nonparametric, and semiparametric weather generators have been used in climate change impact assessment studies in the past (Wilks, 2012, 2010; Dibike and Coulibaly, 2005; Charles et al., 2007; Kwon et al., 2011; Eum and Simonovic, 2012; Hashmi et al., 2011; Richardson, 1981; Soltani and Hoogenboom, 2003; Kuchar, 2004; Craigmile and Guttorp, 2011; Apipattanavis et al., 2007; Semenov and Barrow, 1997; Sharif and Burn, 2006, 2007; Lee et al., 2012; King et al., 2012, 2016; Srivastav and Simonovic, 2015). Downscaling by weather generator technique is performed by first obtaining scaled historical climate data, which is regarded as equivalent to future projections without internal climate variability component. Scaled historical data has been obtained using several approaches in the past. One of the approaches is by estimating change factors using historical and future GCM modeled climate data (Anandhi et al., 2011; Gaur and Simonovic, 2015). Change factors can be calculated for various temporal scales (daily, monthly, seasonal, yearly), can have different mathematical formulations (additive, multiplicative), and can be associated with different quantiles of the distribution (mean, median, or other quantiles). The scaled historical climate is then used as input into a weather generator to obtain future climate projections with internal climate variability. This is done by generating several runs of historical scaled data such that key statistical properties of each run is similar to the input climatic time-series with some perturbation component which incorporates internal climate variability component to the input time-series. In some studies, weather generator parameters have been modified in place of using change factors to include the climate change signal in the weather generator downscaled outputs (Schoof et al., 2007, 2010).

2.3 Statistical Versus Dynamic Downscaling of GCM Projections

Both statistical and dynamic downscaling approaches have strengths and limitations. Statistical downscaling methods present a way to downscale GCM projections quickly and with minimal computational requirements. For this reason they are desirable for uncertainty assessment in future climate projections. They are also particularly useful in cases where location-specific downscaled projections are desired. Dynamic downscaling approaches can provide downscaled projections at a finer resolution grid-scale than the GCMs (typically at 10–20 km spatial resolution), but location-specific projections still need to be extracted from them using some spatial interpolation technique. This can be problematic in regions that have complex, varying topography or land-cover within a downscaled grid-cell. Statistical downscaling approaches are also very useful for real-time applications. They have been found to produce excellent results especially in regions where observed data are abundant and are considered very suitable for downscaling near-term climate predictions since the relationships derived in the historical timeline are considered to be time-invariant in statistical downscaling. This assumption is practically valid in near-term (5–10 years) projections. Finally, only large-scale changes in climate can be accounted for when performing statistical downscaling as regional geophysical characteristics are not considered in these models.

On the other hand, physically based dynamic downscaling methods are suitable for both near-term and long-term predictions. In other words, they can model climate in drastically changed climatic and geophysical environments. They are particularly useful in modeling climate projections in regions where observed data is limited. Dynamic downscaling methods are computationally expensive and use boundary conditions that are extracted from the GCMs. Therefore, bias associated with the boundary conditions is also transferred to the dynamically downscaled climate projections. Statistical downscaling models, on the other hand, offer an opportunity to bias-correct GCM data before they are downscaled.

As evident from above discussion, both statistical and dynamic downscaling methods have areas of strengths and weaknesses. An ideal downscaling model would combine the strengths of both these categories of methods and minimize the weaknesses.

2.4 Effects of Land-Cover and Elevation on Climate

Both land-cover and elevation are known to influence climate both from observational studies and physical sciences. Some of these concepts and observational evidences have been summarized in this section.

2.4.1 Land-Cover and Climate

Regional and local-scale climate, which are most relevant for catchment-scale hydrologic assessment, are shaped not only by large-scale atmospheric processes but also by its physical characteristics like land-cover and topography (Lowry, 1977; Oke, 1982). Observational studies have found detectable influence of both land-cover (Pielke et al., 2006) and elevation (Pepin et al., 2015), which are also supported by the laws of atmospheric and physical sciences.

The influence of land-cover on temperature has been particularly well explored in the case of urban areas. Urban areas are characterized by low albedo and heat absorbing surfaces because of which their near-surface heat transfer is led by an enhanced sensible heat flux and reduced latent heat flux as compared to the nonurban areas. Additionally, urban areas are unable to dissipate heat absorbed during the daytime primarily due to a reduced sky-view factor in the canopy layer. This phenomenon is well documented in the scientific literature and is commonly termed as the urban heat island (UHI) effect (Voogt and Oke, 2003; Oke, 1982). UHI effect has been recorded for many cities across the globe including Colombo, Sri Lanka (Emmanuel and Johansson, 2006), San Juan, Puerto Rico (Velazquez-Lozada et al., 2006), Singapore (Wong and Yu, 2005), Ouagadougou, Burkina Faso (Offerle et al., 2005), Gaborone, Botswana (Jonsson, 2004), Phoenix, Arizona (Fast et al., 2005), Addis Ababa, Ethiopia (Abebe and Magento, 2016), Toronto, Canada (Wang et al., 2015), Noida, India (Kikon et al., 2016), among others.

The effect of land-cover on temperature has also been reported for other land-cover classes. For example, Ge (2010) found agricultural regions to be 2.3°C cooler than the surrounding grassland areas in the growing season during the daytime and 1.61°C warmer than the surroundings after the harvest. Roy et al. (2007) investigated the effect of massive agricultural green revolution on temperature in India using both monthly observational climatic data and regional climate models. They found that during the growing season agriculture can considerably reduce temperatures in the surrounding areas. Boucher et al. (2004) performed a global-scale study to investigate the impact of agricultural growth on atmospheric water vapor and climate. They found a cooling of up to ~0.8K over irrigated areas together with an increase in global mean radiative forcing in the same period due to an enhanced water vapor concentration due to irrigation. McPherson et al. (2004) investigated the impact of winter wheat crops on climate over Oklahoma and found considerable cooling in the growing season and warming after the crops have been harvested. Similar cooling of irrigated areas of Great Plains over nonirrigated areas is also found by Mahmood et al. (2006). Another study has noted a local cooling effect as a result of intensive horticulture due to increase in surface albedo in south-eastern Spain (Campra et al., 2008). Studies have also found decrease in temperatures due to afforestation/deforestation, which results in changes in the surface albedo, leaf area index, roughness lengths, and rooting depths (Sellers, 1992; Jackson et al., 1996; Pitman, 2003; Fahey and Jackson, 1997; Xue and Shukla, 1996; Nosetto et al., 2005). Li et al. (2017) evaluated contribution of historical land-cover change to changes in extreme temperature over Eurasia. They compared temperature simulations obtained from National Centre for Atmospheric Research's Community Atmosphere Model Version 5.0 (NCAR CAM5.0) (Neale et al., 2012) forced with land-covers corresponding to years 1850 and 2000. Results indicated that historically over Eurasia land-cover change has led to changes in extreme indices of both minimum and maximum temperatures; however, the magnitude of change varies significantly among different regions. Other observational studies further corroborate to a relationship between regional land-cover and temperatures (Qiao et al., 2014; Fall et al., 2010a,b; Hale et al., 2006, 2008; Lawrence and Chase, 2010).

Several observational studies have also noted the influence of land-cover on precipitation, especially for urban areas. Li et al. (2011) studied urban signature in strong and weak precipitation events over the Pearl River Delta (PRD) in China using tropical rainfall measuring mission (TRMM) satellite precipitation data. They found that over and around the urban regions "strong" precipitation events have increased with urbanization while "weak" precipitation events have decreased. They also found strengthening of the precipitation intensity, a decrease in rainfall frequency, and an increase in convective rainfall and afternoon precipitation events over and around the urban areas. Similar findings were reported by De and Rao (2004). They analyzed rainfall trends of several Indian megacities (with population more than 1 million) such as New Delhi, Kolkata, Mumbai, and Chennai between 1901 and 2000 and found statistically significant increasing trends in annual and monsoon precipitations. A decreasing trend was also found for a few cities. They found more pronounced increases in precipitation during the period 1951–2000 when rapid industrial development took place over the selected urban locations. Rao et al. (2004) performed a similar analysis on precipitation trends for the duration 1901–2000 and found similar statistically significant increasing trends for the cities analyzed. Kishtawal et al. (2010) analyzed mean and extreme rainfall trends of urban locations within India using observed as well as remotely sensed TRMM precipitation data and identified an increasing trend linked to the pace of urbanization of the cities. Further urban locations were found to have more possibility of witnessing an extreme precipitation event than the surrounding nonurban area. Several other studies have also found evidences of land-cover linkage with rainfall (Kug and Ahn, 2013; Schlüzen et al., 2010; Halfon et al., 2009; Fujibe et al., 2009; Mote et al., 2007; Diem and Mote, 2005; Inoue and Kimura, 2004; Dixon and Mote, 2003; Shepherd et al., 2002). There are three hypotheses as to how urban areas can impact regional precipitation distribution: (1) by modifying the thermodynamic processes such as energy balance and urban heat island induced circulation within and around the city, (2) by causing winds to converge over and downwind of the cities due to roughness of the city elements, and (3) by effecting cloud microphysical processes due to the presence of large amounts of aerosol in the urban air (Han et al., 2014).

The influence of agricultural development has also been found to influence rainfall. Segal et al. (1998) adopted a modelling approach for investigating this and found an increase in precipitation with increase in irrigated agricultural areas over North America. An increase in agricultural areas in the Indian subcontinent has been found to decrease the monsoon rainfall (Lee et al., 2009; Niyogi et al., 2010). This is considered to happen as a result of reduced temperatures over the agricultural areas due to irrigation. This in turn diminishes the land–sea temperature contrast in agricultural areas, which is one of the driving factors for monsoon rainfall in India. A detectable effect of afforestation/reforestation on precipitation has been found and has been found to be dependent on geographical location, regional atmospheric characteristics, and the extent of afforested-reforested area (Xue and Shukla, 1996; Pitman and Narisma, 2005; Pielke et al., 2006 and references therein).

It has also been found that different sections within a city can have different climates based on the geophysical characteristics of their neighborhoods. For example, urban green areas such as parks are generally cooler than their surrounding built-up areas, and can produce temperature differences of up to 7°C, a phenomenon termed as "park cool island" (Oke et al., 1989; Barradas, 1991; Spronken-Smith and Oke, 1998; Jansson et al., 2007; Chang et al., 2007). The cooling effect of these vegetated areas has also been found to extend into the surrounding urban areas as well (Jauregui, 1991; Chen and Wong, 2006; Wong and Yu, 2005; Jonsson, 2004). For instance, after analyzing temperatures at seven gauging stations in Curitiba, Brazil, Kruger and Givoni (2007) found significant thermal effects of neighboring land-cover cells on the temperature recorded at those stations. Sun (2011) analyzed temperatures for two streets in Taichung city, Taiwan, and found them to be significantly correlated with the surrounding green ratio and building ratio in the night. Yokobori and Ohta (2009) too found temperatures in the Tokyo metropolitan area, Japan, to decrease with the total amount of vegetated area present in the neighborhood areas. Yan et al. (2014) found the air temperature in Beijing, China, to be significantly affected by the manmade and tree cover surface composition of the city. They also analyzed this relationship at different spatial scales (from 20 to 300 m in radius) and found that the relationships varied with the spatial scale considered.

Above studies provide evidence that local-scale climate is not only affected by the land-cover properties of the location of interest but also by the land-cover composition of the areas surrounding the location. Further, the effect of neighborhood land-cover composition is also dependent on the spatial scale at which those relationships are analyzed.

2.4.2 Elevation and Climate

The relationship between elevation and temperature is well understood in meteorology. The rate of decrease of temperature with elevation is referred to as the lapse rate. An approximate lapse-rate value of 6.5°C per 1000 m in the troposphere is considered as standard in many meteorological applications (Lydolph, 1985). The relationship between elevation and precipitation is more uncertain than the temperature with some studies finding consistent relationships between the two variables while some studies finding variable relationships. For example, Puvaneswaran and Smithson (1991) found both increasing and decreasing trends while analyzing precipitation-elevation relationships across Sri Lanka and termed the relationship to be complex. On the other hand using geographically weighted regression (GWR), Brunsdon et al. (2001) found a definite relationship between elevation and precipitation over the Great Britain. They highlighted the importance of considering GWR while studying these relationships as they vary in space.

Sensitivity of different regions to climate change has also been found to be elevation-dependent. In general, higher elevation regions are considered to be more susceptible to climate change than the lower elevation regions. Yan and Liu (2014), for instance, found that higher elevation areas (>2000 masl) in the Tibetan Plateau show a higher rate of warming than other lower elevation regions surrounding it. Giorgi et al. (1996) performed modeling experiments using a regional climate model and compared present-day and doubled CO_2 experiment modeled climates over the Alpine region. They obtained more pronounced warming over higher elevation regions than the lower elevation regions from their experiments.

Li et al. (2012) analyzed the variation of temperature and precipitation extremes across southwestern China over the period 1961–2008 and found significant links between elevation and changes in diurnal temperature range, frost days, ice days, cold night frequency and cold day frequency, consecutive dry days, consecutive wet days, wet-day precipitation, and the number of heavy precipitation days in the region during this period. Historical and future projected trends in minimum temperature were analyzed with reference to elevation in the Tibetan plateau by Liu et al. (2009) and greater vulnerability of higher elevation regions over the lower elevation regions was demonstrated. A more detailed review of the observational evidence and plausible operating mechanisms that lead to these elevation-dependent responses to greenhouse gases is provided in Pepin et al. (2015). Climate modification brought due to snow-albedo feedback, more frequent cloud cover, and water vapor–related radiative feedbacks are considered as possible mechanisms for a higher warming rate in the higher elevation regions. It is also recognized that the conclusions made for elevation-dependent changes in climate variables are not well understood because of less data availability at higher elevation regions (Pepin et al., 2015).

The literature review presented above provides observational and physical evidence that land-cover and elevation and their distribution across a region influences its climate. Under rapidly changing projected future geophysical conditions (Hurtt et al., 2011) it is therefore incessant to consider changes in geophysical characteristics when making future climate projections.

2.5 Previous Assessments of Future Land-Cover Change Impact on Climate

The impact of future projected land-cover change on climate has been studied using GCMs at regional to global scales. The typical methodology has been to compare climate projections obtained from the GCMs with and without considering future land-cover change. For example, Hua et al. (2015) analyzed future temperature projections obtained with and without considering LULC change from a GCM: CanESM2 under two RCPs: RCP2.6 and RCP8.5 and found small magnitudes of change at a global scale; however, changes of upto 0.1°C were obtained at regional scales. Quesada et al. (2017) investigated the impact of future LULC change on monsoonal precipitation by analyzing future projections under RCP8.5

from five earth system models (ESMs) and found significant changes in four (out of eight) of those regions. A unique methodology to identify land-cover change–induced climatic changes is proposed by Kumar et al. (2013). They compared climatic projections made for two neighboring sites: one which has projected to experience land-cover change in future and another which is projected to remain unchanged in future. From their analysis using 14 GCMs under RCP8.5, they found substantial land-cover change–induced summer warming in parts of North America and Eurasia.

Studies have also evaluated land-cover change effects on climate at local scales. Malyshev et al. (2015) analyzed the effect of local-scale land-cover heterogeneity in the geophysical fluid dynamics laboratory (GFDL) ESM and found significant effects on climate. Georgescu et al. (2013) simulated mid- and end-of-century temperature for the Arizona city using the weather research and forecast (WRF) modeling system. They found that the projections are highly sensitive to the scenario-based land-cover trajectory. Both mid- and end-century temperature estimates were found to be strongly dependent on the built environment and future emission pathways in the Sun Corridor expansion projections. Chen and Frauenfeld (2015) downscaled future temperature and precipitation projections made by Community ESM under RCP4.5 using WRF modeling system over China. They obtained significant effects of future-projected urbanization on both temperature and precipitation across China. Argueso et al. (2014) used the WRF modeling system to downscale CSIRO MK3.5 GCM outputs to 2 Km grid-scale. They simulated the present (1990–2009) and future (2040–59) climates for Sydney area and concluded that coupling of future urbanization effects and climate change will significantly affect the local climatology. They projected more intense changes in minimum temperatures than in maximum temperatures, particularly in winter and spring season when urban effects will contribute almost equally toward temperature change as the changes in global emissions. Kusaka et al. (2012) used dynamically downscaled projections from three GCMs MIROC3.2-medres, MRI-CGCM2.3.2a, and CSIRO-Mk3.0 climate models to access future (2070s) summertime (August) temperature for three largest urban settlements in Japan: Tokyo, Osaka, and Nagoya. They used WRF model to get high resolution temperature distributions in the cities and found that the city temperatures will increase by 2–3°C in future. Also the impact from UHI effect was found to be comparable to that from greenhouse gas emissions. Hamdi et al. (2014) used a high resolution limited area model ARPEGE-IFS coupled with Town Energy Balance model to estimate changes in UHI magnitudes as projected by the ARPEGE-Climat GCM in the timeslice 2071–2100. The model was run at 4 km spatial resolution for SRES A1B scenario and it was found from the inline run (where TEB was activated for regional as well as urban runs) that the intensity of daytime UHI will decrease by 0.2–0.24°C. Also, strong UHI events were projected to decrease in the future by 1°C. McCarthy et al. (2012) used the latest version of the Hadley Centre Regional Climate model HadRM3 at 25-km resolution coupled to a simple urban land-surface scheme (Best et al., 2006) to assess the sensitivity of UK urban climates to large-scale greenhouse gas–induced climate change, local forcing from urban land use, and anthropogenic heat flux resulting from energy use in the urban areas. Adachi et al. (2012) also calculated future UHI intensities for Tokyo city by using five future projections from climate models downscaled using the TERC-RAMS regional model. After comparing the results obtained with and without incorporating urban effects, they concluded that the temperature change between 1990s and 2070s because of greenhouse gas emissions is projected to be around 2°C while due to land-cover changes is projected to be around 0.5°C. Several other studies have also evaluated local-scale changes in climate due to future projected urbanization (Zong-Ci et al., 2013; Zhao et al., 2013; Cao et al., 2016; Bounoua et al., 2015; Zhang et al., 2016).

2.6 Conclusions From Literature Review

In this section we first presented theoretical background about climate change and land-cover change, and discussed how they have considerably affected earth's climate in the past and are expected to continue doing so in the future. From the discussion provided, it is clear that earth has experienced extensive climate and land-cover over the 20th century and this trend is expected to continue in the 21st century. An introduction to statistical downscaling with a discussion on inherent assumptions, criteria for the selection of predictors, and a review of the state-of-the-art statistical downscaling models used in the past is presented next. From the discussion it can be concluded that a variety of PerfectProg and MOS models have been used in the past to statistically downscale GCM projections. They are built on a set of statistical assumptions. The choice of predictor variables have primarily been guided by the availability of atmospheric and climatic data with the climate models although other minor factors also guide the predictor selection. A summary of the strengths and weaknesses of statistical and dynamic downscaling models is provided next. The discussion provides evidence that both dynamic and statistical downscaling models can be improved in order to make the downscaling process both realistic and practical. A detailed discussion on the observational evidence of linkage between climate and geophysical characteristics of a region is presented next. The discussion provides ample evidence of that land-cover and elevation effects of temperature and precipitation patterns from local to regional scales. Finally, previous studies that have projected future land-cover change–driven climatic changes have been reviewed. It is seen that all such studies have adopted a dynamic downscaling approach for doing so.

The discussion provided in this section clearly highlights that under the backdrop of rapidly changing climatic and geophysical characteristics, the need to develop a new statistical downscaling model that can consider geophysical characteristics of a region like land-cover and elevation when performing downscaling is particularly high. The new model should be fast and robust like traditional statistical downscaling models but should also account for geophysical characteristics of a region as done in dynamic downscaling. The physical scaling model, which is envisioned to have these qualities, is presented in the following section.

3. PHYSICAL SCALING MODEL OF DOWNSCALING

It has been highlighted in previous sections that statistical downscaling methods used currently do not explicitly account for geophysical characteristics of the region of study. This is one of the major drawbacks of statistical downscaling because of which geophysical changes in a region cannot be modeled using statistical downscaling methods. In SP method (Gaur and Simonovic, 2016a,b) this is achieved by including additional geophysical covariates representing land-cover and elevation distribution of the region. The predictant (observed local-scale climate) and predictor (large-scale climate and geophysical covariate) variables are linked using a generalized additive model (GAM) regression relationship. In GAM regression, the predictant variable is connected to smoothed predictor variables using a link function. The smoothing is generally performed using nonparametric algorithms. The regression function is totally nonparametric in nature and hence is suitable for modeling a range of climate variables including precipitation and temperature. A more detailed description of SP model and its extensions is provided below.

3.1 SP Model

SP model approach to downscaling temperature (surface or air) can be mathematically expressed as:

$$g(T_{obs}) = B_0 + f_1(T_{\mathrm{mod}}) + f_2(E_p) + f_3(LC_p) \tag{4.1}$$

where T denotes temperature, E denotes elevation (masl) of the reference pixel, LC denotes categorical land-cover variable associated with the reference pixel, and B denotes regression parameters. Subscript *obs* and mod denote if the climatic data is observed or model-based, respectively. T_{mod} denotes large scale "background" climate data obtained by bilinear interpolation of GCM temperature data at the reference pixel. Subscript p indicates that the data used is a pixel-scale data. Variables g and f represent the link and smoothing functions, respectively. In this study smooth functions are fit using penalized likelihood maximization algorithm. The penalized likelihood maximization algorithm is a variant of maximum likelihood estimation algorithm and applies a tradeoff between model fit wiggliness and goodness of fit by incorporating a penalty function (Wood, 2000).

In the case of precipitation, method involves a two-step process of predicting precipitation occurrence using a logistic regression model and a wet-day precipitation amounts model using a GAM regression model. SP method approach to downscale precipitation (Gaur and Simonovic, 2018) can be mathematically expressed as:

$$\ln\left(\frac{P_{obs}}{1 - P_{obs}}\right) = B_0 + B_1 P_{\mathrm{mod}} + B_2 E_p + B_3 LC_p \tag{4.2}$$

$$g(P_{obs,wet}) = B_0 + f_1(P_{\mathrm{mod},wet}) + f_2(E_p) + f_3(LC_p) \tag{4.3}$$

where notations have previously defined meanings. Additionally, subscript *wet* denotes values on wet days only (i.e., days with more than 0.1 mm of precipitation).

3.2 SP With Surrounding Pixel Information (SPS) Model

SPS method is an extension to SP method where land-cover and elevation properties of the reference, as well as neighborhood, pixels are incorporated into the method formulation (Gaur and Simonovic, 2016b). SPS method for downscaling air temperature can be mathematically expressed as:

$$\begin{aligned} g(T_{obs}) = B_0 + f_1(T_{\mathrm{mod}}) + f_2(E_p) + f_3(LC_p) + \\ f_4(Fr_{W,s}) + \cdots + f_{17}(Fr_{BSV,s}) + f_{18}(R_{E,s}) \end{aligned} \tag{4.4}$$

where, symbols have similar meanings as explained above. Predictors $Fr_{W,s}$, $\ldots Fr_{BSV,s}$ represent the fraction of total area surrounding the reference pixel that is occupied by *Water, …Barren and Sparsely Vegetated* land-cover classes, respectively. The value of predictors: $Fr_{W,s}$, $\ldots Fr_{BSV,s}$ is between 0 and 1 and they add up across all neighborhood land-cover classes to give a value of 1. Neighborhood elevation information is incorporated by including a predictor $R_{E,s}$, which represents the ratio between reference pixel elevation and mean elevation of pixels surrounding the reference pixel at a specific neighborhood scale, s.

For precipitation, SPS downscaling method involves two steps of forming a precipitation occurrence and wet-day precipitation amounts model (Gaur and Simonovic, 2018). The two steps involved in SPS method can be mathematically expressed as:

$$\ln\left(\frac{P_{obs}}{1-P_{obs}}\right) = B_0 + B_1 P_{\mathrm{mod}} + B_2 E_p + B_3 LC_p + B_4 Fr_{W,s} + \cdots + B_{17} Fr_{BSV,s} + B_{18} R_{E,s} \tag{4.5}$$

$$g(P_{obs,wet}) = B_0 + f_1(P_{\mathrm{mod},wet}) + f_2(E_p) + f_3(LC_p) + f_4(Fr_{W,s}) + \ldots. + f_{17}(Fr_{BSV,s}) + f_{18}(R_{E,s}) \tag{4.6}$$

where, symbols have same meanings as explained before.

The choice of spatial scale within which neighborhood geophysical characteristics are analyzed is critical to SPS method application and performance. Four neighborhood scales (represented as s in Eqs. 4.4–4.6): 3x3, 5x5, 7x7, and 9x9 have been chosen before to define neighborhood characteristics (Gaur and Simonovic, 2016b). The selected neighborhood scales are shown in Fig. 4.1 in darker shades of grey for neighborhood scales: 33, 5x5, 7x7, and 9x9, respectively, while the reference pixel is shown in black. Larger neighborhood scales are considered to be inclusive of smaller spatial scales, for instance, neighborhood scale 5x5 encompasses pixels corresponding to neighborhood scale 3x3 and additional darker shaded pixels specific to 5x5 neighborhood scale.

4. ADVANTAGES OF SP AND SPS METHODS OVER OTHER TRADITIONAL STATISTICAL DOWNSCALING MODELS

SP and SPS methods described in this chapter have following advantages over the traditional downscaling methods:

- They can account for geophysical characteristics of the region of interest when used to perform downscaling. As explained before, elevation and land-cover properties of the location of interest, as well as its neighborhoods, can be included into the SP and SPS method definition.

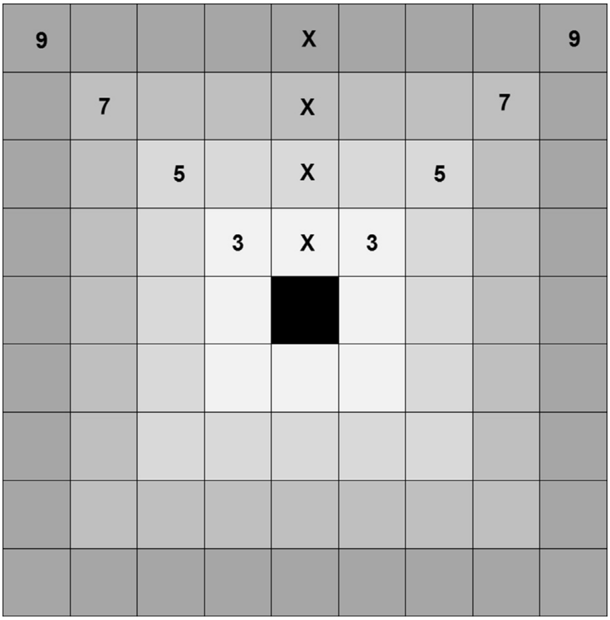

Figure 4.1 Neighborhood scales considered in this study. Increasing neighborhood scales are shown in progressively darker shades of grey. The reference pixel is shown in black.

The inclusion of additional geophysical predictors in the method can produce more accurate results in regions where geophysical characteristics play an important role in shaping local climate.

- The method also provides an opportunity to investigate climatic implications of historically observed and/or future projected geophysical changes for any region. Such analysis is possible with the SP and SPS methods because they include geophysical characteristics of the region of interest into the method definition.
- The method builds on a delicately balanced compromise between statistical and dynamic downscaling approaches and attempts to build on the positives of both downscaling approaches. It provides local-scale climate estimates consistent with the geophysical characteristics of a region and is able to do it using minimal computational and time resources.
- The method provides an opportunity to explore uncertainty in future projected climatic changes under projected geophysical changes. Since the method is fast and computationally inexpensive, it can be used in conjunction with data from multiple GCMs and RCPs to access the uncertainty associated with such changes.

- The method can also be used in real-time applications for downscaling GCM projections, as well as to model climatic effects of changes in geophysical characteristics of a region. Preparation of web-tools (Neset et al., 2016) and serious gaming tools (Savic et al., 2016) for instructional, educational, and entertainment purposes can be one such real-time application of the method.
- The method has been developed and implemented using publicly available datasets (Friedl et al., 2010; Wan, 2006; Mesinger et al., 2006; Taylor et al., 2012) and an open source programming language R (R Development Core Team, 2008). So it can be implemented conveniently and free of charge to obtain downscaled GCM projections for any region across the globe.

5. VALIDATION OF SP AND SPS MODELS

In this section the analysis results and conclusions obtained from extensive assessment of SP and SPS models towards downscaling North American regional reanalysis (NARR) surface temperature, air temperature, and precipitation are presented. A comparison of the performance of SP and SPS models with other state-of-the-art statistical downscaling models is also presented and discussed.

5.1 Study Region

The validation of SP and SPS models is performed at the southern Saskatchewan region of Canada. The physiographic setting of the study region is shown in Fig. 4.2. The study region is representative of the Canadian prairies and is characterized by diverse topography and land-cover. Two major urban centers located in this region are: Saskatoon and Regina. They are shown as black dots in Fig. 4.2. It experiences a continental climate. Large fluctuations in temperature are observed owing to its landlocked position. Significant spatial variability in precipitation is also observed. The region receives almost two-thirds of its precipitation during the summer season, which usually occurs due to large-scale convective and cyclonic systems. Snow cover plays a critical role in shaping the hydrometeorology of this region as it stays snow-covered almost 6 months a year.

5.2 Data Used

List of data used for performing model validation is provided below. Data sources and other specifications have also been summarized.

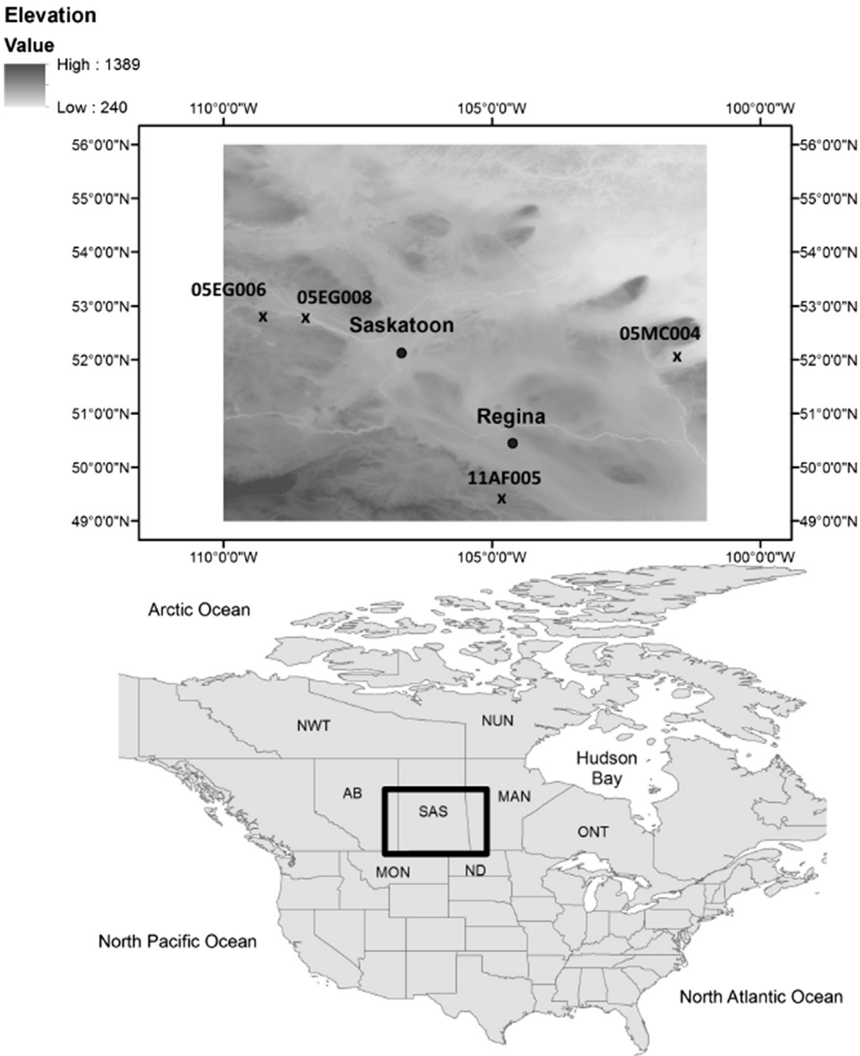

Figure 4.2 The location of southern Saskatchewan region, which is chosen as the study region in Gaur and Simonovic (2016a,b, 2018).

5.2.1 Remotely Sensed Surface Temperature

MODIS surface temperature level 3 Terra (MOD11A1) and Aqua (MYD11A1) products are used in this study (Wan, 2006). Both day-time: Terra-Day (TD), Aqua-Day (AD) and night-time: Terra-Night (TN), Aqua-Night (AN) data are collected for the period 2006–13. These data are available at an approximate spatial resolution of 1 × 1 km. Quality assessment files of the data are assessed and only reliable data are selected for analysis. It was found that a higher percentage of reliable data is available for snow-free months (April to October) than for the snow-covered months (November to March). The reason behind this can be that more cloud-free conditions occur during the summers than in the winters facilitating the sensing of reliable surface temperature data. The distribution of data is also found uneven within a day with higher percentage of data available in the night-time than in the daytime. Overall the data is found to be evenly distributed across the study period.

5.2.2 Remotely Sensed Land-Cover

MODIS recorded level 3 annual land-cover product (MCD12Q1) in University of Maryland (UMD) classification scheme (Friedl et al., 2010) is used in this study. In UMD classification scheme, land-cover is classified into 14 different classes. These land-cover classes as well as the classification codes used to refer them in this study are listed in Table 4.2. The land-cover dataset is available at a 500 m spatial resolution. Land-cover for the period 2006–12 is considered.

5.2.3 Remotely Sensed Elevation

The National Aeronautics and Space Administration (NASA) Shuttle Radar Topographic Mission (SRTM) elevation product is used. This data has a spatial resolution of 90 m.

Table 4.2 Land-Cover Classes as Identified in the UMD Classification System

Classification Code	UMD Classes
W	Water
ENF	Evergreen Needleleaf Forest
EBF	Evergreen Broadleaf Forest
DNF	Deciduous Needleleaf Forest
DBF	Deciduous Broadleaf Forest
MF	Mixed Forest
CS	Closed Shrublands
OS	Open Shrublands
WS	Woody Savannas
S	Savannas
G	Grasslands
C	Croplands
UB	Urban and Built-up
BSV	Barren or Sparsely Vegetated

Classification code used in downscaling models to represent each land-cover class is also provided.

5.2.4 Reanalysis Surface Temperature

North American Regional Reanalysis (NARR) 3-hourly surface (skin) temperature data for the period 2006–13 are used for downscaling. Daily air temperature and precipitation rate data are also used for downscaling. These data are produced by running the NCEP Eta model together with the Regional Data Assimilation System (Mesinger et al., 2006) and have an approximate spatial resolution of 32 km.

5.2.5 Gauged Daily Air Temperature and Precipitation Data

Daily precipitation data gauged at 52 locations within the study region over the period 2006–13 is acquired from the Environment Canada. The data can be accessed at: http://climate.weather.gc.ca/. The list of gauging stations selected for study is provided in Table 4.3. Using MODIS land-cover data it is found that these gauging stations are associated with UMD land-cover classes: S, OS, G, DNF, UB and C.

5.3 Surface Temperature Downscaling

SP model was evaluated for its performance in downscaling NARR surface temperature data over the period 2006–13 in Gaur and Simonovic (2016a) by adopting a cross-validation approach. The model was formulated for each month and for each timeline: AD, TD, AN, and TN using data from reliable MODIS surface temperature and land-cover pixels distributed across the study region. To adjudge if a particular MODIS pixel was reliable, its quality assessment files were accessed and pixels associated with an error of magnitude <1 K were adjudged reliable for model calibration. Similarly, land-cover pixels that were specified as "good quality" in land-cover quality assessment files were chosen for analysis.

SP model considers observed (or remotely sensed in this case) temperature data as predictant variable and elevation, land-cover, and model-derived climate as predictor variables. A GAM model with these predictor and predictant variables was formulated. The method was thereafter calibrated on randomly chosen 90% of the data and validated on the rest 10% of the data. A total of 100 such calibration and validation sets were generated for each month and each timeline giving a total of 4800 (100 sets $\times$ 4 timelines $\times$ 12 months) calibration and validation sets for model evaluation.

Model performance was evaluated for each timeline—AD, TD, AN, and TN—and model performance was found satisfactory in all timelines. The average root mean squared error (RMSE) across all models was found to be 0.05, 0.02, 0.02, and 0.03 K for timelines AD, TD, AN, and TN,

Table 4.3 List of Stations With Data Considered in This Study

S.No.	Station Name	Latitude	Longitude	Elevation
1	Rosetown East	51.57	−107.92	586.00
2	Last Mountain Cs	51.42	−105.25	497.00
3	Bratt's Lake Climate	50.20	−104.71	580.00
4	Wynyard (Aut)	51.77	−104.20	560.10
5	Nipawin	53.33	−104.00	371.90
6	Assiniboia Airport	49.73	−105.95	725.50
7	Hudson Bay(Aut)	52.82	−102.32	358.10
8	Pilger	52.42	−105.15	552.00
9	Prince Albert A	53.22	−105.67	428.20
10	Outlook Pfra	51.48	−107.05	541.00
11	North Battleford	52.77	−108.25	548.00
12	Coronach Spc	49.05	−105.48	756.00
13	Watrous East	51.67	−105.40	525.60
14	Melfort	52.82	−104.60	490.00
15	Elbow Cs	51.13	−106.58	595.00
16	Kindersley A	51.52	−109.18	693.70
17	Meadow Lake A	54.13	−108.52	480.70
18	North Battleford Rcs	52.77	−108.26	548.00
19	Yorkton	51.26	−102.46	498.40
20	Eastend Cypress (Aut)	49.44	−108.99	1059.00
21	Spiritwood West	53.37	−107.55	584.30
22	Yorkton	51.26	−102.46	498.30
23	La Ronge A	55.15	−105.27	379.20
24	Regina Int'l A	50.43	−104.67	577.60
25	Regina Rcs	50.43	−104.67	577.30
26	Saskatoon Intl A	52.17	−106.70	504.10
27	Val Marie Southeast	49.06	−107.59	796.00
28	Broadview	50.37	−102.57	599.80
29	Estevan	49.22	−102.97	580.60
30	Estevan A	49.22	−102.97	580.30
31	Indian Head Cda	50.55	−103.65	579.10
32	Loon Lake Rcs	54.02	−109.14	545.60
33	Lucky Lake	50.95	−107.15	664.70
34	Meadow Lake	54.13	−108.52	481.00
35	Moose Jaw A	50.33	−105.57	576.70
36	Moose Jaw Cs	50.33	−105.54	577.00
37	Nipawin	53.33	−104.02	371.90
38	Nipawin	53.33	−104.01	371.90
39	North Battleford	52.77	−108.24	548.30
40	Rockglen (Aut)	49.17	−105.98	917.00
41	Saskatoon Rcs	52.17	−106.72	504.10

Continued

Table 4.3 List of Stations With Data Considered in This Study—cont'd

S.No.	Station Name	Latitude	Longitude	Elevation
42	Scott Cda	52.36	−108.83	659.60
43	Swift Current	50.29	−107.69	816.90
44	Swift Current A	50.30	−107.68	816.90
45	Swift Current Cda	50.27	−107.73	825.00
46	Weyburn	49.70	−103.80	588.60
47	Yorkton	51.26	−102.46	498.30
48	Cypress Hills Park	49.64	−109.51	1259.00
49	Jimmy Lake Awos	54.91	−109.96	637.10
50	Leader Airport	50.91	−109.50	675.50
51	Mankota	49.10	−107.02	830.00
52	Maple Creek	49.90	−109.47	766.70

respectively. Additionally, superior model performances were obtained in snow-free months than in the snow-covered months, and in the night-time than in the daytime. Further the method was found to perform better in the lower elevation and flat areas of the study region than in areas with complex topography. Based on the obtained results, the method was considered suitable for the downscaling of GCM surface temperature projections across the southern Saskatchewan region.

5.4 Air Temperature Downscaling

SP and SPS models were assessed for their skill in downscaling NARR air temperature data in Gaur and Simonovic (2016b). The assessment was performed over the period 2006–12. A large ensemble of SP and SPS models were evaluated to identify most appropriate functional form, approach, and spatial scale for the application of SP and SPS models.

Two approaches, direct and indirect, for downscaling NARR air temperatures were tested. In the direct approach, NARR air temperature data were used directly to obtain local air temperature whereas in the indirect approach, NARR surface temperature are first used to estimate downscaled surface temperature, which are then used to predict downscaled air temperature using a model that estimates air temperature from surface temperature (referred to as ST→AT model hereafter). Three different functional forms were used to link predictor and predictant variables. The total ensemble of models evaluated in Gaur and Simonovic (2016b) are summarized in Table 4.4. The method, functional form used to link predictor and predictant variables, and predictors used in the ST → AT model, for each model are also summarized.

Table 4.4 Models Evaluated for Air Temperature Downscaling Skill in Gaur and Simonovic (2016b)

S.No.	Approach	Model	Method	Functional Form	Predictors (ST → AT Model)
1	Direct	SP-lm	SP	LR	–
2		SP-qr	SP	QR	–
3		SP-gam	SP	GAM	–
4		SPS3x3-lm	SPS	LR	–
5		SPS5x5-lm	SPS	LR	–
6		SPS7x7-lm	SPS	LR	–
7		SPS9x9-lm	SPS	LR	–
8	Indirect	SP-lm-ST	SP	LR	ST
9		SP-qr-ST	SP	QR	ST
10		SP-gam-ST	SP	GAM	ST
11		SP-lm-ST-LC	SP	LR	ST, LC
12		SP-qr-ST-LC	SP	QR	ST, LC
13		SP-gam-ST-LC	SP	GAM	ST, LC
14		SP-lm-ST-LC-AVs	SP	LR	ST, LC, AVs
15		SP-qr-ST-LC-AVs	SP	QR	ST, LC, AVs
16		SP-gam-ST-LC-AVs	SP	GAM	ST, LC, AVs
17		SPS3x3-lm-ST-LC-AVs	SPS	LR	ST, LC, AVs
18		SPS5x5-lm-ST-LC-AVs	SPS	LR	ST, LC, AVs
19		SPS7x7-lm-ST-LC-AVs	SPS	LR	ST, LC, AVs
20		SPS9x9-lm-ST-LC-AVs	SPS	LR	ST, LC, AVs

The downscaling skills of models listed in Table 4.4 were also compared with the skill of a state-of-the-art statistical downscaling method: SDSM. The performances of SDSM and SP models toward downscaling NARR air temperature data were accessed by using leave-one-out cross-validation technique (Stone, 1974) over 52 gauging stations listed in Table 4.4. In this technique, validation was performed at a reference gauging station while calibration was performed on the remaining gauging stations. This process was repeated until validation was performed at all gauging stations located within the study region. Downscaling methods were evaluated based on two metrics: root mean squared error (RMSE) and Pearson correlation coefficient (r). Additionally, Akaike Information Criteria (AIC) was used to compare the quality of SP methods. AIC statistic (provided in Eq. 4.7) quantifies a regression model quality by meriting it based on the goodness of fit of the regression relationship and demeriting it based on its increased complexity.

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{4.7}$$

where AIC represents AIC statistic value, k represents the number of estimated parameters in a model, and L represents the maximum value of likelihood function for the model.

Downscaling of NARR temperature data by SDSM method was performed by first screening large-scale atmospheric variables for prediction. In Gaur and Simonovic (2016b) the atmospheric variables listed in Table 4.5 were considered for the initial screening process. The choice of atmospheric variables was made taking into account the recommended large-scale atmospheric variables in Wilby et al. (2002) and the availability of atmospheric data in the NARR product. Next, monthly correlations between large-scale atmospheric variables and temperature across all temperature recording stations are examined for each timeline. Results indicated that the correlation between large-scale atmospheric variables and observed

Table 4.5 NARR-Derived Atmospheric Variables Considered for Screening in SDSM Method

S.No.	Short-Name	Long-Name
1	dlwrf	Downward Longwave Radiation Flux
2	dswrf	Downward Shortwave Radiation Flux
3	uswrf	Upward Shortwave Radiation Flux
4	ulwrf	Upward Longwave Radiation Flux
5	cape	Convective Available Potential Energy
6	tke	Total Kinetic Energy
7	hgt500hpa	Geopotential Height at 500 hpa
8	hgt850hpa	Geopotential Height at 850 hpa
9	hgt1000hpa	Geopotential Height at 1000 hpa
10	shum500hpa	Specific Humidity at 500 hpa
11	shum850hpa	Specific Humidity at 850 hpa
12	shum1000hpa	Specific Humidity at 1000 hpa
13	hcdc	High Cloud Area Fraction
14	mcdc	Medium Cloud Area Fraction
15	lcdc	Low Cloud Area Fraction
16	prate	Precipitation Rate
17	prmsl	Mean Sea Level Pressure
18	rhum	Relative Humidity
19	wnd	Wind Speed

temperature depends on the atmospheric variable, month of the year, and time of the day being analyzed. Based on the obtained correlation values and personal judgment, three atmospheric variables, shum1000hpa, hgt850hpa and ulwrf, were selected to model local temperature using SDSM model.

Before performing downscaling by SP methods (listed in Table 4.3), they were tested for the presence of multicollinearity in predictor variables. Variance inflation factors (VIFs) associated with different predictor variables were calculated and their values are found within an agreeable range of 1–2.5 indicating acceptable multicollinearity among the predictor variables (Rogerson, 2001; Pan and Jackson, 2008). From leave-one-out cross-validation of SDSM and SP methods, at more than 50% of all gauging stations the performances of SP methods considered under direct approach were found superior than the SDSM method in terms of both RMSE and correlation. On the other hand, the performance of SP methods considered under indirect approach was found inferior to SDSM method at more than 50% of the stations. A summary of RMSE values obtained for different models assessed is presented in Fig. 4.3. It can be noted from the figure that the lowest RMSE values are associated with direct SP downscaling methods, followed by SDSM, and followed by indirect SP methods.

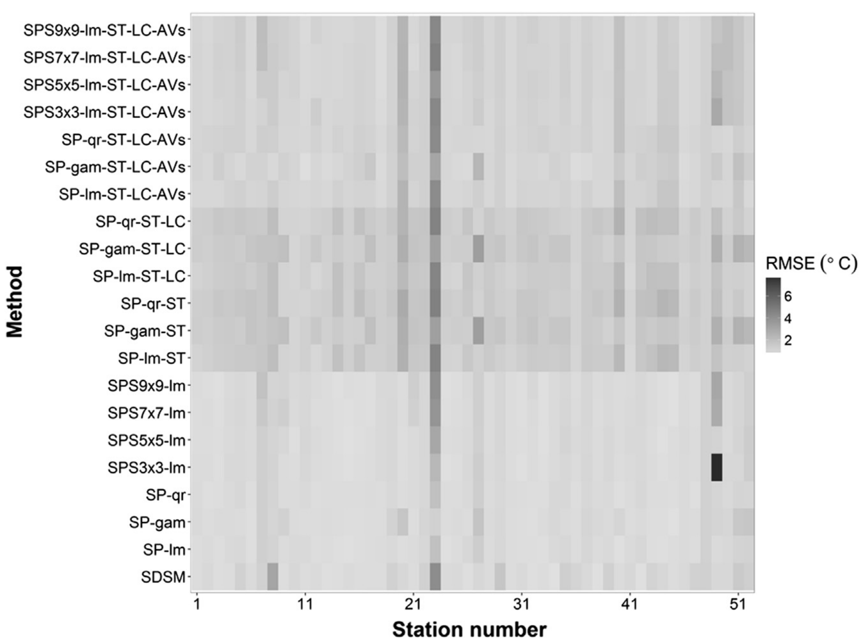

Figure 4.3 RMSE obtained from leave-one-out cross-validation procedure on the 52 stations on which SP and SPS methods are evaluated for air temperature downscaling skill.

Based on the average RMSE across all gauging stations, methods considered under direct SP approach were found to show almost similar performances (RMSE $\sim$ 0.8°C). SDSM method was obtained as the next best performing method (RMSE = 0.9°C). The performances of indirect methods were found to be inferior to both direct SP and SDSM methods. The performance of methods considering atmospheric variables as predictors in the ST → AT method were found to be similar (RMSE $\sim$ 1.1°C), as well as superior to the methods that do not account for them (RMSE $\sim$ 1.5°C). Similar method performances were also observed in terms of correlation.

Relative performance of methods at each gauging station was also accessed. Three best performing methods based on RMSE and correlation were identified at each gauging station and overall composition of first, second and third ranked methods across all gauging stations were analyzed. Among SDSM and SP methods, SP-gam was found to be the most consistent first and second ranked method while SP-lm was obtained as the most consistent third ranked method. The results obtained in Gaur and Simonovic (2016b) suggest that direct SP methods tend to outperform SDSM method while the latter outperforms indirect SP methods in terms of the RMSE and correlation.

The variation of SDSM and SP method performances across different months was also analyzed. In general, superior model performances are obtained in the summer months (April–September) than in the winter months (October–March). This is expected in the case of SP method because the quality of remotely sensed data is known to be higher in cloud and snow-free conditions which usually occur during the summer months.

The qualities of SP methods considered under direct and indirect approaches were further compared in terms of AIC. Data across all gauging stations was used to estimate AIC associated with each SP and SPS model. Among the models considered under direct approach, method SPS5x5-lm was obtained to be of the highest quality while SP-qr was obtained as the lowest quality model. Among the models considered under the indirect approach, method SP-gam-ST-LC-AVs was found to be of the highest quality while SP-qr-ST was obtained to be of the lowest quality. The qualities of SPS models were found to be superior to SP methods in both direct and indirect approaches with method SPS5x5-lm demonstrating highest method quality among SPS methods. The results from this analysis suggested 5×5 (area within 6.25 km^2 from the reference location) to be the best spatial scale at which neighborhood information should be included in the SPS models. Among the methods under indirect approach, methods considering atmospheric variables in the ST → AT models were found to have a higher quality than the other methods.

To summarize, the results obtained in Gaur and Simonovic (2016b) demonstrated that air temperature downscaling skill of direct SP and SPS models is superior to the SDSM model while the skill of indirect SP and SPS models is inferior to the SDSM model for the study region analyzed in the study. Between SP and SPS models, latter have higher quality in the regression relationship than the SP method. A comparison of results obtained at different neighborhood scales suggest 5 × 5 to be the optimum scale at which maximum downscaling skill is demonstrated by the SPS method.

5.5 Precipitation Downscaling

SP and SPS methods were evaluated for their skill in downscaling NARR precipitation over the period 2006–12 in Gaur and Simonovic (2018). The downscaling skill of SP and SPS methods were also compared with the skills of two state-of-the-art precipitation downscaling models: SDSM and GLM, described before in Section 2.2.3. Since SP and GLM-based models both employ GAM as the regression function, the same is used to build relationship between low resolution atmospheric variables and locally observed climatic data in the SDSM model. This is done to maintain regression function consistency among all models being evaluated in this study so that an unbiased evaluation of downscaling methodologies can be performed.

In Gaur and Simonovic (2018) the selection of predictors for SDSM and GLM models was made by analyzing monthly correlations between atmospheric variables and locally observed precipitation data at all gauging stations located within the study region. In this study, specific humidity (shum), high cloud area fraction (hcdc), medium cloud area fraction (mcdc), low cloud area fraction (mcdc), air temperature (air), and geopotential height (hgt) were tested for their correlation with observed precipitation across 57 precipitation gauging stations located in the study region. Highest correlations were obtained in the case of cloud-cover variables: high cloud area fraction (hcdc), medium cloud area fraction (mcdc), and low cloud area fraction (lcdc). Therefore they were selected as atmospheric predictor variables for performing downscaling by SDSM.sig and GLM.sig models (SDSM and GLM models with only significant or relevant atmospheric predictors). Another version of SDSM and GLM models which considers all atmospheric variables listed above for downscaling were also considered

Table 4.6 Models Evaluated in Gaur and Simonovic (2018)

S.No.	Model Name	Model Name (Short)	Predictors
1	SP	M1	P, LC, E
2	SP_LC	M2	P, E
3	SP_LC_elev	M3	P
4	SPS3x3	M4	P, LC, E, NLC_{3x3}, NE_{3x3}
5	SPS5x5	M5	P, LC, E, NLC_{5x5}, NE_{5x5}
6	SPS7x7	M6	P, LC, E, NLC_{7x7}, NE_{7x7}
7	SPS9x9	M7	P, LC, E, NLC_{9x9}, NE_{9x9}
8	SDSM	M8	wnd, rhum, prmsl, lcdc, shum1000hpa,
9	GLM	M9	mcdc, hcdc, air, shum850hpa, shum500hpa, hgt1000hpa, hgt850hpa, hgt500hpa
10	SDSM.sig	M10	lcdc, mcdc, hcdc
11	GLM.sig	M11	

and referred to as "SDSM" and "GLM" models in the study. The list of models evaluated in Gaur and Simonovic (2018) is provided in Table 4.6.

Two different tests of robustness were performed: (1) Test for temporal robustness (TR) and (2) Test for spatial robustness (SR). In the temporal robustness test, downscaling models were calibrated over the period 2006–10 and validated over the period 2011–13. On the other hand, in the spatial robustness test the downscaling models were calibrated across 29 stations located across the study region and validated across the rest of the gauging stations. The downscaled precipitation were evaluated on the basis of the downscaling model's ability to simulate following seven precipitation-based indices: (1) Spearman correlation coefficient between model simulated and observed data (sp.cor), (2) fraction of dry days, that is, fraction of days with less than 0.1 mm of rainfall (ddays), (3) maximum precipitation intensity (ppt.max), (4) mean wet day precipitation (ppt.wet), (5) total number of 1-day precipitation events (p1d), (6) total number of 2–4 day precipitation events (p2 to 4d) and (7) total number of 5 or more day precipitation events (p5d).

While calibrating SP, SPS, and GLM models, probability predictions made in the occurrence model are associated with an occurrence (1) or no-occurrence (0) value using a threshold value such that:

$$f(p) = \begin{Bmatrix} 1 & \text{if} \ \ p >= p_{\text{threshold}} \\ 0 & \text{if} \ p < p_{\text{threshold}} \end{Bmatrix} \tag{4.8}$$

where p denotes predicted probabilities as obtained from the occurrence model and $p_{threshold}$ denotes the threshold probability value chosen for analysis. In Gaur and Simonovic (2018), a series of $p_{threshold}$ values ranging between 0 and 1 were tested to select a threshold probability value that provides maximum prediction accuracy to the SP method and GLM-based models.

The variation of model efficiencies with probability threshold values for SP method–based models and GLM models for both snow-cover states and robustness tests were analyzed. Twenty-one probability threshold values evenly spaced between 0 and 1 at a spacing of 0.05 are considered for analysis. Model efficiency was calculated by evaluating the percentage of total data length correctly predicted by the calibrated model on validation time-series. It was found that occurrence model performance for SP and GLM-based models varied significantly with the choice of probability threshold value. Further, minor variations in model efficiency were also observed with differences in snow-cover state, robustness test, and downscaling model considered. Optimal threshold value for each model, snow-cover state, and robustness test combination was used for making prediction from these models. A summary of these optimal threshold values is presented in Table 4.7. It is noticed that threshold values for GLM models are higher than the threshold values of SP method–based models.

The calibrated models were used to downscale NARR precipitation grid data across the validation period (for TR test) and validation stations (for SR test). The performance of models in terms of average (across all

Table 4.7 Optimum Probability Threshold Values for Different Models for Snow-free (sf) and Snow-Covered (sc) Months, and for TR (SR) tests as Obtained in Gaur and Simonovic (2018)

	Snow	
Model	**sf**	**sc**
GLM	0.55 (0.55)	0.5 (0.6)
GLM.sig	0.55 (0.5)	0.45(0.55)
SP	0.35 (0.35)	0.3 (0.45)
SP-LC	0.35 (0.35)	0.35 (0.45)
SP-LC-elev	0.35 (0.35)	0.3 (0.45)
SPS3x3	0.4 (0.35)	0.3 (0.45)
SPS5x5	0.4 (0.4)	0.4 (0.4)
SPS7x7	0.4 (0.4)	0.4 (0.45)
SPS9x9	0.4 (0.35)	0.4 (0.5)

experiments and snow-cover states) rank correlation between modeled and observed data is shown in Table 4.8. Best performing model in terms of average performance is highlighted in grey. From the results it is clear that SP method–based models majorly outperform both SDSM and GLM models in terms of correlation. From the results obtained in Gaur and Simonovic (2018), it was also found that the performance of SP models was better in the snow-free months ($rho_{avg.} = 0.5$) as compared to snow-covered months ($rho_{avg.} = 0.4$) and better in the TR test ($rho_{avg.} = 0.5$) than the SR test ($rho_{avg.} = 0.4$). Following SP method–based models, SDSM model is found to perform best, followed by GLM model. Further the performance of SDSM.sig and GLM.sig models is found to be inferior than the SDSM and GLM models.

SP models perform best in simulating the fraction of total number of dry days in the validation time-series as evident from Table 4.8. The performance of SP models was again found to be better in snow-free months ($RMSE_{avg.} = 0.1$) than the snow-covered months ($RMSE_{avg.} = 0.5$) and in the TR test ($RMSE_{avg.} = 0.11$) than in the SR test ($RMSE_{avg.} = 0.12$). Among other models, GLM model was also found to perform well ($RMSE_{avg.} = 0.11$) followed by SDSM ($RMSE_{avg.} = 0.5$). Again the performance of SDSM.sig and GLM.sig models toward simulating dry day fraction was found to be inferior than the SDSM and GLM models.

Maximum precipitation intensity was not simulated satisfactorily by all three types of models considered in this study. This can be seen from Table 4.8 where biases associated with maximum precipitation values are presented. SP model was found to perform best followed by SP_LC_elev model. Among the three types of models, SP method–based models are found to perform best, followed by GLM and then by SDSM model. SP model performance was found to be significantly better in snow-free months ($RMSE_{avg.} = 54$ mm) than the snow-covered months ($RMSE_{avg.} = 90$ mm) and in the TR test ($RMSE_{avg.} = 70$ mm) than SR test ($RMSE_{avg.} = 78$ mm). Further the performance of SDSM.sig and GLM.sig models towards simulating maximum precipitation intensity was found to be inferior than SDSM and GLM models.

Wet-day mean precipitation was simulated reasonably well by SP method–based models (Table 4.8). Among the three types of models considered, GLM model was found to perform best ($RMSE_{avg.} = 0.8$ mm), followed by SP method ($RMSE_{avg.} = 1$ mm), and then by SDSM ($RMSE_{avg.} = 2.8$ mm). In the case of wet-day mean precipitation, SP model performance was found to be better in the snow-covered months

Table 4.8 Spearman Correlation Coefficient Between Model Simulated and Observed Precipitation, and RMSE Associated With Dry Day Fraction, Maximum Precipitation, Mean Wet-Day Precipitation, Total 1-Day Precipitation Events, Total 2–4 Days Precipitation Events, and Total >=5 Day Precipitation Events, for Models Considered in This Study

Statistic	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
Spearman correlation coefficient	0.44	0.44	0.44	0.43	0.43	0.42	0.42	0.43	0.40	0.41	0.36
RMSE-dry day fraction	0.09	0.11	0.10	0.10	0.12	0.13	0.12	0.48	0.13	0.55	0.14
RMSE- maximum precipitation	71	71	71	71	71	71	71	82	75	85	83
RMSE-mean wet-day precipitation	0.87	1.00	0.95	1.01	1.06	1.12	1.07	2.69	0.95	2.9	1.03
RMSE-total 1-day precipitation events	65	78	59	173	231	269	278	1615	181	1957	139
RMSE-total 2–4-day precipitation events	442	491	429	548	611	649	611	784	657	1381	673
RMSE-total >=5-day precipitation events	165	175	168	171	174	185	166	1085	177	509	189

Best performing model is highlighted in grey. Models M1 to M11 can be referred from Table 4.6.

($RMSE_{avg.} = 0.9$ mm) than the snow-free months ($RMSE_{avg.} = 1$ mm) and better in the TR test ($bias_{avg.} = 0.8$ mm) than the SR test ($bias_{avg.} = 1$ mm). Again the performance of SDSM.sig and GLM.sig models toward simulating mean wet-day precipitation intensity was found to be inferior to SDSM and GLM models. This suggests that the performance of SDSM and GLM models (formulated by considering all atmospheric variables) deteriorates by the selection of most relevant set of predictors for prediction.

The occurrences of 1-day and 2–4-day precipitation events were best simulated by SP method–based models, whereas GLM model performs best in simulating 5 or more day precipitation events as evident in Table 4.8. SDSM model was found to underestimate the occurrence frequency of 1-day and 2–4 day precipitation events and overestimate the occurrence frequency of 5 or more day precipitation events. The performances of SP models were again found to be superior in the TR test ($RMSE_{avg.} = 154$, 420, 132, respectively, for 1-day, 2–4 days, and 5 or more days precipitation events) than in the SR test ($RMSE_{avg.} = 298$, 743, 223 respectively for 1-day, 2–4 days, and 5 or more days precipitation event). Further SP method–based models were found to perform better in snow-free months than in the snow-covered months for 1-day and 2–4-day precipitation events ($RMSE_{avg.}$ found lower by 32 and 178 for 1 day and 2–4 day precipitation events, respectively). However, they were found to perform better in the snow-covered months than snow-free months for more than 5-day precipitation events ($RMSE_{avg.}$ found lower by 58). The performance of SDSM.sig and GLM.sig models toward simulating 1-day and 2–4-day precipitation events was found to be inferior than the SDSM and GLM models. However, in the case of 5 or more day precipitation events, SDSM.sig model was found to perform better than SDSM model. GLM.sig model still performs inferiorly to GLM model in the case of 5 or more day precipitation events.

Overall SP method–based models were found to perform better than the SDSM and GLM-based models. This was also evident when index specific bias associated with each individual model was normalized and combined. Index specific RMSE values were normalized so that intermodel comparisons can be made taking into consideration all seven indices. Overall, based on average normalized RMSE ($RMSE_{ANB}$), models ranked as: (1) SP_LC_elev ($RMSE_{ANB} = 0.02$), (2) SP ($RMSE_{ANB} = 0.03$), (3) SP_LC ($RMSE_{ANB} = 0.04$), (4) SPS9x9 ($RMSE_{ANB} = 0.16$), (5) SPS3x3 ($RMSE_{ANB} = 0.19$), (6) SPS5x5 ($RMSE_{ANB} = 0.22$), (7) GLM

($RMSE_{ANB} = 0.22$), (8) SPS7x7 ($RMSE_{ANB} = 0.28$), and (9) SDSM ($RMSE_{ANB} = 0.91$) in terms of model performance across indices. It was noted that SP method–based models, SP, SP_LC and SP_LC_elev, perform significantly better than all other models considered for evaluation. This suggested that the inclusion of neighborhood land-cover and elevation configuration didn't considerably improve SP model performance in the study region.

5.6 Conclusions From SP and SPS Model Validation

Results obtained from extensive validation of SP and SPS models have been described in this section. Downscaled NARR climate data obtained from SP and SPS models are found to be associated with acceptable amount of error. Additionally, the models are found to perform better than other state-of-the-art statistical downscaling models: SDSM and GLM. These findings establish the models as reliable downscaling tools for the southern Saskatchewan region. The unique advantage of these models is highlighted in the following section where they are used to identify land-cover change–induced hydroclimatic changes in future. Such an analysis is not possible with other state-of-the-art statistical downscaling models for the reasons discussed in Section 4.

6. CASE STUDY: USE OF SP AND SPS METHODS TO IDENTIFY FUTURE LAND-COVER CHANGE–INDUCED CHANGES IN HYDROCLIMATIC VARIABLES

In this section we present a case study where SP and SPS methods were used to identify hydroclimatic changes projected as a result of future land-cover change in four catchments located in the southern Saskatchewan region of Canada (Gaur and Simonovic, 2018). The locations of the selected catchments are shown by black crosses in Fig. 4.1. HYDAT is a flow database maintained by the Water Survey of Canada, an organization that is responsible for collecting and disseminating flow data in Canada. In Gaur and Simonovic (2018), catchments with HYDAT IDs: 05EG006, 05EG008, 05MC004, and 11AF005 were chosen for analysis. Hydrometric details of the selected catchments are summarized in Table 4.9. It can be noted that all four catchments were small (with catchment area less than $1000\ km^2$) and hence can be modeled as lumped hydrologic systems. Additionally, all are nonregulated catchments which mean that changes in outflow from these

Table 4.9 List of Catchments Analyzed in Gaur and Simonovic (2018)

HYDAT ID	Name	Drainage Area (Km^2)	Elevation (Masl)	Data Length
05EG006	Birling creek near Paynton	426	593	1970–92
05EG008	Page creek near Iffley	921	673	1971–87
05MC004	Conjuring creek near Preeceville	255	594	1965–92
11AF005	Beaver creek at international boundary	387	773	1977–95

catchments can be attributed to changes in external climate forcing and changes in the physical characteristics of the catchments.

6.1 Data Used

NASA satellite elevation and land-cover products were used to satisfactorily represent catchment physical characteristics. To this end, Shuttle Radar Topographic Mission (SRTM) elevation product (Jarvis et al., 2008) and MODIS land-cover data, MCQ12Q1 in the University of Maryland (UMD) classification scheme, were used. The elevation and land-cover data have a spatial resolution of 90 and 500 m respectively. The land-cover data has an annual temporal frequency and is considered for the period 2006–12 in this study.

ANUSPLIN is a 10-km gridded precipitation, maximum temperature, and minimum temperature data developed by applying thin plate spline smoothing algorithms on gauged climate records across Canada (Hopkinson et al., 2011 and Hutchinson et al., 2009). ANUSPLIN data grids falling within the selected catchments were considered for analysis. Daily air temperature and precipitation at 947 climate gauging stations located in Alberta, Saskatchewan, and Manitoba provinces of Canada were acquired from Environment Canada. The gridded and gauged climate data were considered for the periods 1961–2012 and 2006–12, respectively, in this study. It was found that together these gauging stations characterize all of University of Maryland (UMD) land-cover classes (listed in Table 4.2) and hence were considered appropriate for calibrating the SP and SPS downscaling models.

Daily flow data for the selected catchments were collected from the HYDAT database. Flow data available between the period 1961 and 2012 was collected. A summary of the flow data length available at each flow

Table 4.10 List of GCMs Considered for Analysis in Gaur and Simonovic (2018)

GCM	Model	Resolution	Source
1	IAP-FGOALS	$1.66° \times 2.81°$	Institute of Atmospheric Physics, Chinese Academy of Sciences, China
2	MRI-CGCM3	$1.08° \times 2.16°$	Meteorological Research Institute, Japan
3	NorESM1-M	$2° \times 2°$	Norwegian Climate Centre, Norway

gauging station, drainage area, mean elevation, and HYDAT ID is provided in Table 4.9.

Daily reanalysis precipitation rate and near-surface air temperature were acquired for the period 2006–12 from North American Regional Reanalysis (NARR) data repository (Mesinger et al., 2006). The data are spatially interpolated across all gauging stations. The interpolated reanalysis climate data are used to calibrate downscaling models considered in this study.

Future land-cover projections for southern Saskatchewan region corresponding to three climate models listed in Table 4.10 and emission scenarios RCP 2.6 and RCP 8.5 were prepared in Gaur and Simonovic (2016a) by downscaling and reclassifying future harmonized land-use projections from two integrated assessment models: IMAGE and MESSAGE (Hurtt et al., 2011). In Gaur and Simonovic (2018) annual land-cover projection at the selected catchments for the periods 2041–60 and 2081–2100 were acquired from the land-cover data produced in Gaur and Simonovic (2016a). In Gaur and Simonovic (2016a) variations in land-cover for different GCMs were found negligible as compared to variations with different emission scenarios, a single land-cover projection was considered for all three GCMs in this study.

An analysis of future land-cover trajectories over the 21st century across the selected catchments for RCP2.6 and RCP8.5 in Gaur and Simonovic (2018) suggested that parts of catchments 05EG006 and 05EG008 will transition from forested land to BSV land-cover class under both RCPs; however, the transition is expected to be more drastic in case of RCP 8.5 as compared to RCP 2.6. In case of catchment 05EG004 under RCP 2.6, a section of the forested land was projected to transition to grasslands, while under RCP 8.5 a transition from forested land to BSV was expected. Lastly, in the case of catchment 11AF005 under RCP 2.6 a transition from croplands to grasslands was projected while in the case of RCP 8.5 a conversion of grasslands to BSV was projected.

GCM-based daily air temperature and precipitation projections for the period 2006–2100 were collected from the coupled model inter-comparison project-phase 5 (CMIP5) data repository (Taylor et al., 2012). Data corresponding to climate models listed in Table 4.3 and for two Representative Concentration Pathways (RCPs)—RCP 2.6 and RCP 8.5 was acquired. The choice of climate models was made based on the availability of future land-cover data as developed in Gaur and Simonovic (2016a).

6.2 Sacramento Hydrologic Model

Hydrologic models are used to estimate catchment flow response provided by a set of catchment geophysical and climatic characteristics. In Gaur and Simonovic (2018), Sacramento hydrologic model is used to estimate catchment outflow under future projected climatic and land-cover trajectories. The Sacramento model (Burnash, 1995) flow is generated by distributing precipitation falling at a location to overland flow, interflow, and baseflow components accounting for losses due to evapotranspiration and interception. Groundwater movement is modeled by considering upper zone and lower zone storages. Runoff is contributed by five different processes: (1) direct runoff from permanent and temporary impervious areas, (2) surface runoff due to precipitation occurring at a rate faster than percolation and interflow that take place when both upper zone storages are full, (3) interflow resulting from the lateral drainage of a temporary free water storage, (4) supplemental base flow, and (5) primary base flow. The model has 13 free parameters which are optimized in the model.

Routing in this model is performed using exponential form of unit hydrograph with explicit slow and quick flow components. The routing scheme involves three free parameters which are optimized. Snow-melt is modeled offline using a temperature index modeling approach discussed in Walter et al. (2005). Model is calibrated to optimize an objective function which is a weighted sum of coefficient of determination (R^2) and relative bias. The shuffled complex evolution global optimization method (Duan et al., 1993) followed by a local optimization method (Nelder and Mead, 1965) with multistart options is used to calibrate the model. A lumped version of this hydrologic model is considered sufficient for this study because the catchments considered are small (drainage area $<1000\ km^2$).

6.3 Analysis and Results

In Gaur and Simonovic (2018), SP and SPS downscaling models were calibrated over the period 2006–12 using gauged climate data as predictant variable and model specific predictor variables as described before. Precipitation and temperature downscaling models were calibrated separately for each month. The calibrated temperature downscaling models were used to downscale future GCM maximum and minimum temperature, and calibrated precipitation downscaling models were used to downscale GCM precipitation projections. Downscaling was performed for historical (2006–16), 2050s (2041–60), and 2090s (2081–2100) timelines with all MODIS grid cells falling within the selected catchments.

Future projected changes in precipitation, maximum and minimum temperature, and flow variables were analyzed between (1) historical and 2050s; and (2) historical and 2090s, for the selected catchments. An analysis of the projected changes suggested that both minimum and maximum temperatures can be expected to increase across the four catchments in future. Under RCP 2.6, minimum temperatures were projected to increase by 4.1 (2.3) K, 4.1 (2.4) K, 3.7 (2.1) K, and 4.1 (2.5) K by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. Under RCP 8.5 minimum temperatures were projected to increase by 4.1 (7.9) K, 4 (7.9) K, 4.3 (8.1) K, and 4.2 (8.6) K by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. On the other hand, under RCP 2.6 maximum temperatures were projected to increase by 3.9 (2.5) K, 3.7 (2.4) K, 3.9 (2.9) K, and 4.4 (3.4) K by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. Under RCP 8.5 maximum temperatures were projected to increase by 2.4 (6) K, 2.8 (6.6) K, 3.8 (7.6) K, and 3.3 (6.9) K by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. Overall larger changes are projected for minimum temperature as compared to maximum temperature in all four catchments.

Changes in precipitation were investigated by accessing projected changes in mean precipitation, dry days (days with precipitation less than 0.1 mm), frequency of small (lasting 1 day), moderate (lasting 2–4 consecutive days), and heavy (lasting more than 4 consecutive days) precipitation events between historical and future timelines. Mean precipitation was projected to increase slightly across the catchments. Under RCP 2.6 mean precipitation was projected to increase by 0.03 (0.03), 0.02 (0.03), 0.05 (0.06), and 0.01 (0.04) mm/day by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. Under RCP 8.5 mean precipitation was projected to increase by 0.07 (0.13), 0.06 (0.12), 0.05 (0.1), and 0.06 (0.08) mm/day by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. In accordance with these projected changes in mean precipitation, a decrease in the number of dry days and increase in the frequency of small, moderate, and heavy precipitation events were also projected across the selected catchments.

Next, Sacramento hydrologic model was calibrated for each catchment using ANUSPLIN-based gridded climate data and discharge data available at each discharge gauging station. ANUSPLIN gridded precipitation, maximum temperature, and minimum temperature gridded data were used to calibrate the hydrologic model because only a few or none of the climate gauging stations were found to be located within the selected catchments. The available daily discharge data length at each catchment is provided in Table 4.9. The entire flow data length was considered while calibrating the model to obtain robust hydrologic parameters which can then be used to predict future flows in the selected catchments. Nash-Sutcliffe efficiency and correlation between observed and modeled streamflow were analyzed for all catchments and satisfactory calibration results were obtained.

The calibrated Sacramento model was thereafter used to predict catchment outflow using downscaled climate projections. Above mentioned changes in climate variables were also reflected in the projected changes in catchment outflow which was largely projected to increase across the four catchments. Under RCP 2.6 flow rate was projected to change by −0.01 (−0.01), 0.03 (0.02), 0.05 (0.01), and 0.04 (0.20) m^3/s by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. On the other hand, under RCP 8.5 flow rate was projected to change by 0.02 (−0.01), 0.08 (0.04), 0.1 (0.1), and 0.14 (0.16) m^3/s by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. In summary, all three hydroclimatic variables of interest, that is, precipitation, temperature, and flow were projected to increase across the selected catchments in future.

To further investigate land-cover driven hydroclimatic changes in the catchments, quantitative and statistical differences between the projected hydroclimatic changes with and without considering land-cover change were analyzed. Here former (i.e., climatic changes with land-cover change) are presented as mean hydroclimatic change obtained from SP, SPS3x3, SPS5x5, SPS7x7, SPS9x9 downscaling models while latter (i.e., climatic changes without land-cover change) were presented as the mean hydroclimatic change obtained from SP_LC downscaling model. In Gaur and Simonovic (2018) SP_LC model was referred to a version of SP model where land-cover is omitted from the list of predictors used for downscaling in the SP model. In other words only large-scale climate and elevation are used as predictors for downscaling GCM projections in the SP_LC model.

Statistical significance of the influence of land-cover change was explored by performing Wilcoxon signed rank test (Wilcoxon, 1945) on projected temperature, precipitation, and flow time-series obtained from SP_LC and other downscaling models for timelines 2050s and 2090s. Test results identified statistically significant differences in the projected time-series for all three hydroclimatic variables, and for both future timelines.

The magnitudes of land-cover driven climatic changes and its flow-related implications were explored. In the case of mean temperature (average of minimum and maximum temperature), a land-cover change driven warming of −0.01 (−0.01), −0.01 (−0.01), −0.01 (−0.02), and 0.16 (0.19) K, averaged across both emission scenarios was obtained by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. In the case of mean precipitation, a land-cover change driven increase of −0.02 (0.05), 0.01 (0.03), 0.02 (0.01), and −0.06 (−0.11) mm/day, averaged across GCMs and both emission scenarios were obtained by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively. Finally in the case of flow, a land-cover change driven increase of −0.014 (0.001), −0.022 (−0.002), −0.004 (−0.004), and 0.002 (0.002) m^3/s averaged across GCMs and both emission scenarios was obtained by 2050s (2090s) for catchments 05EG006, 05EG008, 05MC004, and 11AF005, respectively.

Finally, the distribution of land-cover–driven changes across different quantiles was also evaluated. It was found that for all catchments, higher quantiles were associated with higher changes. In other words hydro-climatic extremes were found to be more influenced from the land-cover change than the lower climate and flow distribution quantiles.

6.4 Conclusions From SP Model Case Study

In Gaur and Simonovic (2018) climatic effects of land-cover change and their implications on flow were explored. Four small catchments located in the southern Saskatchewan region were selected for investigation. Future climate and hydrologic projections are made by downscaling future precipitation and temperature projections made by three GCMs and two emission scenarios. By analyzing changes projected between baseline (2006–16) and future timelines, 2050s and 2090s, it can be concluded that minimum temperature, maximum temperature, precipitation magnitude and intensity, and flow rate are set to increase in future over the selected catchments.

Climatic projections with future land-cover change are obtained from SP model versions: SP, SPS3x3, SPS5x5, SPS7x7, SPS9x9 and projections without considering future land-cover change are obtained using SP model version: SP_LC. The downscaled climate projections are used to derive future streamflow response from all four catchments using Sacramento hydrologic model. A comparison of these two sets of hydroclimatic projections for 2050s and 2090s showed statistically significant differences between them. These results are in line with the findings from other dynamic downscaling-based studies which have found notable land-cover change–induced climatic changes at local to regional scales (for example, Malyshev et al., 2015). An analysis of the variation of slope of the projected change with quantiles indicates an increasing trend with the quantile value for all three variables analyzed. This suggests higher land-cover–induced changes in hydroclimatic extremes as compared to the means.

Most importantly, this case study clearly demonstrated the usefulness of SP and SPS models in incorporating regional geophysical characteristics in model formulation, as well as their ability to model climatic influences of geophysical changes within a statistical downscaling framework.

7. WORKING EXAMPLE OF SP AND SPS METHODS IN R PROGRAMMING LANGUAGE

In this section, we demonstrate how SP and SPS methods can be applied to downscale GCM projections in the R programming language. First we provide a gentle introduction to the R programming language, and then present a step-by-step demonstration of SP and SPS model downscaling using it.

7.1 Basics of R Programming Language

R is a programming language and software environment for statistical analysis, graphics representation, and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team (R Development Core Team, 2008). R implements many common statistical procedures, as well as provides excellent graphics functionalities through its libraries and packages. It is an open source language which means that the users are able to acquire it free of cost, as well as can contribute towards its research and development. In this subsection, basic features of R programming language have been reviewed with an intention that this can help the readers to get started with R.

7.1.1 Downloading and Installing R

R programming language for windows operating system can be downloaded from https://cran.r-project.org/bin/windows/base/. R for Linux and Mac operating systems can also be downloaded from the R project for statistical computing website: https://www.r-project.org/. It is also common to use R Studio to run and edit R codes as this software provides enhanced code editing, debugging, and visualizing capabilities to the R users. R Studio for Windows, Linux, and Mac operating systems can be downloaded from https://www.rstudio.com/products/rstudio/download/. R and R Studio softwares can then be installed on the computer by running the executables obtained from above sources.

7.1.2 Common Data-Types in R Programming

Some of the most common data-types used in R programming language are vector, data-frame, and list. A vector can be a sequence of numbers, logical values, or character strings. A vector with three numeric values can be defined as follows in the R Studio command-line. Below the user commands are provided in lines commencing with the symbol: ">" while output from the R (if any) is provided in lines commencing with the symbol: "[1]".

```
C1 > c(3, 4, 7)
[1] 3 4 7
```

A vector with three logical values can be defined as follows.

```
C2 > c(TRUE, FALSE, FALSE)
[1] TRUE FALSE FALSE
```

A vector with three character values can be defined as follows. The resulting vector has also been saved into a variable named: vec.char.

```
C3 > vec.char = c("a", "b", "c")
```

A vector element can be extracted using the "[]" brackets with the index of the element to be extracted as shown below.

```
C4 > vec.char[1]
[1] a
```

A data-frame is used to store vectors of equal length in the form of tables. Below is the command to create a data-frame "df" with three columns and four rows representing the marks obtained by four students named: "A," "B," "C," "D" in two subjects: maths and physics.

```
C5>df=data.frame(name=c("A","B","C","D"),maths=c(85,90,97,76),
      physics=c(88,66,76,98))
```

Data-frame element(s) can be extracted using the "[]" brackets with the row number and column number indices of the element to be extracted. The entire row or column can be extracted by specifying the same in the command. An entire column can also be extracted by specifying the column name together with the "$" operator as shown below.

```
C6 > df[1,2]
[1] 85
C7>df[1,]
[1] A 85 88
C8 > df[,1]
[1] A B C D
C9>df$name
[1] A B C D
```

A list can be used to store vectors of equal or unequal lengths. Below is an R command to create a list with 3 numeric, 4 logical, and 5 character elements stored in the first, second, and third elements. The list is stored as a variable named "lst."

```
C10 > lst = list (c(1, 5, 7), c(TRUE, TRUE, FALSE, FALSE), c("a", "b",
    "c", "d", "e"))
```

List elements can be extracted by using the "[[]]" brackets along with the list element number that needs to be extracted. Further, subelements within a list element can also be accessed using "[]" brackets with the index of the subelement number as demonstrated below.

```
C11 > lst[[1]]
[1] 1 5 7
C12 > lst[[1]][3]
[1] 7
```

7.1.3 Relevant R Packages

R packages are a collection of R codes, functions, data, and compiled code in a well-defined format. R comes with a standard set of packages. Other packages can be downloaded and installed separately by the users based on their needs. Here we provide a brief introduction of a few R packages (apart from the standard packages) that are very useful in performing downscaling of GCM data by SP and SPS methods in R. These packages are:

- MODIS and MODISTools: The intended purpose of these packages is to facilitate acquisition and processing of MODIS data-products. MODIS package contains functions to gain automated access to the global online data archives and processing capabilities such as file conversion, mosaicking, subsetting, and time-series filtering (Mattiuzzi, 2016). The package can be downloaded from Comprehensive R Archive Network (CRAN) by running the following command on the command line: install.packages("MODIS", repos="http://R-Forge.R-project.org"). MODISTools package also allows users to extract MODIS data time-series at one or more than one locations without downloading the image rasters (Tuck and Phillips, 2016). The package can be downloaded from Comprehensive R Archive Network (CRAN) by running the following command on the command line: install.packages("MODISTools", repos="http://R-Forge.R-project.org").
- ncdf4: This package is designed to work with NetCDF libraries version 4 in R, which is the most commonly used NetCDF version currently. Another package "ncdf" can be used to access NetCDF version 3 libraries. In "ncdf4" package, utilities like chunking and compression have also been included. The package can be downloaded from CRAN by running the following command on the command line: install.packages("ncdf4", repos="http://R-Forge.R-project.org").
- raster: This package is intended to facilitate raster processing in R. Among other functions, the package contains functions that can read and write rasters, perform raster operations such as reprojection, resampling, filtering, merging, etc.; perform raster calculations; and visualize raster data (Hijmans, 2015). The package can be downloaded from CRAN by running the following command on the command line: install.packages("raster", repos"http://R-Forge.R-project.org").
- lubridate: The lubridate package is intended to facilitate easy handling of date-time data in R. Among others, it contains functions that can be used to extract components of a date-time such as year, month, day, hour, minute, and seconds, and perform algebraic manipulation on the date-time objects (Grolemund and Wickham, 2011). The package can be downloaded from CRAN by running the following command on the command line: install.packages("lubridate", repos =" http://R-Forge.R-project.org").

- reshape2: This R package is extremely useful to transform data between wide and long formats. A wide format has a column for each variable while a long format has a column for possible variable types and another column for the values of these variables (Wickham, 2007). The package can be downloaded from CRAN by running the following command on the command line: install.packages("reshape2", repos="http://R-Forge.R-project.org").
- ggplot2: This package is meant for "declaratively" creating graphic by telling ggplot2 how to map variables to aesthetics and what graphical primitives to use. It produces plots following the grammar of graphics (Wilkinson, 2005) where essential building blocks of a graph, that is, data, aesthetic mapping, geometric object, statistical transformations, scales, coordinate system, position adjustments, and faceting are specified by the user (Wickham, 2009). The package can be downloaded from CRAN by running the following command on the command line: install.packages("ggplot2", repos="http://R-Forge.R-project.org").
- mgcv: This package is very helpful in performing GAM regression. It includes several methods for estimating regression parameters, smoothing functions, and link functions in GAMs in computationally efficient manner (Wood, 2011). The package can be downloaded from CRAN by running the following command on the command line: install.packages("ggplot2", repos="http://R-Forge.R-project.org").

7.1.4 Other Relevant R Functions

Apart from the functions available in aforementioned packages, plenty of useful functions are available in R base library. They have been used extensively while downscaling GCM projections using SP and SPS models. A description of some of those functions that have been most extensively used is provided below. A more detailed description of any of these functions (or any other function) can be obtained by running "?"function-name command in R.

- which(): This function is used to know the position of elements of a logical vector that are TRUE.
- subset(): This function is used to extract section of a data-frame with rows that meet a particular criteria. For instance in a data-frame that stores monthly discharge data time-series, this function can be used to select part of the time-series that corresponds to January or have month values equal to 1.

- sapply/lapply: The sapply() and other similar functions are an alternate for looping in R. Their usage is recommended as their usage can make the R codes run much faster than when for instance for() loops are used. The "apply" family of functions have many variants like sapply() which stands for simplify and apply, lapply() which stands for list and apply, vapply(), tapply(), etc.
- rbind/cbind(): Both rbind() and cbind() are used to combine two data-frames or matrices. It is essential that data-frames (or matrices) have the same number of columns when rbind() is used to combine them. Similarly, data-frames (or matrices) should have equal number of rows when cbind() is used.
- do.call(): The do.call() function executes a function call over all list elements passed to it. This function is commonly used in conjunction with lapply() function where "list" output obtained from lapply() are further analyzed or manipulated using do.call() function.

7.2 Application of SP and SPS Model Downscaling in R

In this section a demonstration of how R programming language can be used to downscale climate model projections using SP and SPS methods is presented. A discussion on how to extract, organize, and prepare remotely sensed and climate model datasets in R is provided first. This is followed by a working example where SP and SPS models are applied to downscale a sample future temperature projection dataset.

Climate model datasets from GCMs or reanalysis products are generally available in Network Common Data From (NetCDF) format. NetCDF is a self-describing, machine independent data format that supports the creation, access, and sharing of array oriented data. By self-describing, it means that information about specifications of the file, the data it stores, and its layout that is stored within the file. In R NetCDF format files can be accessed using "ncdf4" package. All packages need to be downloaded and loaded in the R session prior to their usage. Packages can be loaded by using command "library" (package-name) in the R session. Following commands can be used to access and extract data from NetCDF files in R:

```
C13 > file.read = nc_open(file.name)
```

Above function nc_open reads a NetCDF file. The location of.nc file should be specified in the argument file.name of the function. The details of the.nc file are stored in the variable file.read.

```
C14 > file.read$dim$names
```

The variable file.read can be used to extract import file characteristics such as file dimensions, which conveys the layout of data stored in the.nc file. Above command prints out detailed description of all dimensions associated with the file.read variable.

```
C15 > file.read$var$names
```

Variables convey information about the data stored in the.nc files. Above command prints out detailed description of all dimensions associated with the file.read variable.

```
C16 > ncvar_get(file.read, var.name)
C17 > ncvar_get(file.read, varid=var.name, start=c(1,1,1),
      count=c(dim1,dim2,1))
C18 > sapply(1:10, function(x) ncvar_get(file.read, varid=var.name,
      start=c(1,1,x), count=c(dim1,dim2,1)))
```

The function "ncvar_get" is used to extract data stored in a particular variable in the.nc file. Command C16 will extract the values of a variable with name "var.name" stored in the.nc file. Sometimes due to data volume it is not possible or desirable to import and save all of the data in one variable in one go. In those cases, "ncvar_get" command can also be specified other attributes so that the data is read in manageable chunks. For instance, command C17 provided above reads only the first array element of the variable "var.name". Variables "dim1" and "dim2" denote the x and y dimensions of the data array. Command C18 performs the same operation iteratively for array elements 1 to 10 using the "sapply" function. The indices for which the data needs to be extracted is ascertained by examining the time-indices for which data is provided in the.nc file and then finding indices that contain data for the user-defined time-period of interest.

Above discussion provides a brief introduction on how R programming language can be used to access GCM and reanalysis climate data that is typically available in the.nc format using the ncdf4 package.

Next, we discuss how MODIS data can be extracted using R programming language. MODIS-based climatic and land-cover data are very useful for performing SP and SPS model—based downscaling, and they can be easily extracted and managed using R programming language. We demonstrate the use of two packages: "MODISTools" and "MODIS" towards downloading and analyzing MODIS data in R. The first package

MODISTools can be used to download spatiotemporal MODIS data using "MODISSubsets" function provided below:

```
C19 > MODISSubsets(LoadDat, Products, Bands, Size)
```

In "MODISSubsets" function, argument "LoadDat" reads in a data-frame with details about coordinates and IDs of all locations where data needs to be extracted, as well as the start and end dates of the data to be downloaded. The argument "Products" reads in the product code, which can be obtained from the function GetProducts(). The argument "Bands" is supplied the band names to be downloaded. For a particular product, a list of bands can be obtained using the Getbands() command. The argument "Size" is supplied with the spatial scale at which the data should be extracted. A value of c(0,0) is supplied if only data at the location of interest needs to be extracted. Other values such as c(1,1) provide values spatially averaged over an area of 2 km^2 from the location of interest. Following command C20 will extract the day-time and night-time surface temperature data sensed by Terra satellite over the London (Ontario) city (lon = −81.25, lat = 42.98) for the year 2012.

```
C20>MODISSubsets(LoadDat=data.frame(lat=42.98,long=-81.25,
            start.date=2012,end.date=2012,id=1),
            Products = "MOD11A1"
            Bands = c("LST_Day_1km", "LST_Night_1km"),
            Size = c(0,0))
```

Downloading MODIS data using "MODISTools" is most advantageous when the data needs to be extracted at a limited number of locations. However, if data needs to be extracted over a region or a country, it is beneficial to extract a raster image encompassing the region of interest for analysis. The runGdal function provided in the MODIS package is extremely useful in such cases. Following is a description of the function:

```
C21 > runGdal(product, begin, end, extent)
```

In the "runGdal" function, the argument "product" needs to be supplied with the MODIS product ID, beginning and end dates for data extraction are supplied in the arguments: "begin" and "end" respectively. The area for which the data needs to be extracted is supplied through the "extent" argument. The area to be supplied can be selected interactively, by supplying a shapefile or a raster file, by specifying the country name or an extent object. Other important arguments such as "outProj" which supplies the output raster projection "pixelSize" which can be supplied with the output data spatial resolution, and "dataFormat" which can be used to specify the data-format of output raster image. The following command can be used to extract surface temperature data sensed by Terra satellite over Canada for the year 2012.

```
C22 > runGdal(product="MOD11A1",begin="2,012,001",
      end = "2,012,366",extent="canada")
```

So far the discussion provided in this section dealt with the preparation of geophysical and climatic data required for the application of SP and SPS method downscaling models. Once the climatic and geophysical data are prepared, the downscaling models can be used to downscale GCM projections. We demonstrate this by downscaling future temperature and precipitation projections obtained from a GCM: MRI-CGCM3 under RCP2.6 using SP and SPS models. The data for downscaling model calibration and for making future projections can be downloaded from https://drive.google.com/drive/folders/0B86JMqp3jy6CcmYzT2dQNXlMQW8?usp=sharing. The datasets are provided in .rds format which is a format to save and load R objects as files in the system. File named "Model calibration data.rds" contains data needed to calibrate SP and SPS models and file named "Future prediction data.rds" contains data needed for making downscaled future climatic projections.

Once the folder is placed in a directory (termed as "Fakepath" below), the data can be imported into R using following commands:

```
C23 > cal.data=readRDS("Fakepath/Model calibration data.rds")
C24 > pred.data=readRDS("Fakepath/Future prediction data.rds")
```

The function "readRDS" used above reads in any R object with .rds extension. Here model calibration data are imported into an R data-frame object "cal.data" and prediction data are imported into an R object "pred.data." The datasets can be examined using two very useful R functions: "summary" and "head"/"tail" as shown below.

```
C25 > summary(cal.data)
C26 > head(cal.data, 50)
```

The former function provides a summary of the data stored in each column of the data-frame. The latter functions "head" or "tail" shows user requested number (which in above case is 50) of first or last few lines in a data-frame, respectively. Table 4.11 provides a description of different columns in data-frames: "cal.data" and "val.data."

Table 4.11 Descriptions of Columns Present in the Calibration and Prediction Data

Column Name	Description
elev	Elevation of the gauging stations
P.obs	Precipitation recorded at the gauging stations
T.obs	Temperature recorded at the gauging stations
C.3x3/C.5x5/C.7x7/C.9x9	Cropland fraction in 3x3/5x5/7x7/9x9 neighborhood scale
G.3x3/G.5x5/G.7x7/G.9x9	Grassland fraction in 3x3/5x5/7x7/9x9 neighborhood scale
BSV.3x3/BSV.5x5/BSV.7x7/BSV.9x9	Barren or Sparsely Vegetated fraction in 3x3/5x5/7x7/9x9 neighborhood scale
OS.3x3/OS.5x5/OS.7x7/OS.9x9	Open Shrublands in 3x3/5x5/7x7/9x9 neighborhood scale
W.3x3/W.5x5/W.7x7/W.9x9	Water fraction in 3x3/5x5/7x7/9x9 neighborhood scale
ENF.3x3/ENF.5x5/ENF.7x7/ENF.9x9	Evergreen Needleleaf Forest fraction in 3x3/5x5/7x7/9x9 neighborhood scale
UB.3x3/UB.5x5/UB.7x7/UB.9x9	Urban fraction in 3x3/5x5/7x7/9x9 neighborhood scale
DBF.3x3/DBF.5x5/DBF.7x7/DBF.9x9	Deciduous Broadleaf Forest fraction in 3x3/5x5/7x7/9x9 neighborhood scale
DNF.3x3/DNF.5x5/DNF.7x7/DNF.9x9	Deciduous Needleleaf Forest fraction in 3x3/5x5/7x7/9x9 neighborhood scale
MF.3x3/MF.5x5/MF.7x7/MF.9x9	Mixed Forest fraction in 3x3/5x5/7x7/9x9 neighborhood scale
S.3x3/S.5x5/S.7x7/S.9x9	Savannas fraction in 3x3/5x5/7x7/9x9 neighborhood scale
WS.3x3/WS.5x5/WS.7x7/WS.9x9	Woody Savannas fraction in 3x3/5x5/7x7/9x9 neighborhood scale
CS.3x3/CS.5x5/CS.7x7/CS.9x9	Closed Shrublands fraction in 3x3/5x5/7x7/9x9 neighborhood scale
EBF.3x3/EBF.5x5/EBF.7x7/EBF.9x9	Evergreen Broadleaf Forest fraction in 3x3/5x5/7x7/9x9 neighborhood scale
ef.elev.3x3/ef.elev.5x5/ef.elev.7x7/ef.elev.9x9	Relative elevation fraction in 3x3/5x5/7x7/9x9 neighborhood scale
T.NARR	NARR temperature interpolated at the gauging station location
P.NARR	NARR precipitation interpolated at the gauging station location
LC	Land-cover at the gauging station
GCM.ppt	GCM precipitation interpolated at the gauging station location
GCM.tas	GCM temperature interpolated at the gauging station location
date	Date associated with the data

SP method calibration can be performed using "gam" function available in the package "mgcv." This can be performed with either of the following commands:

```
C28 > SP.mod.T = gam(T.obs ~ s(T.NARR) + LC+ s(elev), data = cal.data)
C29 > SP.mod.T.jan = gam(T.obs ~ s(T.NARR) + LC+ s(elev),
      data = subset(cal.data, month(date) %in% 1))
```

The command C28 fits a GAM function on the observations provided in the "cal.data" object and the resulting model is stored in another R object: "SP.mod." If the calibration needs to be performed using data belonging to all days in the month of January, this can be done using command C29. The code subsets only the data belonging to January month by using "subset" function and checking which months corresponding to the "dates" column equal to 1. The model can be calibrated for other months in a similar fashion.

SPS model for downscaling temperature can be calibrated in a similar way. Command C30 calibrates SPS3x3 model using entire "cal.data" series and stores the calibrated model in an R object named: "SPS3x3.mod.T." Appropriate modifications in predictor variables can be made when calibrating SPS models for other neighborhood scales.

```
C30 > SPS3x3.mod.T = gam(T.obs ~ s(T.NARR) + s(LC) + s(elev) +
                         s(C.3x3) + s(G.3x3) + s(BSV.3x3) + s(OS.3x3) +
                         s(W.3x3) + s(ENF.3x3) + s(UB.3x3) + s(DBF.3x3) +
                         s(DNF.3x3) + s(MF.3x3) + s(S.3x3) + s(WS.3x3) +
                         s(CS.3x3) + s(EBF.3x3) + s(ef.elev.3x3),
                         data = cal.data)
```

Performing downscaling using the calibrated models is also very straightforward and intuitive. For instance, following commands can be used to predict downscaled temperature data using above calibrated SP and SPS3x3 models:

```
C31 > colnames(pred.data)[which(colnames(pred.data) %in% "GCM.tas")]
      ="T.NARR"
C32 > pred.data$SP.pred.T = predict(SP.mod.T, newdata = pred.data)
C33 > pred.data$SPS3x3.pred.T = predict(SPS3x3.mod.T, newdata
      = pred.data)
```

For making predictions using the calibrated GAM model, the predictor variables present in the calibration data should also be present in the validation data. Therefore the column "GCM.tas" which contains interpolated GCM data is renamed as "T.NARR" which contained NARR-derived model data in the calibration dataset. This is performed by command

C31 presented above. Following this, the calibrated models "SP.mod" and "SPS3x3.mod" are used to downscale prediction dataset "pred.data" using commands provided in the commands C32 and C33 above.

Precipitation downscaling by SP and SPS models involves two steps as discussed in Section 3. First, dry and wet-day sequences are predicted using a logistic regression model and then, wet-day precipitation intensity is predicted using a GAM model. The calibration of logistic regression model and GAM models can be performed in R using following set of commands:

```
C34 > cal.data$r.switch=0
C35 > cal.data$r.switch[which(cal.data$P.obs>0.1)]=1
C36 > occ.mod.SP.P=glm(r.switch~P.NARR+elev+LC,data=cal.data,
      family="binomial")
C37 > int.mod.SP.P=gam(P.obs~s(P.NARR)+s(elev)+s(LC),data=
      subset(cal.data,r.switch==1))
C38 > occ.mod.SPS3x3.P = glm(r.switch ~ P.NARR + elev + LC + C.3x3 +
      G.3x3 + BSV.3x3 + OS.3x3 + W.3x3 + ENF.3x3 + UB.3x3 + DBF.3x3 +
      DNF.3x3 + MF.3x3 + S.3x3 + WS.3x3 + CS.3x3 + EBF.3x3 +
      ef.elev.3x3, data=cal.data, family="binomial")
C39 > int.mod.SPS3x3.P = gam(P.obs ~ s(P.NARR) + s(elev) +
      s(LC) + s(C.3x3) + s(G.3x3) + s(BSV.3x3) + s(OS.3x3) +
      s(W.3x3) + s(ENF.3x3) + s(UB.3x3) + s(DBF.3x3) +
      s(DNF.3x3) + s(MF.3x3) + s(S.3x3) + s(WS.3x3) +
      s(CS.3x3) + s(EBF.3x3) + s(ef.elev.3x3), data=subset(cal.data,
      r.switch==1))
```

In command C34 and C35 a new column with a rainfall switch defining rainfall (1) or no-rainfall (0) state is added to the calibration data-frame: "cal.data". All days with daily precipitation magnitude greater than 0.1 mm is allotted a value of 1 in the column "r.switch" whereas all days with precipitation magnitudes less than or equal to 0.1 mm are allotted a value of 0. Command C36 calibrates the logistic regression model for SP model using "r.switch" as the predictant variable and columns "P.NARR", "elev," and "LC" as predictor variables. In R, logistic regression can be performed within the generalized linear modelling (glm) framework. A glm framework in R can model regression functions of many families including gaussian, binomial, poisson, gamma, inverse gaussian, and quasi. For defining a logistic regression, we specify the family of the regression function as "binomial" and link function as "logit" (not specified above as it is the default link function for binomial family in R). Command C37 calibrates a GAM model on rainfall intensities only using data for days when the value of "r.switch" is equal to 1 (in other words only using data for wet days). Commands C38 and C39 perform similar calibration of logistic regression and GAM models for the SPS3x3 downscaling model. The only difference is that in this case all neighborhood predictors corresponding to the scale 3x3 are also used in defining the regression models.

Prediction of downscaled precipitation using SP and SPS models involves first predicting dry and wet-day sequences and secondly predicting wet precipitation magnitudes. This can be performed in R using passing following set of commands:

```
C40 > colnames(pred.data)[which(colnames(pred.data) %in%
      "GCM.ppt")]="P.NARR"
C41 > pred.data$r.switch=as.numeric(predict(occ.mod.SP.P,
      newdata=pred.data, type="response"))
C42 > pred.data$r.switch[which(pred.data$r.switch>=0.5)]=1
C43 > pred.data$r.switch[which(pred.data$r.switch<0.5)]=0
C44 > pred.data$SP.pred.P=0
C45 > pred.data$SP.pred.P[which(pred.data$r.switch>=0.5)] =
      as.numeric(predict(int.mod.SP.P, newdata = subset(pred.data,
      r.switch> = 0.5)))
C46 > pred.data$r.switch=as.numeric(predict(occ.mod.SPS3x3.P,
      newdata=pred.data, type="response"))
C47 > pred.data$r.switch[which(pred.data$r.switch>=0.5)]=1
C48 > pred.data$r.switch[which(pred.data$r.switch<0.5)]=0
C49 > pred.data$SPS3x3.pred.P=0
C50 > pred.data$SPS3x3.pred.P[which(pred.data$r.switch>=0.5)] =
      as.numeric(predict(int.mod.SPS3x3.P, newdata = sub
      set(pred.data, r.switch> = 0.5)))
```

The command C40 again renames the column “GCM.ppt” to “P.NARR” for prediction to ensure that the predictors chosen to calibrate the occurrence and precipitation intensity models are also present in the prediction dataset “pred.data.” The command C41 predicts the probabilities of a particular day to be rainy given the values of predictors “P.NARR”, “LC”, and “elev” for each day. Next, in lines C42 and C43 we choose a probability threshold that can be used to decide on the predictant state given the predicted probabilities. In Gaur and Simonovic (2018) this has been calibrated for different models. For this demonstration a value of 0.5 is chosen as the probability threshold value above which the predictant, that is, “r.switch” is considered as having a rainy (1) state or else it is considered to have a nonrainy (0) state. In lines C44 and C45 the values of precipitation for days with “r.switch” values equals to zero are taken to be zero. For days with “r.switch” equals to 1 (or for rainy or wet days) the precipitation magnitude is predicted using the model “int.mod.SP.P.”

Commands C46 to C50 perform precipitation downscaling using SPS3x3 model using a similar process but using the logistic regression model "occ.mod.SPS3x3.P" and wet-day precipitation magnitude prediction GAM model "int.mod.SPS3x3.P."

Finally the results can be visualized effectively using several useful functions available in the "ggplot2" package. The function "melt" in package "reshape2" is also extremely useful for preparing data to be used in "ggplot2" package. A small example is presented here where yearly maximums of GCM and downscaled precipitation data (from models SP and SPS3x3) stored in the "pred.data" data-frame are plotted using "ggplot2" package.

```
C51 > ymax.P.GCM = sapply(2014:2100, function(x) max(subset
      (pred.data, year(date) %in% x)$P.NARR, na.rm=T))
C52 > ymax.P.SP = sapply(2014:2100, function(x) max(subset
      (pred.data, year(date) %in% x)$SP.pred.P, na.rm=T))
C53 > ymax.P.SPS3x3 = sapply(2014:2100, function(x) max(subset
      (pred.data, year(date) %in% x)$SPS3x3.pred.P, na.rm=T))
C54 > ymax.P=data.frame(year=2014:2100, GCM=ymax.P.GCM,
      SP=ymax.P.SP, SPS3x3=ymax.P.SPS3x3)
C55 > data.plot = melt(ymax.P, id="year")
C56 > ggplot()+geom_line(data=data.plot,aes(x=year,y=value,
      group=variable),size=1)+
              geom_point(data=data.plot,aes(x=year,y=value),
              size=3)+
              facet_wrap(~variable, scales="free")+theme_bw()+
              xlab("Year")+ylab("Precipitation(mm)")+
              theme(legend.title = element_text(size=22,
              face = "italic"),legend.position="bottom")+
              theme(axis.title.y=element_text(face="bold",
              size=22),
                    axis.title.x=element_text(face="bold",
                    size=22),
                    text=element_text(size=22))+
                         theme(axis.text.x=element_text(size=18),
              axis.text.y=element_text(size=18))+theme(legend.
              key.width = unit(3,"cm"))
C57 > ggsave("Fakepath/Sample_plot.png",width=14,height=10)
```

The extraction of yearly maximum precipitation as simulated by GCMs and downscaling models SP and SPS3x3 is performed in commands C51, C52, and C53, respectively. Next the yearly maximum results are aggregated into one data-frame "ymax.P". Thereafter the wide-format data-frame with four columns is converted into a long-format data-frame "ymax.P" with three columns using a function "melt" from "reshape2" package.

The long-format data-frame "data.plot" has three columns named "year", "variable", and "value." These three columns store the year values, that is, 2014 to 2100, variable values, that is, GCM, SP, and SPS3x3, and value of precipitation maximums (in mm) corresponding to each combination of year and variable name. The function ggplot() which we are using to plot the graphs needs an input data-frame in the long-format in order to do the plotting.

The data-frame "data.plot" is used to plot yearly maximum precipitation from GCMs, and SP, SPS3x3 downscaling models using the ggplot() function. As explained before, ggplot2 package is built on the grammar of graphics. It can be noted from command C56 that and data, aesthetic mapping, geometric object, statistical transformations, scales, coordinate system, position adjustments, and faceting arguments are passed along with the ggplot() function. The output generated from this command is shown in Fig. 4.4. In the plot, the three panels show annual maximum precipitation values for GCM, SP, and SPS3x3-based precipitation projections. The generated plot can be saved by using the ggsave() function in C57, where among other arguments, the location where file needs to be saved, file-type, plot dimensions, etc., are specified to save the plot in the system.

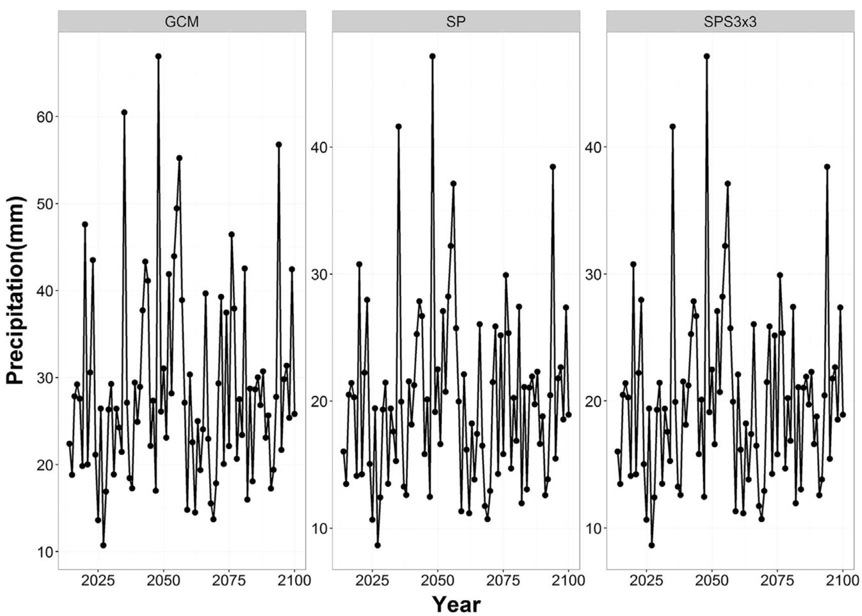

Figure 4.4 Annual precipitation maximum magnitudes (in mm) as plotted by the ggplot() function in R.

8. CONCLUSIONS AND FUTURE WORK

This discussion provided in this chapter is centered on a novel statistical downscaling model: physical scaling (SP) and its extension: SPS model. The motivation behind the development of the model and the theoretical research gap that this model fills is discussed first. The proposed model aims to bridge the gap between dynamic and statistical downscaling by building on the positives of both approaches. The state-of-the-art statistical downscaling models do not account for regional geophysical characteristics when performing downscaling. For this reason, climatic effects of geophysical characteristics as well as of changes in those characteristics can't be quantified currently within a statistical downscaling framework.

This limitation is overcome in SP and SPS models by considering covariates which represent regional geophysical characteristics like elevation and land-cover and their distribution in the model definition. The predictor and predictant variables are linked using a GAM function which is fully nonparametric in nature and can be used to link a range of predictor and predictant variables. Temperature downscaling by SP model is performed in one step by linking locally observed temperature with model-based temperature, land-cover, and elevation of the location of interest. In case of precipitation, downscaling is performed by formulating a logistic regression function which models wet-dry day precipitation sequences, as well as a GAM regression function that models precipitation magnitudes on wet days using model-based precipitation, land-cover, and elevation of the location of interest as predictors. In SPS models, land-cover and elevation characteristics of the location of interest, as well as its neighborhoods, are incorporated into the model formulation. This is done at a chosen neighborhood scale which has been taken as 3x3, 5x5, 7x7, and 9x9 in previous studies.

Both SP and SPS models have been extensively validated for their skill to downscale NARR temperature and precipitation gridded data in previous studies (Gaur and Simonovic, 2016a,b, 2018). This has been done both by quantifying model error over the historical time-period, as well as by comparing its performance with other state-of-the-art downscaling models like SDSM and GLM. The results from validation studies have been summarized in this chapter. They clearly show that the developed SP and SPS models are reliable downscaling tools that can be used to downscale GCM projections in southern Saskatchewan region of Canada.

Apart from being a fast, reliable, and accurate downscaling model, the major advantage of SP and SPS models is that they can be used to identify climatic impacts of future geophysical changes. In this chapter results from a case-study which quantified hydroclimatic impacts of future land-cover change at four catchments in southern Saskatchewan region of Canada are also presented and discussed. The case-study clearly demonstrates a sample analysis that can be performed using SP and SPS models that is not possible using other traditional statistical downscaling models.

Finally, a detailed presentation of how SP and SPS models can be formulated and used using an open source R programming language is made. To make the tutorial accessible for early users of R, an overview of R and its key components has also been provided. The tutorial then presents commands to read climate model—based atmospheric and climatic data from .nc files, and MODIS remotely sensed data directly from R. The tutorial then presents commands and discussion to calibrate SP and SPS models in R, and to use them for downscaling future precipitation and temperature projections from a representative GCM. Finally R is used to plot and compare downscaled future precipitation results with the raw GCM projections. From the tutorial it can be seen that SP and SPS models are very easily implementable in an open source platform and therefore they can be easily used by the scientific community across the globe without any additional costs.

The presented research can be extended in many possible directions in future. SP method can be modified to account for snow-cover during the winter months. Snow-cover is a very important physical parameter which affects the climatology and hydrology of any region and is especially relevant in the Canadian context. An appropriate spatial scale for the calibration of SP method needs to be ascertained. The spatial scale chosen should be a compromise between accuracy and robustness of the downscaled results. The applicability of SP method can be evaluated in other regions of Canada and the globe. It will be interesting to see how the model performs in regions that have more complex physiography than the region considered in this study. Further, case studies can be performed on catchments located in other regions of the globe to better understand the underlying dynamics of the land-cover change driven climatic changes. It will be interesting to compare land-cover change driven hydrologic changes with land-cover change driven climatic changes which in turn affects catchment hydrology. The study can be performed at catchments located in different climatic regions and biomes to compare and contrast the results.

REFERENCES

Abebe, M.T., Megento, T.L., 2016. The City Of Addis Ababa From 'Forest City' To 'Urban Heat Island'. Assessment Of Urban Green Space Dynamics, J. Urban Environ. Eng. 10 (2), 254–262.

Adachi, S.A., Kimura, F., Kusaka, H., Inoue, T., Ueda, H., 2012. Comparison of the impact of global climate changes and urbanization on summertime future climate in the Tokyo Metropolitan Area. J. Appl. Meteorol. Climatol. 51, 1441–1454.

Anandhi, A., Frei, A., Pierson, D.C., Schneiderman, E.M., Zion, M.S., Lounsbury, D., Matonse, A.H., 2011. Examination of change factor methodologies for climate change impact assessment. Water Resour. Res. 47 (3), W03501.

Apipattanavis, S., Podesta, G., Rajagopalan, B., Katz, R.W., 2007. A semiparametric multivariate and multisite weather generator. Water Resour. Res. 43 (11), W11401.

Argueso, D., Evans, J.P., Fita, L., Bormann, K.J., 2014. Temperature response to future urbanization and climate change. Clim. Dyn. 42 (7-8), 2183–2199.

Barradas, V.L., 1991. Air temperature and humidity and human comfort index of some city parks of Mexico City. Int. J. Biometeorol. 35 (1), 24–28.

Beecham, S., Rashid, M., Chowdhury, R.K., 2014. Statistical downscaling of multi-site daily rainfall in a South Australian catchment using a Generalized Linear Model. Int. J. Climatol. 34, 3654–3670.

Bellone, E., Hughes, J.P., Guttorp, P., 2000. A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. Clim. Res. 15, 1–12.

Benestad, R., Hanssen-Bauer, I., Chen, D., 2008. Empirical-Statistical Downscaling. World Scientific Publishing Company, Singapore, p. 228.

Bergant, K., Kajfez-Bogataj, L., 2005. N-PLS regression as empirical downscaling tool in climate change studies. Theor. Appl. Climatol. 81, 11–23.

Best, M.J., Grimmond, C.S.B., Villani, M.G., 2006. Evaluation of the urban tile in MOSES using surface energy balance observations. Bound. Lay. Meteorol. 118, 503–525.

Boucher, O., Myhre, G., Myhre, A., 2004. Direct human influence of irrigation on atmospheric water vapor and climate. Clim. Dyn. 22, 597–603.

Bounoua, L., Zhang, P., Mostovoy, G., Thome, K., Masek, J., Imhoff, M., Shepherd, M., Quattrochi, D., Santanello, J., Silva, J., 2015. Impact of urbanization on US surface climate. Environ. Res. Lett. 10, 101001.

Brunsdon, C., McClatchey, J., Unwin, D.J., 2001. Spatial variations in the average rainfall-altitude relationship in Great Britain: an approach using geographically weighted regression. Int. J. Climatol. 21, 455–466.

Burnash, R.J.C., 1995. The NWS river Forecast system – catchment modeling. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology, Revised edition. Water Resources Publications, Highlands Ranch, Colo. c1995. http://www.wrpllc.com/books/cmwh.html

Campra, P., Carcia, M., Canton, Y., Palacios-Orueta, A., 2008. Surface greenhouse cooling trends and negative radiative forcing due to land use change toward greenhouse farming in southeastern Spain. J. Geophys. Res. 113, D18109.

Cao, Q., Yu, D., Georgescu, M., Wu, J., 2016. Impacts of urbanization on summer climate in China: an assessment with coupled land-atmospheric modeling. J. Geophys. Res. Atmos. 121, 10505–10521.

Cavazos, T., Hewitson, B.C., 2005. Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. Clim. Res. 28, 95–107.

Charles, S.P., Bari, M.A., Kitsios, A., Bates, B.C., 2007. Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, western Australia. Int. J. Climatol. 27, 1673–1690.

Chang, C.R., Li, M.H., Chang, S.D., 2007. A preliminary study on the local cool island intensity of Taipei city parks. Landsc. Urban Plan. 80 (4), 386–395.
Chen, L., Frauenfeld, O.W., 2016. Impacts of urbanization on future climate in China. Clim. Dyn. 47 (1-2), 345–357.
Chen, Y., Wong, N.H., 2006. Thermal benefits of city parks. Energy Build. 38, 105–120.
Chiew, F.H.S., Kirono, D.G.C., Kent, D.M., Frost, A.J., Charles, S.P., Timbal, B., Nguyen, K.C., Fu, G., 2010. Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates. J. Hydrol. 387 (1–2), 10–23.
Clarke, L.E., Edmonds, J.A., Jacoby, H.D., Pitcher, H., Reilly, J.M., Richels, R., 2007. Scenarios of greenhouse gas emissions and atmospheric concentrations. Sub-report 2.1a of Synthesis and Assessment Product 2.1. In: Climate Change Science Program and the Subcommittee on Global Change Research, Washington DC.
Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J., Wehner, M., 2013. Long-term climate change: projections, commitments and irreversibility. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Coulibaly, P., 2004. Downscaling daily extreme temperatures with genetic programming. Geophys. Res. Lett. 31, L16203.
Craigmile, P.F., Guttorp, P., 2011. Space-time modelling of trends in temperature series. J. Time Anal. 32, 378–395.
De, U.S., Rao, G.S.P., 2004. Urban climate trends - the Indian scenario. J. Ind. Geophys. Union 8, 199–203.
Dibike, Y.B., Coulibaly, P., 2005. Hydrologic impact of climate change in the Saugenay Watershed: comparison of downscaling methods and hydrologic models. J. Hydrol. 307, 144–163.
Diem, J.E., Mote, T.L., 2005. Interepochal changes in summer precipitation in the southeastern United States: evidence of possible urban effects near Atlanta, Georgia. J. Appl. Meteorol. 44, 717–730.
Dixon, P.G., Mote, T.L., 2003. Patterns and causes of Atlanta's urban heat island-initiated precipitation. J. Appl. Meteorol. 42, 1273–1284.
Duan, Q., Gupta, V.K., Sorooshian, S., 1993. A shuffled complex evolution approach for effective and efficient optimization. J. Optim. Theory Appl. 76 (3), 501–521.
Emmanuel, R., Johansson, E., 2006. Influence of urban morphology and sea breeze on hot humid microclimate: the case of Colombo, Sri Lanka. Clim. Res. 30, 189–200.
Eum, H.-I., Simonovic, S.P., 2012. Assessment on variability of extreme climatic events for the upper Thames river basin in Canada. Hydrol. Process. 26 (4), 485–499.
Fahey, B., Jackson, R.B., 1997. Hydrological impacts of converting native forests and grasslands to pine plantations, South Island, New Zealand. Agric. Forest Meteorol. 84, 69–82.
Fall, S., Niyogi, D., Pielke, R., Gluhovsky, A., Kalnay, E., Rochon, G., 2010a. Impacts of land use land cover on temperature trends over the continental United States: assessment using the North American Regional Reanalysis. Int. J. Climatol. 30, 1980–1993.
Fall, S., Diffenbaugh, N., Niyogi, D., Peilke Sr, R., Rochon, G., 2010b. Temperature and equivalent temperature over the United States (1979–2005). Int. J. Climatol. 30, 2045–2054.
Fast, J.D., Torcolini, J.C., Redman, R., 2005. Pseudovertical temperature profiles and the urban heat island measured by a temperature datalogger network in Phoenix, Arizona. J. Appl. Meteorol. 44, 3–13.

Fealy, R., Sweeney, J., 2007. Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalized linear modeling approach. Int. J. Climatol. 27, 2083–2094.
Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Review: linking climate change modeling to impacts studies: recent advances in downscaling techniques for hydrological modeling. Int. J. Climatol. 27, 147–178.
Friederichs, P., Hense, A., 2007. Statistical downscaling of extreme precipitation events using censored quantile regression. Mon. Weather Rev. 135, 2365–2378.
Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.
Frost, A.J., Charles, S.P., Timbal, B., Chiew, F.H.S., Mehrotra, R., Nguyen, K.C., Chandler, R.E., McGregor, J.L., Fu, G., Kirono, D.G.C., Fernandez, E., Kent, D.M., 2011. A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. J. Hydrol. 408 (1–2), 1–18.
Fujibe, F., Togawa, H., Sakata, M., 2009. Long-term change and spatial anomaly of warm season afternoon precipitation in Tokyo. Sci. Online Lett. Atmos. 5, 17–20.
Fujino, J., Nair, R., Kainuma, M., Masui, T., Matsuoka, Y., 2006. Multigas mitigation analysis on stabilization scenarios using aim global model. Energy J. 343–354. Special issue #3.
Gaur, A., Simonovic, S.P., 2015. Towards reducing climate change impact assessment process uncertainty. Environ. Process. 2 (2), 275–290.
Gaur, A., Simonovic, S.P., 2016a. Accessing vulnerability of land-cover types to climate change using physical scaling downscaling model. Int. J. Climatol. https://doi.org/10.1002/joc.4887.
Gaur, A., Simonovic, S.P., 2016b. Extension of Physical Scaling method and its application towards downscaling climate model based near surface air temperature. Int. J. Climatol. https://doi.org/10.1002/joc.4921.
Gaur, A., Simonovic, S.P., 2018. Application of physical scaling towards downscaling climate model precipitation data. Theor. Appl. Climatol. 132 (1-2), 287–300.
Ge, J., 2010. MODIS observed impacts of intensive agriculture on surface temperature in the southern Great Plains. Int. J. Climatol. 30, 1994–2003.
Georgescu, M., et al., 2013. Summer-time climate impacts of projected megapolitan expansion in Arizona. Nat. Clim. Change 3, 37–41.
Ghosh, S., Mujumdar, P.P., 2008. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Adv. Water Resour. 31, 132–146.
Giorgi, F., Hurrell, J.W., Marinucci, M.R., Beniston, M., 1996. Elevation dependency of the surface climate change signal: a model study. J. Clim. 10, 288–296.
Giorgi, F., Hewitson, B.C., Christensen, J., Hulme, M., von Storch, H., Whetton, P., Jones, R., Mearns, L., Fu, C., 2001. Regional climate information – evaluation and projections. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J. (Eds.), Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, New York, USA, p. 881.
Grolemund, G., Wickham, H., 2011. Dates and times made easy with lubridate. J. Stat. Softw. 40 (3), 1–25.
Hale, R.C., Gallo, K.P., Owen, T.W., Loveland, T.R., 2006. Land use/land cover change effects on temperature trends at U.S. Climate Normals stations. Geophys. Res. Lett. 33, L11703.
Hale, R.C., Gallo, K.P., Loveland, T.R., 2008. Influences of specific land use/land cover conversions on climatological normals of near-surface temperature. J. Geophys. Res. 113, D14113.

Halfon, N., Levin, Z., Alpert, P., 2009. Temporal rainfall fluctuations in Israel and their possible link to urban and air pollution effects. Environ. Res. Lett. 4, 025001.
Hamdi, R., Van de Vyver, H., De Troch, R., Termonia, P., 2014. Assessment of three dynamical urban climate downscaling methods: Brussels's future urban heat island under an A1B emission scenario. Int. J. Climatol. 34, 978—999.
Han, J.Y., Baik, J.J., Lee, H., 2014. Urban impacts on precipitation. Asia-Pac. J. Atmos. Sci. 50 (1), 17—30.
Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Comparison of LARS-WG and SDSM for simulation and downscaling of extreme precipitation events in a Watershed. Stoch. Environ. Res. Risk Assess. 25, 475—484.
Hertig, E., Jacobeit, J., 2008. Downscaling future climate change: temperature scenarios for the Mediterranean area. Global Planet. Change 63, 127—131.
Hewitson, B.C., Crane, R.G., 1996. Climate downscaling: techniques and application. Clim. Res. 7, 85—95.
Hijioka, Y., Matsuoka, Y., Nishimoto, H., Masui, T., Kainuma, M., 2008. Global GHG emission scenarios under GHG concentration stabilization targets. J. Global Environ. Eng. 13, 97—108.
Hijmans, R.J., 2015. Raster: Geographic Data Analysis and Modeling. R package version 2.5-2. https://CRAN.R-project.org/package=raster.
Hope, P., Timbal, B., Fawcett, R., 2010. Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time. Int. J. Climatol. 30, 1360—1371.
Hopkinson, R.F., McKenney, D.W., Milewska, E.J., Hutchinson, M.F., Papadopol, P., Vincent, L.A., 2011. Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada. J. Appl. Meteorol. Climatol. 50, 1654—1665.
Houghton, R.A., 1999. The annual net flux of carbon to the atmosphere from changes in land use 1850—1990. Tellus 51B, 298—313.
Hua, W., Chen, H., Sun, S., Zhou, L., 2015. Assessing climatic impacts of future land use and land cover change projected with the CanESM2 model. Int. J. Climatol. 35, 3661—3675.
Hughes, J.P., Guttorp, P., 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour. Res. 30, 1535—1546.
Hughes, J., Guttorp, P., Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. Appl. Stat. 48, 15—30.
Hurtt, G.C., Frolking, S., Fearon, M.G., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S.W., Houghton, R.A., 2006. The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood harvest activity, and resulting secondary lands. Global Change Biol. 12, 1208—1229.
Hurtt, G.C., et al., 2011. Harmonization of land-use scenarios for the period 1500—2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Clim. Change 109 (1—2), 117—161.
Hutchinson, M.F., McKenney, D.W., Lawrence, K., Pedlar, J.H., Hopkinson, R.F., Milewska, E., Papadopol, P., 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961—2003. J. Appl. Meteorol. Climatol. 48, 725—741.
Huth, R., 2002. Statistical downscaling of daily temperature in central Europe. J. Clim. 15, 1731—1742.
Inoue, T., Kimura, F., 2004. Urban effects on low-level clouds around the Tokyo metropolitan area on clear summer days. Geophys. Res. Lett. 31, L05103.

Jackson, R.B., Canadell, J., Ehleringer, J.R., Mooney, H.A., Sala, O.E., et al., 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108, 389–411.
Jansson, C., Jansson, P.E., Gustafsson, D., 2007. Near surface climate in an urban vegetated park and its surroundings. Theor. Appl. Climatol. 89 (3–4), 185–193.
Jarvis A., Reuter H. I., Nelson A., Guevara E., Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org), 2008.
Jauregui, E., 1991. Influence of a large urban park on temperature and convective precipitation in a tropical city. Energy Build. 15 (3), 457–463.
Jonsson, P., 2004. Vegetation as an urban climate control in the subtropical city of Gaborone, Botswana. Int. J. Climatol. 24 (10), 1307–1322.
Kigobe, M., McIntyre, N., Wheater, H., Chandler, R., 2011. Multi-site stochastic modelling of daily rainfall in Uganda. Hydrol. Sci. J. 56 (1), 17–33.
Kikon, N., Singh, P., Singh, S.K., Vyas, A., 2016. Assessment of urban heat islands (UHI) of Noida City, India using multi-temporal satellite data. Sustain. Cities Soc. 22, 19–28.
King, L.M., Irwin, S., Sarwar, R., McLeod, A.I., Simonovic, S.P., 2012. The effects of climate change on extreme precipitation events in the upper Thames river basin: a comparison of downscaling approaches. Can. Water Resour. J. 37 (3), 253–274.
King, L.M., McLeod, I., Simonovic, S.P., 2016. Improved weather generator algorithm for multisite simulation of precipitation and temperature. J. Am. Water Resour. Assoc. 51 (5), 1305–1320.
Kishtawal, C.M., Niyogi, D., Tewari, M., Pielke Sr., R.A., Shepherd, J.M., 2010. Urbanization signature in the observed heavy rainfall climatology over India. Int. J. Climatol. 30, 1908–1916.
Kruger, E., Givoni, B., 2007. Outdoor measurements and temperature comparisons of seven monitoring stations: preliminary studies in Curitiba, Brazil. Build. Environ. 42 (4), 1685–1698.
Kuchar, L., 2004. Using WGENK to generate synthetic daily weather data for modelling of agricultural processes. Math. Comput. Simulat. 65, 69–75.
Kug, J.-S., Ahn, M.S., 2013. Impact of urbanization on recent temperature and precipitation trends in the Korean peninsula. Asia-Pac. J. Atmos. Sci. 49, 151–159.
Kumar, S., Dirmeyer, P.A., Merwade, V., DelSole, T., Adams, J.M., Niyogi, D., 2013. Land use/cover change impacts in CMIP5 climate simulations: a new methodology and 21st century challenges. J. Geophys. Res. Atmos. 118, 6337–6353.
Kusaka, H., Hara, M., Takane, Y., 2012. Urban Climate Projection by the WRF Model at 3-km Horizontal Grid Increment: Dynamical Downscaling and Predicting Heat Stress in the 2070's August for Tokyo, Osaka, and Nagoya Metropolises. J. Meteorol. Soc. Jpn. 90B, 47–63.
Kwon, H.H., Sivakumar, B., Moon, Y.I., Kim, B.S., 2011. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. Stoch. Environ. Res. Risk Assess. 25, 567–581.
Lawrence, P.J., Chase, T.N., 2010. Investigating the climate impacts of global land cover change in the Community Climate System Model (CCSM). Int. J. Climatol. 30, 2066–2087.
Lee, E., Chase, T.N., Rajagopalan, B., Barry, R.G., Biggs, T.W., Lawrence, P.J., 2009. Effects of irrigation and vegetation activity on early Indian summer monsoon variability. Int. J. Climatol. 29, 573–581.
Lee, T-c., Chan, K-y, Chan, H-s, Kok, M-h, 2011. Projections of extreme rainfall in Hong Kong in the 21st century. Acta Meteorol. Sin. 25 (6), 691–709.
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.
Li, Y., Smith, I., 2009. A statistical downscaling model for southern Australia winter rainfall. J. Clim. 22, 1142–1158.

Li, Z., He, Y., Theakstone, W.H., Wang, X., Zhang, W., Cao, W., Du, J., Xin, H., Chang, L., 2012. Altitude dependency of trends of daily climate extremes in southwestern China, 1961–2008. J. Geogr. Sci. 22, 416.

Li, X., Chen, H., Liao, H., Hua, W., Sun, S., Ma, H., Li, X., Gao, C., Zhu, S., 2017. Potential effects of land cover change on temperature extremes over Eurasia: current versus historical experiments. Int. J. Climatol. https://doi.org/10.1002/joc.4976.

Li, X., Jiang, F., Li, L., Wang, G., 2011. Spatial and temporal variability of precipitation concentration index, concentration degree and concentration period in Xinjiang, China. Int. J. Climatol. 31, 1679–1693.

Liu, X., Cheng, Z., Yan, L., Yin, Z.-Y., 2009. Elevation dependency of recent and future minimum surface air temperature trends in the Tibetan Plateau and its surroundings. Global Planet. Change 68 (3), 164–174.

Liu, Z., Xu, Z., Charles, S.P., Fu, G., Liu, L., 2011. Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. Int. J. Climatol. 31 (13), 2006–2020.

Lowry, W.P., 1977. Empirical estimation of the urban effects on climate: a problem analysis. J. Appl. Meteorol. 16, 129–135.

Lydolph, P.E., 1985. The Climate of the Earth. Rowman and Littlefield Publishers Inc.

Mahmood, R., Foster, S.A., Keeling, T., Hubbard, K.B., Carlson, C., Leeper, R., 2006. Impacts of irrigation on 20th century temperature in the northern Great Plains. Global Planet. Change 54, 1–18.

Malyshev, S., Shevliakova, E., Stouffer, R.J., Pacala, S.W., 2015. Contrasting local versus regional effects of land-use-change-induced heterogeneity on historical climate: analysis with the GFDL earth system model. J. Clim. 28, 5448–5469.

Manzanas, R., Brands, S., San-Martin, D., Lucero, A., Limbo, C., Gutierrez, J.M., 2015. Statistical downscaling in the tropics can be sensitive to reanalysis choice: a case study for precipitation in the Philippines. J. Clim. 28, 4171–4184.

Masson, D., Knutti, R., 2011. Spatial-scale dependence of climate model performance in the CMIP3 ensemble. J. Clim. 24, 2680–2692.

Mattiuzzi, M., 2016. MODIS: MODIS Acquisition and Processing. R package version 0.10-34/r510. https://R-Forge.R-project.org/projects/modis/.

McCarthy, M.P., Harpham, C., Goodess, C.M., Jones, P.D., 2012. Simulating climate change in UK cities using a regional climate model, HadRM3. Int. J. Climatol. 32 (12), 1875–1888.

McPherson, R.A., Stensrud, D.J., Crawford, K.C., 2004. The impact of Oklahoma's winter wheat belt on the mesoscale environment. Mon. Weather Rev. 132, 405–422.

Mesinger, F., et al., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360.

Mote, T.L., Lacke, M.C., Shepherd, J.M., 2007. Radar signatures of the urban effect on precipitation distribution: a case study for Atlanta, Georgia. Geophys. Res. Lett. 34, L20710.

Neale R. B., et al., Description of the NCAR Community Atmosphere Model CAM 5.0, NCAR Tech. Note NCAR/TN-486+STR, 2012, 289pp., Natl. Cent. For Atmos. Res., Boulder, Colo.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.

Neset, T.-S., Opach, T., Lion, P., Lilja, A., Johansson, J., 2016. Map-based web tools supporting climate change adaptation. Prof. Geogr. 68 (1), 103–114.

Niyogi, D., Kishtawal, C., Tripathi, S., Govindaraju, R.S., 2010. Observational evidence that agricultural intensification and land use change may be reducing the Indian summer monsoon rainfall. Water Resour. Res. 46, W03533.

Nosetto, M.D., Jobbagy, E.G., Paruelo, J.M., 2005. Land-use change and water losses: the case of grassland afforestation across a soil textural gradient in central Argentina. Global Change Biol. 11, 1101–1117.

Offerle, B., et al., 2005. Urban modification of the surface energy balance in the West African Sahel: Ouagadougou, Burkina Faso. J. Clim. 18, 3983–3995.

Oke, T.R., 1982. The energetic basis of the urban heat island. Quart. J. R. Meteorol. Soc. 108, 1–24.

Oke, T.R., Crowther, J.M., McNaughton, K.G., Monteith, J.L., Gardiner, B., 1989. The micrometeorology of the urban forest [and discussion]. Philos. Trans. R. Soc. Lond. B 324 (1223), 335–349.

Pan, Y., Jackson, R.T., 2008. Ethnic difference in the relationship between acute inflammation and serum ferritin in US adult males. Epidemiol. Infect. 136, 421–431.

Pepin, N., Bradley, R.S., Dias, H.F., et al., 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Change 5, 424–430.

Pielke, R.A., Adegoke, J., Beltran-Przekurat, A., Hiemstra, C.A., Lin, J., Nair, U.S., Niyogi, D., Nobis, T.E., 2006. An overview of regional land-use and land-cover impacts on rainfall. Tellus 59B, 587–601.

Pitman, A.J., 2003. The evolution of, and revolution in, land surface schemes designed for climate models. Int. J. Climatol. 23, 479–510.

Pitman, A.J., Narisma, G.T., 2005. The role of land surface processes in regional climate change: a case study of future land cover change over south western Australia. Meteorol. Atmos. Phys. 89, 235–249.

Puvaneswaran, K.M., Smithson, P.A., 1991. Precipitation — elevation relationships over Sri Lanka. Theor. Appl. Climatol. 43, 113.

Qiao, Z., Guangjin, T., Lixiao, Z., Xu, X., 2014. Influences of urban expansion on urban heat island in Beijing during 1989–2010. Adv. Meteorol. 11, 187169.

Quesada, B., Devaraju, N., de Noblet-Ducoudré, N., Arneth, A., 2017. Reduction of monsoon rainfall in response to past and future land use and land cover changes. Geophys. Res. Lett. 44 https://doi.org/10.1002/2016GL070663.

Rao, V.B., Fernandez, J.P.R., Franchito, S.H., 2004. Quasi-stationary waves in the Southern Hemisphere during the El Niño and La Niña events. Ann. Geophys. 22, 789–806.

R Development Core Team, 2008. 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.

Räisänen, J., Ylhäisi, J.S., 2011. How much should climate model output be smoothed in space? J. Clim. 24, 867–880.

Riahi, K., Krey, V., Rao, S., Chirkov, V., Fischer, G., Kolp, P., Kindermann, G., Nakicenovic, N., Rafai, P., 2007. Exploring the consequence of high emission trajectories. Climatic Change 109, 133.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190.

Robertson, A.W., Kirshner, S., Smyth, P., 2004. Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. J. Clim. 17, 4407–4424.

Robertson, A.W., Ines, A.V.M., Hansen, J.W., 2007. Downscaling of seasonal precipitation for crop simulation. J. Appl. Meteorol. Climatol. 46, 677–693.

Rogerson, P.A., 2001. Statistical Methods for Geography. Sage, London.

Roy, S.S., Mahmood, R., Niyogi, D., Lei, M., Foster, S.A., Hubbard, K.G., Douglas, E., Pielke Sr., R., 2007. Impacts of the agricultural Green Revolution–induced land use changes on air temperatures in India. J. Geophys. Res. 112, D21108.

Salathé, E.P., 2003. Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. Int. J. Climatol. 23, 887–901.

Salathé, E.P., 2005. Downscaling simulations of future global climate with application to hydrologic modeling. Int. J. Climatol. 25, 419–436.
Samuels, R., Rimmer, A., Alpert, P., 2009. Effect of extreme rainfall events on the water resources of the Jordan River. J. Hydrol. 375 (3–4), 513–523.
Savic, D.A., Morley, M.S., Khoury, M., 2016. Serious gaming for water systems planning and management. Water 8 (10), 456.
Schlünzen, K.H., Hoffmann, P., Rosenhagen, G., Riecke, W., 2010. Long-term changes and regional differences in temperature and precipitation in the metropolitan area of Hamburg. Int. J. Climatol. 30, 1121–1136.
Schoof, J.T., Pryor, S.C., Robeson, S.M., 2007. Downscaling daily maximum and minimum air temperatures in the midwestern USA: a hybrid empirical approach. Int. J. Climatol. 27, 439–454.
Schoof, J.T., Pryor, S.C., Surprenant, J., 2010. Development of daily precipitation scenarios for the United States based on probabilistic downscaling. J. Geophys. Res. 115, D13106.
Schoof, J.T., 2012. Scale issues in the development of future precipitation scenarios. J. Contemp. Water Res. Educ. 147, 8–16.
Schoof, J.T., 2013. Statistical downscaling in climatology. Geogr. Compass 7, 249–265.
Segal, M., Pan, Z., Turner, R.W., Takle, E.S., 1998. On the potential impact of irrigation areas in North America on summer rainfall caused by large-scale systems. J. Appl. Meteorol. 37, 325–331.
Sellers, P.J., 1992. Biophysical models of land surface processes. In: Trenberth, K.E. (Ed.), Climate System Modeling. Cambridge University Press, pp. 451–490.
Semenov, M.A., Barrow, E.M., 1997. Use of a stochastic weather generator in the development of climate change scenarios. Climatic Change 35, 397–414.
Sharif, M., Burn, D.H., 2006. Simulating climate change scenarios using an improved K-Nearest neighbor model. J. Hydrol. 325, 179–196.
Sharif, M., Burn, D.H., 2007. Improved K-nearest neighbour weather generating model. J. Hydrol. Eng. 12, 42–51.
Sharma, D., Das Gupta, A., Babel, M.S., 2007. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. Hydrol. Earth Syst. Sci. 11, 1373–1390.
Shepherd, J.M., Pierce, H., Negri, A.J., 2002. Rainfall modification by major urban areas: observations from spaceborne rain radar on the TRMM satellite. J. Appl. Meteorol. 41, 689–701.
Smith, S.J., Wigley, T.M.L., 2006. MultiGas forcing stabilization with minicam. Energy J. 373–392. Special issue #3.
Soltani, A., Hoogenboom, G., 2003. Minimum data requirements for parameter estimation of stochastic weather generators. Clim. Res. 25, 109–119.
Spronken-Smith, R.A., Oke, T.R., 1998. The thermal regime of urban parks in two cities with different summer climates. Int. J. Remote Sens. 19 (11), 2085–2104.
Srivastav, R.K., Simonovic, S.P., 2015. Multi-site, multivariate weather generator using maximum entropy bootstrap. Clim. Dyn. 44, 3431–3448.
Stocker, T.F., et al., 2013. Summary for policymakers. In: Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). J. R. Stat. Soc. B 36, 111–147.

Sun, C.Y., 2011. A street thermal environment study in summer by the mobile transect technique. Theor. Appl. Climatol. 106 (3–4), 433–442.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An Overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498.

Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J. Hydrol. 456–457, 12–29.

Timbal, B., Hope, P., Charles, S., 2008. Evaluating the consistency between statistically downscaled and global dynamical model climate change projections. J. Clim. 20, 6052–6059.

Timbal, B., Jones, D.A., 2008. Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. Climatic Change 86, 165–187.

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.

Tuck, S., Phillips, H., 2016. MODISTools: MODIS Subsetting Tools. R package version 0.95.0. http://cran.r-project.org/package=MODISTools.

Turner II, B.L., Clark, W.C., Kates, R.M., Richards, J.F., Mathews, J.T., Meyer, W.B. (Eds.), 1990. The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years. Cambridge Univ Press, Cambridge, p. 713.

Van Vuuren, D.P., Eickhout, B., Lucas, P.L., den Elzen, M.G.J., 2006. Long-term multi-gas scenarios to stabilise radiative forcing—exploring costs and benefits within an integrated assessment framework. Energy J. 27, 201–233.

Van Vuuren, D.P., Den Elzen, M.G.J., Lucas, P.L., Eickhout, B., Strengers, B.J., Van Ruijven, B., Wonink, S., Van Houdt, R., 2007. Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. Climatic Change 81, 119–159.

Van Vuuren, D.P., Edmonds, J., Kainuma, M., et al., 2011. The representative concentration pathways: an overview. Climatic Change 109, 5–31.

Velazquez-Lozada, A., Gonzalez, J.E., Winter, A., 2006. Urban heat island effect analysis for San Juan, Puerto Rico. Atmos. Environ. 40, 1731–1741.

Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climate. Remote Sens. Environ. 86, 370–384.

Walter, M.T., Brooks, E.S., McCool, D.K., King, L.G., Molnau, M., Boll, J., 2005. Process-based snowmelt modeling: does it require more input data than temperature-index modeling? J. Hydrol. 300 (1–4), 65–75.

Wan, Z., 2006. Collection-5 MODIS Land Surface Temperature Products Users Guide.

Wang, Y., Berardi, U., Akbari, H., 2015. The urban heat island effect in the city of Toronto. Procedia Eng. 118, 137–144.

Wickham, H., 2007. Reshaping data with the reshape package. J. Stat. Softw. 21 (12), 1–20.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Widmann, M., Bretherton, C.S., Salathé, E.P., 2003. Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor. J. Clim. 16, 799–816.

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. 21, 530–548.

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002. SDSM – a decision support tool for the assessment of regional climate change impacts. Environ. Model. Softw. 17 (2), 145–157.

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, Supporting Material of the Intergovernmental Panel on Climate Change. Available from: the DDC of IPCC TGCIA, 27.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics 1, 80–83.

Wilkinson, L., 2005. The grammar of graphics. In: Statistics and Computing, second ed. Springer, New York.

Wilks, D.S., 1999. Multisite downscaling of precipitation with a stochastic weather generator. Clim. Res. 11, 125–136.

Wilks, D.S., 2010. Use of stochastic weather generators for precipitation downscaling. WIREs Clim. Change 1 (6), 898–907.

Wilks, D.S., 2012. Stochastic weather generators for climate-change downscaling, part II: multivariable and spatially coherent multisite downscaling. WIREs Clim. Change 3, 267–278.

Wise, M., Calvin, K., Thomson, A., Clarke, L., Bond-Lamberty, B., Sands, R., Smith, S.J., Janetos, A., Edmonds, J., 2009. Implications of limiting CO2 concentrations for land use and energy. Science 324, 1183–1186.

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. B 62 (2), 413–428.

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change 62, 189–216.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. 73 (1), 3–36.

Wong, N.H., Yu, C., 2005. Study of green areas and urban heat island in a tropical city. Habitat Int. 29 (3), 547–558.

Xue, Y., Shukla, J., 1996. The influence of land surface properties on Sahel climate. Part II: Afforestation. J. Clim. 9, 3260–3275.

Yan, H., Fan, S., Guo, C., Hu, J., Dong, L., 2014. Quantifying the impact of land cover composition on intra-urban air temperature variations at a mid-latitude city. PLoS One 9 (7), e102124.

Yan, L., Liu, X., 2014. Has climatic warming over the Tibetan Plateau paused or continued in recent years? J. Earth Ocean Atmos. Sci. 1, 13–28.

Yokobori, T., Ohta, S., 2009. Effect of land cover on air temperatures involved in the development of an intra-urban heat island. Clim. Res. 39 (1), 61–73.

Zhang, P., Bounoua, L., Thome, K., Wolfe, R., 2016. Modeling impact of urbanization in us cities using simple biosphere model SiB2. In: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. https://doi.org/10.1109/IGARSS.2016.7730764.

Zhao, C., Jiang, Q., Sun, Z., Zhong, H., Lu, S., 2013. Projected urbanization impacts on surface climate and energy budgets in the Pearl River Delta of China. Adv. Meteorol. 542086.

Zong-Ci, Z., Yong, L., Jian-Bin, H., 2013. Are there impacts of urban heat island on future climate change? Adv. Clim. Change Res. 4 (2), 133–136.

Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Clim. 12, 2474–2489.

FURTHER READING

Cai, W., Santoso, A., Wang, G., Yeh, S.-W., An, S., Cobb, K.M., Collins, M., Guilyardi, E., Jin, F., Kug, J.-S., Lengaigne, M., McPhaden, M.J., Takahashi, K., Timmermann, A., Vecchi, G., Watanabe, M., Wu, L., 2015. ENSO and greenhouse warming. Nat. Clim. Change 5, 849–859.

Rizwan, A.M., Dennis, L.Y.C., Liu, C., 2008. A review on the generation, determination and mitigation of Urban Heat Island. J. Environ. Sci. 20 (1), 120–128.

Sitch, S., Brovkin, V., von Bloh, W., van Vuuren, D., Eickhout, B., Ganopolski, A., 2005. Impacts of future land cover changes on atmospheric CO2 and climate. Global Biogeochem. Cycles 19, GB2013.