

# Multi-site downscaling of maximum and minimum daily temperature using support vector machine

V. V. Srinivas,<sup>a</sup> Bidroha Basu,<sup>a</sup> D. Nagesh Kumar<sup>a\*</sup> and Sanjay K. Jain<sup>b</sup>

<sup>a</sup> Department of Civil Engineering, Indian Institute of Science, Bangalore, India

<sup>b</sup> National Institute of Hydrology, Roorkee, India

**ABSTRACT:** Climate change impact assessment studies involve downscaling large-scale atmospheric predictor variables (LSAPVs) simulated by general circulation models (GCMs) to site-scale meteorological variables. This article presents a least-square support vector machine (LS-SVM)-based methodology for multi-site downscaling of maximum and minimum daily temperature series. The methodology involves (1) delineation of sites in the study area into clusters based on correlation structure of predictands, (2) downscaling LSAPVs to monthly time series of predictands at a representative site identified in each of the clusters, (3) translation of the downscaled information in each cluster from the representative site to that at other sites using LS-SVM inter-site regression relationships, and (4) disaggregation of the information at each site from monthly to daily time scale using  $k$ -nearest neighbour disaggregation methodology. Effectiveness of the methodology is demonstrated by application to data pertaining to four sites in the catchment of Beas river basin, India. Simulations of Canadian coupled global climate model (CGCM3.1/T63) for four IPCC SRES scenarios namely A1B, A2, B1 and COMMIT were downscaled to future projections of the predictands in the study area. Comparison of results with those based on recently proposed multivariate multiple linear regression (MMLR) based downscaling method and multi-site multivariate statistical downscaling (MMSD) method indicate that the proposed method is promising and it can be considered as a feasible choice in statistical downscaling studies. The performance of the method in downscaling daily minimum temperature was found to be better when compared with that in downscaling daily maximum temperature. Results indicate an increase in annual average maximum and minimum temperatures at all the sites for A1B, A2 and B1 scenarios. The projected increment is high for A2 scenario, and it is followed by that for A1B, B1 and COMMIT scenarios. Projections, in general, indicated an increase in mean monthly maximum and minimum temperatures during January to February and October to December.

**KEY WORDS** downscaling; disaggregation; temperature; support vector machine; CGCM3; IPCC SRES scenarios; Hydroclimatology; Beas river basin

Received 11 August 2012; Revised 5 June 2013; Accepted 8 June 2013

## 1. Introduction

Recently, there is growing scientific evidence that global climate has changed over the past century because of anthropogenic factors. Human activities such as burning of fossil fuels, farming and deforestation have increased concentration of greenhouse gases in the earth's atmosphere, which absorb solar radiation reflected by the earth's surface. This has triggered rise in the average global temperature, which is more pronounced in the past few decades. Assessing implications of projected climate change on the vulnerable resources of earth would be useful to devise strategies for sustainable planning, management and conservation of the same.

Currently, numerical models called general circulation models (GCMs) are regarded as reliable tools available to simulate future climatic conditions on earth. Output

from a GCM is available at centres of grid boxes (usually of size  $>10\,000\text{ km}^2$ ) covering the earth's surface. The performance of GCMs in simulating coarse-scale atmospheric dynamics is reasonable. However, they fail to simulate climate variables at finer (e.g. at-site) scale that is of relevance for several practical applications. To overcome this shortcoming, downscaling methods gained recognition. Review of downscaling methods can be found in Wilby and Wigley (1997), Xu (1999), Fowler *et al.* (2007), Wilby *et al.* (2009) and Wilks (2012). A brief overview of various downscaling approaches available in literature is provided in the following section.

The conventional downscaling methods attempt to downscale GCM simulations to weather variables (e.g. rainfall, temperature, wind speed, humidity) at a single location. In real world scenario, spatial covariance structure is evident between records of different variables in a region. An effective downscaling model must be parsimonious and should ensure simulation of the covariance structure to a reasonable extent, though exact

\* Correspondence to: D. Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India. E-mail: nagesh@civil.iisc.ernet.in

reproduction/mimicking of the structure may not be necessary in a climate change (non-stationary) scenario. A few recent studies attempted to address this issue based on the assumption of spatiotemporal covariance stationarity. However, there is dearth of meaningful efforts especially for multi-site downscaling of temperature. Most of the work was focused to develop stochastic weather generators and to downscale precipitation. Conventional weather generators involve separately downscaling large-scale climate variables to one or more weather variables at each target location. Consequently historical/observed spatiotemporal covariance structure may not be evident in the downscaled weather variables. Wilks (1999) developed a procedure to use weather generator for simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at 62 stations in the western United States. Qian *et al.* (2002) applied the procedure to Guadiana river basin in Portugal for simultaneous simulation of daily precipitation, and daily maximum and minimum temperatures at six sites. Bardossy and Van Mierlo (2000) simultaneously downscaled daily mean temperature to several locations in Aller catchment, Germany, by conditioning the predictand on its previous day's value, circulation patterns and downscaled value of precipitation for the current day using multivariate autoregressive (MAR) approach. Hundecha and Bardossy (2008) investigated the potential of multiple linear regression (MLR) and MAR statistical downscaling models in terms of their ability to construct indices of extremes of daily precipitation and temperatures. The MLR model was used to downscale the predictands on the seasonal basis from seasonal measures of a set of large-scale weather variables, while MAR model involved construction of daily series of the predictands at multiple sites from large-scale atmospheric circulation patterns and other variables.

Khalili *et al.* (2011) have developed a multiple regression-based downscaling model for simulating daily maximum and minimum temperatures at various sites concurrently. Spatial moving average process was considered for modelling stochastic component of the model. For the case with  $n$  sites,  $m$  months and  $q$  predictor variables, this model requires  $m \times [2 \times n \times (q + 1) + (4n^2 + 1)]$  parameters, which would be 6252 for 10 sites, 12 months and 5 predictors. It is thus not parsimonious. Recently, Jeong *et al.* (2012) proposed a statistical downscaling model for simultaneous downscaling of daily maximum and minimum temperatures at multiple sites. It employs multivariate multiple linear regression (MMLR) to simulate deterministic series from large-scale reanalysis data and adds spatially correlated random series to the deterministic series to complement the underestimated variance, and to reproduce the spatial correlation structure of maximum and minimum temperatures, and at-site temporal correlation between maximum and minimum temperatures. In this study the models of Khalili *et al.* (2011) and Jeong *et al.* (2012) are henceforth referred to as multi-site multivariate statistical downscaling (MMSD) and MMLR respectively.

Practical tools available for downscaling single-site scenarios of daily surface weather variables under present and future climate forcing include statistical downscaling model (SDSM 4.2, Wilby and Dawson, 2007) and automated statistical downscaling (ASD, Hessami *et al.*, 2008). The SDSM is a hybrid of the stochastic weather generator and transfer function-based downscaling methods, whereas ASD is a hybrid of a stochastic weather generator and regression-based downscaling methods.

In essence, literature review reveals that currently there is a need to devise effective strategy for multi-site downscaling of weather variables. This study is motivated to develop a least-square support vector machine (LS-SVM) (Vapnik, 1995, 1998) based methodology for multi-site downscaling of maximum and minimum daily temperature series. The LS-SVM has been found to be effective in statistical downscaling of hydro-meteorological variables in earlier studies that focused on univariate downscaling to a single site (e.g. Tripathi *et al.*, 2006; Anandhi *et al.*, 2008, 2009, 2012; Srinivas and Tripathi, 2008).

The remainder of this article is organized as follows: An overview of downscaling approaches is provided in Section 2. The proposed LS-SVM methodology for multi-site downscaling of maximum and minimum daily temperatures is presented in Section 3 and methodologies for MMSD and MMLR models are also provided. Description of the study region and details concerning the data used for the study are provided in Section 4. Results from application of the proposed method to Bias river basin are presented and discussed in Section 5. The results are compared with those from MMSD and MMLR models. Finally summary and concluding remarks are presented in Section 6.

## 2. Overview of downscaling approaches

The approaches available for downscaling output of a GCM to target locations could be broadly classified as dynamic and statistical downscaling. In the dynamic downscaling approach a regional climate model (RCM) is embedded into GCM. The RCM is a numerical model having horizontal grid spacing of about 20–50 km, which is driven by initial conditions, time-dependent lateral meteorological conditions and surface boundary conditions. The time-varying atmospheric boundary conditions are specified by the host GCM. An advantage of RCM is its consistency with GCM, and ability to (1) simulate finer-scale atmospheric processes such as orographic precipitation, or low-level jets better than the host GCM and (2) respond in physically consistent ways to different external forcings (Wilby and Dawson, 2007). The shortcomings of RCM, which restrict its use in climate impact studies, are its complicated design and high computational cost. Output of a RCM will not be readily available at locations of user choice in a study area and it can have substantial errors (Wilks, 2010). Moreover, RCM is inflexible in the sense that expanding the region

or moving to a slightly different region requires redoing the entire experiment.

The statistical downscaling approach involves developing statistical relationships that transform large-scale atmospheric (climate) variables simulated by GCM to local (at-site) scale variables. The methods based on statistical downscaling approach can be classified into three categories: weather generators, weather typing or weather classification schemes and transfer function (Wilby *et al.*, 2004; Wilby and Dawson, 2007). In weather generator methods, future local daily weather scenarios are generated stochastically by modifying parameters of conventional weather generators (e.g. WGEN, LARS-WG or EARWIG) conditional on outputs from a host GCM corresponding to different climate change scenarios. Typically a primary variable (e.g. precipitation occurrence) is simulated using weather generator, and secondary variables (e.g. wet day precipitation amounts, temperatures and solar radiation) are modelled conditional on the primary variable. Shortcomings of this method include unanticipated effects on secondary variables associated with changes to parameters governing primary variables (Wilks, 1992, Wilby *et al.*, 2004), and low skill at reproducing inter-annual to decadal climate variability (Wilby and Dawson, 2007).

Weather typing methods involve grouping of local weather variables in relation to prevailing patterns of atmospheric circulation. Future local weather scenarios are constructed either by resampling from the observed variable distribution (conditioned on the circulation pattern produced by a GCM), or by first generating synthetic sequences of weather pattern using Monte Carlo techniques and then resampling from the observed data. These methods could yield physically interpretable linkages to surface climate. However, they could be insensitive to future climate forcing and may not capture intra-type variation in surface climate (Wilby *et al.*, 2004).

The transfer function method relies on the development of direct quantitative relationship between the local scale weather variable (predictand) and 'large-scale atmospheric variables (LSAVs) that affect the predictand' (predictors) through some form of regression. This method is computationally not as intensive as RCM, and easily adjustable to new areas. Hence it is widely used for downscaling by hydro-meteorologists. Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. Caution must be exercised in making the choice as it significantly affects results from this method (Wilby and Dawson, 2007). Examples of transfer functions that are used to develop predictor–predictand relationship include linear and nonlinear regression, artificial neural network, canonical correlation, least-square and standard SVMs (e.g. Tripathi *et al.*, 2006; Srinivas and Tripathi, 2008) and relevance vector machine (Ghosh and Mujumdar, 2008). LS-SVM provides a computational advantage over standard SVM (Suykens, 2001) and is thus found effective in downscaling studies (Anandhi *et al.*, 2008, 2009, 2012; Srinivas and Tripathi, 2008). In

this study, LS-SVM-based methodology is developed to obtain future daily projections of maximum and minimum temperatures at sites in a river basin.

### 3. Methodology

In this section, first the LS-SVM methodology proposed for multi-site downscaling of maximum and minimum temperatures is presented in Section 3.1. Following that methodologies for MMLR and MMSD models are briefly described in Sections 3.2. and 3.3., respectively.

#### 3.1. LS-SVM methodology for multi-site downscaling of maximum and minimum daily temperatures

The proposed methodology for multi-site downscaling involves the following steps.

1. Examine correlation structure of historical data of maximum and minimum temperatures (predictands) to identify a representative site for which predictand data are highly correlated with those at the other sites in the study area. If such a site cannot be located, delineate the available sites in the study area into clusters such that predictand data at each site in a cluster is highly correlated with those at other sites in the same cluster. Correlation between predictand data of sites in different clusters should be as less as possible. Identify a representative site in each cluster for which predictand data are highly correlated with those at the other sites in the cluster.
2. Identify predictor variables from LSAVs available in both observed/reanalysis data and GCM simulations, such that they are reasonably well correlated with the predictand data at the representative site.

Often observed data on LSAV are unavailable for target locations and hence reanalysis data are considered as their surrogate for the sake of analysis. Further, an issue associated with GCMs is that they run on a sub-daily time scale and the resulting simulations, though available at finer time scale (e.g., 6-hourly), are not considered reliable (Prudhomme *et al.*, 2002, p. 1138; Brown *et al.*, 2008, p. 20; Environment Canada: <http://www.cccsn.ec.gc.ca/?page=temporal-d>). The simulations available at coarser time scale are thus preferable. Another point to note is that the output from reanalysis and GCM are available at centres of grid boxes. Hence for the sake of this analysis, choose the spatial domain of each of the LSAVs as a set of grid points surrounding the geographical location of the representative site. In the present study, reanalysis and GCM data have been considered at monthly time scale. Further, as resolutions of GCM and reanalysis data are different, GCM data have to be interpolated to reanalysis grid points using software such as MATLAB and Grid Analysis and Display System (GrADS; Doty and Kinter, 1993).

3. Standardize the monthly NCEP reanalysis data corresponding to each of the predictors for baseline period

by subtracting its respective mean and dividing by the standard deviation. Similarly standardize each of the GCM predictor variables data using its respective mean and the standard deviation (Wilby *et al.*, 2004). This operation would scale the values (corresponding to GCM as well as reanalysis data) to a smaller unit less range, and they are used as input for SVM-based downscaling. However, one can opt to rescale the standardized values with the means and standard deviations of either observations or reanalysis outputs (Johnson and Sharma, 2012), if the rescaled values are to be used as input to SVM-based downscaling.

4. To avoid redundancy in standardized reanalysis predictor data, process the data using principal component analysis (PCA) to extract principal directions and corresponding principal components (PCs) that are non-redundant, orthogonal and preserve significant percentage of the variance present in the data. The PCs corresponding to baseline period form the training (calibration) set and the remaining PCs form validation set. The principal directions are assumed to be the same for PCs corresponding to calibration and validation sets, and would be subsequently utilized to arrive at PCs from GCM predictor data.
5. Provide PCs in the training set as an input to LS-SVM model to develop regression relationships between the PCs and the contemporaneous observed monthly values of each of the predictands (maximum and minimum temperatures) at the representative site. For calibration of the model, tune the parameters  $\sigma$  and  $C$  (by grid search procedure, Gestel *et al.*, 2004) and identify optimal values of parameters as those for which output of the model (i.e. downscaled monthly values of predictands) correlate well with the contemporaneous observed monthly values of the predictands at the representative site.

The LS-SVM is a machine learning algorithm that implements structural risk minimization principle, which attempts to minimize an upper bound on the generalization error, by striking a right balance between the training error and the capacity of the machine (i.e. the ability of machine to learn any training set without error). The algorithm automatically decides the model architecture. The SVM uses kernel functions for implicitly mapping the input data to a higher, possibly infinite, dimensional space for identifying a linear solution that corresponds to a non-linear solution in the original lower dimensional input space. This makes SVM a feasible choice for downscaling problems that are non-linear in nature. Details of the LS-SVM downscaling algorithm can be found in Tripathi *et al.* (2006) and Anandhi *et al.* (2008).

6. For validating the LS-SVM model, provide PCs in validation set and the model parameters identified in calibration as an input, and compare the model output (downscaled monthly values) with the contemporaneous observed monthly values of the predictands using a performance measure. If the model performs satisfactorily proceed to the next step, else proceed

to step (2) for scrutinizing the chosen predictor variables to assess if they are adequate and if any additional predictor variables are necessary to develop an effective downscaling model.

7. To derive downscaled monthly values at other sites in the cluster from the representative site, develop two LS-SVM inter-site regression relationships (one for maximum temperature and another for minimum temperature) between the observed monthly predictand values for the representative site and those for each of the other sites in the cluster for the baseline period. Subsequently, validate the developed regression relationships using the observed predictand data corresponding to the validation set. Following this, use the relationships to transfer downscaled information from the representative site to the other sites.
8. Disaggregate the downscaled monthly values to daily values using  $k$ -nearest neighbour disaggregation methodology (Nagesh Kumar *et al.*, 2000 and Anandhi *et al.*, 2012) for each of the sites in the cluster. The methodology involves finding nearest neighbour to each of the downscaled monthly values of predictand at a site. A nearest neighbour refers to an observed monthly value of predictand that is deemed closer to the downscaled monthly value of the predictand. Historical relationship between the nearest neighbour and its corresponding daily values for the site is used to disaggregate the downscaled monthly values to daily values.
9. To obtain downscaled values corresponding to GCM simulated monthly values of predictors, extract PCs from the GCM data along principal directions obtained in step (4) from standardized reanalysis predictor data. Provide the PCs as an input to the LS-SVM downscaling models developed in step (5), and obtain the downscaled monthly values of predictands (maximum and minimum temperatures) at the representative site as output. Subsequently, derive the corresponding downscaled monthly values at other sites in the cluster using the LS-SVM inter-site regression relationships developed in step (7). Following this, disaggregate the monthly values to daily values using the  $k$ -nearest neighbour disaggregation methodology described in step (8).

### 3.2. Multivariate multiple linear regression (MMLR) methodology

This section provides a brief description of MMLR methodology proposed by Jeong *et al.* (2012) that is adapted for downscaling monthly maximum and minimum temperatures for each of the sites considered in the present study.

Let  $X$  denote a matrix of dimension  $N \times n$  containing PCs extracted from large-scale atmospheric predictor variables (LSAPVs), where  $n$  denotes the number of PCs and  $N$  refers to record length (in months) for each of the predictor variables corresponding to the calibration

period. Further let  $Y$  denote a matrix of dimension  $N \times 2$  containing data of predictand variables (maximum and minimum temperatures) for a site. Steps involved in the methodology are as follows:

(1) Develop a MMLR relationship for each site as

$$Y_{[N \times 2]} = X_{[N \times n]} \cdot B_{[n \times 2]} + E_{[N \times 2]} \quad (1)$$

where  $B$  is a parameter matrix and  $E$  is residual matrix. The ordinary least squares estimates of parameters is given by,  $\hat{B} = (X^T X)^{-1} X^T Y$ , where T denotes transpose of a matrix.

(2) Determine downscaled values of monthly maximum and minimum temperatures at sites in the study area using the following relationship.

$$\hat{Y}_{[N \times 2]} = X_{[N \times n]} \cdot \hat{B}_{[n \times 2]} \quad (2)$$

(3) Determine residual matrix as

$$E = Y - \hat{Y} \quad (3)$$

(4) Update the downscaled values of monthly maximum and minimum temperatures as,  $\tilde{Y} = \hat{Y} + R$ , where  $R$  is an estimate of  $E$  that accounts for spatial correlation structure of the predictands. The matrix  $R$  is generated from a multivariate normal distribution having zero mean and spatial covariance matrix corresponding to the residual matrix. Further details on this model can be found in Jeong *et al.* (2012).

### 3.3. Multi-site multivariate statistical downscaling (MMSD) methodology

This section provides a brief description of MMSD methodology proposed by Khalili *et al.* (2011) that is adapted for downscaling monthly maximum and minimum temperatures for sites considered in this study.

Let  $T_{[N \times 1]}^{\max} = [T_1^{\max}, \dots, T_N^{\max}]'$  and  $T_{[N \times 1]}^{\min} = [T_1^{\min}, \dots, T_N^{\min}]'$  denote the observed time series of monthly maximum and minimum temperatures, respectively, for a site corresponding to the calibration period. Herein, monthly value of a predictand (maximum or minimum temperature) is considered to represent the average of daily values of the predictand over all the days in a month. Further let  $X_{[N \times n]}$  denote the matrix containing PCs extracted from LSAPVs.

The regression model for maximum temperature for a site  $i$  is given by

$$T_{[N \times 1]}^{\max,i} = X_{[N \times n]} \cdot B_{[n \times 1]}^{\max,i} + E_{[N \times 1]}^{\max,i} \quad (4)$$

Similarly, for the minimum temperature regression model for the site  $i$  can be expressed as

$$T_{[N \times 1]}^{\min,i} = X_{[N \times n]} \cdot B_{[n \times 1]}^{\min,i} + E_{[N \times 1]}^{\min,i} \quad (5)$$

where  $B_{[n \times 1]}^{\max,i}$  and  $B_{[n \times 1]}^{\min,i}$  denote the parameter matrices, and  $E_{[N \times 1]}^{\max,i} = [\varepsilon_1^{\max}, \dots, \varepsilon_N^{\max}]'$  and  $E_{[N \times 1]}^{\min,i} = [\varepsilon_1^{\min}, \dots, \varepsilon_N^{\min}]'$  denote the residual matrices

corresponding to maximum and minimum temperatures for the site, respectively. The residuals of maximum and minimum temperatures for all the  $N_s$  sites in the study area are modelled using spatial moving average process of order-1 as

$$\begin{pmatrix} E_{[N \times 1]}^{\max,1} \\ E_{[N \times 1]}^{\min,1} \\ \vdots \\ E_{[N \times 1]}^{\max,N_s} \\ E_{[N \times 1]}^{\min,N_s} \end{pmatrix} = \begin{pmatrix} U_{[N \times 1]}^{\max,1} \\ U_{[N \times 1]}^{\min,1} \\ \vdots \\ U_{[N \times 1]}^{\max,N_s} \\ U_{[N \times 1]}^{\min,N_s} \end{pmatrix} + \beta \times W \times \begin{pmatrix} U_{[N \times 1]}^{\max,1} \\ U_{[N \times 1]}^{\min,1} \\ \vdots \\ U_{[N \times 1]}^{\max,N_s} \\ U_{[N \times 1]}^{\min,N_s} \end{pmatrix} \quad (6)$$

where  $U_{[N \times 1]}^{\max,i} = [u_1^{\max}, \dots, u_N^{\max}]'$  and  $U_{[N \times 1]}^{\min,i} = [u_1^{\min}, \dots, u_N^{\min}]'$  represent matrices containing random numbers drawn from standard normal distribution;  $\beta$  is the moving average coefficient; and  $W$  is the weight matrix that is given by

$$W = \begin{bmatrix} W^{\max} & W^{\max,\min} \\ W^{\min,\max} & W^{\min} \end{bmatrix} \quad (7)$$

where  $W^{\max}$  and  $W^{\min}$  are weight matrices for the maximum and minimum temperatures respectively, associated with correlations;  $W^{\max,\min} = W^{\min,\max}$  denote weight matrix whose components are the correlations between maximum and minimum temperatures for different pairs of sites in the study area. Further details of this method can be found in Khalili *et al.* (2011).

### 4. Study area and data description

The study region is the catchment of Beas river upstream of Pandoh dam (Figure 1), which has an area of 5278 km<sup>2</sup>. The Beas river is a tributary of Sutlej River, which originates in the Himalayas in central Himachal Pradesh, India. Records of maximum and minimum temperatures were available at daily time scale for four sites (Bhuntar, Larji, Manali and Pandoh) in the study area from Bhakra Beas Management Board (BBMB), India. Details of the data are provided in Table 1 and locations of the sites are shown in Figure 1. It can be noted that the sites have contemporaneous record for the period April 1985 to March 2010.

NCEP/NCAR Reanalysis-1 gridded data prepared by National Centers for Environmental Prediction (NCEP; Kalnay *et al.*, 1996) were considered for developing the downscaling model. The data included pressure level variables, surface variables and surface fluxes. Information on pressure level variables and surface variables was available for monthly time scale at a spatial resolution of 2.5° (with 144 × 73 points extending from 0 to 357.5°E and 90°N to 90°S). The surface fluxes data were available at daily time scale, and the spatial coverage of the same was T62 Gaussian grid with 192 × 94 points extending from 88.542°N to 88.542°S and 0° to 358.125°E. In a Gaussian grid longitudes are equally spaced, whereas latitudes are unequally spaced

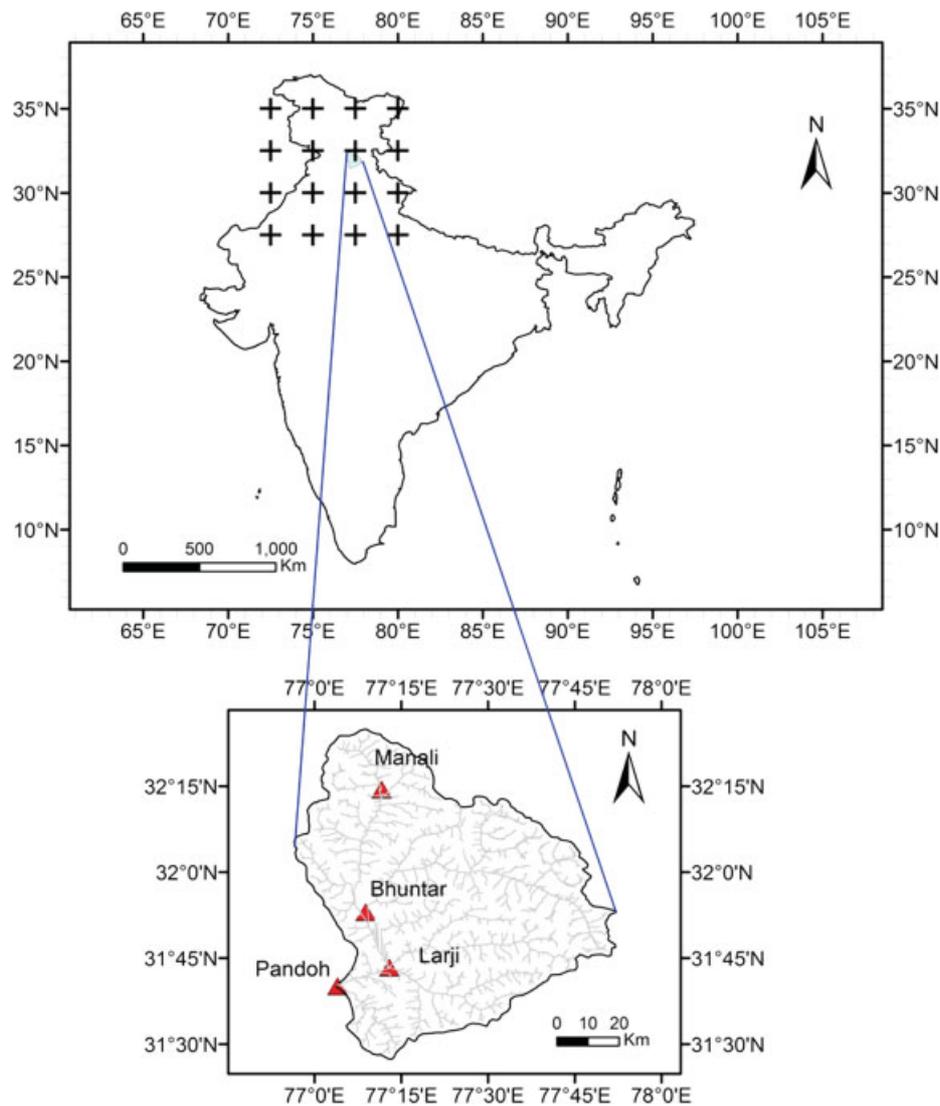


Figure 1. Study area showing location of sites, and grid points of NCEP for which information on large scale atmospheric variables was extracted for the downscaling analysis.

Table 1. Location details of the gauges considered for the study in Beas basin.

S. No.	Station	Latitude	Longitude	Elevation(m)	Duration of record
1	Bhuntar	31°53'02"	77°08'51"	1102	April 1985 to March 2010
2	Larji	31°43'21"	77° 12' 58"	950	April 1985 to March 2010
3	Manali	32°14'26"	77°11'37"	1842	April 1985 to March 2010
4	Pandoh	31°40'08"	77°03'59"	899	January 1971 to March 2010

and are defined by Gaussian quadrature. T62 indicates use of triangular truncation at wave number 62, and this corresponds to horizontal resolution of approximately 209 km. The fluxes data were aggregated to monthly time scale and then interpolated to 2.5° spaced grid points corresponding to the other NCEP variables using GrADS software. Data on pressure level variables were available for at most 17 pressure levels (from among 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10 mb). The data pertaining to the period April 1985 to March 2010 were downloaded

for grid points shown in Figure 1 from the web link: [http://www.esrl.noaa.gov/psd/data/gridded/data.ncep\\_reanalysis.html](http://www.esrl.noaa.gov/psd/data/gridded/data.ncep_reanalysis.html).

The GCM for obtaining future projections of temperature at sites in the Beas river basin has been considered as T63 version of third generation Canadian Coupled Global Climate Model (CGCM3.1/T63). The T63 version of GCM has 31 levels in the vertical. The atmosphere model output was available on a 128 × 64 Gaussian surface grid (resolution is approximately 2.81° lat × 2.81° long) and the ocean model output was available on a 256 × 192

Table 2. Site-to-site cross-correlations of maximum temperature at monthly time scale for observed (historical) data and for downscaled values corresponding to calibration period and four IPCC scenarios (A1B, A2, B1 and COMMIT).

Site	Bhuntar	Larji	Manali	Pandoh	
<b>(a) Historical</b>					
Bhuntar	1.000	0.957	0.884	0.953	
Larji		1.000	0.863	0.906	
Manali			1.000	0.825	
Pandoh				1.000	
Model	Site	Bhuntar	Larji	Manali	Pandoh
<b>(b) Calibration period (April 1985 to December 2000)</b>					
LS-SVM	Bhuntar	1.000	0.999	0.995	0.997
	Larji		1.000	0.997	0.995
	Manali			1.000	0.989
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.994	0.993	0.986
	Larji		1.000	0.991	0.982
	Manali			1.000	0.978
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.959	0.951	0.949
	Larji		1.000	0.957	0.950
	Manali			1.000	0.942
	Pandoh				1.000
<b>(c) A1B scenario</b>					
LS-SVM	Bhuntar	1.000	0.967	0.904	0.989
	Larji		1.000	0.968	0.979
	Manali			1.000	0.912
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.982	0.978	0.969
	Larji		1.000	0.966	0.959
	Manali			1.000	0.944
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.861	0.845	0.857
	Larji		1.000	0.839	0.872
	Manali			1.000	0.819
	Pandoh				1.000
<b>(d) A2 scenario</b>					
LS-SVM	Bhuntar	1.000	0.970	0.912	0.990
	Larji		1.000	0.972	0.980
	Manali			1.000	0.918
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.983	0.979	0.969
	Larji		1.000	0.970	0.961
	Manali			1.000	0.945
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.860	0.855	0.856
	Larji		1.000	0.851	0.874
	Manali			1.000	0.834
	Pandoh				1.000
<b>(e) B1 scenario</b>					
LS-SVM	Bhuntar	1.000	0.969	0.916	0.989
	Larji		1.000	0.972	0.980
	Manali			1.000	0.922
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.982	0.979	0.969
	Larji		1.000	0.968	0.963
	Manali			1.000	0.946
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.854	0.842	0.833
	Larji		1.000	0.845	0.863
	Manali			1.000	0.819
	Pandoh				1.000

Table 2. (Continued.)

Model	Site	Bhuntar	Larji	Manali	Pandoh
(f) COMMIT scenario LS-SVM	Bhuntar	1.000	0.960	0.900	0.987
	Larji		1.000	0.969	0.979
	Manali			1.000	0.916
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.981	0.978	0.967
	Larji		1.000	0.968	0.958
	Manali			1.000	0.943
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.863	0.846	0.847
	Larji		1.000	0.834	0.862
	Manali			1.000	0.813
	Pandoh				1.000

grid (i.e.  $2 \times 3 = 6$  oceanic grid boxes underlying each atmospheric grid box). Simulated monthly climate data of the GCM were extracted from Canadian Center for Climate Modelling and Analysis (CCCma) web site: <http://www.cccma.ec.gc.ca/data/cgcm3> for Intergovernmental Panel on Climate Change (IPCC) 20th Century Experiment (20C3M) pertaining to the period 1850–2000, and for four IPCC SRES emission scenarios (A1B, A2, B1 and COMMIT) pertaining to the period 2001–2100. For 20C3M scenario, equivalent atmospheric carbon-dioxide concentrations and other input data are based either on historical records or on estimates beginning around the time of the Industrial revolution. On the other hand, A1B, A2 and B1 scenarios assume diverse pathways for the equivalent atmospheric carbon dioxide concentrations that would reach 720, 850 and 550 ppm in the year 2100, respectively.

The A1B scenario describes a future world of very rapid economic growth, a global population that peaks in mid-century and declines thereafter, and rapid introduction of new and more efficient technologies that do not rely too heavily on either fossil-intensive or non-fossil energy sources. B1 scenario describes a convergent world, with the same global population as A1, but with more rapid changes in economic structures towards a service and information economy, with reductions in material intensity and the introduction of clean and resource-efficient technologies. B2 scenario describes a heterogeneous world with intermediate population and economic growth, diverse technological change, with emphasis on local solutions to economic, social and environmental sustainability. On the other hand, A2 scenario describes a very heterogeneous world with high population growth, regionally oriented economic development, slower and fragmented technological change and per capita economic growth. In COMMIT scenario, atmospheric carbon dioxide concentrations are held at year 2000 levels and this experiment is based on conditions that already exist (IPCC, 2007).

## 5. Results and discussion

Statistical consistency checks were performed on the observed data of predictands (maximum and minimum temperatures). To identify representative site for downscaling analysis, site-to-site monthly cross-correlations for each of the predictands were examined. Results indicated that maximum and minimum temperatures at Bhuntar site were highly correlated with those at the other three sites (see Tables 2(a) and 3(a)). Consequently, Bhuntar was chosen as the representative site.

Monthly LSAV data of GCM were interpolated to NCEP grid points using 3rd order (4-point) Bessel interpolation procedure available in GrADS software. Subsequently predictor variables corresponding to each predictand were so chosen from LSAVs available in both NCEP reanalysis data and CGCM3 simulations (corresponding to 20C3M scenario) that those variables are reasonably well correlated with monthly data of predictands at the Bhuntar site. For the sake of this analysis, the spatial domain of each of the predictor variables was chosen as 16 NCEP grid points surrounding the representative site. A few of the typical correlation plots that have been scrutinized in this analysis are presented in Figure 2, for brevity. The predictor variables that have been chosen based on this analysis were: (1) temperature at 100 and 850 mb pressure levels (Ta-100 and Ta-850); (2) geopotential height at 500 and 1000 mb pressure levels (hgt-500 and hgt-1000); (3) specific humidity at 700 mb pressure level (Shum-700); (4) zonal wind velocity at 100 and 925 mb pressure levels (uwind-100 and uwind-925); (5) meridional wind velocity at 50 and 600 mb pressure levels (vwind-50 and vwind-600); (6) latent heat net flux (Lhf); (7) sensible heat net flux (Shf); (8) net shortwave radiation (Swr), and (9) net longwave radiation (Lwr). The number of LSAPVs corresponding to these 13 predictors over the spatial domain considered for the analysis (i.e. 16 NCEP grid points) is 208 ( $13 \times 16$ ). Spatial average values of monthly correlations computed between the NCEP/NCAR reanalysis predictor variable data and predictands (maximum and minimum temperatures) are presented alongside those estimated

Table 3. Site-to-site cross-correlations of minimum temperature at monthly time scale for observed (historical) data and for downscaled values corresponding to calibration period and four IPCC scenarios (A1B, A2, B1 and COMMIT).

Site	Bhuntar	Larji	Manali	Pandoh	
<b>(a) Historical</b>					
Bhuntar	1.000	0.960	0.974	0.940	
Larji		1.000	0.955	0.915	
Manali			1.000	0.921	
Pandoh				1.000	
Model	Site	Bhuntar	Larji	Manali	Pandoh
<b>(b) Calibration period (April 1985 to December 2000)</b>					
LS-SVM	Bhuntar	1.000	0.995	0.999	0.972
	Larji		1.000	0.996	0.988
	Manali			1.000	0.973
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.994	0.997	0.957
	Larji		1.000	0.996	0.975
	Manali			1.000	0.960
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.966	0.961	0.937
	Larji		1.000	0.961	0.949
	Manali			1.000	0.930
	Pandoh				1.000
<b>(c) A1B scenario</b>					
LS-SVM	Bhuntar	1.000	0.985	0.997	0.970
	Larji		1.000	0.995	0.995
	Manali			1.000	0.984
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.965	0.984	0.751
	Larji		1.000	0.974	0.865
	Manali			1.000	0.771
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.802	0.802	0.570
	Larji		1.000	0.805	0.677
	Manali			1.000	0.559
	Pandoh				1.000
<b>(d) A2 scenario</b>					
LS-SVM	Bhuntar	1.000	0.984	0.997	0.965
	Larji		1.000	0.994	0.993
	Manali			1.000	0.980
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.963	0.984	0.748
	Larji		1.000	0.974	0.867
	Manali			1.000	0.773
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.828	0.835	0.626
	Larji		1.000	0.821	0.713
	Manali			1.000	0.612
	Pandoh				1.000
<b>(e) B1 scenario</b>					
LS-SVM	Bhuntar	1.000	0.986	0.997	0.973
	Larji		1.000	0.995	0.995
	Manali			1.000	0.986
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.969	0.986	0.764
	Larji		1.000	0.977	0.868
	Manali			1.000	0.786
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.800	0.805	0.572
	Larji		1.000	0.793	0.685
	Manali			1.000	0.556
	Pandoh				1.000

Table 3. (Continued.)

Model	Site	Bhuntar	Larji	Manali	Pandoh
(f) COMMIT scenario LS-SVM	Bhuntar	1.000	0.988	0.998	0.981
	Larji		1.000	0.996	0.997
	Manali			1.000	0.991
	Pandoh				1.000
MMLR	Bhuntar	1.000	0.966	0.985	0.760
	Larji		1.000	0.975	0.868
	Manali			1.000	0.782
	Pandoh				1.000
MMSD	Bhuntar	1.000	0.786	0.793	0.537
	Larji		1.000	0.792	0.672
	Manali			1.000	0.543
	Pandoh				1.000

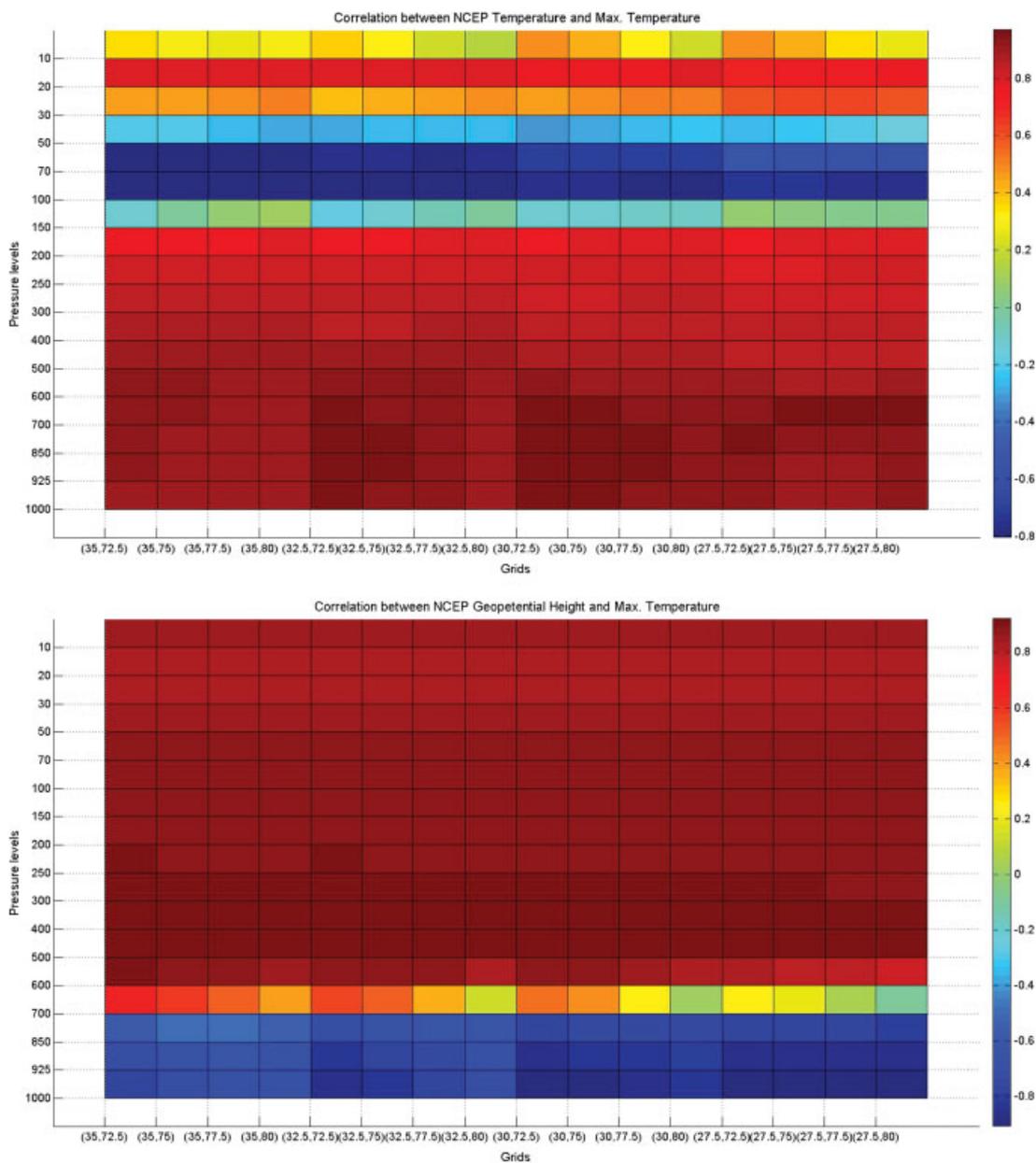


Figure 2. Typical correlation plots that have been scrutinized to identify predictors for downscaling maximum temperature.

Table 4. Correlation estimated between NCEP/NCAR reanalysis/CGCM3 predictor variable data and predictands for the Bhuntar station.

Predictor name	Correlation between NCEP/NCAR predictor data and the predictand:		Correlation between CGCM3 predictor data and the predictand:	
	Maximum temperature	Minimum temperature	Maximum temperature	Minimum temperature
Ta-100	-0.77	-0.85	-0.81	-0.85
Ta-850	0.95	0.95	0.95	0.94
hgt-500	0.92	0.89	0.90	0.87
hgt-1000	-0.84	-0.93	-0.89	-0.97
Shum-700	0.80	0.97	0.68	0.88
uwind-100	-0.84	-0.96	-0.80	-0.94
uwind-925	0.10	0.20	0.56	0.51
vwind-50	-0.28	-0.21	-0.73	-0.70
vwind-600	-0.75	-0.62	-0.53	-0.59
Lhf	0.80	0.93	0.55	0.73
Shf	0.65	0.62	0.80	0.75
Swr	-0.88	-0.83	-0.88	-0.83
Lwr	-0.19	-0.56	-0.42	-0.18

between CGCM3 20C3M predictor variable data and predictands for the Bhuntar station in Table 4. Scrutiny of the correlation values indicates that predictors are, in general, well correlated with the predictands, irrespective of whether predictor data is based on CGCM3 or NCEP/NCAR.

Monthly NCEP and GCM data on each of the 208 LSAPVs were standardized. For this purpose, mean and standard deviation of each LSAPV were computed for a pre-defined baseline period (April 1985 to December 2000) from NCEP data. Standardization of NCEP as well as GCM data corresponding to a predictor variable involved subtraction of respective mean and division by the respective standard deviation of the predictor variable for the baseline period. Principal directions and PCs were extracted from the standardized NCEP-based LSAPV data. The use of PCs as an input to a downscaling model helps in making the model more stable and at the same time reduces its computational burden. The PCs were used to calibrate and validate the LS-SVM downscaling model, as described in steps (5) and (6) of the methodology presented in Section 3.1. The model outputs (i.e. downscaled maximum and minimum temperatures) were compared with the contemporaneous observed monthly values of the predictands using four performance measures: average bias (A-bias), root mean square error (RMSE), normalized mean square error (NMSE) and Pearson product moment correlation.

$$\text{A-bias} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i) \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (9)$$

$$\text{NMSE} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sigma_x^2} \quad (10)$$

$$\text{Correlation} = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - \bar{\hat{x}})(x_i - \bar{x})}{\sigma_{\hat{x}} \sigma_x} \quad (11)$$

where  $\hat{x}_i$  and  $x_i$  denote downscaled and observed values respectively, and  $N$  represents the number of values;  $\sigma_{\hat{x}}$  and  $\sigma_x$  denote standard deviation of the downscaled and observed values respectively. If the downscaling model is perfect (best), the downscaled values would be exactly equal to their historical counterparts and that corresponds to zero value for A-bias, RMSE and NMSE, and unity (one) for correlation.

In model calibration, sensitivity analysis was performed to examine the effect of the number of PCs on the performance of the downscaling model and optimal number of PCs was decided as 3. Further, in the grid search procedure, the domain (feasible range) for searching optimal values of LS-SVM model parameters was chosen as [1, 2000], and the model performance was found to be comparable for a few alternate combinations of parameters. Estimates of the performance measures corresponding to the parameters  $\sigma = 1$  and  $C = 200$  were deemed acceptable for the Bhuntar site and they are presented in Tables 5 and 6 for maximum and minimum temperatures respectively. Correlations between the observed and the downscaled monthly values of maximum and minimum temperatures for calibration period (April 1985 to December 2000) for Bhuntar site were found to be 0.98 and 0.99 respectively. The same for validation period (January 2001 to March 2010) were 0.86 and 0.94 respectively. Time series of the predictands resulting from the downscaling analysis were compared with their real world counterparts in Figures 3 and 4. The results indicate that

Table 5. Estimates of performance measures for NCEP-based downscaled monthly maximum temperature values resulting from LS-SVM, MMLR and MMSD models.

Station name	Performance measure	Model	Calibration period (April 1985 to December 2000)	Validation period (January 2001 to March 2010)	Total period (April 1985 to March 2010)
Bhuntar	A-bias (°C)	LS-SVM	0.00	-1.14	-0.42
		MMLR	-0.09	-0.60	-0.28
		MMSD	-0.04	-1.23	-0.48
	RMSE (°C)	LS-SVM	1.16	3.26	2.18
		MMLR	3.11	3.13	3.12
		MMSD	2.65	2.70	2.67
	NMSE	LS-SVM	0.04	0.31	0.13
		MMLR	0.08	0.09	0.27
		MMSD	0.07	0.08	0.20
Correlation	LS-SVM	0.98	0.86	0.93	
	MMLR	0.87	0.87	0.87	
	MMSD	0.90	0.92	0.90	
Larji	A-bias (°C)	LS-SVM	NA	NA	-0.46
		MMLR	-0.08	-1.26	-0.52
		MMSD	-0.06	-2.05	-0.80
	RMSE (°C)	LS-SVM	NA	NA	2.88
		MMLR	3.58	4.14	3.80
		MMSD	2.89	4.05	3.37
	NMSE	LS-SVM	NA	NA	0.18
		MMLR	0.07	0.11	0.32
		MMSD	0.06	0.10	0.25
Correlation	LS-SVM	NA	NA	0.91	
	MMLR	0.87	0.82	0.85	
	MMSD	0.91	0.85	0.88	
Manali	A-bias (°C)	LS-SVM	NA	NA	-0.40
		MMLR	-0.05	3.44	1.24
		MMSD	-0.08	2.74	0.96
	RMSE (°C)	LS-SVM	NA	NA	3.25
		MMLR	3.37	5.17	4.13
		MMSD	2.73	4.69	3.58
	NMSE	LS-SVM	NA	NA	0.25
		MMLR	0.09	0.11	0.40
		MMSD	0.08	0.10	0.30
Correlation	LS-SVM	NA	NA	0.87	
	MMLR	0.84	0.83	0.80	
	MMSD	0.89	0.83	0.85	
Pandoh	A-bias (°C)	LS-SVM	NA	NA	-0.37
		MMLR	-0.12	-0.69	-0.33
		MMSD	0.03	-1.41	-0.50
	RMSE (°C)	LS-SVM	NA	NA	2.44
		MMLR	3.51	3.72	3.59
		MMSD	3.00	3.20	3.07
	NMSE	LS-SVM	NA	NA	0.19
		MMLR	0.11	0.13	0.42
		MMSD	0.09	0.11	0.31
Correlation	LS-SVM	NA	NA	0.90	
	MMLR	0.81	0.79	0.80	
	MMSD	0.85	0.86	0.85	

The downscaling model is calibrated and validated for only Bhuntar station. NA is mentioned for cases where calibration and validation are not applicable.

the downscaling model performed fairly well in downscaling LSAPVs to monthly values of the predictands for the representative site (Bhuntar). Comparison of these results with those obtained using MMLR and MMSD models (presented in Tables 5 and 6) indicates effectiveness of the developed LS-SVM model. Development of MMLR and MMSD models was done following the procedure given in Sections 3.2. and 3.3., respectively. The

parameter matrix **B** in the case of MMLR model was found to be

$$\begin{bmatrix} -0.5057 & 0.2680 & -0.2314 \\ -0.5838 & -0.0833 & 0.1460 \end{bmatrix}^T,$$

$$\begin{bmatrix} -0.5691 & 0.3793 & -0.1394 \\ -0.6052 & -0.0190 & 0.0432 \end{bmatrix}^T,$$

Table 6. Estimates of performance measures for NCEP-based downscaled monthly minimum temperature values resulting from LS-SVM, MMLR and MMSD models.

Station name	Performance measure	Model	Calibration period (April 1985 to December 2000)	Validation period (January 2001 to March 2010)	Total period (April 1985 to March 2010)
Bhuntar	A-bias (°C)	LS-SVM	0.00	-0.42	-0.16
		MMLR	0.00	-0.62	-0.23
		MMSD	-0.02	-0.54	-0.21
	RMSE (°C)	LS-SVM	0.73	2.21	1.47
		MMLR	1.64	1.74	1.68
		MMSD	1.99	1.91	1.96
	NMSE	LS-SVM	0.01	0.11	0.05
		MMLR	0.04	0.04	0.07
		MMSD	0.05	0.04	0.09
Correlation	LS-SVM	0.99	0.94	0.97	
	MMLR	0.97	0.97	0.97	
	MMSD	0.95	0.96	0.95	
Larji	A-bias (°C)	LS-SVM	NA	NA	-0.13
		MMLR	0.02	-2.67	-0.98
		MMSD	0.07	-2.61	-0.93
	RMSE (°C)	LS-SVM	NA	NA	2.41
		MMLR	2.35	3.62	2.89
		MMSD	2.36	3.80	2.98
	NMSE	LS-SVM	NA	NA	0.12
		MMLR	0.05	0.08	0.18
		MMSD	0.05	0.08	0.19
Correlation	LS-SVM	NA	NA	0.94	
	MMLR	0.94	0.93	0.92	
	MMSD	0.94	0.92	0.91	
Manali	A-bias (°C)	LS-SVM	NA	NA	-0.16
		MMLR	0.03	-1.23	-0.44
		MMSD	0.02	-1.20	-0.43
	RMSE (°C)	LS-SVM	NA	NA	1.95
		MMLR	2.04	2.71	2.31
		MMSD	2.26	2.99	2.56
	NMSE	LS-SVM	NA	NA	0.09
		MMLR	0.05	0.06	0.13
		MMSD	0.06	0.06	0.16
Correlation	LS-SVM	NA	NA	0.95	
	MMLR	0.95	0.94	0.93	
	MMSD	0.93	0.92	0.92	
Pandoh	A-bias (°C)	LS-SVM	NA	NA	-0.12
		MMLR	-0.02	-0.12	-0.06
		MMSD	0.04	0.10	0.06
	RMSE (°C)	LS-SVM	NA	NA	2.19
		MMLR	2.87	2.40	2.70
		MMSD	2.70	2.03	2.47
	NMSE	LS-SVM	NA	NA	0.11
		MMLR	0.07	0.05	0.17
		MMSD	0.07	0.05	0.14
Correlation	LS-SVM	NA	NA	0.94	
	MMLR	0.90	0.93	0.91	
	MMSD	0.91	0.95	0.92	

$$\begin{bmatrix} -0.4964 & 0.2260 & -0.1579 \\ -0.5413 & -0.0945 & 0.0921 \end{bmatrix}^T$$

and

$$\begin{bmatrix} -0.4391 & 0.3217 & -0.2549 \\ -0.5477 & 0.2289 & 0.0254 \end{bmatrix}^T$$

for the sites Bhuntar, Larji, Manali and Pandoh respectively. The optimal value of the parameter  $\beta$  in the case of MMSD model was found to be 0.1041 for the study area.

To translate downscaled information from Bhuntar to other sites, inter-site LS-SVM regression relationships were fitted between Bhuntar site and each of the other sites using observed values of predictands at those sites. This involved calibration of the relationships for the baseline period (April 1985 to December 2000) and their validation for the period January 2001 to March 2010. The error involved in this operation was quantified using the four performance measures given in Equations (8)–(11), with  $\hat{x}_i$  and  $x_i$  representing value

Table 7. Results obtained from calibration and validation of inter-site regression relationships developed using LS-SVM model for maximum and minimum temperatures.

Station name	Performance measure	Maximum temperature		Minimum temperature	
		Calibration	Validation	Calibration	Validation
Larji	A-bias	0.00	0.37	0.00	1.94
	RMSE	1.48	2.61	1.23	2.73
	NMSE	0.99	1.01	0.99	2.00
	Correlation	0.98	0.92	0.98	0.96
Manali	A-bias	0.00	-3.94	0.00	0.63
	RMSE	1.43	5.17	1.11	1.91
	NMSE	0.99	2.36	0.99	1.11
	Correlation	0.97	0.87	0.98	0.97
Pandoh	A-bias	0.00	0.14	0.00	-0.99
	RMSE	1.70	1.48	1.76	1.53
	NMSE	0.99	1.00	0.99	1.71
	Correlation	0.95	0.96	0.96	0.98

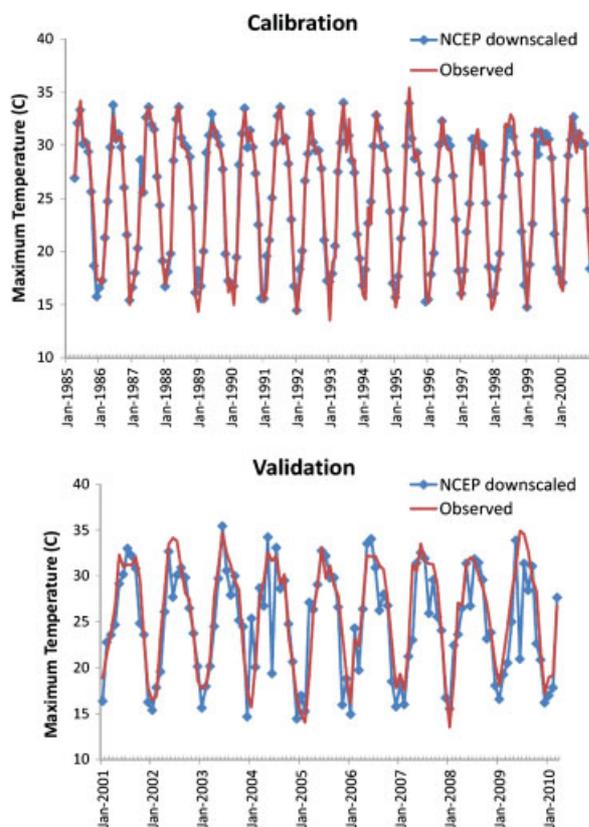


Figure 3. Results pertaining to calibration and validation of LS-SVM downscaling model developed for downscaling NCEP data (on large-scale predictor variables) to monthly maximum temperature at Bhuntar station.

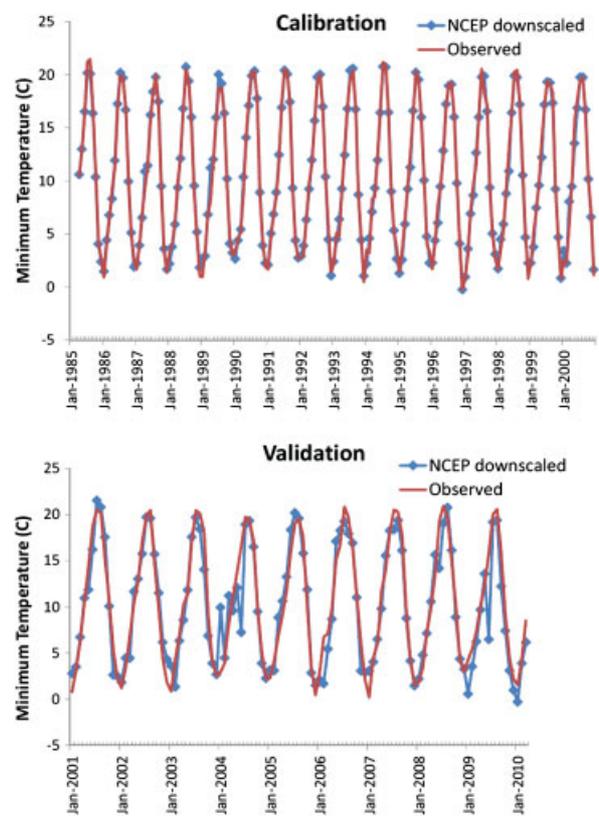


Figure 4. Results pertaining to calibration and validation of LS-SVM downscaling model developed for downscaling NCEP data (on large scale predictor variables) to monthly minimum temperature at Bhuntar station.

generated using regression relation and observed value respectively. The model performance was found to be comparable for a few alternate combinations of parameters. Results corresponding to the parameters  $\sigma = 1$  and  $C = 50$  were deemed acceptable and are shown in Table 7. Scrutiny of the results shows that overall LS-SVM model performed fairly well in capturing inter-site regression relationships in the case of both maximum and minimum temperatures, though some inflation is

noted for maximum temperature derived at station Manali. Following this, the downscaled monthly time series of maximum and minimum temperatures were translated from Bhuntar to those at the other sites in the study area by using the developed inter-site regression relationships. The estimates of performance measures for Larji, Manali and Pandoh stations are shown in Tables 5 and 6 (alongside those already presented for Bhuntar station) for maximum and minimum temperatures

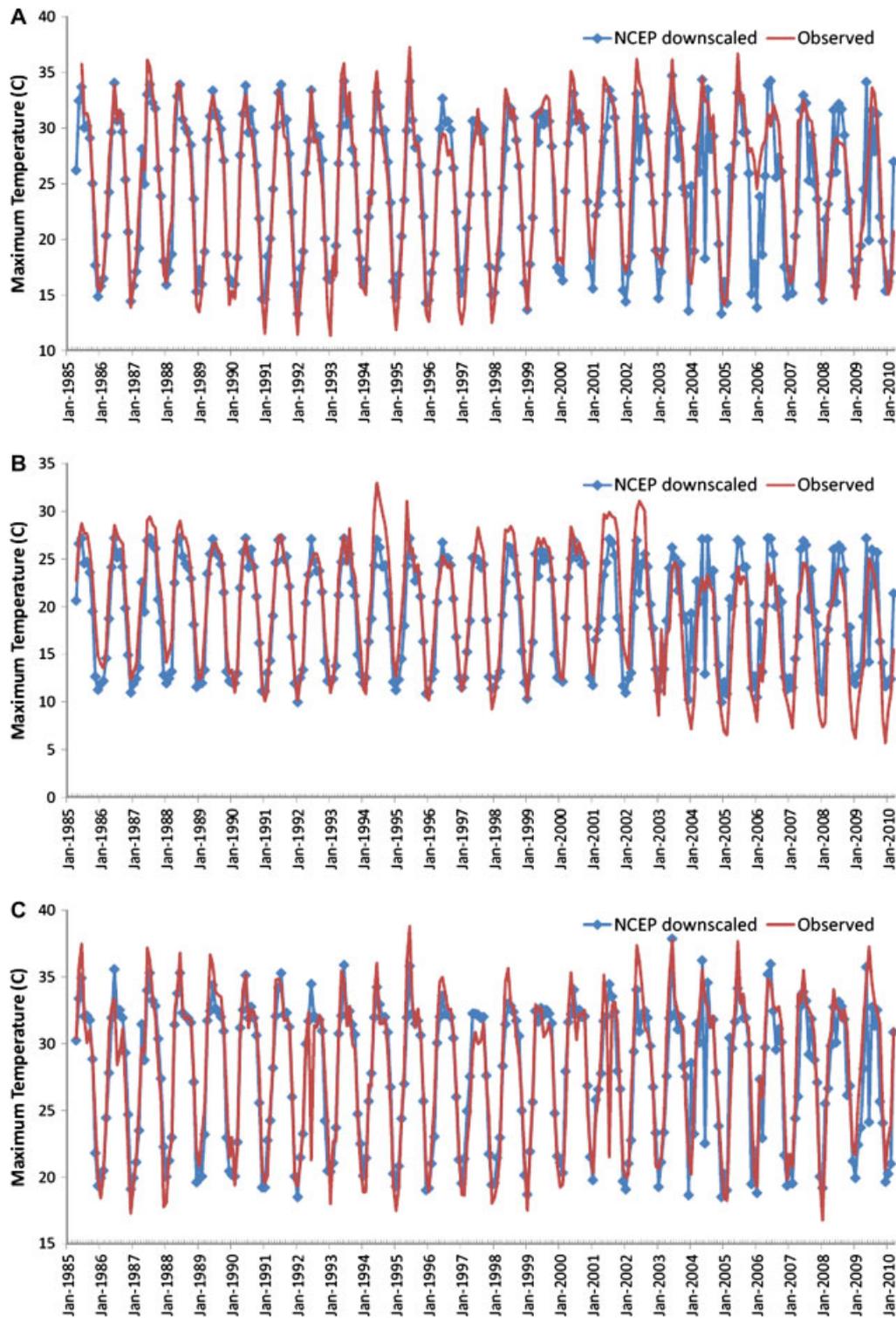


Figure 5. Observed and downscaled values of monthly maximum temperature for (A) Larji (B) Manali and (C) Pandoh sites. The downscaled values for each of these three sites were obtained from those for Bhuntar site, using LS-SVM regression relationship fitted between the observed monthly maximum temperatures at Bhuntar and the respective site.

respectively. In the case of maximum temperature, the A-bias value for the three stations varied between  $-0.46$  and  $-0.37^{\circ}\text{C}$ , RMSE varied between  $2.44$  and  $3.25^{\circ}\text{C}$ , while NMSE varied from  $0.18$  to  $0.25$ . The same for minimum temperature varied between  $-0.16$  and  $-0.12^{\circ}\text{C}$ ,  $1.95$  and  $2.41^{\circ}\text{C}$ , and  $0.09$

and  $0.12$ , respectively. Further correlations between the observed and the downscaled monthly values of maximum temperature for the sites Larji, Manali and Pandoh during the period April 1985 to March 2010 were found to be  $0.91$ ,  $0.87$  and  $0.90$  respectively (Table 5 and Figure 5). The same for minimum

Table 8. Cross-correlations between time series of minimum and maximum temperatures at monthly time scale for observed/historical data (April 1985 to March 2010), NCEP-based downscaled values corresponding to calibration period, and GCM-based downscaled values for four IPCC scenarios (A1B, A2, B1 and COMMIT).

Time series of minimum temperature at site	Time series of maximum temperature at site				
	Bhuntar	Larji	Manali	Pandoh	
<b>(a) Historical</b>					
Bhuntar	0.893	0.869	0.844	0.824	
Larji	0.918	0.894	0.802	0.857	
Manali	0.903	0.877	0.840	0.833	
Pandoh	0.928	0.897	0.884	0.900	
Model	Time series of minimum temperature at site	Time series of maximum temperature at site			
		Bhuntar	Larji	Manali	Pandoh
<b>(b) Calibration period (April 1985 to December 2000)</b>					
LS-SVM	Bhuntar	0.908	0.911	0.919	0.893
	Larji	0.934	0.936	0.944	0.922
	Manali	0.912	0.915	0.925	0.898
	Pandoh	0.956	0.956	0.959	0.950
MMLR	Bhuntar	0.853	0.847	0.858	0.791
	Larji	0.869	0.860	0.868	0.811
	Manali	0.845	0.836	0.848	0.782
	Pandoh	0.884	0.886	0.878	0.840
MMSD	Bhuntar	0.906	0.904	0.916	0.858
	Larji	0.914	0.929	0.932	0.878
	Manali	0.883	0.895	0.923	0.852
	Pandoh	0.947	0.957	0.954	0.943
<b>(c) A1B scenario</b>					
LS-SVM	Bhuntar	0.772	0.807	0.811	0.771
	Larji	0.831	0.864	0.864	0.832
	Manali	0.802	0.836	0.838	0.802
	Pandoh	0.845	0.875	0.872	0.844
MMLR	Bhuntar	0.366	0.311	0.375	0.268
	Larji	0.428	0.380	0.419	0.344
	Manali	0.339	0.283	0.346	0.251
	Pandoh	0.509	0.497	0.468	0.460
MMSD	Bhuntar	0.481	0.425	0.505	0.375
	Larji	0.624	0.564	0.648	0.531
	Manali	0.467	0.406	0.507	0.371
	Pandoh	0.814	0.819	0.803	0.790
<b>(d) A2 scenario</b>					
LS-SVM	Bhuntar	0.766	0.799	0.800	0.766
	Larji	0.825	0.855	0.854	0.827
	Manali	0.795	0.827	0.826	0.796
	Pandoh	0.840	0.867	0.863	0.841
MMLR	Bhuntar	0.425	0.375	0.425	0.339
	Larji	0.473	0.435	0.465	0.398
	Manali	0.391	0.344	0.393	0.313
	Pandoh	0.544	0.544	0.520	0.500
MMSD	Bhuntar	0.537	0.472	0.567	0.427
	Larji	0.649	0.606	0.680	0.560
	Manali	0.522	0.471	0.567	0.420
	Pandoh	0.814	0.822	0.827	0.795
<b>(e) B1 scenario</b>					
LS-SVM	Bhuntar	0.793	0.827	0.827	0.794
	Larji	0.842	0.873	0.872	0.845
	Manali	0.818	0.851	0.850	0.820
	Pandoh	0.849	0.878	0.878	0.850
MMLR	Bhuntar	0.365	0.303	0.366	0.262
	Larji	0.423	0.367	0.409	0.332
	Manali	0.338	0.277	0.339	0.241
	Pandoh	0.510	0.495	0.474	0.459

Table 8. (Continued.)

Model	Time series of minimum temperature at site	Time series of maximum temperature at site			
		Bhuntar	Larji	Manali	Pandoh
MMSD	Bhuntar	0.496	0.426	0.538	0.380
	Larji	0.631	0.582	0.660	0.541
	Manali	0.484	0.409	0.521	0.360
	Pandoh	0.802	0.810	0.805	0.779
(f) COMMIT scenario LS-SVM	Bhuntar	0.809	0.843	0.841	0.812
	Larji	0.852	0.883	0.877	0.857
	Manali	0.832	0.864	0.860	0.836
	Pandoh	0.858	0.885	0.879	0.860
MMLR	Bhuntar	0.380	0.319	0.370	0.286
	Larji	0.425	0.373	0.405	0.340
	Manali	0.352	0.292	0.341	0.264
	Pandoh	0.497	0.485	0.457	0.447
MMSD	Bhuntar	0.456	0.400	0.514	0.339
	Larji	0.635	0.585	0.663	0.539
	Manali	0.472	0.414	0.521	0.365
	Pandoh	0.802	0.805	0.801	0.783

Table 9. Correlation between the observed daily time series of predictands and those downscaled based on NCEP reanalysis data for the four sites considered in the study.

Site	Correlation of maximum temperature for the period			Correlation of minimum temperature for the period		
	April 1985 to December 2000	January 2001 to March 2010	April 1985 to March 2010	April 1985 to December 2000	January 2001 to March 2010	April 1985 to March 2010
Bhuntar	0.79	0.66	0.74	0.93	0.87	0.90
Larji	0.81	0.69	0.73	0.91	0.85	0.88
Manali	0.80	0.66	0.72	0.87	0.85	0.86
Pandoh	0.76	0.66	0.72	0.90	0.87	0.89

temperature for the sites were 0.94, 0.95 and 0.94 respectively (Table 6 and Figure 6). Comparison of these results with those obtained using MMLR and MMSD models (presented in Tables 5 and 6) indicate effectiveness of the developed LS-SVM-based methodology for multi-site downscaling of maximum and minimum daily temperature series.

For the sake of comparing correlation structure of the observed predictand data with that of the downscaled monthly values corresponding to calibration period and each of the four IPCC scenarios (A1B, A2, B1 and COMMIT), the following two types of correlations were computed.

1. Cross-correlations between time series of a predictand for all possible pairs of sites.
2. Cross-correlations between time series of different predictands at each site and for all possible pairs of sites.

All the three models exhibited satisfactory performance in accounting for inter-site correlation structure of maximum as well as minimum temperatures for the calibration period, though some inflation is noted (Tables

2(b) and 3(b)). However in the case of downscaled values obtained for IPCC scenarios, inflation in the correlations is noted in the case of results obtained from MMLR and LS-SVM models, whereas deflation is noted in the case of those from MMSD model (parts (c)–(f) of Tables 2 and 3). The deviation (inflation and deflation) is more in the case of maximum temperature. Furthermore, it has been noted that the LS-SVM model exhibited satisfactory performance in accounting for cross-correlations between time series of different predictands for calibration period as well as IPCC scenarios (Table 8). However, considerable deflation in correlations was noted in the case of downscaled values resulting from MMLR and MMSD models, though their performance was comparable with that of LS-SVM for calibration period.

Overall, the foregoing results indicate that the performance of LS-SVM model is better among the three models. Consequently, downscaled values obtained from only LS-SVM model were considered for subsequent analysis. The *k*-nearest neighbour disaggregation methodology was used for temporal disaggregation of the downscaled monthly values to daily values at each of the sites, as mentioned in step (8) of Section 3.1. Correlations computed between the observed daily time series of

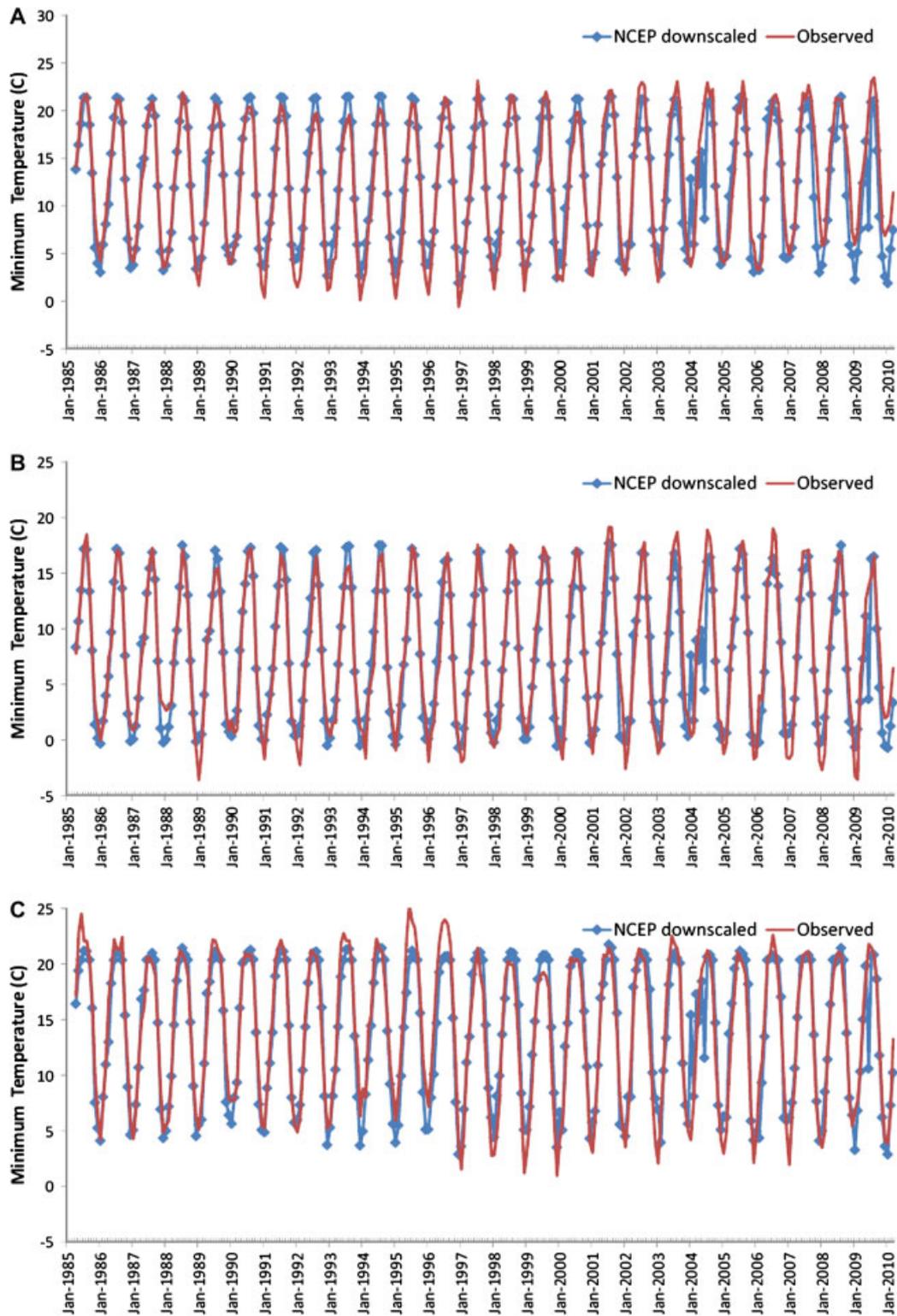


Figure 6. Observed and downscaled values of monthly minimum temperature for (A) Larji (B) Manali and (C) Pandoh sites. The downscaled values for each of these three sites were obtained from those for Bhuntar site, using LS-SVM regression relationship fitted between the observed monthly minimum temperatures at Bhuntar and the respective site.

predictands and those obtained by disaggregation are presented in Table 9 for calibration (April 1985 to December 2000), validation (January 2001 to March 2010) and combined calibration and validation periods (April 1985 to March 2010). The results indicate that the proposed

methodology to arrive at daily values of the predictands is effective. The correlations for the period April 1985 to December 2000 are better than those for the period January 2001 to March 2010, as expected. The performance of the methodology in deriving daily minimum

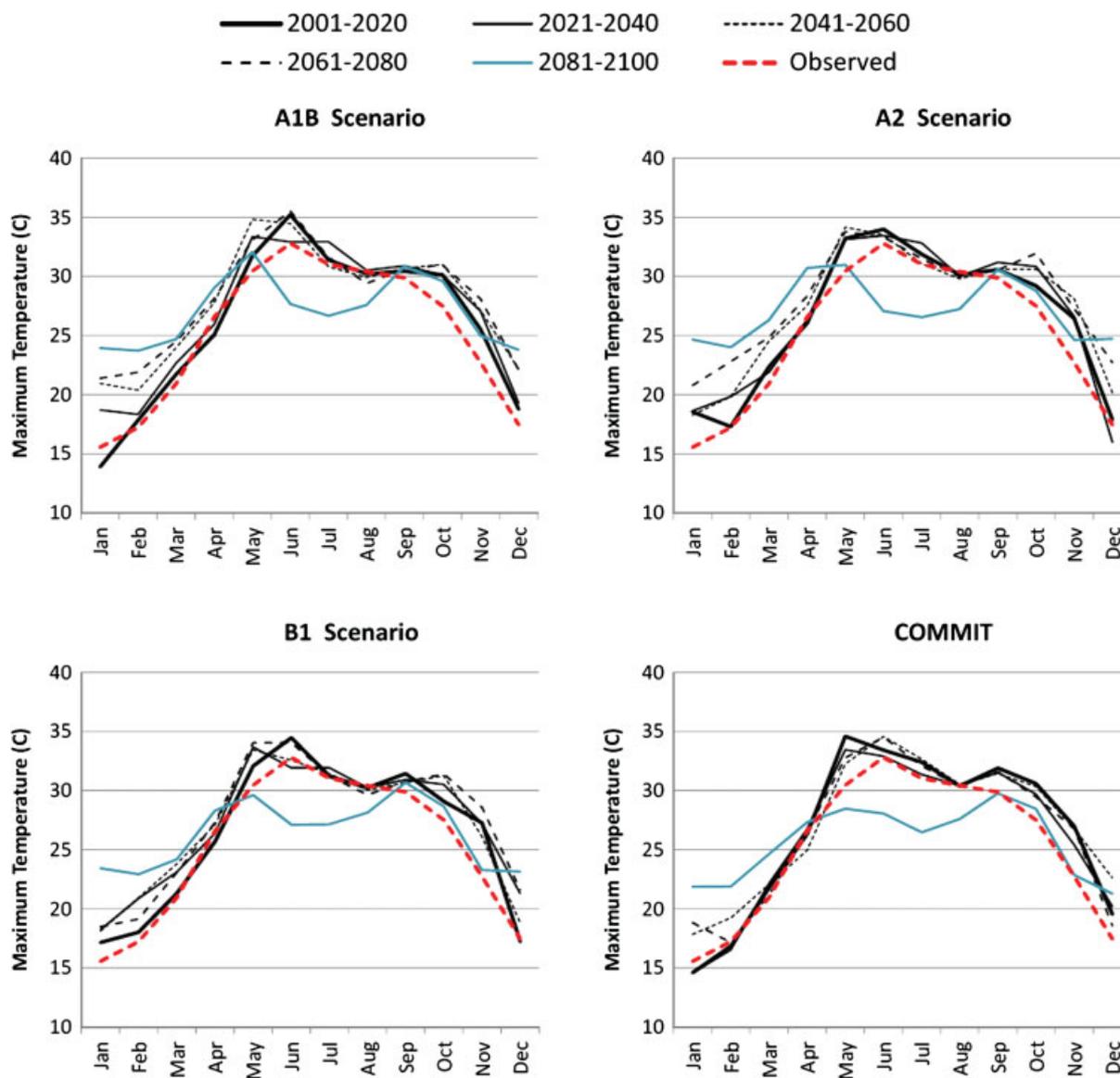


Figure 7. Observed (historical) and projected/downscaled values of average monthly maximum temperature at Bhuntar for climate change scenarios.

temperature is better when compared with that in deriving daily maximum temperature.

Downscaled daily values of the predictands corresponding to GCM-based predictor data for 20C3M and each of the four emission scenarios (i.e. A1B, A2, B1 and COMMIT) were obtained using the procedure described in step (9) of Section 3.1.

Correlations between the observed and the GCM-based downscaled monthly values of maximum temperature for the sites Bhuntar, Larji, Manali and Pandoh during the period April 1985 to March 2010 were found to be 0.83, 0.81, 0.83 and 0.79 respectively. The same for minimum temperature for the sites were 0.85, 0.82, 0.84 and 0.76 respectively. These results indicate that CGCM3 was able to simulate climate of the Beas basin reasonably well for the historical period, and hence future projections based on the GCM can be considered reliable.

To discern information contained in the downscaled future values of predictands, future projections of annual average values of a predictand for each of the sites were computed based on the downscaled monthly values of the predictand for the sites. Results indicated an increase in annual average maximum and minimum temperatures at all the sites for A1B, A2 and B1 scenarios. The projected increment is high for A2 scenario, and it is followed by that for A1B, B1 and COMMIT scenarios.

To scrutinize future changes projected in mean monthly values of a predictand for each climate change scenario, the projected/downscaled future monthly values of the predictand were divided into five sets (2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100) and average monthly values of the predictand were computed for each of the sets. Mean monthly maximum temperature is projected to increase in future at Bhuntar site

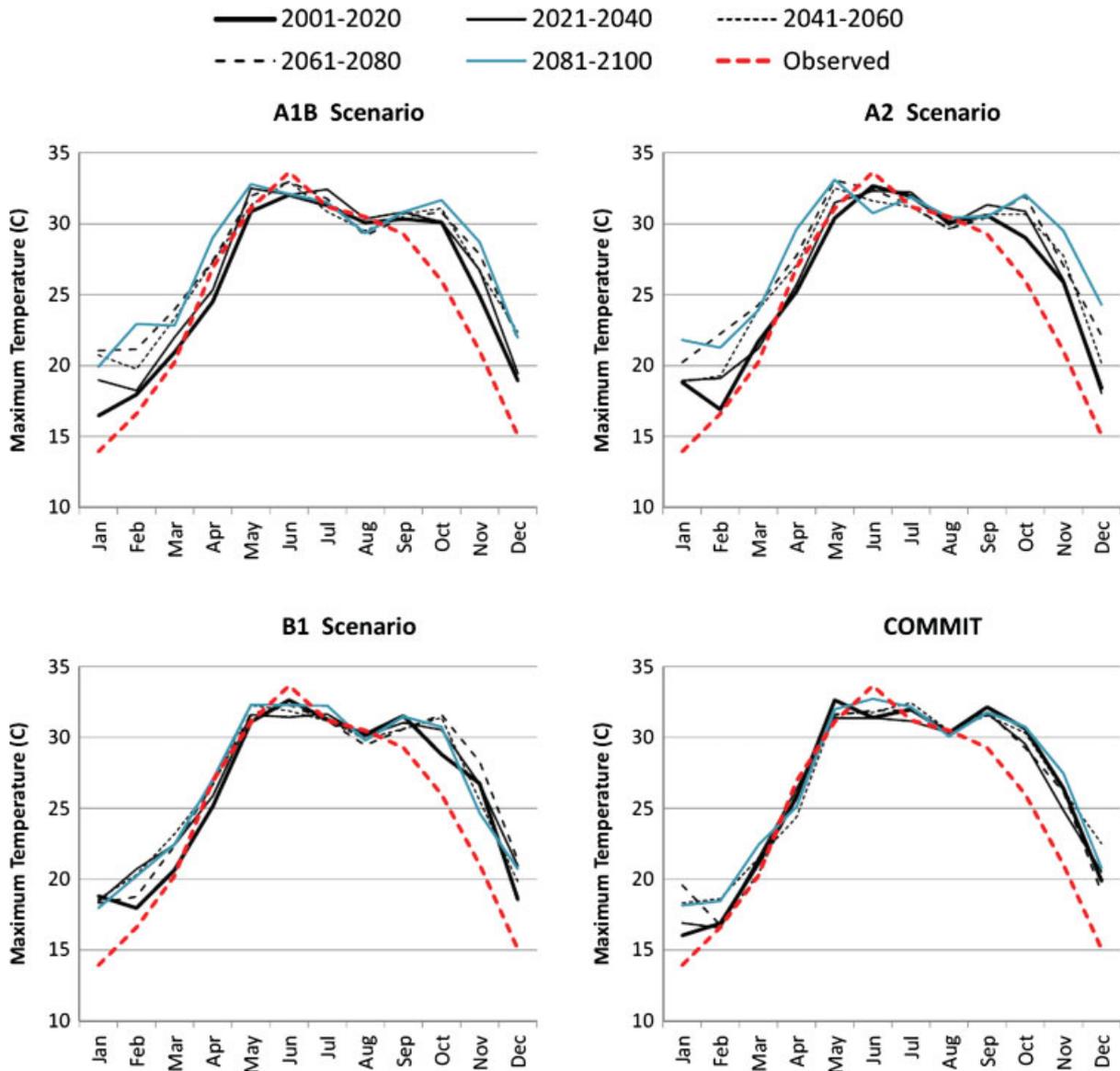


Figure 8. Observed (historical) and projected/downscaled values of average monthly maximum temperature at Larji for climate change scenarios.

during January to April for A1B, A2 and B1 scenarios. Further, decrease in the value of the statistic is projected for June to August for all the scenarios over the period 2081–2100 (Figure 7). At Larji site mean monthly maximum temperature is projected to increase in future during January to March and September to December (Figure 8). At Manali, future mean monthly maximum temperature is projected to decrease during April to August, and increase during January to February (over the period 2021–2080) and October to December (Figure 9). At Pandoh site, value of the statistic is projected to increase in future for almost all the months for A1B, A2 and B1 scenarios, and during June to December for COMMIT scenario (Figure 10). Further, mean monthly minimum temperature is, in general, projected to decrease in future during June to July and increase during January to February and October to December for all the scenarios (not shown for brevity).

**6. Summary and Concluding Remarks**

LS-SVM-based methodology has been developed for multi-site downscaling of maximum and minimum daily temperature series. The effectiveness of the methodology is demonstrated by application to four sites in the catchment of Beas river basin, India. Predictor variables influencing predictands in the study area have been identified from LSAVs in NCEP reanalysis data, and Canadian Coupled Global Climate Model (CGCM3.1/T63) simulations for 20C3M scenario. The GCM simulations for IPCC SRES scenarios were downscaled to future monthly projections of the predictands at Bhuntar site. The downscaled monthly information was translated from Bhuntar to that at other three sites, and then temporally disaggregated to arrive at daily time series of predictands. The performance of the model was found to be better when compared with recently proposed

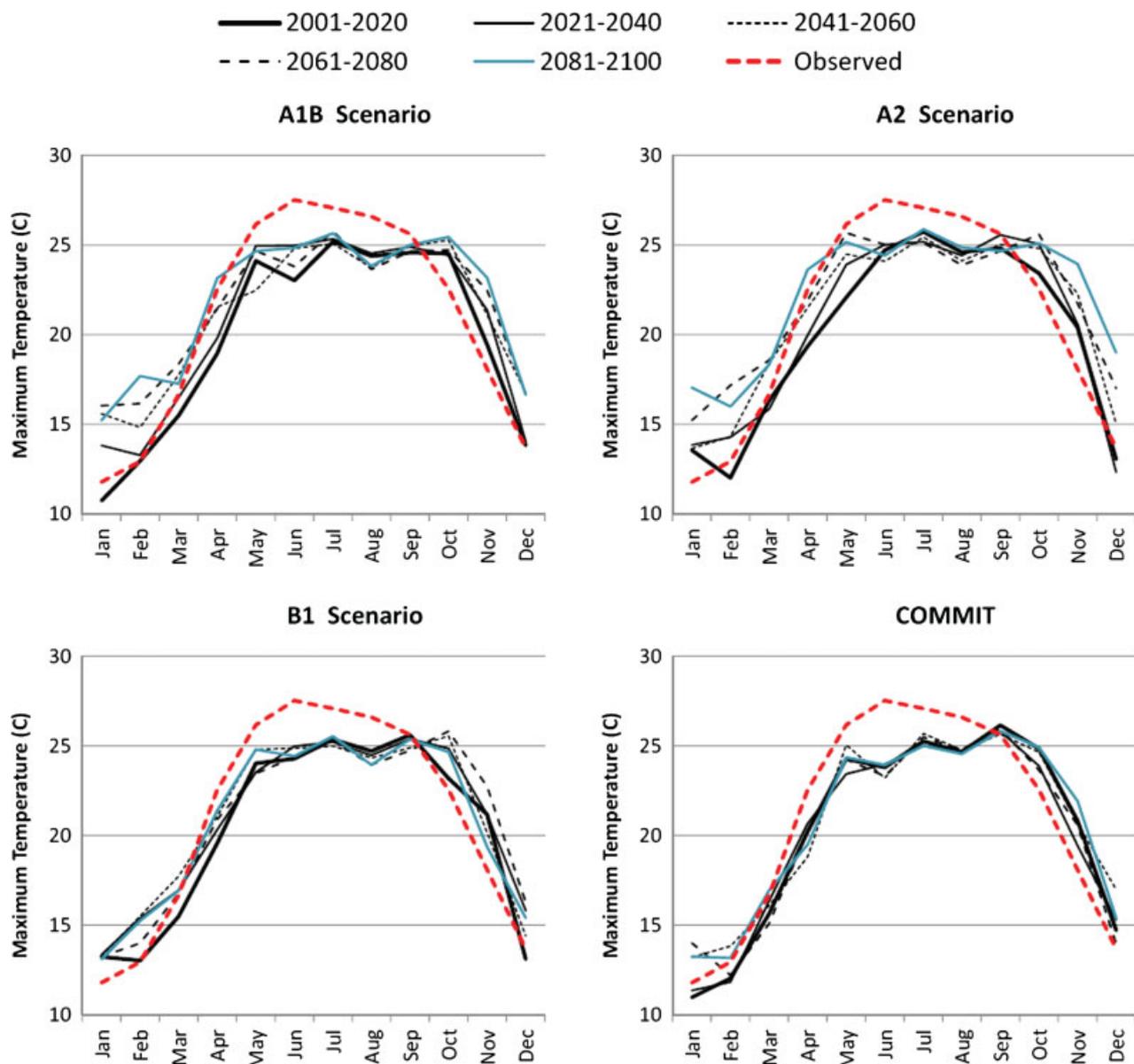


Figure 9. Observed (historical) and projected/downscaled values of average monthly maximum temperature at Manali for climate change scenarios.

MMLR-based downscaling method and MMSD method. Results indicated an increase in annual average maximum and minimum temperatures at all the sites for A1B, A2 and B1 scenarios. The projected increment is high for A2 scenario, and it is followed by that for A1B, B1 and COMMIT scenarios. Further, in general, future mean monthly maximum and minimum temperatures are projected to increase during January to March and October to December. It is necessary to apply the procedure suggested in this work to simulations from other GCMs for strengthening the conclusions drawn in this study.

In implementing the proposed methodology a modeller can explore the option of developing separate model for different seasons/months to see if it offers better results keeping in view parsimony considerations. Further scope involves accounting for biases in GCMs. Assumptions

are made in the development of GCMs (in terms of parameterizations and empirical equations) owing to incomplete knowledge on underlying geophysical processes, and those result in biases for climate variables between GCM output and observed records. Therefore bias correction of GCM outputs is necessary before they are used in downscaling studies for regional impact assessment (Frost *et al.*, 2011). Various approaches are in use for this purpose that include simple options such as mean correction and standardization (Wilby *et al.*, 2004), and relatively advanced options such as bias correction and spatial disaggregation method (Wood *et al.*, 2002) and its variations (e.g. Ahmed *et al.*, 2013), quantile-based mapping (Li *et al.*, 2010), and bias correction and constructed analogues method (Maurer and Hidalgo, 2008). In the present study, standardization is considered for bias correction. It would be interesting to verify if

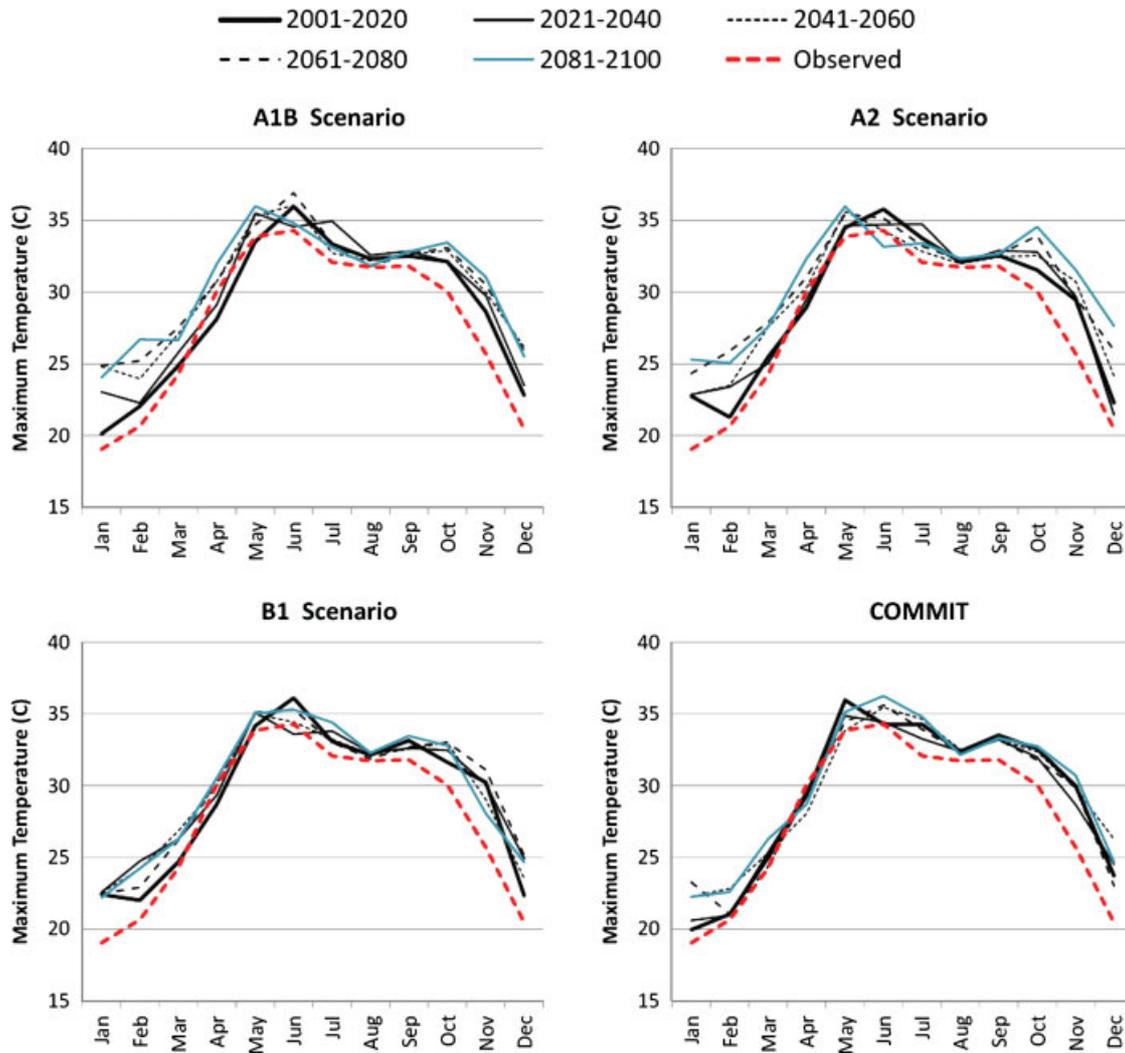


Figure 10. Observed (historical) and projected/downscaled values of average monthly maximum temperature at Pandoh for climate change scenarios.

implementation of a more advanced option for bias correction would lead to different downscaling results. This is deferred for future research.

### Acknowledgements

The authors would like to express their gratitude to the anonymous reviewers for their constructive and helpful comments that were helpful in improving quality of the manuscript. The authors are thankful to National Institute of Hydrology (NIH), Roorkee, India, for sponsoring the study, and to Bhakra Beas Management Board (BBMB), Pandoh, India, for providing the meteorological data used in this study.

### References

Ahmed KF, Wang G, Silander J, Wilson AM, Allen JM, Horton R, Anyah R. 2013. Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global and Planetary Change* **100**: 320–332.

- Anandhi A, Srinivas VV, Nanjundiah RS, Nagesh Kumar D. 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28**(3): 401–420. DOI: 10.1002/joc.1529
- Anandhi A, Srinivas VV, Nagesh Kumar D, Nanjundiah RS. 2009. Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **29**(4): 583–603. DOI: 10.1002/joc.1719
- Anandhi A, Srinivas VV, Nagesh Kumar D, Nanjundiah RS. 2012. Daily relative humidity projections in an Indian river basin for IPCC SRES scenarios. *Theoretical and Applied Climatology* **108**: 85–104. DOI: 10.1007/s00704-011-0511-z
- Bardossy A, Van Mierlo JMC. 2000. Regional precipitation and temperature scenarios for climate change. *Hydrological Sciences Journal des Sciences Hydrologiques* **45**(4): 559–575.
- Brown C, Greene A, Block P, Giannini A. 2008. Review of downscaling methodologies for Africa climate applications. IRI Technical Report 08-05, International Research Institute for Climate and Society, Columbia University, USA.
- Doty B, Kinter JL III. 1993. The Grid Analysis and Display System (GrADS): a desktop tool for earth science visualization. In *American Geophysical Union. 1993 Fall Meeting*, San Francisco, CA, 6–10 December.
- Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modeling to impacts studies: recent advances in downscaling techniques for hydrological modeling. *International Journal of Climatology* **27**: 1547–1578. DOI: 10.1002/joc.1556

- Frost AJ, Charles SP, Timbal B, Chiew FHS, Mehrotra R, Nguyen KC, Chandler RE, McGregor JL, Fu G, Kirono DGC, Fernandez E, Kent DM. 2011. A comparison of multi-site daily rainfall downscaling techniques under Australian conditions. *Journal of Hydrology* **408**: 1–18.
- Gestel TV, Suykens JAK, Baesens B, Viaene S, Vanthienen J, Dedene G, Moor BD, Vandewalle J. 2004. Benchmarking least squares support vector machine classifiers. *Machine Learning* **54**(1): 5–32.
- Ghosh S, Mujumdar PP. 2008. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources* **31**(1): 132–146. DOI: 10.1016/j.advwatres.2007.07.005
- Hessami M, Gachon P, Ouarda T, St-Hilaire A. 2008. Automated regression-based statistical downscaling tool. *Environmental Modelling & Software* **23**: 813–834.
- Hundecha Y, Bardossy A. 2008. Statistical downscaling of extremes of daily precipitation and temperature and construction of their future scenarios. *International Journal of Climatology* **28**: 589–610.
- IPCC. 2007. *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: New York.
- Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P. 2012. A multivariate multi-site statistical downscaling model for daily maximum and minimum temperatures. *Climate Research* **54**: 129–148.
- Johnson FM, Sharma A. 2012. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Research* **48**(1): W01504.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* **77**(3): 437–471.
- Khalili M, Nguyen VTV, Gachon P. 2011. A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. *International Journal of Climatology*. DOI: 10.1002/joc.3402
- Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research* **115**: D10101. DOI: 10.1029/2009JD012882
- Maurer EP, Hidalgo HG. 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences* **12**: 551–563.
- Nagesh Kumar D, Lall U, Peterson MR. 2000. Multi-site disaggregation of monthly to daily streamflow. *Water Resources Research* **36**(7): 1823–1833. DOI: 10.1029/2000WR900049
- Prudhomme C, Reynard N, Crooks S. 2002. Downscaling of global climate models for flood frequency analysis: where are we now? *Hydrological Processes* **16**: 1137–1150.
- Qian B, Corte-real J, Xu H. 2002. Multisite stochastic weather models for impact studies. *International Journal of Climatology* **22**: 1377–1397.
- Srinivas VV, Tripathi S. 2008. Statistical downscaling of regional precipitation under climate change scenarios using support vector machines. In *Hydrology and Hydraulics*, Singh VP (ed). Water Resources Publications, Highlands Ranch, CO; Chapter 15: 533–586.
- Suykens JAK. 2001. Nonlinear modelling and support vector machines. In *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, 287–294.
- Tripathi S, Srinivas VV, Nanjundiah RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330**(3–4): 621–640. DOI: 10.1016/j.jhydrol.2006.04.030
- Vapnik VN. 1995. *The Nature of Statistical Learning Theory*. Springer Verlag: New York.
- Vapnik VN. 1998. *Statistical Learning Theory*. Wiley: New York.
- Wilby RL, Dawson CW. 2007. Statistical downscaling model, SDSM, Version 4.2 – A decision support tool for the assessment of regional climate change impacts. User Manual.
- Wilby RL, Wigley TML. 1997. Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography* **21**: 530–548.
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA), 1–27.
- Wilby RL, Troni J, Biot Y, Tedd L, Hewitson BC, Smith DM, Sutton RT. 2009. A review of climate risk information for adaptation and development planning. *International Journal of Climatology* **29**: 1193–1215.
- Wilks DS. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change* **22**: 67–84.
- Wilks DS. 1999. Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agricultural and Forest Meteorology* **96**: 85–101.
- Wilks DS. 2010. Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change* **1**: 898–907.
- Wilks DS. 2012. Stochastic weather generators for climate-change downscaling, part II: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change* **3**(3): 267–278.
- Wood AW, Maurer EP, Kumar A, Lettenmaier DP. 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research* **107**(D20): 4429. DOI: 10.1029/2001JD000659
- Xu CY. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography* **23**(2): 229–249.