A Multivariate Statistical Approach to Downscaling of Daily Extreme Temperature Processes for Different Locations

CCTC 2013 Paper Number 1569695357

Mahzabeen Rahman[1], Malika Khalili[2] and Van Thanh Van Nguyen[3]
[1] PhD student, McGill University, Department of Civil Engineering and Applied Mechanics
[2] Research Associate, McGill University, Department of Civil Engineering and Applied Mechanics.
[3] Professor, McGill University, Department of Civil Engineering and Applied Mechanics.

Abstract

A multivariate statistical downscaling technique has been developed for modeling concurrently the daily maximum (Tmax) and minimum (Tmin) temperature series at different locations in the context of climate change. The proposed approach consists of a combination of a multiple regression model to represent the linkage between global climate predictors and local Tmax and Tmin series, and a Singular Value Decomposition (SVD) technique to describe the observed statistical properties of the regression model stochastic components. Results of a numerical application using daily extreme temperature data over Bangladesh have indicated the feasibility and accuracy of the suggested multisite statistical downscaling procedure.

Keywords: multisite downscaling, daily temperature extremes, statistical modeling, singular value decomposition

Résumé

Une technique de réduction d’échelle statistique à plusieurs variables a été élaborée pour la modélisation simultanée de séries de températures maximales (T maxi) et minimales (T mini) journalières à différents endroits dans le contexte des changements climatiques. La méthode proposée consiste en une combinaison d’un modèle de régression multiple visant à représenter le lien entre les prédicteurs relatifs au climat global et les séries de T maxi et T mini locales d’une part et, d’autre part, d’une technique de décomposition en valeurs singulières (SVD) afin de décrire les propriétés statistiques observées des composants stochastiques du modèle de régression. Les résultats d’une application numérique utilisant les données de températures extrêmes journalières au Bangladesh confirment la faisabilité et l’exactitude de la procédure de réduction d’échelle statistique multisite suggérée.

Mots clés: réduction d’échelle multisite, températures extrêmes journalières, modélisation statistique, décomposition en valeurs singulières

1. Introduction

Spatial resolution of Global Climate Model (GCM) output is too coarse to resolve regional scale effect and to be used directly in climate change impact studies. Hence, over the past years, the scientific community has focused on the development of innovative techniques to bridge the gap of scales, known as ‘downscaling’ [1,2]. Several extensive research works have been carried out to develop and assess the performance of various statistical downscaling techniques. In general, a majority of these downscaling methods were developed for the downscaling of hydrologic processes at a single site, but very few studies have dealt with multi-site downscaling to preserve simultaneously the spatial and temporal dependence. However, daily meteorological data especially temperature exhibit a strong temporal persistence at a given site and a high spatial dependence between different sites over a given area. Ignoring this temporal and spatial dependence in the downscaled daily Tmax and Tmin series at different sites may affect the accuracy of the climate change impact assessment results. For example, in the study by [3], it was shown that the hydrological modeling using the single-site weather generator resulted in a significant underestimation of extreme stream flows especially in summer and autumn seasons. Therefore, for accurate assessments of the effects of climate change, considerable interest is growing for multisite downscaling. For instance, Khalili et al [4] have developed a multi-site multivariate statistical downscaling approach for simulating daily maximum and minimum temperatures at many locations simultaneously based on a multiple regression model, in which a spatial moving average method was used to simulate the stochastic components of the proposed multiple regression model. This approach was capable to accurately reproduce the spatial and temporal dependence and produced better results than those of the Canadian Regional Climate Model (CRCM).

In the present paper, an efficient multisite multivariate statistical downscaling technique was applied in order to simulate simultaneously both daily Tmax and Tmin series at different locations in the context of climate change. The proposed approach consists of a combination of a multiple regression model to represent the linkage between global climate predictors and local temperature series, and SVD technique to reproduce the observed statistical properties of the regression model stochastic component. This approach was first developed by [5] and has been applied to the daily temperature series over southwest region of Quebec and southeast region of Ontario in Canada. The suggested method was found effective in reproducing accurately the statistical characteristics of daily temperature series at a given site as well as the temporal persistence, and spatial dependence among these stations. In the present paper, this method was tested using the daily temperature data over Bangladesh, which represents completely different climatic and physiographic conditions. Results of this test have indicated the feasibility of the proposed multisite multivariate statistical downscaling method. In addition, it was found that results of this downscaling method were more accurate than those given by the popular single-site statistical downscaling procedure called Statistical DownScaling Model (SDSM) [6].

2. Methodology

As mentioned above, the concept of the proposed multisite multivariate statistical downscaling approach is based on the combination of a multiple regression model and SVD technique. Equation 1 represents the regression model for daily Tmax [5].

$$\begin{pmatrix} T\max_{i,m,1} \\ T\max_{i,m,2} \\ . \\ . \\ T\max_{i,m,n} \end{pmatrix} = \begin{pmatrix} \alpha_{T\max_{0,m,1}} \\ \alpha_{T\max_{0,m,2}} \\ . \\ . \\ \alpha_{T\max_{0,m,n}} \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^{q} \alpha_{T\max_{j,m,1}} \, p_{T\max_{ij,m,1}} \\ \sum_{j=1}^{q} \alpha_{T\max_{j,m,2}} \, p_{T\max_{ij,m,2}} \\ . \\ . \\ \sum_{j=1}^{q} \alpha_{T\max_{j,m,n}} \, p_{T\max_{ij,m,n}} \end{pmatrix} + \begin{pmatrix} E_{T\max_{i,m,1}} \\ E_{T\max_{i,m,2}} \\ . \\ . \\ E_{T\max_{i,m,n}} \end{pmatrix} \tag{1}$$

Where, $T\max_{i,m,k}$ is the maximum temperature on day i in month m and at station k from a network of n stations; $\alpha_{T\max_{j,m,k}}$ is the j[th] regression parameter for month m and at station k; $p_{T\max_{ij,m,k}}$ is the value of the j[th] predictor, from a total of q monthly significant predictors, on day i in month m and at station k; $E_{T\max_{i,m,k}}$ is the residual value on day i in month m and at station k. The above equation is only for Tmax but the similar representation is applicable for Tmin as well.

In this study, stepwise regression with backward selection [7] is used to select the significant predictors for both Tmax and Tmin, for each month, and for each station. In order to address the issue of multicollinearity and model parsimony, only five predictors are selected for the regression model [5].

Residuals of Tmax and Tmin, which are the stochastic components of the regression models have been generated simultaneously by SVD technique. In fact, after estimating the regression parameters, the general matrix of the residual values E is obtained by the concatenation of the matrices $E_{T\max}$ and $E_{T\min}$, whose elements are the residuals obtained for daily Tmax and Tmin, respectively :

$$E = \left[E_{T\max} \;\; E_{T\min}\right] \tag{2}$$

The rows of the residual matrix E are the records in years and the columns represent the residuals for each day, for each station, and for both Tmax and Tmin. Therefore, for 20-year calibration period ranging from 1981 to 2000, for the used four stations (Table 1), and for both daily Tmax and Tmin, the dimension of the residual matrix E will be 20×2920. Short description of the procedural steps of generating the residual matrix E by SVD technique is given in the following section.

i) Compute the $m \times n$ E_s matrix of the standardized values of E.

$$E_{sij} = \frac{E_{ij} - \overline{E}_j}{S_j} \tag{3}$$

Where, E_{ij} is the value of E at the i^{th} row and j^{th} column. $\overline{E}_j$ and S_j are the mean and standard deviation of the j^{th} column respectively.

ii) Compute the $n \times n$ correlation matrix P_{E_s} of matrix E_s .

$$P_{E_s} = \frac{1}{m-1} \times {E_s}^T \times E_s \tag{4}$$

iii) Decompose the matrix P_{E_s} using SVD technique [8, 9] to obtain the $n \times n$ matrix $D_{P_{E_s}}$, which is the matrix of the eigenvalues of P_{E_s}, and the $n \times n$ matrix $V_{P_{E_s}}$, which is the matrix of the corresponding eigenvectors of P_{E_s}.

$$P_{E_s} = V_{P_{E_s}} \times D_{P_{E_s}} \times {V_{P_{E_s}}}^T \tag{5}$$

If there are only r non-zero eigenvalues (r is the rank of E_s), the $(r \times r)$ matrix of eigenvalues $D_{P_{E_s}}$ is used in the computation of the $n \times r$ matrix of eigenvectors $V_{P_{E_s}}$.

iv) Given these matrices, compute the $m \times r$ standardized principal components matrix Z of E_s [10].

$$Z = E_s \times V_{p_{E_s}} \times {D_{p_{E_s}}}^{-\frac{1}{2}} \tag{6}$$

Here, Z is the matrix of the standardized principal components of E_s transforming the large number of correlated variables into a small number of uncorrelated variables, which account for the maximum of variance. The principal components are the linear combinations of the standardized variables and therefore, are closer to the normal distribution. Hence, depending on this principle, it is possible to generate the principal components independently following a standard normal distribution with zero mean and unit variance.

v) The simulated matrix of the principal components $\hat{Z}$ is then used to obtain the simulated standardized residual matrix $\hat{E}_s$ [10].

$$\hat{E}_s = \hat{Z} \times {D_{p_{E_s}}}^{\frac{1}{2}} \times {V_{p_{E_s}}}^T \tag{7}$$

vi) Finally, compute the non-standardized simulated matrix $\hat{E}$.

$$\hat{E}_{ij} = \hat{E}_{sij} \times S_j + \overline{E}_j \tag{8}$$

3. Numerical Application

3.1 Study Data

As mentioned previously, the feasibility of the proposed downscaling method is tested using the daily extreme temperature records, Tmax and Tmin, available for the 1961-2000 period from a network of four weather stations located in Bangladesh (Table1). In addition, the climate predictors used in the present study are the National Centers for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR) re-analysis data [11], which have been linearly interpolated to match the Gaussian grids of the third version of the Canadian Centre for Climate Modelling and Analysis Coupled Global Climate Model (CGCM3) [12]. Furthermore, the data for the 1981-2000 period were used for calibration, and the remaining data for the 1961-1980 period were used for the validation purposes.

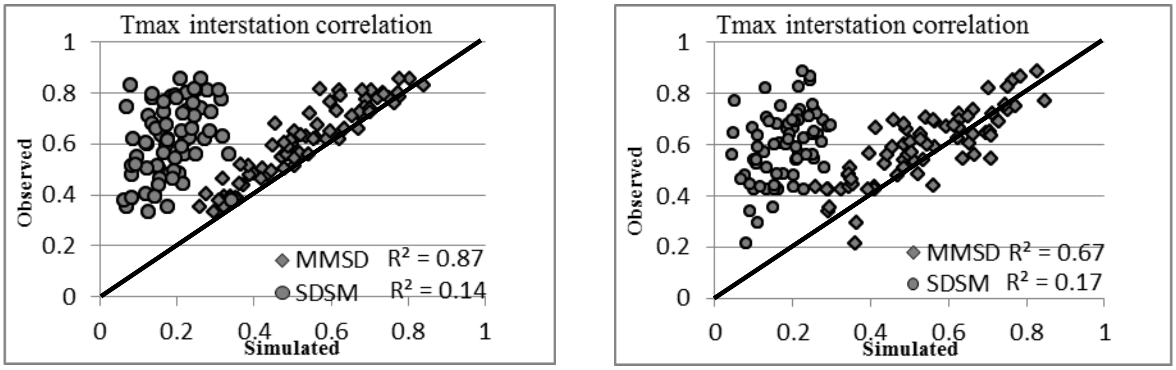
Table 1: Names and coordinates of four weather stations used in this study.

Station No	Station Name	Latitude	Longitude	Altitude (metres above sea
Station 1	Bogra	24.85 ^{0} N	89.37^0 E	17.9
Station 2	Jessore	23.2^0 N	89.33^0 E	6
Station 3	Dhaka	23.78^0 N	90.38^0 E	6.5
Station 4	Chittagong	22.35^0 N	91.82^0 E	33.2

Source: [13]

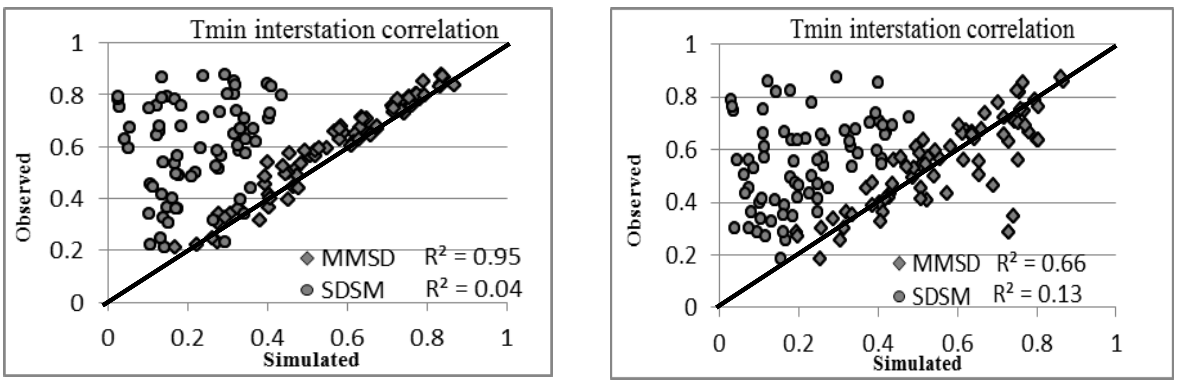
3.2 Results

Using the proposed multisite multivariate statistical downscaling approach, a set of 100 simulations of daily Tmax and Tmin time series for all stations and for both calibration and validation periods was generated. Results regarding spatial dependence are shown in the scatter plots (Figures 1 and 2), which represent the observed versus simulated monthly interstation correlation for daily Tmax and Tmin respectively. It was found that the monthly inter-station correlations are well reproduced for both daily Tmax (R^2=0.87) and Tmin (R^2=0.95) for the calibration period as well as for the validation period (R^2=0.67 and 0.66 for daily Tmax and Tmin, respectively) (Figures 1 and 2). In addition, it can be seen that the single-site downscaling SDSM (Statistical DownScaling Model) procedure is not capable to capture the spatial dependency among the stations. In fact, for daily Tmax, the R^2 values between the observed and simulated monthly interstation correlations are 0.14 and 0.17 for the calibration and validation periods, respectively (Figure 1). As shown by Figure 2, lower R^2 values are obtained for Tmin (0.04 and 0.13 for the calibration and validation periods, respectively).



(a) (b)

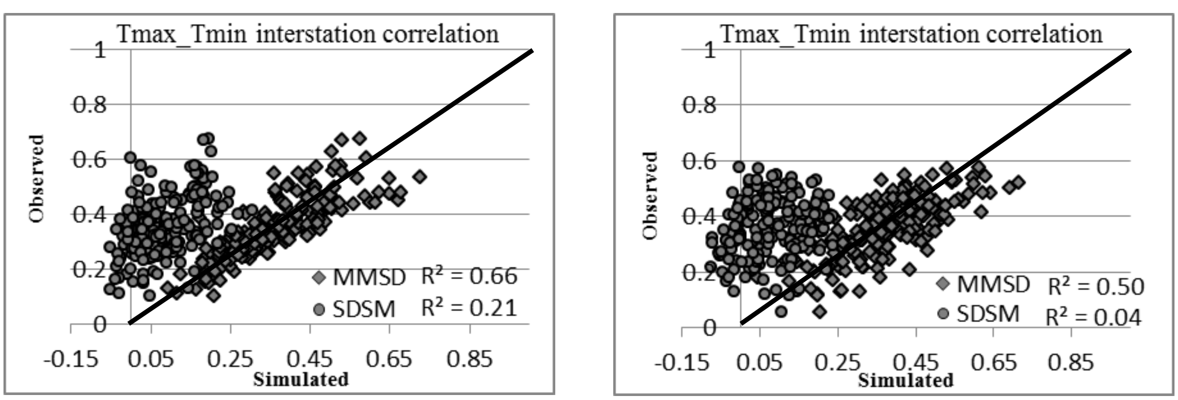
Figure 1: Monthly interstation correlations for daily Tmax over (a) calibration and (b) validation period. Here, MMSD represents Multisite Multivariate Statistical Downscaling technique.



(a) (b)

Figure 2: Same as Figure 1 but for daily Tmin.

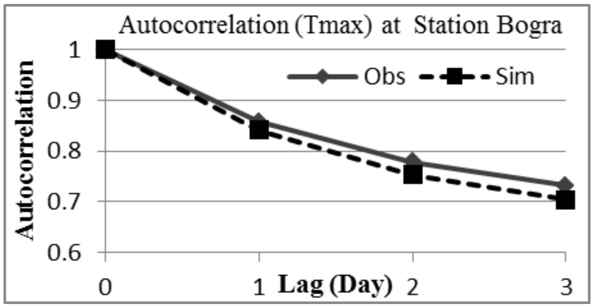
Furthermore, Figure 3 shows the monthly correlations between daily Tmax and Tmin at each station and at each pair of stations over both the calibration and validation periods. It can be seen that the proposed approach can reproduce quite well these correlations and can provide better results than the SDSM.

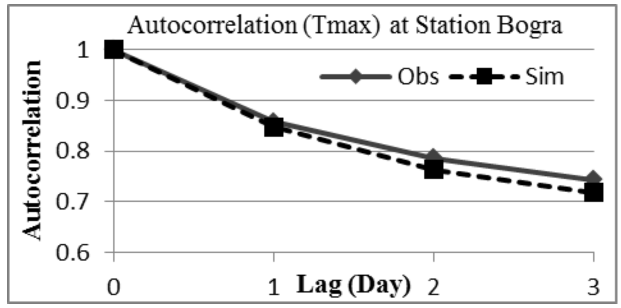


(a) (b)

Figure 3: Monthly correlations between daily Tmax and Tmin at each station and pair of stations for (a) calibration and (b) validation period.

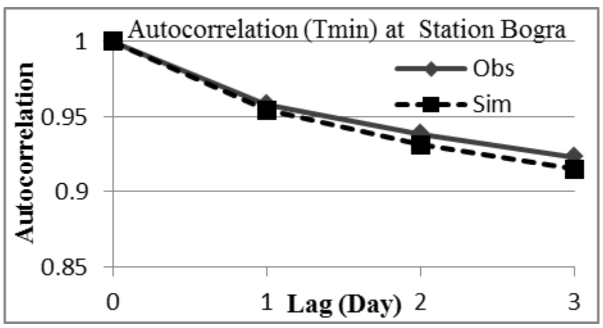
Temporal dependencies are evaluated in terms of reproducing the autocorrelations in the downscaled temperature series at each station. The observed versus simulated autocorrelations of lags from one to three are presented in Figures 4 and 5 for Tmax and Tmin, respectively, at station Bogra. These figures illustrate the ability of the proposed approach to reproduce very well the observed autocorrelation for Tmax and Tmin series for both calibration and validation periods. Similar results were also obtained for the other stations.

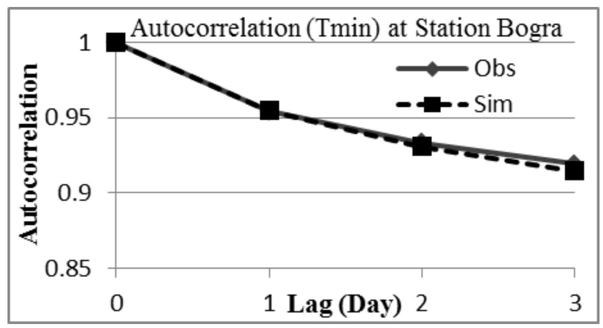




(a) (b)

Figure 4: Autocorrelation of Tmax for (a) calibration and (b) validation period at station Bogra.





(a) (b)

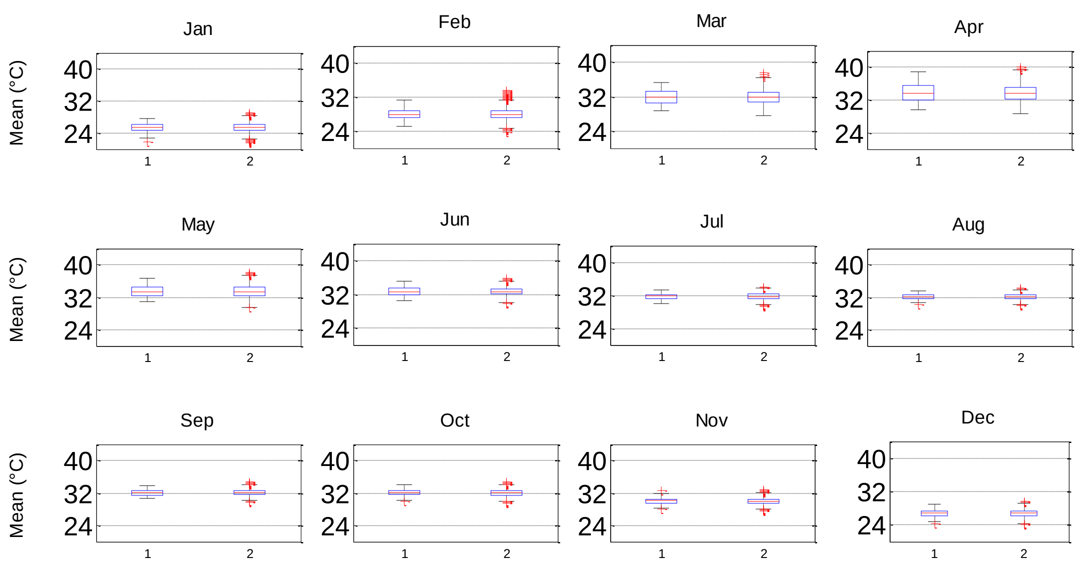
Figure 5: Same as Figure 4 but for daily Tmin.

Four temperature indices, the mean (Mean) and standard deviation (STD) of Tmax and Tmin, the 90th percentile of daily Tmax (Tmax90p), and the 10th percentile of daily Tmin (Tmin10p) were computed for each month, and for both the calibration and validation periods. The suggested downscaling model can reproduce well the monthly means and standard deviations of daily Tmax and Tmin for both calibration and validation periods. For Tmax90p and Tmin10p indices, a very good agreement between observed and downscaled values was also found. However, there are some differences in the median and variability of the downscaled series compared with the observed ones, especially over the validation period. In general, the proposed approach has been found capable to reproduce these basic statistics. For purposes of illustration, only Figures 6 and 7 are presented here to show the good agreement between the observed and estimated boxplots of the monthly mean of Tmax and the STD of Tmin, respectively. Finally, Table 2 shows the monthly average values of the coefficient of determination (R^2), the Mean Absolute Error (MAE), and the Root Mean Square Error (RMSE) based on 100 simulations of daily Tmax and Tmin time series over both the calibration and validation periods for all four stations. The expressions of R^2, MAE and RMSE are given by the equation 9, 10 and 11 respectively. It can be seen that very accurate results were obtained as indicated by the high R^2 value (and the very low values of MAE and RMSE) for both Tmax and Tmin over the calibration and validation periods.

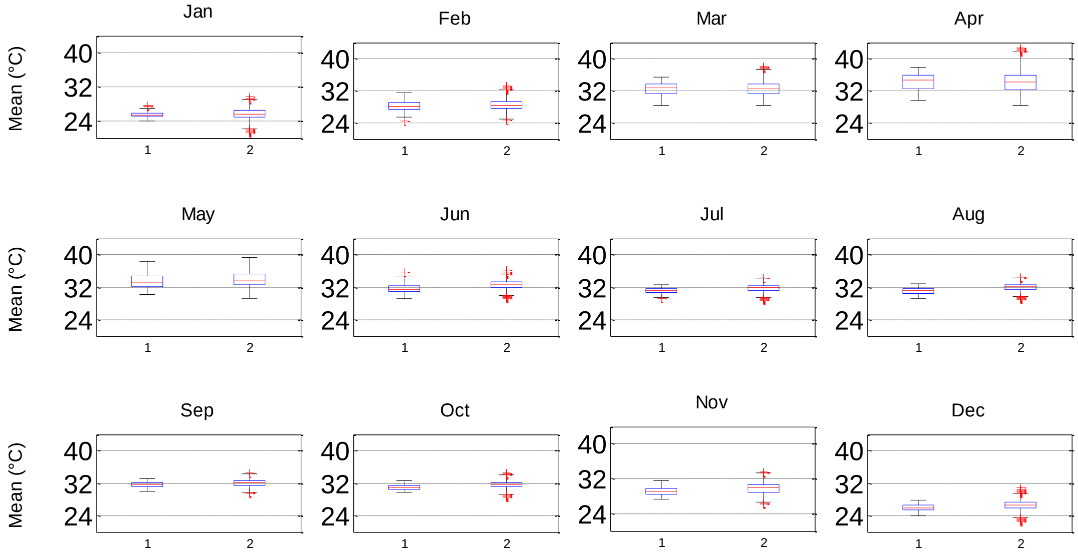
$$R^2 = 1 - \frac{\sum_{i=1}^{t}\left(X_{Obs,i,s} - X_{Sim,i,s}\right)^2}{\sum_{i=1}^{t}\left(X_{Obs,i,s} - \overline{X}_{Obs}\right)^2} \quad (9)$$

$$MAE = \frac{1}{t}\sum_{i=1}^{t}\left|X_{Obs,i,s} - X_{Sim,i,s}\right| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{t}\sum_{i=1}^{t}\left(X_{Obs,i,s} - X_{Sim,i,s}\right)^2} \quad (11)$$

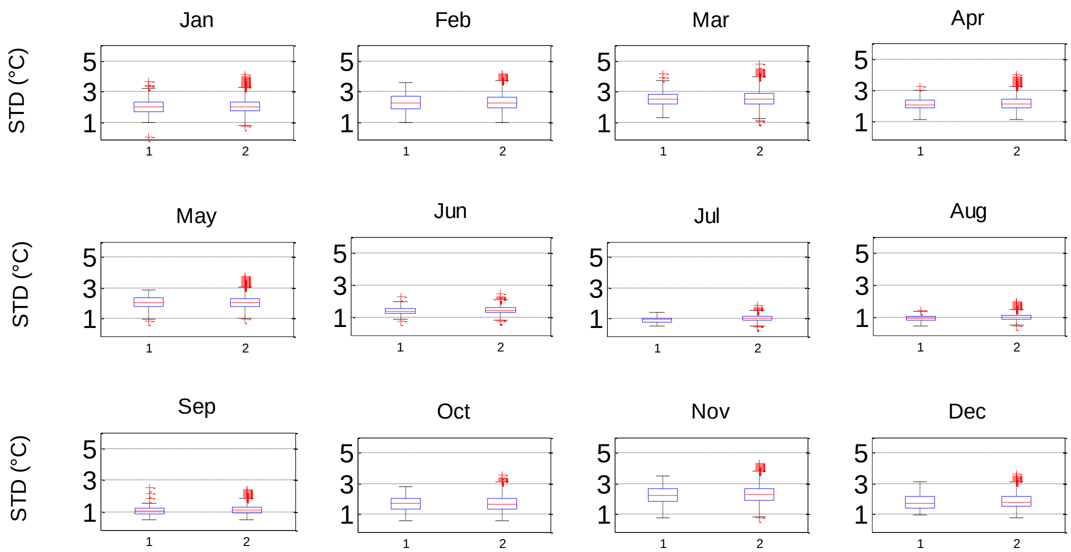


(a)

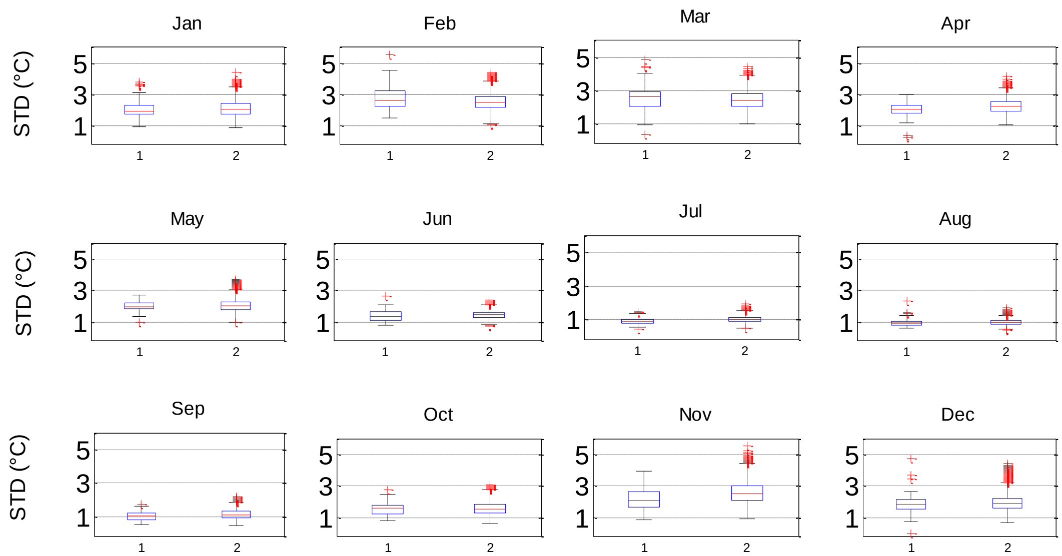


(b)

Figure 6: Boxplots of the mean of Tmax for the observed and downscaled series for (a) the calibration and (b) the validation period. Each boxplot represents the daily time series obtained at all stations aggregated over each period of 20 years. Here, 1 and 2 represent observed and simulated data, respectively. The band in the middle of each box represents the median value, the boxes and whiskers represent the inter-quartile range (IQR) and $1.5 \times IQR$ respectively. Crosses beyond the whiskers represent outliers.



(a)



(b)

Figure 7: Same as Figure 6 but for STD of Tmin.

Table 2: Coefficients of determination (R^2), MAE, and RMSE between the observed and simulated monthly Tmax and Tmin for both the calibration and validation periods at all stations.

Station	Tmax (Calibration)			Tmax (Validation)			Tmin (Calibration)			Tmin (Validation)		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
Station1	1	0.00	0.00	0.94	0.58	0.69	1	0.07	0.10	0.99	0.30	0.40
Station2	1	0.z00	0.00	0.94	0.64	0.74	1	0.04	0.06	0.99	0.47	0.60
Station3	1	0.00	0.00	0.90	0.79	0.85	1	0.09	0.14	0.98	0.53	0.75
Station4	1	0.00	0.00	0.94	0.44	0.49	1	0.06	0.09	0.99	0.27	0.43

4 Conclusion

In the present study, a multivariate statistical technique has been applied for downscaling daily extreme temperature series at many sites concurrently. The approach consists of a combination of multiple regression models to represent the linkage between global climate predictors and local Tmax and Tmin series, and the SVD technique to describe the observed statistical properties of the regression model stochastic component. Results of a numerical application using data from a network of four weather stations located in Bangladesh have indicated the feasibility and accuracy of this multisite multivariate statistical downscaling method. In particular, it has been shown that the SVD technique was able to preserve accurately the observed basic statistical properties (mean, standard deviation, correlations) of the historical daily extreme temperature records at a single site as well as for different locations. In addition, it has been demonstrated that the proposed multisite multivariate model could be quite efficient in the simulation of daily temperature extremes for a large number of sites based on the computational efficiency of the SVD method in the simulation of large matrices of correlated vectors.

5. References

1] Fowler H.J., Blenkinsop S., Tebaldi, C. , “Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling”,*International Journal of Climatology*, Vol.27, Iss.12, 2007, pp1547-1578.

[2] Xu C., “From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches”', *Progress in Physical Geography*, Vol. 23, Iss. 2, 1999, pp 229-249.

[3] Khalili M., Brissette F. , Leconte, R.,. “Effectivemness of multi-site weather generator for hydrological modeling”,*Journal of the American Water Resources Association*, Vol. 47, Iss. 2, 2011,pp 303-314.

[4] Khalili, M., Nguyen, V. T. V, Gachon, P., “A statistical approach to multi-site multivariate downscaling of daily extreme temperature series”, *International Journal of Climatology,* Vol. 33, Iss. 1, 2013, pp 15-32. Article first published online: 3 NOV 2011. DOI: 10.1002/joc.3402.

[5] Khalili, M., Nguyen, V. T. V., “A statistical approach to multi-site multivariate downscaling of daily extreme temperature series for climate-related impact assessment studies”, HIC 2012- 10th International Conference on Hydroinformatics, "Understanding Changing Climate and Environment and Finding Solutions", 14-18 July, Hamburg, Germany, 2012.

[6] Wilby, R., Dawson, C,. “Using SDSM Version 3.1–A decision support tool for the assessment of regional climate change impacts”, User Manual. London, UK, 2007.

[7] McCuen, R. H., “Modeling hydrologic change: statistical methods”, CRC Press, 2003.

[8] Salkind N J, Rasmussen K., “Encyclopedia of measurement and statistics*”,* Thousand Oaks, Calif.: SAGE Publications, 2007.

[9] Joreskog K. G., Klovan JE, Rayment RA., “Geological Factor Analysis”, Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York, 1978.

[10] Green, P. E., “Analyzing multivariate data”, Dryden Press (Hinsdale, Ill.), 1978.

[11] Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, Dool Hvd, Jenne R, Fiorino M, “The NCEP/NCAR 50-Year Reanalysis”,*Bulletin of the American Meteorological Society,* Vol. 82, Iss. 2, 2001, pp 247-267.

[12] DAI CGCM3 Predictors, 2008: ''Sets of Predictor Variables Derived From CGCM3 T47 and NCEP/NCAR Reanalysis'', version 1.1, April 2008, Montreal, QC, Canada, 15pp.

[13] Islam, A. S., "Analyzing Changes of Temperature Over Bangladesh Due to Global Warming Using Historic Data". Young scientists of Asia conclave, Jawaharlal Nehru Centre for Advanced Scientific Research Jakkur, 15-17 January, Bangalore, 2009.

6. Acknowledgements

The authors would like to acknowledge the Data Access Integration (DAI) Team for providing the data. The DAI Portal (http://loki.qc.ec.gc.ca/DAI/) is made possible through collaboration among the Global Environmental and Climate Change Centre (GEC3), the Adaptation and Impacts Research Division (AIRD) of Environment Canada, and the Drought Research Initiative (DRI). The authors also acknowledge Institute of Water and Flood Management, University of Engineering and Technology (IWFM, BUET), Bangladesh and Bangladesh Meteorological Department (BMD) for historical data of daily temperatures. The authors also appreciate the public release of SDSM. SDSM 4.1 was supplied by Drs Wilby and Dawson, on behalf of the Environment Agency of England and Wales.

7. Biographies

Mahzabeen Rahman earned master's degrees from Asian Institute of Technology, Thailand and Institute of Water and Flood Management, BUET, Bangladesh. Currently Ms. Rahman is pursuing PhD program in the Department of Civil Engineering and Applied Mechanics at McGill University under the supervision of Professor Van-Thanh-Van Nguyen.

Malika Khalili graduated (Ph. D) from the École de technologie supérieure of Montreal. Her thesis developed a new multi-site generation approach of climate data for efficient assessment of climate change impacts on the hydrology regime of medium and large size river basins. She is currently research associate at McGill University where she did her postdoctoral fellowship. Her research concerns modeling of hydrologic processes for climate-related impact assessment studies under the supervision of Professor Van-Thanh-Van Nguyen.

Van Nguyen is Professor and Chair of Department of Civil Engineering and Applied Mechanics as well as Director of the Brace Centre for Water Resources Management at McGill. His scientific and professional contributions over more than 30 years have been mostly in the areas of Hydrology and Water Resources Management. He is author or co-author of over 200 articles in refereed journals, specialized monographs and conference proceedings.