Evaluation of Rule and Decision Tree Induction Algorithms for Generating Climate Change Scenarios for Temperature and Pan Evaporation on a Lake Basin

Manish Kumar Goyal[1] and C. S. P. Ojha[2]

Abstract: Climate change scenarios generated by general circulation models (GCMs) have too coarse a spatial resolution to be useful in planning disaster risk reduction and climate change adaptation strategies at regional to river/lake basin scales. This paper investigates the performances of existing state-of-the-art rule induction and tree algorithms, namely, single conjunctive rule learner, decision table, M5P model tree, decision stump, and REPTree. Downscaling models are developed to obtain projections of mean monthly maximum and minimum temperatures (*Tmax* and *Tmin*) as well as pan evaporation to lake-basin scale in an arid region in India using these algorithms. The predictor variables, such as air temperature, zonal wind, meridional wind, and geo-potential height, are extracted from the National Centers for Environmental Prediction (NCEP) reanalysis data set for the period 1948–2000 and from the simulations using third-generation Canadian coupled global climate models for emission scenarios for the period 2001–2100. A simple multiplicative shift was used for correcting predictand values. The performances of various models have been evaluated on several statistical performance parameters such as correlation coefficient, mean absolute error, and root mean square error. The M5P model tree algorithm was found to yield better performance among all other learning techniques explored in the present study. An increasing trend is observed for *Tmax* and *Tmin* for emission scenarios, whereas no trend has been observed for pan evaporation in the future. **DOI: 10.1061/(ASCE)HE.1943-5584.0000795.** © *2014 American Society of Civil Engineers.*

Author keywords: Climate change; Intergovernmental Panel on Climate Change (IPCC) scenarios; India; Pichola Lake; Statistical downscaling.

Introduction

General Circulation Models (GCMs) are an important tool in the assessment of climate change. These models are based on well-established physical principles and have been demonstrated to reproduce observed features of recent climate and past climate changes (Fowler et al. 2007; Chu et al. 2010; Ojha et al. 2010). However, they remain relatively coarse in resolution and are unable to resolve significant subgrid scale features such as topography, clouds, and land use (Wilby et al. 2002; Anandhi et al. 2009; Goyal et al. 2012). Since confidence in the projection of GCMs decreases at smaller scales, other techniques, such as the use of regional climate models and downscaling methods, have been specifically developed for the study of regional-scale and local-scale climate change. The methods used to convert GCM outputs into local meteorological variables required for reliable hydrological modeling are usually referred to as *downscaling* techniques. Two fundamental approaches exist for the downscaling of large-scale GCM output to a finer spatial resolution. The first of these is a dynamical approach where a higher resolution climate model is embedded within a GCM. The second approach is to use statistical methods to establish empirical relationships between GCM-resolution climate variables and local climate. A dynamic downscaling approach has superior capability in complex terrain or with changed land cover (Kite 1997; Wang et al. 2004). However, this method entails higher computation cost and relies strongly on the boundary conditions provided by GCMs (Chu et al. 2010). In contrast, statistical downscaling gains local predictands by appropriate statistical or empirical relationships with surface or troposphere atmospheric predictors (Xu 1999). Since this method is comparatively cheap and computationally efficient and is as powerful as its dynamic competitor, it has been widely employed in climate change impact assessments (Wilby and Wigley 1997; Fowler et al. 2007). However, its drawback is that it requires a reliable observed historical data series for calibration and in order to build the appropriate statistical relationship (Chu et al. 2010).

Statistical downscaling methods are generally classified into three groups: (1) regression models, (2) weather typing schemes, and (3) weather generators (WGs). Each group covers a range of methods, all relying on the fundamental concept that regional climates are largely a function of the large-scale atmospheric state (Fowler et al. 2007). This relationship may be expressed as a stochastic and/or deterministic function between large-scale atmospheric variables (known as predictors) and local climate variables (known as predictands). In regression models, the transfer function

[1]Assistant Professor, Dept. of Civil Engineering, Indian Institute of Technology, Guwahati 781039, India; Research Fellow, DHI-NTU Water and Environment Research Centre and Education Hub, School of Civil and Environmental Engineering, Nanyang Ave., Nanyang Technological Univ., Singapore 639798; and Research Scholar, Indian Institute of Technology, Roorkee 247667, India (corresponding author). E-mail: vipmkgoyal@gmail.com

[2]Professor, Dept. of Civil Engineering, Indian Institute of Technology, Roorkee 27667, India.

Note. This manuscript was submitted on November 8, 2011; approved on January 25, 2013; published online on January 29, 2013. Discussion period open until September 1, 2014; separate discussions must be submitted for individual papers. This technical note is part of the ***Journal of Hydrologic Engineering***, Vol. 19, No. 4, April 1, 2014. © ASCE, ISSN 1084-0699/2014/4-828-835/$25.00.

is used to describe methods that directly quantify a relationship between the predictand and a set of predictor variables (Kang et al. 2007). Weather typing or classification schemes relate the occurrence of particular *weather classes* to local climate (Conway et al. 1996; Fowler et al. 2000, 2007). Weather generators in the simplest form are stochastic models, based on daily precipitation with a two state first-order Markov chain dependent on transition probabilities for simulating precipitation occurrence, and a gamma distribution for precipitation amounts (Mason 2004; Fowler et al. 2007). Several studies (Fowler et al. 2007; Anandhi et al. 2009) have shown that the statistical downscaling method is simple to handle and has, by and large, superior capability and is therefore widely applied (Wilby and Harris 2006). Downscaling has found wide application in hydroclimatology for scenario construction and simulation/prediction of (1) monthly mean temperature (Benestad 2001), (2) daily minimum and maximum temperature (Wilby et al. 2002), (3) daily mean, minimum and maximum temperature (Kettle and Thompson 2004), and (4) monthly minimum and maximum air temperature (Anandhi et al. 2009).

This paper explores existing state-of-the-art rule induction and tree algorithms, namely: (1) single conjunctive rule learner, (2) decision table, (3) M5 model tree, (4) decision stump, and (5) REPTree as a downscaling methodology to study climate change impact over the Pichola Lake basin in an arid region. In the literature, there is no evidence of any study dealing with simultaneous evaluation of various decision learning approaches. In light of this, the objective of this study is to (1) rank various rule induction and tree algorithms, (2) downscale mean monthly maximum temperatures (*Tmax*) and minimum temperatures (*Tmin*) as well as pan evaporation using the best available approach from simulations of the third-generation Canadian Coupled Global Climate Model (CGCM3) for the latest IPCC scenarios (IPPC 2007), and (3) apply a simple multiplicative shift to correct the GCM bias and compare the results with and without bias correction.

Study Region and Data Extraction

The study area is the Pichola Lake catchment in the Rajasthan province in India that is situated from 72.5°E to 77.5°E and 22.5°N to 27.5°N. The Pichola Lake basin, located in the Udaipur district, is one of the major sources for water supply for this region. The mean monthly *Tmax* in the catchment varies from 19°C to 39.5°C, and the mean annual *Tmax* is 30.6°C. The mean monthly *Tmin* ranges from 3.4°C to 29.8°C based on decadal (1990–2000) observed values. This region receives an average annual precipitation of 597 mm. For other details, readers are referred to studies by Khobragade (2009) and Goyal and Ojha (2012).

The monthly mean atmospheric variables were derived from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) (hereafter called NCEP) reanalysis data set (Kalnay et al. 1996) for the period of January 1975 to December 2000. The NCEP and NCAR accomplished different re-analysis projects that aimed to generate global data sets for a long time period for different atmospheric parameters (Kalnay et al. 1996). The data have a horizontal resolution of 2.5° latitude by 2.5° longitude and 17 constant pressure levels in the vertical. The atmospheric variables, such as air temperature, zonal wind, meridional wind, and geo-potential height, are extracted for nine grid points whose latitude ranges from 22.5°N to 27.5°N, and longitude ranges from 72.5°E to 77.5°E at a spatial resolution of 2.5°.

Rule Induction and Tree Algorithms

Single Conjunctive Rule Learner

Single conjunctive rule learner is one of the data mining learning algorithms and is normally known as inductive learning. The objective of rule induction is generally to induce a set of rules from data that confines all generalizable knowledge within that data and at the same time is as small as possible (Cohen 1995). Classification in rule induction classifiers is typically based on the firing of a rule on a test instance, triggered by matching feature values at the left-hand side of the rule (Clark and Niblett 1989). Rules can be of various normal forms and are typically ordered; with ordered rules, the first rule that fires establishes the classification outcome and halts the classification process (Mohd and Thomas 2007).

Decision Table

Decision table employs the wrapper method to find a good subset of attributes for inclusion in the table. This is done using a best-first search. The decision table algorithm constructs a decision rule using a simple decision table majority classifier, as proposed by (Kohavi 1995). It reviews the data set with a *decision table*, which contains the same number of attributes as the original data set.

M5 Model Tree

Model tree technique presents a structural representation of the data and a piecewise linear fit of the class (Quinlan 1992). Like conventional decision tree learners, M5 builds a tree by splitting the data based on the values of predictive attributes. Instead of selecting attributes by an information theoretic metric, M5 chooses attributes that minimize intra-subset variation in the class values of instances that go down each branch. The variability is measured by the standard deviation of the values that reach that node from the root through the branch with calculating the expected reduction in error as a result of testing each attribute at that node. The attribute that maximizes the expected error reduction is chosen. The splitting stops if the values of all instances that reach a node vary slightly, or only a few instances remain (Bhattacharya and Solomatine 2005; Goyal and Ojha 2011; Ajmera and Goyal 2012).

Decision Stump

Decision stumps are simple classifiers, in which the final decision is made by only a single hypothesis or feature. A decision stump is simply a one node decision tree based on a co-occurrence feature, whereas the majority classifier assigns the most frequent sense in the training data to every occurrence of that word in the test data. The decision stump algorithm builds simple binary decision *stumps* for both numeric and nominal classification problems (Witten et al. 1999).

REPTree

The REPTree algorithm is a fast decision tree learner that constructs a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with back-fitting) (Senthil kumar et al. 2012). The algorithm only sorts values for numeric attributes once (Daud and Corne 2007).

Weka (Witten and Frank 2000) implementation of the various learning algorithms is used here. Weka is written in Java and is freely available from http://www.cs.waikato.ac.nz/~ml.

J. Hydrol. Eng. 2014.19:828-835.

Table 1. Cross-Correlation Computed between Probable Predictors in NCEP and GCM Data Sets

Correlation	Ta925	Ua925	Va925	Va200	Ta20	Zg200	Ua200	Ta500	Zg500
P	0.83	0.79	0.67	−0.18	0.66	0.81	0.23	0.81	0.60
S	0.68	0.56	0.43	−0.14	0.46	0.64	0.57	0.64	0.39
K	0.87	0.76	0.61	−0.20	0.68	0.85	0.73	0.85	0.59

Note: K = Kendall's tau; P = product moment correlation; S = Spearman's rank correlation.

Table 2. Various Performance Statistics of Model Using Various Approaches

		CC		MAE		RMSE	
Approach	Model	Training	Validation	Training	Validation	Training	Validation
Single conjunctive rule learner	TmaxSCRL	0.69	0.73	2.78	2.60	3.33	3.09
Decision table	TmaxDT	0.97	0.73	0.66	2.46	1.04	3.46
M5 model tree	TmaxMT	0.96	0.96	0.69	1.03	1.24	1.36
Decision stump	TmaxDS	0.70	0.64	2.57	2.92	3.30	3.56
REPTree	TmaxREP	0.85	0.84	1.68	1.86	2.46	2.54
Single conjunctive rule learner	TminSCRL	0.89	0.83	2.72	3.42	3.32	4.34
Decision table	TminDT	0.95	0.91	1.84	2.39	2.34	3.25
M5 model tree	TminMT	0.98	0.95	1.07	1.61	1.43	2.41
Decision stump	TminDS	0.89	0.86	2.74	3.43	3.30	4.28
REPTree	TminREP	0.98	0.93	1.05	1.98	1.62	2.96
Single conjunctive rule learner	EvapSCRL	0.78	0.63	1.24	1.57	1.75	2.28
Decision table	EvapDT	0.98	0.72	0.30	1.30	0.45	1.95
M5 model tree	EvapMT	0.96	0.94	0.57	0.78	0.77	1.00
Decision stump	EvapDS	0.70	0.69	1.48	1.47	2.01	2.06
REPTree	EvapREP	0.56	0.69	2.31	2.07	2.78	2.82

Results and Discussions

The most relevant probable predictor variables necessary for developing the downscaling models are identified by using the three measures of dependence. The cross-correlations enable verifying the reliability of the simulations of the predictor variables by the GCM and are provided in Table 1. Seven predictor variables, namely air temperature at 925, 500, and 200 hPa; zonal wind (925 hPa); meridional wind (925 hPa); and geopotential height (500 and 200 hPa) at 9 NCEP grid points with a dimensionality of 63, are used because of better correlation as the standardized data of potential predictors. Principal Component Analysis (PCA) is performed to transform the set of correlated N-dimensional predictors ($N = 63$) into another set of N-dimensional uncorrelated vectors (called principal components) by linear combination, such that most of the information content of the original data set is stored in the first few dimensions of the new set. It is observed that the four leading principal components (PCs) of the PCA method explain about 97% of the information content (or variability) of the original predictors. PCs are extracted to form feature vectors from the standardized data of potential predictors. These feature vectors are provided as input to the various downscaling models. Results of the different models for predictands are tabulated in Table 2. It can be observed during calibration from Table 2 that the performance of models using M5P tree and decision table algorithms for mean monthly *Tmax* and *Tmin*, as well as pan evaporation, is clearly superior to that of REPTree, decision stump, and single conjunctive rule learner-based models in the training data set. For the *Tmax* and *Tmin* predictands, the results of the decision stump and single conjunctive rule learner -based models are quite similar. It can be inferred that models *Model3* and *Model1*, using decision table, performed the best for predictands *Tmax* and *pan evaporation*, respectively, while *Model8* using M5P performed the best for the predictand *Tmin*. It can also be inferred from validation results that the performance of M5P tree models (namely *TmaxMT*, *TminMT*, and *EvapMT*) for the predictands (*Tmax* and *Tmin* as well as *pan evaporation*) is clearly superior to that of decision table, REPTree, decision stump, and single conjunctive rule learner-based models in the validation data set.

It can also be observed from Table 2 that the performance of M5P models for all predictands is clearly superior to that of decision table, REPTree, decision stump, and single conjunctive rule learner-based models in the training and validation data set, barring few exceptions. Models using decision table performed better in the training data set for the predictands *Tmax* and *pan evaporation*. It can be inferred that models *TmaxMT*, and *TminMT* using algorithm M5 model tree performed best for predictands *Tmax* and *Tmin*, respectively, whereas model *EvapMT* using algorithm M5 model tree performed best for the predictand *pan evaporation*. A multiplicative shift (Ines and Hansen 2006) is used to correct the GCM bias of models *TmaxMT, TminMT*, and *EvapMT* corresponding to *Tmax* and *Tmin* as well as *pan evaporation*, respectively. All the corrected models performed better than uncorrected in terms of various performance measures, as given in Table 3, barring few exceptions. A comparison of mean monthly observed *Tmax, Tmin*, and *pan evaporation* with *Tmax, Tmin, and pan evaporation* simulated using M5 model tree algorithm models *TmaxMT* (corrected),

Table 3. Various Performance Statistics of Model Using Bias Correction

	CC		MAE		RMSE	
Model	Training	Validation	Training	Validation	Training	Validation
TmaxMT (corrected)	0.97	0.96	0.69	1.01	1.14	1.23
TminMT (corrected)	0.98	0.97	0.97	0.91	1.12	1.81
EvapMT (corrected)	0.96	0.95	0.52	0.68	0.57	1.05

TminMT (corrected) and *EvapMT* (corrected) is shown in Figs. 1–3 for the validation period. It can be inferred that these models are not able to mimic a few low observed events for the *Tmax* and *Tmin* predictands while a few maxobserved events for the *pan evaporation* predictand.

Once the downscaling models have been calibrated and validated, the next step is to use these models to downscale the scenario simulated by the GCM. The GCM simulations are run through the calibrated and validated models (viz *TmaxMT, TminMT*, and *EvapMT*) to obtain future simulations of the predictands. The predictands' (viz *Tmax* and *Tmin* as well as *pan evaporation*) patterns are analyzed with box plots for 20 year time slices. Typical results of downscaled predictands (*Tmax*, *Tmin*, and *pan evaporation*) obtained from the predictors are presented in Figs. 4–6. In part (a)

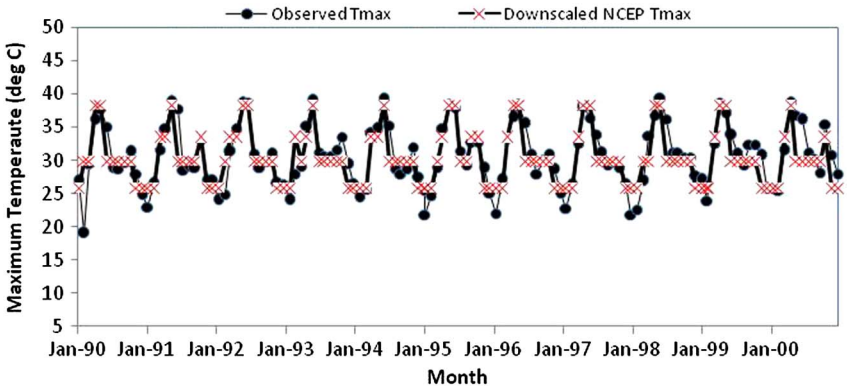

Fig. 1. Typical results for comparison of the monthly observed *Tmax* with *Tmax* simulated using M5 model tree downscaling model *TmaxMT* for NCEP data for validation period (1990 to 2000)

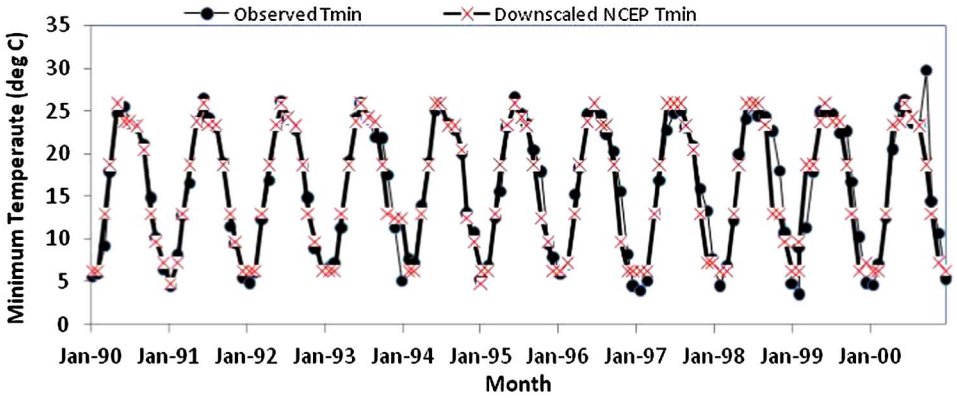

Fig. 2. Typical results for comparison of the monthly observed *Tmin* with *Tmin* simulated using M5 model tree downscaling model *TminMT* for NCEP data for validation period (1990 to 2000)

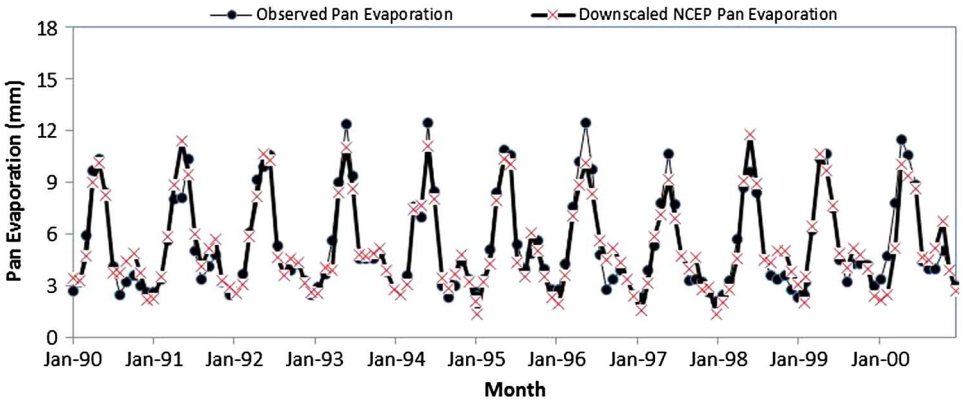

Fig. 3. Typical results for comparison of the monthly observed *pan evaporation* with *pan evaporation* simulated using M5 model tree downscaling model *EvapMT* for NCEP data for validation period (1990 to 2000)

J. Hydrol. Eng. 2014.19:828-835.

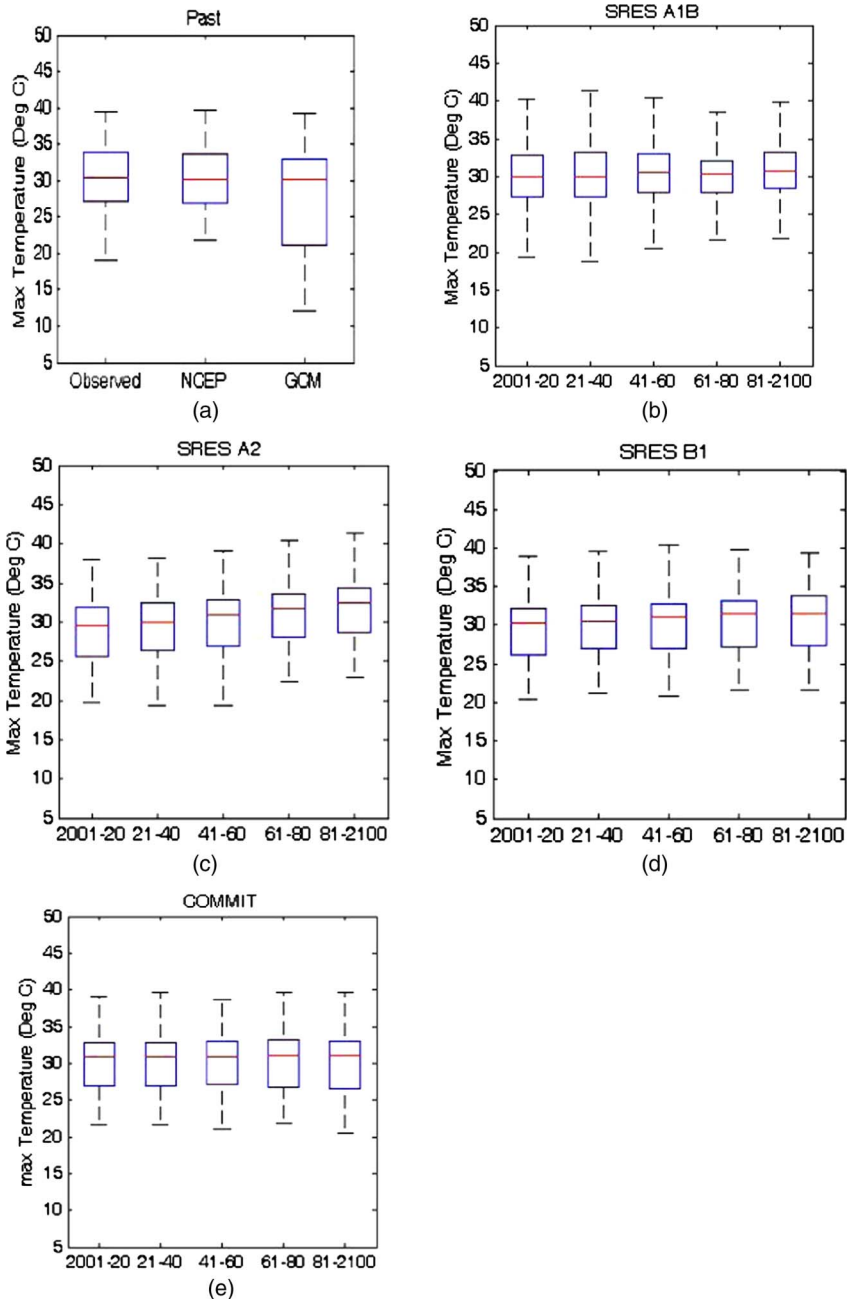

Fig. 4. Box plots results from the M5 model tree based downscaling model *TmaxMT* model for the predictand *Tmax*

of these figures, the *Tmax*, *Tmin, and pan evaporation* downscaled using NCEP and GCM data sets are compared with the observed *Tmax*, *Tmin*, and *pan evaporation* for the study region using box plots. The projected *Tmax*, *Tmin*, and *pan evaporation* for 2001–2020, 2021–2040, 2041–2060, 2061–2080, and 2081–2100, for the four scenarios A1B, A2, B1, and COMMIT are shown in (b), (c), (d) and (e), respectively.

From the box plots of downscaled predictands (Figs. 4 and 5), it can be observed that *Tmax* and *Tmin* are projected to increase in the future for the A1B, A2, and B1 scenarios, whereas no trend is discerned from the COMMIT scenario by using predictors. The projected increase in predictands is high for the A2 and A1B scenarios, whereas it is least for the B1 scenario. This may be because among the scenarios considered, the scenarios A1B and A2 have the highest concentration of atmospheric carbon dioxide (CO_2), equal to 720 and 850 ppm, and the same for the B1 and COMMIT scenarios, which have concentrations equal to 550 ppm and approximately 370 ppm, respectively. Increase in the concentration of CO_2 in the atmosphere causes the earth's average temperature to rise, which in turn causes an increase in evaporation, especially at lower latitudes (Anandi et al. 2009). In the COMMIT scenario, where the emissions are held the same as in the year 2000, no significant trend in the pattern of projected future precipitation could be discerned. From the box plot of pan evaporation (Fig. 6), it can be concluded that the change pattern of pan evaporation is not very clear in the future for A1B, B1, and COMMIT scenarios. This may be because a greater number of factors contribute to evaporation. However, an increasing trend has been observed for the A2 scenario. The appendix shows pruned model trees obtained by using the M5 model tree algorithm for the predictands.

J. Hydrol. Eng. 2014.19:828-835.

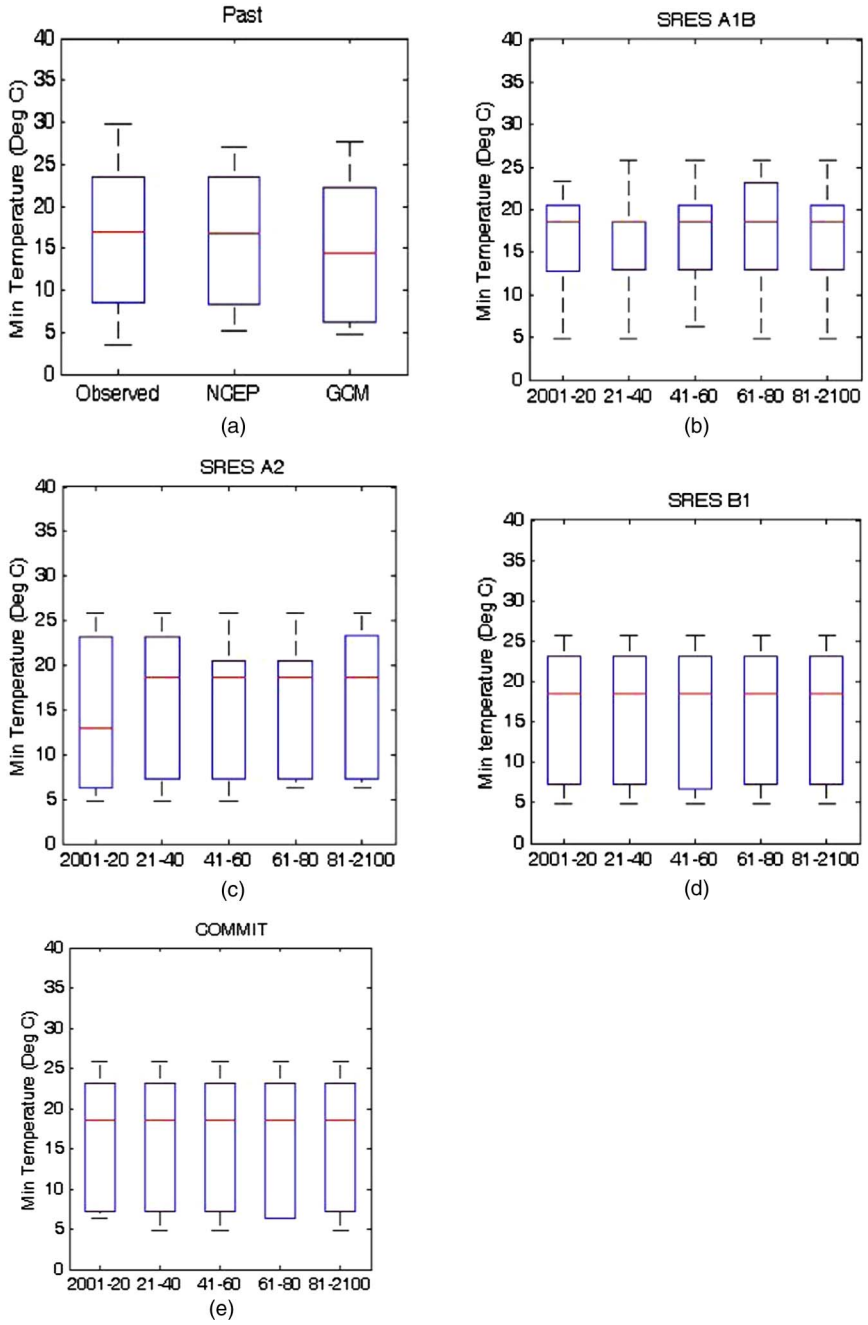

Fig. 5. Box plots results from the M5 model tree based downscaling model *TminMT* model for the predictand *Tmin*

Furthermore, a physical-based empirical method for computation of evaporation is employed to analyze the trend. For this purpose, the Hargreaves method (Hargreaves and Samani 1985) is a simple, empirical approach that has been used in cases in which the availability of weather data is limited. This approach requires only measurements of maximum and minimum temperatures, with extraterrestrial radiation calculation as a function of latitude and day of the year (Itenfisu et al. 2003). The Mann–Kendall nonparametric trend test was performed on the computed evaporation values using the Hargreaves method for all the scenarios using downscaled *Tmax* and *Tmin* with extraterrestrial radiation. A value of 0.05 was chosen as the local significance level. It is observed that there is no significant trend, either positive or negative, for the predictand *pan evaporation*. Hence, this physical-based model produced similar results to those of this study.

Comparison with Previous Downscaling Studies

Anandhi et al. (2009) developed downscaling models using a support vector machine (SVM) for obtaining projections of monthly mean maximum and minimum temperatures (*Tmax* and *Tmin*) for the catchment of the Malaprabha reservoir in India. The resulting models produced similar results to those of this study. For example, the results of downscaling show that *Tmax* and *Tmin* are projected to increase in the future for the A1B, A2, and B1 scenarios, whereas no trend is discerned from the COMMIT scenario, and *Tmax* was better simulated than *Tmin* between the two predictands. A comparison for pan evaporation has been made to a similar study carried out in the semi arid Haihe River basin in China. Chu et al. (2010) developed the downscaling models for pan evaporation using statistical

J. Hydrol. Eng. 2014.19:828-835.

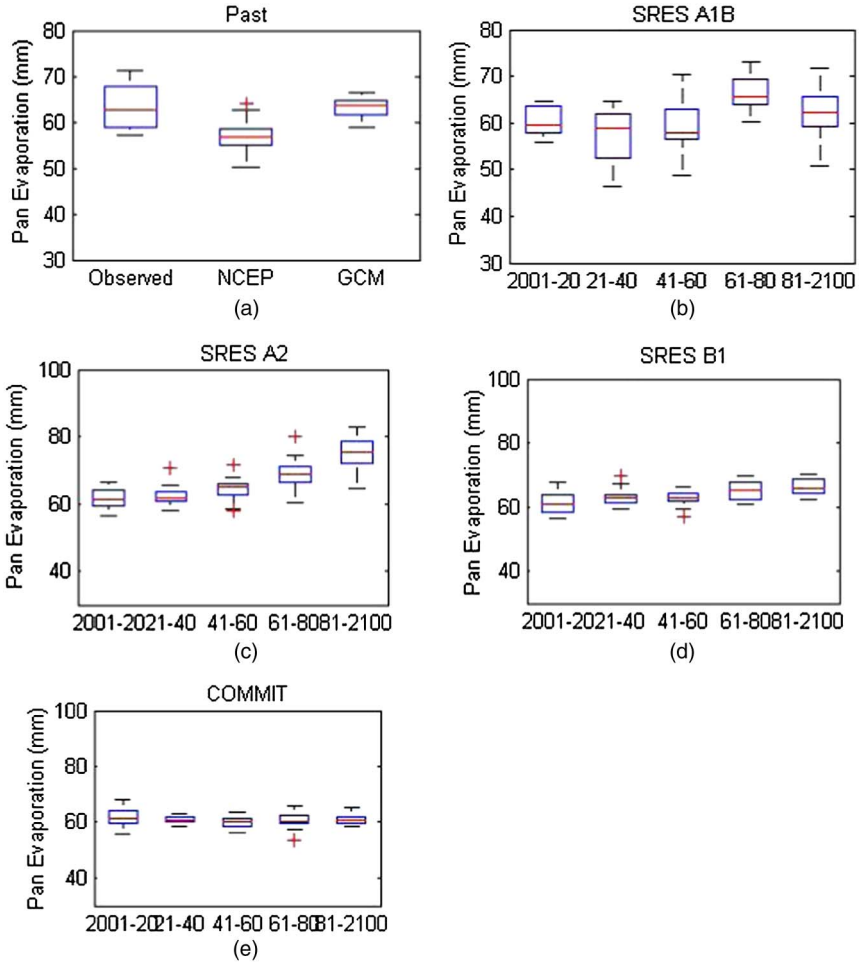

Fig. 6. Box plots results from the *EvapMT* model based downscaling model for the predictand *pan evaporation*

downscaling method, and results produced are similar to those of this study.

Conclusions

This paper explores the suitability of various available rule and decision tree learning algorithms, namely the single conjunctive rule learner, decision table, M5 model tree, decision stump, and REPTree approaches to downscale mean monthly maximum temperature (*Tmax*), minimum temperature (*Tmin*), and *pan evaporation* from GCM output to local scale. The M5P model tree performed best for the predictands *Tmax, Tmin*, and *pan evaporation*. The M5P model tree was followed by REPTree in the case of *Tmax* and *Tmin* but followed by decision table in the case of *pan evaporation*. The single conjunctive rule learner performed worst in the case of *Tmin* and pan evaporation, whereas the decision stump algorithm performed worst in the case of *Tmax*. GCM bias correction procedure improved the overall predictability of predictands. The results of downscaling models using the M5P model tree algorithm show that *Tmax* and *Tmin* are projected to increase in the future for A1B, A2, and B1 scenarios, whereas no trend is discerned from the COMMIT scenario using predictors. It is interesting to note that the change in the pattern of *pan evaporation* is not obvious in the future for the A1B, B1, and COMMIT scenarios but is apparent in the A2 scenario, which corresponds to an increasing trend.

However, data used from a number of GCMs could produce different results in a given location. Therefore, caution should be exercised in interpreting the outcome of such impact analysis for practical applications since downscaling results in this research using the outputs of a single GCM.

Appendix. Pruned Model Tree

Pruned model trees obtained using the M5 modeling approach are as follows:

1. Pruned model tree obtained by using M5 modeling approach for maximum temperature
   - M5 pruned model tree: (using smoothed linear models).
   - LM1 (72/26.882%).
   - LM num: 1.
     $$\text{MaxTemp} = 0.7908 \times \text{pc1} - 1.0665 \times \text{pc2} - 1.0403 \times \text{pc3} - 0.4 \times \text{pc4} + 30.3012$$
   - Number of Rules: 1.
2. Pruned model tree obtained by using M5 modeling approach for minimum temperature
   - M5 pruned model tree: (using smoothed linear models).
     pc1 <= −0.623: LM1 (36/21.891%)
     pc1 > −0.623:
     |pc1 <= 3.497: LM2 (17/14.41%)
     |pc1 > 3.497: LM3 (19/14.916%)
   - LM num: 1.

J. Hydrol. Eng. 2014.19:828-835.

MinTemp = 1.1205 × pc1 − 0.6049 × pc2 − 1.2277 × pc3 − 0.432 × pc4+13.3791

- LM num: 2.
  MinTemp = 0.8571 × pc1 + 0.4189 × pc2 − 0.3184 × pc3 + 0.2432 × pc4 + 18.3162
- LM num: 3.
  MinTemp = 1.0488 × pc1 + 0.1258 × pc2 − 0.3184 × pc3 + 0.2289 × pc4 + 18.8965
- Number of Rules: 3.

3. Pruned model tree obtained by using M5 modeling approach for pan evaporation
   - M5 pruned model tree: (using smoothed linear models).
     pc1 <= −2.735: LM1 (26/13.378%)
     pc1 > −2.735: LM2 (46/32.736%)
   - LM num: 1.
     Pan_Evaporation = 0.2927 × pc1 − 0.5117 × pc2 − 0.2963 × pc3 + 0.1108 × pc4 + 4.6447
   - LM num: 2.
     Pan_Evaporation = 0.5937 × pc1 − 0.7731 × pc2 − 0.5083 × pc3 + 4.7801
   - Number of Rules: 2.

Acknowledgments

The authors are grateful to four unknown reviewers in addition to associate editor and section editor for their insightful comments and suggestions that improved the paper.

References

Ajmera, T. K., and Goyal, M. K. (2012). "Development of stage discharge rating curve using model tree and neural networks: An application to Peachtree Creek in Atlanta." *Expert Syst. Appl.*, 39(5), 5702–5710.

Anandhi, A., Srinivas, V. V., Kumar, D. N., and Nanjundiah, R. S. (2009). "Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine." *Int. J. Clim.*, 29(4), 583–603.

Benestad, R. E. (2001). "A comparison between two empirical downscaling strategies." *Int. J. Clim.*, 29, 1645–1668.

Bhattacharya, B., and Solomatine, D. P. (2005). "Neural networks and M5 model trees in modelling water level–discharge relationship." *Neurocomputing*, 63, 381–396.

Chu, J. T., Xia, J., Xu, C.-Y., and Singh, V. P. (2010). "Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China." *Theor. Appl. Climatol.*, 99(1–2), 149–161.

Clark, P., and Niblett, T. (1989). "The CN2 rule induction algorithm." *Mach. Learn.*, 3(4), 261–284.

Cohen, W. (1995). "Fast effective rule induction." *Proc., 12th Int. Conf. on Machine Learning*, Morgan Kaufmann, Tahoe City, CA, 115–123.

Conway, D., Wilby, R. L., and Jones, P. D. (1996). "Precipitation and air flow indices over the British Isles." *Clim. Res.*, 7(2), 169–183.

Daud, M. N. R., and Corne, D. W. (2007). "Human readable rule induction in medical data mining: A survey of existing algorithms." *WSEAS European Computing Conf.*, Athens, Greece.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling." *Int. J. Clim.*, 27(12), 1547–1578.

Fowler, H. J., Kilsby, C. G., and O'Connell, P. E. (2000). "A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions." *Hydrol. Earth Syst. Sci.*, 4, 261–280.

Goyal, M. K., and Ojha, C. S. P. (2011). "Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree." *Water Resour. Manage.*, 25(9), 2177–2195.

Goyal, M. K., and Ojha, C. S. P. (2012). "Downscaling of surface temperature for Lake Catchment in arid region in India using linear multiple regression and neural networks." *Int. J. Clim.*, 32(4), 552–566.

Hargreaves, G. H., and Samani, Z. A. (1985). "Reference crop evapotranspiration from Temperature." *Trans. ASAE*, 1(2), 96–99.

Ines, V. M. A., and Hansen, J. W. (2006). "Bias correction of daily GCM rainfall for crop simulation studies." *Agric. Forest Meteorol.*, 138, 44–53.

Itenfisu, D., Elliott, R. L., Allen, R. G., and Walter, I. A. (2003). "Comparison of reference evapotranspiration calculations as part of the ASCE standardization effort." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)0733-9437(2003)129:6(440), 440–444.

Kalnay, E., et al. (1996). "The NCEP/NCAR 40-year reanalysis project." *Bull. Am. Meteorol. Soc.*, 77(3), 437–471.

Kang, H. W., An, K. H., Park, C. K., Solis, A. L. S., and Stitthichivapak, K. (2007). "Multimodel output statistical downscaling prediction of precipitation in the Philippines and Thailand." *Geophys. Res. Lett.*, 34(15), L15710.

Kettle, H., and Thompson, R. (2004). "Statistical downscaling in European mountains: Verification of reconstructed air temperature." *Clim. Res.*, 26(2), 97–112.

Khobragade, S. D. (2009). "Studies on evaporation from open water surfaces in tropical climate." Ph.D. thesis, Indian Institute of Technology, Roorkee, India.

Kite, G. W. (1997). "Simulating Columbia river flows with data from regional-scale climate models." *Water Resour. Res.*, 33(6), 1275–1285.

Kohavi, R. (1995). "The power of decision tables." *Proc., European Conf. on Machine Learning, Lecture Notes in Artificial Intelligence*, Vol. 914, Springer, Berlin, 174–189.

Mason, S. J. (2004). "Simulating climate over western North America using stochastic weather generators." *Clim. Change*, 62(1–3), 155–187.

Mohd, F. b. O., and Thomas, M. S. Y. (2007). "Comparison of different classification techniques using WEKA for breast cancer." *Biomed 06: IFMBE Proc.*, Vol. 15, 520–523.

Ojha, C. S. P., Goyal, M. K., and Adeloye, A. J. (2010). "Downscaling of precipitation for Lake Catchment in arid region in India using linear multiple regression and neural networks." *Open Hydrol. J.*, 4, 122–136.

Quinlan, J. R. (1992). "Learning with continuous classes." *Proc., 5th Australian Joint Conf. on Artificial Intelligence*, World Scientific, Singapore, 343–348.

Senthil kumar, A. R., Ojha, C. S. P., Goyal, M. K., Singh, R. D., and Swamee, P. K. (2012). "Modelling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic and decision tree algorithms." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000445, 394–404.

Wang, Y. Q., Leung, L. R., Mcgregor, J. L., Wang, W. C., Ding, Y. H., and Kimura, F. (2004). "Regional climate modeling: Progress, challenges, and prospects." *J. Meteorol. Soc. Jpn.*, 82(6), 1599–1628.

Wilby, R. L., Dawson, C. W., and Barrow, E. M. (2002). "SDSM—A decision support tool for the assessment of climate change impacts." *Environ. Model. Software*, 17(2), 147–159.

Wilby, R. L., and Harris, I. (2006). "A framework for assessing uncertainties in climate change impacts: Low-flow scenarios." *Water Resour. Res.*, 42(2), W02419.

Wilby, R. L., and Wigley, T. M. L. (1997). "Downscaling general circulation model output: A review of methods and limitations." *Prog. Phys. Geog.*, 21(4), 530–548.

Witten, I. H., and Frank, E. (2000). *Data mining: Practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann, San Francisco, CA.

Witten, I. H., Frank, E., Trigg, L., Hall, M., Holmes, G., and Cunningham, S. J. (1999). "Weka: Practical machine learning tools and techniques with Java implementations." *Emerging Knowledge Engineering and Connectionist-Based Info. Systems*, 192–196.

Xu, C.-Y. (1999). "From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches." *Prog. Phys. Geog.*, 23(2), 229–249.