

# Uncertainties in downscaled relative humidity for a semi-arid region in India

AAVUDAI ANANDHI

*CUNY Institute for Sustainable Cities, City University of New York, New York, NY 10065, USA.  
e-mail: aswamy@hunter.cuny.edu*

Monthly scenarios of relative humidity ( $R_H$ ) were obtained for the Malaprabha river basin in India using a statistical downscaling technique. Large-scale atmospheric variables (air temperature and specific humidity at 925 mb, surface air temperature and latent heat flux) were chosen as predictors. The predictor variables are extracted from the (1) National Centers for Environmental Prediction reanalysis dataset for the period 1978–2000, and (2) simulations of the third generation Canadian Coupled Global Climate Model for the period 1978–2100. The objective of this study was to investigate the uncertainties in regional scenarios developed for  $R_H$  due to the choice of emission scenarios (A1B, A2, B1 and COMMIT) and the predictors selected. Multi-linear regression with stepwise screening is the downscaling technique used in this study. To study the uncertainty in the regional scenarios of  $R_H$ , due to the selected predictors, eight sets of predictors were chosen and a downscaling model was developed for each set. Performance of the downscaling models in the baseline period (1978–2000) was studied using three measures (1) Nash–Sutcliffe error estimate ( $E_f$ ), (2) mean absolute error (MAE), and (3) product moment correlation ( $P$ ). Results show that the performances vary between 0.59 and 0.68, 0.42 and 0.50 and 0.77 and 0.82 for  $E_f$ , MAE and  $P$ . Cumulative distribution functions were prepared from the regional scenarios of  $R_H$  developed for combinations of predictors and emission scenarios. Results show a variation of 1 to 6%  $R_H$  in the scenarios developed for combination of predictor sets for baseline period. For a future period (2001–2100), a variation of 6 to 15%  $R_H$  was observed for the combination of emission scenarios and predictors. The variation was highest for A2 scenario and least for COMMIT and B1 scenario.

---

## 1. Introduction

Researchers and decision-makers who assess the impact of climate change in river basin development, agriculture, hydrology, irrigation management, etc., require future scenarios of relative humidity ( $R_H$ ). In a river basin,  $R_H$  is one of the variables used in the (1) estimation of evapotranspiration (ET) by Penman–Monteith method which is recommended as a standard method by the FAO (Allen *et al* 1998). Precipitation and ET

play key roles in the development of irrigation management programmes which require knowledge of when to irrigate and the amount of water to apply (Castellví *et al* 1996, 1997; Perez *et al* 2008); (2) estimation of evaporation from a variety of surfaces, such as lakes, rivers, soils and wet vegetation (Allen *et al* 1998); since evaporation depends on the energy supply, vapour pressure gradient (in terms of humidity) and wind; (3) assessment of crop growth (Castellví *et al* 1996) and stomatal conductance (Wang *et al* 2009). Changes in

**Keywords.** Climate change; downscaling; hydroclimatology; relative humidity; multi-step linear regression; Intergovernmental Panel on Climate Change Special Report on Emission Scenarios (IPCC SRES); Malaprabha river basin.

humidity are important as they impact the hydrological cycle and surface energy budget (Vincent *et al* 2007).  $R_H$  is identified as one of the six cardinal variables used for quantitative climate impact studies based on variables (IPCC 2001) and regional shifts in its climatological distribution can have the potential to influence climate on a wide range of scales (Wright *et al* 2010). The semi-arid regions which contribute about 17.7% of the total land surface area are highly sensitive to climate change and its variability (Ragab and Prudhomme 2002; Legesse *et al* 2003; Abdulla *et al* 2009) and so are affected by changes in humidity. Hence, there is need to develop future climate scenarios of  $R_H$  in a river basin.

The future is uncertain and impossible to predict (Schenk and Lensink 2007). A scenario provides a plausible description of how the future may develop. To assess the uncertainty and knowledge gaps associated with future, a spectrum of alternative futures (scenarios) are used. The uncertainty associated with future greenhouse gas emissions has been assessed in the Special Report on Emission Scenarios (SRES; Nakicenovic *et al* 2000) of the Intergovernmental Panel on Climate Change (IPCC) using the four SRES ‘families’. Global climate models (GCMs) are used to construct scenarios of future climate change under a specified scenario of emissions. The output scenarios from the GCMs cannot be directly used at a regional scale due to the mismatch of spatial scales between GCMs on the one hand, and local observations and local impact assessments on the other hand.

In the past, future climate scenarios at a regional scale have been derived (i) based on analogies with different climatic zones or historical time periods or (ii) from GCMs using simple manipulation of current climate observations (e.g., change factor methodology, or CFM) and (iii) more sophisticated statistical and dynamical downscaling methodologies (Wilby *et al* 2000; Mearns *et al* 2001; Anandhi *et al* 2011). The statistical downscaling technique derives empirical relationships that transform large-scale features of the GCM (LF) to regional-scale variables or site-specific variables (RSV).

$$RSV = g(LF). \quad (1)$$

There are three types of statistical downscaling namely, weather classification methods, weather generators and transfer functions. The review of downscaling techniques can be obtained from Fowler *et al* (2007) and Anandhi (2010). Future scenarios are derived in this study using transfer functions as they are conceptually simple and the most common statistical downscaling approach (Schoof *et al* 2007). Transfer functions capture the linear or nonlinear relationships between

the large-scale atmospheric variables (predictors) and the local meteorological variable of interest (predictand).

A vast majority of the statistical downscaling studies concentrate on temperature and precipitation and only a few studies have dealt with other atmospheric quantities such as humidity (Huth 2005). In the past, only a handful of studies have statistically downscaled humidity using transfer functions (Enke and Spekat 1997; Huth 2005).

The objective of this study is to develop regional scenarios of  $R_H$  for a semi-arid region and investigate the uncertainties in the developed scenarios for the selected predictors and emission scenarios. The chosen methodology is tested for the Malaprabha river basin in Karnataka, India. Multilinear regression with stepwise screening (MSR) is the downscaling technique used in this study. A1B, A2, B1 and COMMIT are the emission scenarios considered.

The rest of this paper is structured as follows: Section 2 provides the description of the study region and details of the data used. Section 3 describes the methodology developed in the study to arrive at scenarios of  $R_H$  at river basin scale for different combinations of predictor and emission scenarios. Section 4 presents the results and discussion. Finally, section 5 provides summary of the work presented and conclusions drawn from the study.

## 2. Study region and data

The catchment of Malaprabha reservoir in Karnataka was considered for the study. It has an area of 2564 km<sup>2</sup> situated between 15°30'N and 15°56'N latitudes and 74°12'E and 75°15'E longitudes. The Malaprabha reservoir is a major source of water for the semi-arid region of north Karnataka. The reservoir supplies water for an irrigable area of 218191 hectares in north Karnataka. The location map of the study region is shown in figure 1.

The observed records of  $R_H$  (predictand) were available on a monthly timescale from a gauging station near the region. It is situated in Gadag district (administrative division) at 15°25'N latitude and 75°38'E longitude, and has records for the period January 1978 to December 2000. Primary source of the data is the Water Resources Development Organization, Government of Karnataka, India. The  $R_H$  is high during the monsoons (June to September) and low during the post-monsoon period. In summer, the weather is dry and humidity low. The mean monthly daytime  $R_H$  in the basin during the monsoons ranges from 83 to 87%, while the same during post-monsoon (October to January) and summer (February to May) is 67 to 78% and 57 to 74%, respectively.

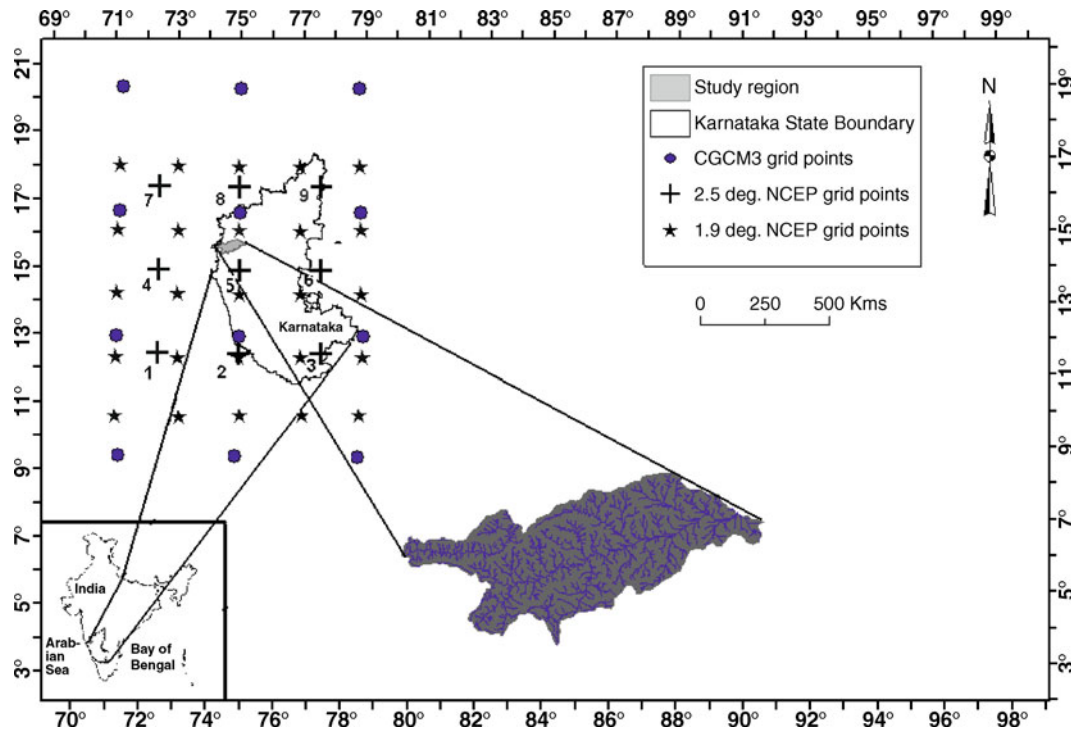


Figure 1. Location of the study region in Karnataka, India. The latitude, longitude and scale of the map refer to Karnataka. The global climate model simulations (CGCM) are extracted at 12 grid points (represented as circles). The latent heat flux data is obtained from reanalysis datasets at 25 grid points spaced at  $1.9^\circ$  (represented as stars) are re-gridded to the nine  $2.5^\circ$  NCEP grid points (represented as +). Among the nine grid points 1, 4 and 7 are in the Arabian Sea, and the remaining points are on land.

The monthly mean atmospheric variables and latent heat (predictors) were extracted from the database prepared by National Centres for Environmental Prediction (NCEP; Kalnay *et al* 1996; Kistler *et al* 2001) and simulations of the third generation Canadian coupled global climate model (CGCM).

The NCEP reanalysis data were extracted for the period January 1978 to December 2000 and available at two different spatial scales. The data of atmospheric variables were collated for nine grid points whose latitude ranges from  $12.5^\circ\text{N}$  to  $17.5^\circ\text{N}$ , and longitude ranges from  $72.5^\circ\text{E}$  to  $77.5^\circ\text{E}$  at a spatial resolution of  $2.5^\circ$ . The data of latent heat flux was available at a different spatial resolution of approximately  $1.9^\circ$  and extracted for 25 grid points whose latitude ranges from  $10.48^\circ\text{N}$  to  $18.1^\circ\text{N}$  and longitude ranges from  $71.25^\circ\text{E}$  to  $78.75^\circ\text{E}$ .

For obtaining future projections of  $R_H$ , monthly simulations of CGCM were considered for four emission scenarios (A1B, A2, B1 and COMMIT) that are relevant to the IPCC's fourth assessment report (AR4) (Alley *et al* 2007). CGCM simulations for the period 1978–2000 were downloaded for the 20C3M scenario. The CGCM data were obtained (through website <http://www.ccma.bc.ec.gc.ca/>) for the period January 1978 to December

2100, for 12 grid points whose latitude ranges from  $9.28^\circ\text{N}$  to  $20.41^\circ\text{N}$  and longitude from  $71.25^\circ\text{E}$  to  $78.75^\circ\text{E}$ . The spatial resolution of CGCM output is  $3.75^\circ$  along the longitude and approximately  $3.75^\circ$  along the latitude.

The GCM data and the information on atmospheric flux were re-gridded to a common  $2.5^\circ$  grid using grid analysis and display system (GrADS; Doty and Kinter 1993).

### 3. Methodology

This section briefly presents the various steps followed in the study (figure 2). The theory behind some of these steps is discussed in detail in the next section.

1.  $R_H$  is the predictand and its observed data collected.
2. Predictors are selected such that the variables are as follows:
  - readily available from archives of GCM output and reanalysis datasets
  - physically related with the predictand
  - based on previous studies

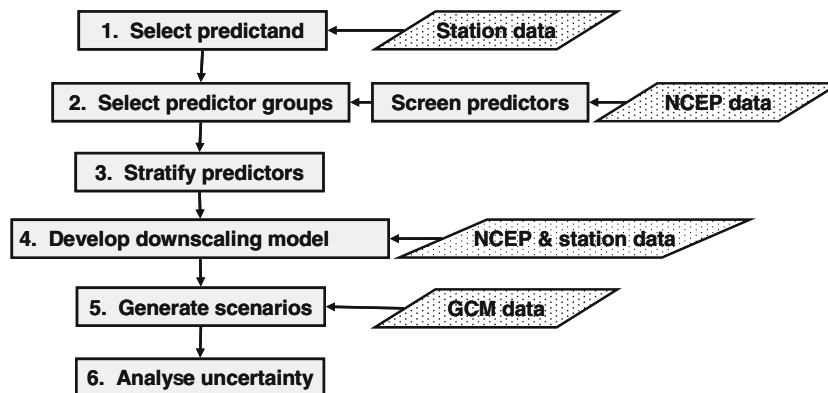


Figure 2. Brief methodology followed in this study.

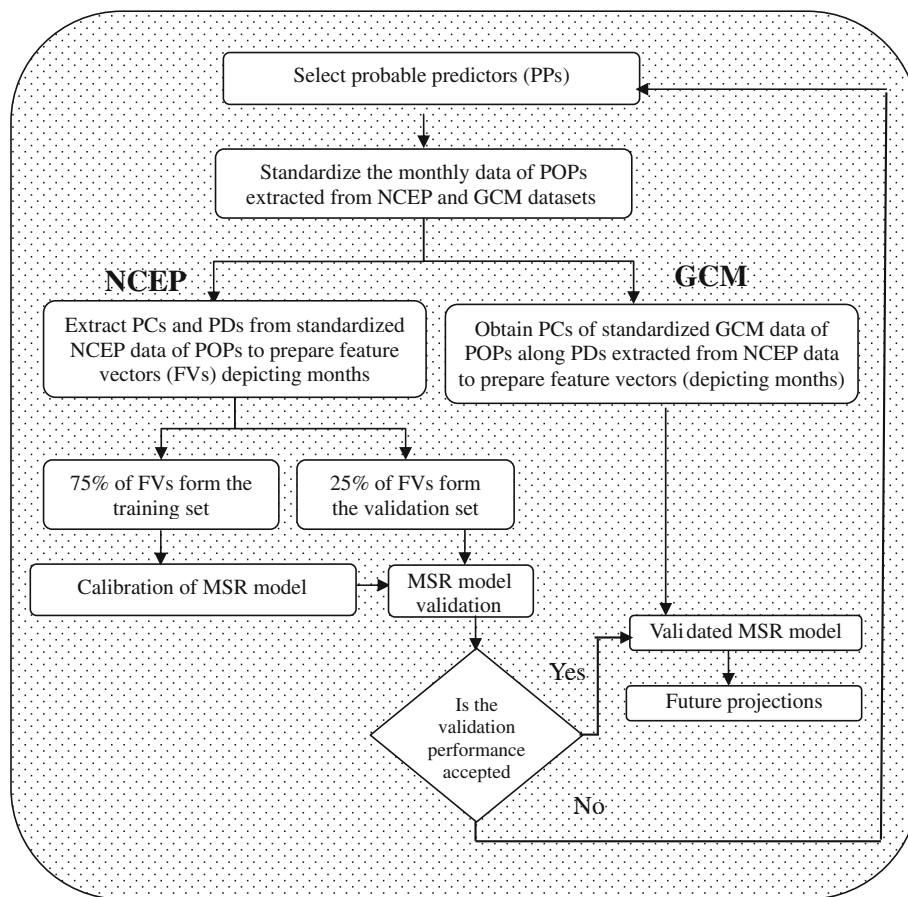


Figure 3. Methodology followed for MSR downscaling. In the figure, PCs and PDs denote principal components and principal directions, respectively. PPs and POPs denote probable predictors and potential predictors, respectively.

Predictor variables are then screened using scatter plots and cross-correlations computed between (a) the predictor variables in NCEP and GCM data (b) the NCEP predictor variables and the predictand.

3. Predictors are stratified based on location. Among the nine grid points chosen for the study,

six grid points lie on land and three grid points lie in the sea (figure 1). The separation of predictors into two groups (land and sea) is referred to as stratification based on location in this study.

4. Downscaling model is developed with MSR using MatLab (<http://www.mathworks.com>). Steps 4 and 5 are shown in figure 3. The predictors are standardized and principal components (PCs)

which preserve more than 98% of the variance in the data were extracted, and corresponding principal directions are noted. Feature vector for each month was formed using the PCs extracted for the month. The feature vectors are inputs to the model and downscaled predictand is the output. The feature vectors formed are partitioned into a training set (calibration) and a testing set (validation) for the periods 1978–1993 and 1994–2000, respectively.

A stepwise screening procedure is applied in model development. This is a systematic method for adding and removing terms from a multilinear model based on their statistical significance 10% in a regression. The method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models. At each step, the  $p$ -value of an  $F$ -statistic is computed to test models with and without a potential term. If a term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the term is added to the model. Conversely, if a term is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the term is removed from the model. The method terminates when no single step improves the model.

5. Generate future scenarios for the period 2001 to 2100. To form feature vectors for GCM data, the PCs are extracted along principal directions obtained for the NCEP data. These feature vectors are processed through the validated downscaling model to obtain future projections of the predictand, for each of the four emission scenarios considered (i.e., SRES A1B, A2, B1 and COMMIT).
6. Analyse uncertainty
  - (a) Uncertainty in the downscaling model due to predictors is estimated for the period 1978–2000, and
  - (b) Uncertainty in the future scenarios due to predictors and emission scenarios are considered for the period 2001–2100.

While analysing the first part (a), the uncertainty due to predictors is selected as the source of uncertainty and then it is quantified by propagating it to the model outputs. In this study, the uncertainty due to predictors are quantified in terms of model performance using goodness-of-fit measures such as mean absolute error, Nash–Sutcliffe coefficient of efficiency and correlation coefficient explained in equations (2–4).

- (1) Nash–Sutcliffe coefficient of efficiency ( $E_f$ , Nash and Sutcliffe 1970) is defined as:

$$E_f = 1 - \left[ \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \right]. \quad (2)$$

- (2) Mean absolute error (MAE, Johnson *et al* 2003) is given as:

$$\text{MAE} = 1 - \left[ \frac{\sum_{i=1}^N |(y_i - \hat{y}_i)|}{\sum_{i=1}^N |(y_i - \bar{y}_i)|} \right]. \quad (3)$$

- (3) Product moment correlation ( $P$ , Pearson 1896) is defined as:

$$P = \frac{\sum_{i=1}^N (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{N\sigma_{y_i}\hat{\sigma}_{y_i}} \quad (4)$$

where  $N$  represents the number of feature vectors prepared from the NCEP record,  $y_i$  and  $\hat{y}_i$  denote the observed and the downscaled (or generated) values of predictand, respectively,  $\bar{y}$  and  $\sigma_{y_i}$  are the mean and the variance of the observed predictand, respectively.  $\bar{\hat{y}}_i$  and  $\hat{\sigma}_{y_i}$  are the mean and the variance of the downscaled predictand, respectively.

Earlier studies have used the goodness-of-fit measures such as mean absolute error, Nash–Sutcliffe coefficient of efficiency, root mean square error, index of agreement to quantify uncertainty (Harmel and Smith 2007; Harmel *et al* 2009; Krueger *et al* 2009, 2010). In this study,  $R_H$  observations are central to locating and quantifying uncertainty. For a particular predictor set, the value of the different measure represents the deviation of downscaled  $R_H$  from the observed  $R_H$ . In this study, when the observation and the downscaling method remain unchanged, using different predictor sets could result in different scenarios of downscaled  $R_H$  which can be attributed to uncertainty in predictor sets. The range of the variation in the measure is used to represent the typical range of uncertainty that can be expected by using different combinations of predictors. Thus instead of using a single-valued prediction with unique explanations for a particular predictor set, this study explicitly gives the uncertainty (range for each measure). Here the uncertainty arising due to errors in predictor data, observations and downscaling models are not included.

Uncertainty across the different percentiles of downscaled  $R_H$  is calculated using empirical cumulative distribution functions (CDF). The

downscaled  $R_H$  is arranged in an increasing order and the CDF is defined using equation (5).

$$\hat{F}_n(x) = \sum_{i=1}^n \prod (X_i \leq x) / n, \quad (5)$$

where

$$\prod (X_i \leq x) = \begin{cases} 1 & \text{if } X_i \leq x \\ 0 & \text{if } X_i > x. \end{cases}$$

CDF is the proportion of observations less than or equal to  $x$ , and the function ( $\hat{F}_n(x)$ ) that assigns probability  $1/n$  to each of  $x_i$  ( $R_H$  value). For the baseline period  $n=276$ , [12×23 years (1978–2000)], while for future period  $n=1200$ , [12×100 years (2001–2100)].

Eight sets of predictor are chosen to study the uncertainty in predictors. They are, the set (1) consists of all the four predictors (Hus 925, Ta 925, Ta sur, and LH) for land grid points, set (2) comprises Ta 925 and Ta sur for land grid points, set (3) has three predictors (Hus 925, Ta 925 and Ta sur) for land grid points, whereas set (4) has only Hus 925 for land grid points. Sets (5–8) have the same four groups of predictors, but there is no stratification and so all the nine grids were considered.

There would be eight CDFs for downscaled  $R_H$  using NCEP predictors and eight more CDFs

using CGCM predictors for the baseline period. Similarly for future period, each of the four emission scenarios would have eight CDFs. From the CDFs, the maximum and minimum value of  $R_H$  [ $\max(R_{H,p})$ ,  $\min(R_{H,p})$ ] are estimated for each of the percentiles ( $p$ ;  $y$ -axis of CDFs×100). Uncertainty range for a percentile ( $p$ ) is the difference between the maximum and minimum value of  $R_H$  (equation 6) and  $p$  varies between 0 and 100.

$$UR_P = \max(R_{H,p}) - \min(R_{H,p}). \quad (6)$$

There would be two series of uncertainty ranges for the baseline period (one each for NCEP and CGCM predictors). Similarly, there would be four series of uncertainty ranges, one each for the four emission scenarios. The uncertainty ranges for the percentiles is plotted and its statistics such as mean, median, standard deviation, maximum and minimum for each series are calculated.

## 4. Results and discussion

### 4.1 Predictor selection and stratification

Large-scale atmospheric variables (LSAV) namely air temperature and specific humidity at 925 mb,

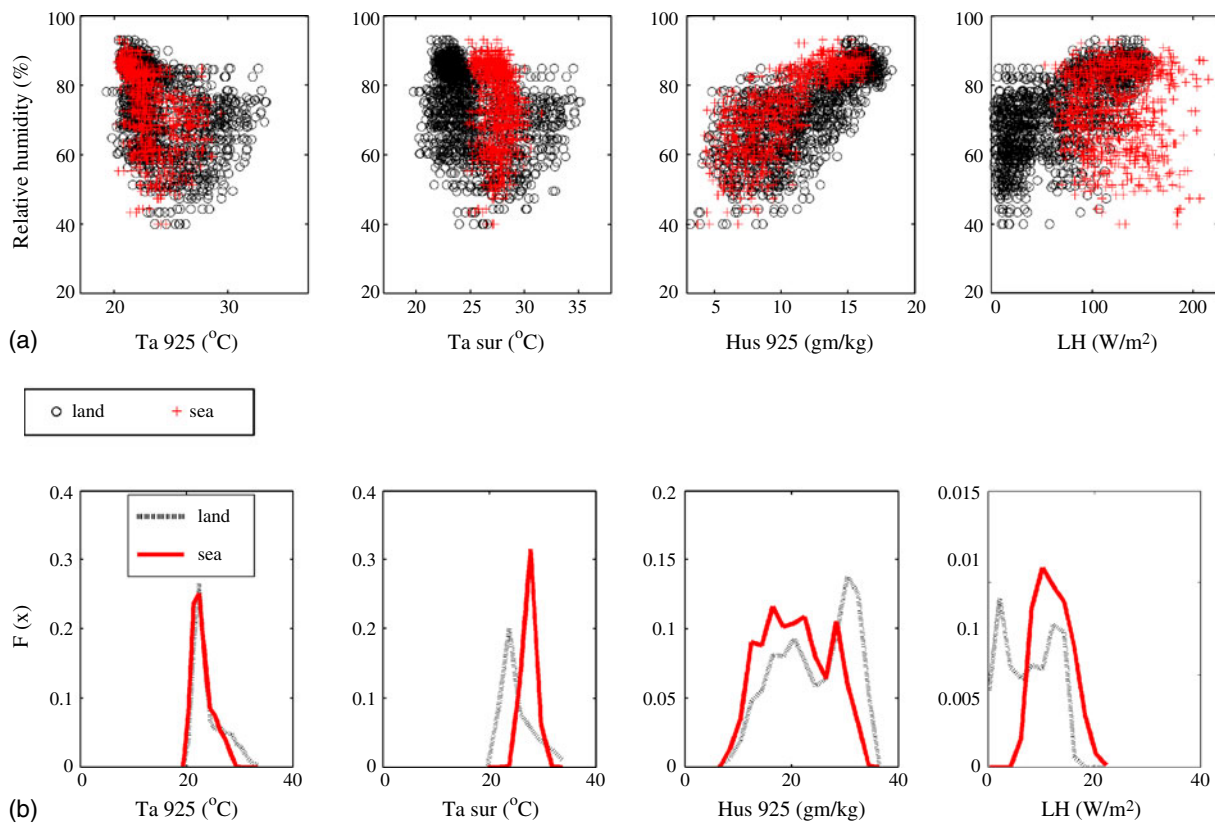


Figure 4. (a) Scatter plots prepared between probable predictor variables in NCEP data ( $x$ -axis) and predictand ( $y$ -axis). (b) Probability density function (PDF) of predictors whose grid points lie on land and sea. Grid points 1, 4 and 7 from figure 1 lie in sea while the rest of the points lie on land.

surface air temperature and latent heat flux were considered as the probable predictors. They are denoted by  $T_a$  925,  $H_{us}$  925,  $T_a$  sur, and LH, respectively. Temperature and humidity are associated with local thermodynamic stability and hence are useful as predictors. Temperature affects the moisture-holding capacity and the pressure at a location. The pressure gradient affects the circulation, which in turn affects the moisture brought into the place and hence the humidity. At 925 mb pressure height, the boundary layer (near surface effect) is important. LH indicates the amount of moisture going from the surface to atmosphere. The amount of moisture held in the atmosphere is related to temperature through the ‘Clausius–Clapeyron’ equation. The Clausius–Clapeyron relation describes the exponential increase of the water-holding capacity of the atmosphere with increasing temperature ( $T$ ) such that

for a 1K increase in  $T$ , surface saturated specific humidity or saturated vapour pressure should increase approximately by 7%, increasing with latitude and altitude (Willett *et al* 2008).

The scatter plots and cross-correlations computed between the NCEP predictor variables and the predictand are shown in figure 4(a). From the perusal of the figure and the cross-correlations computed it was observed that  $H_{us}$  925 has the highest correlation with predictand (0.75), while the rest of the predictors have comparatively lesser correlations. The temperature predictors are negatively correlated ( $-0.41$  to  $-0.31$ ) with the predictand while  $H_{us}$  925 and LH had a positive correlation with the predictand.

Stratification was carried out because the factors governing changes in  $R_H$  over land and sea are different and it would be interesting to study this effect on downscaling. Large-scale atmospheric

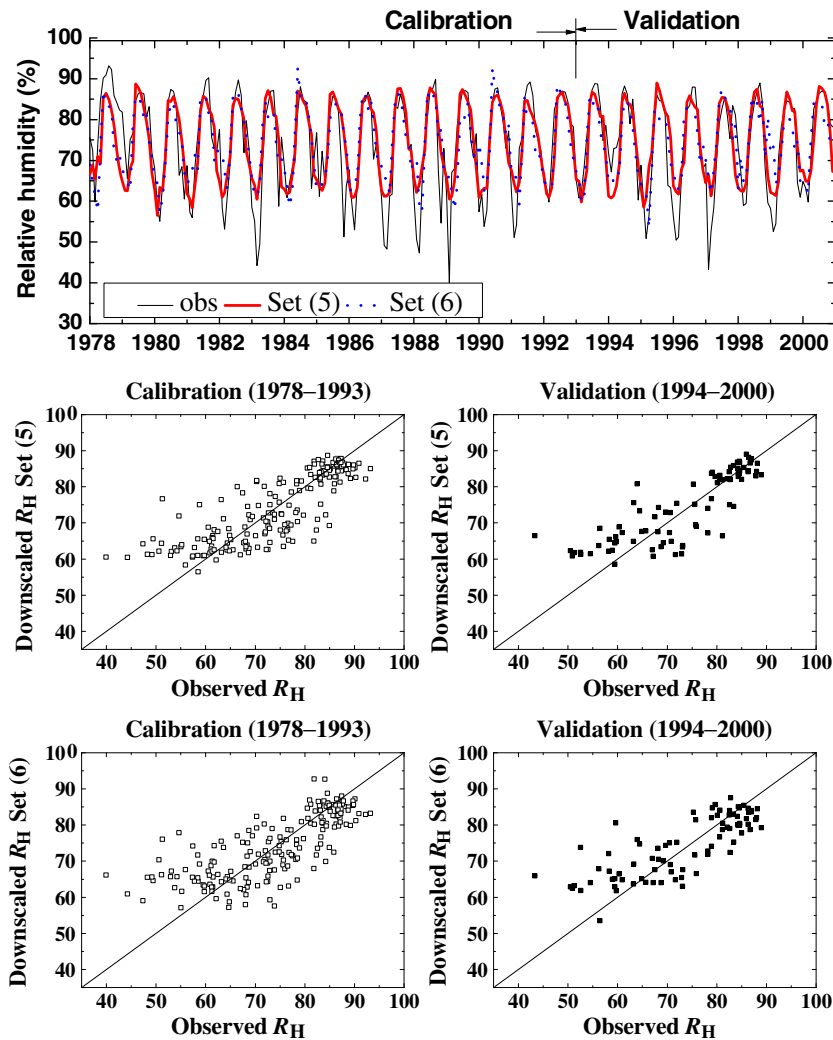


Figure 5. Calibration (1978–1993) and validation (1994–2000) results for monthly downscaled  $R_H$  using set (5) and set (6). All  $R_H$  values are in %. The first plot has the following three time series (i) downscaled  $R_H$  using set (5) (solid red lines), (ii) downscaled  $R_H$  using set (6) (dotted blue lines) and observed  $R_H$  (solid black line). The scatterplots show the monthly downscaled  $R_H$  versus the observed monthly  $R_H$  for calibration and validation periods. The lines on the scatterplots show 1:1 relationship.

circulation patterns are the primary drivers of day-to-day and inter-annual variations in surface climate, bringing with them local to regional-scale changes in features of surface weather such as humidity, precipitation and temperature that are characteristic to those patterns (Vrac *et al* 2007). The PDFs of the predictors in land and sea are shown in figure 4(b). It can be observed from the figure that the differences in the distribution between land and sea vary with predictors. The difference is less for temperature at 925 mb pressure level than at the surface.

To study the differences among predictors, they are divided into four groups, namely all the four predictors (Hus 925, Ta 925, Ta sur, and LH), three predictors (Hus 925, Ta 925, Ta sur), temperature predictors only (Ta 925, Ta sur) and only Hus 925 predictor. The cross-correlations computed between predictors in land and the predictand have slightly higher values, indicating that the predictor–predictand relationship improved marginally due to location-based stratification. The results of the effect of stratification on downscaling are discussed in the next few subsections.

#### 4.2 Downscaling model

$R_H$  estimated using eight downscaling models (one for each predictor set) is compared with the observed  $R_H$  for the calibration (1978–1993) and validation (1994–2000) periods. Among the models, the model developed using set (5) had the highest Nash–Sutcliffe coefficient of efficiency ( $E_f$ ) of 0.68. This set consists of all the four predictors (Hus 925, Ta 925, Ta sur and LH) without stratification. Model developed using set (6) comprising only temperature predictors (Ta 925 and Ta sur) without stratification had the least  $E_f$  of 0.58. The time series of downscaled  $R_H$  using models with

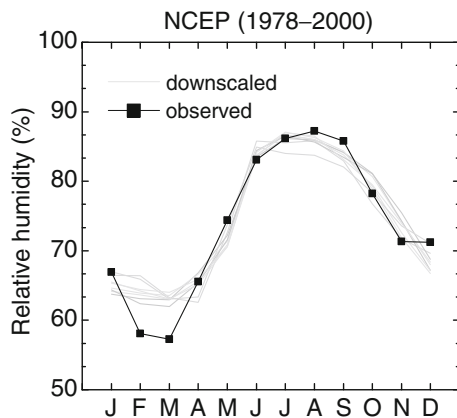


Figure 6. The mean monthly plot showing the downscaled  $R_H$  (grey lines) versus the observed (black line). There are eight grey lines for the eight downscaling models developed for the eight predictor sets.

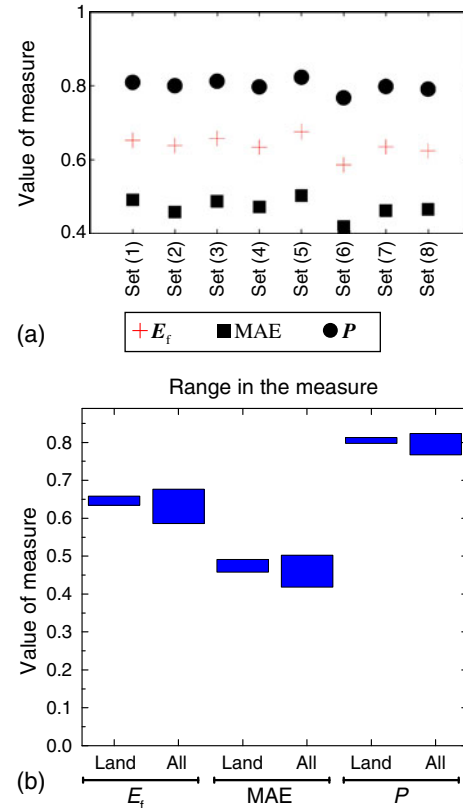


Figure 7. (a) Value of the performance measures ( $E_f$ , MAE and  $P$ ) obtained by comparing the downscaled  $R_H$  values with the observed  $R_H$  values for the period 1978–2000. Maximum values of  $E_f$ , MAE and  $P$  indicate better performance. (b) Differences in the range in performance measure obtained for predictors with and without stratification (Land and All, respectively).

highest and least  $E_f$  are shown in figure 5. It can be observed from the figure that the downscaling models did not capture the troughs (low  $R_H$ ) in the time series. The mean value of  $R_H$  for each month was calculated for the observed and downscaled  $R_H$  for the period 1978–2000 (figure 6). From the figure, it can be observed that February and March months have lower  $R_H$ , which is not captured well by the downscaling models. Stratification of predictors did not improve the performance of the downscaling model (figure 7a).

#### 4.3 Uncertainty in downscaled $R_H$ for baseline period (1978–2000)

The uncertainty range that can be expected in the performance of downscaled  $R_H$  due to the predictors selected from NCEP datasets in this study is between 0.59 and 0.68 for  $E_f$ , 0.42 to 0.50 for MAE and 0.77 to 0.82 for  $P$  (figure 7b). Although the performance of the downscaling models did not improve with stratification, the uncertainty range decreased with stratification (figure 7b). The

ranges in the performance measures due to stratification are 0.63 to 0.66 for  $E_f$ , 0.46 to 0.49 for MAE and 0.80 to 0.81 for  $P$ .

The uncertainty range in the downscaled  $R_H$  across the different percentiles due to the predictors selected is shown in figure 8. It can be observed from the figure and the statistics calculated that:

- (i) uncertainty range is between 1 and 5.4%  $R_H$  due to the predictors selected from NCEP datasets and it increases slightly from 1.5% to 6.5%  $R_H$  for GCM predictor sets,
- (ii) while using NCEP and GCM predictor sets, uncertainty range statistics in the downscaled values are: the average 2.2 and 4.4%  $R_H$ , the standard deviation 0.7 and 1.4%  $R_H$ , median 1.5 and 4.1%  $R_H$ , respectively,
- (iii) uncertainty range statistics are higher for  $R_H$  downscaled using GCM predictors when

compared to the NCEP. This is due to the differences between the predictors obtained from the two sources, and

- (iv) the uncertainty ranges are higher at the extreme values of  $R_H$  (close to 0th and 100th percentiles).

#### 4.4 Uncertainty in downscaled $R_H$ for SRES scenarios (2001–2100)

The CDFs and the uncertainty range in the downscaled  $R_H$  due to the predictors selected and four emission scenarios (A1B, A2, B1 and COMMIT) are shown in figure 9. For the four emission scenarios, the minimum uncertainty range in  $R_H$  is observed for the  $F(x)$  values  $\sim 0.3$  to  $0.5$  ( $\sim 30$  to  $50$  percentiles), while the maximum ranges in  $R_H$  are observed close to 1st and 90th percentiles. This

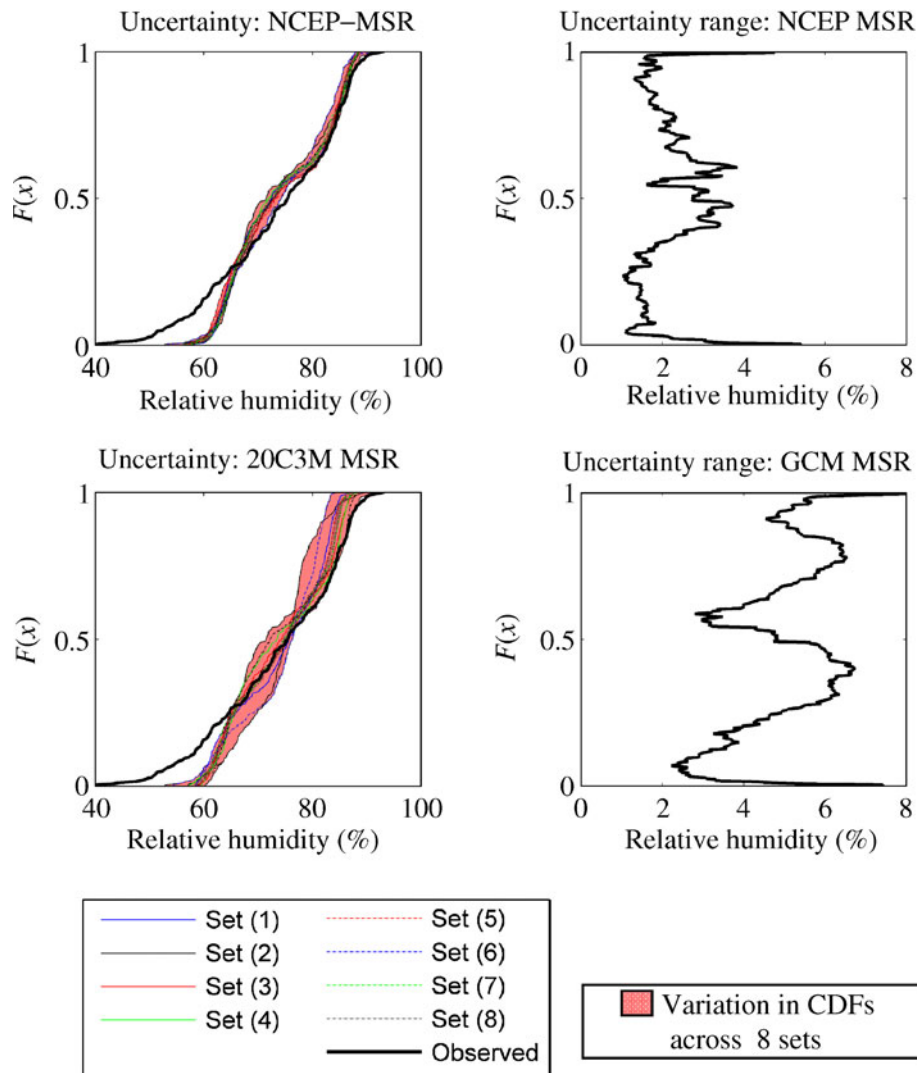


Figure 8. The cumulative distribution functions (CDFs) (column 1), estimated from observed and downscaled  $R_H$  for eight sets of predictor for the baseline period (1978–2000). The uncertainty range in the downscaled  $R_H$  (column 2). The legend is for CDFs (column 1).

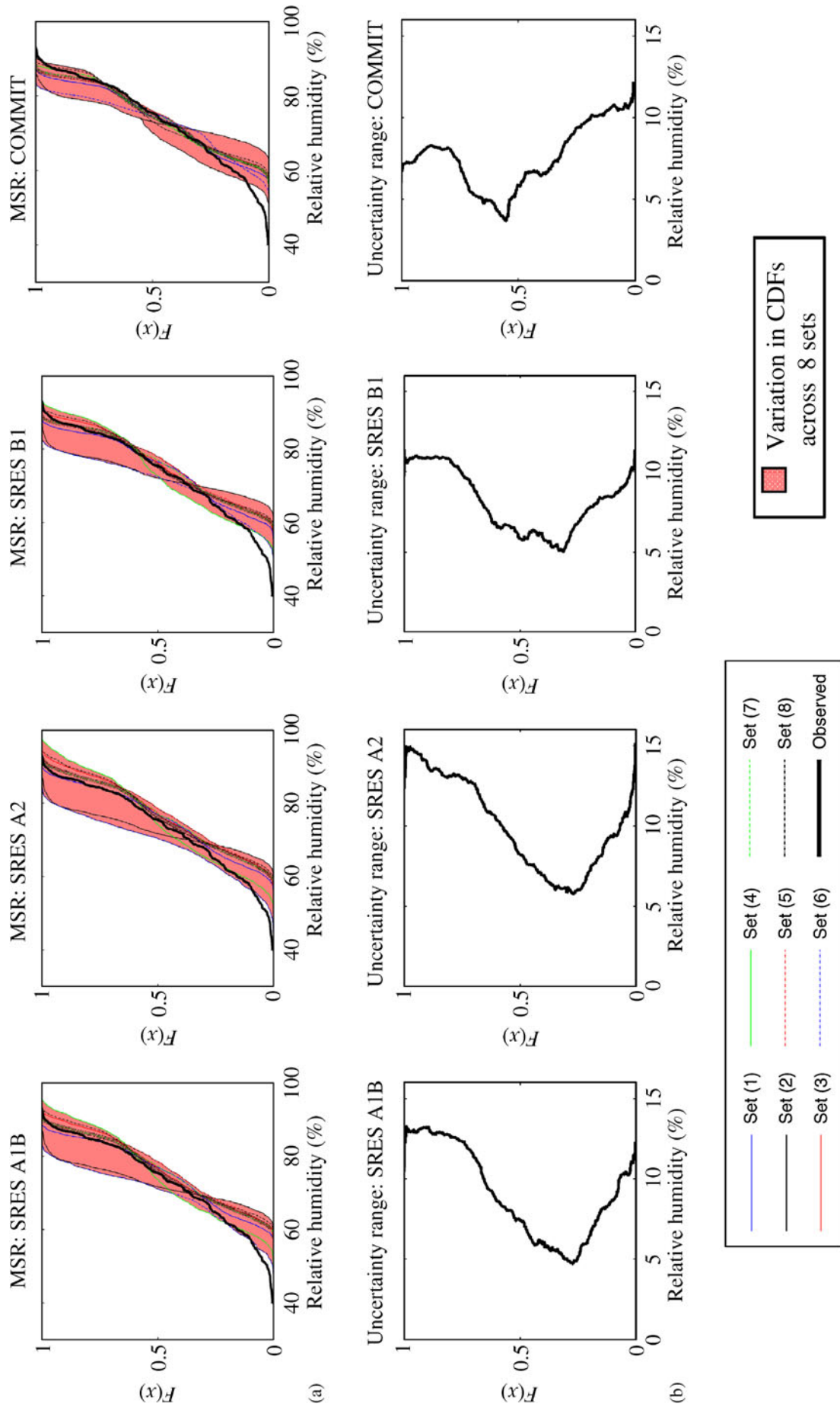


Figure 9. The cumulative distribution functions (CDFs) (a) and uncertainty range (b) obtained from the future scenarios of downscaled  $R_H$  for combinations of predictors (eight sets) and emission scenarios (A1B, A2, B1 and COMMIT) for the period 2001–2100. Columns represent the IPCC SRES scenarios.

indicates that for the four emission scenarios, the differences in the CDFs from the eight predictor sets is less around the median values of downscaled  $R_H$  (less uncertainty range), while the differences in CDFs were more for the extreme values of downscaled  $R_H$  (more uncertainty range). The maximum ranges in the A2, A1B, B1 and COMMITs scenarios are between 11 and 15%  $R_H$ . The mean range is between 8.5 and 10%  $R_H$ . The standard deviations of the ranges were 0.5 and 1.8%  $R_H$ . All the statistics of the range show that the uncertainty due to the predictors is maximum for A2 scenario, while B1 and COMMIT scenarios had the least uncertainty.

## 5. Summary

The future scenarios of monthly relative humidity ( $R_H$ ) for Malaprabha river basin, was estimated using a transfer function (a type of statistical downscaling technique). For this purpose, MLR is used.

The objective of this study was to estimate scenarios of  $R_H$  and investigate the uncertainties in the estimated scenarios for the choice of emission scenarios and predictors.

The predictors considered in this study are large-scale atmospheric variables namely air temperature and specific humidity at 925 mb, surface air temperature and latent heat flux. The predictor variables are obtained from (1) NCEP reanalysis dataset for the period 1978–2000, and (2) simulations of the third generation CGCM for the period 1978–2100. Four emission scenarios pertaining to the IPCC fourth assessment report (AR4) were used in this study. They were SRES A1B, A2, B1 and COMMIT.

The predictors were stratified based on the location of the grid point (land or sea) to study its effect on downscaling. Out of the nine grid points chosen in this study, six grid points are on land and rest are in sea. The predictors were grouped into eight sets (four sets for land grid points and four without stratification). An MLR downscaling model was developed for each predictor set. For the baseline period (1978–2000), the performance of the downscaled  $R_H$  obtained for eight models are compared using three performance measures (1) Nash–Sutcliffe error estimate ( $E_f$ ), (2) mean absolute error (MAE), and (3) product moment correlation ( $P$ ). Results show for the predictors selected in this study, the performance of the downscaled  $R_H$  did not improve with stratification. The  $E_f$  would vary between 0.59 and 0.68 without stratification. However, the variation in the performance of the downscaling models were reduced with stratification ( $E_f$  between 0.63 and 0.66). Similar

results were observed in MAE and  $P$  performance measures.

CDFs were calculated using the downscaled  $R_H$  for combinations of predictor sets (8), predictor data source (2) and emission scenarios (4). For the baseline period, the uncertainty range in the downscaled  $R_H$  using NCEP and CGCM varies between 1 and 6.6%  $R_H$ . For the future period (2001–2100), for the four emission scenarios, the uncertainty range in the downscaled  $R_H$  varies between 5 and 15%  $R_H$ .

Further investigation is required in predicting  $R_H$  using other GCMs and downscaling methods. These are deferred for future studies.

## Acknowledgements

The author thanks Profs. D Nagesh Kumar, V V Srinivas, Ravi Nanjundiah and Dr Soni Pradhanang for their support, encouragement and valuable inputs.

## References

- Abdulla F, Eshtawi T and Assaf H 2009 Assessment of the impact of potential climate change on the water balance of a semi-arid watershed; *Water Resour. Manag.* **23** 2051–2068.
- Allen R, Pereira L, Raes D and Smith M 1998 Crop evapotranspiration: Guidelines for computing crop requirements; FAO Irrigation and Drainage Paper No. 56. FAO, Rome, Italy.
- Alley R, Berntsen T, Bindoff N, Chen Z, Chidthaisong A, Friedlingstein P, Gregory J, Hegerl G, Heimann M, Hewitson B, Hoskins B, Joos F, Jouze I J, Kattsov V, Lohmann U, Manning M, Matsuno T, Molina M, Nicholls N, Overpeck J, Qin D, Raga G, Ramaswamy V, Ren J, Rusticucci M, Solomon S, Somerville R, Stocker T, Stott P, Stouffer R, Whetton P, Wood R and Wratt D 2007 Climate change 2007: The physical science basis. Summary for policymakers; Geneva, CH, pp. 1–18.
- Anandhi A 2010 Assessing impact of climate change on season length in Karnataka for IPCC scenarios; *J. Earth Syst. Sci.* **119** 447–460.
- Anandhi A, Frei A, Pierson D C, Schneiderman E M, Zion M S, Lounsbury D and Matonse A H 2011 Examination of change factor methodologies for climate change impact assessment; *Water Resour. Res.* **47** W03501.
- Castellví F, Perez P J, Stockle C O and Ibañez M 1997 Methods for estimating vapor pressure deficit at a regional scale depending on data availability; *Agri. Forest Meteorol.* **87** 243–252.
- Castellví F, Perez P J, Villar J M and Rosell J I 1996 Analysis of methods for estimating vapor pressure deficits and relative humidity; *Agri. Forest Meteorol.* **82** 29–45.
- Doty B and Kinter J I 1993 The grid analysis and display system (grads): A desktop tool for earth science visualization; American Geophysical Union 1993 Fall Meeting, San Francisco, CA, 6–10 December.
- Enke W and Spekat A 1997 Downscaling climate model outputs into local and regional weather elements by classification and regression; *Climate Res.* **8** 195–207.

- Fowler H J, Blenkinsop S and Tebaldi C 2007 Linking climate change modeling to impacts studies: Recent advances in downscaling techniques for hydrological modeling; *Int. J. Climatol.* **27** 1547–1578.
- Harmel R D and Smith P K 2007 Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling; *J. Hydrol.* **337** 326–336.
- Harmel R D, Smith D R, King K W and Slade R M 2009 Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications; *Environ. Model. Softw.* **24** 832–842.
- Huth R 2005 Downscaling of humidity variables: A search for suitable predictors and predictands; *Int. J. Climatol.* **25** 243–250.
- IPCC ed 2001 Climate change 2001: The scientific basis contributions of working group I to the third assessment report of the Intergovernmental Panel on Climate Change; Cambridge, United Kingdom and New York, Cambridge University Press, 881 p.
- Johnson M S, Coon W F, Mehta V K, Steenhuis T S, Brooks E S and Boll J 2003 Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern us: A comparison of hspf and smr; *J. Hydrol.* **284** 57–76.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K C, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R and Joseph D 1996 The NCEP/NCAR 40-year reanalysis project; *Bull. Am. Meteorol. Soc.* **77** 437–471.
- Kistler R, Collins W, Saha S, White G, Woollen J, Kalnay E, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R and Fiorino M 2001 The ncep–ncar 50-year reanalysis: Monthly means cd-rom and documentation; *Bull. Am. Meteorol. Soc.* **82** 247–267.
- Knight C G, Knight S H E, Massey N, Aina T, Christensen C, Frame D J, Kettleborough J A, Martin A, Pascoe S, Senserson B, Stainforth D A and Allen M R 2007 Association of parameter, software, and hardware variation with large-scale behaviour across 57,000 climate models; *PNAS* **104** 12,259–12,264.
- Krueger T, Quinton J N, Freer J, Macleod C J A, Bilotta G S, Brazier R E, Butler P and Haygarth P M 2009 Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer; *J. Environ. Qual.* **38** 1137–1148.
- Krueger T, Freer J, Quinton J N, Macleod C J A, Bilotta G S, Brazier R E, Butler P and Haygarth P M 2010 Ensemble evaluation of hydrological model hypotheses; *Water Resour. Res.* **46** W07516.
- Legesse D, Vallet-Coulomb C and Gasse F 2003 Hydrological response of a catchment to climate and land use changes in tropical Africa: Case study south central ethiopia; *J. Hydrol.* **275** 67–85.
- Mearns L, Hulme M, Carter T, Leemans R, Lal M and Whetton P 2001 Climate scenario development; Chapter 13. In: *Climate Change 2001: The Scientific Basis* (eds) Houghton J *et al* (Cambridge, UK: Cambridge University Press), 739–768.
- Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenham J, Gaffin S, Gregory K, Grübler A, Jung T, Kram T, La Rovere E, Michaelis L, Mori S, Morita T, Pepper W, Pitcher H, Price L, Raihi K, Roehrl A, Rogner H, Sankovski A, Schlesinger M, Shukla P, Smith S, Swart R, van Rooijen S, Victor N and Dadi Z 2000 IPCC special report on emissions scenarios; Cambridge, United Kingdom and New York, NY, USA, 599.
- Nash J E and Sutcliffe J V 1970 River flow forecasting through conceptual models. Part I – a discussion of principles; *J. Hydrol.* **10** 282–290.
- Palmer T N, Doblus-Reyes F J, Hagedorn R and Weisheimer A 2005a Probabilistic prediction of climate using multi-model ensembles: From basics to applications; *Phil. Trans. Roy. Soc. B* **360** 1991–1998.
- Palmer T N, Shutts G, Hagedorn R, Doblus-Reyes F J, Jung T and Leutbecher M 2005b Representing model uncertainty in weather and climate prediction; *Ann. Rev. Earth Planet. Sci.* **33** 163–193.
- Pearson K 1896 Mathematical contributions to the theory of evolution, III: Regression heredity and panmixia; *Phil. Trans. Roy. Soc. London Ser.* **187** 253–318.
- Perez P J, Castellvi F and Martínez-Cob A 2008 A simple model for estimating the bowen ratio from climatic factors for determining latent and sensible heat flux; *Agri. Forest Meteorol.* **148** 25–37.
- Ragab R and Prudhomme C 2002 Sw–soil and water: Climate change and water resources management in arid and semi-arid regions: Prospective and challenges for the 21st century; *Biosyst. Eng.* **81** 3–34.
- Schenk N and Lensink S 2007 Communicating uncertainty in the IPCC's greenhouse gas emissions scenarios; *Climatic Change* **82** 293–308.
- Schoof J, Pryor S and Robeson S 2007 Downscaling daily maximum and minimum temperatures in the USA: A hybrid empirical approach; *Int. J. Climatol.* **27** 439–454.
- Tebaldi C and Knutti R 2007 The use of the multi-model ensemble in probabilistic climate projections; *Phil. Trans. Roy. Soc. A* **365** 2053–2075.
- Vincent L A, van Wijngaarden W and Hopkinson R 2007 Surface temperature and humidity trends in Canada for 1953–2005; *J. Climate* **20** 5100–5113.
- Vrac M, Hayhoe K and Stein M 2007 Identification and inter-model comparison of seasonal circulation patterns over North America; *Int. J. Climatol.* **27** 603–620.
- Wang S, Yang Y, Trishchenko A P, Barr A G, Black T A and McCaughey H 2009 Modeling the response of canopy stomatal conductance to humidity; *J. Hydrometeorol.* **10** 521–532.
- Wilby R, Hay L, Gutowski W, Arritt R, Takle E, Pan Z, Leavesley G and Clark M 2000 Hydrological responses to dynamically and statistically downscaled climate model output; *Geophys. Res. Lett.* **27** 1199–1202.
- Willett K M, Jones P D, Gillett N P and Thorne P W 2008 Recent changes in surface humidity: Development of the HadCRUH dataset; *J. Climate* **21** 5364–5383.
- Wright J S, Sobel A and Galewsky J 2010 Diagnosis of zonal mean relative humidity changes in a warmer climate; *J. Climate* **23** 4556–4569.