Hydro-meteorological predictions from GCM simulations: downscaling techniques and uncertainty modelling

PRADEEP MUJUMDAR[1], SUBIMAL GHOSH[2] & DEEPASHREE RAJE[1]

1 *Department of Civil Engineering, Indian Institute of Science, Bangalore 560012, India*
pradeep@civil.iisc.ernet.in

2 *Department of Civil Engineering, Indian Institute of Technology, Powai, Mumbai 400076, India*

Abstract Hydrological implications of global climate change are usually assessed by downscaling appropriate predictors simulated by General Circulation Models (GCMs). Results from GCM simulations are subject to a number of uncertainties due to incomplete knowledge about the underlying geophysical processes of global change (GCM uncertainties) and uncertain future scenarios (scenario uncertainties). Disagreement between projections of regional climate change suggests that reliance on a single GCM with a few selected scenarios could lead to inappropriate planning and adaptation responses. This paper summarizes recent published work by the authors. The following methods and tools for statistical downscaling are discussed: (a) Fuzzy Clustering, (b) Relevance Vector Machine (RVM) and (c) Conditional Random Fields (CRFs). Uncertainty modelling with non-parametric methods and possibility theory are discussed. Applications of the methodologies are demonstrated by projection of the meteorological drought in the Orissa subdivision, India, and by predictions of the inflow to Hirakud Dam, Mahanadi River basin in India.

Key words downscaling; uncertainty; fuzzy clustering; relevance vector machine; possibility; conditional random fields

INTRODUCTION

Modelling the hydrological impacts of climate change involves using simulation results from General Circulation Models (GCMs), which are the most credible tools designed to simulate a time series of climate variables globally, accounting for the effects of greenhouse gases in the atmosphere. Despite the significant progress made in modelling future climate, uncertainties still exist (Mitchell & Hulme, 1999). Reliable results are not yet available at the spatial and temporal resolutions required for many impact studies (GCM outputs are typically at a spatial resolution of around 3° latitude and 4° longitude). This is because GCMs were not primarily designed for climate-change impact studies, and hence are not well suited for answering questions of primary interest to hydrologists concerning regional hydrological variability (e.g. Xu, 1999). A key challenge to hydrologists is thus to express the GCM results at a scale more relevant to hydrological studies, i.e. to downscale GCM experiment outputs. There are two approaches to downscaling:

(a) Dynamic downscaling uses a physical model whose grid over a limited domain is nested within the coarse grid of a GCM (Jones *et al*., 1995). The major drawback of dynamic downscaling is its complicated design and high computational cost. It is also inflexible, i.e. the experiment has to be repeated on expanding the region or moving to a slightly different region.

(b) Statistical downscaling derives empirical relationships between large-scale GCM variables (predictors) and regional-scale variables (predictands) such as precipitation and streamflow. There are three implicit assumptions involved in statistical downscaling (Hewitson & Crane, 1992): (i) the predictors are variables of relevance and are realistically modelled by the host GCM, (ii) the empirical relationship continues to be valid under altered climatic conditions, and (iii) the predictors employed fully represent the climate change signal. A detailed discussion of different downscaling models may be found in Prudhomme *et al*. (2002).

The results of downscaling depend on the accuracy of the driving GCM, hence it is essential in regional impact assessment to consider uncertainty stemming from several sources. Different levels of uncertainty are related to: (1) GCM uncertainty or inter-model variability; (2) scenario uncertainty or inter-scenario variability; (3) different realizations of a given GCM due to parameter uncertainty (intra-model variability); and (4) uncertainty due to downscaling methods. Simonovic & Li (2003, 2004) have shown the uncertainty in studies of climate change impacts on flood protections resulting from selection of GCMs and scenarios. Use of several GCMs and scenarios leads to a wide spread in the downscaled hydrological projection, especially in years far into the future, leading to uncertainties as to which among the several possible predictions should be used in developing responses. This paper discusses three applications: (a) a non-parametric method for modelling GCM/scenario uncertainty in projections of the standardized precipitation index, SPI-12, using fuzzy clustering for downscaling precipitation in the Mahanadi basin, India; (b) a possibilistic method for modelling GCM and scenario uncertainty in monsoon streamflow projections using the relevance vector machine for downscaling streamflow in the Mahanadi River; and (c) a conditional random field model applied to downscaling monsoon precipitation in the Mahanadi basin.

DROUGHT ASSESSMENT USING NON-PARAMETRIC METHODS FOR GCM/SCENARIO UNCERTAINTY

In this approach, fuzzy clustering-based downscaling (Ghosh & Mujumdar, 2006) is used for modelling future precipitation using circulation pattern, projected with the available GCM outputs. The standardized precipitation index (SPI) developed by McKee *et al*. (1993) is used as a drought index which requires precipitation as an input variable. Assuming future SPI to be a random variable at every time step, methodologies based on kernel density and orthonormal systems are used to determine the non-parametric pdf of SPI. Probabilities for different categories of future drought are computed from the estimated pdf. Details of the methodology may be found in Ghosh & Mujumdar (2007). The methodology is applied to the case study of the Orissa meteorological subdivision in India to analyse the severity of different degrees of drought in the future.

Fuzzy clustering-based downscaling

A statistical relationship based on fuzzy clustering and linear regression is developed between mean sea level pressure (MSLP) and precipitation, with reanalysis data of MSLP as predictor and observed precipitation as predictand. Gridded MSLP data used in the downscaling are obtained from the National Center for Environmental Prediction/ National Center for Atmospheric Research (NCEP/NCAR) reanalysis project (Kalnay *et al*., 1996). Monthly average MSLP outputs from 1948 to 2002 were obtained for a region spanning 15°–25°N in latitude and 80°–90°E in longitude that encapsulates the study region. Figure 1 shows the NCEP grid points superposed on the map of Orissa meteorological subdivision. The method involves training NCEP data of circulation pattern with observed precipitation and use of the resulting regression relationship in modelling future precipitation from GCM projections. The training involves three steps (Ghosh & Mujumdar, 2006): PCA, fuzzy clustering, and linear regression with seasonality terms.

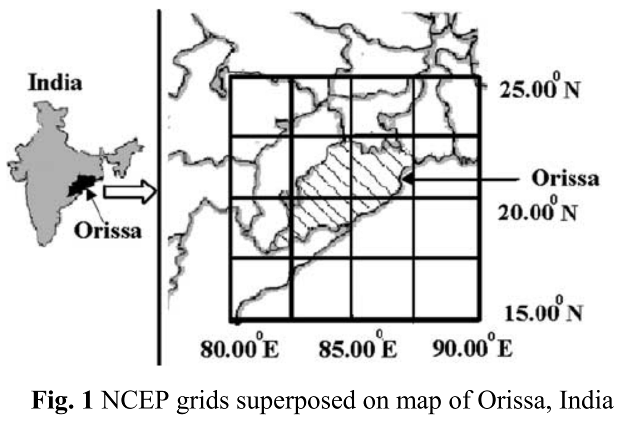


Fig. 1 NCEP grids superposed on map of Orissa, India (Ghosh & Mujumdar, 2006).

Standardization (Wilby *et al*., 2004) is used prior to statistical downscaling to reduce systematic biases in the mean and variances of GCM predictors relative to the observations or NCEP/NCAR data. The procedure typically involves subtraction of mean and division by standard deviation of the predictor variable for a pre-defined baseline period (1960–1990) for both NCEP/NCAR and GCM outputs. PCA is used to convert predictors (MSLP at 25 grid points) into a set of uncorrelated variables, with the first three principal components explaining 99.7% of the variability of the original data set. Fuzzy clustering is used to classify the principal components into classes or clusters. Fuzzy clustering assigns membership values of the classes to various data points. The important parameters required for the fuzzy clustering algorithm are the number of clusters (c) and the fuzzification parameter (m), which are determined from cluster validity indices such as the Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE). Linear regression is used to model the monthly precipitation with principal components, membership values of the principal components in each of the clusters, and the cross product of membership values and principal components as regressors. An appropriate seasonality term is used to capture the seasonality. The linear regression equation is given by:

$$P_t = C + \sum_{i=1}^{I-1} \beta_i \times \mu_{it} + \sum_{k=1}^{K} \gamma_k \times pc_{kt} + \sum_{i=1}^{I-1}\sum_{k=1}^{K} \rho_{ik} \times \mu_{it} \times pc_{kt} \tag{1}$$

with

$$C = C^o + C^1 \times \sin(2\pi p/12) + C^2 \times \cos(2\pi p/12) \tag{2}$$

$$\beta_i = \beta_i^o + \beta_i^1 \times \sin(2\pi p/12) + \beta_i^2 \times \cos(2\pi p/12) \tag{3}$$

$$\gamma_k = \gamma_k^o + \gamma_k^1 \times \sin(2\pi p/12) + \gamma_k^2 \times \cos(2\pi p/12) \tag{4}$$

$$\rho_{ik} = \rho_{ik}^o + \rho_{ik}^1 \times \sin(2\pi p/12) + \rho_{ik}^2 \times \cos(2\pi p/12) \tag{5}$$

where P_t is the precipitation at time t, pc_{kt} is the kth principal component of circulation pattern at time t, and μ_{it} is the membership in cluster i of the principal components at time t. Parameters K and I are the number of principal components used and the number of clusters, respectively; β_i, γ_k, and ρ_{ik} are the coefficients of μ_{it}, pc_{kt}, and their product terms, respectively; and C is the constant term used in the equation. The membership values μ_{it} in each cluster are assigned to the different points based on fuzzy c-means algorithm. Seasonality is incorporated by equations (2)–(5), where p is the serial number of the month within a year (p = 1, 2, … , 12). The correlation coefficient (r) between the observed and predicted precipitation is used to measure the goodness of fit of the regression model. Here the r value obtained is 0.924. The long-term mean and median of observed *vs* model-predicted precipitation for the wet (JJAS) and dry period shows a good match.

GCM output pre-processing

GCM grid points do not match with the NCEP grid points and, hence, interpolation is performed with a linear inverse square procedure using spherical distances (Willmott *et al*., 1985) to obtain the GCM output at NCEP grid points. The eigenvectors or principal directions obtained from NCEP data are used as a reference to convert the gridded standardized GCM output to the corresponding principal components.

Bias removal

The bias of annual mean of precipitation as downscaled from different standardized GCM outputs is compared to observed data for the baseline period and it is seen that, even after standardization, the bias is not significantly reduced. To remove the biases, the 1961–1990 simulated mean is subtracted, and the observed baseline period mean is added, so that all the models have the same mean in the historic period, and thus the resulting uncertainty is solely due to GCM and scenario uncertainty and not due to biases present in the GCMs.

The downscaling model significantly underestimates the inter-annual variability, most notably in the wet season. A reason for this may be the insensitivity of MSLP in correctly modelling precipitation. MSLP can partially explain historic rainfall variation, but an improvement of the model is possible if moisture content or humidity is incorporated. In the present study the analysis is only limited with MSLP because for most of the GCMs used, the outputs of moisture content or humidity are not available. The precipitation, thus computed for all the GCMs with scenarios, is converted into a suitable drought indicator for examining future drought scenario.

Uncertainty modelling

The severity of future drought may be studied by estimating the evolution of the pdf of a drought indicator. The drought indicator, SPI-12 (McKee *et al.*, 1993) values computed with downscaled precipitation from GCMs are considered as the realizations of the random variable SPI-12 in each year The pdf is estimated based on: (a) assumption of normal distribution, (b) a kernel density estimation, and (c) an orthornormal series.

Kernel density estimation entails a weighted moving average of the empirical frequency distribution of the data. Most non-parametric density estimators can be expressed as kernel density estimators (Scott, 1992; Tarboton *et al.*, 1998). It involves the use of kernel function ($K(x)$), defined by a function having the following property:

$$\int_{-\infty}^{\infty} K(x)\mathrm{d}x = 1 \tag{6}$$

A pdf can therefore be used as a kernel function. A normal kernel (i.e. a Gaussian function with mean 0 and variance 1) is used here. A kernel density estimator ($\hat{f}(x)$) of a pdf at x is defined by:

$$\hat{f}(x) = (nh)^{-1} \sum_{l=1}^{n} K((x - x_l)/h) \tag{7}$$

where n is the number of observations (here number of available GCM outputs), x_l is the lth observation (here SPI-12), and h is the smoothing parameter known as bandwidth, which is used for smoothing the shape of the estimated pdf.

A pdf from a small sample can be estimated using the orthonormal series method, which is essentially a series of orthonormal functions obtained from the sample. The summation of the series with coefficients results in the desired pdf. For this work, the orthonormal series as the subset of the Fourier series consisting of cosine functions is selected:

$$\phi_o(x) = 1 \text{ and } \phi_j(x) = \sqrt{2}\cos(\pi j x) \quad j = 1,2,3... \tag{8}$$

The pdf of SPI-12 computed using the three methods is presented in Fig. 2 along with frequency distribution of the sample for three arbitrarily chosen years (2007, 2041 and 2093) selected from the three time slices of the years 2000–2010, 2040–2050 and 2090–2100. For all the cases, it is clear from the figure that a normal pdf fails to model the samples of SPI-12, in particular the feature of multimodality, in all the three cases. The pdf obtained using orthogonal series closely resembles the shape generated by the frequency distribution. From the overall trend in probabilities of all categories of drought, it may be concluded that the probability of near-normal condition will decrease, and the probabilities of mild, severe, and extreme droughts will increase over time.

POSSIBILISTIC APPROACH TO GCM AND SCENARIO UNCERTAINTY

Dissimilarities between the bias-corrected GCM simulations under different scenarios after the year 1990 (end of baseline period) result in different system performance measures which do not validate the assumptions of equi-predictability of GCMs and equi-possibility of scenarios, which are made in the earlier non-parametric analysis. Details of the methodology may be found in

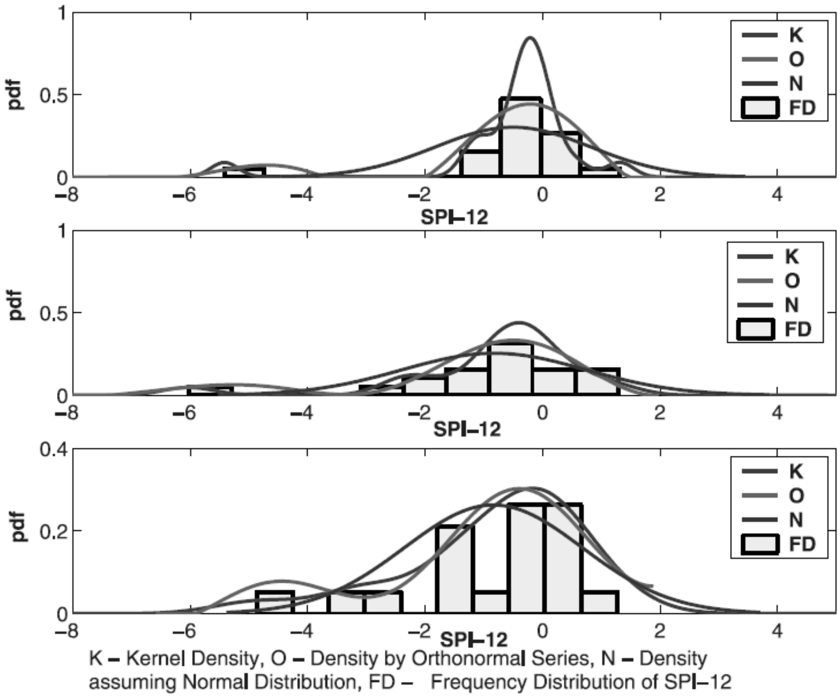


Fig. 2 Estimation of pdf of SPI-12 for the years 2007, 2041 and 2093 (Ghosh & Mujumdar, 2007).

Mujumdar & Ghosh (2008). A downscaling method based on fuzzy clustering and Relevance Vector Machine (RVM) is applied to project monsoon streamflow from three GCMs with two greenhouse emission scenarios. Possibility theory is an uncertainty theory devoted to addressing partially inconsistent knowledge and linguistic information based on intuition. Unlike probability, possibility is not computed from a frequency resulting from a sample, but is assigned to an event based on intuitive argumentation (Scott, 1999). This intuition about the future hydrological condition is derived based on the performance of GCMs with associated scenarios in modelling the streamflow of the recent past (1991–2005), when there are signals of climate forcing. Application of the possibilistic model is demonstrated with the monsoon streamflow of Mahanadi at Hirakud Dam.

Downscaling to streamflow with relevance vector machine

A statistical downscaling model based on PCA, fuzzy clustering and Relevance Vector Machine (RVM) is developed to predict the monsoon streamflow of Mahanadi River at Hirakud Reservoir, from GCM projections of large-scale climatological data. Surface air temperature at 2 m, mean sea level pressure (MSLP), geopotential height at 500 hPa and surface specific humidity are considered as the predictors for modelling Mahanadi streamflow in the monsoon season.

The RVM (Tipping, 2001) is a statistical tool which is capable of capturing the nonlinear relationship between the predictors and predictand with minimum overfitting. The mathematical structure of an RVM model is similar to the Support Vector Machine developed by Vapnik (1995).

For each GCM and scenario, the downscaling model is applied to give a future streamflow projection. Interpolation, PCA and fuzzy clustering are performed in the same way as described for the earlier work. Principal components and cluster membership of GCM output are then used in the developed RVM regression model to project the monsoon streamflow of Mahanadi in the future.

Bias correction

For validation purposes, the monsoon streamflow is also computed for the baseline period of 1961–1990 with the GCM output. It is seen that there is considerable bias near zero flow values and in extreme cases, in spite of standardization. To remove such bias from a given downscaled output, the following methodology is used: (a) cdfs are obtained with the downscaled GCM-generated and observed streamflow for the years 1961–1990, using the Weibull probability plotting position formula, to act as a reference; (b) for a given value of GCM-generated streamflow (*X*GCM), the value of cdf (cdfGCM) is computed; (c) corresponding to cdfGCM, the observed streamflow value is obtained from the cdf of observed data; (d) the GCM-generated streamflow is replaced by the observed data, thus computed, having the same cdf value; and (e) based on the reference cdfs obtained in (a), the correction is applied to the streamflow values obtained from the GCM for the future.

Possibilistic uncertainty modelling

Possibility theory, founded by Zadeh (1978), is an uncertainty theory devoted to addressing incomplete information and partially inconsistent knowledge (Dubois, 2006). Complete ignorance about climate forcing will lead to assignment of equal possibility to all the GCMs and scenarios. With time, using the growing evidence from signals of climate forcing, it should be relevant to assign a possibility distribution to the GCMs and scenarios based on their performance in the period where climate change is visible.

The bias-corrected streamflow projections with their corresponding cdfs for four time slices, 1991–2005, 2020s, 2050s and 2080s are presented in Fig. 3. The figure shows that the cdf of streamflow downscaled from one GCM is entirely different from that of another, and also that dissimilarity exists among two scenarios of any particular GCM although all scenarios project a reduction in monsoon flow. The amount of uncertainty also increases with time: in the 2080s it is higher than the other time slices.

The possibility distribution (or more appropriately, possibility mass function) obtained for the GCMs and scenarios (normalized values) is presented in Fig. 4. The difference between the possibility values of two GCMs for a given scenario is higher than that between the possibility values for two scenarios of a given GCM, which indicates that the uncertainty due to selection of GCM is greater than scenario uncertainty.

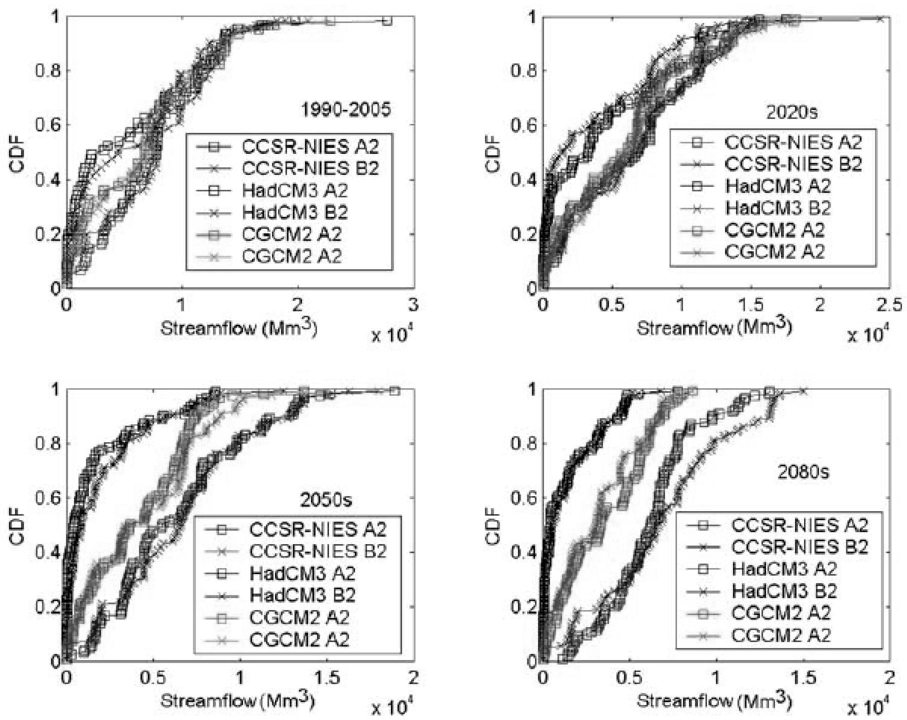


Fig. 3 The cdfs of bias-corrected streamflow projections (Mujumdar & Ghosh 2008).

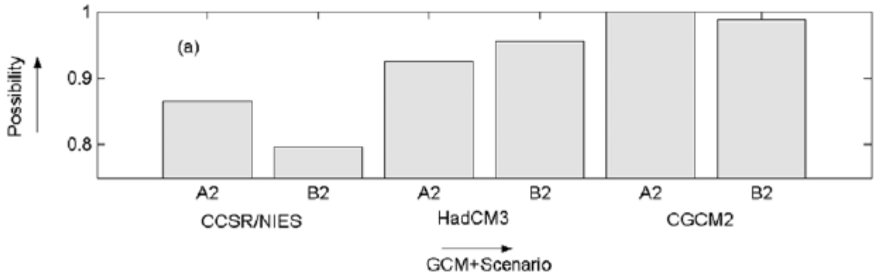


Fig. 4 Possibility distribution of GCMs and scenarios (Mujumdar & Ghosh, 2008).

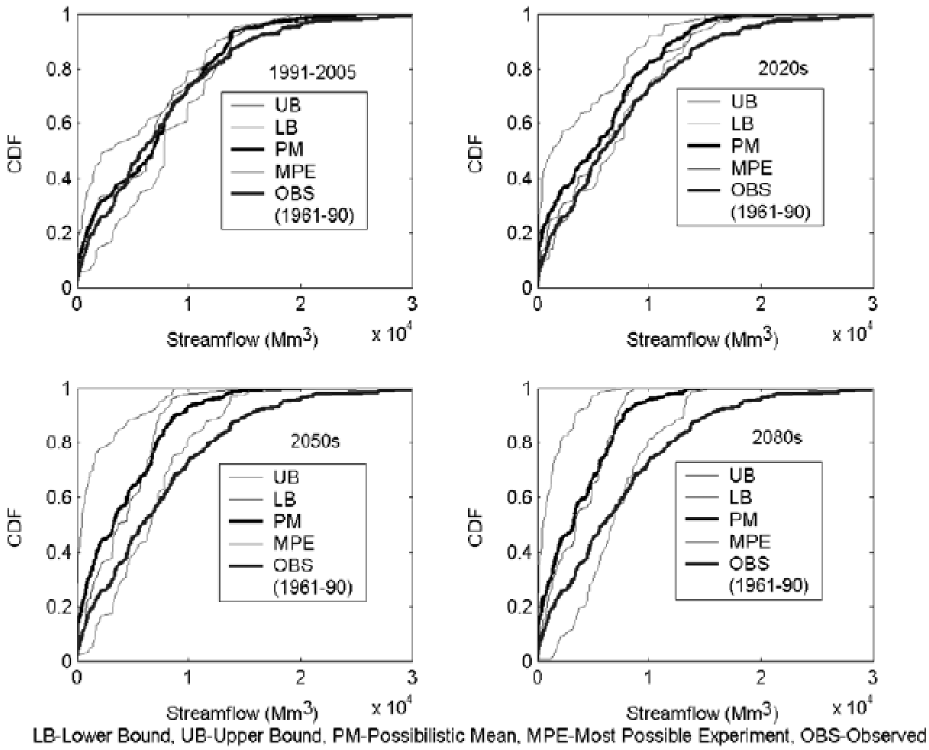


Fig. 5 Upper bound, lower bound and possibilistic mean cdf (Mujumdar & Ghosh, 2008).

The possibility values obtained for each GCM and scenario are used as weights to compute the possibilistic mean cdf (F_{pm}) for the time slices 1991–2005, 2020s, 2050s, and 2080s.

$$F_{pm} = \frac{\sum_g \sum_s \Pi(g,s) \times F_{gs}}{\sum_g \sum_s \Pi(g,s)} \tag{9}$$

where $\Pi(g, s)$ and F_{gs} are the possibility and cdf associated with gth GCM and sth scenario. The upper and lower bounds, possibilistic mean cdf and the most possible cdf (cdf for the GCM/scenario with possibility 1) are presented in Fig. 5 for 1991–2005, 2020s, 2050s and 2080s. Figure 5 shows a reduction in the probability of occurrence of extreme high-flow events in the future. Significant changes are observed in the low-flow conditions.

CONDITIONAL RANDOM FIELD DOWNSCALING

Conditional random field (CRF) downscaling is a new downscaling method in which the precipitation sequence and atmospheric variables are represented as a linear chain CRF to downscale to precipitation in a probabilistic framework. CRFs are discriminative, undirected graphical models that are very powerful for modelling relational information (Lafferty *et al*., 2001). By directly modelling the conditional probability of the output variables given the observations rather than the joint probability, CRFs avoid the difficult task of specifying a generative model for observations. As a result, CRFs can handle complex dependencies between observations, enabling them to use high-dimensional feature vectors.

CRF-downscaling model

Let the daily precipitation sequence at a site be represented by y, and the observed daily atmospheric variable sequence by x. The graphical structure for these random variables is represented as a linear chain CRF. Precipitation is discretized for computational purposes into a number of classes, including a class for zero precipitation. Hence, we can write the conditional distribution of the precipitation sequence y as (Lafferty *et al*., 2001):

$$p(\mathbf{y}\,|\,\mathbf{x}) = \frac{1}{Z(\mathbf{x})}\exp\left\{\sum_{t=1}^{T}\sum_{k=1}^{K}\lambda_k f_k(y_t, y_{t-1}, \mathbf{x})\right\} \tag{10}$$

where $\{\lambda_k\}$ is a parameter vector, and f_k is a set of real valued feature functions defined on pairs of consecutive precipitation values and the entire sequence of atmospheric data. The feature functions are the key components of a CRF. Each feature function can depend on observations from any time step.

Data pre-processing

The CRF-downscaling method is applied to monsoon (June–September) daily precipitation at eight sites in the Mahanadi basin in Orissa. In this application, large-scale atmospheric predictors of sea level pressure, specific humidity at 500hPa, precipitation flux, surface air temperature at 2 m, maximum surface air temperature at 2 m, minimum surface air temperature at 2 m, surface U-wind (zonal/eastward) and surface V-wind (meridional/northward) are chosen. Data of daily values for atmospheric variables are obtained from the NCEP/NCAR reanalysis data for the period 1951–2000 (Kalnay *et al*., 1996) and are used for training the model. High-resolution gridded daily precipitation data for 1951–2000 on a 1° × 1° grid interpolated from station data are obtained from the India Meteorological Department (IMD) for training the model. The MIROC3.2 medium resolution GCM from the Center for Climate System Research, Japan with the IPCC Assessment Report 4 A1B scenario is used for prediction.

Testing results (e.g. Fig. 6) show that the model is able to reproduce the distribution of daily precipitation well in terms of number of dry days (probability of zero precipitation) and wet day amounts. The model is also able to reproduce the pdf of wet and dry spell lengths for the testing period quite well.

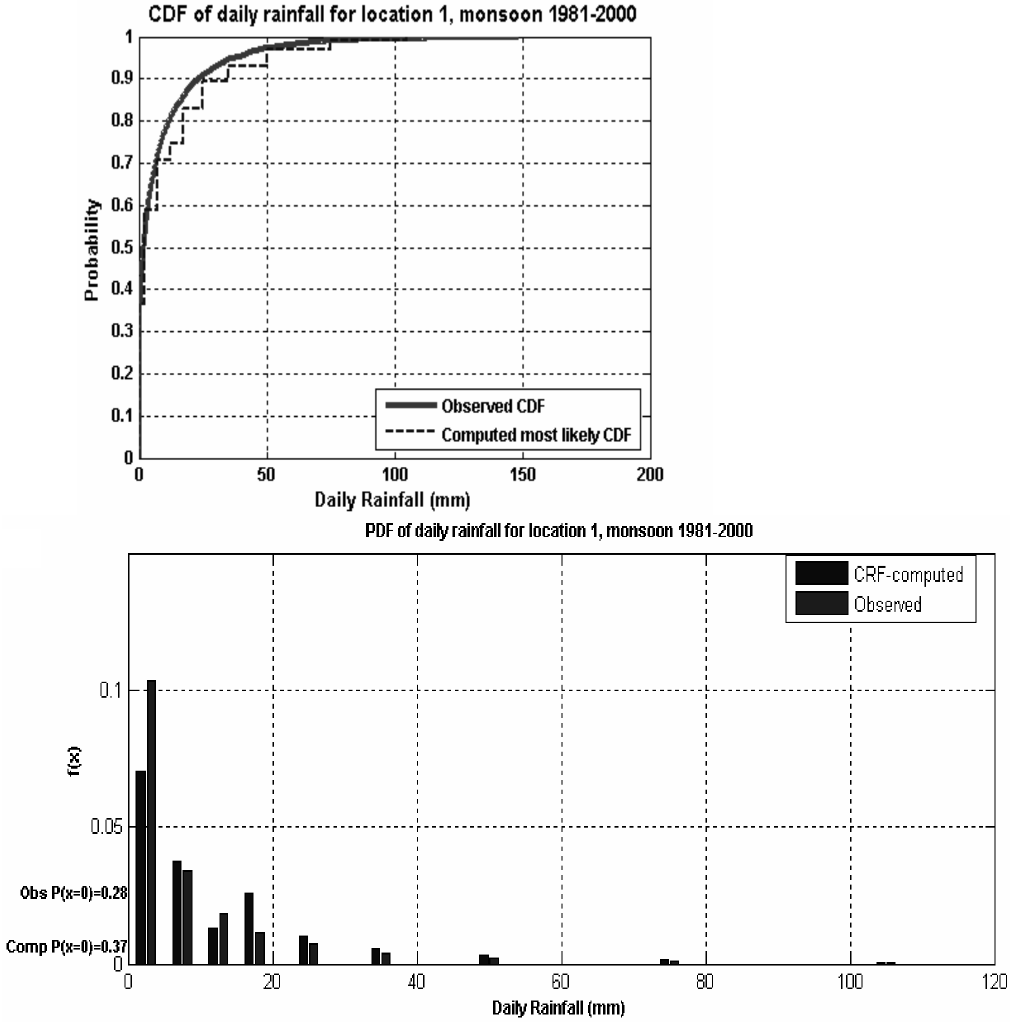
Projection of precipitation time series

The PCs of the standardized MIROC model daily outputs interpolated at NCEP grid points for the chosen predictor variables are used to predict the most likely precipitation sequences at each site. Projections for the A1B scenario (Fig. 7) for years 2046–2065 and 2081–2100 show an increase in the number of wet days for the monsoon season at two sites.

CONCLUDING REMARKS

This paper presents an overview of the published work of the authors on statistical downscaling of GCM simulations and addressing uncertainties in hydrological predictions arising from the GCMs. There are several advantages and limitations associated with the methodologies discussed in this paper. The bias removal methodology presented here ensures that GCM projections present the

(a)



(b)

Fig. 6 Observed *vs* computed most likely (a) cdf; and (b) pdf for 1981–2000. (In (b) the observed data is the right-hand bar of each pair of bars.)

uncertainty due to modelled climate change and not due to inherent bias. The non-parametric method for addressing uncertainty does not consider uncertainty due to parameterization and the structure of the impact model (GCM) itself, and those due to starting conditions used in GCM simulations and the downscaling techniques. For water resources management it is important to know the effectiveness of the GCMs in modelling climate change and which of the scenarios best represent the present situation under global warming. The possibilistic mean cdf provides a way of incorporating such information in projection of future uncertainty by assigning weights to GCMs and scenarios. Even though significant difference between the possibilities assigned to different scenarios may not be observed in the near future, there will be a growing difference between the possibility values assigned to GCMs with passage of time. Such a growing difference of the possibility values for different GCMs will increase the importance of the possibilistic model with time in future. A limitation of this approach is that uncertainties due to choice of downscaling method are not addressed in the methodology. CRF-downscaling is a new stochastic technique which does not need assumptions about independence of input atmospheric variables or their distribution and, hence, the method has substantial flexibility in using rich, overlapping features of the observations to model the conditional distribution. The limitations of the model are that it is computationally intensive and its implementation with discretization of precipitation involves loss of information. The results of the model are highly dependent on the accuracy of the ability of the driving GCM to accurately simulate atmospheric patterns.

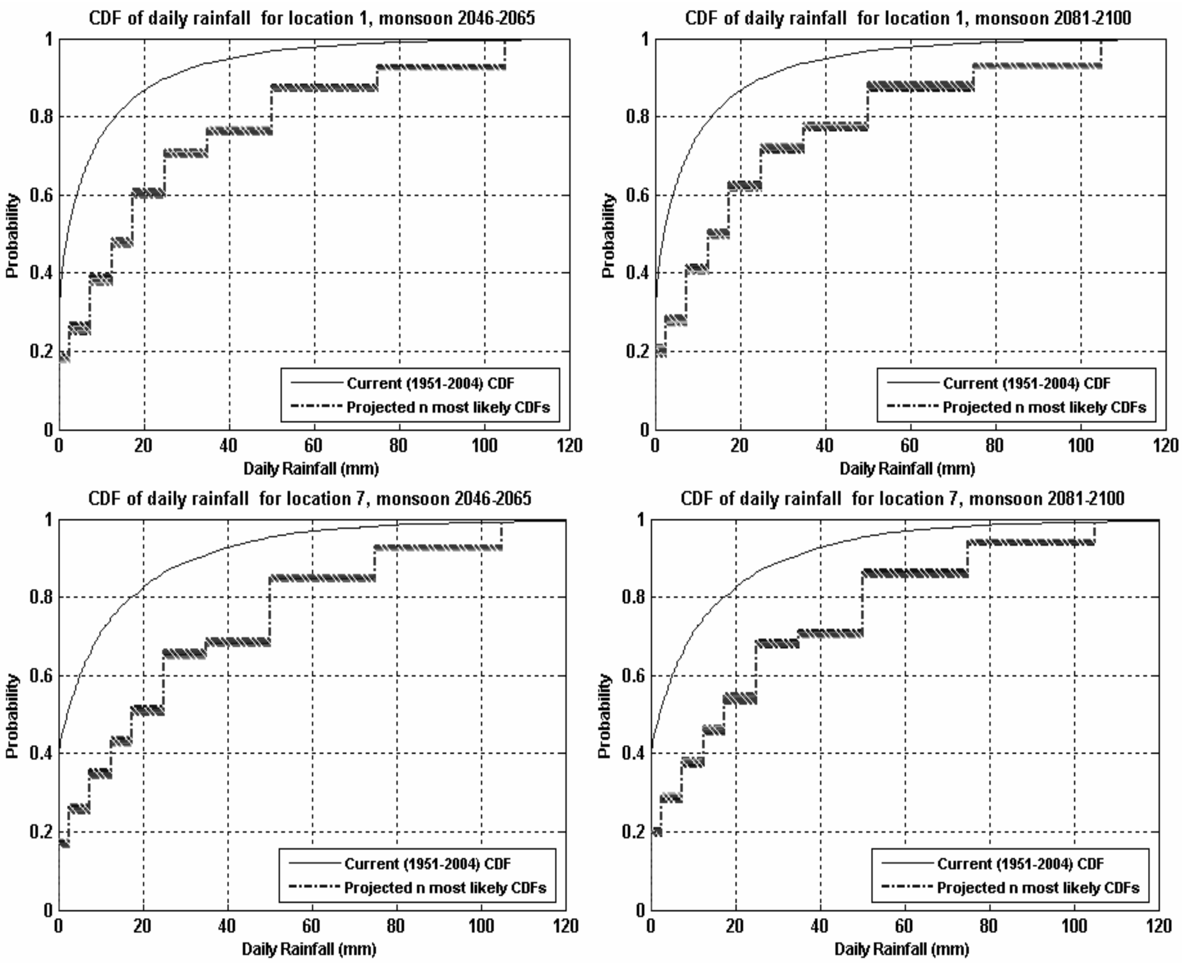


Fig. 7 Future projected most likely cdfs *vs* current (1951–2004) cdf (Raje & Mujumdar, 2009).

REFERENCES

Dubois, D. (2006) Possibility theory and statistical reasoning. *Comput. Stat. Data Anal.* **51**, 47–69.

Ghosh, S. & Mujumdar, P. P. (2006) Future rainfall scenario over Orissa with GCM projections by statistical downscaling. *Curr. Sci.* **90**(3), 396–404.

Ghosh, S. & Mujumdar, P. P. (2007) Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. *Water Resour. Res.* **43**, W07405, doi:10.1029/2006WR005351.

Hewitson, B. C. & Crane, R. G. (1992) Large-scale atmospheric controls on local precipitation in tropical Mexico. *Geophys. Res. Lett.* **19**(18), 1835–1838.

Jones, P. D., Murphy, J.M. & Noguer, M. (1995) Simulation of climate change over Europe using a nested regional-climate model: assessment of control climate, including sensitivity to location of lateral boundaries. *Quart. J. Roy. Met. Soc.* **121**, 1413–1449.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R. & Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Met. Soc.* **77**, 437–471.

Lafferty, J., McCallum, A. & Pereira, F. (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proc. 18th Int. Conf. on Machine Learning*, 282–289. Morgan Kaufmann, San Francisco, California, USA.

McKee, T. B., Doesken, N. J. & Kleist, J. (1993) The relationship of drought frequency and duration to time scale. In: *Proc. Eighth Conf. on Applied Climatology*, 179–184. American Meteorological Society.

Mitchell, T. D. & Hulme, M. (1999) Predicting regional climate change: living with uncertainty. *Progr. Phys. Geogr.* **23**(1), 57–78.

Mujumdar, P. P. & Ghosh, S. (2008) Modeling GCM and scenario uncertainty using a possibilistic approach: application to the Mahanadi River, India. *Water Resour. Res.* **44**, W06407, doi:10.1029/2007WR006137.

Prudhomme, C., Reynard, N. & Crooks, S. (2002) Downscaling of global climate models for flood frequency analysis: where are we now? *Hydrol. Processes* **16**, 1137–1150.

Raje, D. & Mujumdar, P. P. (2009) A conditional random field based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin. *Water Resour. Res.* 2008WR007487R (under review).

Scott, D. W. (1992) *Multivariate Density Estimation, Theory, Practice, and Visualization.* John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Simonovic, S. P. & Li, L. (2003) Methodology for assessment of climate change impacts on large-scale flood protection system. *J. Water Resour. Plan. Manage.* **129**(5), 361–371.

Simonovic, S. P. & Li, L. (2004) Sensitivity of the red river basin flood protection system to climate variability and change. *Water Resour. Manage.* **18**, 89–110.

Tarboton, D. G., Sharma, A. & Lall, U. (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour. Res.* **34**(1), 107–119.

Tipping, M. E. (2001) Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res*. **1**, 211–244.

Vapnik, V. N. (1995) *The Nature of Statistical Learning Theory.* Springer Verlag, New York, USA.

Willmott, C. J., Rowe, C. M. & Philpot, W. D. (1985) Small-scale climate map: a sensitivity analysis of some common assumptions associated with the grid point interpolation and contouring. *Am. Cartogr.* **12**, 5–16.

Xu, C. Y. (1999) From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progr. Phys. Geogr.* **23**(2), 229–249.

Zadeh, L. A. (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Systems* **1**(1), 3–28.