Sl. no. 35

Assessment of Water Resources under Climate Change Scenarios at River Basin Scale

No. 23/52/2006-R&D

PROJECT COMPLETION REPORT
Submitted to
Indian National Committee on Hydrology (INCOH), Roorkee

P P Mujumdar
D. Nagesh Kumar
V. V. Srinivas

Department of Civil Engineering
Indian Institute of Science
Bangalore - 560 012

May 2009

Contents

Assessment of Water Resources under Climate Change Scenarios at River Basin Scale

No. 23/52/2006-R&D

1. Name and Address of the Institute: Indian Institute of Science
Bangalore 560 012

2. Name and Address of the P.I: Prof. P. P. Mujumdar
Professor and Chairman
Department of Civil Engineering
Indian Institute of Science
Bangalore 560 012
Ph : (080) 2293 2323; 2293 2669 (Off)
2360 2668 (Res)
Fax : (080) 2360 0404; 2360 0290

3. Title of the Scheme Assessment of Water Resources under Climate Change Scenarios at River Basin Scale

4. Financial Details Pl. see Appendix 4.

5. Original Objectives and Methodology as in the Sanctioned Proposal

(a) **Objectives** (Reproduced from the sanctioned proposal):

The following are the approved objectives of the project.

1. To analyze long-term rainfall and runoff processes, water demands and extreme hydrological events in Mahanadi and Krishna river basins of India.
2. To identify a set of climate variables affecting the magnitude, temporal and spatial variability of streamflow and evapotranspiration
3. To develop stochastic/statistical relationships between the climate variables and the two hydrologic variables (streamflow and evapotranspiration) for the two river basins.
4. To construct long term future hydrologic scenarios by downscaling GCM outputs to hydrologic variables at basin scale for a diverse range of climate change scenarios.

5. To study the implications of climate change on water resources in the two river basins in terms of changes in water availability, water demands, and magnitude and frequency of hydrologic extremes.

6. To suggest measures for sustainable management of surface water resources in the selected river basins, based on key findings, and

7. To provide guidelines to the policy makers regarding adaptation of water resource projects to mitigate the impact of climate change.

(b) **Methodology Proposed** (reproduced from the sanctioned proposal)

The broad methodology consists of (a) carrying out a base line study on the hydrology of two river basins situated in different hydro-climatic regions in the country, (b) identifying a set of climate variables that may have influence on the precipitation, runoff and evapotranspiration in the basins, (c) establishing statistical relationships between the climate and the hydrologic variables , (d) developing downscaling models to scale down the GCM outputs to basin scale, (e) generating future plausible hydrologic scenarios corresponding to a number of GCM scenarios, and (f) assessing the implications for water availability and water use in the river basins, with a view to provide measures for sustainable management of surface water resources.

The following steps indicate the specific methodology to be adopted in the project.

1. The Mahanadi and the Krishna river basins will be chosen as pilot case studies to develop and demonstrate methodologies to assess the climate change impact on water availability. These two basins are of reasonably large size and would enable development of downscaling models from the GCM outputs. The Mahanadi basin is mainly flood prone, and the Krishna basin is predominantly drought prone, thus together providing an opportunity to develop an understanding of the climate change implications on water availability in two hydrologically extreme situations. The investigators have a reasonably good acquaintance with the two river basins through their earlier research. A significant amount of data needed for the proposed study, however, is not available with the investigators and needs to be collected. It is expected that all hydrologic data would be made available through the International Hydrology Program (IHP), by National Institute

of Hydrology (NIH). The following hydrologic and physical data will be collated for the two basins: (a) Precipitation and streamflow data, (b) Flood magnitudes and frequencies (b) Evaporation (or evapotranspiration) data at several locations in the basins, (c) Planned cropping patterns (d) Seasonal water utilisations, (e) Details of all major control structures (such as storage reservoirs) in the basins, and (f) Any other data useful for the studies. It is emphasized that collection of data would not be a major activity of the proposed project, and it is expected that all relevant hydrologic and physical data will be supplied by NIH through IHP.

2. The climate related data at the basin scale will be collected from sources such as the India Meteorological Department (IMD). The coarser resolution climate data, for use with GCMs in downscaling, will be collected from sources such as the National Center for Environmental Prediction (NCEP), USA, National Oceanic and Atmospheric Administration and Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES), International Research Institute for Climate Change (IRI), Columbia University, and the Climate Diagonistic Center, Boulder, Colorado. These data available from different sources are, in general, available at different grid sizes, and therefore, suitable interpolations will be necessary to relate them to the particular basins. The downscaling models will be calibrated with concurrent coarse resolution and fine resolution data (at basin scale), with a view to obtain basin scale data for a given coarse resolution data. This downscaling is necessary because all GCM outputs (future climate scenarios) are available only at coarse resolutions of about 3^0. The climate variables which would be used as predictors to obtain future scenarios of predictands (hydrologic variables) include air and sea surface temperature, relative and specific humidities, zonal, meridional and vertical components of wind velocity, geopotential height, sea level pressure among others. The coarse resolution data on these variables are available at a spatial resolution of 3^0 or higher, at various temporal resolutions (e.g., daily, monthly and annual) on the web sites of the sources mentioned. One of the investigators has recently worked with one such data set through a student dissertation (Tripathi, 2004).

3. Statistical and time series analysis of the hydrologic and climatic data will be carried out to understand the inter-annual variability of precipitation and water availability in the two river basins, and to relate the flood events with climatic events. A main focus in the

project would, however, be on larger time scales of a season comprising of several months (e.g., the monsoon, post-monsoon and summer seasons, typically comprising of 4 months each). Since the interest is in water availability assessment, the total seasonal flow in the basin would be one of the important variables investigated, avoiding suitably the need to account for controlled flows from reservoirs. An implicit assumption in such an approach is that the reservoirs have a seasonal cycle of emptying and filling up, which is generally valid in the country.

4. A set of predictors will be identified from the statistical analysis of climate variables (such as those listed in step 2 above) for downscaling of predictors to hydrologic variables (predictands) such as rainfall, streamflow and evapotranspiration. Screening of the climate variables would be done to identify those variables which have the maximum influence on the hydrologic variables. The following guidelines are available (Wilby and Wigley, 1998) for identifying climate predictors: (a) The predictor (e.g., sea surface temperature, surface pressure, wind velocity etc.) should have high correlation with the predictand (e.g., streamflow and evapotranspiration) and the dependence should have a physical meaning, (b) The predictor should be sensitive to climate change signal, (c) The relationship between predictor and predictant should be robust and should not change due to possible climate change, and (d) The predictor should be realistically simulated by the General Circulation Models (GCMs). The methodology for identifying the predictor variables would consist of the following steps: (a) Cross correlation analysis between climate predictors and hydrologic variables, for annual and seasonal time scales (b) Cross correlation analysis across all climate predictors to screen independent predictor variables, (c) Identification of spatial domain of each predictor by studying the spatial variation of the cross correlation structures between predictands and the hydrologic variables, and (d) Qualitative analysis of nonlinear relationships through scatter plots and curve fitting between predictor and predictands (hydrologic variables). This step is necessary because the correlation analysis gives a measure of only linear relationships.

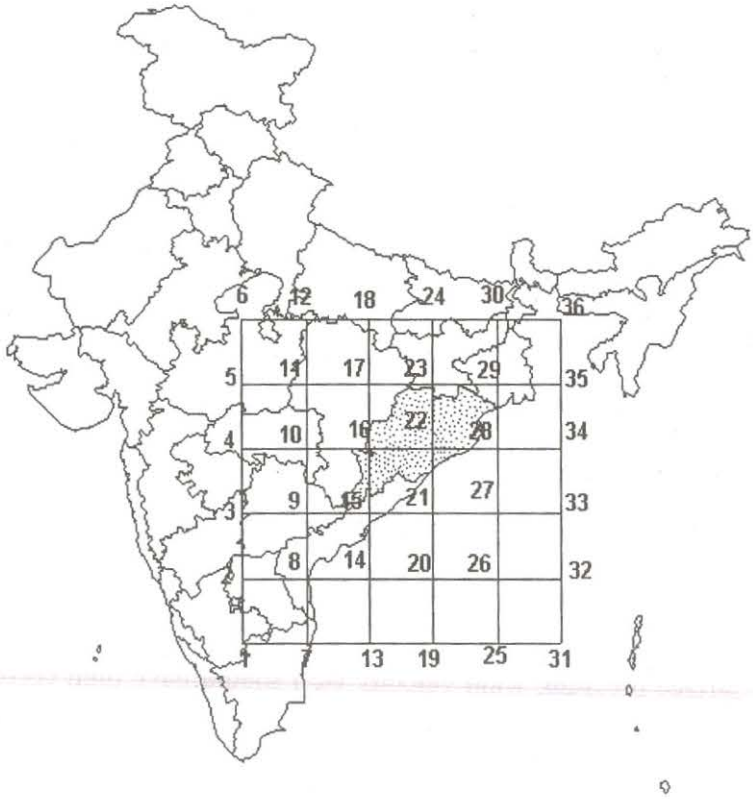

Fig 1. NCEP grid points surrounding the meteorological sub-division Orissa

5. Suitable downscaling techniques will be used/developed to downscale the GCM outputs to the river basin scales. General Circulation Models (GCMs) are the most widely used tools for studying climate change impact on various physical phenomena. Despite significant progress in developing GCMs, however, the resolution of the currently available GCMs is about 3^0, which is of the order of several thousand square kilometers. There is thus a spatial scale mismatch between the GCM scenarios of climate change and the operational scale of interest to hydrologists (such as a river basin scale). Several downscaling methods have been developed to address this problem. A recently developed method of non-parametric approach (Maity and Nagesh Kumar, 2008) will be explored for developing the downscaling models.

Statistical downscaling methods, which are based on empirical relationships that transform large scale outputs of the GCMs to regional / basin scales will also be used for comparison. Use of Artificial Neural Networks (ANNs) for downscaling will also be explored. A recent study at IISc (Tripathi, 2004) has successfully demonstrated the use of

ANNs for such a purpose. The end-product of downscaling is a set of hydrologic variables (or predictands) at a finer spatial resolution (such as river basin/watershed/sub-watershed scale), for a given set of climate variables simulated at a coarse spatial scale by GCM. Simulations provided by GCMs for several projected climate change scenarios specified by the IPCC (e.g., IPCC, 2000) such as the A1, B1, A2 and B2 scenarios, would be downscaled to the river basin scale to generate corresponding possible hydrologic scenarios, for the two river basins, with the statistical relationships developed in step 4 above.

The hydrologic scenarios will be generated for several future periods, such as 2025, 2050 etc., for which the GCM outputs are available. Among outputs available from a number of GCMs, it is proposed to use those from the second generation Coupled General Circulation Model (CGCM2) of the Canadian Center for Climate Modeling and Analysis and the second generation Global climate model of Hadley's Centre (HadCM2). It must be noted that running a GCM is much beyond the scope of the proposed project, and simulations of GCMs developed by eminent International climatologists that could be readily procured, will be used in the project.

6. From the hydrologic scenarios of streamflow and evaporation (or evapotranspiration), implications of climate change on seasonal water availability and water use will be assessed. The water use scenarios will be generated for the two basins from the generated evapotranspiration values and (projected) cropping patterns. Where available, satellite remote sensing imageries would be used to obtain land use/land cover information within the river basins. GIS would be used appropriately for spatial data base analysis and presentation of results. Essential software & hardware for image analysis and GIS is available at the Institute.

From the key findings of the study, suggestions for sustainable use of water in the basins would be made, and guidelines to policy makers regarding adaptation of water resource projects to mitigate the impact of climate change will be provided.

6. Any changes in the objectives during the operation of the scheme:

Objectives (1), (2) and (6) could not be dealt with, in the project because of shortage of time. Also, the implications of climate change were studied mainly on rainfall and

streamflow. Implications on water demand (through evapotranspiration) could not be studied. However, the methodologies developed in the project are general and are readily applicable to other variables and other regions. The work that could not be completed in this project could be taken up through subsequent projects, either by the same investigators or by others.

7. Data Collected and Used in the Analysis, with sources of data:

A considerable effort has been put in collecting hydrologic/meteorological data of the Mahanadi (Fig. 2) and Krishna (Fig. 3) river basins. The data have been collected from Central Water Commission (CWC), Hyderabad, IMD, Hirakud Project Office, Orissa, and the Water Resources Development Organisation (WRDO), Karnataka. The following data have been collected: Daily rainfall data from three IMD rain gauge stations within Mahanadi basin upstream of Hirakud reservoir: Jharsuguda Aero (1949-2000), Champa (1951-84) and Raipur (1979-99); Daily discharge data at Basantpur (1972-73 to 2003-04) and Seorinarayan (1986-87 to 2002-03), both upstream of Hirakud reservoir. Gauged streamflow data have been obtained at the gauges shown in Fig. 3, for the Krishna basin. The details of these data are given in Table 1. In addition, a CD-ROM from IMD containing High Resolution (one degree grid) Daily Rainfall Data for 53 years (1951-2003) has been obtained. Meteorological data such as the rainfall, maximum and minimum temperatures, relative humidity, wind speed and evaporation rate are acquired for several stations in Krishna basin from WRDO, Bangalore, and CRIDA, Hyderabad: Santhebastwadi (1991-2003), Badami (1991-2003), Bellary (1967-1987), Bijapur (1995-2003), Puttur (1991-2001) and Gadag (1978-2000).

In addition, depending on the region and the GCM used, appropriate data is downloaded from the websites of IPCC datacenter and the National Center for Environmental Prediction (NCEP), USA. These details are given in the methodology section.

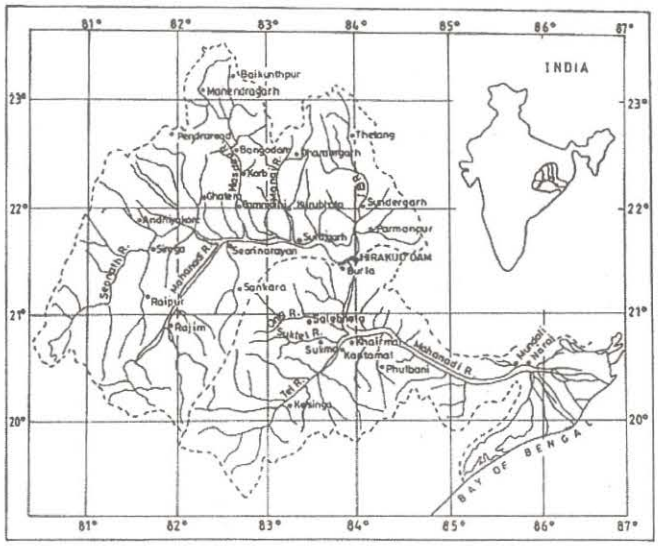

Fig 2 Mahanadi Basin

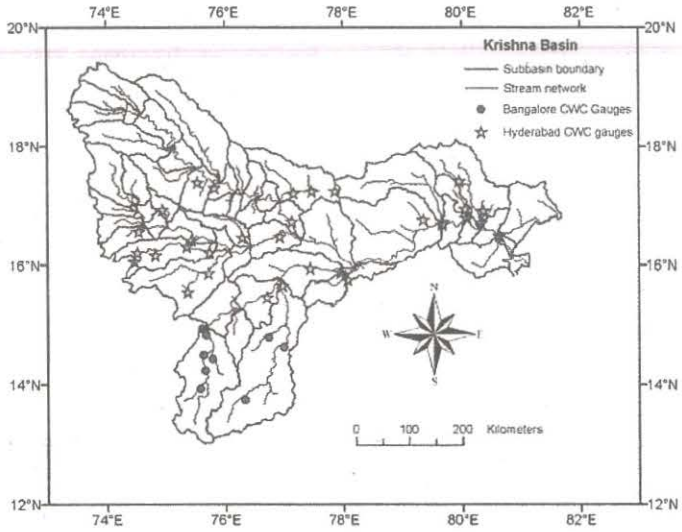

Fig 3 Krishna Basin showing Stream Gauge sites

Table 1. Details of Daily Streamflow Data Collected (Krishna Basin)

S.No.	Site / Code No.	State	District	Tributary	Latitude N	Longitude E	Record Duration
1	Bawapuram (AKLOOB8)	Andhra Pradesh	Kumool	Tungabhadra	15° 53'00"	77 ° 57'00"	11/07/1965 to 29/5/2001
2	Cholachguda(AKSOOHI)	Karnataka	Bijapur	Malaprabha	15°52"00"	75 ° 43' 00"	01/06/1982 to 29/5/2001
3	Daddi (AKTOOU8)	Karnataka	Belgaum	Ghataprabha	16°04'00"	74 ° 28' 00"	01/06/1979 to 30/5/2001
4	Dameracherla (AKFOOA7)	Andhra Pradesh	Nalgonda	Musi	16° 44'00"	79 ° 40'00"	27/07/1968 to 30/5/2001
5	Galgali (AKOOOR3)	Karnataka	Bijapur	-	16°25'00"	75 ° 26"00"	21/02/1976 to 30/5/2001
6	Gokak falls (AKTOOP9)	Karnataka	Belgaum	Ghataprabha	16°10'00"	74 ° 49'00"	13/07/1971 to 30/5/2001
7	Gotur (AKT20C3)	Karnataka	Belgaum	Ghataprabha	16 "13'00"	74° 31'00"	17/06/1980 to 30/5/2001
8	Halia (AKHOOG2)	Andhra Pradesh	Nalgonda	Halia	16°46'00"	79° 21'00"	11/07/1984 to 30/5/2001
9	Huvinhedgi (AKOOON2)	Karnataka	Raichur	-	16°29'07"	76 ° 55'07"	01/02/1976 to 30/5/2001
10	Jewangi (AKPIOM5)	Andhra Pradesh	Ranga Reddy	Bhima	17°15'00"	77 ° 28' 00"	26/11/1978 to 30/5/2001
11	K Agraharam (AKOOOK6)	Andhra Pradesh	M'boob Nagar	.	17°15'00"	77 °51'00"	30/08/1981 to 30/5/2001
12	Keesara (AKAOOB4)	Andhra Pradesh	Krishna	Munneru	16°43'00"	80° 19'00"	26/06/1964 to 30/5/2001
13	Kokangaon (AKP7006)	Karnataka	Bijapur	Bhima	17°18'00"	75 ° 49' 00"	21/07/1979 to 30/5/2001
14	Lakshmipuram (AKL10B7)	Andhra Pradesh	Kumool	Tungabhadra	15°45'30"	78 ° 04' 30"	13/09/1984 to 19/5/2001
15	Madhira (AKAIOD2)	Andhra Pradesh	Khammam	Munneru	16°55'00"	80° 21'00"	07/06/1984 to 30/5/2001
16	Malkhed(AKPIOEI)	Karnataka	Gulbarga	Bhima	17°12'35"	77 ° 09' 25"	1/08/1990 to 30/5/2001
17	Mantralayam (AKLOOF2)	Andhra Pradesh	Kumool	Tungabhadra	15° 57'00"	77°26'00"	30/06/1972 to 30/5/2001
18	Mudhol	Karnataka	Belgaum	Ghataprabha	16°1900"	75°21'00"	25/7/2000 to 30/5/2001
19	Navalgund (AKS20I2)	Karnataka	Dharwad	Malaprabha	15°33'00"	75 ° 22'00"	05/06/1991 to 30/5/2001
20	Oollenur(AKLOOKI)	Karnataka	Raichur	Tungabhadra	15°28'00"	76 ° 42' 00"	10/07/1972 to 30/5/2001
21	P.S.Gudem (AKA20F5)	Andhra Pradesh	Warangal	Munneru	17 ° 25'00"	79 ° 57' 00"	01/09/1987 to 30/5/2001
22	Paleru Bridge (AKCOOD3)	Andhra Pradesh	Krishna	Paleru	16°57'00"	80 ° 03" 00"	06/09/1964 to 30/5/2001
23	Pandegaon (AKUOOK6)	Karnataka	Belgaum	Agrani	16°56'00"	74° 56'00"	06/10/1979 to 29/5/2001
24	Pondugala (AKOOOE7)	Andhra Pradesh	Guntur	-	16° 41'00"	79° 40'00"	13/11/1975 to 30/5/2001
25	Sadalga (AKVOOD3)	Karnataka	Belgaum	Dudhganga	16°34'00"	74 ° 32'00"	24/06/1969 to 29/5/2001
26	Shirdhon (AKP80B8)	Karnataka	Bijapur	Bhima	17°24'00"	75 ° 32' 00"	30/08/1979 to 30/5/2001
27	T. Rampuram (AKLSOA6)	Karnataka	Ballary	Tungabhadra	15°39'33"	76 ° 57' 58"	18/12/1965 to 30/5/2001
28	Talikot (AKQOOE2)	Karnataka	Bijapur	Don	16°28'22"	76°17'23"	21/09/1995 to 30/5/2001
29	Vijayawada (AKOOOB9)	Andhra Pradesh	Krishna	-	16° 30'00"	80 ° 37' 00"	02/09/1964 to 30/5/2001
30	Wadenapalli (AKOOOD5)	Andhra Pradesh	Nalgonda	,	16°48'00"	80 ° 04' 00"	10/12/1965 to 30/5/2001
31	Yadgir(AKPOOB6)	Karnataka	Gulbarga	Bhima	16°44'03"	77°07'18"	01/08/1965 to 30/5/2001

8. Methodology Followed:

The following subsections provide the details of the work carried out in the project. The discussion is organized as follows: The need for downscaling, in the context of regional climate change impact assessment is discussed first, followed by a discussion on the downscaling and projections of streamflow in Mahanadi river basin using the Relevance Vector Machines (RVMs). Work carried out on assessing the impacts of climate change on meteorological drought in the Orissa meteorological subdivision, and addressing uncertainties thereof, is discussed next. Since the downscaled outputs from different GCMs and different scenarios lead to different projections for the future, weighted probability distributions are provided for the Mahanadi streamflow, the weights having been based on the performance of the particular GCMs and the scenarios during the recent past, 1991-2005. This methodology, along with discussion on the results is presented next. Extensive studies have been carried out in the project on identification of climate predictors for modeling the hydrologic variables, in the Malaprabha sub-basin of the Krishna basin. Details of these studies are presented in the last subsection.

8.1 Statistical downscaling of GCM simulations to streamflow using relevance vector machine

Modeling hydrologic impacts of climate change involves simulation results from General Circulation Models (GCMs), which are the most credible tools designed to simulate time series of climate variables globally, accounting for the effects of greenhouse gases in the atmosphere. GCMs perform reasonably well in simulating climatic variables at larger spatial scale ($>10^4$ km^2), but poorly at the smaller space and time scales relevant to regional impact analyses (Bates et al, 1998). Such poor performances of GCMs at local and regional scales have led to the development of Limited Area Models (LAMs) in which fine computational grid over a limited domain is nested within the coarse grid of a GCM (Jones et al, 1995). This procedure is also known as dynamic downscaling. The major drawback of dynamic downscaling, which restricts its use in climate change impact studies, is its complicated design and high computational cost. Moreover, it is inflexible

in the sense that expanding the region or moving to a slightly different region requires redoing the entire experiment (Crane and Hewitson, 1998). Another approach to downscaling, termed statistical downscaling, involves deriving empirical relationships that transform large scale features of the GCM (Predictors) to regional scale variables (Predictands) such as precipitation and streamflow. There are three implicit assumptions involved in statistical downscaling (Hewitson and Crane, 1992). Firstly, the predictors are variables of relevance and are realistically modeled by the host GCM. Secondly, the empirical relationship is valid also under altered climatic conditions. Thirdly, the predictors employed fully represent the climate change signal.

Statistical downscaling methodologies can be broadly classified into three categories (Murphy, 1999): weather generators, weather typing and transfer function. Weather generators are statistical models of observed sequences of weather variables. They can also be regarded as complex random number generators, the output of which resembles daily weather data at a particular location (Katz and Parlange, 1996). There are two fundamental types of daily weather generators, based on the approach to model daily precipitation occurrence: the Markov chain approach (Hughes and Guttrop, 1994) and the spell-length approach (Mehrotra and Sharma, 2005). In the Markov chain approach, a random process is constructed which determines a day at a station as rainy or dry, conditional upon the state of the previous day, following given probabilities. In case of spell-length approach, instead of simulating rainfall occurrences day by day, spell-length models operate by fitting probability distribution to observed relative frequencies of wet and dry spell lengths. In either case, the statistical parameters extracted from observed data are used along with some random components to generate a similar time series of any length. Weather typing approaches (Brown and Katz, 1995) involve grouping of local, meteorological variables in relation to different classes of atmospheric circulation. Future regional climate scenarios are constructed either by resampling from the observed variable distribution (conditioned on the circulation pattern produced by a GCM), or by first generating synthetic sequences of weather pattern using Monte Carlo techniques and then resampling from the generated data. The mean or frequency distribution of the local climate is then derived by weighting the local climate states with the relative frequencies of the weather classes. The most popular approach of downscaling is the use of transfer

function which is a regression based downscaling method (Crane and Hewitson, 1998), (Cannon and Whitfield, 2002), (Wilby et al, 2002), (Tripathi et al, 2006) that relies on direct quantitative relationship between the local scale climate variable (predictand) and the variables containing the large scale climate information (predictors) through some form of regression. Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. Todate, linear and nonlinear regression, Artificial Neural network (ANN), canonical correlation, etc. have been used to derive predictor–preditand relationship. Among them, ANN based downscaling techniques have gained wide recognition owing to their ability to capture nonlinear relationships between predictors and predictand (Crane and Hewitson, 1998), (Hewitson and Crane, 1996).

Despite a number of advantages, the traditional neural network models have several drawbacks including possibility of getting trapped in local minima and subjectivity in the choice of model architecture (Suykens, 2001). Recently, Vapnik (1995, 1998) pioneered the development of a novel machine learning algorithm, called Support Vector Machine (SVM), which provides an elegant solution to these problems. The SVM has found wide range of applications in the fields of classification and regression analysis. SVM has some drawbacks of rapid increase of basis functions with the size of training data set and absence of probabilistic interpretation (Govindaraju, 2005). Recently Tipping (2001) developed Relevance Vector Machine (RVM), a new methodology for classification and regression using the concept of probabilistic bayesian learning framework, which can predict accurately utilizing dramatically fewer basis functions than a comparable SVM while offering a number of additional advantages.

In a recent study (Tripathi et al, 2006), SVM has been used as a downscaling technique for predicting subdivisional precipitation of different regions in India. In that study, the GCM generated large scale output (predictors) are converted into principal components using Principal Component Analysis (PCA) and used directly as an input to SVM with Gaussian RBF as the kernel function. Ghosh and Mujumdar (2006) found that a heuristic classification of large scale GCM outputs based on fuzzy clustering, prior to regression, improves the model performance and thus in the present study both SVM and RVM coupled with PCA and fuzzy clustering are used to downscale GCM output to

streamflow. The flowchart of the model is presented in Fig. 8.1.1. The large scale GCM outputs are converted into principal components using PCA, which is further classified into fuzzy clusters using fuzzy c-mean clustering. The membership in each of the clusters along with the principal components is used as input to SVM/RVM. The relationship between the climate variables and streamflow is complex and nonlinear. Standard regression methods such as linear regression fail to model such nonlinear processes, and therefore SVM and RVM are used in the present study. Gaussian RBF, Laplacian RBF and heavy tailed RBF have been used as the kernel functions to compare the results. The National Center for Environmental Prediction/ National Center for Atmospheric Research (NCEP/NCAR) reanalysis data have been used for training the downscaling model and GCM output is used for projecting future streamflow with the trained model. The performance of RVM is compared with SVM for downscaling in the present study.

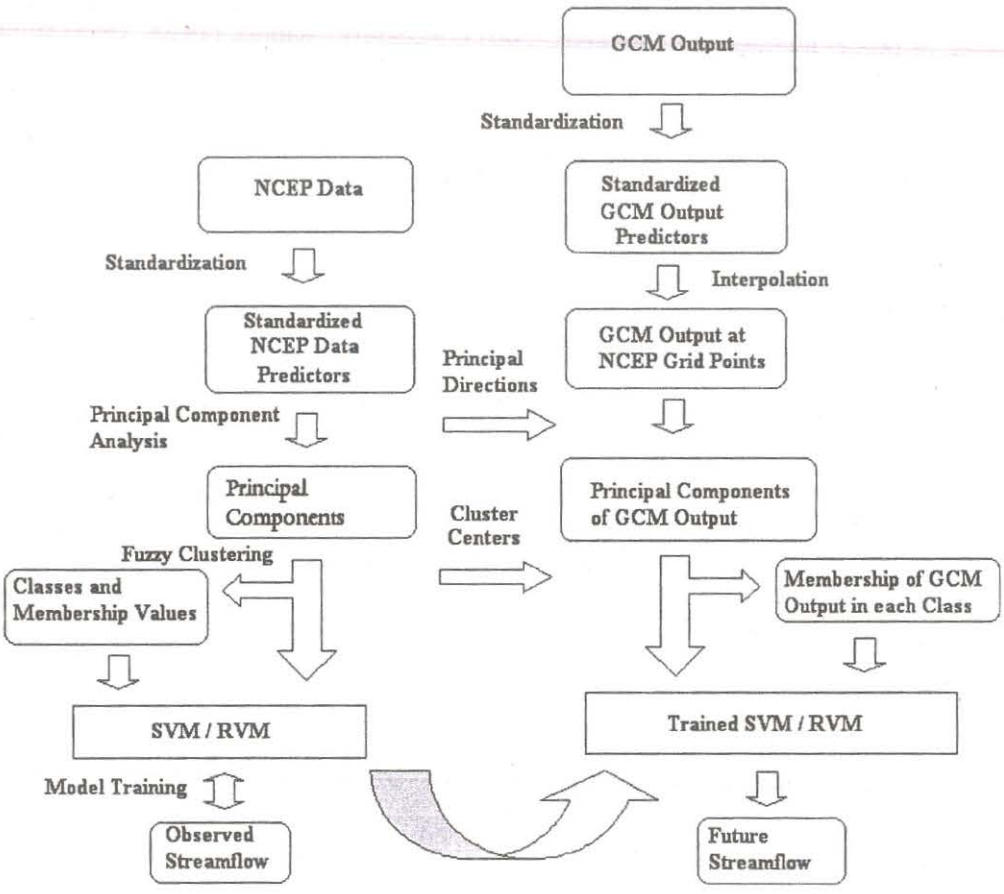

Fig. 8.1.1 Flowchart of the proposed model

Results are obtained with different kernel functions. The model is applied to the case study of Mahanadi river basin in India to model the reservoir inflow to the Hirakud dam from large scale GCM output. Details of the case-study, data and the analysis performed prior to the training of SVM or RVM is presented in the following section.

8.1.1 Data and input to vector machine

The Hirakud dam is located at Mahanadi River in Orissa at east coast of India (Fig. 8.1.2). The latitude and the longitude of the location are 21.32° N and 83.45°E, respectively. The monthly inflow to Hirakud dam from 1961 to 1990 is obtained from Department of Irrigation, Government of Orissa, India. Due to the absence of any major control structure upstream to Hirakud dam, the inflow to the dam is considered as unregulated flow. Mahanadi is a rain-fed river with high streamflow in monsoon (June, July, August and September) due to heavy rainfall and therefore the ground water component with infiltration is insignificant compared to the streamflow during the monsoon season. In the non-monsoon season, infiltration to ground water is quite significant in absence of rainfall, resulting in low streamflow in Mahanadi with almost dry conditions. Thus, only for the monsoon season the streamflow can be modeled with the climatological variables without considering ground water component. Therefore the monthly monsoon flow data of Mahanadi from year 1961 to year 1990 is used in the downscaling model as predictand. Selection of predictor is an important step in statistical downscaling. The predictors, used for downscaling (Wetterhall et al, 2005) should be: (1) reliably simulated by GCMs, (2) readily available from archives of GCM outputs, and (3) strongly correlated with the surface variables of interest. Cannon and Whitfield (2002) have used MSLP, 500 hPa geopotential height, 800 hPa specific humidity, and 100–500 hPa thickness field as the predictors for downscaling GCM output to streamflow. Monsoon streamflow can be considered broadly as the resultant of rainfall and evaporation. Rainfall is a consequence of Mean Sea Level Pressure (MSLP) (Bardossy et al, 1995), geopotential height and humidity whereas evaporation is mainly guided by temperature and humidity. Therefore, the present study considers 2 m surface air temperature, MSLP, 500 hPa geopotential height and surface specific humidity as the predictors for modeling Mahanadi streamflow in monsoon season. It is worth mentioning that land use is the single most important factor in generating the flow from the rainfall.

In the present study, land use pattern is assumed to remain the same in future and therefore the statistical relationship between the predictors and the streamflow will remain unaltered in future. Gridded climate variables are obtained from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis project (Kalnay et al, 1996) (http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml). Reanalysis data are outputs from a high resolution atmospheric model that has been run using data assimilated from surface observation stations, upper-air stations, and satellite-observing platforms. Results obtained using these fields therefore represent those that could be expected from an ideal GCM (Cannon and Whitfield, 2002). Monthly climatological data from 1961 to 1990 were obtained for a region spanning 15°N–25°N in latitude and 80°E–90°E in longitude. Fig. 8.1.2 shows the NCEP grid points superposed on the map of Mahanadi river basin. A statistical relationship based on fuzzy clustering and vector machine is developed between large scale climatic variables and inflow to Hirakud dam, with reanalysis data as regressor and observed streamflow as regressand. This relationship is used to model the future streamflow using GCM output. GCM developed by Center for Climate System Research/ National Institute for Environmental Studies (CCSR/NIES), Japan, with B2 scenario is used for projection of future streamflow. The grid size of the GCM is 5.5°latitude x 5.625°longitude.

The monthly output for B2 scenario is extracted for CCSR–NIES GCM for the region of interest covering all the NCEP grid points extending from 13.8445°N to 30.4576°N in latitude and 78.7500°E to 95.6250°E in longitude from IPCC data distribution center (http://www.mad.zmaw.de/IPCC_DDC/html/ddc_gcmdata.html).

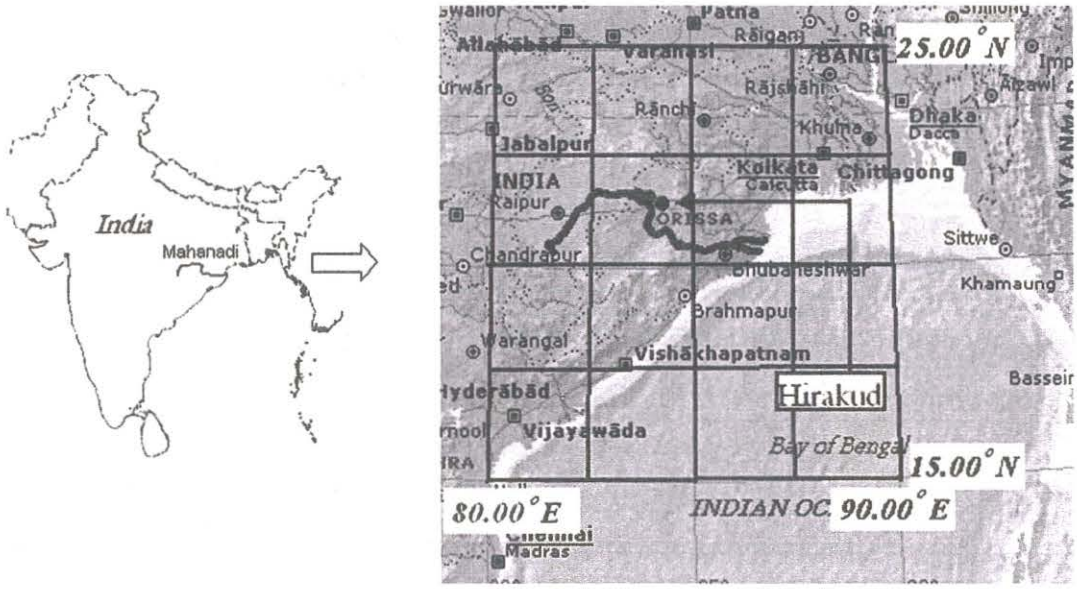

Fig 8.1.2 NCEP grids superposed on Mahanadi river basin

Standardization (Wilby et al, 2004) is used prior to statistical downscaling to reduce systematic biases in the mean and variances of GCM outputs 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 predefined baseline period for both NCEP/NCAR and GCM output. The period 1961–1990 is used as a base-line because it is of sufficient duration to establish a reliable climatology, yet not too long, nor too contemporary to include a strong global change signal (Wilby et al, 2004). A major limitation of standardization is that it considers the bias in only mean and variance. There is a possibility that the reanalysis data and GCM output may deviate from normal distribution, and there may exist bias in other statistical parameters. For Mahanadi river basin, four predictor variables (MSLP, 2m surface air temperature, specific humidity, and 500hPa geopotential height) at 25 NCEP grid points with a dimensionality of 100, are used which are highly correlated with each other. Principal Component Analysis (PCA) (Hughes et al, 1993) is performed to transform the set of correlated N-dimensional predictors (N = 100) 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. In the present study, it is observed that first 10 Principal Components (PCs) represent 98.1% of the information content (or variability) of the original predictors, and therefore they are used

in downscaling. The advantage of PCA is that it reduces the dimensionality of the predictors and at the same time there is no redundant information and correlation among the predictors, which may lead to multicollinearity. 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, and it is more generalized and useful to describe a point not by a crisp cluster, but by its membership values in all the clusters (Ross, 1997), (Gűler and Thyne, 2004). The important parameters required for fuzzy clustering algorithm are number of clusters (c) and fuzzification parameter (m). Fuzzification parameter controls the degree of fuzziness of the resulting classification, which is the degree of overlap between clusters. The minimum value of m is 1 which implies hard clustering. Number of clusters and fuzzification parameter are determined from cluster validity indices like Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE) (Roubens, 1982). FPI estimates the degree of fuzziness generated by a specified number of classes and is given by

$$FPI = 1 - \frac{cF - 1}{c - 1} \tag{8.1.1}$$

where

$$F = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T}(\mu_{it})^2 \tag{8.1.2}$$

μ_{it} is the membership in cluster i of the principal components in month t. NCE estimates the degree of disorganization created by a specified number of classes and given as

$$NCE = \frac{H}{\log c} \tag{8.1.3}$$

$$\text{where } H = \frac{1}{T}\sum_{i=1}^{c}\sum_{t=1}^{T} -\mu_{it} \times \log(\mu_{it}) \tag{8.1.4}$$

The optimum number of classes/clusters is established on the basis of minimizing these two measures given by Eqs. (8.1.1)–(8.1.3). The FPI and NCE attain their minimum values when the number of clusters is 3 for almost all cases with different m values. The value of FPI should be chosen in such a way that the resulting clustering is neither too fuzzy nor too hard. The clustering becomes non-fuzzy when FPI = 0 and fully fuzzy when FPI = 1. Guler and Thyne (2004) have recommended an FPI value of 0.25 for the

purpose of selection of number of clusters and fuzzification parameter in fuzzy clustering. In this work, FPI and NCE are plotted with number of clusters c, for different values of fuzzification parameter, m (Fig. 8.1.3). It is found that FPI value of almost 0.25 is achieved for $m = 1.4$ and $c = 3$. These values are used for fuzzy clustering. The sum of the membership of a data point in three clusters is equal to 1 and thus the membership of only two clusters will automatically fix the other and are sufficient to be used as an input to vector machine. Thus, the number of input variables used in the SVM and RVM is 12 (10 principal components along with two memberships). The following section presents the Support Vector (SV) regression used for statistical downscaling with training and testing.

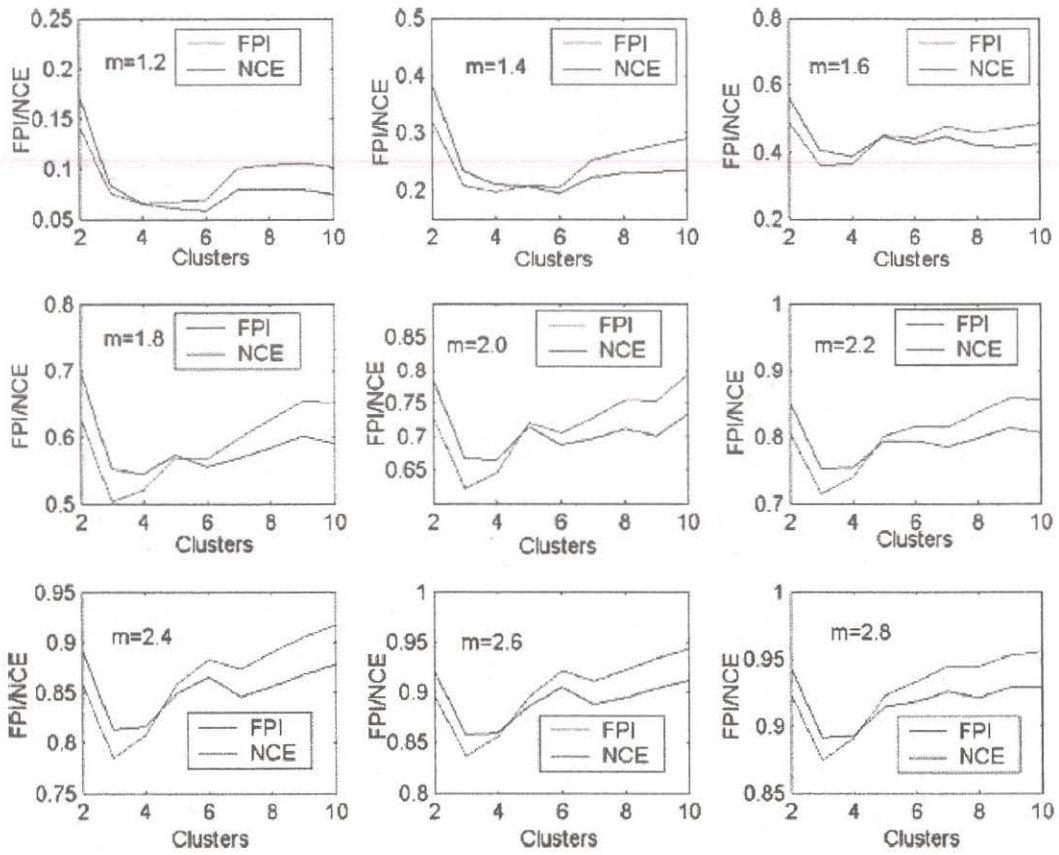

Fig. 8.1.3 Cluster validity test

8.1.2 Training and testing with support vector machine

The foundations of Support Vector Machine (SVM) have been developed by Vapnik (1995) and are gaining popularity due to many attractive features, and promising

empirical performance. The formulation embodies Structural Risk Minimization (SRM) principle, which has been proved to be superior (Gunn et al, 1997) to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. This difference equips SVM with a greater ability to generalize, which is the goal of statistical learning. A brief introduction to statistical learning with the concept of SRM may be found in Vapnik (1998) and Dibike et al. (Dibike et al, 2001).

Given a training data $\{(x_1, y_1), ..., (x_l, y_l), X \in \Re^n, Y \in \Re\}$, the SV regression equation can be given by (Smola, 1996)

$$y = f(x) = \sum_{i=1}^{l} w_i \times K(x_i, x) + b \tag{8.1.5}$$

where $K(x_i, x)$ and w_i are the kernel functions and the corresponding weights used in the SV regression. b is a constant known as bias. The ith input x_i for training is called support vector if $w_i \neq 0$ for that particular i. Naturally, in Eq. (8.1.5) the inputs other than support vectors will be vanished. The architecture of an SVM is presented in Fig. 8.1.4. The loss function considered for SVM is an ε-insensitive loss function (Fig. 8.1.5) described as

$$|\xi|_\varepsilon = |y - f(x)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x)| \leqslant \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{8.1.6}$$

The methodology for computation of weights and bias is presented in Appendix 1. Gaussian, Laplacian and heavy tailed Radial Basis Functions (RBF) are used as kernel in the present study. Details of these kernel functions are presented in Appendix 2. Selection of a suitable RBF is an important task in SVM as it has a high sensitivity on model performance (Chapelle et al, 1999). All the above mentioned kernels are used in the present analysis to compare and select the SVM regression model with the best kernel for downscaling purpose.

SVM regression models with all the kernels are trained to determine the relationship between NCEP/NCAR output of large scale climate variables and Mahanadi monsoon streamflow where the principal components, membership in fuzzy clusters and seasonal components are used as input. For training and testing K-fold cross validation (K = 10) procedure is used. According to this methodology, the training set is partitioned into K

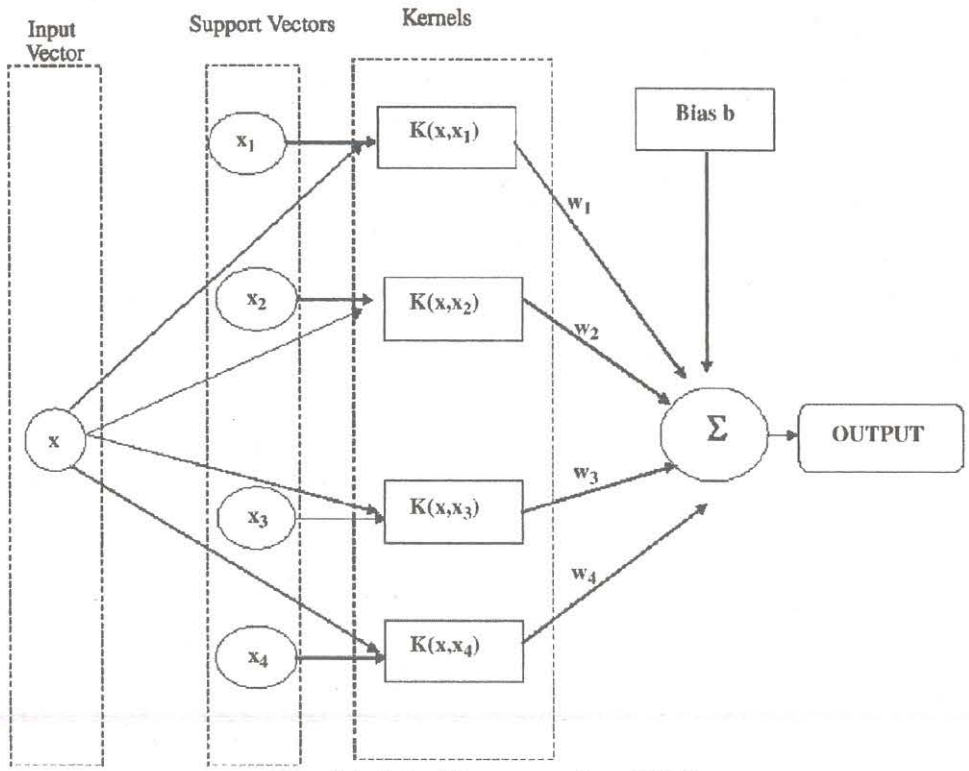

Fig. 8.1.4 Architecture of an SVM

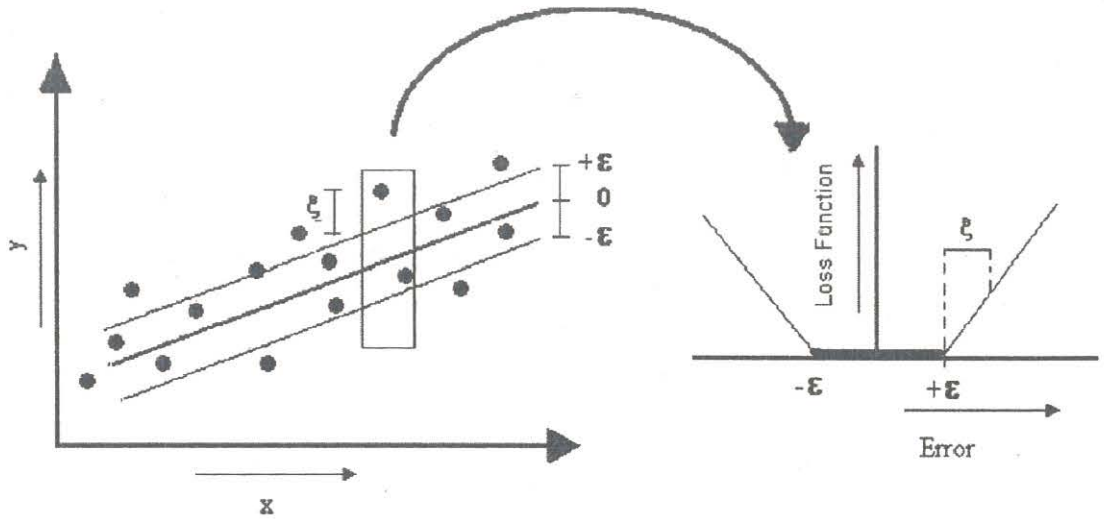

Fig. 8.1.5 ε-Insensitive loss function

8.1.3 Relevance vector machine

Despite of excellent modeling performance of SVM, it has some practical and significant drawbacks. They are (Tipping, 2001):

- Although relatively sparse, SVMs make unnecessarily liberal use of basis functions since the number of support vectors required typically grows linearly with the size of the training set. Some form of post-processing is often required to reduce computational complexity.
- Predictions are not probabilistic. In regression the SVM outputs a point estimate, whereas the conditional distribution of target given input ($p(t|x)$) is desired.
- There is no straightforward method to estimate C and ε. Sometimes cross validation is used to estimate them which is wasteful for both data and computation.
- The kernel function $K(x, x_i)$ must satisfy Mercer's condition.

Relevance Vector Machine developed by Tipping (2001) is a Bayesian treatment of Eq. (8.1.5) which does not suffer from any of the limitations stated above. In RVM, a fully probabilistic framework is adopted and introduced a priori over the model weights governed by a set of hyperparameters, associated with weights, whose most probable values are iteratively estimated from the data. Sparsity is achieved because in practice the posterior distributions of many of the weights are sharply (indeed infinitely) peaked around zero. The remaining training vectors associated with nonzero weights are termed as relevance vectors. The most compelling feature of the RVM is that, while capable of generalization performance comparable to an equivalent SVM, it typically utilizes dramatically fewer kernel functions. Following Tipping (2001), the mathematical background of RVM is presented here. RVMs have identical functional form as SVMs (Eq. (8.1.5)), but use kernel terms that correspond to fixed nonlinear basis function (Tipping, 2001).

Seeking to forecast y for given x according to $y = f(x) + \varepsilon_n$, involving weights $w = (w_0, w_1, ..., w_l)^T$, where $\varepsilon_n \approx N(0, \sigma_{\varepsilon_n}^2)$, the likelihood of the complete data set can be written as

$$p(\mathbf{y}|\mathbf{w}, \sigma_{\varepsilon_n}^2) = (2\pi\sigma_{\varepsilon_n}^2)^{-l/2} \exp\left\{-\frac{1}{2\sigma_{\varepsilon_n}^2}\|\mathbf{y} - \Phi_w\|^2\right\} \tag{8.1.7}$$

where

$\Phi(x_i) = [1, K(x_i, x1), K(x_i, x_2), \ldots, K(x_i, x_l)]^T$

Maximum likelihood estimation of w and $\sigma_{\varepsilon_n}^2$ often results in severe overfitting. Tipping (2001) suggested imposition of some prior constraints on the parameters, w, by adding a complex penalty to the likelihood or the error function. This is a prior information that controls the generalization ability of the learning system. Typically, higher-level hyperparameters are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights, w:

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^{I} N(w_i|0, \alpha_i^{-1}) \tag{8.1.8}$$

where α is a hyperparameter vector that controls how far from zero each weight is allowed to deviate. For completion of hierarchical prior specifications, hyperpriors over α and the noise variance, $\sigma_{\varepsilon_n}^2$, are defined. Consequently, using Bayes rule, the posterior overall unknowns could be computed given the defined noninformative prior distributions:

$$p(\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2) \cdot p(\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2)}{\int p(\mathbf{y}|\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2) p(\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2) \mathrm{d}\mathbf{w}\, \mathrm{d}\alpha\, \mathrm{d}\sigma_{\varepsilon_n}^2} \tag{8.1.9}$$

Computation of $p\left(w, \alpha, {\sigma_{\varepsilon_n}}^2 | y\right)$ in Eq. (8.1.9) is not possible directly as the integral in the right-hand side can not be performed. Instead the posterior can be decomposed as

$$p(\mathbf{w}, \alpha, \sigma_{\varepsilon_n}^2|\mathbf{y}) = p(\mathbf{w}|\mathbf{y}, \alpha, \sigma_{\varepsilon_n}^2) p(\alpha, \sigma_{\varepsilon_n}^2|\mathbf{y}) \tag{8.1.10}$$

The posterior distribution of the weight can be given by

$$p(\mathbf{w}|\mathbf{y}, \alpha, \sigma_{\varepsilon_n}^2) = \frac{p(\mathbf{y}|\mathbf{w}, \sigma_{\varepsilon_n}^2) \cdot p(\mathbf{w}|\alpha)}{p(\mathbf{y}|\alpha, \sigma_{\varepsilon_n}^2)} = (2\pi)^{-I/2} |\Sigma|^{-1/2} \times \exp\left\{ -\frac{1}{2} (\mathbf{w} - \mu)^{\mathrm{T}} \Sigma^{-1} (\mathbf{w} - \mu) \right\} \tag{8.1.11}$$

where the posterior co-variance and mean are respectively

$$\Sigma = (\sigma_{\varepsilon_n}^{-2} \Phi^{\mathrm{T}} \Phi + \Lambda)^{-1} \tag{8.1.12}$$

$$\mu = \sigma_{\varepsilon_n}^{-2} \Sigma \Phi^{\mathrm{T}} \mathbf{y} \tag{8.1.13}$$

with $A = diag(\alpha_0, ..., \alpha_I)$. Therefore, machine learning becomes a search for the hyperparameter posterior most probable, i.e., the maximization of $p(\alpha, \sigma_{\varepsilon_n}^2|\mathbf{y}) \propto p(y|\alpha, \sigma_{\varepsilon_n}^2)\, p(\alpha) p(\sigma_{\varepsilon_n}^2)$

with respect to α and $\sigma_{\varepsilon_n}^2$, For uniform hyperpriors, it is required to maximize the term $p\left(y \mid \alpha, \sigma_{\varepsilon_n}{}^2\right)$, which is computable and given by

$$\begin{aligned} p(\mathbf{y}|\alpha, \sigma_{\varepsilon_n}^2) &= \int p(\mathbf{y}|\mathbf{w}, \sigma_{\varepsilon_n}^2) p(\mathbf{w}|\alpha) \mathrm{d}\mathbf{w} \\ &= (2\pi)^{-l/2} |\sigma_{\varepsilon_n}^2 I + \Phi A^{-1} \Phi^{\mathrm{T}}|^{1/2} \\ &\quad \times \exp\{-\frac{1}{2} \mathbf{y}^{\mathrm{T}} (\sigma_{\varepsilon_n}^2 I + \Phi A^{-1} \Phi^{\mathrm{T}})^{-1} \mathbf{y}\} \end{aligned} \tag{8.1.14}$$

Tipping (2001) contended that all the evidence from several experiments suggests that this predictive approximation is very effective. Bayesian models refer to Eq. (8.1.9) as the marginal likelihood, and its maximization is known as the type II-maximum likelihood method (Berger, 1985). As argued by Tipping (2001), MacKay (2003) refers to this term as the evidence for hyperparameter and its maximization as the evidence procedure. Hyperparameter estimation is typically carried out with an iterative formula such as a gradient ascent on the objective function (Tipping, 2001).

At convergence of the hyperparameter estimation procedure, predictions can be made based on the posterior distribution over the weights, conditioned on the maximized most probable values of α and $\sigma_{\varepsilon_n}^2$, α_{MP} and σ_{MP}^2 respectively. The predictive distribution for a given x_* can be computed using Eq. (8.1.11):

$$p(y_*|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma_{\mathrm{MP}}^2) = \int p(y_*|\mathbf{w}, \sigma_{\mathrm{MP}}^2) p(\mathbf{w}|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma_{\mathrm{MP}}^2) \mathrm{d}\mathbf{w} \tag{8.1.15}$$

Since both terms in the integrand are Gaussian, this can be readily computed, giving

$$p(y_*|\mathbf{y}, \alpha_{\mathrm{MP}}, \sigma_{\mathrm{MP}}^2) = N(y_*|t_*, \sigma_*^2) \tag{8.1.16}$$

with

$$t_* = \mu^{\mathrm{T}} \Phi(x_*) \tag{8.1.17}$$

$$\sigma_*^2 = \sigma_{\mathrm{MP}}^2 + \Phi(x_*)^{\mathrm{T}} \Sigma \Phi(x_*) \tag{8.1.18}$$

The outcome of the optimization involved in RVM (i.e. maximization of $p\left(y \mid \alpha, \sigma_{\varepsilon_n}{}^2\right)$), is that many elements of a go to infinity such that *w* will have only a few nonzero weights that will be considered as relevant vectors. The relevant vectors (RVs) can be viewed as counterparts to support vectors (SVs) in SVMs; therefore, the resulting model enjoys the

properties of SVMs (i.e., sparsity and generalization) and, in addition, provides estimates of uncertainty bounds in the predictions they make (Khalil et al, 2006).

Table 8.1.2 Results obtained from training and testing of RVM based downscaling model

Kernel used	*R* value for training	*R* value for testing	Number of relevant vectors (% of training data set)
Gaussian RBF	0.9423	0.6019	71.30
Laplacian RBF	0.8417	0.6418	25.56
Heavy tailed RBF	0.7937	0.6998	8.06

8.1.3.1. Training and testing with RVM

Principal components and fuzzy cluster memberships derived from NCEP/NCAR reanalysis data are used as input to RVM. Similar to SVM, Gaussian RBF, Laplacian RBF and Heavy tailed RBF are used as kernels in the RVM regression model for downscaling with K-fold cross validation. The results obtained from training and testing are presented, with $\sigma = 1$, in Table 8.1.2. Compared to SVM, RVM involves very few relevant vectors for the regression with all the kernels and thus minimizing the possibility of overtraining as well as computational time. This is reflected in the differences between R values for training and testing with all the kernels. A comparatively small difference between the R value of training and testing shows the reduction of overtraining which is not achieved by SVM. Among the RVM kernels the model with heavy tailed RBF shows the highest R value for testing among all the RVM models and thus selected as the best model for downscaling. The selection of the width of the kernel is one of the major criterion in selecting the appropriate model. The kernel width can not be computed with the Bayesian treatment of RVM and therefore a post-modeling sensitivity analysis is required to compute kernel width that results in minimum overfitting. Sensitivity analysis of the training and testing R values and the number of RVs involved in the model is carried out, with variation in the kernel width, and presented in Fig. 8.1.6. As RVM involves only kernel width as a parameter, the computational effort of postmodeling sensitivity analysis is significantly less compared to SVM. It is observed that the testing R value achieved its maximum at a kernel width of 1.9, involving minimum number of

RVs. The training and testing R values are obtained as 0.7745 and 0.7256, respectively, using only 7.41% of the data set as relevant vectors. Therefore RVM using heavy tailed RBF with the width of 1.9 is used for statistical downscaling in the present study. After the selection of model the whole data set is trained using RVM based regression with heavy tailed RBF as the kernel. The overall R value is obtained as 0.8226. The observed and predicted monsoon streamflow from June 1961 to August 1990 with scatter plot are presented in Fig. 8.1.7. It is clear that even RVM is not able to mimic the extreme rainfall observed in the record. Possibly this could be because regression based statistical downscaling models often cannot explain entire variance of the downscaled variable (Wilby et al, 2004). The goodness of fit of the model is also tested with Nash–Sutcliffe coefficient (Nash and Sutcliffe, 1970), which has been recommended by ASCE Task Committee on definition of Criteria for evaluation of watershed models of the watershed management committee, Irrigation and Drainage Division (ASCE, 1993). The Nash–Sutcliffe coefficient (E) is given by

$$E = 1 - \frac{\sum_t (Q_{ot} - Q_{pt})^2}{\sum_t (Q_{ot} - \overline{Q_o})^2} \tag{8.1.19}$$

where Q_{ot} and Q_{pt} are the observed and predicted streamflow in time t, and $\overline{Q}_o$ is the mean observed streamflow. Nash–Sutcliffe coefficient can vary from 0 to 1 with 0 indicating that the model predics no better than the average of the observed data, and 1 indicating a perfect fit. It is obtained as 0.67 for the present model which is satisfactory. Wetterhall et al. (2005) have tested the long term seasonal mean, and standard deviation for verification of a downscaling model. In the present analysis also, similar test has been performed. The long term mean and standard deviation of observed streamflow are 7332.0 Mm^3 and 5995.6 Mm^3 and those of predicted streamflow are 7384.1 Mm^3 and 4607.6 Mm^3, which shows a good match in mean but difference in standard deviation. This may be because the regression based statistical downscaling models often cannot explain entire variance of the downscaled variable (Wilby et al, 2004) and therefore the present model can not mimic the high streamflow in 1961. Other limitation of the method is that assuming constant error variance (homoscedasity) may prove to be a limitation when streamflow is the response variable as it is often observed that streamflow error variance is related to the magnitude of the flow, and, the hierarchical structure cannot

accommodate Markovian dependence in the flows easily. After the verification, the RVM regression model is used for modeling of future streamflow time series from the predictor variables as projected by GCM developed by CCSR/NIES with B2 scenarios.

8.1.4 Future streamflow projection

GCM developed by Center for Climate System Research/National Institute for Environmental Studies (CCSR/NIES), Japan, with B2 scenario is used for projection of future streamflow. The grid size of the GCM is 5.5° latitude x 5.625° longitude. The monthly output for B2 scenario is extracted for CCSR/NIES GCM for the region of interest covering all the NCEP grid points extending from 13.8445°N to 30.4576°N in latitude and 78.7500°E to 95.6250°E in longitude from IPCC data distribution center. GCM grid points do not match with NCEP grid points and thus interpolation is required to obtain the GCM output at NCEP grid points. Interpolation is performed with a linear inverse square procedure using spherical distances (Willmott et al, 1985). The predictor variables for CCSR/NIES GCM are then interpolated to the 25 NCEP grid points. Using the principal directions or eigen vectors obtained from PCA of NCEP data, principal components are obtained for the GCM output.

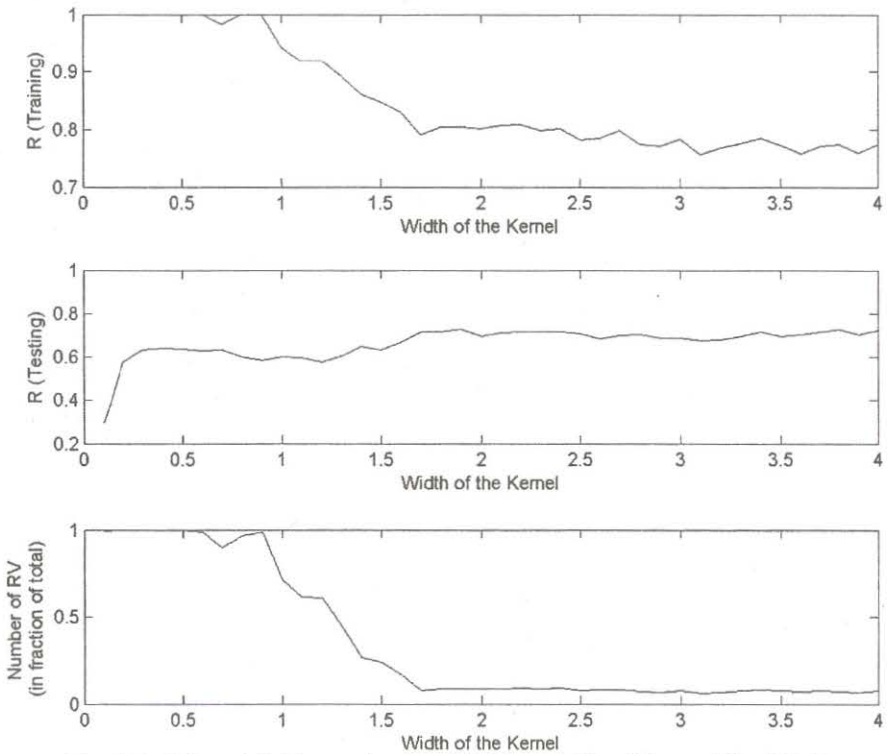

Fig. 8.1.6 Sensitivity analysis with the width of kernel in RVM

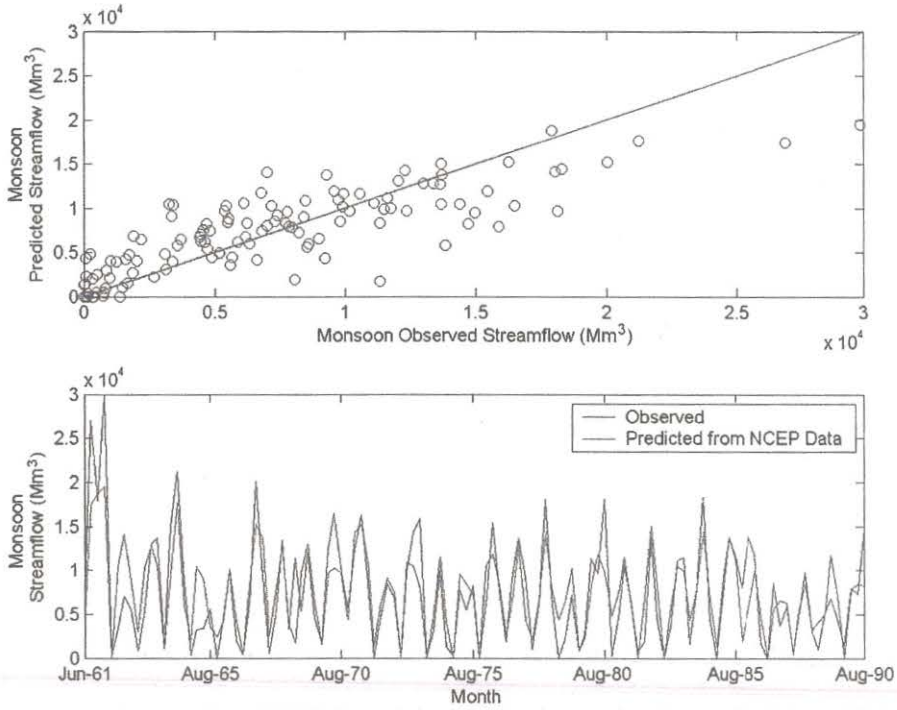

Fig. 8.1.7 Observed and RVM modeled monthly streamflow of Mahanadi river

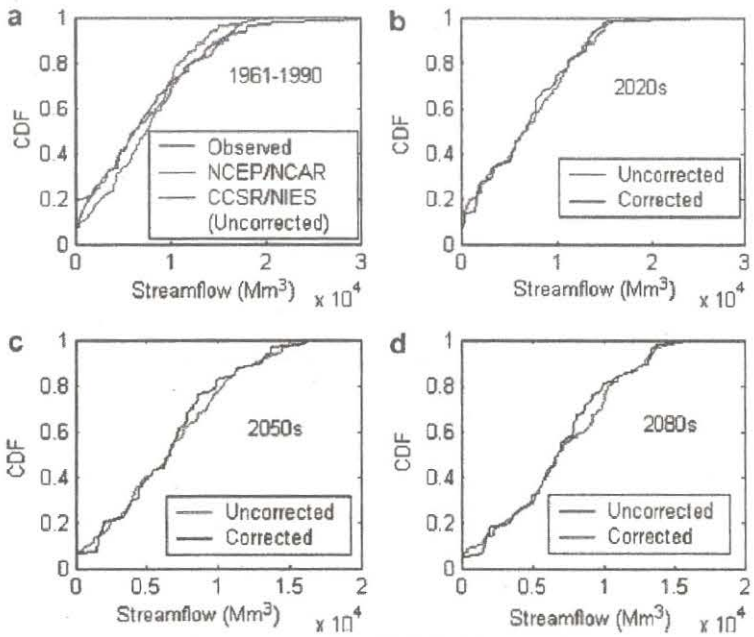

Fig. 8.1.8 Projected future streamflow for CCSR/NIES GCM with B2 scenario using RVM based downscaling

The membership of the principal components of GCM output in each of the fuzzy clusters are then computed using the cluster centers obtained from fuzzy clustering.

Principal components and cluster membership of GCM output are then used in the developed RVM regression model to project the monsoon streamflow of Mahanadi for future. For validation purpose, the monsoon streamflow is also computed for the base-line period of years 1961–1990 with the GCM output. The CDFs obtained from NCEP data, GCM output and the observed data, using Weibull's probability plotting formula, are presented in Fig. 8.1.8(a). Although the CDF obtained from GCM matches quite well, there is considerable bias near zero flow values and at the extreme cases. This is because standardization may reduce the bias in the mean and variance of the predictor variable but it is much harder to accommodate the biases in large scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter-variable relationships (Wilby and Dawson 2004). Moreover, regression based statistical downscaling models often cannot explain entire variance of the downscaled variable, which is also reflected in terms of bias near zero flow and high flow conditions. While modeling monsoon streamflow such biases should be taken care otherwise it will propagate in the computation of future seasons (Ghosh and Mujumdar, 2006a). To remove such bias from a given downscaled output the following methodology is used:

- CDFs are obtained with the downscaled GCM generated and observed streamflow for the years 1961–1990 using Weibull's probability plotting position formula.
- For a given value of GCM generated streamflow (XGCM), the value of CDF (CDFGCM) is computed.
- Corresponding to CDFGCM the observed streamflow value is obtained from the CDF of observed data.
- The GCM generated streamflow is replaced by the observed data, thus computed, having the same CDF value.
- The CDFs of GCM generated and observed streamflow, obtained for the years 1961–1990, act as reference, and based on them the correction is applied to the streamflow values obtained from GCM for future.

A major drawback of the method described above is that, if the future GCM streamflow is out of the range of historical GCM streamflow, the methodology of bias correction with Weibull's plotting position will fail. If such cases appear, then different parametric

probability distribution (with upper bound of random variable as 1) can be fitted to the observed GCM streamflow and the best pdf can be selected among them with Akaike Information Criteria (AIC) or v2-test. As the new range is now extended to 1, it is possible to perform the quantile transformation even if the future GCM streamflow is out of the range of historical GCM streamflow. In the present case, the future GCM streamflows are all within the range of observed GCM streamflow and therefore the bias is corrected with Weibull's plotting position and quantile transformation. The long term mean and standard deviation of observed streamflow are 7332.0 Mm^3 and 5995.6 Mm^3 and those of GCM projected streamflow before bias correction were 7194.2 Mm^3 and 5607.2 Mm^3. After bias correction mean and standard deviation of GCM projected streamflow are 7331.7 Mm^3 and 6009.4 Mm^3, respectively, which shows bias has been significantly reduced. The CDFs projected future streamflow is plotted for standard 30 year time slices 2020 s, 2050 s and 2080 s in Fig. 8.1.8b–d, which clearly shows a decrease in the high flows of the monsoon season in Mahanadi. The occurrence of extreme high flow events will reduce significantly and therefore there is a decreasing trend in the monthly peak flow. The projection of CCSR/NIES GCM with B2 scenario presents a favorable condition for Hirakud dam in future for flood control operation. Earlier study (Rao, 1995) on Mahanadi river also revealed decrease in monsoon streamflow for the historic period with an increasing trend in surface temperature. It is concluded in that study, that due to increase in temperature, the water yields in the river is adversely affected. Following the study, it can be inferred that one of the probable reason of such decreasing trend in streamflow may be significant increase in temperature due to climate warming.

Analysis of instrumental climate data revealed that the mean surface temperature over India has warmed at a rate of about 0.4 °C per century (Hingane et al., 1985) which is statistically significant. The increasing trend of temperature in Mahanadi river basin due to climate change is more severe. Rao (1992) found that the surface air temperature over this basin is increasing at a rate of 1.1 °C per century, which is more than double of that of entire India. Fig. 8.1.9a presents box plots of the temperature projected by CCSR/NIES with B2 scenario for historic base period (1961–1990), 2020s, 2050s and 2080s. The box plot presents the median, upper and lower quartiles and the outliers. The

middle line of the box gives the median whereas the upper and lower edges give the 75 percentile and 25 percentile of the data set, respectively. A significant increasing trend is observed in the surface air temperature. The corresponding box plots for the monsoon streamflow are presented in Fig. 8.1.9b. The result shows that although there is no significant change in the median of the monsoon flow, the occurrence of high flows will reduce significantly because of high surface warming and therefore there is a decreasing trend in the monthly peak flow. The reason may be the insensitivity of climatic variables towards low flow because of significant ground water component and therefore only the effect on high flow, which is of interest, is reflected in the results. It is worth mentioning that the projected streamflow presented, is due to a single GCM using a single scenario and it is widely acknowledged that disagreements between different GCMs over regional climate changes represents a significant source of uncertainty (Wilby and Harris, 2006). Therefore over-reliance on a single GCM could lead to inappropriate planning and adaptation responses. Thus future decision making should incorporate all the GCMs with scenarios to model the underlying GCM and scenario uncertainty.

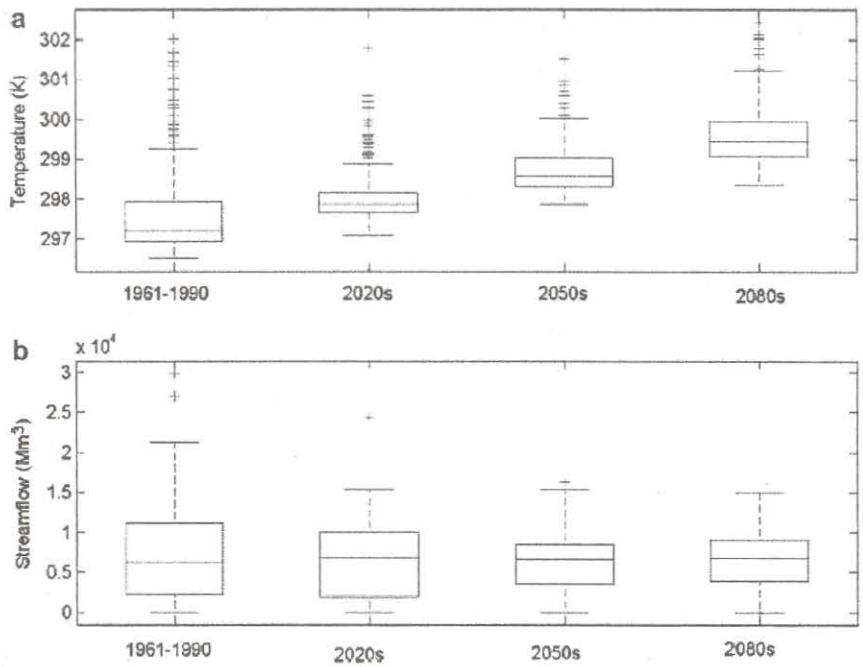

Fig. 8.1.9. (a) Box plot of projected temperature of the case-study area by CCSR/NIES GCM with B2 Scenario and (b) box plot of downscaled streamflow from CCSR/NIES GCM output with B2 Scenario.

8.1.5 Remarks

Downscaling of GCM output to monsoon streamflow is performed using SVM and RVM in the present analysis. Standardisation is performed to remove the biases present in the mean and the variances of the predictor variables. PCA and fuzzy clustering are performed prior to training to improve the model performance. As the model is a combination of classification and regression, it can be categorized into a hybrid model of weather typing and transfer function. It has been observed that RVM not only involves probabilistic reasoning but also outperforms SVM for regression based statistical downscaling in terms of goodness of fit. RVM involves fewer numbers of relevant vectors and the chance of overfitting is less than that of SVM. The model developed in the present study is capable of producing a satisfactory value of goodness of fit in terms of R value and Nash–Sutcliffe coefficient. However, from Fig. 8.1.7 it is found that even RVM is not able to mimic the extreme streamflow observed in the record. Possibly this could be because regression based statistical downscaling models often cannot explain entire variance of the downscaled variable (Wilby et al, 2004). Bias resulting from the drawback is corrected at the end of downscaling. The GCM CCSR/NIES with B2 scenario projects a decreasing trend in future monsoon streamflow of Mahanadi. In Rao (1995), a decreasing trend in the streamflow of Mahandi River with an increasing trend in surface temperature is observed, and it is concluded in that study that due to increase in temperature, the water yields in the river is adversely affected. Following the study, it can be inferred that one of the possible reasons for such a decrease in Mahanadi River streamflow may be increase in surface temperature. Such a decrease in streamflow may cause a critical situation for Hirakud dam in meeting the future irrigation and power demand. The methodology developed can be used to project the streamflow for other GCMs and scenarios also and there is a possibility of mismatch in the projections resulting GCM and scenario uncertainty. Modeling of such uncertainty is necessary for future decision making. The methodology presented, does not limit its usefulness only for modeling streamflow. It is adaptable and can be used to model any other hydrologic variable, viz. precipitation, evaporation, etc. to assess the impact of climate change on hydrology.

8.2 Uncertainties due to GCMs and Climate Scenarios in Projecting Mahanadi Streamflow

Climate change estimates on regional or local spatial scales are burdened with a considerable amount of uncertainty, stemming from several sources. Huth (2004) stated ``For estimates based on downscaling of General Circulation Model (GCM) outputs, different levels of uncertainty are related to: (i) GCM uncertainty or inetrmodel variability, (ii) scenario uncertainty or interscenario variability, (iii) different realizations of a given GCM due to parameter uncertainty (intramodel variability) and (iv) uncertainty due to downscaling methods". Uncertainty in initial conditions will also give rise to different GCM realizations. This work focuses on the first two sources of uncertainties in assessment of climate change impact on streamflow and its application to the Mahanadi basin in India. GCM uncertainty, which is due to incomplete knowledge about the underlying geophysical processes of global change, coarse grid resolutions and unresolved processes leads to limitations in the accuracy of the models. Scenario uncertainty results from unpredictability in the forecast of future socio-economic and human behavior resulting in future green house gas (GHG) emission scenarios. Downscaled outputs of a single GCM with a single climate change scenario represent a single trajectory among a number of realizations derived using various scenarios with GCMs. Such a single trajectory alone can not represent a future hydrologic scenario, and will not be useful in assessing hydrologic impacts due to climate change. Simonovic and Li (2003, 2004) have shown the uncertainty lying in climate change impact studies on flood protection resulting from selection of GCMs and scenarios. Use of several GCMs and scenarios leads to a wide spread in the downscaled hydrologic 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.

Research into probabilistic forecasts of climate change has been advancing rapidly on several fronts. New and Hulme (2000) developed a model for scenario uncertainty using Bayesian Monte-Carlo approach assuming a prior distribution of the uncertain parameters of the climate models. GCM uncertainty is presented in terms of sensitivity of climate change model outputs to streamflow. A similar methodology for sensitivity analysis and

GCM uncertainty modeled in those studies is due to the inherent bias present in the GCMs.

Ghosh and Mujumdar (2007a) have used nonparametric methods in modeling GCM and scenario uncertainty for future drought assessment in Orissa meteorological subdivision, India. Samples of a drought indicator are generated with downscaled precipitation from available GCMs and scenarios. In that study the bias has been corrected for each GCM with respect to the observed data of baseline period (years 1961-1990) and it is assumed that bias free GCM simulations are equally accurate across all GCMs and all the scenarios are equally possible. With this assumption, nonparametric methods such as kernel density estimation and orthonormal series methods are used to determine the pdf of the drought indicator. Scenario uncertainty is considered in the model by incorporating simulations of different scenarios. The information generated through the pdf of the drought indicator in a future year, can be used in long term planning decisions. A limitation in the model is that all scenarios are not available under all GCMs, and therefore, outputs of some of the scenarios for a few GCMs are missing which may lead to partial ignorance. Moreover, the set of available scenarios may not fully compose the universal sample space, Ω, which is defined to contain all possible scenarios and thus precise or conventional probability is not expressive enough for application to scenarios (Tonn, 2005). To model partial ignorance resulting from the above mentioned reasons, the methodology is further extended (Ghosh and Mujumdar, 2007b) with the concept of imprecise probability or interval probability. A normal distribution is assumed for the drought indicator for each year, with imprecision inherent in it. Uncertainty underlying in this assumption and that due to partial ignorance about future scenarios are modeled by fitting the normal distribution to drought indicator with interval regression leading to a imprecise normal distribution resulting in probabilities of events in terms of interval grey number, a number with known lower and upper bounds but unknown distribution information.

Dissimilarities between the bias-corrected GCM simulations under different scenarios after the year 1990 (end of base-line 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 analysis by Ghosh and Mujumdar

(2007a,b). An evaluation of climate change impact, in terms of quantification of change in hydrologic and climatological variables is performed with respect to the base-line period 1961-1990 (*http : //sedac.ciesin.columbia.edu/ddc/baseline/index.html*). Following this it is assumed in the present study that the impact of climate forcing will be visible after 1990, or in other words after 1990 the change in the climate and hydrologic variable will be quantified with respect to those of the base-line period. For appropriate planning and adaptation responses, with the passage of time, it is relevant to assess the effectiveness of the GCMs in best modeling climate change and also to judge which of the scenarios best represent the present situation under climate forcing. The objective of this study is to model the uncertainty in climate change derived from different GCMs and scenarios by assigning possibility distribution to different GCMs and scenarios, measured in terms of their ability in modeling climate change based on their performance in the recent past (years 1991-2005) under climate forcing. To do this, we use possibility theory, which is an uncertainty theory devoted to addressing devoted to the handling of partially inconsistent knowledge and linguistic information based on intuitions. Unlike probability, possibility is not computed from a frequency resulting from a sample, but is assigned to an event based on intuitive argumentation (Spott, 1999). In the present study, such intuition about the future hydrologic condition, is derived based on the performance of GCMs with associated scenarios. Based on such intuition, a possibility mass function is derived with possibility values assigned to the GCMs and scenarios. `possibility assigned to a GCM' is interpreted here as the possibility with which the future hydrologic variable of interest is modeled best by the downscaled output of the GCM. Similarly, `possibility assigned to a scenario' denotes the possibility with which the scenario best represents the climate forcing resulting in the change in the hydrologic variable. The possibility values thus computed are used as wights in deriving a possibilistic mean CDF (weighted CDF) of future hydrologic variable for time slices 2020s (years 2006-2035), 2050s (years 2036-2065), and 2080s (years 2066-2095). The following section presents a brief overview on data used and the methodology used in the present study.

(*http://www.mad.zmaw.de/IPCC_DDC/html/SRES_TAR/index.html*). Possibilities are assigned to GCMs and scenarios based on their performances in predicting the streamflow during years 1991-2005, when signals of climate forcing are visible. The possibilities are used as weights for deriving the possibilistic mean CDF for the three standard time slices of 2020s, 2050s and 2080s. The following subsection presents the details of case-study are and the data used.

8.2.1.1 Study Area and Observed Streamflow Data

The Mahanadi river of eastern India, rises on the Amarkantak plateau in the Eastern Ghats in central India in Chhatishgarh. It drains most of the state of Chhattisgarh, much of Orissa, and portions of Jharkhand and flows east to the Bay of Bengal. The data considered for this case-study are the inflow to the the Hirakud dam, which is located on Mahanadi river in Orissa (21.32^{o}N, 83.45^{o}E) at east coast of India (Fig. 8.1.2). The monthly inflow to Hirakud dam from 1961 to 2005, is obtained from the Department of Irrigation, Government of Orissa, India. Due to an absence of any major control structure upstream of the Hirakud reservoir, the inflow to the dam is considered as unregulated flow. The Mahanadi river is a rain-fed river with high streamflow during June to September due to monsoon rainfall, with insignificant contribution from ground water during this season. In the non-monsoon season, low rainfall results in low flow conditions, compared to which ground water component is significant. Moreover, the monsoon flows are important in Hirakud reservoir to meet the demands during the year. Thus, the monsoon streamflow is only modeled here using the atmospheric variables without considering ground water component. The monthly monsoon flow data of Mahanadi at the Hirakud reservoir from year 1961 to year 2005 is used in the analysis. Fig. 8.2.2 presents the monsoon flow of the Mahanadi river for the period 1961-2005. Box plots are plotted separately for the base line periods (1961-1990) and the recent past (1991-2005). It shows a decrease in the stream flow in the recent past with respect to that of baseline period which can be considered as an impact of "climate signal".

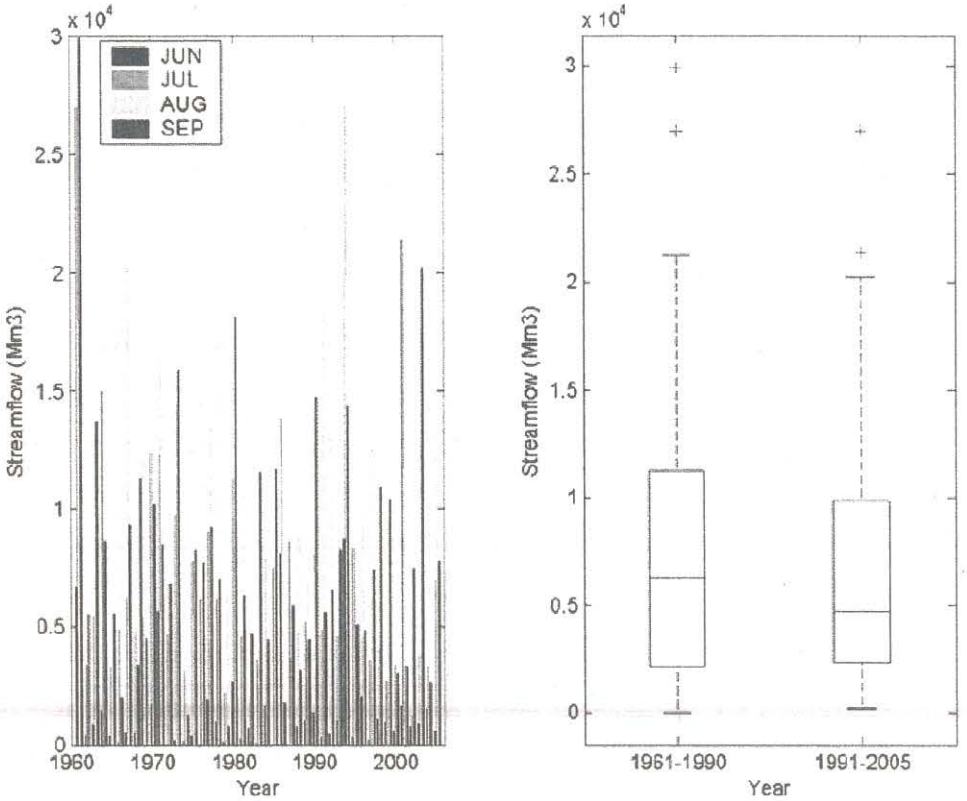

Fig. 8.2.2 Monsoon streamflow of Mahanadi river at Hirakud

8.2.1.2 Development of the Downscaling Model

The statistical downscaling model (Ghosh and Mujumdar, 2007c) used in present study consists of PCA, fuzzy clustering and relevance vector machine. Selection of the predictor is an important step in statistical downscaling. The predictors used for downscaling should be (Wilby et al., 1999, Wetterhall et al., 2005): (1) reliably simulated by GCMs, (2) readily available from archives of GCM outputs, and (3) strongly correlated with the surface variables of interest. Cannon and Whitfield (2002) have used MSLP, 500 hPa geopotential height, 800 hPa specific humidity, and 100-500 hPa thickness field as the predictors for downscaling GCM output to streamflow. Monsoon streamflow can be considered broadly as a resultant of rainfall and evaporation. Rainfall is a consequence of Mean Sea Level Pressure (MSLP)(Bardossy and Plate, 1991; Bardossy et al., 1995; Hughes and Guttorp, 1994; Wetterhall et al., 2005), geopotential height and humidity whereas evaporation is mainly influenced by temperature and humidity. Therefore, the present study considers 2m surface air temperature, MSLP,

analysis is limited only to the three GCMs (CCSR/NIES, Japan; CGCM2, Canada; and HadCM3, U.K.) and the two scenarios (IPCC TAR scenarios A2 and B2). The GCM outputs are extracted from the IPCC data distribution center (*http://www.mad.zmaw.de/IPCC_DDC/html/ddc_gcmdata.html*), for the region covering all the NCEP grid points. For baseline period 1961-1990, run for A2 and B2 scenarios are same as they are forced with the same 20th century forcing. The A2 storyline and scenario family describes a very heterogeneous world. The underlying theme is self-reliance and preservation of local identities. Fertility patterns across regions converge very slowly, which results in continuously increasing population. Economic development is primarily regionally oriented and per capita economic growth and technological change more fragmented and slower than other storylines. The B2 storyline and scenario family describes a world in which the emphasis is on local solutions to economic, social and environmental sustainability. It is a world with continuously increasing global population, at a rate lower than A2, intermediate levels of economic development, and less rapid and more diverse technological change. While the scenario is also oriented towards environmental protection and social equity, it focuses on local and regional levels. The expected increase in global temperature for the next century for scenarios A2 and B2 are nearly 3.4^0 C and 2.4^0 C (IPCC, 2000).

GCM grid points do not match with NCEP grid points and thus interpolation is required to obtain the GCM output at NCEP grid points. Interpolation is performed with a linear inverse square procedure using spherical distances (Willmott et al., 1985). For example, for the GCM developed by CCSR/NIES, Japan, the grid size is 5.5° latitude × 5.625° longitude. The ouput is extracted for the Mahanadi river basin at 16 grid points extending from 13.8445° N to 30.4576° N and 78.7500° E to 95.6250° E. These values are then interpolated to the 25 NCEP grid points. Standardization is performed after interpolation, prior to downscaling. The eigen vectors or principal directions obtained from NCEP data are used as reference to convert the gridded standardized GCM output to the corresponding principal components. Cluster memberships are computed for GCM outputs using the cluster centers obtained from NCEP/NCAR reanalysis data. The statistical relationship based on RVM developed between climatological variables and

streamflow is then applied to the principal components and cluster memberships to predict the inflow to Hirakud reservoir.

It is observed in Ghosh and Mujumdar (2007c) that even after standardisation, the bias is not significantly reduced because the methodology may reduce the bias in the mean and variance of the predictor variable but it is much harder to accommodate the biases in large-scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter-variable relationships (Wilby and Dawson, 2004). To remove such bias from a given downscaled output, for all the GCMs and scenarios, the following methodology (Ghosh and Mujumdar, 2007c) is used, which is similar to the method used by Wood et al. (2002) for removing biases from the predictors:

- CDFs are calculated for the downscaled GCM-generated and observed streamflow for the years 1961-1990 using Weibull's probability plotting position.
- For a given value of GCM-generated streamflow (X_{GCM}), the value of the CDF (CDF_{GCM}) is computed.
- The observed streamflow value is obtained from the observed CDF corresponding to CDF_{GCM}.
- The GCM-generated streamflow is replaced by this observed value.
- The CDFs of GCM-generated and observed streamflow, obtained for the years 1961-1990, act as reference, and based on these, the correction is applied to the streamflow values obtained from the GCM for future.

The correction for bias involved here is based on equi-probability transformation. From the CDFs of GCM simulated variables and observed variables for base-line period 1961-1990, the rule for the transformation (bias correction with equi-probability transformation) is derived, and then used in the future hydrologic scenarios for computation of bias free estimates of the hydrologic variable of interest. It should be noted that the assumption in this methodology for bias correction is that the bias in GCMs remain same in future. After the bias corrections the GCM projections under A2 and B2 scenarios are used for modeling GCM and scenario uncertainty.

8.2.1.4 Modeling Uncertainty with Possibility Theory

Background to Uncertainty

Modeling of GCM and scenario uncertainty necessitates use of a number of GCM outputs of different scenarios for risk based studies of future hydrologic extremes. One major assumption in modeling scenario uncertainty in most available literature (Giorgi and Mearns, 2003; Wilby and Harris, 2006; Ghosh and Mujumdar, 2007a,b) is that all scenarios are equally likely. This assumption is necessary because of ignorance about climate forcing. It is argued here that the signals of climate forcing, following the IPCC definition of base-line period (*http : //sedac.ciesin.columbia.edu/ddc/baseline/index.html*), would be visible because of global warming after the year 1990. For appropriate planning and adaptation responses, with the passage of time, it is relevant to assess the effectiveness of GCMs in modeling climate change and also to judge which of the scenarios represent the present situation best under climate forcing. A methodology based on possibility distributions is developed here to model GCM and scenario uncertainty with an objective of assignment of possibility values to GCMs and scenarios depending on their performance in modeling signals of climate forcing. As a pre-requisite, a brief overview of possibility theory is given in the following subsection.

Possibility Theory

Possibility theory, founded by Zadeh (1978), is an uncertain theory devoted to addressing incomplete information, and partially inconsistent knowledge (Dubois, 2006). It is related to the theory of fuzzy sets as a fuzzy restriction which acts as an elastic constraint on the values that may be assigned to a variable (Zadeh, 1978). More specifically, if F is a fuzzy subset of a universe of discourse $\Omega = u$ which is characterized by its membership function μ_F, then a proposition of the form "X is F", where X is a variable taking values in Ω, induces a possibility distribution Π_X which equates the possibility of X taking the value u to $\mu_F(u)$ - the compatibility of u with F. In this way, X becomes a fuzzy variable which is associated with possibility distribution Π_X in much the same way as a random variable is associated with a probability distribution (Zadeh, 1978). A main feature of possibility that distinguishes it from probability is that it is mainly ordinal and is not

generated by a GCM will not perfectly match the observed CDF, a C value of 1 is nearly impossible. Therefore, the results obtained from Eq. (8.2.6) can not be used directly as the possibility for a particular GCM and scenario, because according to the properties of possibility distribution there should be at least one scenario simulated by any of the GCMs with a possibility value 1. To satisfy the property, the results obtained from Eq. (8.2.6) for all the three GCMs and associated scenarios, are normalized by dividing the C values with the maximum value of C and the normalized value thus obtained is used as the corresponding possibility value.

8.2.2 Results and Discussion

8.2.2.1 Predicted Streamflow for 1961-1990 using Reanalysis Data

The observed and predicted (from RVM) monsoon streamflow from June 1961 to August 1990, along with the scatter plot are presented in Fig. 8.2.3. Wetterhall et al. (2005) have tested the long term seasonal mean, and standard deviation for verification of a downscaling model. In the present analysis a similar test was performed. The long term mean and standard deviation of observed streamflow are 7332.0 Mm^3 and 5995.6 Mm^3 and those of predicted streamflow are 7384.1 Mm^3 and 4607.6 Mm^3, which shows an acceptable match in central tendency (mean) but a significant difference in standard deviation. RVM based downscaling underestimates the observed high flows. One reason for this could be that the regression based statistical downscaling models often cannot explain the entire variance of the observed variable (Wilby and Dawson, 2004; Tripathi et al., 2006). The bias which is generally observed in the hydrologic variable downscaled with GCM outputs is the sum of the bias present in the downscaling model (in the RVM based statistical relationship) and in the GCM output.Both of them is adjusted at the end of downscaling using CDF matching approach (Subsection 8.2.1.3).

8.2.2.2 Predicted Streamflow using GCM data

The calibrated RVM model developed with reanalysis data is used to predict streamflow from the outputs of GCMs CCSR/NIES, CGCM2 and CSIRO-MK2 under A2 and B2 scenario. For validation of the downscaled GCM projections, the CDF obtained using

Weibulls's plotting position for the base-line period (1961-1990) with the downscaled GCM projections is plotted with that of observed streamflow (Fig. 8.2.4). In Fig. 8.2.4, CDFs of the downscaled variable derived from different GCM outputs have significant deviations from that of the observed data which suggests that bias is not completely corrected using standardization and in such a condition, if the bias is not removed, the resulting uncertainty in future will not be solely due to modeled climate change but also due to the biases present in the GCMs. The bias is removed using the methodology of equiprobability transformation presented in Subsection 8.2.1.3. The bias corrected streamflow projections with their corresponding CDFs for four time slices, 1991-2005, 2020s, 2050s and 2080s are presented in Fig. 8.2.5. 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. Another interesting feature in Fig. 8.2.5 is the increased dissimilarity between the GCMs with time. The amount of uncertainty in 2080s is higher than those of the other time slices. This may point to different climate sensitivity among the models due to ignorance about the underlying geophysical processes. Such ignorance is addressed here with possibility theory (Zadeh, 1978; Dubois, 2006).

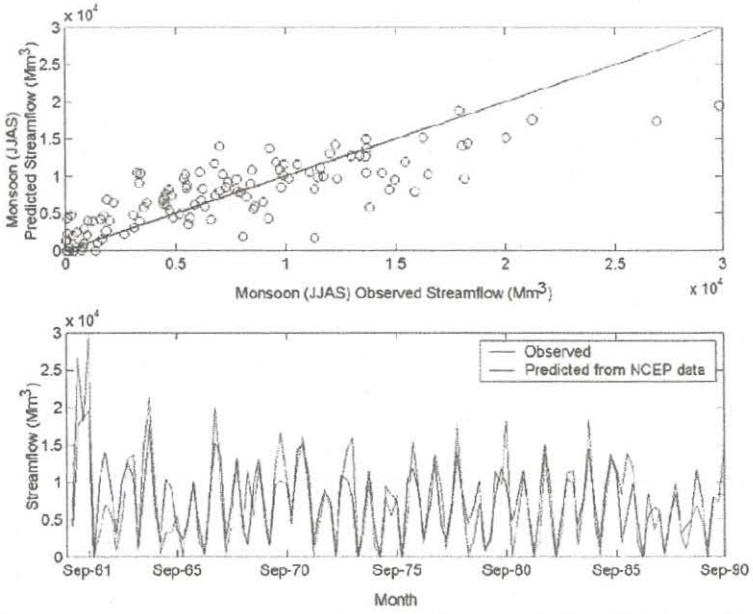

Fig. 8.2.3 Observed and predicted streamflow (JJAS) of Mahanadi river

pronounced in the initial time period (1991-2005) and therefore the results obtained by modeling climate forcing by GCMs are not significantly different from each other. With the passage of time, and with a stronger signal of climate change the possibility distribution information will be more useful in assessing which of the GCMs is able to model the climate change the best and which of the scenarios the regional or local climate is actually following. This information, however is conditional on the downscaling method used and a change in downscaling model may change the resultant possibility distribution.

Using the axioms of possibility distribution given in Eq. (8.2.3) to (8.2.5) the possibility distributions of the GCMs and scenarios are computed separately. For example, the possibility of GCM CCSR/NIES is given by:

$$\Pi(CCSR/NIES) = \Pi((CCSR/NIES, A2) \cup (CCSR/NIES, B2)) \quad (8.2.7)$$

$$= sup(\Pi(CCSR/NIES, A2), \Pi(CCSR/NIES, B2)) \quad (8.2.8)$$

Similarly the possibility of a scenario (say A2) is given by:

$$\Pi(A2) = \Pi((CCSR/NIES, A2) \cup (HadCM3, A2) \cup (CGCM2, A2)) \quad (8.2.9)$$

$$= sup(\Pi(CCSR/NIES, A2), \Pi(HadCM3, A2), \Pi(CGCM2, A2)) \quad (8.2.10)$$

The possibility distributions of GCMs and scenarios are plotted separately in Fig. 8.2.6(b) and 8.2.7(c), which show CGCM2 to be the GCM having highest possibility value with A2 as the most possible scenario for use in regional climate change impact assessment for streamflow in the Mahanadi river basin. It should be noted that projection of a hydrologic variable other than streamflow may result in a different possibility distribution for the same region. A GCM/scenario with a possibility 1 does not imply that the particular GCM/ scenario perfectly projects climate change, but in this case, it points to an ignorance of existence of any better GCMs or scenarios in modeling climate change impact on streamflow at the river basin scale. 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.

$$F_{pm} = \frac{\sum_g \sum_s \Pi(g,s) \times F_{gs}}{\sum_g \sum_s \Pi(g,s)} \quad (8.2.11)$$

where, $\Pi(g,s)$ and F_{gs} are the possibility and CDF associated with g^{th} GCM and s^{th} scenario. We also calculate the range in predictions from the GCM/scenario combinations to compare with the possibilistic mean CDF as follows. For each of the discrete streamflow values at equal intervals, maximum and minimum CDF values are obtained from the CDFs generated using the projections with three GCMs and two scenarios. The maximum and minimum CDF values are considered as upper and lower bounds of the CDF ($[F^+, F^-]$), resulting in an imprecise CDF. The interval between F^+ and F^- is known as the probability box. Without any information regarding signals of climate forcing, i.e., in absence of observed streamflow for years 1991-2005, ($[F^+, F^-]$) represents the band of imprecise CDF within which, all the CDFs generated by various GCMs and scenarios have equal possibility (all equal to 1) signifying complete ignorance about climate forcing and future scenarios. 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. 8.2.6 for years 1991-2005, 2020s, 2050s and 2080s. It is observed that the value of streamflow at which the possibilistic mean CDF reaches the value of 1 for years 2020s, 2050s and 2080s are lower than that of base-line period 1961-1990 and also reduces with time, which shows reduction in probability of occurrence of extreme high flow events in future and therefore there is likely to be a decreasing trend in the monthly peak flow. A discussion on these results is presented in the following subsection.

Table 8.2.1: Performance Measure C for the three GCMs and the two Scenarios

GCM	Scenario	C
CCSR/NIES	A2	0.8178
	B2	0.7533
HadCM3	A2	0.8743
	B2	0.9024
CGCM2	A2	0.9454
	B2	0.9327

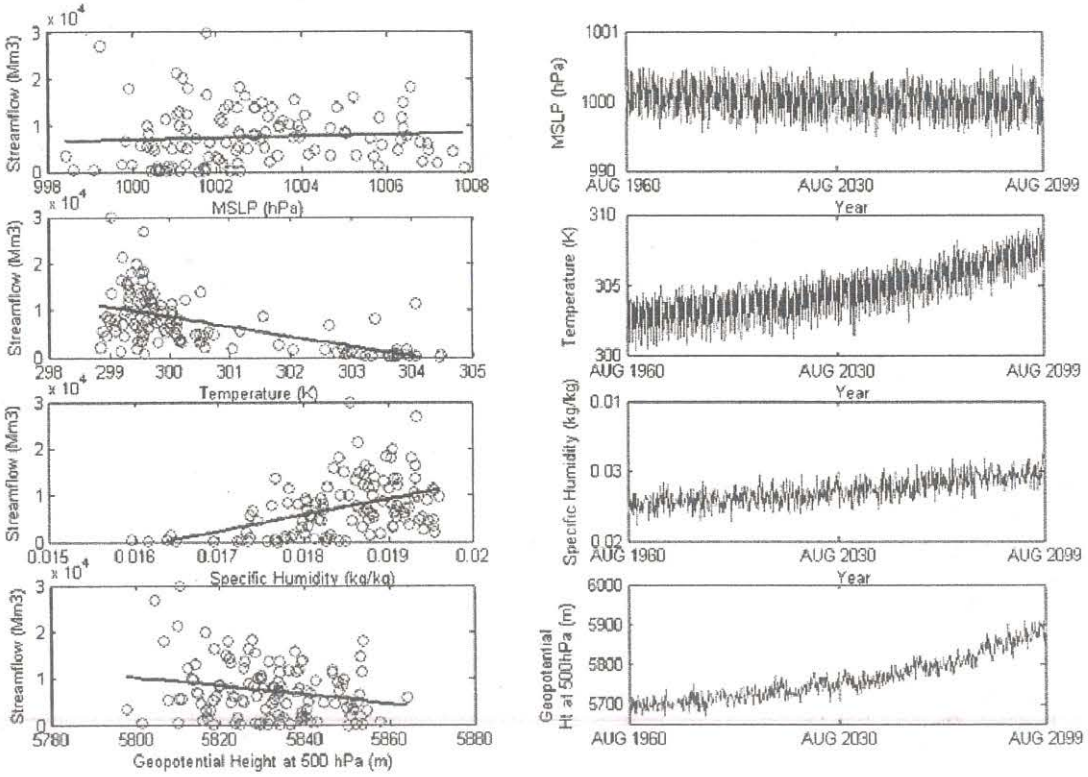

Fig. 8.2.8 Effect of variations in predictor variables on Mahanadi streamflow

Table 8.2.2: Uncertainty in Stremflow Projections

	Quantile [Streamflow, (Mm^3)]								
	2020s			2050s			2080s		
CDF value	UB CDF	LB CDF	Difference	UB CDF	LB CDF	Difference	UB CDF	LB CDF	Difference
0.25	131	2732	2601	63	3524	3461	76	4508	4433
0.50	1381	6807	5426	393	6576	6183	491	6690	6199
0.75	7329	10144	2815	1639	8623	6984	1638	9120	7482
0.90	9811	13412	3601	5584	13009	7425	3375	13070	9695
0.95	11313	15482	4169	7395	13667	6272	4675	13263	8586

UB: Upper Bound, LB: Lower Bound

Table 8.2.3: Streamflow (in Mm^3) Derived from Possibilistic Mean CDF for Years 2020s, 2050s and 2080s

CDF Value	1961-1990	2020s		2050s		2080s	
	Streamflow	Streamflow	Change*	Streamflow	Change*	Streamflow	Change*
0.25	2063	911	-55.84%	774	-62.48 %	791	-61.66%
0.50	6283	4926	-21.60%	3254	-48.21%	3180	-49.39%
0.75	11273	8480	-24.78 %	6757	-40.06 %	6018	-46.61 %
0.90	15430	12170	-21.28%	8800	-27.69%	7788	-36.01%
0.95	18148	13773	-24.11%	11350	-37.46%	9725	-46.41%

*- Change is measured with respect to the streamflow (Col.2) derived from the CDF of observed flow for the period 1961-1990.

Table 8.2.4: Effects of the Change in Predictor Variables on Streamflow

Predictors			Change in Streamflow (Mm^3)	
(GCM simulated)	1961-1990 average	2080s average	due to unit change in predictor variable	due to total change (Col3-Col2) in predictor variable
(1)	(2)	(3)	(4)	(5)
MSLP (hPa)	1000.82	999.6	186.49	-227.52
Temperature (K)	303.14	306.63	-2103.00	-7339.47
Specific Humidity (kg/kg)	0.0156	0.0190	3.44×10^{6}	11696.00
Geopotential Height (m)	5801.06	5869.22	-93.71	-6387.27
Total change in streamflow				-2258.26

Multiple CDFs derived with different GCMs and scenarios are therefore not useful in decision making and an appropriate aggregation of the ensembles resulting in a single CDF is desirable. The possibilistic mean CDF is a resultant of all the CDFs derived with different GCMs and scenarios with their associated weights. It should be noted that an arithmetic mean CDF may also serve the same purpose but it assigns equal weights to all the GCMs and the scenarios. The advantage of using possibilistic mean CDF over arithmetic mean CDF is that the possibilistic mean CDF assigns weights to GCMs and scenarios based on their performances in recent years under climate forcing. Most possible CDF (i.e., CDF with the highest possibility value) is also another option to be used in decision making, but this does not consider the projections derived with other GCMs and scenarios and at the same time it is also not guaranteed that the GCM under a scenario which performs best in the recent past of fifteen years will always perform better than other GCMs in future. In the present study, assignment of weights based on performance, resembles the Bayesian approach developed by Tebaldi et al. (2004, 2005). The differences between the two approaches are: (1) Bayesian model assigns weights to GCMs based on the bias in their simulations for baseline period, whereas the possibilistic

model corrects for the bias in GCM simulations and assigns weights based on their performance in the recent past after the baseline period when the signals of climate forcing are visible, and (2) Bayesian approach does not assign weights to scenarios and considers equal possibilities of scenarios, whereas the present model assigns different possibility values to different scenarios.

Table 8.2.3 presents the values of streamflow corresponding to possibilistic mean CDF values of 0.25, 0.5, 0.75 and 0.9 for the periods 2020s, 2050s and 2080s. The results show that the monsoon flow of Mahandi river is likely to reduce in future. The reduction of the flow is quantified with respect to the observed flow of base-line period 1961-1990. Significant changes are observed in the low flow conditions for the periods 2020s, 2050s and 2080s. For the high flow condition (flow corresponding to the CDF value of 0.95) the change is most significant for the period 2080s. An earlier study (Rao, 1995) on Mahanadi river also observed a decrease in monsoon streamflow for the historic period. One possible reason for such a decreasing trend reported in that study is the significant increase in temperature due to climate warming. Analysis of instrumental climate data has revealed that the the mean surface temperature over India has increased at a rate of about 0.4^{o} C. per century (Rao, 1995), which is statistically significant. The increasing trend of temperature in the Mahanadi river basin due to climate change is even more severe. Rao and Kumar (1992) have found that the surface air temperature over this basin is increasing at a rate of 1.1^{o} C per century, which is more than double the rate of increase for entire India. In the present study, the effects of the possible changes in predictor variables MSLP, geopotential height at 500hPa, surface specific humidity and surface temperature on the streamflow are analyzed individually and are presented in Fig. 8.2.8 for the most possible experiment, CGCM2 under A2 scenario. Significant change is not observed in streamflow due to the change of MSLP. Also there is no significant trend in the time-series of MSLP simulated by the GCM for the Mahanadi basin. The correlations of streamflow with temperature and geopotential height are negative whereas it has a positive correlation with specific humidity. The time series plots of temperature, specific humidity and geopotential height have a high increasing trend. Therefore the effect of temperature and geopotential height are negative and the effect of specific humidity is positive towards the change in monsoon streamflow of Mahanadi river.

Details of the analysis are tabulated in Table 8.2.4 with the change in the average values of predictor variables in 2080s with respect to that of base-line period. It is observed that the summation of individual effects of the predictor variables results in a net decrease in streamflow which is also reflected in Fig. 8.2.7 and Table 8.2.3. It should be noted that the analysis presented in Table 8.2.4 presents approximate change in streamflow and the possible reasons behind such change. As the correlation between the predictor variables is not considered in Table 8.2.4 and also the average of the predictors over all the GCM grid points on Mahanadi basin is considered without accounting for them individually, this analysis can not give the accurate estimates. It is however helpful in pointing out the possible reasons of decrease in streamflow. The analysis suggests that increases in temperature and geopotential height are possible reasons for decrease in streamflow. With the increase of surface temperature, the specific humidity increases but such an increase in humidity is not sufficient to nullify the effect of change in the other predictor variables. In a recent study for the same region (Orissa meteorological subdivision), Ghosh and Mujumdar (2007a, b) have also found an increasing trend of extreme meteorological drought which resembles the trend in projections of Mahanadi streamflow in the present study. Simultaneous occurrence of reduction in Mahandai streamflow and increase in extreme drought pose a major challenge for water resources engineers in meeting water demands in future.

The results presented in this work are obtained for the RVM based downscaling model and it should be noted that a change in the downscaling technique may alter the results. Use of multiple downscaling techniques in modeling downscaling uncertainty should therefore be incorporated in assessment of hydrologic impacts of climate change. A limitation of the work presented here is that the methodology does not consider the uncertainty due to the use of multiple downscaling models. Another limitation of the model is that the Third Assessment Report (TAR) data have been used in the present study which have very recently been replaced by Assessment Report 4 (AR4) data. Use of AR4 data involves substantially larger multi-model ensembles (of 17 GCMs) which may result in a more credible outcome. The difference between the possibility values for different GCMs and scenarios is very low because of the low dissimilarity between the projections simulated by different climate models for the validation period 1991-2005.

are constructed either by resampling from the observed variable distribution (conditional on the circulation pattern produced by a GCM), or by first generating synthetic sequences of weather pattern using Monte Carlo techniques and resampling from the generated data. The mean, or frequency distribution of the local climate is then derived by weighting the local climate states with the relative frequencies of the weather classes. Bardossy et al. (1995) used a fuzzy rule based technique for classification of Circulation Patterns (CP) into different states. Stochastic models such as Markov Chains may be used to predict precipitation from different states of classified circulation patterns (Bardossy and Plate 1991). The most popular approach of downscaling is the use of transfer function which is a regression based downscaling method that rely on direct quantitative relationship between the local scale climate variable (predictand) and the variables containing the large scale climate information (predictors) through some form of regression. Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. Todate, linear and non-linear regression (Wilby et al., 1998a), artificial neural network (Wilby et al., 1998b, Tripathy and Srinivas, 2005), fuzzy Rule based system (Bardossy et al., 2005), Suuport Vector Machine (Tripathy et al., 2006), analogue method (Wetterhall et al.; 2004, Gutierrez et al. 2004) etc. have been used to derive predictor-preditand relationship. A combination of classification based weather typing and transfer function method for downscaling may be found in Ghosh and Mujumdar (2006), where, Principal Component Analysis (PCA), fuzzy clustering and linear regression with seasonality term have been used for downscaling mean sea level pressure to precipitation. A completely different and unique approach of inverse modeling may be found in Cunderlik and Simonovic (2004) and Prodanovic et al. (2005), where critical meteorological situations are found from critical hydrologic events. In the final stage, the frequency of critical weather situations is investigated under future climatic conditions obtained from GCM. Since the analysis of GCM outputs is one of the last steps in this methodology, the approach allows easy updating, when new and improved GCM outputs become available. Detailed discussions on different models used for downscaling, may be found in Leavesley (1994) and Prudhomme et al. (2002).

Climate change impact assessment models developed based on GCM output are subjected to a range of uncertainties due to both 'incomplete knowledge' and 'unknowable future scenario' (Hulme and Crater, 1999; New and Hulme, 2000). 'Incomplete knowledge' mainly arises from inadequate information and understanding about the underlying geophysical process of global change, leading to limitations in the accuracy of GCMs. This can also be termed as GCM uncertainty. Uncertainty due to 'unknowable future scenario' is associated with the unpredictability in the forecast of future socio-economic and human behavior resulting in future Green House Gas (GHG) emission scenarios, and can also be termed as scenario uncertainty. Scenarios are alternative images of how the future might unfold and are an appropriate tool with which to analyze how driving forces may influence future emission outcomes and to assess the associated uncertainties. A basic assumption in the development of a scenario is that all such scenarios are equally possible in future. The choice of impact model (structure and parameterisation) is also another important source of uncertainty that is increasingly recognised. Downscaled outputs of a single GCM with a single climate change scenario represents a single trajectory among a number of realizations derived using various scenarios with GCMs. Such a single trajectory alone, therefore, can not represent a future hydrologic scenario, and will not be useful in assessing hydrologic impacts due to climate change. No quantified probability is attached to the simulated outcome of a single GCM for a single scenario and thus the approach of downscaling a single GCM output is not particularly useful for risk adaption studies (New and Hulme, 2000). In Benestad (2004), regional temperature scenarios were presented for northern Europe in the form of probability distributions, based on spatially interpolated empirically downscaled trends, derived using a multi-model ensemble as well as various downscaling options and it was found that spatial warming rate patterns, derived from the individual models exhibit large differences. Simonovic and Li (2003, 2004) have shown the uncertainty lying in climate change impact studies on flood protection resulting from selection of GCMs and scenarios. Available GCM outputs have been used for assessing effectiveness of flood protection system by them and it has been concluded that different GCMs provide different estimates of the hydrologic parameters. Using several climate change scenarios

with several GCMs provides the user of the impact study with a range of possible outcomes, but again with no attached probabilities (New and Hulme, 2000).

New and Hulme (2000), developed a model for scenario uncertainty using Bayesian Monte-Carlo approach assuming a prior distribution of the uncertain parameters of the climate models. GCM and scenario uncertainty is presented in terms of sensitivity of climate change model outputs to streamflow. Similar methodology for sensitivity analysis and risk assessment of irrigation demand may be found in Jones (2000). A simple probabilistic energy balance model, that samples uncertainty in greenhouse gas emissions, the climate sensitivity, the carbon cycle, ocean mixing, and aerosol forcing, has been used by Dessai et al. (2005), to quantify uncertainty in regional climate change projections. Assignment of global mean temperature probabilities in GCMs through pattern-scaling technique has been suggested in that study. In order to combine the resulting probabilities, regional skill scores for each GCM, season, and climate variable (surface temperature, and precipitation) are devised in 23 world regions, based on model performance and model convergence. A range of sensitivity experiments are carried out with different skill score schemes, climate sensitivities, and emission scenarios for performing sensitivity analysis of regional climate change probabilities.

The above-mentioned literature on modeling GCM and scenario uncertainty limit themselves in representing uncertainty by performing sensitivity analysis of hydrologic events to climatic parameters. However, implications of such uncertainty in estimating the severity of future extreme events, such as floods and droughts, with a probabilistic approach, has not been addressed there. Research into probabilistic forecasts of climate change has been advancing rapidly on several fronts. For example, there have been systematic evaluations of uncertainties due to climate model projections using multimodel ensembles (Raisanen and Palmer, 2001; Giorgi and Mearns, 2003); multiensemble experiments with one GCM (Murphy et al., 2004). Bayesian methods have been applied to multimodel ensembles to characterize uncertainty and probability distribution functions (PDFs) for future climate changes at regional scales (Tebaldi et al., 2004, 2005). In a more recent study Wilby and Harris (2006) developed a probabilistic framework for modeling GCM and scenario uncertainty, where, GCMs were weighted according to an index of reliability for downscaled effective rainfall. A Monte Carlo

approach was then used to explore components of uncertainty affecting projections for the River Thames by the 2080s. It was found that the resulting cumulative distribution functions (CDFs) of low flows were most sensitive to uncertainty in the climate change scenarios and downscaling of different GCMs.

The present study attempts to answer the specific question of interpreting the available outputs from GCMs with different scenarios in assessing the severity of future drought, addressing both GCM and scenario uncertainty. Uncertainty due to structure and parametrization is not considered in this work, to keep the focus of the work on modeling GCM and scenario uncertainty. An overview of the methodology proposed in this work is presented in Fig. 8.3.1. Fuzzy clustering based downscaling (Ghosh and Mujumdar, 2006) is used for modeling future precipitation using circulation pattern, projected with the available GCM outputs. 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 nonparametric pdf of SPI, as it is very unlikely that the small sample of available GCM outputs will follow a particular parametric distribution. Probabilities for different categories of future drought are computed from the estimated pdf. The methodology is applied to the case study of Orissa meteorological subdivision in India to analyze the severity of different degrees of drought in future.

The following sub-section presents details of the case study area, data extraction and downscaling technique used for the analysis.

8.3.1 Data Extraction and Statistical Downscaling

The Orissa meteorological subdivision located on the eastern coast of India, extended from 17° N to 22° N in latitude, and 82° E to 87° E in longitude. The monthly area weighted precipitation data of Orissa meteorological subdivision in India, from January, 1950 to December 2002, is obtained from Indian Institute of Tropical Meteorology, Pune (*http://www.tropmet.res.in*). This data set is used in the downscaling as predictand. Primary source of this data is the India Meteorological Department (IMD). Selection of predictor is an important step in statistical downscaling.

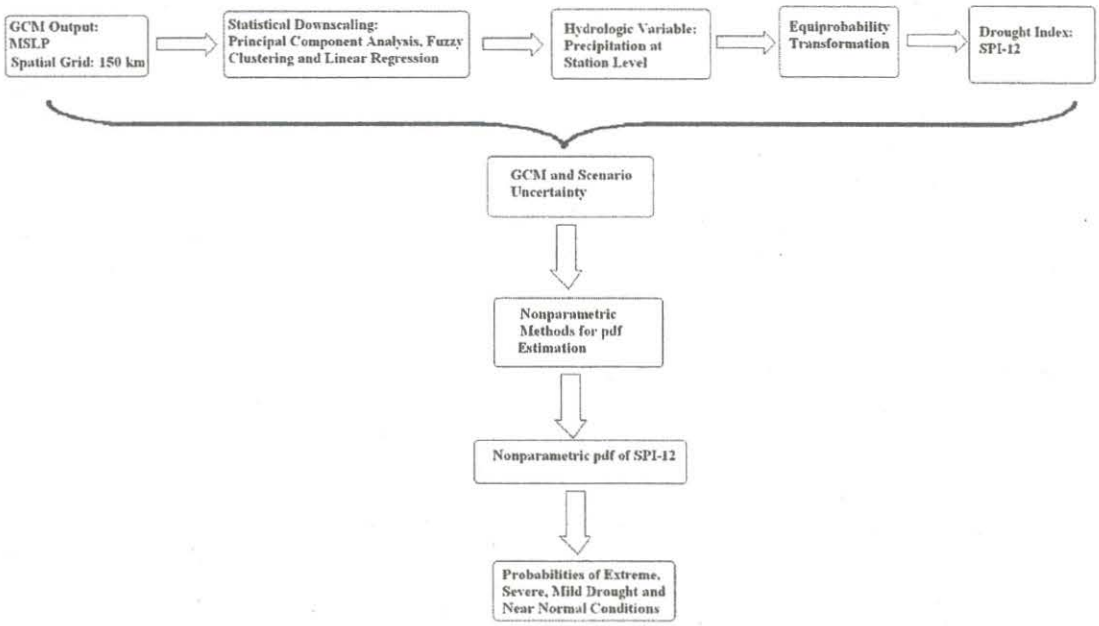

Fig 8.3.1 Overview of the Method

The predictors, used for downscaling (Wilby et al., 1999; Wetterhall et al., 2005; Tripathi et al., 2006) should be: (1) reliably simulated by GCMs, (2) readily available from archives of GCM outputs, and (3) strongly correlated with the surface variables of interest. Precipitation can be related to air mass transport and thus related to atmospheric circulation, which is a consequence of pressure differences and anomalies (Bardossy, 1997) and thus circulation pattern is used as the predictor for downscaling in most of the earlier models (e.g., Bardossy and Plate, 1991; Hughes and Guttorp, 1994; Bardossy et al., 1995; Wetterhall et al., 2005). Based on these studies, the present methodology uses Mean Sea Level Pressure (MSLP) as predictor for downscaling. 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; *http : //www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml*). Reanalysis data are outputs from a high resolution atmospheric model that has been run using data assimilated from surface observation stations, upper-air stations, and satellite-observing platforms. Results obtained using these fields therefore represent those that could be expected from an ideal GCM (Cannon and Whitfield, 2002). Monthly average MSLP from 1948 to 2002 were obtained for a region spanning $15^0 N - 25^0 N$ in latitude and

$80^0 E - 90^0 E$ in longitude that encapsules the study region. Fig. 8.1.2 shows the NCEP grid points superposed on the map of Orissa meteorological subdivision. A statistical relationship based on fuzzy clustering and linear regression is developed between MSLP and precipitation with reanalysis data of MSLP as predictor and observed precipitation as predictand. This relationship is used to model the future precipitation using available GCM projections of MSLP. Table 8.3.1 gives a list of GCMs with available scenarios. The output of MSLP of GCMs with scenarios, as given in Table 8.3.1, are extracted from the IPCC data distribution center, for the region covering all the NCEP grid points (*http://www.mad.zmaw.de/IPCC_DDC/html/ddc_gcmdata.html*).

An overview of the statistical downscaling technique used here to model future precipitation from GCM projected circulation pattern is presented in Fig. 8.3.2. The method involves training NCEP data of circulation pattern with observed precipitation and use of the resulting regression relationship in modeling future precipitation from GCM projections. The training involves three steps (Ghosh and Mujumdar, 2006): Principal Component Analysis (PCA), fuzzy clustering and linear regression with seasonality terms. 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 predefined baseline period for both NCEP/NCAR and GCM output. The period 1961-1990 is used as a base-line because it is of sufficient duration to establish a reliable climatology, yet not to long, nor to contemporary to include a strong global change signal (Wilby et al., 2004). For Orissa meteorological subdivision, MSLP values at 25 NCEP grid points are used as predictor which are highly correlated with each other. PCA is used to convert them into a set of uncorrelated variables. It was found that 99.7% of the variability of original data set is explained by the first 3 principal components and therefore only the first three principal components are used for modeling precipitation. 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, and it is more generalized and useful to describe a point not by a crisp cluster, but by its membership values in all the clusters.

The important parameters required for the fuzzy clustering algorithm are the number clusters (c) and the fuzzification parameter (m). The fuzzification parameter controls the degree of the fuzziness of the resulting classification, which is the degree of overlap between clusters. The minimum value of m is 1 which implies hard clustering. Number of clusters and fuzzification parameter are determined from cluster validity indices like Fuzziness Performance Index (FPI) and Normalized Classification Entropy (NCE) (Roubens, 1982). FPI estimates the degree of fuzziness generated by a specified number of classes and given by:

$$FPI = 1 - \frac{cF - 1}{c - 1} \tag{8.3.1}$$

where,

$$F = \frac{1}{T} \sum_{i=1}^{c} \sum_{t=1}^{T} (\mu_{it})^2 \tag{8.3.2}$$

μ_{it} is the membership in cluster i of the principal components in time t. NCE estimates the degree of disorganization created by a specified number of classes and given as:

$$NCE = \frac{H}{logc} \tag{8.3.3}$$

where,

$$H = \frac{1}{T} \sum_{i=1}^{c} \sum_{t=1}^{T} -\mu_{it} \times log(\mu_{it}) \tag{8.3.4}$$

The optimum number of classes/ clusters is established on the basis of minimizing these two measures. The clustering becomes non-fuzzy when FPI=0 and fully fuzzy when FPI=1 (Guler and Thyne 2004). The value of FPI should be chosen in such a way that the resulting clustering is neither too fuzzy nor too hard. Guler and Thyne (2004) have recommended a FPI value of 0.25 for the purpose of selection of number of clusters and fuzzification parameter in fuzzy clustering. In this work, FPI and NCE are plotted with number of clusters c, for different values of fuzzification parameter, m (Fig. 8.3.3). It is found that FPI value of 0.25 is achieved for m=2.0 and c=2. Ross (1997) also recommends a default value of m 2.0. Therefore, the number of clusters is selected as 2 and in clustering algorithm the value of m is considered as 2.0.

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 & principal components as regressors. An appropriate seasonality term is used to capture the seasonality. The linear regression equation is given by:

$$P_t = C + \sum_{i=1}^{I-1} \beta_i \times \mu_{it} + \sum_{k=1}^{K} \gamma_k \times pc_{kt} + \sum_{i=1}^{I-1} \sum_{k=1}^{K} \rho_{ik} \times \mu_{it} \times pc_{kt} \qquad (8.3.5)$$

with

$$C = C^0 + C^1 \times \sin(2\pi p/12) + C^2 \times \cos(2\pi p/12) \qquad (8.3.6)$$

$$\beta_i = \beta_i^0 + \beta_i^1 \times \sin(2\pi p/12) + \beta_i^2 \times \cos(2\pi p/12) \qquad (8.3.7)$$

$$\gamma_k = \gamma_k^0 + \gamma_k^1 \times \sin(2\pi p/12) + \gamma_k^2 \times \cos(2\pi p/12) \qquad (8.3.8)$$

$$\rho_{ik} = \rho_{ik}^0 + \rho_{ik}^1 \times \sin(2\pi p/12) + \rho_{ik}^2 \times \cos(2\pi p/12) \qquad (8.3.9)$$

where, P_t is the precipitation in time t, pc_{kt} is the k^{th} principal component of circulation pattern in time t, and μ_{it} is the membership in cluster i of the principal components in time t. K and I are the number of principal components used and number of clusters respectively. β_i, γ_k and ρ_{ik} are the coefficients of μ_{it}, pc_{kt} and their product terms, respectively. C is the constant term used in the equation. The membership values μ_{it} in each cluster are assigned to different points based on fuzzy c-means algorithm. These membership values lie between 0 and 1. $(I-1)$ number of clusters are adequate to model the regression equation as the sum of the membership values in all the clusters at time t is one and thus $(I-1)$ memberships will automatically fix the value of I^{th} membership and therefore the I^{th} membership will be a redundant input variable to the regression model. Seasonality is incorporated by Eqs. (8.3.6-8.3.9), where, p is the serial number of the month within a year $(p = 1,2,3,\ldots,12)$. Correlation Coefficient (r) between the observed and predicted precipitation is considered as the goodness of fit of the regression model. Here the r value is obtained as 0.924. Linear regression without fuzzy clustering, i.e., only with the principal components obtained from NCEP reanalysis data of MSLP, results in a lower r value of 0.803, which shows the importance of fuzzy clustering in the improvement of downscaling model fit.

period (June, July, August and September) the long term mean and median of observed precipitation are 281.4 mm/month and 281.9 mm/month respectively and those of predicted precipitation are 281.5mm/month and 283.3mm/month, which shows a good match. Similar results are also obtained for dry period. For dry period (months other than June, July, August and September) the long term mean and median of observed precipitation are 74.9 mm/month and 73.8 mm/month respectively and those of predicted precipitation are 74.3 mm/month and 73.6 mm/month. After this verification the model (Eqs. 8.3.5-8.3.9) is used for modeling of future precipitation time series for different GCMs with different scenarios.

Table 8.3.1: GCMs Used and Available Scenarios

GCM	Organization	Scenarios Available
CCSR/NIES Coupled GCM	Center for Climate Research Studies (CCSR) and National Institute for Environmental Studies (NIES), Japan	A1, A2, B1, B2
Second Generation Coupled Global Climate Model (CGCM2)	Canadian Center for Climate Modelling and Analysis, Canada	IS92a, A2, B2
HadCM3	Hadley Centre for Climate Prediction and Research (HCCPR), UK	IS95a, (GHG+ Ozone+Sulphate), A2
ECHAM4/OPYC3	Max Planck Institute für Meteorologie, Germany.	IS92a, A2, B2
CSIRO-MK2	Australia's Commonwealth Scientific and Research Organisation (CSIRO)	(IS92a+Sulph), IS92a, A1, A2, B1, B2

Table 8.3.2 Bias of Downscaled Observed Precipitation Relative to Observed Data

Mean of Observed Annual Precipitation (1961–1990), mm	Mean of NCEP Downscaled Annual Precipitation (1961–1990), mm	GCM	Mean of GCM Downscaled Annual Precipitation (1961–1990), mm	Bias = Observed Mean Annual Precipitation – Downscaled Mean Annual Precipitation, mm
1394.2	1441.6	CGCM2	1558.3	–164.1
		CCSR/NIES	1542.2	–148.0
		HadCM3	1121.4	272.8
		ECHAM4/OPYC3	1387.8	6.4
		CSIRO-MK2	1322.6	71.6

GCM grid points do not match with NCEP grid points and thus interpolation is required to obtain the GCM output at NCEP grid points. Interpolation is performed with a linear inverse square procedure using spherical distances (Willmott et al., 1985). For example, for GCM developed by CCSR/NIES, Japan, the grid size is 5.5° latitude × 5.625° longitude. The MSLP ouput is extracted for Orissa meteorological subdivision at 16 grid points extending from 13.8445° N to 30.4576° N in latitude and 78.7500° E to 95.6250°

E in longitude. These MSLP values are then interpolated to the 25 NCEP grid points. Statistical relationship (Eqs. 8.3.5-8.3.9) obtained between MSLP and precipitation is then applied to to these interpolated NCEP gridded GCM output to model precipitation.Similar to CCSR/NIES, all other GCMs as given in Table 8.3.1, are used to simulate the precipitation for historic period and to project future precipitation of Orissa meteorological subdivision. The eigen vectors or principal directions obtained from NCEP data are used as reference to convert the gridded standardized GCM output to the corresponding principal components. Therefore weights, or principal directions/ eigen vectors are not calculated separately with the output of each of the GCMs, rather the same reference principal directions or eigen vectors as obtained from NCEP output are used for all of the GCMs. Another alternative approach may be to blend of GCM outputs with the NCEP re-analysis data for obtaining a universal set of principal components. For validation the bias of annual mean of precipitation as downscaled from differnet standardised GCM output relative to observed data for base-line period is presented in Table 8.3.2. It is seen that even after standardisation, the bias is not significantly reduced because the methodology may reduce the bias in the mean and variance of the predictor variable but it is much harder to accommodate the biases in large-scale patterns of atmospheric circulation in GCMs (e.g. shifts in the dominant storm track relative to observed data) or unrealistic inter-variable relationships (Wilby and Dawson, 2004). Discussion on biases of differnet GCMs after downscaling may also be found in Wilby and Harris (2006). To remove the biases, the mean of 1961-1990 simulated mean is subtracted and 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. Fig. 8.3.4 shows the future projection of precipitation for wet (June, July, August and September) and dry period separately for CCSR/NIES GCM with B2 scenario. It is observed that 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 modeling 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

listed in Table 8.3.1, outputs of moisture content or humidity are not available. Fig 8.3.4 clearly indicates a slight increase in wet period precipitation and severe decrease in dry period precipitation, for the particular scenario. The precipitation thus computed for all the GCMs with scenarios are converted into suitable drought indicator for examining future drought scenario.

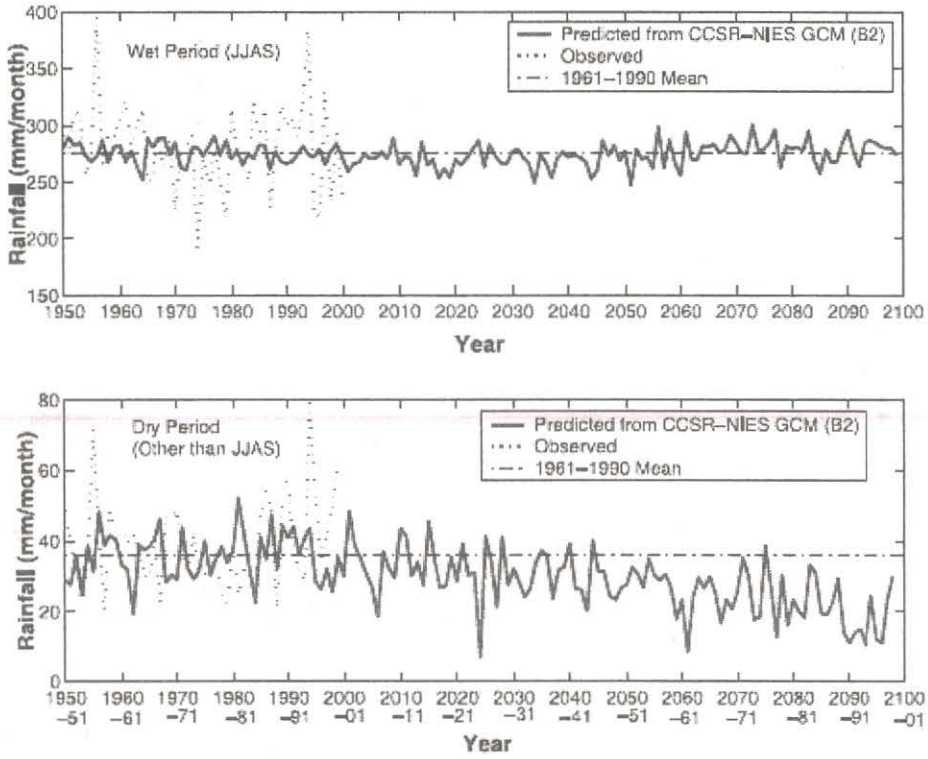

Fig 8.3.4 Rainfall for Wet and Dry Period with CCSR/NIES-B2 Projection

8.3.2 Drought Indicators

A drought indicator, briefly defined, is a variable to identify and assess drought conditions (Steinemann, 2003). Common indicators are based on meteorologic and hydrologic variables such as precipitation, streamflow, soil moisture, reservoir storage, and ground water levels. A drought trigger is a threshold value of the drought indicator that distinguishes a drought category, and determines when drought response actions should begin or end. Drought categories typically represent levels of severity, such as "mild, moderate, severe, or extreme drought. Commonly used drought indicators include Standardized Precipitation Index (SPI), Palmer Drought Severity Index (PDSI), Crop

Moisture Index (CMI), Surface Water Supply Index (SWSI), Reclamation Drought Index (RDI), and Deciles. (*http : //www.drought.unl.edu/whatis/indices.htm*).

Most of the drought indicators stated above require multiple input data such as precipitation, available water content of soil, temperature, snowpack, reservoir storage etc. The Standardized Precipitation Index (SPI) is the simplest one which requires only precipitation as input and is generally computed for 3, 6, 12, 48 months, with notation of SPI-3, SPI-6, SPI-12, SPI-48, respectively. Because of the computational simplicity and least input requirement SPI is used for drought assessment in the present work. The analysis is performed for annual drought and thus SPI-12 ia used for examining the drought scenario. A brief overview of SPI is given in the following subsection.

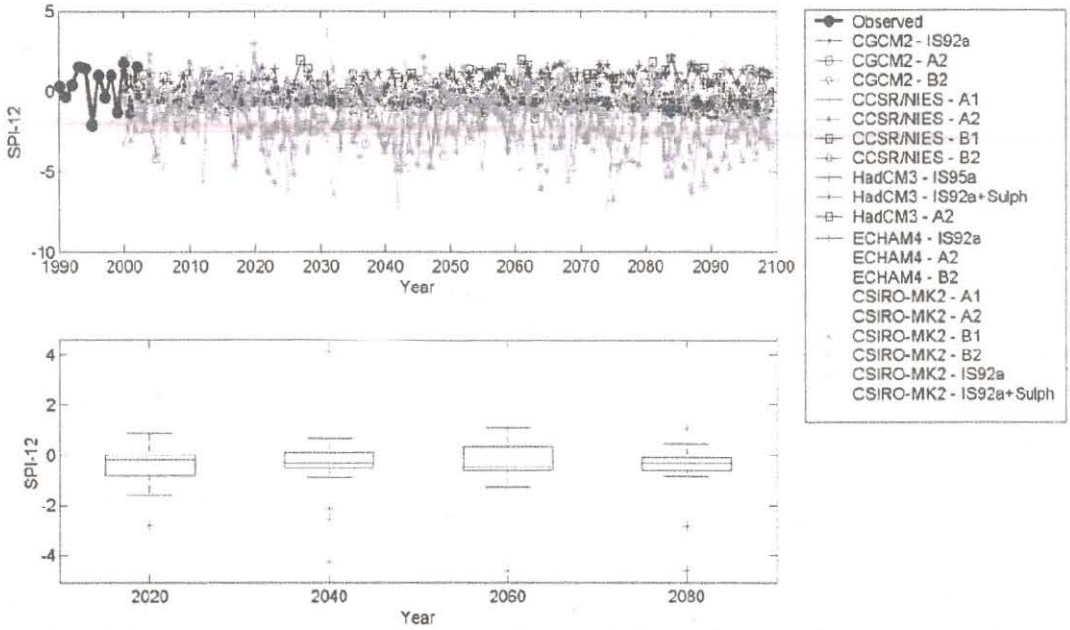

Fig 8.3.5 Predicted SPI-12 from GCM Projections with Different Scenarios. The Lower Figure gives the Box Plots for Four Years.

8.3.2.1 Standardized Precipitation Index

McKee et al. (1993) developed the Standardized Precipitation Index (SPI) for the purpose of defining and monitoring drought. SPI is based on the probability distribution of precipitation and requires only precipitation as input data. SPI can be defined by the value of standard normal deviate corresponding to the Cumulative Distribution Function (CDF) value of a precipitation event with a known probability distribution. A common

procedure adopted for computing SPI is to fit a gamma distribution to the precipitation data, although the Pearson III has also been recommended, and then to transform the data to an equivalent SPI value based on the standard normal distribution (Steinemann, 2003). Details of the methodology for calculation of SPI may be found in the website of Colorado Climate Center, Colorado State University (*http : //ccc.atmos.colostate.edu/pub/spi.pdf*). The standard procedure is as follows:

1. Fit a Gamma distribution to the time series of non-zero precipitation for each time scale of interest (e.g., 3 months, 12 months, 24 months, 48 months, etc.) without overlapping of data segments. Compute the parameters of the Gamma distribution.
2. Compute the value of CDF (G(x)) corresponding to each value of non-zero precipitation (x).
3. Compute the zero precipitation probability (q) from the historical time series. The value of CDF (H(x)) for a specific precipitation (x) will be:

$$H(x) = q + (1-q) \times G(x) \qquad (8.3.11)$$

4. Compute the value of standard normal deviate corresponding to the value of CDF (H(x)). This is the SPI value for the precipitation(x).

Based on the value, the severity of drought can be assessed and categorized into different classes. Table 8.3.3 presents the categories of drought corresponding to their SPI values (McKee et al., 1993; Steinemann, 2003).

The parameters required for estimation of SPI, viz., parameters of Gamma distribution and non zero precipitation probability, are estimated based on the observed annual precipitation by fitting it to Gamma distribution. Using these parameters, the future annual precipitation (computed from monthly precipitation), downscaled from GCM output, is converted into SPI-12. The SPI-12 is calculated for all GCMs for available scenarios. The projected SPI-12 thus computed in Fig. 8.3.5, which shows that SPI-12 time series downscaled from one GCM is entirely different from that of another and also a considerable dissimilarity exists among two scenarios of any particular GCM. The box plot presented in Fig. 8.3.5 presents the sparseness of the SPI-12 values computed from different GCMs with scenarios for the years 2020, 2040, 2060 and 2080. A single time series of SPI-12 generated from a GCM for a particular scenario represents a single trajectory among a number of realizations derived using various scenarios with GCMs

and can not by itself represent the future drought condition . Such uncertainty due to GCMs and scenarios are modeled in a probabilistic framework to assess the severity of possible droughts in future.

Table 8.3.3: Drought Categories

Drought Category	SPI values
Near Normal	0 to -0.99
Mild to Moderate Drought	-1.00 to -1.49
Severe Drought	-1.50 to -1.99
Extreme Drought	-2.00 or less

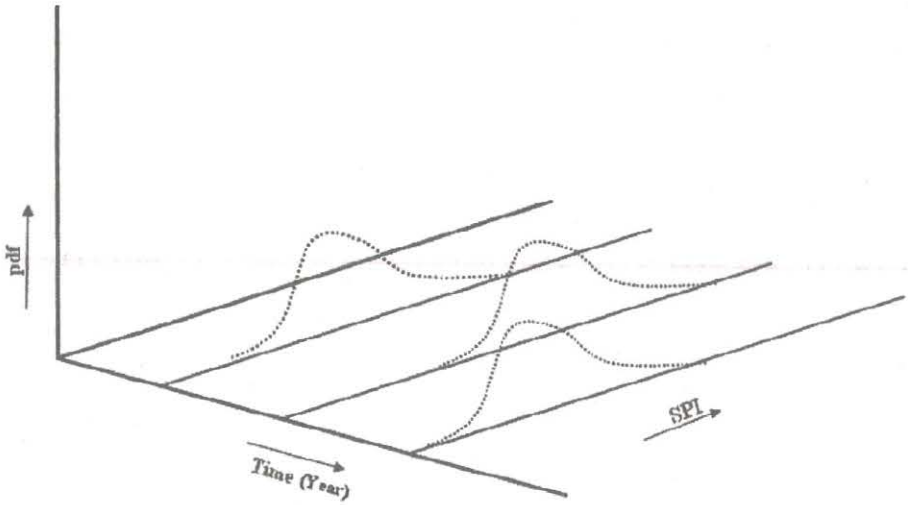

Fig 8.3.6 Pdf of SPI-12 at Each Time Step

8.3.3 Modeling GCM and Scenario Uncertainty

Climate change impact studies on hydrology, based on GCMs, are characterized by GCM and Scenario uncertainty. The source of GCM uncertainty lies in inadequate information and understanding about the underlying geophysical process of global change leading to varied assumptions and limitations in GCM outputs. Unpredictability in the forecast of future socio-economic and human behavior resulting different Green House Gas (GHG) emission scenarios leads to scenario uncertainty. Modeling of GCM and scenario uncertainty necessitates use of a number of GCM outputs of different scenarios for risk based studies of future hydrologic extremes.

In the present work the SPI-12 values computed with downscaled outputs from GCMs, are considered as the realizations of the random variable SPI-12 in each year where there exists a pdf of SPI-12 in each year (Fig. 8.3.6). The severity of future drought may be

studied by estimating the evolution of the pdf of a drought indicator. The simplest methodology of such analysis is based on the assumption of normal distribution for future SPI-12 in each year. However, it is very unlikely that SPI-12 will follow a normal distribution and thus such analysis may lead to erroneous conclusion. In such cases nonparametric pdfs estimated by a kernel density function with a suitable smoothing parameter are useful, as prior assumption of the data to follow a particular distribution can be avoided (Lall et al., 1993). Applications of kernel density estimation for determination of pdf for hydrologic variables may be found in Lall (1995), Lall et al. (1996), Sharma et al. (1997), and Tarboton et al. (1998). Small sample size, however, may not result in accurate estimation of nonparametric pdf, using kernel function. The methodology based on orthonormal series (Efromovich, 1999) for determination of nonparametric pdf from a small sample, may be used to overcome this drawback. Here, we discuss the use of all the three methods (viz., use of a Normal distribution, kernel density estimation and orthonormal series) for examining implications on future drought scenarios.

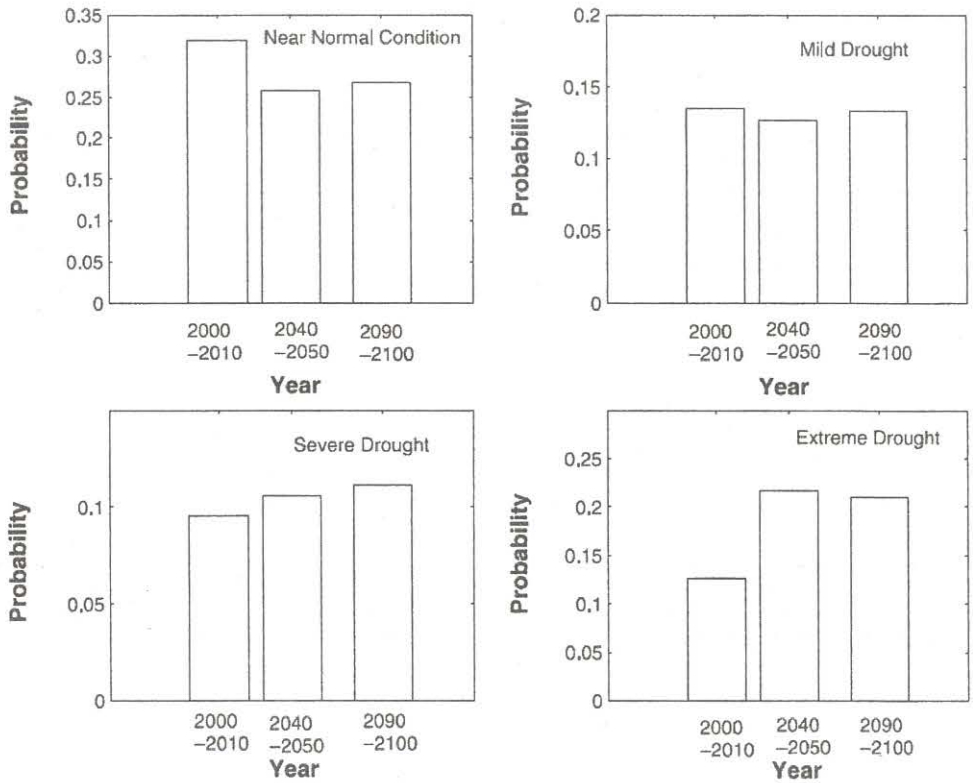

Fig 8.3.7 : Probability of Droughts with Normal Distribution for SPI-12

8.3.3.1 Assumption of Normal Distribution

The simplest method of modeling a sample of data without prior knowledge of distribution is with an assumption of normal distribution. In the present case, we assume no prior information regarding the future distribution of SPI-12 and, for simplicity, assume a normal distribution. The results for each GCM and emissions scenario is taken as the set of independent realizations of SPI-12 and that this set is used at each time step to establish the probability distribution. The values of the parameters of the normal distribution, i.e., mean and variance are considered as the sample estimates and are obtained from the of SPI-12 projected from different GCMs with scenarios at a particular year. As SPI-12, less than -2 indicates extreme drought, the CDF value of SPI-12 at -2 will give the probability of extreme drought.

$$P(ExtremeDrought) = F_{SPI}(-2) \tag{8.3.12}$$

Similarly the probability of other categories of drought at a particular year can be estimated from the CDF of the SPI-12 at that time. The probability of severe drought, mild to moderate drought and near normal condition are given by:

$$P(SevereDrought) = F_{SPI}(-1.5) - F_{SPI}(-2) \tag{8.3.13}$$

$$P(MildDrought) = F_{SPI}(-1.0) - F_{SPI}(-1.5) \tag{8.3.14}$$

$$P(NearNormal) = F_{SPI}(0) - F_{SPI}(-1.0) \tag{8.3.15}$$

where P(E) denotes the probability of an event E, and $F_{SPI}(x)$ denotes the value of CDF of SPI at x. A major limitation of this method is that there is no guarantee that SPI will follow normal distribution. This may lead to erroneous results, but an idea about the trend of severity, i.e, whether the probability of extreme events increases or decreases may be gathered from this analysis.

Fig. 8.3.7 shows the average probabilities of drought events for three time slices, years 2000-2010, 2040-2050 and 2090-2100. Considerable variation in the probabilities of near normal condition and extreme drought are seen from years 2000-2010 to 2040-2050. The probability of near normal condition is reduced and that of extreme drought is increased significantly in the years 2040-2050. Probabilities for mild and severe drought remain almost same. Variations in the probabilities of different drought are not significant in the later years, 2040-2050 to 2090-2100. This may mean that the assumption of normal

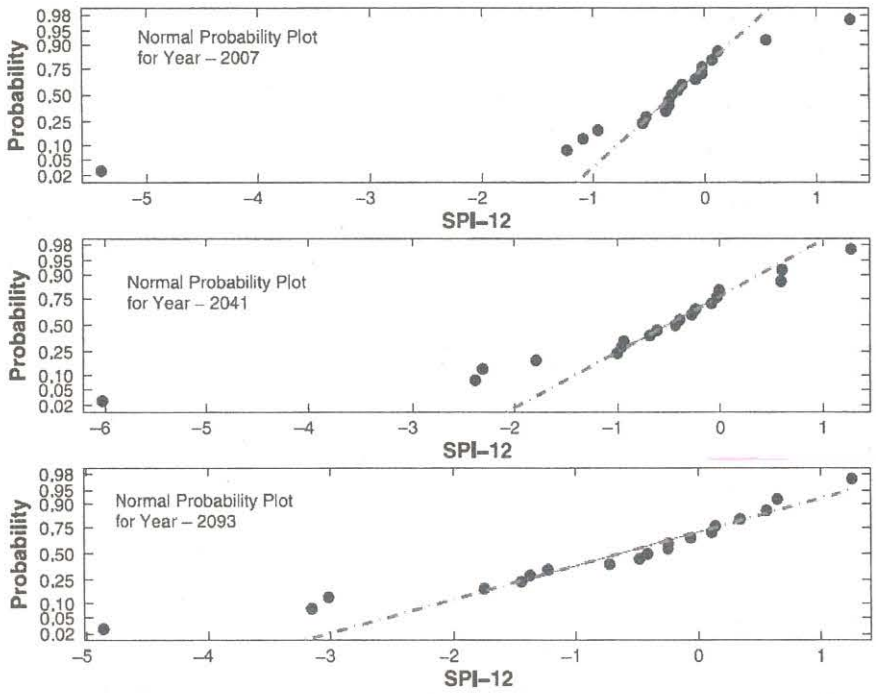

Fig 8.3.8 Normal Probability Plot of SPI-12 for Years 2007, 2041 and 2093

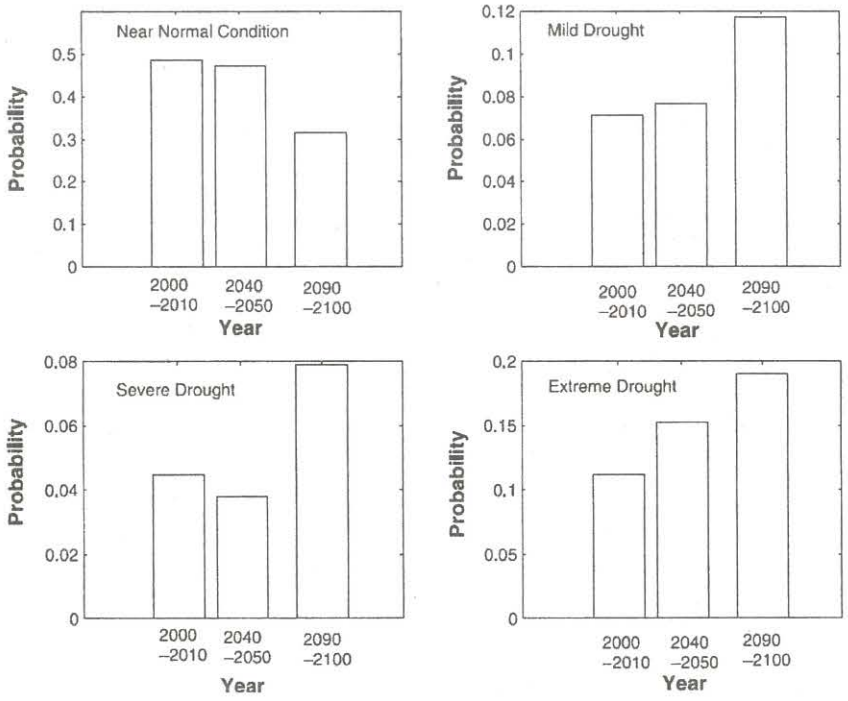

Fig 8.3.9 Probability of Droughts by Kernel Density Estimation

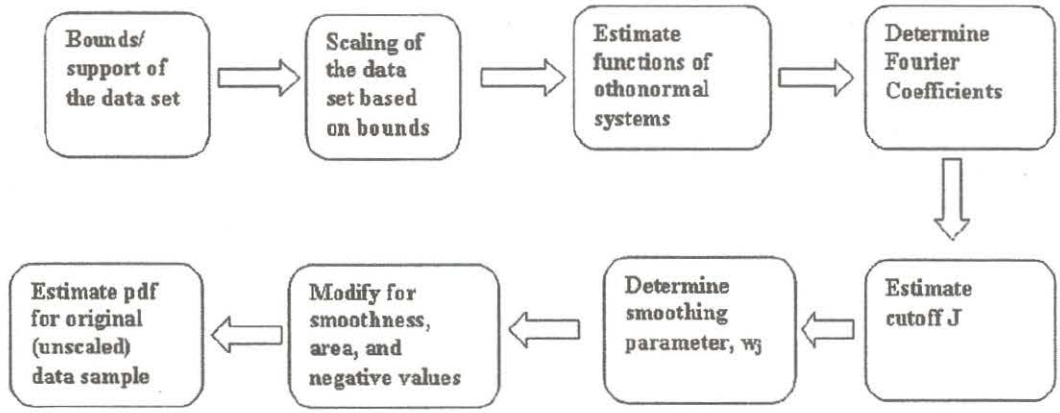

Fig 8.3.10 Algorithm for pdf Estimation Using Orthonormal Series Method

distribution does not result in a correct assessment of drought impacts of climate change years farther in future. Fig. 8.3.8 presents the normal probability plot of SPI-12 for three arbitrarily chosen years 2007, 2041 and 2093 from the three time slices. For all the cases the SPI-12 values deviate significantly from the normal distribution. A similar observation may be expected for other years also and thus the probability represented in Fig. 8.3.7 is not accurate. To determine the pdf of SPI-12 in a year more accurately, kernel density estimation method is used to obtain the nonparametric pdf of SPI-12 for each year in future.

8.3.3.2 Kernel Density Estimation

Kernel density estimation entails a weighted moving average of the empirical frequency distribution of the data. Most nonparametric 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 following property:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \qquad (8.3.16)$$

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

$$\hat{f}(x) = (nh)^{-1} \sum_{l=1}^{n} K((x - x_l)/h) \qquad (8.3.17)$$

where n is the number of observations (here, number of available GCM outputs) and x_i is the i^{th} observation (here, SPI-12), h is the smoothing parameter known as bandwidth, which is used for smoothening the shape of the estimated pdf. The selection of bandwidth is an important step in kernel estimation method. A change in bandwidth may dramatically change the shape of the kernel estimate (Efromovich, 1999).

Bandwidth for kernel estimation may be evaluated by minimizing the deviation of the estimated pdf from the actual one. When the actual pdf is unknown, the conventional method is to assume a normal distribution. Although, there are other methods like plug-in estimates (Polansky and Baker, 2000) and least square cross validation (Scott, 1992; Tarboton et al., 1998), in the present study the bandwidth is estimated based on normal distribution for computational simplicity at each of the time step. Thus the optimal bandwidth (h_0) is given by:

$$h_0 = (1.587)\sigma n^{-\frac{1}{3}} \qquad (8.3.18)$$

For non-normal densities σ is given by (Silverman, 1986):

$$\sigma = \min\{S, IQR/1.349\} \qquad (8.3.19)$$

where S is the the sample standard deviation and IQR is the interquartile range. The value of bandwidth thus evaluated is used to estimate the pdf of the data series using Eq. (8.3.17). This methodology is used to derive the nonparametric pdf for SPI at different time steps. By numerical integration the CDF value at SPI of -2, -1.5, -1.0, and 0 is estimated. These are used for finding out the probability of different classes of drought in future.

Fig. 8.3.9 presents the probabilities of drought conditions in the years 2000-2010, 2040-2050 and 2090-2100, as obtained using the kernel density estimation. Although an apparent increase in the probability of extreme drought are observed from both the methodologies (viz., methods based on the assumption of normal distribution and the kernel density estimation), the resulting probabilities are quite different. The difference between the probabilities of near normal condition for the years 2000-2010 and 2040-2050 using the method based on kernel density estimation is not significant whereas a larger change has been found for the model based on the assumption of normal distribution. A significant change is found for the years 2090-2100 from 2040-2050 in the

probabilities of near normal condition from kernel density estimation method, which is absent in the plots obtained from the model based on the assumption of normal distribution. The probability of extreme drought has a continuous increasing trend in Fig 8.3.9, which was absent in Fig. 8.3.7. Significant changes are also observed in the probabilities of mild and severe droughts. Although the methodology of kernel estimation, used in the present work is computationally simple, there are some drawbacks:

1. A large sample can give a better estimate of kernel density estimator. In the present analysis, the sample size (19) is small, consisting only of the downscaled SPI of the available GCM output.
2. The bandwidth is estimated by assuming the actual density is normal, which may not be valid.

To overcome these drawbacks a methodology based on orthonormal series is used which is an ideal method for estimation of nonparametric pdf from a small sample (Efromovich, 1999). The next subsections presents the details of the methodology for estimation of pdf using using orthonormal series.

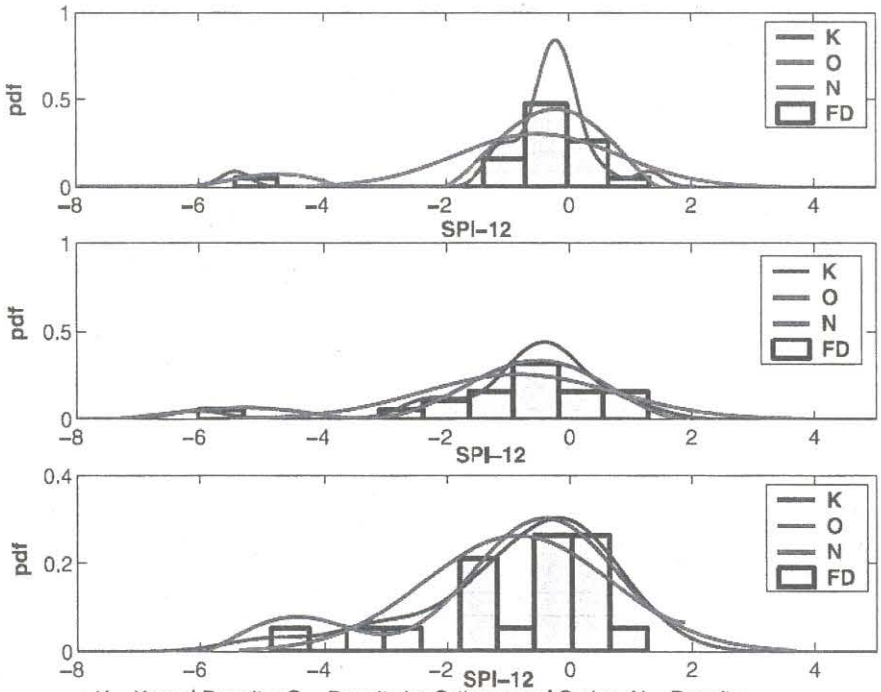

Fig 8.3.11 Pdf of SPI-12 for Years 2007, 2041 and 2093

8.3.3.3 Method of Orthonormal Series

A pdf from a small sample can be estimated using orthonormal series method which is essentially a series of orthonormal functions, obtained from the sample. The summation of the series with co-efficients results in the desired pdf. The following subsection presents the details of the concept and methodology of the method.

Concept of Orthonormal Series and Density Estimation

The methodology based on orthonormal series is used to estimate univariate density of data set with small sample size. Mathematical development of the methodology presented in this subsection is taken from Efromovich (1999). An orthonormal series is a series of orthonormal functions, $\Phi_s(x)$ and $\Phi_j(x)$ satisfying Eqns (8.3.20-8.3.21):

$$\int \Phi_s(x)\Phi_j(x)dx = 0 \quad \forall s \neq j \tag{8.3.20}$$

$$\int (\Phi_j(x))^2 dx = 1 \quad \forall j \tag{8.3.21}$$

Typically a univariate density function of a random variable X may be well approximated by an orthonormal series $\tilde{f}_J(x)$:

$$\tilde{f}_J(x) = \sum_{j=0}^{J} \theta_j \Phi_j(x) \tag{8.3.22}$$

where J is called the cut-off, $\Phi_j(x), j = 0,1,...$ are the functions of orthonormal system, and $\theta_j, j = 0,1,...$ are the coefficients corresponding to each function. In our case, X is SPI-12 simulated values from climate models. For this work, we select as the orthonormal series the subset of the Fourier series consisting of cosine functions:

$$\Phi_0(x) = 1 \tag{8.3.23}$$

$$\Phi_j(x) = \sqrt{2}\cos(\pi j x), \quad j = 1,2,3,.... \tag{8.3.24}$$

The algorithm involved in estimating probability density function based on orthonormal series is presented in Fig. 8.3.10. The detailed methodology is discussed in Appendix 3. After estimating the pdf numerical integration is performed for evaluating the CDF values at critical points of SPI-12 equal to -2, -1.5, -1.0, and 0.0. These are used for examining the severity of future drought.

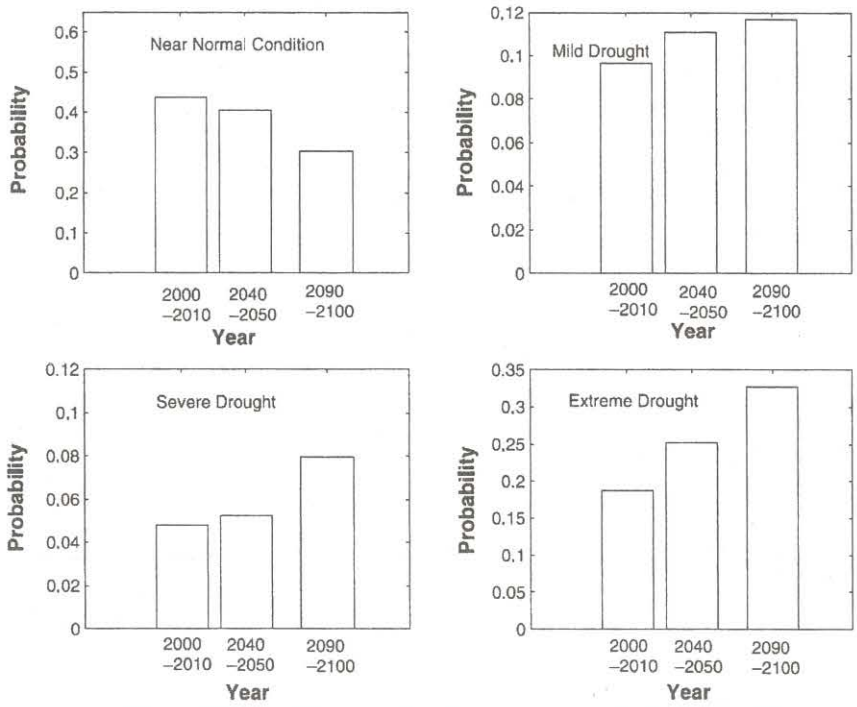

Fig 8.3.12 Probability of Droughts Using Orthonormal Series

Application and Results

The pdf of SPI-12 computed using the orthonormal series method is presented in Fig. 8.3.11 along with frequency distribution of the sample and the pdfs resulting from the other two methods for three arbitrarily chosen years 2007, 2041 and 2093 selected from the three time slices of 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, especially the feature of multi-modality, in all the three cases. The pdf obtained using orthogonal series closely resembles the shape generated by the frequency distribution. As an example, for the year 2007, around zero value of SPI-12 the kernel density estimator overestimates the pdf, whereas the pdf generated by orthonormal series estimates it reasonably accurately. A similar result is also obtained for the year 2041. For the year 2093, the pdfs obtained from both the nonparametric methods are nearly the same. One possible reason for the difference between the pdfs obtained from kernel density estimation and orthonormal series is the improper selection of bandwidth in kernel density estimator which may oversmooth or undersmooth the generated pdf.

Fig. 8.3.12 presents the probabilities of drought conditions in the years 2000-2010, 2040-2050 and 2090-2100, as obtained using orthonormal series based density estimation. The results are by and large similar to those of kernel density estimation except for the probabilities of mild drought. Kernel density estimation procedure projects a sudden increase in the probability of mild drought for the years 2090-2100, whereas such significant change is not observed in the results obtained from orthonormal series method. 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 drought will increase over time, which projects the Orissa meteorological subdivision to be more drought-prone in future. From Fig. 8.3.12, it is observed that significant increases in probabilities are indicated in cases of severe and extreme droughts only, which implies that climate change impact is more prominent on the extreme hydrologic events. As indicated by these results, impact of climate change may also be more severe for the Orissa meteorological subdivision because of it's position at the coast of the Bay of Bengal. A slight change in the pressure anomaly of the sea can have a severe impact on the precipitation of Orissa, which results in increase of hydrologic extremes in that region. Recent past records of Orissa with a fluctuating weather condition and the high occurrence of hydrologic extremes show that this is the most affected region of India due to climate change (www.cseindia.org/programme/geg/pdf/orissa.pdf).

Some recent studies on the trend of severity of drought for different regions around the world suggest that drought-prone areas are increasing worldwide. Andreadis and Lettenmaier (2006) have examined drought characteristics over the conterminous United States (U.S.) and found an increasing trend of drought in the southwest and parts of the interior of the West US. From the monthly data set of global PDSI (Palmer Drought Severity Index), Dai et al. (2004) have observed that most parts of Eurasia, Africa, Canada, Alaska, and eastern Australia became drier from 1950 to 2002 as large surface warming has occurred since 1950 over these regions. Dry area has more than doubled (from 12% to 30%) since the 1970s, with a large jump in the early 1980s due to precipitation decreases and subsequent expansion primarily due to surface warming. Without the warming, the PDSI decreases would have been much smaller and less

pervasive. Surface warming due to climate change and green house gas effect may lead to a severe drought condition in the present case study area Orissa. Dai et al. (1997) and Dai and Wigley (2000) have found that the decrease in precipitation is occurring mainly over El Nino Southern Oscillation (ENSO) sensitive regions. There is an established evidence of climatic teleconnection between ENSO and Indian rainfall (Maity and Nagesh Kumar, 2006) and thus the impact is more severe for India, especially for the Orissa region, because of its coastal position. Therefore, global warming with high surface warming in Orissa, sensitivity of precipitation to ENSO and coastal position are possible reasons for the trend in the probabilities obtained in this study for different categories of drought. The probabilities obtained from the analysis will be useful in computing the expected future damage due to drought, and to prepare the policy makers in generating appropriate responses.

8.3.4 General Observations

The methodology presented in this work deals with the problem of uncertainties due to GCMs and scenarios in a climate change impact assessment study. For examining the future drought scenario of Orissa meteorological subdivision in India, timeseries of SPI-12 are obtained from the projections of available outputs of several GCMs with several available scenarios. To model the uncertainty in a probabilistic framework it is assumed that there exists a pdf of the SPI-12 in each year of simulation. The pdf is estimated using nonparametric methods to obtain the probability of different categories of drought in the future. The results show an increasing trend in the probabilities of extreme and severe drought, in the region in future.

A limitation of the methodology presented here is that it does not consider the uncertainty due to parameterization and the structure of the impact model (GCM) itself, that is increasingly recognized in recent years. The other two sources of uncertainty not considered here are those due to starting conditions used in GCM simulations and the downscaling techniques. Given the hydrologic variable of interest projected from climate model runs with different parameter values of impact models or by using different downscaling techniques, the basic concepts of the proposed methodology may be used for uncertainty modeling. It may be noted that all scenarios are not available under all

GCMs, or, in other words, in the IPCC data distribution center does not provide outputs for all the scenarios for all GCMs. This leads to some implicit weighting of the GCMs.

Nonparametric pdf is used in the present study to estimate the probability of occurrence of different categories of droughts, under future climate change scenarios. Use of parametric pdf, such as the normal distribution does not lead to precise or accurate estimate of such probabilities. Modeling SPI assuming normal probability distribution should be considered only when the resulting imprecision is modeled. Uses of nonparametric pdf by kernel function, orthonormal system have also imprecision associated with the smoothing of kernel estimate, and determination of support of orthonormal system. Theoretically, SPI-12 can vary from $-\infty$ to $+\infty$, but in the present analysis, by using a heuristic method its support is fixed and used for pdf estimation. This may lead to imprecision in its estimate which is not modeled in the present analysis. In such cases, simultaneous accounting of randomness and imprecision or fuzziness, in a single integrated model by using the concept of imprecise probability (Zadeh, 2002) is useful, where the parameters of the probability distribution are considered as interval numbers or fuzzy numbers. Modeling GCM and scenario uncertainty with imprecise probability may give more generalized estimates of the probabilities of drought conditions, useful in examining future conditions.

Future scope of research of the present study includes the use of probabilities estimated for different categories of drought in water resources systems planning and operation. Steinmann (2003) has pointed out that drought indicators and triggers often lack statistical integrity, consistency among drought categories, and correspondence with desired management goals and thus evaluation of indicator for compatibility, consistency and applicability is a must before using them in water resources systems models. For example, for direct minimization of expected damage through reservoir operation, Surface Water Supply Index (SWSI) is more useful than SPI-12 as SWSI includes reservoir storage as input variable, but such analysis may involve a downscaling model to predict the reservoir inflow also. Use of imprecise probability in modeling GCM and scenario uncertainty along with the uncertainties due to parameterization, structure, initial value; and use of suitable drought indicator in decision making will be a useful direction of research on climate change impact assessment.

8.3.5 Remarks

A methodology of modeling GCM and scenario uncertainty for examining the severity of future drought is presented in this work. The drought indicator, SPI-12, projected from GCMs is considered to have a pdf in each year in future. The pdfs are estimated with nonparametric statistical techniques of kernel estimates and orthonormal system. Results are presented in terms of probabilities of different categories of drought in future. The methodology does not only represent uncertainty due to different GCM projections, but also incorporates it in examining the future drought scenario. Models based on software such as MAGICC (www.cgd.ucar.edu/cas/ACACIA/projects/magicc.html) require an assumption of prior probability distribution of GHG emission and concentration, which may also lead to high imprecision. The methodology presented here does not assume such prior probability. The methodology results in an increasing trend in the probability of severe and extreme drought for Orissa meteorological subdivision with a decrease in the probability of near normal condition. It may be concluded from the results that the region will be more drought prone due to the effect of climate change. Sources of uncertainty other than GCMs and scenarios (e.g., uncertainty due to parameterization, initial value, structure, and downscaling method) are ignored in the present study. The methodology presented here does not limit its usefulness only for drought prediction. Given a suitable index for a hydrologic event the methodology, may be used to examine the hydrologic implications of the GCM simulations for the particular event in future.

8.4 Streamflow Projections for Malaprabha River Basin for IPCC SRES Scenarios using SVM Based Downscaling and SWAT Model

Streamflows in a river basin are likely to vary in future owing to changes in climate and consequent changes in hydrometeorological processes, changes in land-use/ land-cover patterns and topography of watersheds plausibly due to natural hazards and/or anthropogenic factors. This section is concerned with development of a downscaling methodology to obtain future projections of streamflows in a river basin. The methods in vogue for projecting future streamflows under climate change scenarios can be broadly classified into **four categories, 1-4**:

interest today. Thus, the climates constructed by analog scenarios are not reliable representations of the greenhouse gas induced warmer conditions projected for future.

The downscaled scenarios are constructed by translating GCM simulated information from coarser scale to finer watershed scale using spatial downscaling models, based on the assumption that regional climate is conditioned by climate on relatively larger scale (e.g., continental). The spatial downscaling techniques can be broadly classified into dynamic downscaling and statistical downscaling. In the dynamic downscaling approach a Regional Climate Model (RCM) is embedded into GCM. The RCM is essentially a numerical model in which GCMs are used to fix boundary conditions. The major drawback of RCM, which restricts its use in climate impact studies, is its complicated design and high computational cost. Moreover, RCM is inflexible in the sense that expanding the region or moving to a slightly different region requires redoing the entire experiment (Crane and Hewitson, 1998). Whereas, the statistical downscaling involves deriving empirical relationships that transform large-scale features of the GCM (*LF*) to regional-scale variables (*RSV*)

$RSV = g\,(LF)$

where *RSV* represents predictands such as precipitation, temperature and streamflow; *LF* refers to predictors such as air temperature, relative and specific humidities, geo-potential height, and wind velocities at various pressure levels; and g is a downscaling function which could be deterministic or stochastic.

The classical statistical downscaling techniques include weather classification methods, weather generators and transfer functions. The simple and commonly used statistical downscaling approaches are based on transfer functions, which model relationships between predictors and predictand using methods such as linear and nonlinear regression, artificial neural networks, canonical correlation and principal component analysis (PCA). In this section, the transfer function based statistical downscaling method is chosen for determining plausible future scenarios of hydrometeorological variables to drive hydrological model, owing to the inherent advantages.

In methods of **category 4**, large scale atmospheric variables (LSAV) simulated by GCM are directly downscaled to streamflow at river basin scale for the present and future climate scenarios using methods such as nonlinear regression and artificial neural

networks. These methods do not take into account dynamics of regional hydrological processes and the mechanisms governing streamflow generation in a watershed. Hence their use may lead to unrealistic simulation of streamflows (e.g., Diaz-Nieto and Wilby, 2005). Nonetheless, the performance of a couple of models of this category is investigated in the present study.

In this section, a comparison is presented between future projections of streamflows in Malaprabha catchment of Krishna river basin, which are obtained for different climate change scenarios, hydrological models and methods for projecting streamflows. Four hydrological models (two each based on downscaling methods of categories 3 and 4), and four climate change emission scenarios (A1B, A2, B1 and COMMIT) are considered for the analysis. Description of each of the scenarios is given in special report on emission scenarios (SRES; Nakicenovic et al., 2000), and all the scenarios are relevant to IPCC's fourth assessment report (AR4) (Alley et al., 2007). Simulations for each of the scenarios are obtained from the Canadian third generation coupled GCM.

The two hydrological models considered to determine future projections of streamflows by methods of category 3 are Soil and water assessment tool (SWAT) (Neitsch et al., 2000) and support vector machine (SVM) (Vapnik, 1995, 1998) based empirical model (SBEM). The SVM has the advantage of implementing the structural risk minimization principle, which is the ability to learn any training set without error. Further, the global optimum solution is possible with SVM.

The two models developed based on methods of category 4 involved use of SVM for downscaling LSAV. One of the two models directly downscales LSAV to streamflows in the target watershed, and is referred to as direct downscaling model (DDSM). The other model downscales LSAV to streamflows in two stages, and is called two-stage deterministic downscaling model (TSDDSM).

8.4.1 Study region and data used

The catchment of Malaprabha reservoir in the Karnataka state of India is considered for case study. It has an area of 2564 km^2 situated between 15°30′ N and 15°56′ N latitudes and 74°12′ E and 75°15′ E longitudes. The source of water for the catchment of Malaprabha reservoir is precipitation. The Malaprabha reservoir is one of the major lifelines for the arid regions of north Karnataka (possibly the largest arid region in India

(1) The data of large set of predictors required for downscaling all the six hydrometeorological variables (precipitation, maximum and minimum temperature, relative humidity, wind speed and solar radiation) considered in the current study are readily available from simulations of CGCM3.

(2) To develop downscaling model, the NCEP data of predictors form input, and the observed data of hydrometeorological variables and streamflows form output. For projecting future sequences of hydrometeorological variables and streamflows, it is necessary to consider GCM data that is consistent with NCEP data. Investigations undertaken as part of previous studies (Tripathi et al., 2006; Anandhi, 2007; Anandhi et al., 2008, 2009) and current study revealed that the CGCM3 data of predictors are consistent with those in NCEP data, and are hence considered reliable for use in climate impact study.

(3) Earlier version of the model (CGCM2) was used to successfully downscale precipitation for India in a previous study (Tripathi et al., 2006) and there is a natural interest to investigate changes in projections using simulations of the latest version of the Canadian model.

The GCM data and the information on atmospheric flux are re-gridded to a common 2.5° NCEP grid using Grid Analysis and Display System (GrADS) (Doty and Kinter, 1993).

Land use/land cover map and soil map are procured from the Karnataka State Remote Sensing Application Center (KSRSAC), Bangalore, India, prepared based on panchromatic (PAN) and linear imaging self scanner (LISS) III merged, Indian remote sensing (IRS) satellite images. Modified data of Shuttle Radar Topography mission (SRTM) in the form of Digital Elevation Model (DEM) for the study region is procured from the International Water Management Institute (IWMI), Hyderabad, India. Topomaps of the study region are procured from Survey of India (SOI) at available finest scale 1:50,000.

8.4.2 Methodology

The methodology followed to arrive at future projections of streamflows in the Malaprabha catchment using SWAT, SBEM, DDSM and TSDDSM is briefly described in this section. Each of the models was calibrated using observed data for the period 1978 – 1993 and validated for the period 1994 – 2000. Subsequently, future streamflows were

generated using the developed model for the period 2001 – 2100, for each of the four emission scenarios (A1B, A2, B1 and COMMIT) considered.

8.4.2.1 Downscaling LSAV to streamflows through SWAT model

The LSAV were downscaled to streamflows through SWAT model in three stages (1-3). In stage 1, six SVM models were developed to downscale monthly sequences of LSAV to monthly sequences of hydrometeorological variables (precipitation, maximum and minimum temperature, relative humidity, wind speed and cloud cover) at target locations in the Malaprabha catchment. For precipitation, the target location is centroid of the watershed, whereas the same is meteorological station for each of the other five variables. In stage 2, *k*-nearest neighbor (*k*-NN) technique is used to disaggregate the monthly sequences of hydrometeorological variables to daily sequences. Finally, in stage 3, SWAT hydrological model is driven by the daily sequences of hydrometeorological variables to arrive at future projections of streamflows in the Malaprabha catchment. Details pertaining to each of these stages are as follows:

Table 8.4.1 Probable predictors selected for downscaling predictands.

Sl. No.	Predictand	Probable predictor(s) selected from NCEP and CGCM3 monthly data sets for downscaling the predictand
1	Precipitation	Ta 925, Ta 700, Ta 500, Ta 200, Zg 925, Zg 500, Zg 200, Hus 925, Hus 850, Ua 925, Ua 200, Va 925, Va 200, prw, ps
2	Maximum temperature	Ta 925, Ua 925, Va 925, LH, SH, SWR, LWR
3	Minimum temperature	Ta 925, Ua 925, Va 925, LH, SH, SWR, LWR
4	Wind speed	Ua 925, Va 925
5	Relative humidity	Ta 925, Hus 925, LH
6	Cloud cover	prw
7	Streamflow	Ta 925, Ta 700, Ta 500, Ta 200, Zg 925, Zg 500, Zg 200, Hus 925, Hus 850, Ua 925, Ua 200, Va 925, Va 200, prw, ps

Stage 1: Downscaling hydrometeorological variables

To develop a downscaling model for each hydrometeorological variable (predictand), a separate set of probable predictors were identified from the LSAV (Table 8.4.1). The steps involved in developing each of the six SVM models are as follows:

study, grid search procedure (Gestel et al., 2004) is used to find the optimum range for each of these two parameters. Subsequently, the optimum values of the parameters are obtained from the selected ranges, using the stochastic search technique of genetic algorithm (Haupt and Haupt, 2004). It is worth mentioning that sensitivity of SVM model's performance to the choice of kernel function is yet to be explored.

Step 9: Validation of the developed SVM model using feature vectors from the test set as input to calibrated model, and comparing model output with contemporaneous observed values of predictand. The 'normalized mean square error (NMSE)' is used as an index to assess the performance of the model.

Step 10: Obtaining future monthly projections of the predictand using feature vectors prepared from GCM simulations as input to the validated SVM downscaling model, for each of the four emission scenarios.

Step 11: Determining the trend in the projected monthly values of the predictand for each scenario by dividing the projected time sequence of predictand into five parts (2001-2020, 2021-2040, 2041-2060, 2061-2080 and 2081-2100).

Stage 2: Disaggregation of hydrometeorological variables

The SWAT hydrological model chosen for arriving at streamflow projections requires daily sequences of hydrometeorological variables as input. For this purpose, the monthly sequences of hydrometeorological variables obtained using SVM downscaling models in stage 1 are disaggregated to daily scale using *k*-nearest neighbor (*k*-NN) technique.

The disaggregation step can be avoided if daily sequences of LSAV can be directly downscaled to daily sequences of hydrometeorological variables in stage 1. However, this option was not preferred because (1) some of the LSAV are not simulated by GCM at daily scale and (2) the monthly sequences of atmospheric variables simulated by the GCM are more reliable than those simulated at daily scale (Prudhomme et al., 2003).

Before proceeding to describe the proposed disaggregation algorithm, variables are defined herein. Let the historical (past) and projected (future) values of predictand be denoted by $u^{h}_{\upsilon h,\tau,j}$ and $u^{p}_{\upsilon p,\tau,j}$, respectively, where the subscripts υh and υp are indices for the past (historical) and future (projected) years (υh=1,...,N_h; υp=1,...,N_p), τ denotes the index for the month within the year (τ = 1,2,...,ω), and j refers to the

index for the day within the month τ . N_h refers to the number of years of historical record (herein N_h = 23 for data from 1978 to 2000), N_p refers to the projected period (herein, N_p = 100 for data from 2001-2100), and ω represents the number of months (=12) in a year. Further, let $u^h_{\upsilon h,\tau,\bullet}$ denote the monthly mean value of the predictand computed using the observed daily values of the predictand in the month τ of year υh. Similarly, let the monthly mean value projected for the predictand in the month τ of year υp be $u^p_{\upsilon p,\tau,\bullet}$.

$$u^h_{\upsilon h,\tau,\bullet} = \frac{\sum_{j=1}^{D_\tau} u^h_{\upsilon h,\tau,j}}{D_\tau} \qquad \tau = 1,\ldots,t,\ldots,\omega \quad \upsilon h = 1,\ldots,N_h \tag{8.4.1}$$

where D_τ denotes the number of days in month τ . For the calibration period (which could be less than or equal to the length of the historical record), the observed value of the predictand on day j in month τ is expressed as a fraction of the monthly mean value of the predictand as:

$$w_{\upsilon h,\tau,j} = \frac{u^h_{\upsilon h,\tau,j}}{u^h_{\upsilon h,\tau,\bullet}} \qquad j = 1,\ldots,D_\tau \quad \tau = 1,\ldots,\omega \quad \upsilon h = 1,\ldots,N_h \tag{8.4.2}$$

Let $\boldsymbol{W}^h_{\upsilon h,\tau} = \{w_{\upsilon h,\tau,1},\ldots,w_{\upsilon h,\tau,D_\tau}\}$ denote the vector containing the fractions of the daily values of predictand in month τ of year υh.

The key steps of the proposed algorithm are as follows:

Step 1: For every projected value of the predictand, the calendar month τ is identified.

Step 2: The conditioning set $\mathbf{z}_\tau$ for each month τ consists of historical state vectors prepared from each of the years of observed record. Historical state vector prepared for month τ of year υh comprises of the observed monthly values of the predictand for the calendar months falling in a window of size M centered on the month τ and extending over all the years of the historical record. For example, for a window of size one month, the conditioning set is $z_\tau = \{u^h_{1,\tau,\bullet},\ldots,u^h_{\upsilon h,\tau,\bullet},\ldots,u^h_{N_h,\tau,\bullet}\}$. The window size depends on the projected trend of the downscaled hydrometeorological variable at monthly scale. Window of size one month could be considered if the SVM model projects no change in

trend. In the presence of trend, a larger size window (e.g, 3 months) is recommended to explore the temporal relationship between a wide range of monthly and daily values of the predictand for use in disaggregation.

Step 3: To disaggregate the projected value of predictand in month τ , $u^{p}_{\upsilon p,\tau,\bullet}$, select its k-nearest neighbors from the conditioning set z_{τ} based on the Euclidean distance $\xi^{ph}_{\upsilon p,\upsilon h,\tau}$ between $u^{p}_{\upsilon p,\tau,\bullet}$ and $u^{h}_{\upsilon h,\tau,\bullet}$ expressed as

$$\xi_{\upsilon h,\upsilon p} = \left\| u^{p}_{\upsilon p,\tau,\bullet} - u^{h}_{\upsilon h,\tau,\bullet} \right\| \quad \text{for } \upsilon h = 1,\ldots, N_h \tag{8.4.3}$$

The number of neighbors k is a smoothing parameter. Lall and Sharma (1996) suggest using k equal to $\sqrt{N_h}$, as a rule of thumb. The k nearest neighbors selected are those most similar to $u^{p}_{\upsilon p,\tau,\bullet}$.

Step 4: Weights are assigned to each of the k nearest neighbors using the discrete probability mass function $p(i)$ (Lall and Sharma, 1996) given by Eq.(8.4.4). This discrete kernel was developed through a Poisson approximation of the probability distribution function of neighbors, in the space defined by historical state vectors (i.e., state space neighbors). Randomly select a nearest neighbor to $u^{p}_{\upsilon p,\tau,\bullet}$ by constructing cumulative density function using $p(i)$ values.

$$p(i) = \frac{1/i}{\sum_{j=1}^{k}(1/j)} \qquad i = 1,\ldots,k \tag{8.4.4}$$

It is to be noted that $p(i)$ is the same for all the months in the projected period. Let $u^{h}_{\upsilon h,\tau,\bullet}$ denote the nearest neighbor.

Step 5: The projected daily values of the predictand for the month τ in year υp are obtained by multiplying $u^{p}_{\upsilon p,\tau,\bullet}$ with $W^{h}_{\upsilon h,\tau}$ corresponding to the nearest neighbor $u^{h}_{\upsilon h,\tau,\bullet}$.

Stage 3: Projecting streamflows using SWAT model

In this stage, SWAT hydrological model is driven by daily sequences of hydrometeorological variables (determined using disaggregation model) to obtain past and future projections of streamflows. The SWAT is a river basin scale hydrological model developed for the United States Department of Agriculture (USDA), Agricultural Research Service. Being a physically based, semi-distributed, continuous time model, it requires numerous inputs and parameters that represent weather, hydrology, soil properties, plant growth, nutrients, pesticides, bacteria and pathogens, and land management. Since its development in the early 1990s, SWAT has undergone continual review and expansion of capabilities. In SWAT, a watershed is divided into multiple subwatersheds, which are then further subdivided into hydrologic response units (HRUs) that consist of homogeneous land use, management, and soil characteristics. The overall hydrological balance is simulated for each subwatershed.

The SWAT model simulates streamflow using the modified SCS runoff curve number method (USDA-NRCS, 2004) or the Green-Ampt method. To apply the SWAT model for streamflow estimation using curve number method, the observed climate variables, namely precipitation, maximum and minimum temperature, wind speed and relative humidity are used as inputs in addition to other processed inputs such as Digital Elevation Model (DEM), land use and soil map. For this analysis, the other inputs to SWAT model, such as DEM, landuse/land cover, and soil type are considered to be the same for the past and the future scenarios. The ArcView GIS interface of SWAT provides an easy-to-use graphical user interface for organizing all the required inputs. Delineating the subwatershed boundaries, defining the HRUs, generating SWAT input files, creating agricultural management scenarios, executing SWAT simulations, and reading and charting of results were all carried out by the various tools available in the interface.

The SWAT model is calibrated using the data for the period January 1978 to December 1993, and validated using data for the period 1994 – 2000. For the sake of calibrating the model, sensitive parameters are identified using Latin Hypercube One-factor-At-a-Time (LH-OAT) (Morris, 1991) method. The sensitive parameters are those that cause large changes in the streamflow generated by the model, when perturbed. The details of the LH-OAT method can be found in Griensven (2005).

To determine optimal values of parameters, auto calibration of the sensitive parameters was carried out using Monte Carlo analysis. For this purpose, the ranges of the selected sensitive parameters are divided into initial input range and behavioral range. The initial input range is based on theoretical limits of the parameters. Thousand parameter sets, containing a fairly broad range of parameter values, are generated from the initial input range using independent uniform distributions. SWAT is driven by each of the parameter sets and daily data on hydrometeorological variables (i.e, precipitation, maximum and minimum temperatures, relative humidity and wind speed) for the period 1978-1993 to generate streamflows. For each parameter set, the generated streamflows are compared with contemporaneous historical values using eight performance measures (described in section 8.4.2.6) to determine the behavioral range of each of the parameters. The behavioral range denotes the parameter range for which majority of the performance measures have optimal values. Following this, ten thousand parameter sets are generated from the behavioral parameter ranges, using independent uniform distributions. Subsequently, for each of the ten thousand parameter sets, streamflows generated by SWAT for the period 1978-1993 are compared with contemporaneous historical values using the eight performance measures. Since there is a possibility of producing acceptable results with different parameter sets, the generated streamflows and parameter set(s) corresponding to optimal value of each of the performance measures are collated. The generated streamflows (corresponding to each one of the collated parameter sets) are visually interpreted by comparison with contemporaneous historical streamflows for the period 1978-1993. The parameter set for which generated streamflows compared well with the historical streamflows is decaled as optimal parameter set.

For validation, the SWAT model is driven by the optimal parameters and historical daily data on hydrometeorological variables for the period 1994 – 2000, and the generated streamflows are compared with contemporaneous historical values using the eight performance measures. The model is declared acceptable after verifying that the generated streamflows compare well with the historical streamflows.

Subsequently, for each of the four climate change scenarios, the SWAT model is driven by the optimal parameters and the daily sequences of downscaled hydrometeorological variables for the period 2001 – 2100 (obtained using *k*-NN disaggregation technique) to

generate future projections of daily streamflows. The daily streamflow sequences were aggregated to obtain monthly streamflow sequences.

8.4.2.2 Downscaling LSAV to streamflows using SVM based empirical model (SBEM)

The proposed method has two stages. The first stage involves downscaling LSAV to future projections of hydrometeorological variables (precipitation, maximum and minimum temperature, relative humidity, wind speed and cloud cover) at target locations in the Malaprabha catchment, following the procedure described for stage-1 in subsection 8.4.2.1. In the second stage, the hydrometeorological variables are further downscaled to streamflows in the Malaprabha catchment using SVM based empirical model (SBEM).

The SBEM captures empirical relationship between historical values of hydrometeorological variables and streamflows. For developing SBEM, hydrometeorological variables having significant correlation with streamflows are identified using historical data for the period 1978-1993. The historical data on the identified hydrometeorological variables are standardized for the period 1978-2000 and principal components (PCs), which preserve more than 98% of the variance in the data, are extracted. Along the principal directions obtained from the principal component analysis, the PCs of future projections of hydrometeorological variables (obtained from stage 1) are extracted. The procedure for preparing feature vectors, and calibration and validation of the SBEM is same as that described in steps 6 to 9 for stage-1 in subsection 8.4.2.1. The validated SBEM is driven by optimal parameters and PCs of future projections of hydrometeorological variables to obtain monthly streamflow sequences as outputs.

8.4.2.3 Downscaling streamflows using deterministic downscaling model (DDSM)

The DDSM involves directly downscaling LSAV to streamflow at target location in Malaprabha river basin for the present and future climate scenarios using SVM technique. The downscaling procedure is similar to that described for stage-1 in subsection 8.4.2.1. Observed streamflow (instead of precipitation) is considered as predictand, while the predictors comprise LSAV from NCEP and GCM datasets (listed in Table 8.4.1).

8.4.2.4 Downscaling streamflows using two-stage deterministic downscaling model (TSDDSM)

The TSDDSM executes the task of DDSM in two stages. The information concerning future scenarios of hydrometeorological variables and streamflows are obtained from first and second stages respectively.

The first stage involves development of six SVM models to downscale monthly sequences of LSAV to monthly sequences of hydrometeorological variables at target locations in the Malaprabha catchment for the past (1978-2000) and future (2001-2100) periods, following the procedure described for stage-1 in subsection 8.4.2.1.

The second stage involves development of SVM model to capture empirical relationship between model generated data on hydrometeorological variables for the past period (1978-1993) (obtained from first stage) and contemporaneous historical streamflows. The procedure is similar to that described for development of SBEM in subsection 8.4.2.2, except that generated data on hydrometeorological variables replaces its historical counterpart. The developed SVM model is used to downscale future projections of hydrometeorological variables to future projections of monthly streamflow sequences.

8.4.2.5 Least-Square Support Vector Machine (LS-SVM)

The Least-Square Support Vector Machine (LS-SVM) has been used in this study for downscaling. The LS-SVM provides a computational advantage over standard SVM (Suykens, 2001; Srinivas and Tripathi, 2008). This sub section presents the underlying principle of the LS-SVM.

Consider a finite training sample of N patterns $\{(\mathbf{x}_i, y_i), i = 1,\ldots,N\}$, where $\mathbf{x}_i$ is the "i-th" pattern in n-dimensional space (i.e., $\mathbf{x}_i = [x_{1i},\ldots,x_{ni}] \in \Re^n$), and it constitutes input to LS-SVM, whereas $y_i \in \Re$ is the corresponding value of the desired model output. Further, let the learning machine be defined by a set of possible mappings $\mathbf{x} \mapsto f(\mathbf{x},\mathbf{w})$, where $f(\cdot)$ is a deterministic function which, for a given input pattern $\mathbf{x}$ and adjustable parameters $\mathbf{w}$ ($\mathbf{w} \in \Re^n$), always gives the same output. The training phase of the learning machine involves adjusting the parameters $\mathbf{w}$. These parameters are estimated by minimizing the cost function $\Psi_L(\mathbf{w},e)$.

$$\Psi_L(\mathbf{w},e) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}C\sum_{i=1}^{N} e_i^2 \tag{8.4.5}$$

subject to the equality constraint

$$y_i - \hat{y}_i = e_i \qquad i = 1,...,N \tag{8.4.6}$$

where C is a positive real constant and $\hat{y}_i$ is the actual model output. The first term of the cost function represents weight decay or model complexity-penalty function. It is used to regularize the weight sizes and to penalize the large weights. This helps in improving the generalization performance (Hush and Horne, 1993). The second term of the cost function represents penalty function.

The solution of the optimization problem is obtained by considering the Lagrangian as

$$L(\mathbf{w},b,\mathbf{e},\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}C\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N}\alpha_i\{\hat{y}_i + e_i - y_i\} \tag{8.4.7}$$

where α_i are Lagrange multipliers and b is the bias term. The conditions for optimality are given by

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}\alpha_i\phi(\mathbf{x}_i) = 0 \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{N}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = \alpha_i - Ce_i = 0 \qquad i = 1,...,N \\ \dfrac{\partial L}{\partial \alpha_i} = \hat{y}_i + e_i - y_i = 0 \qquad i = 1,..,N \end{cases} \tag{8.4.8}$$

The above conditions of optimality can be expressed as the solution to the following set of linear equations after elimination of $\mathbf{w}$ and e_i .

$$\begin{bmatrix} 0 & \vec{\mathbf{1}}^T \\ \vec{\mathbf{1}} & \boldsymbol{\Omega} + C^{-1}\mathbf{I} \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{8.4.9}$$

$$\text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}; \vec{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{N\times 1} ; \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}; \mathbf{I} = \begin{bmatrix} 1\,0\ldots 0 \\ 0\,1\ldots 0 \\ \vdots\,\vdots\,\vdots\,\vdots \\ 0\,0\ldots 1 \end{bmatrix}_{N\times N} \tag{8.4.10}$$

In Eq.(8.4.9), $\boldsymbol{\Omega}$ is obtained from the application of Mercer's theorem.

$$\Omega_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^{\mathrm{T}} \phi(\mathbf{x}_j) \qquad \forall i, j \tag{8.4.11}$$

where $\phi(\cdot)$ represents nonlinear transformation function defined to convert a non-linear problem in the original lower dimensional input space to linear problem in a higher dimensional feature space.

The resulting LS-SVM model for function estimation is:

$$f(\mathbf{x}) = \sum \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^* \tag{8.4.12}$$

where α_i^* and b^* are the solutions to Eq.(8.4.9) and $K(\mathbf{x}_i, \mathbf{x})$ is the inner product kernel function defined in accordance with Mercer's theorem (Mercer, 1909; Courant and Hilbert, 1970) and b^* is the bias. There are several choices of kernel functions, including linear, polynomial, sigmoid, splines and Radial basis function (RBF). The linear kernel is a special case of RBF (Keerthi and Lin, 2003). Further, the sigmoid kernel behaves like RBF for certain parameters (Lin and Lin, 2003). In this study RBF is chosen to map the input data into higher dimensional feature space, which is given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2}{\sigma} \right) \tag{8.4.13}$$

where, σ is the width of RBF kernel, which can be adjusted to control the expressivity of RBF. The RBF kernels have localized and finite responses across the entire range of predictors.

The advantage with RBF kernel is that it maps the training data non-linearly into a possibly infinite-dimensional space and thus it can effectively handle the situations when the relationship between predictors and predictand is nonlinear. Moreover, the RBF is computationally simpler than polynomial kernel, which requires more parameters. It is

with local thermodynamic stability and hence are useful as predictors. Temperature affects the moisture holding capacity of the atmosphere and the pressure at the point. The pressure gradient affects the circulation, which in turn affects the moisture brought into the place and hence the precipitation. Higher precipitable water in the atmosphere means more moisture, which in turn causes statically unstable atmosphere leading to more vigorous overturning, resulting in more precipitation. Lower pressure leads to more winds and so more precipitation. At 925 hPa pressure height, the boundary layer (near surface) effect is prominent. The 850 hPa pressure height is the low level flow response to regional precipitation. The 200 hPa pressure level depicts the global scale effects. Temperature at 500 hPa represents the heating process of the atmosphere due to monsoonal precipitation which is maximum at mid-troposphere at a constant pressure height. Geopotential height represents the pressure gradient which is related to the moisture brought into the place and hence the precipitation. Considering these aspects, fifteen probable predictors are extracted from the NCEP reanalysis and CGCM3 data sets. These are the air temperature at 925hPa (Ta 925), 700hPa (Ta 700), 500hPa (Ta 500) and 200hPa (Ta 200) pressure levels, geo-potential height at 925hPa (Zg 925), 500hPa (Zg 500) and 200hPa (Zg 200) pressure levels, specific humidity at 925hPa (Hus 925) and 850hPa (Hus 850) pressure levels, zonal (Ua) and meridional wind velocities (Va) at 925hPa (Ua 925, Va 925) and 200hPa (Ua 200, Va 200) pressure levels, precipitable water (prw) and surface pressure (Ps).

Scatter plots and cross-correlations were used to assess the relationship between (1) predictors in NCEP and GCM data sets, and (2) predictors in NCEP data sets and streamflow for each of the 2.5° NCEP grid points considered in the study (Fig. 8.4.1). The cross-correlations are computed using the three measures of dependence(product moment correlation, Spearman's rank correlation and Kendall's tau), to verify if they are consistent. It is observed that, (1) there is no distinct non-linearity between the predictor variables in NCEP and GCM data sets; (2) all three measures of dependence show near equal ranking of probable predictors. Therefore, in the following discussion only product moment correlation values are described, without loss of generality; (3) the correlation between predictors in NCEP and GCM data sets is generally greater than 0.5 (except for Va 200 and Zg 500), indicating that the predictor variables are realistically simulated by

the GCM; (4) Ta 700, Ta 500, Ta 200, Ua 925, Va 925, Zg 200, Hus 925, Hus 850 and prw have positive correlation with the streamflow, while Ta 925, Va 200, Ua 200, Zg 925 and Zg 500 have negative correlation with the same; (5) the predictor variables Ta 925, Ta 700, Ta 500, Ta 200, Zg 500, Zg 200, Hus 925, Hus 850 and prw show an increasing trend. The projected increase in trend is high for A2 scenario, while it is least for B1 scenario, whereas no trend is discerned with the COMMIT scenario.

8.4.2.8 Results

Downscaling LSAV to streamflows through SWAT model

The LSAV are downscaled to monthly sequences of the six hydrometeorological variables (precipitation, maximum and minimum temperature, relative humidity, wind speed and cloud cover) at target locations in Malaprabha catchment, following the procedure described for stage-1 in subsection 8.4.2.1. The sequences of each of the downscaled hydrometeorological variables obtained for the past (1978-2000) and future periods (2001-2100) are compared with those observed in the study region using mean statistics (Fig. 8.4.2). In the figure, results for the past period (1978–2000) are shown in (i), whereas the projected variable for the periods 2001-2020, 2021-2040, 2041-2060, 2061-2080 and 2081-2100, for the four scenarios (A1B, A2, B1 and COMMIT) are shown in (ii), (iii), (iv) and (v) respectively. The mean statistic of the observed data is modeled fairly well by the downscaling models, but not the variability. Possibly this could be because regression based statistical downscaling models often cannot explain the entire variance of the downscaled variable (Wilby et al., 2004). Exploration of a larger data set for calibration and validation of the SVM model could possibly provide more insight into this problem. However, in the present study, investigation in this direction is constrained by the limitations posed by the available data.

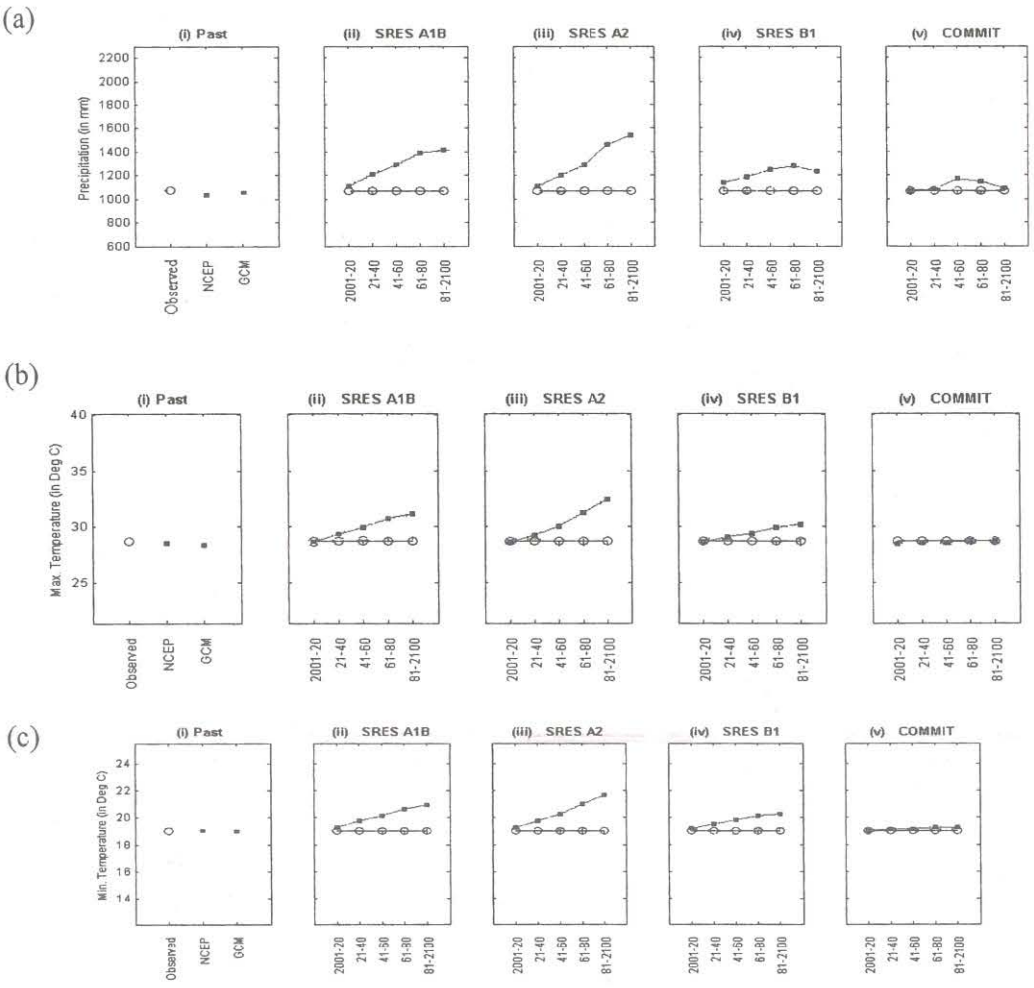

Fig 8.4.2 Typical results from the SVM based downscaling model. Results pertaining to hydrometeorological variables namely precipitation, maximum and minimum temperature, relative humidity, wind speed and cloud cover are represented as a, b, c, d, e and f respectively. The circles denote the observed mean monthly value of a hydrometeorological variable for 1978-2000, and the darkened square represents the mean monthly value of downscaled variable. In the Fig.(i), NCEP and GCM denote SVM generated values of the variable for the past period (1978-2000), for input from predictor variables in NCEP and GCM data respectively. In (ii), (iii), (iv) and (v) the solid line that joins the circles indicates the historical mean monthly value of the variable, while the line connecting the solid squares depicts the mean trend of the variable projected by GCM.

of the SWAT and the model was run forming 14 drainage sub-basins in the study region. The land-use/land-cover and soil maps of the region are overlaid on each other to identify Hydrological Response Units (HRUs). Subsequently, the data of observed climate variables are fed into the SWAT, and it is calibrated and validated to model daily streamflows.

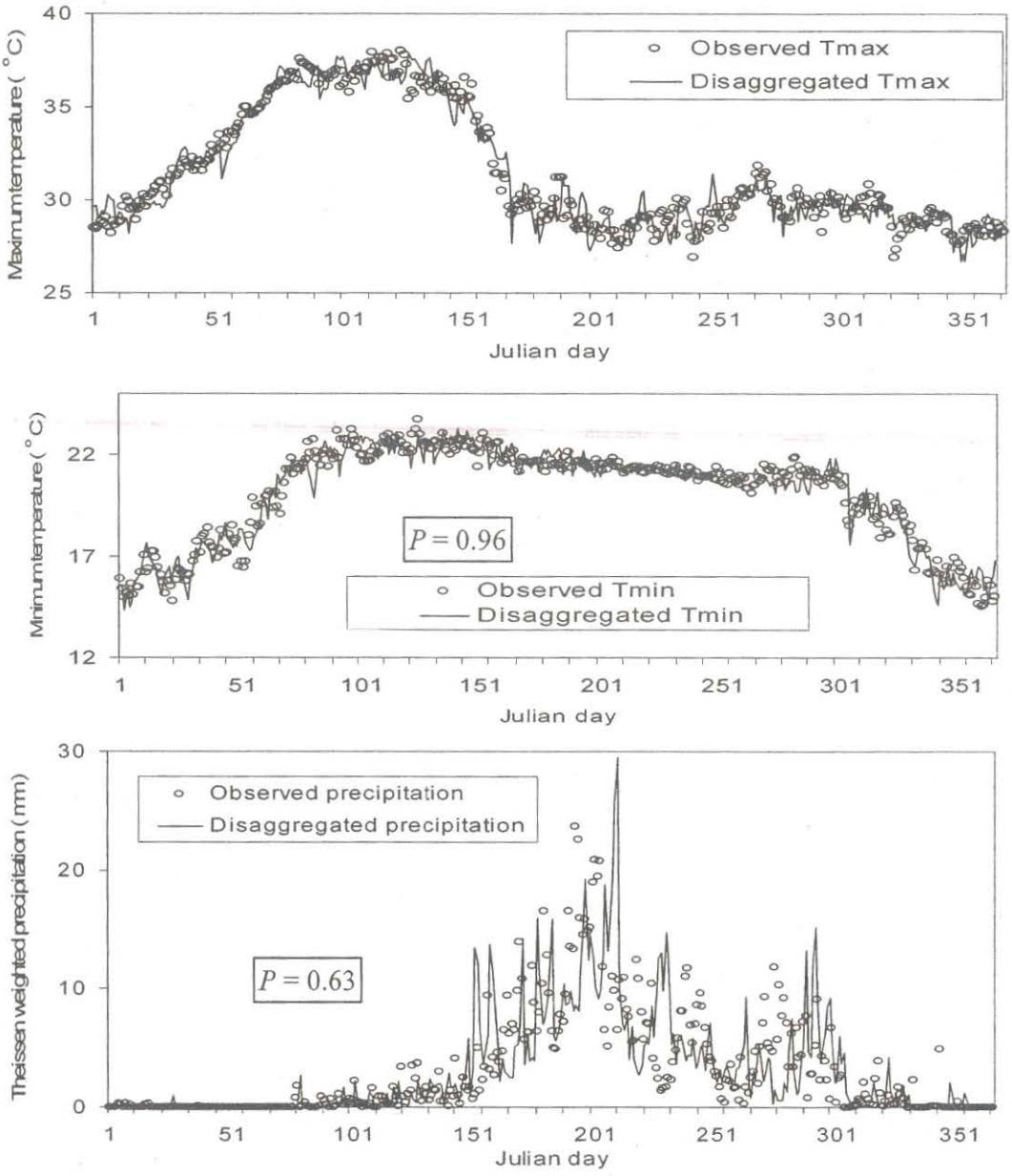

Fig 8.4.3 Observed and disaggregated mean daily maximum and minimum temperatures, and precipitation for the validation period from 1994 to 2000

In the calibration phase, the performance of SWAT is assessed by comparing streamflow generated by the model with that observed at the Malaprabha reservoir at monthly time scale. The model, in general, over-predicted streamflow for several months. This could be attributed to retention storage in natural depressions/tanks of the catchment, which goes unaccounted for in estimating inflows into Malaprabha reservoir every water-year. Sensitivity analysis carried out using LH-OAT (described in stage 3 of subsection 8.4.2.1) revealed that curve number (CN), available water capacity (AWC) of soil, plant uptake compensation factor (EPCO) and the soil evaporation compensation factor (ESCO) are the sensitive parameters. The initial range (Neitsch et al., 2000) and behavioral range of the parameters are listed in Table 8.4.2.

Table 8.4.2. Ranges of different sensitive parameters considered for calibration of SWAT model and optimal values selected for the same.

Sensitive parameter	Initial range	Behavioral range	parameter value selected during calibration
Curve number (CN)	- 20% to + 20% of CN (upper limit 100)	-5% to +5% of CN	75*
AWC	-0.04 to +0.04 of AWC	+1 to +0.04	0.04*
EPCO	0.01 to 1	0.5 to 0.8	0.75
ESCO	0.01 to 1	0.1 to 0.5	0.4

* represents the weighted curve number and AWC for the different combinations of soil and landuse.

The possible options to improve the model performance include decreasing the CN and increasing the AWC, EPCO and ESCO (Neitsch et al., 2002). The model was calibrated empirically to account for retention storage in the region. For each combination of these parameters, the streamflows obtained from the SWAT model are compared with those observed at the Malaprabha reservoir for the calibration period (January 1978-December 1993), in terms of eight model performance indicators (described in section 8.4.2.6) to arrive at optimal set of parameters. Parameters thus obtained are used for the model validation. The correlation between SWAT generated and observed streamflows for the

validation period (January 1994 - December 2000) was found to be 0.98 (Fig. 8.4.4(i) and Table 8.4.3) for the overall time period.

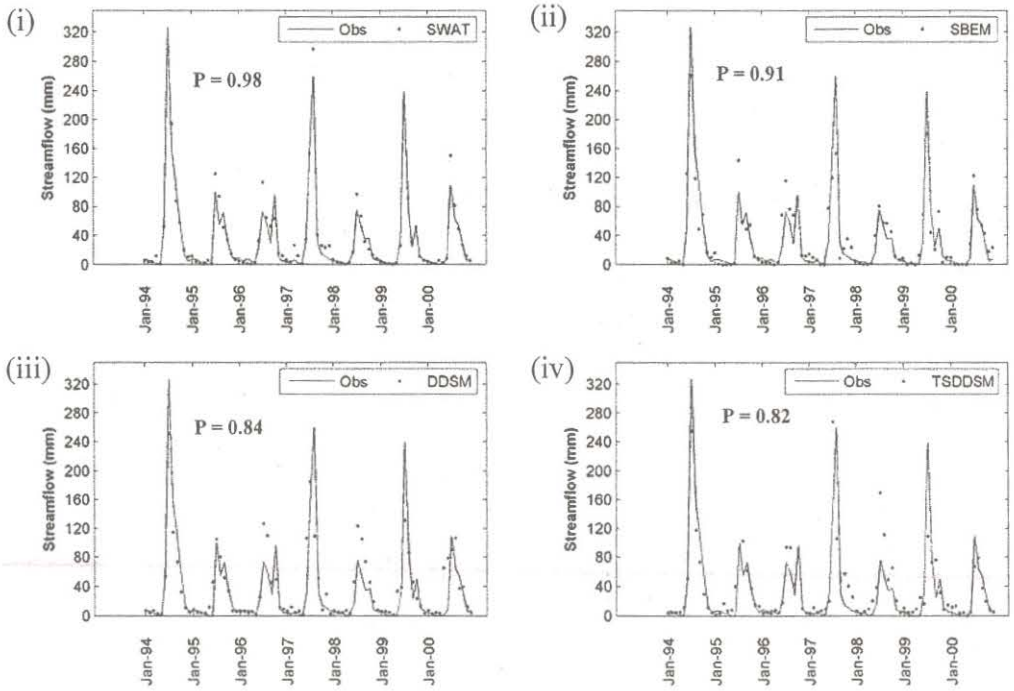

Fig 8.4.4. Comparison of observed and model simulated monthly streamflows into Malaprabha reservoir for the validation period (1994 –2000). Four models namely SWAT, SBEM, DDSM and TSDDSM are used to simulate streamflows. P denotes product moment correlation, also called Pearson's correlation coefficient.

Table 8.4.3. Error statistics for monthly streamflow generated by SWAT, SBEM, DDSM and TSDDSM for the calibration and validation periods. Minimum values of SSE, MSE, RMSE, NMSE and maximum values of E_f, MAE, MCE and *P* indicate optimal values of error statistics.

Time period	Performance Measure	Calibration (1978-1993)				Validation (1994 -2000)			
		SWAT	SBEM	DDSM	TSDDSM	SWAT	SBEM	DDSM	TSDDSM
Overall	SSE	25647.913	95014.228	64383.331	99521.205	15897.076	49049.153	84934.899	95880.647
	MSE	133.583	494.866	335.330	518.340	189.251	583.918	1011.130	1141.436
	RMSE	11.558	22.246	18.312	22.767	13.757	24.164	31.798	33.785
	NMSE	0.048	0.177	0.120	0.185	0.056	0.172	0.298	0.336
	E_f	0.952	0.822	0.880	0.814	0.944	0.826	0.698	0.660
	MAE	0.819	0.654	0.739	0.660	0.771	0.618	0.525	0.509
	MCE	0.961	0.969	0.973	0.975	0.888	0.934	0.901	0.925
	P	0.976	0.908	0.938	0.903	0.976	0.913	0.838	0.815
Monsoon	SSE	20932.072	88910.696	61688.810	95378.471	12727.466	45353.383	74914.366	88121.524
	MSE	327.064	1389.230	963.888	1490.289	454.552	1619.764	2675.513	3147.197
	RMSE	18.085	37.272	31.047	38.604	21.320	40.246	51.725	56.100
	NMSE	0.079	0.337	0.234	0.362	0.073	0.262	0.433	0.509
	E_f	0.919	0.657	0.762	0.632	0.924	0.728	0.551	0.472
	MAE	0.781	0.479	0.577	0.467	0.702	0.432	0.286	0.292
	MCE	0.980	0.979	0.974	0.979	0.897	0.994	0.923	0.999
	P	0.959	0.823	0.874	0.797	0.969	0.868	0.761	0.695
Non-monsoon	SSE	4715.841	6103.533	2694.521	4142.734	3169.610	3695.771	10020.533	7759.122
	MSE	36.843	47.684	21.051	32.365	56.600	65.996	178.938	138.556
	RMSE	6.070	6.905	4.588	5.689	7.523	8.124	13.377	11.771
	NMSE	0.418	0.541	0.239	0.367	0.208	0.243	0.659	0.510
	E_f	0.579	0.454	0.759	0.630	0.788	0.753	0.329	0.480
	MAE	0.351	0.204	0.531	0.315	0.508	0.398	0.271	0.156
	MCE	0.867	0.919	0.964	0.958	0.846	0.709	0.812	0.651
	P	0.793	0.809	0.878	0.803	0.894	0.930	0.620	0.762

SSE: Sum of squares of errors; MSE: mean square error; RMSE: root mean square error, NMSE: normalized mean square error, E_f: Nash-Sutcliffe error estimate; MAE: mean absolute error; MCE: mean cumulative error; *P*: product moment correlation.

Results from SBEM

The SBEM is developed following the procedure given in section 8.4.2.2. During the calibration of SBEM for the period January 1978– December 1993, the optimal SVM parameters σ and C are estimated as 1650 and 2050 respectively, using NMSE (given in section 8.4.2.6) as performance measure. Subsequently, the calibrated model is validated for the period January 1994 - December 2000 using the eight performance measures (given in section 8.4.2.6). It can be noted from the results presented in Table 8.4.3 and Fig. 8.4.4 that the correlation between SBEM generated and observed streamflows for the validation period is 0.91 for the overall time period.

Results from DDSM

The DDSM is developed following the procedure described in section 8.4.2.3. Scatterplots and correlation plots are prepared for selecting potential predictors from LSAV, and the trend in LSAV is also examined. The potential predictors selected based on this analysis are given in Table 8.4.4.

Table 8.4.4. Potential predictors selected as inputs for directly downscaling streamflow using DDSM

Probable Predictors	Potential predictors selected	
	Name	Serial number of 2.5° NCEP grid point shown in Fig. 8.4.1
Ta 925, Ta 700, Ta 500, Ta 200, Zg 925, Zg 500, Zg 200, Hus 925, Hus 850, Ua 925, Ua 200, Va 925, Va 200, prw, ps	Ta 200	2
	Ua 925	1 2 3 4 5 6 7 8 9
	Ua 200	1 2 3 4 5 6 7 8 9
	Va 925	1
	Hus 925	1 2 3 4 6 7
	Hus 850	1 2 3 4 5 6 9
	prw	1 2 3 4 5 6 7 8 9

For the calibration period (January 1978– December 1993), the optimal SVM parameters σ and C are estimated as 39 and 1627 respectively, using NMSE (given in section 8.4.2.6) as performance measure. Adopting these values, the model is validated for the period January 1994 - December 2000 using the eight performance measures (given in section 8.4.2.6). It can be noted from the results presented in Table 8.4.3 and Fig. 8.4.4

that the correlation between DDSM generated and observed streamflows for the validation period (January 1994 - December 2000) is 0.84 for the overall time period.

Results from TSDDSM

The TSDDSM is developed following the procedure described in section 8.4.2.4. Scatter plots and cross-correlations were used for the selection of potential predictors. They indicated (1) no distinct non-linearity between the NCEP- and GCM-downscaled predictor variables, and product moment correlation to be sufficient for the analysis; (2) high correlation between NCEP-downscaled and GCM-downscaled predictors; (3) positive correlation of streamflow with precipitation, relative humidity, wind speed and cloud cover, and negative correlation with maximum temperature. The correlation of minimum temperature with streamflow is found to be negligible.

For the calibration period (January 1978– December 1993), the optimal SVM parameters σ and C are estimated as 37 and 2040 respectively, using NMSE (given in section 8.4.2.6) as performance measure. Adopting these values, the model is validated for the period January 1994 - December 2000 using the eight performance measures (given in section 8.4.2.6). It can be noted from the results presented in Table 8.4.3 and Fig. 8.4.4 that the correlation between the observed monthly streamflows, and those simulated from TSDDSM is 0.82 for the overall time period.

Comparison of model results in reproducing the historical streamflows

The capacity of each of the four models to predict long-term monthly mean streamflow in the Malaprabha basin is evaluated for monsoon and non-monsoon seasons using eight performance measures (described in section 8.4.2.6) and visual comparisons based on time series plots, mean monthly plots and log-log scatter plots (Figs. 8.4.4- 8.4.7). The values of performance measures in Table 8.4.3 indicate that, all the models produced good results. For the monsoon season, among the four models SWAT produced best results for both calibration and validation periods. The performance of SWAT is followed by that of DDSM, SBEM and TSDDSM for the calibration period, whereas for validation period the performance of SWAT is followed by that of SBEM, DDSM and TSDDSM.

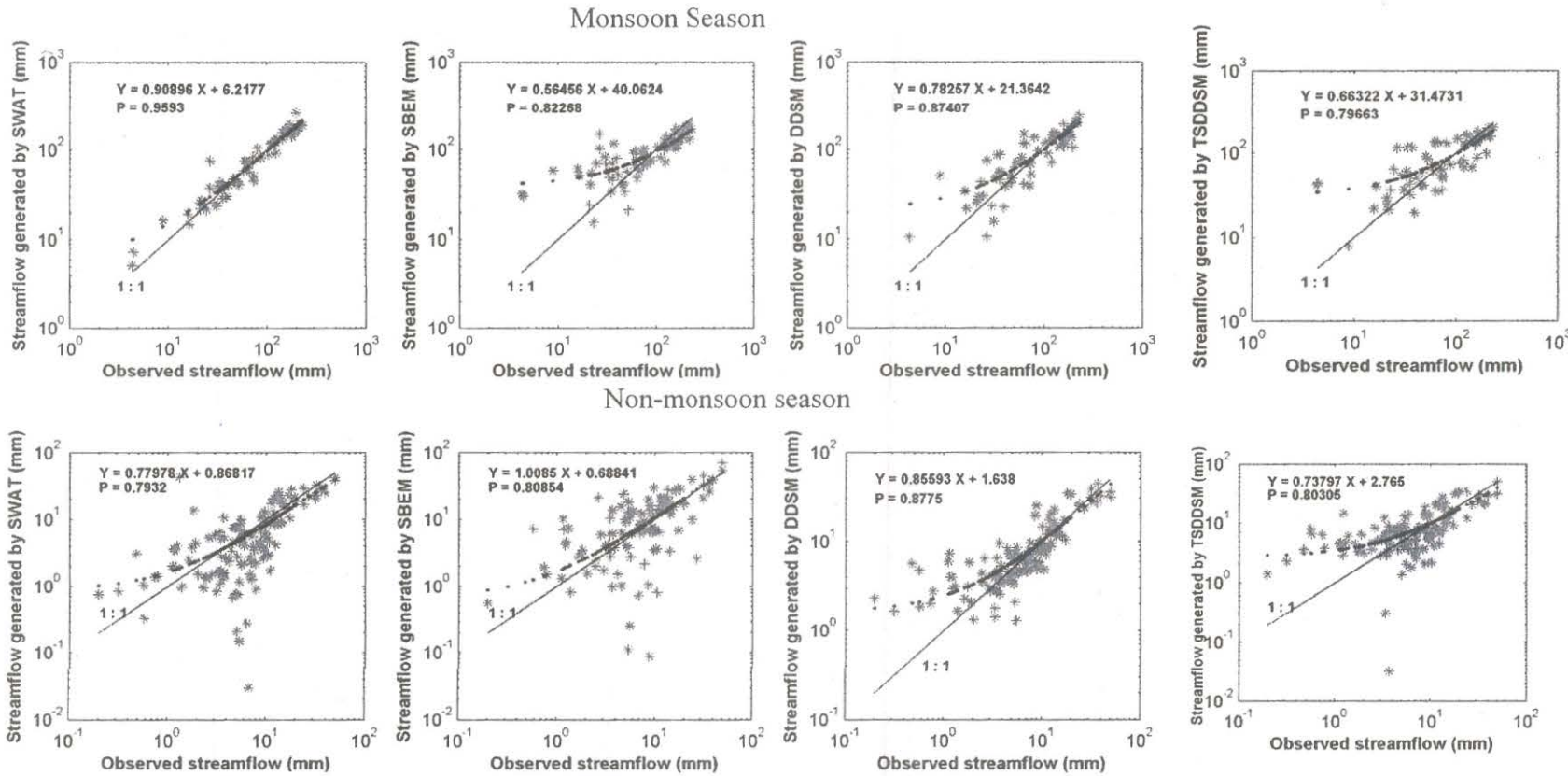

Fig 8.4.6. Scatter plots comparing monthly streamflows for monsoon and non-monsoon seasons predicted by SWAT, SBEM, DDSM and TSDDSM with the observed streamflows for the calibration period 1978-1993. Regression equation between the generated and observed streamflows, and product moment correlation (P) between the same are also shown.

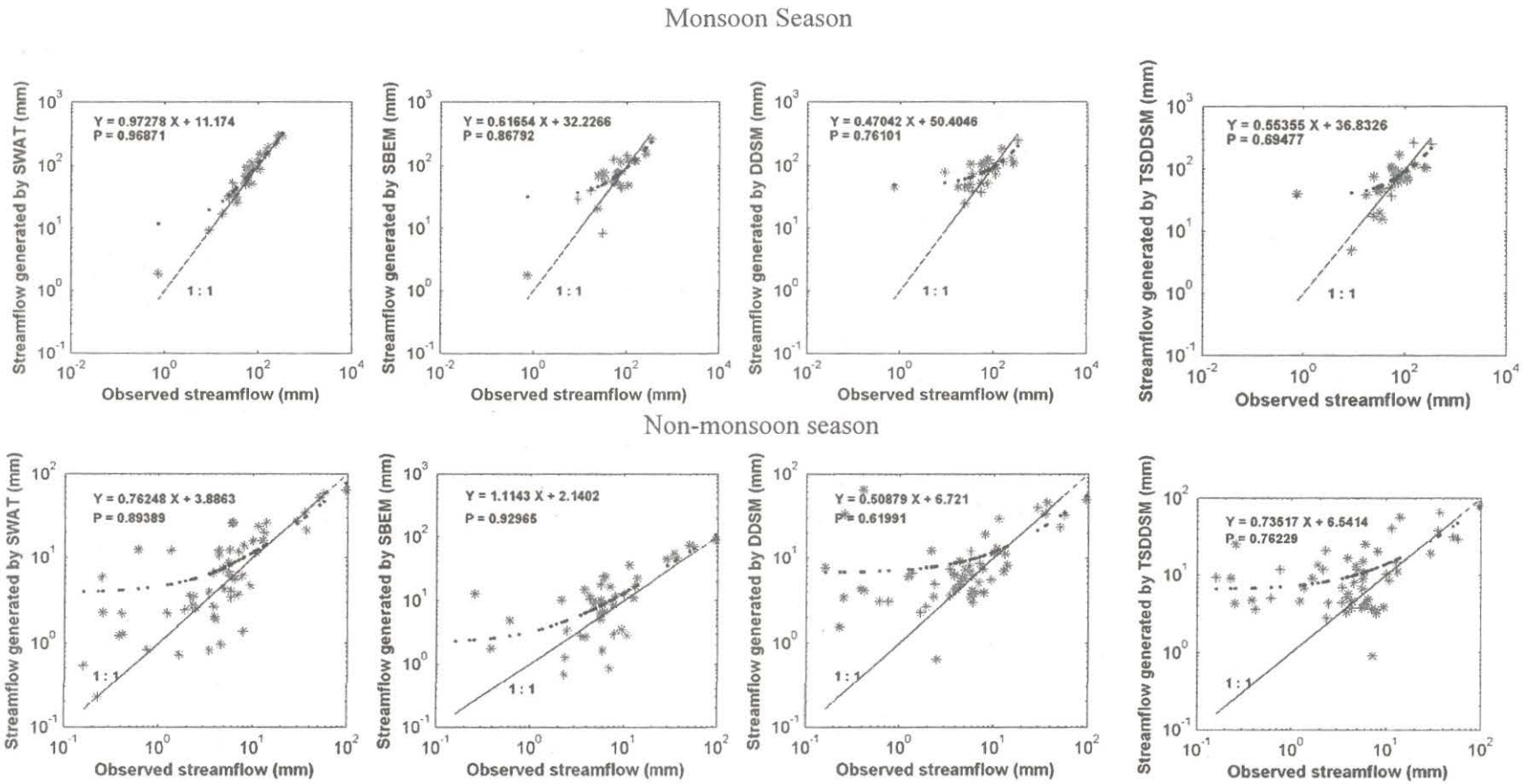

Fig 8.4.7. Scatter plots comparing monthly streamflows for monsoon and non-monsoon seasons, predicted by SWAT, SBEM, DDSM and TSDDSM with the observed streamflows for the validation period 1994 - 2000. Regression equation between the generated and observed streamflows, and product moment correlation (P) between the same are also shown.

In order to compare the models' capabilities in simulating the dynamics of monthly streamflow series, the monthly streamflow values generated using different models are analyzed and correlated with observed values using a linear regression equation $Y = mX + c$. In the equation, Y represents the model generated streamflow, and X is the observed streamflow, and m and c are constants representing the slope and intercept, respectively. The log-log scatter plots and the results of regression analysis are shown in Figs. 8.4.6 and 8.4.7. Scrutiny of results shows that:

(1) For all the four models, generated streamflows correlated well with observed streamflows. Among the models, SWAT's performance was the best (with P = 0.96) during monsoon season, while SBEM's performance was the best during the non-monsoon season, in both calibration and validation periods. For the calibration period, the performance of SWAT is followed by that of DDSM, SBEM and TSDDSM, which had P values of 0.87, 0.82 and 0.80 respectively. The SWAT and SBEM (methods of category 3) performed better than DDSM and TSDDSM (methods of category 4) during monsoon season in validation period. In general, during non-monsoon season in calibration and validation periods, the P values obtained for models are greater than 0.75 in all cases except DDSM for which P is equal to 0.62.

(2) The differences in values of slope and intercept are not significant for the models. The value of intercept is generally less than 15% of the mean seasonal streamflow value for all the models. SWAT has the least value of intercept for monsoon season, while SBEM has the least value of intercept for non-monsoon season during the calibration and validation periods. The slopes are all within ±11% of the 1:1 line for SWAT for the monsoon season and SBEM for the non-monsoon season. A close look at the differences among the models reveals that the SWAT and SBEM models have smaller bias in the regression slope and intercept, when compared to DDSM and TSDDSM.

The results show that among the four models, SWAT, SBEM and DDSM can reproduce historical (observed) monthly streamflow series with an acceptable accuracy compared to TSDDSM. It must, however, be noted that the purpose here is not to discuss in detail which model is superior in simulating historical streamflow values. The main purpose is

to check how diverse the model results are with respect to historical and future climate scenarios.

Comparison of model results in predicting future streamflows for changed climates

The mean annual, mean monsoon and mean non-monsoon streamflows projected in Malaprabha catchment for four climate change scenarios (A1B, A2, B1 and COMMIT) are analyzed for the period 2001-2100. For this purpose, the future projections of streamflow for each scenario are divided into five parts (2001-2020, 2021-2040, 2041-2060, 2061-2080 and 2081-2100). For each part, the percent change in projected mean streamflow with respect to its historical counterpart is computed for the four climate change scenarios. The results are shown in Figs. 8.4.8 and 8.4.9, and Table 8.4.5.

Table 8.4.5. Percentage change in mean annual, mean monsoon and mean non-monsoon streamflows projected with respect to their historical counterparts for the different models and scenarios used (A1B, A2, B1 and COMMIT).

Season	Model Name	Percentage change in streamflow from 2001-2100			
		A1B	A2	B1	COMMIT
Annual	SWAT	5 to 108	5 to150	12 to 60	-3 to 20
	SBEM	1 to 44	-1 to 65	7 to 19	-2 to 8
	DDSM	0 to 37	-4 to 59	1 to 13	0 to -2
	TSDDSM	11 to 364	14 to 581	14 to 175	-6 to 24
Monsoon	SWAT	-19 to 57	-17 to 96	-16 to 21	-24 to -13
	SBEM	0 to 46	1 to 69	3 to 20	-5 to 3
	DDSM	-7 to 30	-10 to 53	-2 to 12	-7 to5
	TSDDSM	13 to 403	16 to 641	12 to 198	-8 to 28
Non-monsoon	SWAT	117 to 346	105 to 391	142 to 245	98 to 174
	SBEM	9 to 32	-9 to 46	26 to 16	13 to 30
	DDSM	33 to 71	21 to 88	16 to 18	29 to 11
	TSDDSM	-1 to 185	6 to 34	25 to 63	2 to 5

In Fig. 8.4.8 the horizontal axis is divided into four branches. Each branch represents the result for one climate change scenario and three models. The three lines shown against each branch represent the results of SWAT, SBEM and DDSM models. It can be seen form the figure that at the annual level there is a wide range of differences between streamflows simulated by the models for each climate scenario considered. Results of

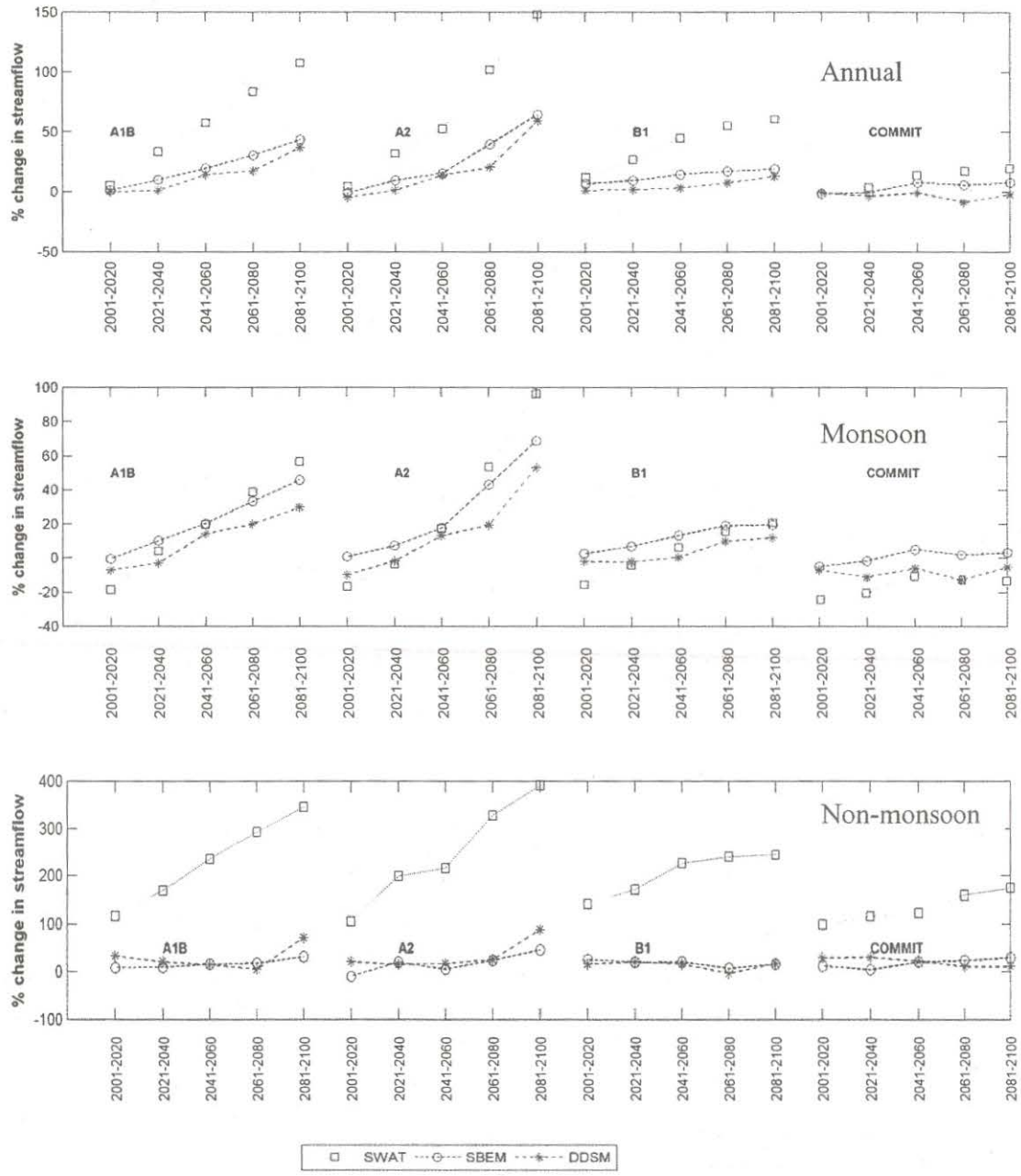

Fig 8.4.8. Comparison of changes projected across models (SWAT, SBEM and DDSM) for mean annual, mean monsoon and mean non-monsoon streamflows for SRES A1B, A2, B1 and COMMIT scenarios.

The percent change in projected mean streamflow with respect to its historical counterpart for all the four scenarios are plotted for each of the three models namely SWAT, SBEM and DDSM for mean annual, mean monsoon and mean non-monsoon streamflows in Fig. 8.4.9. It can be observed from these plots that, among the three models, SWAT model has the maximum variability between the four scenarios considered at annual, monsoon and non-monsoon seasons. The variability of SWAT is followed by that of SBEM and DDSM. It can be observed that the increase in trend is more after 2060 for A1B and A2 scenarios.

The projected increase in streamflow is high for A2 scenario, whereas it is least for B1 scenario. The scenario A2 has the highest concentration of equivalent carbon dioxide (CO_2) equal to 850 ppm, while the same for A1B, B1 and COMMIT scenarios are 720 ppm, 550 ppm and ≈ 370 ppm respectively. Rise in the concentration of equivalent CO_2 in the atmosphere causes the earth's average temperature to increase, which in turn causes increase in evaporation especially at lower latitudes. The evaporated water would eventually precipitate. Increase in precipitation results in increased streamflow. In the COMMIT scenario, where the emissions are held the same as in the year 2000, no significant trend in the projected future streamflow could be discerned.

The statistical significance of trend in the projected streamflow scenarios are assessed using null hypothesis considering 99% confidence level. For the test it is assumed that the variances of past and projected streamflows are unknown and unequal. In the null hypothesis the mean streamflow for the past (1978-2000) and the projected future periods (2001-2020, 2021-2040, 2041-2060, 2061-2080 and 2081-2100) are assumed to be equal. The test statistic, T, is computed using Eq. (8.4.22) (Kottegoda and Rosso, 1998), which has an approximate t-distribution with υ degrees of freedom given by Eq.(8.4.23).

$$T = \frac{(\bar{y} - \bar{\hat{y}}_p)}{\sqrt{(\sigma_y^2/N) + (\sigma_p^2/N_p)}} \qquad (8.4.22)$$

$$\upsilon = \frac{[(\sigma_y^2/N)+(\sigma_p^2/N_p)]^2}{\left[(\sigma_y^2/N)^2/(N-1)\right]+\left[(\sigma_p^2/N_p)^2/(N_p-1)\right]} \quad (8.4.23)$$

where $\bar{y}$ and $\bar{\hat{y}}_p$ are estimated means of the observed and the projected streamflows respectively, σ_y^2 and σ_p^2 are respectively the standard deviations of the observed and projected streamflows, N_p (=20) is sample size of projected streamflows considered for estimating T.

Results are presented at annual scale in Table 8.4.6. It can be seen that on annual scale, the streamflow is projected to increase during 2021-2100 for A1B, A2, and B1 scenarios, and during 2061-2100 for COMMIT scenario by SWAT. In the case of SBEM, the streamflow is projected to increase during 2041-2100 for A1B scenario, and during 2061-2100 for A2 and B1 scenarios. In the case of DDSM the streamflow is projected to increase during 2061-2100 for A2 scenario, and during 2081-2100 for A1B scenario. No trend was discerned with B1 and COMMIT scenarios.

Table 8.4.6. Results of trend analysis for annual streamflows using null hypothesis considering 99% confidence level. The symbols + and 0 denote 'increase' and 'no change' respectively.

Model	Scenario	Duration (in years)				
		2001-2020	2021-2040	2041-2060	2061-2080	2081-2100
SWAT	SRES A1B	0	+	+	+	+
	SRES A2	0	+	+	+	+
	SRES B1	0	+	+	+	+
	COMMIT	0	0	0	+	+
SBEM	SRES A1B	0	0	+	+	+
	SRES A2	0	0	0	+	+
	SRES B1	0	0	0	+	+
	COMMIT	0	0	0	0	0
DDSM	SRES A1B	0	0	0	0	+
	SRES A2	0	0	0	+	+
	SRES B1	0	0	0	0	0
	COMMIT	0	0	0	0	0

8.4.2.9 Summary

Owing to the availability of a number of GCMs, climate change scenarios and methods used for translating GCM information from coarser to finer scale, there is uncertainty in their choice. In this section, a comparison is presented between future projections of streamflows in Malaprabha catchment of Krishna river basin, which are obtained for four models (two each based on methods of categories 3 and 4), and four climate change emission scenarios (A1B, A2, B1 and COMMIT). Simulations for each of the scenarios are obtained from the Canadian third generation coupled GCM. The downscaling methods of category 3 use hydrological models, whereas those of category 4 do not use a hydrological model. Support vector machine (SVM) is chosen as downscaling technique, owing to its inherent advantages.

The two hydrological models considered to determine future projections of streamflows by methods of category 3 are Soil and water assessment tool (SWAT) and support vector machine (SVM) based empirical model (SBEM). The two models developed based on methods of category 4 involved use of SVM for downscaling LSAV. One of the two models directly downscales LSAV to streamflows in the target watershed, and is referred to as direct downscaling model (DDSM). The other model downscales LSAV to streamflows in two stages, and is called two-stage deterministic downscaling model (TSDDSM).

Among the models, SWAT's performance was found to the best during monsoon season, while SBEM's performance was the best during the non-monsoon season, in both calibration and validation periods. For the calibration period, the performance of SWAT is followed by that of DDSM, SBEM and TSDDSM. The SWAT and SBEM (methods of category 3) performed better than DDSM and TSDDSM (methods of category 4) during monsoon season in validation period.

On an annual scale, the streamflow is projected to increase during 2021-2100 for A1B, A2, and B1 scenarios, and during 2061-2100 for COMMIT scenario by SWAT. In the case of SBEM, the streamflow is projected to increase during 2041-2100 for A1B scenario, and during 2061-2100 for A2 and B1 scenarios. In the case of DDSM the

streamflow is projected to increase during 2061-2100 for A2 scenario, and during 2081-2100 for A1B scenario and no trend was discerned with B1 and COMMIT scenarios.

A large uncertainty is observed in predictions of future streamflows in the river basin due to the choice of method for generating streamflow projections, climate scenario and hydrological model considered, even though the models perform equally well in reproducing the monthly historical streamflows.

The present study brings out the uncertainties existing in the choice of method for generating streamflows, climate scenario and hydrological model considered. The uncertainties in the choice of downscaling methods and GCMs should also be considered to provide more general conclusions on change in hydrological response of the river basin in climate change scenarios. Extended research in this direction is underway.

9. Conclusions and Recommendations for Future Studies

Methodologies are developed in this project work to assess the impact of climate change on rainfall and streamflow at river basin scales, and to address the uncertainties involved in such assessment. Results are provided with an aim to help the policy makers in developing adaptive responses to climate change. Specifically the following problems have been dealt with in this project:

a. Impact of climate change on Mahanadi streamflow,
b. Impact of climate change on rainfall and meteorological droughts in Orissa meteorological subdivision,
c. Impact of climate change on streamflow of Malaprabha river, in the Krishna riverbasin, and
d. Uncertainties associated with the future projections.

The following conclusions are drawn from the studies:

1. The methodologies for downscaling and addressing uncertainties developed in this project, doe not limit their applicability only for the specific regions and specific variables. They are adaptable and can be used to model any other hydrologic variable also and in any other region in the

country (and elsewhere) to assess the impact of climate change on hydrology.

2. The downscaling model developed based on the RVM is capable of producing a satisfactory value of goodness of fit in terms of R value and Nash–Sutcliffe coefficient. However, it is observed that the model is not able to reproduce the extreme streamflow observed in the record. Possibly this could be because regression based statistical downscaling models often cannot explain entire variance of the downscaled variable.
3. The GCM CCSR/NIES with B2 scenario projects a decreasing trend in future monsoon streamflow of Mahanadi, under climate change. This observation is consistent with an earlier study reported by Rao (1995). Such a decrease in streamflow may cause a critical situation for Hirakud dam in meeting the future irrigation and power demand.
4. The streamflow in Malaprabha river is projected to increase on an annual scale during 2021-2100 for A1B, A2, and B1 scenarios, and during 2061-2100 for COMMIT scenario by the SWAT model. In the case of SBEM model, the streamflow is projected to increase during 2041-2100 for A1B scenario, and during 2061-2100 for A2 and B1 scenarios. In the case of DDSM model, the streamflow is projected to increase during 2061-2100 for A2 scenario, and during 2081-2100 for A1B scenario and no trend was discerned with B1 and COMMIT scenarios. Thus, with all the models and most scenarios considered, the streamflow is projected to increase in Malaprabha river, although the extent of increase differs from projections of one model to another.
5. A large number of uncertainties exist in climate change impact assessments. These uncertainties stem from various GCMs available and future plausible climate scenarios used. The downscaling relationships themselves introduce another source of uncertainty. To address such uncertainties, results are provided through possible changes in the probability distributions of the meteorological drought and the streamflow.

6. The results indicate an increasing trend in the probability of severe and extreme drought for Orissa meteorological subdivision with a decrease in the probability of near normal condition. It may be concluded from the results that the region will be more drought prone due to the effect of climate change.
7. The results show that the CDF of Mahanadi 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.
8. An increased dissimilarity between the GCMs with time, is observed in projecting Mahanadi streamflow. The amount of uncertainty in 2080s is higher than those of the other time slices. This may point to different climate sensitivity among the models due to ignorance about the underlying geophysical processes. Such ignorance is addressed here with possibility theory.
9. For water resources management it is important to know the effectiveness of the GCMs in modeling climate change and which of the scenarios best represent the present situation under global warming. Possibilities are assigned to GCMs and scenarios based on their system performance measure in predicting the streamflow during years 1991-2005, when signals of climate forcing are visible. A decreasing trend in future monsoon streamflow is projected for the Mahandi even with a possibility weighted CDF.
10. A limitation of the work presented here is that the methodologies do not consider the uncertainty due to the use of multiple downscaling models. Another limitation of the work is that the Third Assessment Report (TAR) data have been used in most models developed in the present study, except the ones used for Malaprabha streamflow, which use the recently released Assessment Report 4 (AR4) data.

Recommendations for Future Studies

Starting with the methodologies and the results provided in this report, the following logical extensions are possible, and should be taken up:

(a) Developing adaptive long term policies for reservoir operation with the probability distributions as input, to offset the impact of climate change.

(b) Assessing impacts of climate change on irrigation water demands, with the methodologies of downscaling and uncertainties presented in the work.

(c) Assessing impacts on river floods, by developing relationships between climate variables and historical floods and using such relationships with future projections from climate models, and

(d) Reworking the Intensity-Duration-Frequency (IDF) relationships for major river basins in India under climate change scenarios.

10. Publications Resulting from the Work Carried out in the Project

1. Anandhi, A, V.V. Srinivas, R.S. Nanjundiah, D. Nagesh Kumar (2008) Downscaling Precipitation to River Basin in India for IPCC SRES Scenarios Using Support Vector Machine, International Journal of Climatology, Wiley InterScience on behalf of Royal Meteorological Society (RMetS), Vol. 28, No. 3, March 2008, pp. 401-420, DOI: 10.1002/joc.1529

2. Anandhi, A, V.V. Srinivas, D. Nagesh Kumar and R.S. Nanjundiah (2009) Role of Predictors in Downscaling Surface Temperature to River Basin in India for IPCC SRES Scenarios using Support Vector Machine, International Journal of Climatology, Wiley InterScience on behalf of Royal Meteorological Society (RMetS), Vol. 29, No. 4, Mar 2009, pp. 583-603, DOI: 10.1002/joc.1719

3. Aavudai Anandhi, V.V. Srinivas and D. Nagesh Kumar, (2009) Water Resources Assessment in a River Basin using GIS and DEM, In GeoHydroinformatics: Integrating GIS and Water Engineering, Eds: Suchith Anand, Mark Ware, Mike Jackson, Kalanithy Vairavamoorthy, Robert J. Abrahart, CRC Press, UK, ISBN 9781420051209, in press.

4. Ghosh, S. and Mujumdar P.P. (2006) Future rainfall scenario over Orissa with GCM projections by statistical downscaling. *Current Science*, 90(3), 396-404.

5. Ghosh and Mujumdar (2007) Statistical downscaling of GCM simulations to streamflow using Relevance Vector Machine. *Advances in Water Resources* 31/1 pp. 132-146 *doi :* 10.1016/j.advwatres. 2007.07.005

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

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

Appendix 1. Support vector regression

The basic concept of SV regression is discussed in the present section first with a linear model and then it is extended to a nonlinear model using Kernels. Given a training data $\{(x_1, y_1), ..., (x_l, y_l), X \in \Re^n, Y \in \Re\}$, the SV regression eqn is be given by (Smola, 1996)

$$y = f(x) = \langle w, x \rangle + b, \qquad w \in X,\ b \in \Re \tag{A.1}$$

where, $\langle \cdot, \cdot \rangle$ denoted the dot product in X. The objective of SVM regression is to find the function f(x) with minimum value of loss function and at the same time is as flat as possible (Smola and Schoelkopf, 1998). Flatness mathematically denotes the smaller value of w and one way to ensure this is to minimize the norm, i.e. $\|w\|^2 = \langle w, w \rangle$. Thus the model can be expressed as the following convex optimization problem.

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C\left(\sum_{i}^{l} \xi_i^* + \sum_{i=1}^{l} \xi_i\right) \tag{A.2}$$

$$\text{Subject to} \quad y_i - \langle w, x \rangle - b \leqslant \varepsilon + \xi_i \tag{A.3}$$

$$\langle w, x \rangle + b - y_i \leqslant \varepsilon + \xi_i^* \tag{A.4}$$

$$\xi_i, \xi_i^* \geqslant 0 \tag{A.5}$$

where C is a pre-specified value which determines the tradeoff between the flatness of *f(x)* and the amount up to which deviations larger than ε is tolerated (ξ_i and ξ_i^*), which correspond to ε -insensitive loss function as presented in Eq. (8.1.6). The optimization model presented in Eqs. (A.2)–(A.5) can be solved using Lagrange multipliers. A dual set of variables are introduced to construct the Lagrange function, which is given below:

$$\begin{aligned} L = {} & \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l} \xi_i^* + \sum_{i=1}^{l} \xi_i\right) - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^{l} \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x \rangle + b) \\ & - \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x \rangle - b) \end{aligned} \tag{A.6}$$

where L is the Lagrangian and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrangian multipliers satisfying the positivity constraints:

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0 \tag{A.7}$$

From the saddle point condition, the partial derivatives of L with respect to the primal variables $\left(w, b, \xi_i, \xi_i^*\right)$ have to vanish for optimality:

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 \tag{A.8}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i = 0 \tag{A.9}$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{A.10}$$

where $\xi_i^{(*)}, \alpha_i^{(*)}, \eta_i^{(*)}$ refer to ξ_i and ξ_i^*; α_i and α_i^*; η_i^* and η_i^* respectively.

Substituting Eqs. (A.8)–(A.10) in Eq. (A.6) the following dual optimization problem is formulated.

Maximize

$$-\frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*) \tag{A.11}$$

subject to

$$\sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \tag{A.12}$$

$$\alpha_i, \alpha_i^* \in [0, C] \tag{A.13}$$

Equation (A.9) can be re-written as

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i \tag{A.14}$$

and thus, from Eq (A.1):

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{A.15}$$

This is called the Support Vector Expansion for linear model which is used in SV regression. b can be computed by using Karush Kuhn Tucker (KKT) condition (Smola and Schoelkopf, 1998).

For most of the hydrologic analysis linear regression is not appropriate and thus a nonlinear mapping using kernel K is used to map the data into a higher dimensional feature space, where, with the kernel, linear analysis is performed. Using the kernel, the regression equation (Eq. (A.15)) can be modified to (Eq. (8.1.5))

Appendix 2. Kernel functions

Kernel functions are used in SVM for nonlinear mapping of the original data or input into a high dimensional feature space. Kernel function used in a SVM should follow Mercer's theorem, according to which it can be written that:

$$\int_{X\times X} K(x,x')f(x)f(x')\mathrm{d}x\,\mathrm{d}x' \geqslant 0 \quad \forall f \in L_2(X) \tag{A.16}$$

Some of the valid kernel functions satisfying the above mentioned condition are given below.

A. *Linear kernel*: The linear kernels are the simplest kernels used in SVM for linear regression. They can be given by

Homogeneous kernel:

$$K(x,x') = \langle x,x' \rangle \tag{A.17}$$

Non-homogeneous kernel:

$$K(x,x') = (\langle x,x' \rangle + 1) \tag{A.18}$$

The performance of SVM with linear kernel function, being similar to that of linear regression, is not capable of modeling complicated and nonlinear relationship between climatological variables and streamflow and therefore such kernels are not used in the present study.

B. *Gaussian Radial Basis Function*: Radial Basis Functions (RBFs) have received significant attention, most commonly with Gaussian form,

$$K(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \tag{A.19}$$

where σ is the width of Gaussian RBF kernel, giving an idea about the smoothness of the derived function. A large kernel width acts as a low-pass filter in frequency domain, attenuating higher order frequencies and thus resulting in a smooth function. Alternatively, RBF kernel with small kernel width retains most of the higher order

frequencies leading to an approximation of a complex function by learning machine (Smola and Schoelkopf, 1998).

C. *Laplacian or Exponential Radial Basis Function*: Laplacian or Exponential RBF of the form,

$$K(x,x') = \exp\left(-\frac{\|x-x'\|}{2\sigma^2}\right) \tag{A.20}$$

produces a piecewise linear solution which can be attractive when discontinuities are acceptable.

D. *Heavy tailed or Sublinear Radial Basis Function*: Heavy tailed RBFs or Sublinear RBFs are introduced by Chapelle et al. (1999) which sometimes outperform traditional Gaussian or RBFs. They can be given by:

$$K(x,x') = \exp\left(-\frac{\|x-x'\|^{0.5}}{2\sigma^2}\right) \tag{A.21}$$

It is worth mentioning that a generalized RBF can be given by

$$K(x,x') = \exp\left(-\frac{\|x^a-x'^a\|^b}{2\sigma^2}\right) \tag{A.22}$$

and it will satisfy Mercer's condition if and only if $0 \le b \le 2$. The choice of a has no impact on Mercer's condition.

Appendix 3: Algorithm for Density Estimation using Orthonormal series

The algorithm for density estimation using orthonormal series involves the following steps:

Determination of Support and Scaling of Data-set

The methodology presented for estimation of pdf using orthonormal systems is valid when the bound of random variable is $[0,1]$. The random variable of interest may have different bounds (say, $[a,b]$) and thus need to be converted to a variable y having interval of [0,1] by scaling the data. The scaled variable y may be given by:

$$y = (x-a)/(b-a) \tag{A.23}$$

Considering the minimum and maximum values from the data set as the two bounds ($[a,b]$) is not a realistic method as there is no guarantee that unobserved values will not

cross these bounds. Methodology for determination of support from a data set may be found in Efromovich (1999). According to that methodology, if x_1, x_2,...., x_n are ordered observations ($x_1 \le x_2 \le .... \le x_n$), then

$$a = x_1 - d_1 \tag{A.24}$$

$$b = x_n + d_2 \tag{A.25}$$

where,

$$d_1 = (x_{1+s} - x_1)/s; \quad and \quad d_2 = (x_n - x_{n-s})/s \tag{A.26}$$

s is a small positive integer assuming that the density is flat near the boundaries of its support. The default value of s is 1, which is considered in the present work.

Estimation of Orthonormal Series with Coefficients

The Functions involved in the orthonormal series can be obtained using Eq. (8.3.23-8.3.24). The coefficients $\theta_j, j = 0,1,...$ presented in Eq. (8.3.22) can be given by Eq. (A.27), where f(y) is the pdf of the scaled random variable Y.

$$\theta_j = \int_{-\infty}^{\infty} f(y)\Phi_j(y)dy \tag{A.27}$$

It follows that θ_j is the expected value of $\Phi_j(y)$, (Eq. A.28), which in turn may be approximated from a finite sample of n observations ($y_j, j = 1, n$) as (Eq. A.29).

$$\theta_j = E[\Phi_j(y)] \tag{A.28}$$

$$\theta_j = \frac{1}{n}\sum_{l=1}^{n}\Phi_j(y_l) \tag{A.29}$$

In our case, where Y is the scaled SPI-12 with a bound [0,1], the y_j are the scaled values of SPI-12 simulated from each climate model.

Estimation of Cutoff J

Determination of an appropriate cut-off J (Eq. 8.3.22) is important in the method based on orthonormal series. The choice of J depends on the goodness of fit, which can be determined by Mean Integrated Square Error (MISE) using Parseval's inequality. Following Effromovich (1999) J can be computed as:

$$J = \underset{0 \le \tilde{J} \le J_n}{\mathbf{argmin}} \quad \sum_{j=0}^{\tilde{J}}(2d_j n^{-1} - \theta_j^2) \tag{A.30}$$

where $J_n = [C_{J0} + C_{J1}\ln(n)]$. The default values of C_{J0} and C_{J1} are 4 and 0.5 (Efromovich, 1999).

Smoothing of Estimated pdf

In many cases it is worthwhile to smooth the fourier coefficients by multiplying them with some constants that takes values between 0 and 1. After smoothing a modified pdf is given by:

$$\tilde{f}_J(y) = \sum_{j=0}^{J} w_j \theta_j \Phi_j(y),\ 0 \le y \le 1,\ 0 \le w_j \le 1 \qquad \text{(A.31)}$$

The weights (w_j) used for smoothing fourier coefficients may be given by:

$$w_0 = 1; \qquad \text{(A.32)}$$

$$w_j = (1 - d/(n\theta_j^2))_+ \qquad \forall j > 0 \qquad \text{(A.33)}$$

Here $(1 - d/(n\theta_j^2))_+ = \max(0, (1 - d/(n\theta_j^2)))$, i.e. the positive part of $(1 - d/(n\theta_j^2))$. Other than first J number of fourier coefficients, a density function also requires a relatively large number of coefficients for a fair visualization. Thus high frequency terms are added, which are shrunk by a hard threshold procedure. After adding these extra terms Eq. (A.31) is modified to:

$$\tilde{f}_J(y) = \sum_{j=0}^{J} w_j \theta_j \Phi_j(y) + \sum_{j=J+1}^{C_{JM} \times J_n} I_{\{\theta_j^2 > C_T d \ln(n)/n\}} \theta_j \Phi_j(y) \qquad \text{(A.34)}$$

where, C_{JM} and C_T are parameters for hard threshold procedure that define the maximal number of elements included in the estimate of pdf. The default values are 6 and 4 respectively (Efromovich, 1999). I is an indicator variable such that $I_{\{A\}}$ has the value of 1 if A is true, zero otherwise. A high frequency term is included only if the corresponding Fourier coefficient is extremely large and thus the procedure does not reduce the smoothness of the estimate.

Modification for Area under pdf and Negative Values

An improvement in the pdf, thus estimated is necessary when it takes negative values at some of the points/ regions. For such cases, the following algorithm is developed in the present study, which ensures that the properties of pdf are satisfied by the estimated pdf.

1. If there exists negative values at some points, find the maximum negative value.

2. Add this to $\tilde{f}_J(y)$ to make the value of the function positive, everywhere.

3. Check the area under the curve. If it is less than 1, find out c by numerical methods, such that

$$\int_{-\infty}^{\infty}(\tilde{f}_J(y)+c)dy=1 \qquad \text{(A.35)}$$

$\tilde{f}_J(y)$ is now modified by adding c, as obtained from Eq (A.37), to it.

$$\tilde{f}_J(y)=\tilde{f}_J(y)+c \qquad \text{(A.36)}$$

4. If the area under the curve is greater than 1, find out $c1$ in a similar procedure:

$$\int_{-\infty}^{\infty}(\tilde{f}_J(y)-c1)dy=1 \qquad \text{(A.37)}$$

Subtracting $c1$ from $\tilde{f}_J(y)$ may lead to a negative value of pdf at some points which is not desirable. In such cases after the subtraction take only the positive part of $\tilde{f}_J(y)$.

$$\tilde{f}_J(y)=(\tilde{f}_J(y)-c1)_+ \qquad \text{(A.38)}$$

where $(\tilde{f}_J(y)-c1)_+=\max(0,(\tilde{f}_J(y)-c1))$. Check the area again and if it is not nearly equal to 1, go to step 3, else stop.

The other way of making adjustment in the estimated pdf to make the area under the curve equal to 1, is the use of multiplicative factor. In either case, the result will come almost similar as this adjustment methodology is not much sensitive to the final pdf.

Estimation of pdf for Unscaled Data-set

The scaled data/observations are distributed according to a density $f_Y(y)$, where $y\in[0,1]$.

The pdf thus obtained corresponds to the scaled data set y over the interval of $[0,1]$. The estimate of $f_X(x)$ of orginal dataset x over interval $[a,b]$ may be given by:

$$f_X(x)=(b-a)^{-1}f_Y(y);\quad x\in[a,b] \qquad \text{(A.39)}$$

APPENDIX 4

STATEMENT OF RECEIPT AND EXPENDITUE
IN RESPECT OF THE SCHEME
ASSESSMENT OF WATER RESOURCES UNDER CLIMATE CHANGE SCENARIOS AT RIVER BASIN SCALE
UPTO: 29.02.2009

Scheme code : MWR0003
Investigator's Name : Dr. P.P.MUJUMDAR
Duration of Scheme : 28/02/2006 to 27/02/2009
Scheme Reference No. : 23/52/2006-R&D/287-98 dtd.23.01.06

Sanction Heads	Amount of Sanction	<..... Expenditure.....> 2005-2006	2006-2007	2007-2008	2008-2009*	Commitment	Total Expenditure	Sanction Balance
SALARY	2,65,000-00	0-00	54,333-00	31,767-00	1,43,000-00	0-00	2,29,100-00	35,900-00
EQUIPMENT	2,80,000-00	0-00	2,72,872-00	4,004-00	0-00	0-00	2,76,876-00	3,124-00
EXPERIMENTAL CHARGES	2,97,000-00	0-00	38,500-00	1,06,093-00	1,20,912-00	0-00	2,65,505-00	1,495-00
TRAVEL	68,000-00	0-00	19,828-00	33,996-00	44,176-00	0-00	98,000-00	0-00
TOTAL	9,10,000-00	0-00	3,85,533-00	1,75,860-00	3,08,088-00	0-00	8,69,481-00	40,519-00

* PROVISIONAL & SUBJECT TO AUDIT

Total Receipts : 9,10,000-00
Total Expenditure : 8,69,481-00*
Receipts Balance : 40,519-00

P.P. Mujumdar
PROJECT INVESTIGATOR

ACCOUNTS OFFICER

References

Allen, M. R., P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. L. Delworth. 2000. Quantifying the uncertainty in forecasts of anthropogenic climate change, Nature, 407, 617– 620.

Akhtar, M., Ahmad, N., Booij, M.J. 2008 The impact of climate change on the water resources of Hindukush–Karakorum–Himalaya region under different glacier coverage scenarios. Journal of Hydrology 355, 148– 163

Alley, R., Berntsen, T., Bindoff, N.L., Chen, Z., Chidthaisong, A., Friedlingstein, P., Gregory, J., Hegerl, G., Heimann, M., Hewitson, B., Hoskins, B., Joos, F., Jouzel, 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.F., Stott, P., Stouffer, R.J., Whetton, P., Wood, R.A., Wratt, D. 2007. Climate Change 2007: The Physical Science Basis. Summary for Policymakers. IPCC eds. Intergovernmental Panel on Climate Change. Geneva, CH.

Anandhi, A. 2007. Impact assessment of climate change on hydrometeorology of Indian river basin for IPCC SRES scenarios. PhD thesis, Indian Institute of Science, India.

Anandhi, A., Srinivas, V.V., Nanjundiah, R.S., Kumar, D.N., 2008. Downscaling Precipitation to River Basin in India for IPCC SRES Scenarios using Support Vector Machine. International Journal of Climatology 28(3), 401-420.

Anandhi, A., Srinivas, V.V., Kumar, D.N., Nanjundiah, R.S., 2009. Role of Predictors in Downscaling Surface Temperature to River Basin in India for IPCC SRES Scenarios using Support Vector Machine. International Journal of Climatology 29(4), 583–603.

Andreadis, K. M. and D. P.Lettenmaier. 2006. Trends in 20th century drought over the continental United States, Geophys. Res. Lett., 33(10), L10403, doi:10.1029/2006GL025711

Arora, V.K., Boer, G.J. 2001. Effects of simulated climate change on the hydrology of major river basins, Journal of Geophysical Research. 106(D4): 3335-3348.

ASCE Task Committee on definition of Criteria for Evaluation of Watershed Models of the Watershed Management Committee, Irrigation and Drainage Division. Criteria for Evaluation of Watershed Models. J Irrig Drain Eng 1993;119(3):429–442.
ASCE Task Committee on definition of Criteria for Evaluation of Watershed Models of the Watershed Management Committee, Irrigation and Drainage Division (1993), Criteria for evaluation of watershed models, J. Irrig. Drain. Eng., 119(3), 429–-442.

Bardossy A, Plate EJ. 1991. Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. J Hydrol ;122:33–47.

Gunn SR, Brown M, Bossley KM. 1997. Network performance assessment for neuro fuzzy data modelling. In: Liu X, Cohen P, Berthold M, editors. Intelligent data analysis. Lecture notes in computer science, vol. 1208. Berlin Heidelberg: Springer-Verlag; p. 313–23.

Gutierrez, J. M., A. S. Cofino, R. Cano,, and M. A. Rodriguez. 2004 Clustering methods for statistical downscaling in short-range weather forecasts, Monthly Weather Review, 132, 2169--2183.

Haupt, R.L., Haupt, S.E., 2004. Practical Genetic Algorithm. Wiley, New Jersey, pp. 253.

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess. 2006 Downscaling heavy precipitation over the UK: a comparison of dynamical and statistical methods and their future scenarios, International Journal of Climatology, 26, 1397--1415.

Hewitson BC, Crane RG. 1996. Climate downscaling: techniques and application. Climate Res ;7:85–95.

Hewitson BC, Crane RG. 1992. Large-scale atmospheric controls on local precipitation in tropical Mexico. Geophys Res Lett;19(18):18351838.

Hingane LS, Rupakumar K, Ramanamurthy BV. 1985. Long-term trends of surface air temperature in india. J Climatol;5:521–8.

Hughes JP, Guttorp P, Charles SP. 1999. A non-homogeneous hidden Markov model for precipitation occurrence. Appl Statist;48(1):15–30.

Hughes JP, Lettenmaier DP, Guttorp P. 1993. A stochastic approach for assessing the effect of changes in synoptic circulation patterns on Gauge precipitation. Water Resour Res;29(10):3303–15.

Hughes, J. P., and P. Guttorp. 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, Water Resources Research, 30(5), 1535--1546.

Hulme, M. and T. C. Carter. 1999. Representing uncertainty in climate change scenarios and impact studies - ECLAT-2 red workshop report, eds. T. Carter, M. Hulme, and D. Viner, Climatic Research Unit, Norwich.

Hush, D.R., Horne, B.G., 1993. Progress in supervised neural networks: What's new since Lippmann? IEEE Signal Processing Magazine 10, 8–39.

Huth, R. 2004. Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors, Journal of Climate, 17, 640--651.

IPCC . 2000. Emission Scenarios, A special report of IPCC Working Group III, Cambridge University Press, Cambridge, UK.

IPCC-TGCIA .1999. Guidelines on the use of scenario data for climate impactand adaption assesment, Version 1, 69.

Johnson, M.S., Coon, W.F., Mehta, V.K., Steenhuis, T.S., Brooks, E.S., 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. Journal of Hydrology. 284: 57–76.

Jones, P. D., J. M. Murphy, and M. Noguer. 1995. Simulation of climate change over europe using a nested regional-climate model, i: assessment of control climate, including sensitivity to location of lateral boundaries, Q. J. R. Meteorol. Soc., 121, 1413--1449.

Jones, R. N. 2000. Analysing the risk of climate change using an irrigation demand model, Clim. Res., 14, 89--100.

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., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77 (3), 437–471.

Katz RW, Parlange MB. 1996. Mixtures of stochastic processes: applications to statistical downscaling. Clim Res;7:185–93.

Keerthi, S.S., Lin, C.-J., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Computation 15 (7), 1667–1689.

Khalil A, McKee M, Kemblowski M, Asefa T. 2005. Sparse Bayesian learning machine for real-time management of reservoir releases. Water Resour Res;41:W11401. doi:10.1029/2004WR003891.

Khalil AF, McKee M, Kemblowski M, Asefa T, Bastidas L. 2006. Multiobjective analysis of chaotic dynamic systems with sparse learning machines. Adv Water Resour;29(1):72–88.

Kim, M.K., Kang, I.S., Park, C.K., Kim, K.M., 2004. Superensemble prediction of regional precipitation over Korea. International Journal of Climatology 24 (6), 777–790.

Kottegoda, N.T., Rosso, R., 1998. Statistics, Probability, and Reliability for Civil and Environmental Engineers. McGraw-Hill, Singapore.

Lall, U. 1995. Recent advances in nonparametric function estimation, Supplement, U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys., 33, 1093--1102

Lall, U., B. Rajagopalan, D. G. Tarboton. 1996. A nonparametric wet/dry spell model for resampling daily precipitation, Water Resour. Res., 32(9), 2803--2823

Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for time series resampling. Water Resourses Research 32(3), 679–693.

Lall, U., Y.-I. Moon, and K. Bosworth. 1993. Kernel flood frequency estimators: Bandwidth selection and kernel choice, Water Resour. Res., 29(4), 1003--1016

Leavesley, G. H. 1994. Modeling the effects of climate change on water resources - a review, Climatic Change, 28, 159--177.

Li, L., Zhen-Chun Hao, Z.C., Wang, J.H, Zhen-Hua Wang, Z.H., Zhong-Bo Yu, Z.B. 2008. Impact of Future Climate Change on Runoff in the Head Region of the Yellow River. Journal of Hydrologic Engineering. 13(5), 347-354

Lin, H.-T., Lin, C.-J. 2003. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report, Department of Computer Science and Information Engineering, National Taiwan University.

MacKay D. 2003. Information theory, inference, and learning algorithms. Cambridge University Press

Maity, R., and D. Nagesh Kumar. 2006. Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño–Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO), J. Geophys. Res., 111, D07104, doi:10.1029/2005JD006539.

Maity,R. and D. Nagesh Kumar. 2008. Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty, , Journal of Geophysical Research - Atmospheres, American Geophysical Union, Vol. 113, D14105

Maurer, E.P. 2007. Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios Climatic Change 82, 309–325.

McKee, T. B., N. J. Doesken, and J. Kleist, (1993), The relationship of drought frequency and duration to time scale, Eighth Conference on Applied Climatology, American Meteorological Society, 179--184.

Mehrotra R, Sharma A. A nonparametric nonhomogeneous hidden Markov model for downscaling of multi-site daily rainfall occurrences. J Geophys Res – Atmos 2005;110(D16):16108.

Mercer, J., 1909. Functions of positive and negative type and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society, London, A 209, 415–446.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Murphy J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins and D. A. Stainforth (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768–772.

Murphy JM. An evaluation of statistical and dynamical techniques for downscaling local climate. J Clim 1999;12:2256–84.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T.Y., Kram, T., La Rovere, E.L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Riahi, K., Roehrl, A., Rogner, H.H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., Dadi, Z., 2000. IPCC Special Report on Emissions Scenarios (SRES), Working Group III, Intergovernmental Panel on Climate Change (IPCC), 595 pp, http://www.grida.no/climate/ipcc/emission/index.htm Cambridge University Press, Cambridge

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models, Part-1: A discussion of principles, J. Hydrol., 10,282–290

Neitsch, S.L., Arnold, J., Kiniry, J.R., Williams, J.R. 2000. Soil and Water Assessment Tool User's Manual Version 2000, Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

Neitsch, S.L., Arnold, J., Kiniry, J.R., Williams, J.R., King, K.W., 2002. Soil and Water Assessment Tool – Theoretical Documentation, Texas Water Resources Institute (TWRI) Report TR-191, College Station, Texas.

New, M. and M. Hulme (2000), Representing uncertainty in climate change scenarios: a Monte Carlo approach, Integrated Assessment, 1, 203--213

Pearson K. 1896. Mathematical contributions to the theory of evolution III regression heredity and panmixia. Philosophical Transactions of the Royal Society of London Series 187: 253–318.

Polansky, A. M. and E. R. Baker (2000), Multistage plug-in bandwidth selection for kernel distribution function estimates. Journal of Statistical Computation and Simulation, 65, 63--80

Prodanovic, P., J. Cunderlik, and S.P. Simonovic (2005), "Synthetic Storm Model for Climate Change Impact Modelling", Proceedings, 17th Canadian Hydrotechnical Conference, CSCE, Edmonton, 8

Prudhomme, C, Jakoba, D., Svenssona, C.2003. Uncertainty and climate change impact on the flood regime of small UK catchments Journal of Hydrology 277 1–23.

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

Raisanen, J., and T. N. Palmer (2001), A probability and decision-model analysis of a multimodel ensemble of climate change simulations, J. Clim., 14, 3212--3226.

Rao, P. G., (1995) Effect of cliamte change on streamflows in the Mahanadi river basin, India, Water International, 20, 205-212.

Rao, P. G. and K. K. Kumar (1992) Climatic shifts over Mahanadi river basin, Current Science, 63, 192-196.

Ross, T. J. (1997), Fuzzy Logic With Engineering Applications, McGraw-Hill International Edition, 379--396.

Roubens, M. 1982. Fuzzy clustering algorithms and their cluster validity, Eur. J. Oper. Res., 10, 294– 301.

Salathe, E.P. 2005. Downscaling simulations of future global climate with application to hydrologic modeling. International Journal of Climatology 25: 419-436.
San Juan River Basin, Colorado. J Hydrol 1999;225:67–91.

Scott, D. W. 1992. Multivariate Density Estimation, Theory, Practice, and Visualization, John Wiley, New York.

Sharma, A., D. G. Tarboton, and U. Lall. 1997. Streamflow simulation: A nonparametric approach, Water Resour. Res., 33(2), 291—308

Shivam Tripathi, 2004, Downscaling of general circulation models to assess the impact of climate change on rainfall of India, ME dissertation, IISc, Bangalore.

Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis, 1st edition, Chapman and Hall, London.

Simonovic, S. P. and E. G. R. Davies. 2006. Are we modelling impacts of climatic change properly?, Hydrol. Process., 20, 431—433.

Simonovic, S. P. and L. Li . 2003. Methodology for Assessment of Climate Change Impacts on Large-Scale Flood Protection System, J. Water Resour. Plann. Manage., 129(5), 361--371.

Simonovic, S. P. and L. Li. 2004. Sensitivity of the Red River Basin Flood Protection System to Climate Variability and Change, Water Resources Management, 18, 89--110.

Smola AJ, Schoelkopf B. 1998. A tutorial on support vector regression, NeuroCOLT2. Technical Report NC2-TR-1998-030, Royal Holloway College, University of London, UK

Smola AJ. 1996. Regression estimation with support vector learning machines. Munchen, Germany: Technische Universitat Munchen
Spott, M. 1999. A theory of possibility distributions, Fuzzy Sets and Systems, 102(2), 135--155.

Srinivas V.V., Tripathi, S., 2008. Statistical Downscaling of Regional Precipitation Under Climate Change Scenarios Using Support Vector Machines. In: Singh, V. P. (Ed.), Hydrology and Hydraulics, Water Resources Publications, Highlands Ranch, Colorado, USA, Chapter 15, pp 533-586.

Srivastava HN, Devan BN, Dixit SK, Prakasarao GS, Singh SS, Rao RK. 1992. Decadal trends in climate over India. Mausam; 43:7–20.

Stehlik, J. and A. Bardossy. 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, Journal of Hydrology, 28(5), 1247--1259

Steinemann, A. 2003 Drought indicators and triggers: a stochastic approach to evaluation, Journal of the American Water Resources Association, 39(5), 1217--1233.

Suykens, J.A.K., 2001. Nonlinear modelling and support vector machines. In: Proceedings of IEEE Instrumentation and measurement technology conference, Budapest, Hungary, pp. 287–294.

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

Tebaldi, C., L. O. Mearns, D. Nychka, and R. L. Smith. 2004. Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations, Geophys. Res. Lett., 31, L24213, doi:10.1029/ 2004GL021276.

Tebaldi, C., R. Smith, D. Nychka, and L. O. Mearns. 2005. Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles, J. Clim., 18, 1524– 1540.

Thapliyal V, Kulshrestha UC. 1991. Climate changes and trends over India. Mausam;42:333–8.

Tipping, M. E. 2001. Sparse Bayesian learning and the relevance vector machine, Journal of Machine Learning Reserach, 1, 211-244.

Tonn, B. 2005. Imprecise probabilities and scenarios, Futures, 37, 767-775.

Toth, B., Pietroniro, A., Conly, F.M., Kouwen, N. 2006. Modelling climate change impacts in the Peace and Athabasca catchment and delta: I—hydrological model Application. Hydrological Processes 20, 4197–4214

Trigo RM, Palutikof JP. 1999. Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim Res;13:45–59.

Tripathi, S., 2004. Downscaling of general circulation models to assess the impact of climate change on rainfall of India. ME thesis, Indian Institute of Science, Bangalore, India.

Tripathi, S. and V. V. Srinivas. 2005. Downscaling of General Circulation Models to assess the impact of climate change on rainfall of India, Proceedings of International Conference on Hydrological Perspectives for Sustainable Development (HYPESD - 2005), 23- 25 February, IIT Roorkee, India, 509--517.

Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach. Journal of Hydrology 330, 621-640.

USDA-NRCS, 2004. Part 630: Hydrology. Chapter 10: Estimation of direct runoff from storm rainfall: Hydraulics and hydrology: Technical references. In NRCS National Engineering Handbook.Washington, D.C.: USDA National Resources Conservation Service. Available at: www.wcc.nrcs.usda.gov/hydro/

Vapnik, V.N., 1995. The Nature of statistical learning Theory, New York: Springer.

Vapnik, V.N., 1998. Statistical Learning Theory, New York, Wiley.

Wahba G. 1985. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline-smoothing problem. Ann Stat;4:1378402.

Wetterhall, F., S. Halldin, and C. Xu. 2005. Statistical precipitation downscaling in central Sweden with the analogue method, Journal of Hydrology, 306, 174--190.

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting

material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA). <http://ipccddc.cru.uea.ac.uk/guidelines/StatDown_Guide.pdf>

Wilby RL, Dawson CW, Barrow EM. 2002. SDSM – a decision support tool for the assessment of regional climate change impacts. Environ Modell Softw;17(2):147–59.

Wilby RL, Dawson CW. 2004. Using SDSM version 3.1 A decision support tool for the assessment of regional climate change impacts, User Manual.

Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, et al. 1998. Statistical downscaling of general circulation model output: a comparison of methods. Water Resour Res;34: 2995–3008.

Wilby, R. L., and I. Harris. 2006. A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42, W02419, doi:10.1029/2005WR004065.

Wilby, R. L., L. E. Hay, and G. H. Leavesly. 1999. A comparison of downscaled and raw gcm output: implications for climate change scenarios in the San Juan river basin, Colorado, Journal of Hydrology, 225, 67--91.

Wilby, R.L., H. Hassan, K. Hanaki. 1998. Statistical downscaling of hydrometeorological variables using general circulation model output, Journal of Hydrology, 205, 1--19.

Wilby, R.L., L.E. Hay, W.J. Gutowski, R.W. Arritt, E.S. Takle, Z.T. Pan, G.H. Leavesley, and M.P. Clark. 2000. Hydrological responses to dynamically and statistically downscaled climate model output, Geophysical Research Letters, 27(8), 1199--1202.

Wilks, D. S., 1999. Multisite downscaling of daily precipitation with a stochastic weather generator, Climate Reserach, 11, 125--136.

Willmott, C. J., C. M. Rowe, and W. D. Philpot. 1985. Small-scale climate map: a sensitivity analysis of some common assumptions associated with the grid-point interpolation and contouring, American Cartographer, 12, 5-16.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier. 2002. Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys. Res., 107(D20), 4429, doi:10.1029/2001JD000659

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

Zhang, B., Govidaraju, R.S., 2000. Prediction of watershed runoff using bayesian concepts and modular neural network. Water Resources Research 36 (3), 753–762.