

# Assessing impact of climate change on season length in Karnataka for IPCC SRES scenarios

AAVUDAI ANANDHI

*CUNY Institute for Sustainable Cities/Hunter College, City University of New York,  
New York, United States of America.*

Changes in seasons and season length are an indicator, as well as an effect, of climate change. Seasonal change profoundly affects the balance of life in ecosystems and impacts essential human activities such as agriculture and irrigation. This study investigates the uncertainty of season length in Karnataka state, India, due to the choice of scenarios, season type and number of seasons. Based on the type of season, the monthly sequences of variables (predictors) were selected from datasets of NCEP and Canadian General Circulation Model (CGCM3). Seasonal stratifications were carried out on the selected predictors using K-means clustering technique. The results of cluster analysis revealed increase in average, wet season length in A2, A1B and B1 scenarios towards the end of 21st century. The increase in season length was higher for A2 scenario whereas it was the least for B1 scenario. COMMIT scenario did not show any change in season length. However, no change in average warm and cold season length was observed across the four scenarios considered. The number of seasons was increased from 2 to 5. The results of the analysis revealed that no distinct cluster could be obtained when the number of seasons was increased beyond three.

## 1. Introduction

Changes in seasons and its length are an indicator, as well as an effect, of climate change. They reflect the variations that are occurring in the cycling of energy in the global environment and profoundly affect the balance of life in ecosystems, impact agriculture and water balance of a region. Hence, knowing the future changes in season length in developing countries like India helps policy makers and general public for realistic assessment, management and mitigation of natural disasters, and for sustainable development. However, there is dearth of attempts in published literature that studies the variability in the future season length in this part of the world.

Global climate models (GCMs) are widely used to simulate climate conditions on earth, several decades into the future, for each of the constructed climate change scenarios. As GCMs are

generally run at coarser scale to cover the whole globe, they are inefficient in simulating regional climatic conditions on earth. For this purpose, weather classification methods have been used in the past to transfer the information on large-scale patterns provided by GCMs (predictors) to seasons (predictand) at regional scales.

Owing to the availability of a number of GCMs, climate change scenarios, classification methods, predictors, definitions of seasons, number of seasons, etc., prediction of season length is subject to a number of uncertainties. Hence, there is a need to study the changes in season length of a region by the various alternatives available.

The objective of this study is to investigate the variability of future season length to some of the uncertainties such as the choice of climate scenario, predictors, definition of seasons. The chosen methodology is tested for the Karnataka state in India.

**Keywords.** General circulation model (GCM); third generation Canadian coupled global climate model (CGCM3); SRES A1B, A2, B1 and COMMIT scenarios; K-means clustering; uncertainty.

## 2. Background

In order to obtain season length, the seasons are to be defined. In general, depending on the application, seasons are defined in a number of ways, such as astronomical seasons, meteorological seasons (e.g., Argiriou *et al* 2004) and standard seasons (Tuller 1990). In this study seasons are divided based on meteorological variables (namely, rainfall, temperature and wind), as they affect agriculture and irrigation of the region and also they are commonly used in climate change impact studies.

A number of studies have analyzed the rainfall and temperatures simulated by the IPCC AR4 GCMs and regional climate model (RCM) for different regions of India (Rupa Kumar *et al* 2006; Kripalani *et al* 2007; Yadav *et al* 2010a). These studies show an increase in mean monsoon rainfall and temperature towards the end of 21st century under scenarios of increasing greenhouse gas concentrations and sulphate aerosols. Though a number of studies have analyzed the changes in rainfall and temperatures for the different seasons in India, there is a lacuna in research to study the changes in season length for the future scenarios.

The season length may be defined as conventional (fixed) length or ‘floating’ length. In a fixed season length, the starting dates and length of seasons remain the same for every year. In contrast, in a ‘floating’ season length, the date of onset and duration of each season is allowed to change from year to year. Studies have shown that floating seasons reflect ‘natural’ seasons contained in the climate data better than fixed seasons, especially under changing climate conditions (Winkler *et al* 1997; Anandhi *et al* 2008). Therefore, in this study, floating season length is used to effectively capture the changes in the future climate conditions.

The weather classification methods are used in this study to define floating meteorological season length, link large-scale circulation pattern and surface weather by grouping the daily/monthly datasets of large-scale circulation or surface climate variables into a finite number of discrete weather types or ‘seasons’. The classification may be subjective (e.g., Lamb 1972), objective (e.g., Tripathi *et al* 2006), or hybrid (e.g., Sailor *et al* 2008). In subjective classification, the stratifications were carried out manually using empirical rules where the meteorological scientist’s experience is applied and reflect the knowledge of meteorologists. These stratifications cannot be easily replicated and are labour-intensive. In objective classification methods, a variety of automated techniques were developed using computers to group the seasons, and there is subjectivity in the choice of the algorithm.

Hybrid techniques combine elements of empirical and automated procedures for grouping seasons, thereby avoiding time delay and enabling the production of easily reproducible and interpretable results. In this study the hybrid approach using K-means clustering (MacQueen 1967) was used since it had the advantages of both subjective and objective approaches.

In this study, season length is proposed to be investigated for:

- each of the four scenarios specified in Special Report on Emission Scenarios (SRES; Nakicenovic *et al* 2000) that are relevant to IPCC’s fourth assessment report (AR4) (Alley *et al* 2007) namely A1B, A2, B1 and COMMIT,
- the seasons defined based on the three meteorological variables,
- the predictors used in the classification of seasons, and
- varying number of seasons.

## 3. Study region and data used

The chosen methodology is tested for the Karnataka state of India (figure 1). It has an area of 191791 km<sup>2</sup> situated between 11°N to 18°N latitudes and 74°E and 78°E longitudes. The state has the second largest arid and semi-arid regions in the country after Rajasthan.

The predictors used for this study comprise of the reanalysis data extracted from database prepared by National Centers for Environmental Prediction (NCEP) (Kalnay *et al* 1996), and the data extracted from simulations by Canadian Center for Climate Modeling and Analysis’s (CCCma) third generation Coupled Global Climate Model (CGCM3). The details of the data used are furnished in table 1. A brief description of the four future scenarios is provided in table 2. The predictors at each of the nine NCEP grid points and twelve CGCM3 grid points which are within and surrounding the study region are used in the study. The list of predictors and their description used are provided in table 3. In addition to the predictors, the surface average temperature and rainfall were also extracted from NCEP and CGCM3 database. 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).

## 4. Methodology

The methodology adopted for stratification into seasons is presented in the form of a flowchart

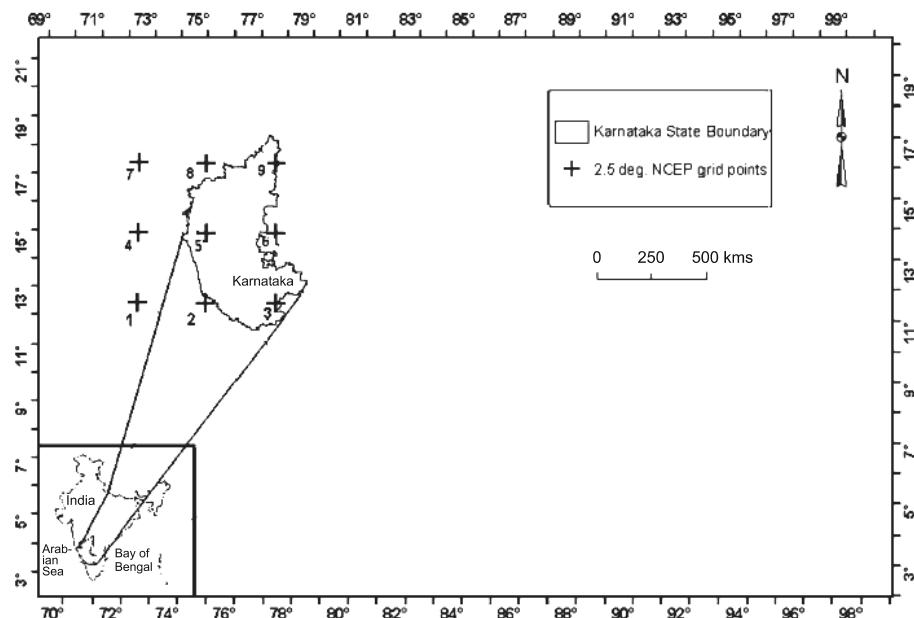


Figure 1. Location of Karnataka state of India. The latitude, longitude and scale of the map refer to the Karnataka state. The data extracted from GCMs are re-gridded to the nine  $2.5^{\circ}$  NCEP grid points. Among the nine grid points 1, 4 and 7 are over Arabian Sea, and the remaining points are on land.

Table 1. Details of meteorological data used in the study.

| Data type                                      | Source of data  | Period   | Details  | Time scale |
|--|---|--|--|------------|
| CGCM3 T/47 data on atmospheric variables       | <a href="http://www.cccma.bc.ec.gc.ca/cgi-bin/data/cgcm3">http://www.cccma.bc.ec.gc.ca/cgi-bin/data/cgcm3</a> | 1948–2100<br>baseline: 20C3M – 1948 to 2000.<br>future: SRES A1B, A2, B1 and COMMIT – 2001 to 2100 | 12 grid points for atmospheric variables, with grid box $\approx 3.75^{\circ}$ .<br>Latitudes range: $9.28^{\circ}\text{N}$ – $20.41^{\circ}\text{N}$ .<br>Longitudes range: $71.25^{\circ}\text{E}$ – $78.75^{\circ}\text{E}$ . | Monthly    |
| NCEP re-analysis data of atmospheric variables | Kalnay <i>et al</i> (1996)  | 1948–2000  | 9 grid points for atmospheric variables with grid box $2.5^{\circ}$ .<br>Latitudes range: $12.5^{\circ}\text{N}$ – $17.5^{\circ}\text{N}$ .<br>Longitudes range: $72.5^{\circ}\text{E}$ – $77.5^{\circ}\text{E}$ .               | Monthly    |
| Observed historical rainfall for Karnataka     | High-resolution gridded daily rainfall data (Rajeevan <i>et al</i> 2005, 2006)                                | 1951–2001  | 20 grid points with grid box $1^{\circ}$ in Karnataka are considered.  | Monthly    |

(figure 2). The steps involved in the stratification into seasons are as follows:

*Step 1.* Selection of appropriate predictors from regridded NCEP and GCM monthly datasets as probable predictors at each of the NCEP grid points within and surrounding the study region (figure 1).

*Step 2.* Preparation of scatter plots and correlation plots using three measures of dependence

(Pearson's product moment correlation and non-parametric rank correlations namely Spearman's rho and Kendall's tau explained in Annexure) between predictor variables in NCEP and GCM datasets after removing significant autocorrelations and trends. Then, examine the correlation between predictors in NCEP and GCM datasets.

*Step 3.* Selection of highly correlated predictors (potential predictors) with reference to a specified threshold  $T_{ng}$  for correlation between predictor

Table 2. *Explanation of the scenarios considered in the study.*

| Dataset  | Description   | Duration  |
|--|---|-----------|
| Climate of the 20th Century (20C3M)            | Atmospheric CO <sub>2</sub> concentrations and other input data are based on historical records or estimates beginning around the time of the Industrial Revolution.  | 1870–2000 |
| Year 2000 equivalent CO <sub>2</sub> (COMMIT)  | Atmospheric CO <sub>2</sub> concentrations are held at year 2000 levels. This experiment is based on conditions that already exist (e.g., ‘committed’ climate change).  | 2001–2100 |
| SRES B1 (550 ppm equivalent CO <sub>2</sub> )  | Atmospheric CO <sub>2</sub> concentrations reach 550 ppm in the year 2100 in a world characterized by low population growth, high GDP growth, low energy use, high land-use changes, low resource availability and medium introduction of new and efficient technologies.             | 2001–2100 |
| SRES A1B (720 ppm equivalent CO <sub>2</sub> ) | Atmospheric CO <sub>2</sub> concentrations reach 720 ppm in the year 2100 in a world characterized by low population growth, very high GDP growth, very high energy use, low land-use changes, medium resource availability and rapid introduction of new and efficient technologies. | 2001–2100 |
| SRES A2 (850 ppm equivalent CO <sub>2</sub> )  | Atmospheric CO <sub>2</sub> concentrations reach 850 ppm in the year 2100 in a world characterized by high population growth, medium GDP growth, high energy use, medium/high land-use changes, low resource availability and slow introduction of new and efficient technologies.    | 2001–2100 |

Table 3(a). *Probable predictors selected.*

| Sl. no. | Predictand  | Probable predictor(s) selected from NCEP and CGCM3 monthly datasets downscaling predictand                        |
|---------|-------------|---|
| 1       | Rainfall    | 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       | Temperature | Ta 925, Ua 925, Va 925  |
| 3       | Wind speed  | Ua 925, Va 925  |

Table 3(b). *Abbreviations used to define predictors.*

| Sl. no. | Abbreviations | Long form of abbreviations     |
|---------|---------------|--------------------------------|
| 1       | Hus 850       | Specific humidity at 850 hPa   |
| 2       | Hus 925       | Specific humidity at 925 hPa   |
| 3       | Prw           | Precipitable water content     |
| 4       | Ps            | Surface pressure               |
| 5       | Ta 200        | Air temperature at 200 hPa     |
| 6       | Ta 500        | Air temperature at 500 hPa     |
| 7       | Ta 700        | Air temperature at 700 hPa     |
| 8       | Ta 925        | Air temperature at 925 hPa     |
| 9       | Ua 200        | Zonal wind at 200 hPa          |
| 10      | Ua 925        | Zonal wind at 925 hPa          |
| 11      | Va 200        | Meridional wind at 200 hPa     |
| 12      | Va 925        | Meridional wind at 925 hPa     |
| 13      | Zg 200        | Geopotential height at 200 hPa |
| 14      | Zg 500        | Geopotential height at 500 hPa |
| 15      | Zg 925        | Geopotential height at 925 hPa |

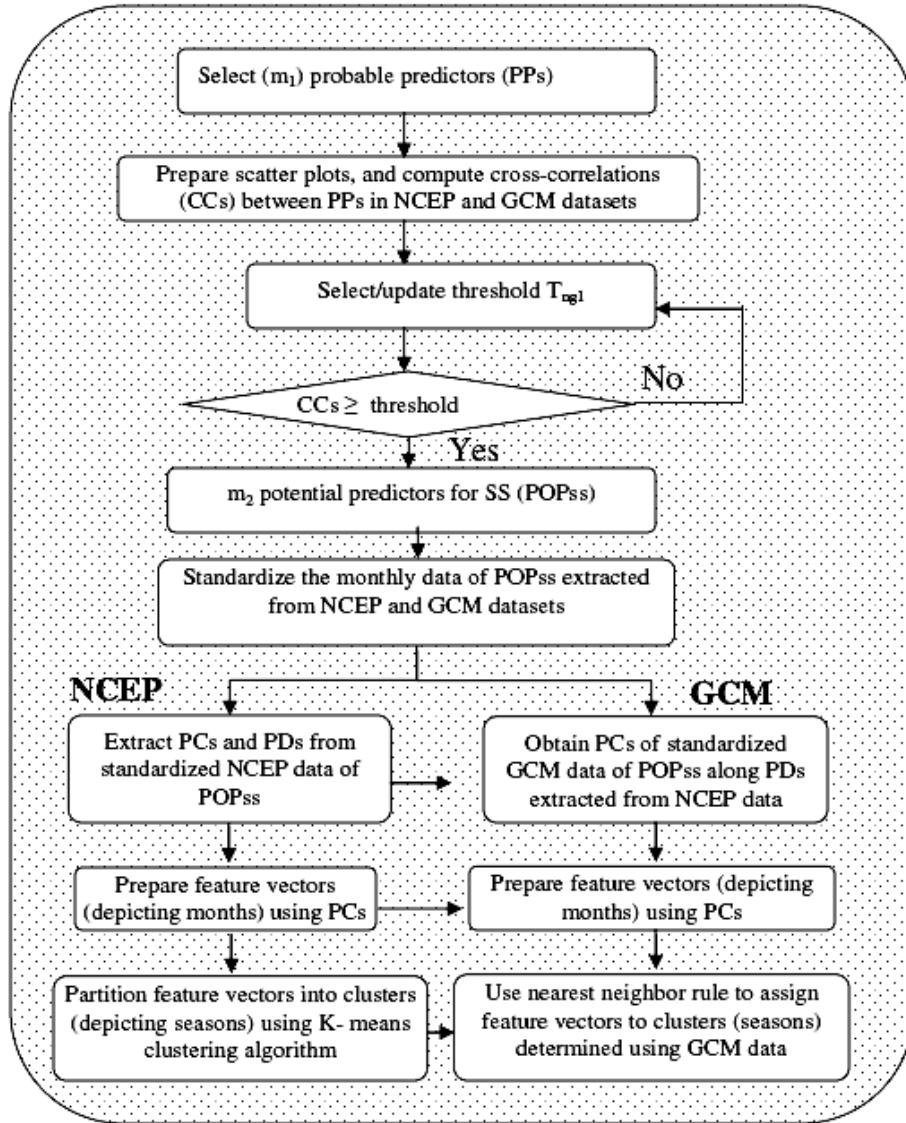


Figure 2. Flow chart of methodology adopted for stratification into seasons. In the figure,  $T_{ng}$  represents threshold for cross-correlation between predictor variables in NCEP and GCM datasets. PCs represents principal components, PDs denotes principal directions and SS denotes seasonal stratification.

variables in NCEP and GCM datasets. The selection of threshold is subjective. Using correlation as a metric to select predictors captures the change in direction of predictors. Other metrics may be used in the place of correlation to capture the changes in pattern as well as magnitude of predictor variables.

*Step 4.* Standardization of potential predictors for the baseline period (1948–2000) to reduce systemic bias (if any) in the mean and variance of the predictors in the GCM data, relative to those of the predictors in the NCEP reanalysis data. This step typically involves subtraction of mean and division by the standard deviation of the predictor for the period.

*Step 5.* Principal component (PC) analysis of standardized NCEP predictor data is carried out to extract PCs. PCs which preserve more than 98% of the variance in the data are selected. The PCs of GCM predictor data are extracted along principal directions of NCEP predictor data. The PCs account for most of the variance in the input and also remove the correlations, if any, among the input data. PCA was carried out in this study as they make the model more stable and reduce the computational load.

*Step 6.* Formation of feature vectors using PCs. The PCs for each month form a feature vector. The feature vectors form the input to the K-means clustering algorithm, and the seasons is its output.

*Step 7.* Partitioning of NCEP feature vectors into clusters, depicting seasons, by using K-means clustering (MacQueen 1967), while the GCM dataset partitioned into two clusters by using nearest neighbour rule (Fix and Hodges 1951).

*Step 8.* In this analysis, each feature vector (representing a month) of the NCEP data is treated as an object having a location in space. The feature vectors are partitioned into clusters such that the feature vectors within each cluster are as close to each other as possible in space, and are as far as possible in space from the feature vectors in the other clusters.

*Step 9.* The distance between feature vectors in space is estimated using Euclidian measure. Subsequently, each feature vector (representing a month) of the NCEP data is assigned a label that denotes the cluster (season) to which it belongs.

*Step 10.* The feature vectors prepared from GCM simulations (past and future) are labeled using the nearest neighbour rule. As per this rule, each feature vector formed using the GCM data is assigned the label of its nearest neighbour from among the feature vectors formed using the NCEP data. To determine the neighbors for this purpose, the distance between NCEP and GCM feature vectors is computed using Euclidean measure.

*Step 11.* From the seasons, depending on the months falling in each season, the season length is determined in terms of month.

*Step 12.* For a season, estimate the average season length and the uncertainty limits.

The procedure to estimate the average season length and their uncertainty limits is explained using an example, wet-dry season based on rainfall for a period say 1948–2000. For a particular set of predictors, selected based on threshold  $T_{ng}$ , the months in the period may be classified as wet or dry month. This classification of wet or dry month may be different for a different  $T_{ng}$ . In other words, the month May 1981 may be classified as a wet month for a set of predictors while the same month (May 1981) may be classified as a dry month for another set of predictors (due to a different  $T_{ng}$ ). In this case for 15 combinations of predictors we get 15 series of months labeled wet – dry. For each month in the period 1948–2000, from the different combinations of  $T_{ng}$  (15 combinations in this case), the number of times the month is classified as wet or dry is counted. If the count is more for wet than dry then the month is assigned as wet and *vice versa*. Thus a single series is obtained for the period 1948–2000 where each

month is assigned a wet or dry label and the uncertainty in the fifteen combinations are accounted by giving equal weightage. From this single series, the average season length and its uncertainty limits are estimated. Now for the period 1948–2000, all the January's (53 in number) are counted for wet or dry months. To estimate the average season length the frequency of the month as dry or wet is estimated. If the frequency of dry month is more than wet then January is a dry month. In the similar way, the rest of the 12 months in the period are assigned wet or dry month depending on the frequency. From the number of months in the wet season the length of average wet season is estimated and from the number of months in the dry season the length of average dry season is estimated. The upper and lower uncertainty limits in a season may be obtained from the frequency of wet and dry months from January to December. For wet season the lower uncertainty limit is the month in which the frequency of wet is 47 of the 53 (at least 90 percent) and frequency of dry is 6 or less (10 per cent). In other words, when at least 47 June months are wet, the month qualifies in the lower uncertainty limit. The upper uncertainty limit is obtained when at least 6 months from the 53 months is wet. In other words, when at least 6 months from 53 June months are wet, the month qualifies in the upper uncertainty limit. The lower uncertainty limit is the lower limit for season length and comprises of least number of months that was in a particular season. The upper uncertainty limit is the upper limit for season length and comprises of maximum number of months in a season. The upper and lower uncertainty limits encompass the range of changes in season length.

## 5. Results and discussion

The sources of uncertainty associated with assessment of future season length were examined using numerous simulations of the season length, for various choices of climate scenario, types of predictors and seasons, and number of seasons. The results of predictor selection and uncertainty analysis are presented and discussed in this section.

In this study, the seasons are classified based on meteorological variables as:

- wet and dry seasons based on rainfall;
- warm and cold seasons based on temperature;
- windy and non-windy seasons based on wind;
- and their combinations.

### 5.1 Predictor selection

Selection of appropriate predictor variables is one of the salient steps for translating large-scale

patterns to various seasons using classification methods. Some of the examples of predictors (variables and indices) chosen in previous studies in predicting Indian Monsoon rainfall are sea surface temperatures (SSTs), mean sea level pressure (MSLP), geopotential height and wind fields, Eurasian snow cover and surface temperatures, the Niño-3 SST anomaly, the Europe pressure gradient, the south Indian Ocean SST index, and north Atlantic SST (Yadav *et al* 2007, 2009a, 2009b, 2010b).

The predictors used in this study include some of the predictors used in previous studies and also based on the work by Anandhi (2007). The selected predictors have a physically meaningful relationship with rainfall, temperature and wind are selected for clustering into different meteorological seasons (table 3a). The physical basis for selecting the predictors in table 3 is explained below. As the rainfall in the region is dependent on dynamics through advection of water from the surrounding sea, and thermodynamics through effects of moisture and temperature, both of which can modify the local vertical static stability (Anandhi 2007). The physical basis for selection of these predictors is given. For example, winds during monsoon season advect moisture into the region, while temperature and humidity are associated 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 rainfall. Higher precipitable water in the atmosphere means more moisture, which in turn causes statically unstable atmosphere leading to more vigorous overturning, resulting in more rainfall. Lower pressure leads to more winds and so more rainfall. 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 rainfall. The 200 hPa pressure level captures the global scale effects. Temperature at 500 hPa represents the heating process of the atmosphere due to monsoonal rainfall 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 rainfall.

## 5.2 Uncertainties in wet and dry season length

Based on rainfall, the study region can be divided into two seasons namely wet and dry seasons. The sources of uncertainty in wet and dry season length were studied, for choice of predictors and climate scenario.

### 5.2.1 Uncertainties in wet and dry season length: Choice of predictors

For projecting future seasons it is necessary that GCM data is consistent with NCEP data. The cross-correlations were used to assess the relationship between predictors in NCEP and GCM datasets. The results reveal that

- Pearson's correlation coefficient and rank correlations (Spearman and Kendall's tau) lead to the same conclusions. All three measures of dependence showed near equal ranking of probable predictors. Therefore, in the following discussion only product moment correlation values were used, without loss of generality;
- the correlation between predictors in NCEP and GCM datasets is generally greater than 0.57 (except for Va 200), indicating that the predictor variables are realistically simulated by the GCM.

Depending on a threshold  $T_{ng}$  for the correlation coefficient, potential predictors are selected for classification into seasons and details are furnished in table 4. It can be observed from the table that the potential predictors selected were subjective to the threshold  $T_{ng}$  concerned. When the values of the predictors in NCEP and 20C3M GCM datasets have the same pattern of variation, then the values of threshold is equal to one ( $T_{ng} = 1$ ). Using such datasets will produce similar seasons during the time period 1948–2000. As the values of threshold decreases, the values of the predictors in NCEP and 20C3M GCM datasets are not of the same pattern of variation and the clusters formed using these two datasets could be different depending on the values of predictors. Higher  $T_{ng}$  represents selecting predictors which have GCM data highly consistent with NCEP data, while lower  $T_{ng}$  represents selecting predictors which have GCM data less consistent with NCEP data.

To estimate the uncertainty in season length for wet and dry seasons, the classification into seasons was performed using cluster analysis for each of the threshold values. From these classifications, the average season length (figure 3a) and uncertainty limits (figure 3b and c) are calculated using procedure explained in methodology section and results are shown in figure 3. It can be observed from the figure that the average wet season length for the various predictors selected is May to October during the period 1948–2000, while the average dry season length during the same period is November to May. The lower uncertainty limit in wet season length is June to September (figure 3b; NCEP, 20C3M), while the limit for dry season length is October to April (figure 3c; NCEP, 20C3M). The upper uncertainty limit in wet and dry season

Table 4. Potential predictors selected for classification into seasons based on rainfall for different values of  $T_{ng}$  for product moment correlation between probable predictors in NCEP and CGCM3 datasets.

| Sl. no. | $T_{ng}$  | R         | Potential predictors selected   |
|---------|-----------|-----------|---|
| 1       | 1.0–0.95  | 1.0–0.95  | —   |
| 2       | 0.93      | 0.93      | Ua 925  |
| 3       | 0.92      | 0.92      | Ua 925, Ps  |
| 4       | 0.91      | 0.91      | Ua 925, Ps, Ua 200  |
| 5       | 0.86–0.90 | 0.86–0.90 | Ua 925, Ps, Ua 200, Zg 925  |
| 6       | 0.84–0.85 | 0.84–0.85 | Ua 925, Ps, Ua 200, Zg 925, Prw   |
| 7       | 0.83      | 0.83      | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925  |
| 8       | 0.80–0.82 | 0.80–0.82 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850   |
| 9       | 0.77–0.79 | 0.77–0.79 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200   |
| 10      | 0.76      | 0.76      | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925   |
| 11      | 0.71–0.75 | 0.71–0.75 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925, Ta 500, Ta 200                                 |
| 12      | 0.70      | 0.70      | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925, Ta 500, Ta 200, Ta 700                         |
| 13      | 0.62–0.69 | 0.62–0.69 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925, Ta 500, Ta 200, Ta 700, Zg 500                 |
| 14      | 0.55–0.61 | 0.55–0.61 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925, Ta 500, Ta 200, Ta 700, Zg 500, Va 925         |
| 15      | 0.26–0.54 | 0.26–0.54 | Ua 925, Ps, Ua 200, Zg 925, Prw, Hus 925, Hus 850, Zg 200, Ta 925, Ta 500, Ta 200, Ta 700, Zg 500, Va 925, Va 200 |

length for the period 1948–2000 is April to October and September to May, respectively.

### 5.2.2 Uncertainties in wet and dry season length: Different future scenarios

The average season length for the future four scenarios (A2, A1B, B1 and COMMIT) are estimated for five time periods 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100 respectively and the results are provided in figure 3. From the figure it can be inferred that the average wet season length increase in future for A2, A1B and B1 scenarios for the time period 2080–2100. The increase was more for A2 scenario, where the wet season length increased from April–November during the period. This increase was followed by A1B and B1 scenarios, in which case the average wet season length increased from May–November during the period 2081–2100. No change in season length was observed in COMMIT scenario. With the increase in wet season length there is a decrease in dry season length during the period 2081–2100. The decrease is complimentary to wet season length. Similar pattern of increase, is observed in upper and lower uncertainty limits in wet season length. The increase in wet season is very high with the entire year classified as wet season during the period 2081–2100 for A2 scenario. These results

compliment the results from an earlier study by Rupa Kumar *et al* (2006). They showed an increase in rainfall towards the end of the 21st century in this region in SRES A2 and B1 scenarios. Hence, the increase in wet season length towards the end of the 21st century could be due to the increase in rainfall during the period.

The annual cycle of rainfall for observed historical, NCEP and GCM scenarios (20C3M, A1B, A2 and B1) are shown in the first column of figure 4 for the five time periods considered. This figure gives us general idea about the wet/dry seasons in the region.

The variation in future season length among scenarios is high for A2 scenario, whereas it is least for B1 scenario. This is because among the scenarios considered, the scenario A2 has the highest concentration of carbon dioxide ( $\text{CO}_2$ ) equal to 850 ppm, while the same for A1B, B1 and COMMIT scenarios are 720 ppm, 550 ppm and  $\approx 370$  ppm, respectively. Any rise in the concentration of  $\text{CO}_2$  in 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. In the COMMIT scenario, where the emissions are held the same as in the year 2000, no significant increase in the pattern of projected future season length could be discerned.

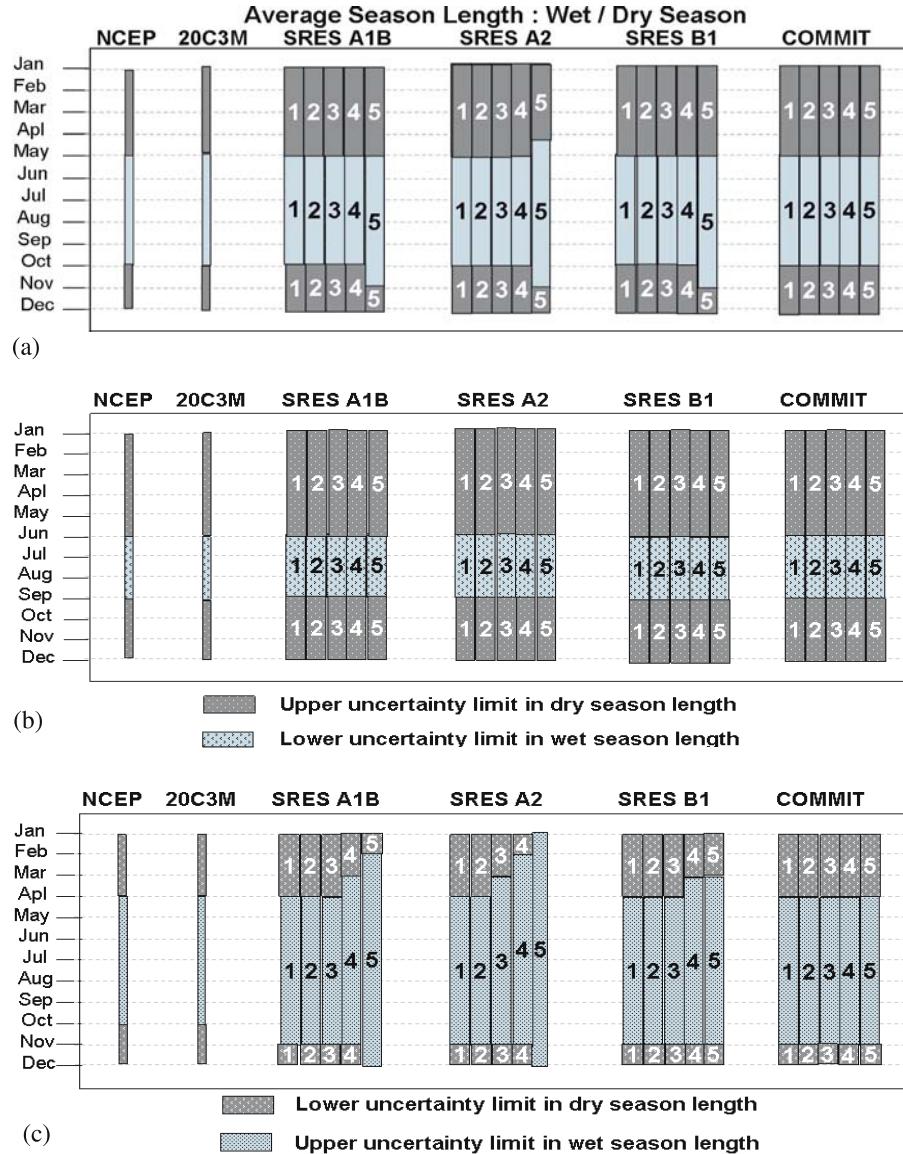


Figure 3. Typical results of classification into two seasons (wet and dry) performed using cluster analysis for different combinations of the fifteen predictors. (a) Shows the average wet and dry season length. (b and c) Show the upper and lower uncertainty limits in season length due to predictors selected. The period for NCEP and 20C3M is 1948–2000. The numbers 1 to 5 in the SRES scenarios represent the five time periods considered in the study namely 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100, respectively. The grey colour (darker shade) represents the dry season while the light blue colour (lighter shade) represents the wet season.

### 5.3 Uncertainties in warm and cool season length

Based on temperature the study region can be divided into two seasons namely warm and cold seasons. The sources of uncertainty in warm and cold season length were studied, for choice of predictors and climate scenario.

The potential predictors selected for classification into seasons for different values of  $T_{ng}$  for product moment correlation between probable predictors in NCEP and GCM datasets are shown in table 5. The classification into seasons was

performed for each of the threshold values shown in the table. The uncertainty in the months selected as a warm or cold month is subjective to predictor selected for clustering (figure 5).

The average season length and the uncertainty limits in the season length are estimated quantitatively using the various simulations of seasons obtained for the different combinations of predictor variables (figure 5). The typical results of uncertainty in the season length for the various combinations of predictors selected are shown in figure 5(b and c). It can be observed from the figure that on an average the warm and cold

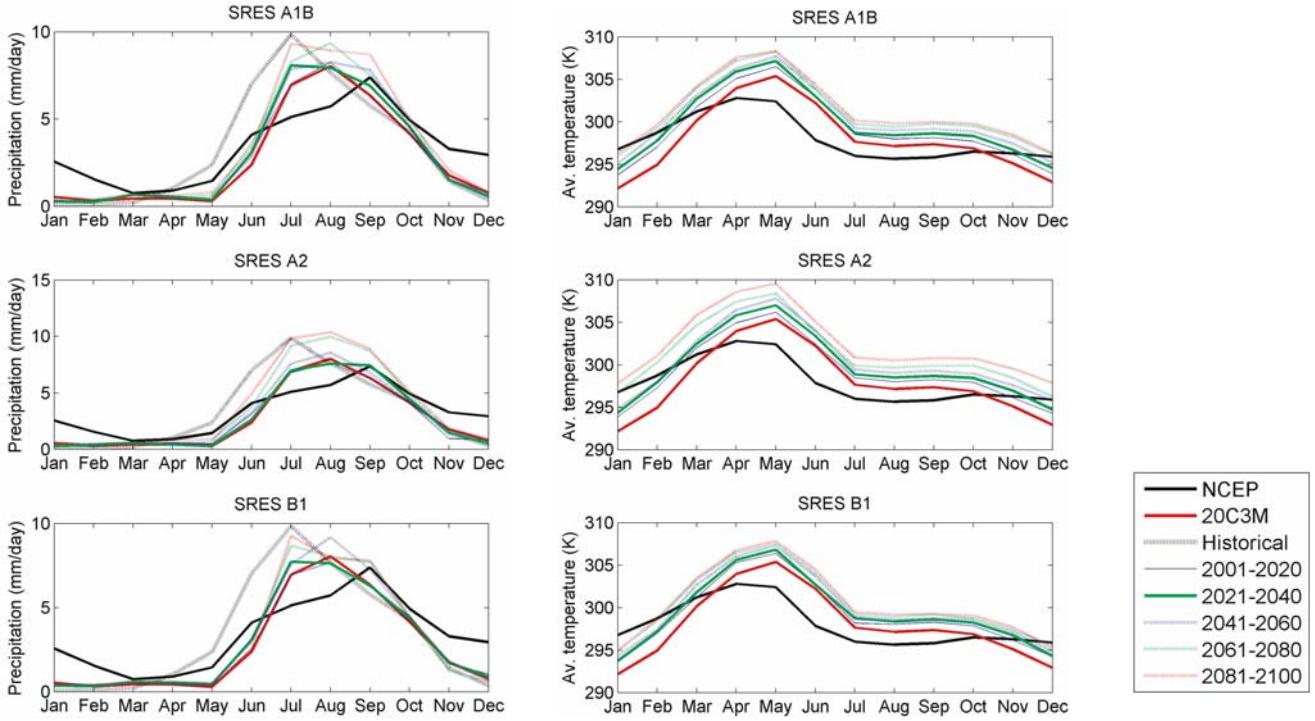


Figure 4. Monthly mean rainfall and average temperatures calculated from NCEP and GCM datasets. The 20C3M and NCEP data are averaged for the period 1948–2000. While for the future SRES scenarios, the five time periods are for the period 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100, respectively. Monthly mean observed rainfall is calculated for the period 1951–2000.

seasons coincided with wet (June–September) and dry months (rest of them) and there is no change in future warm and cold season length for the four scenarios (A1B, A2, B1 and COMMIT) considered. However, there is a decrease in warm season length in the lower uncertainty limit in all the four scenarios considered. The decrease in season length was higher for A2 scenarios when compared with the rest of the three scenarios. This decrease in warm season length may be related to increase in wet season length, during the same period as observed in the previous section. The annual cycle of average temperature for NCEP and GCM scenarios (20C3M, A1B, A2 and B1) are shown in the second column of figure 4 for the five time periods considered. This figure gives us general idea about the warm/cold seasons in the region.

#### 5.4 Uncertainties in wind and non-windy season length

Based on wind, the study region can be divided into two seasons namely windy and non-windy seasons. The classification into seasons was performed for the each of the threshold values to observe the variability in the length of windy and non-windy seasons. The results of the cluster analysis

Table 5. Potential predictors selected for classification into seasons based on temperature for different values of  $T_{ng}$  for product moment correlation between probable predictors in NCEP and CGCM3 datasets.

| Sl. no. | $T_{ng}$  | Potential predictors selected |
|---------|-----------|-------------------------------|
| 1       | 1.0–0.95  | —                             |
| 2       | 0.93      | Ua 925                        |
| 3       | 0.76–0.92 | Ua 925, Ta 925                |
| 4       | 0.56–0.75 | Ua 925, Ta 925, Va 925        |

(not presented for brevity) envisaged that windy and non-windy seasons coincided with wet (June–September) and dry months (rest of them).

Thus the seasons, classified on the basis rainfall, temperature and wind using the predictors from large-scale atmospheric variables, showed that the region had predominant wet and dry seasons. Hence for studying the variation in season length for varying number of seasons, seasons based on rainfall was used.

#### 5.5 Variation in season length with the number of seasons

The variation in season length with increase in the number of seasons was studied. For this purpose, the number of seasons was increased from two to five. It was observed that no distinct seasons

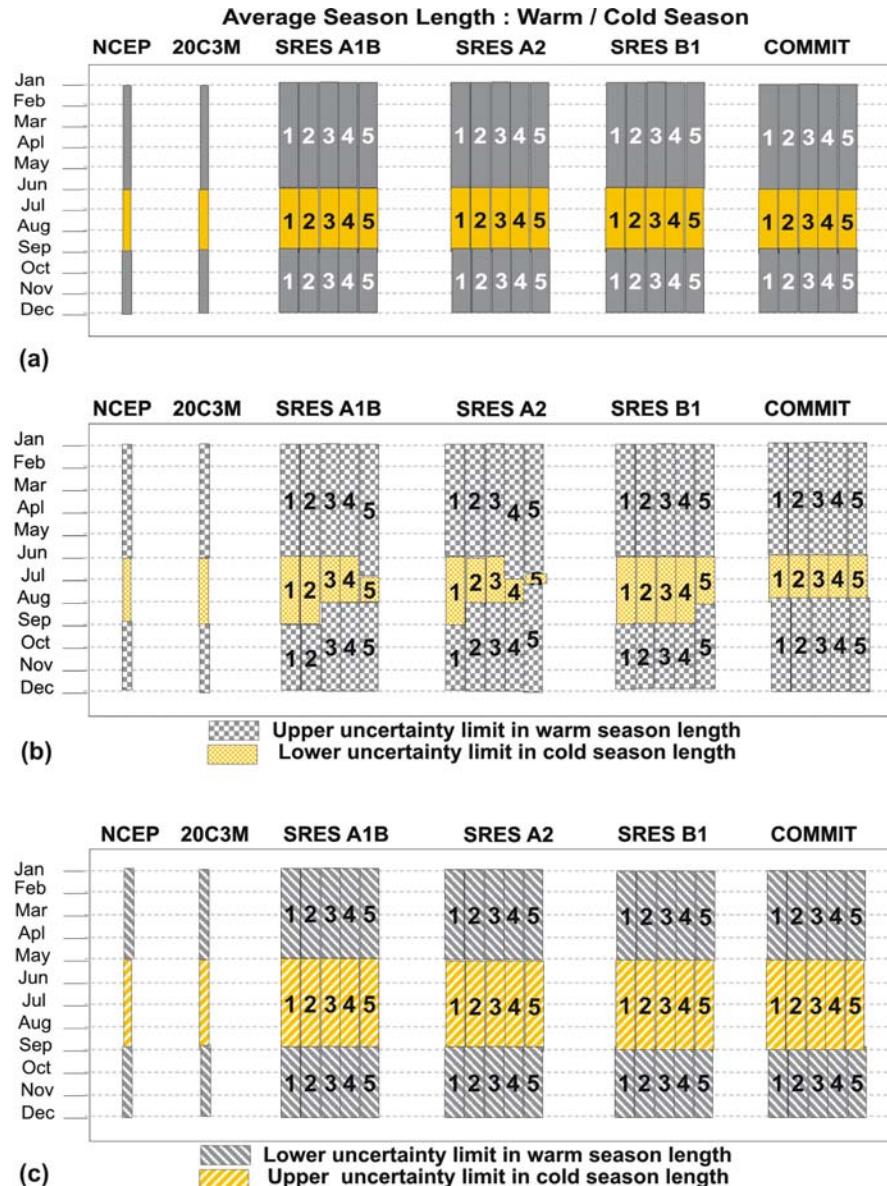


Figure 5. Typical results of classification into two seasons (warm and cold) performed using cluster analysis for different combinations of the predictors. (a) Shows the average warm and cold season length. (b) and (c) Show the upper and lower uncertainty limits in season length due to predictors selected. The period for NCEP and 20C3M is 1948–2000. The numbers 1 to 5 in the SRES scenarios represent the five time periods considered in the study namely 2001–2020, 2021–2040, 2041–2060, 2061–2080 and 2081–2100, respectively. The grey colour (darker shade) represents the cold season while the light orange colour (lighter shade) represents the warm season.

were obtained using cluster analysis when the number of seasons was increased more than three. When the region was classified into three clusters (clusters A, B and C), the cluster A was found to coincide with cold season and clusters B and C together coincide with wet season and cluster C coincide with warm season. This could be attributed to the fact that there are only two distinct seasons for the study region namely wet and dry seasons. It is observed that the wet season obtained was divided into two seasons (clusters B and C) from which one season coincided with warm season (June–September). Among the three

seasons, there was no distinct change in seasons defined by clusters A and C, whereas there is an increase in season length in season defined by cluster B.

#### 5.6 Implications of the varying seasonal length

The variation in season length affects agriculture which is related to changes in crop and forest phenology, increases in insect and pest diseases thereby affecting yield of crops (Patterson *et al* 1999; Peng *et al* 2004; Jones *et al* 2005). Longer growing season could increase plant uptake thereby

increasing evapotranspiration, reducing soil moisture, reducing the availability of water and air pollution. Longer growing seasons, and especially earlier springs, also lead to changes in the hydrological cycle with serious negative implications which could lead to water shortage in summer, when the demand is greater (Christidis *et al* 2007).

The consequences of change in warm and cold season length are of great importance to agriculture since this is closely related to the growth period of crops, crop water requirements, the life cycles of pests (insects and parasites) infecting the crops. The evaporation from water bodies are influenced by the change in warm seasons especially in arid and semi-arid regions. Temperature influences plant growth and increase evaporation/evapotranspiration of crops.

The most likely impact of climate change in plant diseases will be felt in three areas:

- (1) losses from plant diseases,
- (2) efficacy of disease management strategies, and
- (3) the geographical distribution of plant diseases.

Climate change could have positive, negative or no impact on individual plant diseases (Chakraborty *et al* 2000).

The ill effects of increase in windy season include increase in the rate of evaporation of water, increase in aridity, poor crop growth and reduction in yield due to mechanical damage to crops (e.g., stripping of leaves, abrasion in plant canopies through rubbing, Cleugh *et al* 1998).

## 6. Summary and conclusions

Changes in seasons and season length of a region are driven by shifts in the intensity of sunlight reaching earth's surface (insolation), latitude, continental and marine climate, wind direction, and local geographical features. Since change in season length is an indicator, as well as an effect, of climate change and is less studied, the objective of this study was set to investigate the variability of future season length to some of the uncertainties such as the choice of climate scenario, predictors, definition of seasons and number of seasons.

For this purpose four climate scenarios relevant to IPCC's AR4 report namely A1B, A2, B1 and COMMIT were used. The seasons were defined based on the three meteorological variables (namely, rainfall, temperature and wind) using large scale atmospheric variables (predictors) which have physical relationship to these variables. Wet and dry seasons were defined based on rainfall, warm and cold seasons were defined based on temperature, windy and non-windy

seasons based on wind. The predictors from NCEP reanalysis (period: 1948–2000) and Canadian GCM (CGCM3.1, period: 1948–2100) were used in this study. The classifications into seasons were carried out using K-means clustering technique and nearest neighbour rule. The chosen methodology was tested for the Karnataka state in India.

The results showed:

- Uncertainty in season length for the choice of predictors for the different seasonal definitions used in the study.
- The region has predominant wet and dry seasons.
- Average wet season increased in future towards the end of the 21st century for A2, A1B and B1 scenarios. The variation is high for A2 scenario, whereas it is least for B1 scenario. No change in COMMIT scenario. Similar pattern of increase is observed in upper and lower uncertainty limits in wet and dry season length.
- No change in average and upper uncertainty limits for warm and cold seasons in future for all the four scenarios. However, there is a decrease in warm season length in the lower uncertainty limit. The decrease in season length was higher for A2 scenarios when compared with the rest of the scenarios.
- The windy and non-windy seasons coincided with wet and dry seasons.
- When the number of seasons were increased from two to five, the results of the analysis revealed that the distinct cluster could not be obtained when the number of seasons were increased beyond three. When the region was classified into three clusters, three distinct seasons were obtained and they coincide with cold, warm and wet seasons.

The present study brings out the variability in season length across scenarios, season type, and when the number of seasons is changed. The uncertainties in the season length to the choice of clustering methods and GCMs should also be considered to provide more general conclusions about the variation in the study region which would help policy makers and general public for realistic assessment, management and mitigation of natural disasters, and for sustainable development. Investigating these uncertainties is a future scope of study.

## Acknowledgements

The author thanks Prof. D Nagesh Kumar, Dr V V Srinivas, Prof. Ravi Nanjundiah, Dr R Ravi, Dr Muthuswamy Murugan, Prof. Gautham Sethi and Dr R K Yadav and the anonymous

reviewers for their support, encouragement and valuable inputs. The support from the Indian Institute of Science to carry out the research work is acknowledged.

### Annexure: Dependence measures

Three dependence measures used in this study are product moment correlation (Pearson 1896), Spearman's rank correlation (Spearman 1904a, 1904b), and Kendall's tau (Kendall 1951).

Let the probable predictor and predictand for month  $t$  be denoted as  $X_t$  and  $Y_t$ , respectively. Then the product moment correlation which measures the linear relationship between probable predictor and predictand is given by:

$$P = \frac{\sum_{t=1}^N (X_t - \bar{X})(Y_t - \bar{Y})}{N\sigma_X\sigma_Y},$$

where  $N$  refers to the number of months in the datasets;  $\bar{X}$  and  $\bar{Y}$  represent the means of predictor and predictand respectively, while  $\sigma_X$  and  $\sigma_Y$  represent the standard deviations of the same.

Spearman's rank correlation and Kendall's tau are the two nonparametric correlations used in this study which are estimated based on ranks assigned to data points in predictor and predictand datasets. The advantage of these rank correlations over the linear correlation stems from the use of ranks rather than numerical values of the predictor and the predictand variables for estimation of the correlations (Press *et al* 1992). Ranks are assigned to the  $N$  data points in each dataset after arranging them in increasing order of magnitude, such that the least value in the data has the first rank. Spearman's rank correlation ( $\rho$ ) is computed using the difference between the ranks of contemporaneous values of predictor and predictand ( $D_i$ ).

$$\rho = 1 - \frac{6 \sum_{i=1}^N D_i^2}{N(N^2 - 1)},$$

where  $D_i = \text{Rank of } X_t - \text{Rank of } Y_t$ .

Estimation of the Kendall's tau ( $\tau$ ) for a pair of predictor and predictand datasets involves preparation of  $N$  pairs of data ranks  $\{(u_i, v_i), i = 1, \dots, N\}$ , where  $u_i$  and  $v_i$  denote ranks of contemporaneous data points in the predictor and predictand datasets at  $i$ th time step respectively. Let two pairs of ranks be  $(u_j, v_j)$  and  $(u_k, v_k)$ . The two pairs are concordant if  $u_j > u_k$  and  $v_j > v_k$ , or if  $u_j < u_k$  and  $v_j < v_k$ , for which  $(u_j - u_k)(v_j - v_k) > 0$ . The two pairs are discordant, if  $u_j > u_k$  and  $v_j < v_k$ , or if  $u_j < u_k$  and

$v_j > v_k$ , for which  $(u_j - u_k)(v_j - v_k) < 0$ . A tied pair is neither concordant nor discordant, i.e.,  $(u_j - u_k)(v_j - v_k) = 0$ . The Kendall's  $\tau$  is calculated using the formula given below.

$$\tau = \frac{4\lambda}{N(N - 1)/2},$$

where  $\lambda$  is the difference between the number of concordant pairs and the number of discordant pairs. So, a high value of  $\lambda$  means that most pairs are concordant, indicating that the two rankings are consistent. Further,  $N(N - 1)/2$  is the total number of possible pairs of ranks. If there are a large number of tied pairs it should be adjusted accordingly. A positive value of  $\tau$  indicates that the ranks of both the variables increase together, whilst a negative correlation indicates that as the rank of one variable increases the rank of the other decreases. The Kendall coefficient has two advantages over the Spearman coefficient (Leach 1979). The first advantage is that it is appropriate when a large number of ties are present within ranks. The second advantage is its direct and simple interpretation in terms of probabilities of observing concordant and discordant pairs. The Spearman's coefficient can be considered as the regular Pearson's correlation coefficient in terms of the proportion of variability accounted for, whereas Kendall's coefficient represents a probability, i.e., the difference between the probabilities that the observed data are in the same order and the observed data are not in the same order. The advantages of Kendall coefficient makes it useful to effectively interpret the relationship between the predictors in NCEP and GCM data sets and between predictors in NCEP and the predictand.

### References

- 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 and Wratt D 2007 Climate Change 2007: The Physical Science Basis. Summary for Policy makers; Report of the 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 and Kumar D N 2008 Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine; *Int. J. Climatol.* **28(3)** 401–420.

- Argiriou A A, Kassomenos P A and Lykoudis S P 2004 On the methods for the delimitation of seasons: Water, Air, and Soil Pollution; *Focus* **4** 65–74.
- Chakraborty S, Tiedemann A V and Teng P S 2000 Climate change: Potential impact on plant diseases; *Environmental Pollution* **108** 317–326.
- Christidis N, Stott P A, Brown S, Karoly D J and Caesar J 2007 Human contribution to the lengthening of the growing season during 1950–99; *J. Climate* **20** 5441–5454.
- Cleugh H A, Miller J M and Böhm M 1998 Direct mechanical effects of wind on crops; *Agroforestry Systems* **41(1)** 85–112, 10.1023/A:1006067721039.
- Doty B and Kinter J L III 1993 The grid analysis and display system (GrADS): A desktop tool for earth science visualization; American Geophysical Union 1993 Fall Meeting, San Francisco, CA, 6–10 December.
- Fix E and Hodges J L 1951 Discriminatory analysis: Non-parametric discrimination: Consistency properties; USAF School of Aviation Medicine, Project 21-49-004, Report 4.
- Jones G S, Jones A, Roberts D L, Stott P A and Williams K D 2005 Sensitivity of global scale climate change attribution results to inclusion of fossil fuel black carbon aerosol; *Geophys. Res. Lett.* **32** L14701, doi: 10.1029/2005GL023370.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K C, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R and Joseph D 1996 The NCEP/NCAR 40-year reanalysis project; *Bull. Amer. Meteor. Soc.* **77(3)** 437–471.
- Kendall M G 1951 Regression structure and functional relationship, Part I; *Biometrika* **38** 11–25.
- Kripalani R H, Oh J H, Kulkarni A, Sabade S S and Chaudhari H S 2007 South Asian summer monsoon precipitation variability: Coupled climate model simulations and projections under IPCC AR4; *Theoretical and Applied Climatology* **90** 133–159.
- Lamb H H 1972 British Isles weather types and a register of daily sequence of circulation patterns, 1861–1971; *Geophysical Memoir* **116** HMSO, London, pp. 85.
- Leach C 1979 Introduction to statistics: A nonparametric approach for the social sciences (New York: Wiley).
- MacQueen J 1967 Some methods for classification and analysis of multivariate observation; In: *Proceedings of the fifth Berkeley Symposium on mathematical statistics and probability*, (eds) Le Cam L M and Neyman J (Berkeley: University of California Press) **1** 281–297.
- Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenmann J, Gaffin S, Gregory K, Grübler A, Jung T Y, Kram T, La Rovere E L, Michaelis L, Mori S, Morita T, Pepper W, Pitcher H, Price L, Raihi K, Roehrl A, Rogner H H, Sankovski A, Schlesinger M, Shukla P, Smith S, Swart R, van Rooijen S, Victor N and Dadi Z 2000 IPCC Special report on emissions scenarios, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 599.
- Patterson D T, Westbrook J K, Joyce R J V, Lingren P D and Rogasik J 1999 Weeds, insects and diseases; *Climatic Change* **43** 711–727.
- 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.
- Peng S, Huang J, Sheeshy J E, Laza R C, Visperas R M, Zhong X, Centeno G S, Khush G S and Cassman K G 2004 Rice yields decline with higher night temperature from global warming; *Proceedings of National Academy of Science, USA* **101** 9971–9975.
- Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 Numerical recipes in Fortran 77: The art of scientific computing (New York: Cambridge University Press).
- Rajeevan M, Bhate J, Kale J D and Lal B 2005 Development of a high resolution daily gridded rainfall data for the Indian Region (version 2), Meteorol. Monogr. Climatol. 22/2005, India Meteorol. Dept., New Delhi.
- Rajeevan M, Bhate J, Kale J D and Lal B 2006 High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells; *Curr. Sci.* **91(3)** 296–306.
- Rupa Kumar K, Sahai A K, Krishna Kumar K, Patwardhan S K, Mishra P K, Revadekar J V, Kamala K and Pant G B 2006 High-resolution climate change scenarios for India for the 21st century; *Curr. Sci.* **90** 334–344.
- Sailor D J, Smith M and Hart M 2008 Climate change implications for wind power resources in the Northwest United States; *Renewable Energy* **33** 2393–2406.
- Spearman C E 1904a ‘General intelligence’ objectively determined and measured; *Am. J. Psychol.* **5** 201–293.
- Spearman C E 1904b Proof and measurement of association between two things; *Am. J. Psychol.* **15** 72–101.
- Tripathi S, Srinivas V V and Nanjundiah R S 2006 Downscaling of precipitation for climate change scenarios: A support vector machine approach; *J. Hydrol.* **330** 621–640, doi: 10.1016/j.jhydrol.2006.04.030.
- Tuller S E 1990 Standard seasons; *Int. J. Biomet.* **34** 181–188.
- Winkler J A, Palutikof J P, Andresen J A and Goodess C M 1997 The simulation of daily temperature time series from GCM output Part II: Sensitivity analysis of an empirical transfer function methodology; *J. Climate* **10(10)** 2514–2532.
- Yadav R K, Rupa Kumar K and Rajeevan M 2007 Role of Indian Ocean sea surface temperatures in modulating northwest Indian winter precipitation variability; *Theoretical and Applied Climatology* **87** 73–83, doi: 10.1007/s00704005-0221.
- Yadav R K, Rupa Kumar K and Rajeevan M 2009a Increasing influence of ENSO and decreasing influence of AO/NAO in the recent decades over northwest India winter precipitation; *J. Geophys. Res.-Atmos.* **114** D12112, doi: 10.1029/2008JD011318.
- Yadav R K, Rupa Kumar K and Rajeevan M 2009b Out-of-phase relationships between convection over north-west India and warm-pool region during winter season; *Int. J. Climatol.* **29** 1330–1338, doi: 10.1002/joc.1783.
- Yadav R K, Rupa Kumar K and Rajeevan M 2010a Climate change scenarios for northwest India winter season; *Quaternary International* **213** 12–19.
- Yadav R K, Yoo J H, Kucharski F and Abid M A 2010b Why is ENSO influencing northwest India winter precipitation in recent decades?; *J. Climate* (in press).