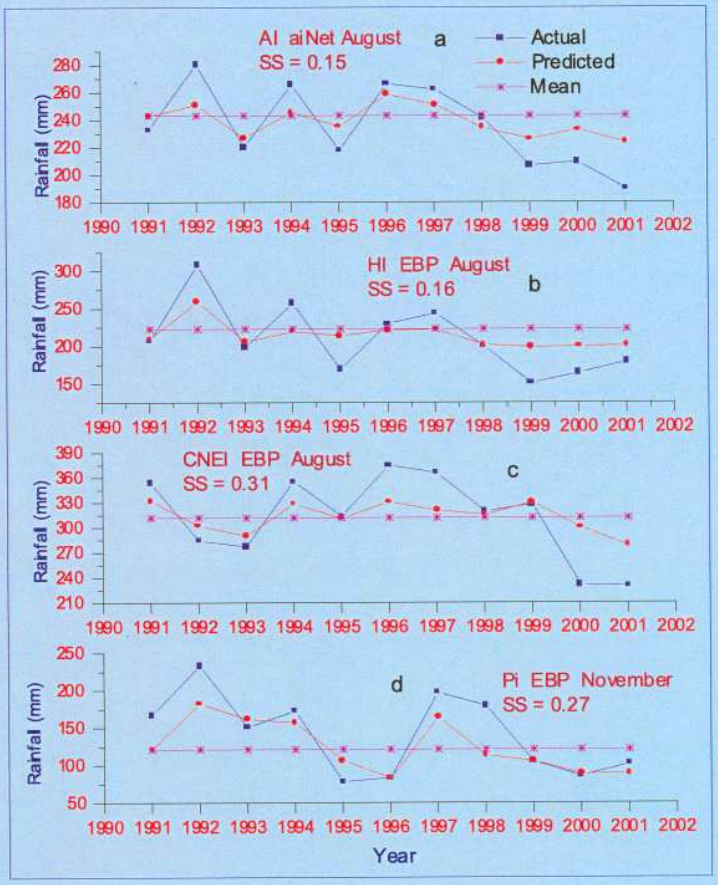
ISSN 0252-1075
Contribution from IITM
Research Report No. RR-101

Multimodel Scheme for Prediction of Monthly Rainfall over India



J.R. Kulkarni, Savita G. Kulkarni, Yaogesh Badhe
Sanjeev S. Tambe, Bhaskar D. Kulkarni and G.B. Pant

December 2003

Indian Institute of Tropical Meteorology
Pune - 411 008, India

ISSN 0252-1075
Contribution from IITM
Research Report No. RR-101

Multimodel Scheme for Prediction of Monthly Rainfall over India

J.R. Kulkarni, Savita G. Kulkarni, Yogesh Badhe,
Sanjeev S. Tambe, Bhaskar D. Kulkarni
and
G.B. Pant

December 2003

Indian Institute of Tropical Meteorology
Dr. Homi Bhabha Road, Pashan Pune - 411 008
Maharashtra, India

E-mail : lip@tropmet.ernet.in
Web : http://www.tropmet.ernet.in
Fax : 91-020-5893825
Telephone : 91-020-5893600

Contents

Multimodel Scheme for Prediction of Monthly Rainfall over India

J. R. Kulkarni, Savita G. Kulkarni*, Yogesh Badhe*, Sanjeev S. Tambe*,
Bhaskar D. Kulkarni* and G. B. Pant
Indian Institute of Tropical meteorology, Pashan Pune 411008

Abstract

Multimodel scheme for the prediction of monthly rainfall over India has been developed. It is a linear combination of three different nonlinear time series models. The models are EBP (based on error back propagation algorithm of Artificial Neural Networks ANN), (2) aiNet (based on probabilstic neural network theory) and SVR (based on Support Vector Machine theory). Separate schemes are developed for (1) all India and (2) seven macro climatic zones (homogeneous, core, northwest, west central, central northeast, northeast and peninsular) of India. Monthly rainfall data of these 8 regions for the period 1871-2001 have been used in the study. The data for the period 1871-1990 has been used for training and data for the period 1991-2001 has been used for testing. The data is preprocessed by applying principle component analysis to overcome multicollinearirty problem and dimensionality reduction. The models showed capability of simulating the monthly rainfall variability over all the regions quite well. The performance of the models has been found to vary from month to month and region to region. The performance of the models in predicting monthly rainfall has been tested by computing the skill scores. The highest skill score acieved is 0.31. Taken all the 8 regions together, this multimodel approach shows skills in the monthly predictions in 54%, 46%, 63% and 50% months in the winter, premonsoon, monsoon and post monsoon season, respectively. The skills scores are low, but they are positive in 56% of cases.

*National Chemical Laboratory, Pashan, Pune 411008

1. Introduction

India's agriculture, water resources and industrial production depend substantially on the summer monsoon season (June - September) rainfall. The prediction of Indian summer monsoon Rainfall (ISMR) has been found immense useful in taking early actions for country's water management strategy, mitigating ensuing drought conditions etc. The ISMR predictions have been carried out for more than century in India. The subject has been reviewed from time to time by various workers (Normand 1953, Shukla 1987, Krishna Kumar et al. 1995, Webster et al. 1998).

The approaches used for ISMR prediction can be classified into three groups: (1) statistical (2) dynamical and (3) machine learning / artificial intelligence. The statistical approach has a long history and has been most widely used. Diagnostic studies of historical data sets have discovered several parameters for ISMR prediction. These parameters represent different components of the coupled atmosphere-ocean-land system. Using these parameters, a variety of regression models have been developed. (Blanford 1884, Walker 1908, 1918, 1923, Thapliyal 1990, Shukla and Mooley 1987, Bhalme et al 1986, Pathasarathy et al. 1988, Gowariker et al. 1989, 1991). The reliability of the statistical models depends crucially on the stability of the relationships between the predictand and predictor parameters. The existence of secular variation in the strength of correlations between ISMR and some predictor parameters has been known since Walker's time. Therefore there has been always updates of the statistical models. The most successful, widely referred as 16 parameter Gowariker model (Gowariker et al. 1991), has been changed to 8 parameter model from April 2003 (www.imd.gov.in).

The General Circulation Models (GCMs) are being used for the dynamical prediction of ISMR. GCMs are basically the mathematical formulations of different atmosphere, ocean and land surface processes. These formulations are based on the classical physical principles. They represent a potentially powerful tool for the long-range weather and climate predictions of the earth-atmosphere system. Although, the day to day predictability of the weather is limited to two weeks, the seasonal prediction of ISMR is possible because the interannual variability of ISMR is mostly governed by slowly varying boundary conditions. The GCMs are integrated in time with prescribed initial and boundary conditions to generate the circulations in the future times. ISMR predictions using GCMs have been carried out in IITM (Indian Institute of Tropical Meteorology) since 1997. Two GCMs viz. COLA (Center for Ocean Land Atmosphere USA) and HadAM2 b (from Hadley center for climate prediction and research U.K.) have been integrated in the ensemble mode to generate the ISMR forecasts (Kulkarni et al. 2001). The ISMR predictions by this dynamical method agree qualitatively well with the observed rainfall, and therefore they are found very useful in supplementing the operational statistical predictions. Recently Krishnamurti et al. (1999) introduced the concept of multimodel forecasting for dynamical prediction approach. Each model has strengths and weaknesses due to usages of variety of formulations, methodologies, assumptions etc. The multimodel approach tries to combine the strengths of the individual models. In this approach, the predictions from the different models are regressed against the observed fields Graham et al. (2000) showed advantages of this approach using PROVOST data set.

The third and recent approach of ISMR prediction is based on the techniques in the machine learning or artificial intelligence area. In this approach 'Artificial Neural Networks' (ANN) are being used (Navone and Ceccatto 1994, Goswami and Srividya 1996, Venkateshan et al. 1997, Sahai et al. 2000). The error back propogation algorithm is used for estimating the

weights of the neurons in ANN. The performance of ANN based models in ISMR predictions are found quite encouraging.

In addition to seasonal monsoon prediction, there is a strong need for monthly rainfall prediction over entire India as well as smaller sub regions of the country. These types of predictions will not only supplement the seasonal monsoon forecast but will provide more utility than the seasonal forecast. It is well known fact that when ISMR for country as whole is excess, there are some parts of the country where the seasonal rainfall is below normal or scanty. Similarly when ISMR for country as a whole is deficient, some parts have excess rainfall. The monthly rainfall prediction over smaller areas will be able to provide the regionality in the rainfall distribution. At present no such method is available for monthly rainfall prediction over India. In the present research report a novel scheme is proposed for the monthly rainfall prediction. The rainfall in the post monsoon and winter months is generally less but for the certain parts of the country, it gives a significant contribution. The monthly rainfall scheme proposed here predicts rainfall for all months, which includes the pre monsoon, post monsoon and winter months also. The scheme is based on the three techniques viz. (1) EBP (based on error back propagation algorithm of Artificial Neural Networks ANN), (2) SVR (based on Support Vector Machine theory) and (3) aiNet (based on probabilstic neural network theory).
The monthly prediction scheme has an additional advantage. The accuracy of ANN based prediction models very much depends upon the data size (Haykins 1999). With the available ISMR data from 1971-2001, the size of the training and test together is limited to 131 points, which imposes restrictions in the selection of best possible ANN architecture Increased data size, permits us to try all possible combinations (number of hidden layers, number of neurons in each layer) to increase the accuracy of the predictions.

The mutimodel scheme is developed using the linear combination of these three models. The methods are described in the following sections.

2. Error Back Propagation (EBP)

Error back propagation method (EBP), proposed by Rumelhart et al. (1986), is the most widely used algorithm in ANN for time series predictions. The detailed description of EBP technique can be found in Haykin (1994); and the references given in Navone and Ceccatto (1994), Venkateshan et al. (1997). However for the sake of completeness, the technique is briefly described here. It is based on a gradient descent technique which minimizes the average-squared-error between the actual and network predicted output by moving down the error surface. An EBP network consists of at least three layers of neurons, namely an input layer, an output layer and a middle layer (known as hidden layer). Each neuron in a layer is fully connected to the neurons of cosecutive layer and the strength of the connection is known as weights. At the beginning of network training, the weights are randomly initialized. Each node in the hidden layer first computes the weighted-sum of inputs which is passed through a nonlinear transfer function to arrive at the output. The common choice of nonlinear transfer function is either the logistic sigmoid or hyperbolic tangent. The output values of the hidden layer neurons serve as inputs to the next layer neurons. The output of the output layer nodes forms the network output. Subsequently, the network outputs are compared with the target (desired) values and the prediction error is estimated. This error is used to correct the connection weights between the hidden and output layers and subsequently those between the hidden and input layers. The changes in the weights are made by presenting the input-output patterns of the training data repeatedly until a prespecified error criterion is fulfilled. At this stage the weights are said to be converged. In

order to speed up the convergence of the weights, a factor 'momentum term' is included in the formualtion. Other important parameter, required to determine is the 'learning rates'. EBP model is said to be developed when the values of connecting weights, leaning rates and momentum terms are determined.

3. Support Vector Regression (SVR)

In recent years, support vector regression (SVR) Vapnik (1995, 1996, 1998) , which is a statistical learning theory based machine learning formalism is gaining popularity due to its many attractive features and promising empirical performance. The salient features of SVR are: (i) like ANNs, SVR is an exclusively data based nonlinear modeling paradigm, (ii) SVR-based models are based on the principle of structural risk minimization, which equips them with greater potential to generalize, (iii) parameters of an SVR model are obtained by solving a quadratic optimization problem, (iv) the objective function in SVR being of quadratic form, it possesses a single minimum thus avoiding the heuristic procedure involved in locating the global or the deepest local minimum on the error surface, and (v) in SVR, the inputs are first nonlinearly mapped into a high dimensional feature space wherein they are correlated linearly with the output. Although the foundation of the SVR paradigm was laid down in mid 1990s, its applications in the field of atmospheric sciences have not started yet.

3.1 SVR-based modeling

Support vector regression is an adaptation of a recent statistical learning theory based classification paradigm, namely *support vector machines* (Vapnik 1995). The SVR formulation follows structural risk minimization (SRM) principle, as opposed to the empirical risk minimization (ERM) approach which is commonly employed within statistical machine learning methods and also in training ANNs. In SRM, an upper bound on the generalization error is minimized as opposed to the ERM, which minimizes the prediction error on the training data. This equips the SVR with a greater potential to generalize the input-output relationship learnt during its training phase for making good predictions for new input data. The SVR is a linear method in a high dimensional feature space, which is nonlinearly related to the input space. Though the linear algorithm works in the high dimensional feature space, in practice it does not involve any computations in that space, since through the usage of kernels, all necessary computations are performed directly in the input space. In the following, the basic concepts of SVR are introduced. A more detailed description of SVR can be found in Vapnik (1995, 1998), Burges (1998), Smola et al. (1998), and Schölkoff et al. (2001).

Consider a training data set $\gamma = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_P, y_P)\}$, such that $\mathbf{x}_i \in \upsilon^N$ is a vector of input variables and $y_i \in \upsilon$, is the corresponding scalar output (target) value. Here, the modeling objective is to find a regression function, $y = f(\mathbf{x})$, such that it accurately predicts the outputs $\{y\}$ corresponding to a new set of input examples, $\{\mathbf{x}\}$, which are drawn from the same underlying joint probability distribution, $P(\mathbf{x}, y)$, as the training set. To fulfill the stated goal, SVR considers following linear estimation function.

$$f(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle + b \qquad (1)$$

where $\mathbf{w}$ denotes the weight vector; b refers to a constant known as "bias"; $\Phi(\mathbf{x})$ denotes a function termed *feature*, and $\langle \mathbf{w}, \Phi(\mathbf{x}) \rangle$ represents the dot product in the feature space, ι (such that Φ: $\mathbf{x} \rightarrow \iota$, $\mathbf{w} \in \iota$). In SVR, the input data vector, $\mathbf{x}$, is mapped into a high dimensional feature space, ι, via a nonlinear mapping function, Φ, and a linear regression is

performed in this space for predicting y. Thus, the problem of nonlinear regression in lower dimensional input space υ^N is transformed into a linear regression in the high dimensional feature space, ι. Accordingly, the original optimization problem involving nonlinear regression is transformed into finding the flattest function in the feature space ι and not in the input space, $\mathbf{x}$. The unknown parameters $\mathbf{w}$ and b in Eq. 1 are estimated using the training set, γ. To avoid over fitting and thereby improving the generalization capability, following regularized functional involving summation of the empirical risk and a complexity term $\|\mathbf{w}\|^2$, is minimized (Burges 1998):

$$\begin{aligned} R_{reg}[f] &= R_{emp}[f] + \lambda \|\mathbf{w}\|^2 \\ &= \sum_{i=1}^{P} C(f(\mathbf{x}_i) - y_i) + \lambda \|\mathbf{w}\|^2 \end{aligned} \tag{2}$$

where, R_{reg} and R_{emp} denote the regression and empirical risks, respectively; $\|\mathbf{w}\|^2$ is the Euclidean norm; $C(.)$ is a cost function measuring the empirical risk, and $\lambda > 0$, is a regularization constant. For a given function, f, the regression risk (test set error), $R_{reg}(f)$, is the possible error committed by the function f in predicting the output corresponding to a new (test) example input vector drawn randomly from the same sample probability distribution, $P(\mathbf{x}, y)$, as the training set. The empirical risk $R_{emp}(f)$, represents the error (termed "training set error") committed in predicting the outputs of the training set inputs. Minimization task described in Eq. 2 involves: (i) minimization of the empirical loss function $R_{emp}(f)$ and, (ii) obtaining as small a $\mathbf{w}$ as possible, using the training set γ. The commonly used cost function is the "ε-insensitive loss function" given as (Vapnik 1998):

$$C(f(\mathbf{x}) - y) = \begin{cases} |f(\mathbf{x}) - y| - \varepsilon \text{ for } |f(\mathbf{x}) - y| \geq \varepsilon \\ 0 \qquad \text{otherwise} \end{cases} \tag{3}$$

where ε is a precision parameter representing the radius of the tube located around the regression function (see Figure 1); the region enclosed by the tube is known as "ε-insensitive zone". The SVR algorithm attempts to position the tube around the data as shown in figure 1. The optimization criterion in Eq. 3 penalizes those data points whose y values lie more than ε distance away from the fitted function, $f(\mathbf{x})$. In figure 1, the size of the stated excess positive and negative deviations are depicted by ξ and ξ^*, which are termed "slack" variables. Outside of the $[-\varepsilon, \varepsilon]$ region, the slack variables assume non-zero values. The SVR fits $f(\mathbf{x})$ to the data in a manner such that: (i) the training error is minimized by minimizing ξ_i and ξ_i^* and, (ii) $\|\mathbf{w}\|^2$ is minimized to increase the flatness of $f(\mathbf{x})$ or to penalize over complexity of the fitting function. Vapnik (1998) showed that the following function possessing finite number of parameters can minimize the regularized function in Eq. 2.

$$f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \sum_{i=1}^{P} (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x}_i) + b \tag{4}$$

where, α_i and α_i^* (≥ 0) are the coefficients (Lagrange multipliers) satisfying $\alpha_i \alpha_i^* = 0$, $i = 1, 2, \ldots, P$, and $K(\mathbf{x}, \mathbf{x}_i)$ denotes the so called 'kernel' function describing the dot product in the feature space. The kernel function is defined in terms of the dot product of the mapping function as given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \tag{5}$$

The advantage of this formulation (Eqs. 4 and 5) is that for many choices of the set $\{\Phi_i(\mathbf{x})\}$, including infinite dimensional sets, the form of K is analytically known and very simple (Mukherjee et al. 1997). Accordingly, the dot product in the feature space ι can be computed without actually mapping the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ into that space (i.e., computation of $\Phi(\mathbf{x}_i)$ and $\Phi(\mathbf{x}_j)$). There exist several choices for the kernel function K; the two commonly used kernel functions, namely, radial basis function (RBF) and nth degree polynomial are defined below in Eqs. (6) and (7), respectively.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{6}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left[1 + (\mathbf{x}_i, \mathbf{x}_j)\right]^n \tag{7}$$

In Eq. (4), the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are obtained by solving following quadratic programming problem.
Maximize:

$$R(\boldsymbol{\alpha}^*, \boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i,j=1}^{P}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon\sum_{i=1}^{P}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{P} y_i(\alpha_i^* - \alpha_i) \tag{8}$$

Subject to constraints: $0 \le \alpha_i, \alpha_i^* \le C$, $\forall i$ and $\sum_{i=1}^{P}(\alpha_i^* - \alpha_i) = 0$. Having estimated $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^*$ and b, using a suitable quadratic programming algorithm, the SVR-based regression function takes the form

$$f(\mathbf{x}, \mathbf{w}) = f(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \sum_{i=1}^{P}(\alpha_i^* - \alpha_i)K(\mathbf{x}_i, \mathbf{x}) + b \tag{9}$$

where, vector $\mathbf{w}$ is described in terms of the Lagrange multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$. Owing to the specific character of the above-described quadratic programming problem, only some of the coefficients, $(\alpha_i^* - \alpha_i)$, are non-zero and the corresponding input vectors, $\mathbf{x}_i$, are called support vectors (SVs). The SVs can be thought of as the most informative data points that compress the information content of the training set. The coefficients α_i and α_i^* have an intuitive interpretation as forces pushing and pulling the regression estimate $f(\mathbf{x}_i)$ towards the measurements, y_i (Muller et al. 1997).

In Eq. (9), the bias parameter, b, can be computed as follows;

$$b = \begin{cases} y_i - f(\mathbf{x}_i)_{b=0} - \varepsilon & for\ \alpha_i \in (0, C) \\ y_i - f(\mathbf{x}_i)_{b=0} + \varepsilon & for\ \alpha_i^* \in (0, C) \end{cases} \tag{10}$$

where, $\mathbf{x}_i$ and y_i respectively denote the ith support vector and the corresponding target output, respectively. In the SVR formulation, C and ε are two user-specified free parameters; while C represents the trade-off between the model-complexity and the approximation error, ε signifies the width of the ε-insensitive zone used to fit the training data. The stated free parameters together with the specific form of the kernel function control the accuracy and generalization performance of the regression estimate. The procedure of judicious selection of C and ε is explained by Cherkassky and Ma (2002).

4 aiNet

aiNet is recently developed software tool for solving the prediction problems (Kranjnc 1997). It is based on the concept of joint probability distribution between input and output vectors. The formulations of aiNet model has been described briefly in the following paragraphs. The detailed statistical and mathematical treatment can be found in the model documentation paper (aiNet manual 1997). The following description of aiNet is mostly taken from Guh (1998).

Suppose we have a number of cases to set up a model. Each case (a model vector) consists of the same variables. The variables consists of the input variables and output variables:

$$X = \left\{ \underbrace{x_1, x_2 \ldots\ldots\ldots\ldots x_n}_{input} \; \underbrace{x_{n+1}, x_{n+2} \ldots\ldots\ldots\ldots}_{output} \right\} = \{Xi, Xo\} \qquad (11)$$

aiNet predicts the output part based on the input part. The aiNet algorithm is based on a few assumptions.

Firstly, all the model vectors are equally important with the same accuracy. Suppose we have a model vector P that is only known for the input part. Its output part is calculated based on the model M (called "C") that has the same input part as P. According to this assumption the output part of P will be the same as that of C. However, it is very unlikely that a perfect match exits in reality. Thus a second probability-based assumption is needed that states that if the input parts of the model vectors P and C are "near", there is a high probability that the output part of C is similar to the output part of P. Conversely, if the input parts of P and C are "far", there is only a low probability (figure 2).

The term "probabilty" is replaced by "similarity" hereafter. The qualitative descriptive terms "near" and "far" are quantified using some measure of distance between the two vectors. The two vectors are near if the vector norm of their difference is a small value. Usually, a Euclidean norm is used in such cases. The Euclidean norm for the difference (distance) between the two vectors P and C is given by

$$d_{pc} = \sqrt{\sum_i (x_{pi} - x_{ci})^2} \qquad (12)$$

When the distance between the input parts of the model vectors are defined, a probability (Gaussian) function can be selected. Moreover, if the probability function and the distance are known, the similarity can be calculated. The similarity between P and C is represented by

$$S_{pc} = e^{-d_{pc}^2 / \alpha} \qquad (13)$$

The equation (13) is a slightly modified form of the Gaussian function where the standard deviation (SD) is replaced by a constant α. This constant is called as "penalty coefficient" and is selected a priori. It has a significant influence on the shape of the probability function as shown in the figure 3.

If α is very small, then the Gaussian function is so steep that even at a small distance d_{pc} the similarity of their output parts will be low -close to zero. However, if α is very large, then the similarity of their output parts will be very high -close to 1.

Similarly, for each model vector M we can calculate its distance from P and their similarity. Finally, we determine how each model vector contributes to the output part of P. It is known that each model vector from M contributes according to its calculated similarity to P. To simplfy the the final calculation of the output part of P, the similarity coefficients must be normalised; that is, their sum must be equal to 1:

$$\bar{s}_{px} = \frac{s_{px}}{\sum_i s_{pi}}; \quad \sum \bar{s}_{pi} = 1.0 \tag{14}$$

The index 'i' in equation (14) runs over all the model vectors and the index 'x' can represent any model vector in the model. Once the similarity coefficients are normalized the final result is obtained by a combination of the output parts all the model vectors. The multipliers in this combination are the normalised similarities of each model vector with respect to the vector P. The output of the model vector P_o is given by

$$P_o = \sum_x \bar{s}_{px} X_o \tag{15}$$

Knowing the right value of the penalty coefficient solves the prediction problem. The attempts are made with different values of it and finally the selection is made of the most succesfull one. According to theory, there is only one optimal value for the penalty coefficient. However the results are not sensitive to the optimal value. Another features associated features of aiNet are:

- Ability to dynamically change the "knowledge base". This means that new data may be added to the neural network (or old data removed), additional variables may be added (or old ones removed) and answers obtained right away - there is no time consuming learning phase.
- aiNet provides a way to estimate the rate of error in the prediction. If the problem is smooth i.e. without noise, then this error will represent an estimation for the error in the predicted result. If the data is noisy, then this will represent an estimation for the noise around the predicted result. This means that error estimation behaves locally.
- aiNet can cope with missing values in the data. In real life it is usually very difficult to find a perfectively assembled knowledge base—there is always some data missing. aiNet handles missing data automatically without requiring representation of the missing data.

5. Data

Monthly rainfall data for (1) All India (AI), (2) Homogeneous India (HI) (3) Core India (CI) and (4) five macro climatic zones of India have been used in the study. The details of the rainfall data can be found from Parthasarathy et al. (1995). Figure 4a (from Parthasarathy et al.1995) shows the meteorological sub-divisions considered for the preparation of AI rainfall time series. Figure 4b, c shows the sub divisions considered for HI and CI rainfall time series respectively. Figure 2d shows five macro climatic zones of India. These are: (1) Northwest India (NWI) (2) West central India (WCI) (3) Central northeast India (CNEI) (4) Northeast India (NEI) and (5) Peninsular India (PI) respectively (Parthasarathy et al. 1995). All these regions are referred as 8 regions in the following discussions. The period of data used is from 1871-2001.

6. Results and Discussion

Figure 5 shows the mean monthly rainfall variation in the year for 8 regions. There is a strong annual cycle with peak in the month of July (associated with monsoon rainfall) and minimum in the month of January for all the regions except PI where a bimodal distribution is seen. The rainfall is very small in the months January through May, and there is a steep increase in the month of June with the onset of monsoon for all the regions, except NEI, where a good amont of rainfall is observed in the winter and pre monsoon months, and the transition to monsoon season is smooth. The annual cycle of rainfall over NEI is well behaved. The second peak in the month of October over PI corresponds to the post monsoon rainfall due to withdrawal of summer monsoon and cyclonic activity.

For the development of the models, the data is divided into training period 1871-1990 and test period 1991-2001. It is well known fact that accuracy of models increase with the increase in the size of the training data set. The partition of the data into training and test has been done to accommodate the sufficiently large data set for the training. The results are found to be stable with varying sizes of the training data. As monthly rainfall data for each of the regions contain large mutual correlations, it is pre processed by applying principle component analysis (PCA). In addition to overcome the muticolinearity problem, PCA does the dimensionality reduction.The uncorrelated input variables accelerates the EBP learning process (Haykin 1999). First 8 principle components, which contain more than 90% of the variance, are used in the development of the models. Table 1 provides the distributions of variances in the 8 PCs for the 8 regions considered in the study. Figure 6 shows first 8 eigen vectors for AI rainfall time series. Similar variations of the eigen vectors are found for other regions. It is seen that the pair of first two eigen vectors represent the annual cycle in the rainfall. These contain bulk of the variances of the PCs. The pair of eigen vectors 3-4, and 5-6 represent semiannual and seasonal cycles respectively. The eigen vector pair 7-8 represents the noise part. Considering all the 8 PCs, the maximum total variance 97.9% is found in AI time series, and minimum 91.0% in NWI rainfall series. In the NEI rainfall, first two PCs capture maximum variance of magnitude 86%, the reason for which may be attributed to the well behaved rainfall annual cycle over the region. As the loadings have both positive and negative signs, they are further scaled to have values between 0-1. These scaled PCs are the basic inputs to the models.

6.1 Simualtion of Annual Rainfall Variability

Using the training data set, three models viz. EBP, SVR and aiNet are developed for the 8 regions. Table 2 shows the architectures of different EBP models developed for 8 regions. The learning rate and momentum coefficients are 0.05 and 0.9 respectively for all the models. SVR implementation known as "ε-SVR" in the LIBSVM software library (Chang and Lin 2001), has been used to develop the 8 SVR-based models. The LIBSVM package utilizes a fast and efficient method known as sequential minimal optimization (SMO) (Joachims 1998, Platt 1998) for solving large quadratic programming problems and thereby estimating function parameters α, α^* and b (see Eq. 9). To obtain an optimal SVR model, it is necessary to examine the effects of kernel function and other algorithm-specific parameters; the three kernel functions that were tested are, polynomial, RBF and sigmoid. Among these, RBF resulted in the least RMSE values for the training and test sets of the outputs, y_1 and y_2. The numbers of support vectors used by the SVR algorithm for fitting the rainfall are given in the Table 3. The optimal values of the four SVR-specific parameters namely, width of RBF kernel (σ), cost coefficient (C), loss function parameter (ε_{loss}) and tolerance for termination

criterion (ε_{tol}) that minimized the E_{trn} and E_{tst} are listed in Table 4. Using aiNet software, 8 models are developed for different regions. The penalty coefficients used in these models are given in the Table 5.

The annual rainfall shows strong seasonal cycle, with maximum in the monsoon season and minimum in the winter season over most of the zones. The standard deviations (SDs) and coefficient of variability (CV) of monthly rainfalls are very high compared to that of monsoon seasonal rainfall. The mean, SDs and CVs of monsoon seasonal rainfall are 852.0 mm, 80.0 mm and 9.4% whereas the corresponding values of AI monthly rainfall are 90.88 mm, 95.77 mm and 105% respectively. The columns 4 and 5 in Table 6 provide the distribution of CVs for the 8 regions. The highest CV (158.8%) for monthly rainfall has been observed over NWI and the lowest CV (81.5 %) is found over PI for the training period 1871-1990. The performances of the models have been tested by examining how well this large variability has been simulated. The columns 6 and 7 in Table 6 provide the distribution of CVs by the model simulated rainfalls for 8 regions. The model simulated CV values are found in agreement with the observed CV values. For the test period, model simulated CV values are found smaller in comparison with the observed CV values. Table 7 shows the distributions of (1) correlation coefficients (CCs) between observed and predicted, (2) means and (3) SDs of predicted and observed monthly rainfalls for all the regions. The CCs are very high for both training and test data sets for all the regions. For NWI CCs for the test period are found to decrease drastically compared to that in the training period. Similarly the means and the SDs are nearly same for both training and test periods for all the zones except for NWI. High CV over NWI may be probable reason for large differences between CCs, SDs for training and test periods. In general the models are capable of simulating the monthly rainfall variability quite well. All the three models show nearly equal performances.

6.2 Multimodel Prediction of Monthly Rainfall

Combining the three models, the multimodel scheme is developed. The regression equation for the multimodel scheme is of the form:

$$MMR(I) = a \times EBP(I) + b \times SVR(I) + c \times aiNet(I) \qquad (20)$$

where MMR(I) is the monthly predicted rainfall by multimodel scheme in the month I, EBP(I), SVR(I), and aiNet(I) are the monthly predicted rainfalls by EBP, SVR and aiNet models for the I th month. a, b, and c are the regression coefficients. Table 8 shows the distribution of coefficients for all the months and for all the regions.

The model skill in the prediction has been tested by computing the parameter 'skill score' 'SS'.

$$SS = 1 - \frac{MSE_P}{MSE_m} \qquad (21)$$

where MSEp and MSEm are the mean-square-errors computed using observed minus predicted and climatological-normal minus predicted values (Wilks 1995). The model has skill in prediction if SS > 0, SS = 1 indicates perfect prediction. The positive value of SS gives model's capability in the predictions above the climatological forecasts.

Table 9 shows the distribution of SS values for different months for the test period for all the 8 regions. The months are arranged according to seasons. The upper values in the table show SS values by the multimodel scheme. It has been observed that the multimodel

approach does not necessarily improve the performance over the individual models. The lower values show the skills in cases where the performance of the individual model is better than the combined one. Here 'a', 'e' and 's' represent aiNet, EBP, and SVR respectively. It is seen that the highest SS value is 0.31 by aiNet and EBP for March and August months for CNEI and for April month by aiNet for PI. For winter months (December, January, February) the maximum SS value (0.15) is observed for January over HI and CI. For post monsoon season (October, November), the maximum SS (0.27) has been observed in November over PI by EBP model. In the monsoon season, the SS values are 0.31, 0.16 and 0.15 for the month August over CNEI, HI and AI, respectively. Figure 7 a, b show the observed and predicted values of monthly rainfalls for these regions. The figure also shows climatological monthly mean value. It is clearly seen that the predicted values are lying in between the observed and climatological mean values. It shows that the positive skill in the prediction is consistent throughout the test period.

The column 15 shows the number of months for which the models have skills in the predictions for each region. The upper values show the number of months for which skills are produced by multimodel approach and lower values show combined skills scores. As stated above, it is seen that sometimes, individual model shows better skill than the multimodel scheme. The lower values in the column 15 include skills by both individual and multimodel scheme taken together. It is seen that the maximum skill is in 11 months for HI and the lowest skill in 3 months for NWI. For 8 regions there are total 96 months in a year. This approach shows skill in 52 months. For the winter months the multimodel approach has skill in 10 months out of 24 months and multimodel plus individual models have skills in 13 months. In the premonsoon months, these values are 5 / 24 and 11 / 24 respectively. In the monsoon months, the multimodel has skills in the 13 months and multimodel plus individual model has skills in 20 months out of 32 months. In the post monsoon months these values are 5 / 16 and 8 / 16 respectively. Taken all the 8 regions together, this multimodel approach shows skills in the monthly predictions in 54%, 46%, 63% and 50% months in the winter, premonsoon, monsoon and post monsoon season, respectively. Considering all the regions and all the months, this approach shows skills in 54 months out of 96 months. The skills scores are low, but they are positive in 56 % of cases. Thus the results clearly demonstrate the potential of the this new approach in the monthly predictions. Nevertheless, more efforts are required to improve the models for achieving better skill scores.

7. Conclusions

Monthly rainfall predictions for 8 regions over India have been attempted using multimodel approach. In this approach, the outputs of the three different models viz. EBP, AiNet and SVR are combined to generate final predictions. The models show capability of simulating the large annual variability consistently well. It is observed that no single model can be selected as the best model, and the performance of the model varies from month to month and region to region. The multimodel generally performs well over individual models. The highest skill score achieved is 0.31. Taken all the 8 regions together, this multimodel approach shows skills in the monthly predictions in 54%, 46%, 63% and 50% months in the winter, premonsoon, monsoon and post monsoon season, respectively. The skills scores are low, but they are positive in 56% of cases. More work is required to improve the performance of the models.

8. References

Bhalme H N, Jadhav S K, Mooley D A, and Ramana Murty Bh V. 1986 Forecasting of monsoon performance over India, J. Climatol., 6, 347-354.

Blanford H H 1884 On the connection of Himalayan snowfall and seasons of drought in India, Proc. R. Soc. London, 37, 3-22.

Burges, C.,1998 A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2(2) 1-47.

Chang C. and C. Lin, 2001 "LIBSVM: a library for support vector machines," software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm .

Cherkassky V., and Y. Ma, 2002 Practical selection of SVM parameters and noise estimation for SVM regression, Submitted to *Neurocomputing* (special issue on SVM).

Goswami P, Srividya 1996 A novel neural network design for long-range prediction of rainfall pattern, Curr. Sci., 70, 447-457.

Gowariker V, Thapaliyal V, Sarker R P, Mandal G S, and Sikka D R, 1989 Prametric and power regression models - New approach to long range forecasting, Mausam, 40, 115-122.

Gowariker V, Thapaliyal V, Kulshreshtha S M, Mandal G S, Sen Roy N, Sikka D R, 1991 A power regression model for long-range forecast of southwest monsoon rainfall over India, Mausam 42, 125-130.

Graham R J, Evans A D L, Mylne K R, Harrison M S J, and Robertson K B, 2000 An assessment of seasonal predictability using atmospheric GCMs, Q J, R. Meteorol. Soc., 126, 2211-2240.

Guh J Y, Yang C U, Yang J M, Chen L M, and Lai Y H 1998 Prediction of Equilibrated Postdialysis BUN by an Artificial Neural Network in High-Efficiency Hemodialysis, American Journal of Kidney Diseases, 31,638-646.

Haykin S 1999: Neural Networks, A Comprehensive Foundation, Prentice Hall International, Inc. New Jersey USA.

Joachims, T.,1998 Making large-scale SVM learning practical, In B. Schölkopf, C. J. C. Burges, and A. J. Smola (Eds.) *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge, MA.

Krishnamurti T N, Kishtawal C M, Zhang Z, LaRow T, Bachiochi D, Williford E, Gadgil S and Surendran S 1999b Improved weather and seasonal climate forecasts from multimodel superensemble, Science 285, No 5433, 1548-1550.

Krishnamurti T N, Kishtawal C M, Zhang Z, LaRow T, Bachiochi D, Williford E, Gadgil S and Surendran S 1999a Multimodel super ensemble forecasts for weather and seasonal climate, FSU (Florida State University) Report 99-8 July 1999.

Krishna Kumar K, Soman M K, and Rupa Kumar K, 1995 Seasonal forecasting of Indian summer monsoon rainfall: A review, Weather, 50, 449-467.

Parthasarathy B, Munot A A, and Kothawale D R 1995 Monthly and seasonal rainfall series for All-India homogeneous regions and meteorological subdivisions: 1871-1994, IITM Research Report No RR-065.

Kulkarni J R, Mujumdar M, Mandake S K, Krishnan, R. Gharge, S P. and Satyan V. 2001 Monsoon-2001: Dynamical Prediction using AGCMs paper presented at Annual Monsoon Workshop of Indian Meteorol. Soc., Pune Chapter, 20 December, 2001.

Krajnc, A 1997 aiNet Manual, Trabarjeva 42, SI-3000 Celje, Slovenia.

Mukherjee, S., E. Osuna and F. Girosi, 1997 Nonlinear prediction of chaotic time series using Support Vector Machines, In *Proc. IEEE Workshop on Neural Networks for Signal Processing 7* (IEEE NNSP'97), 511-519.

Müller, K. R., A. Smola, G. Ratsch, B. Schölkopf, J. Kohlmorgen and V. Vapnik, 1997 Predicting time series with support vector machines, In W. Gerstner, A. Germond, M. Hasler and J-D. Nicoud (Eds.) *Artificial Neural Networks – ICANN'97*, Berlin, Springer Lecture Notes on Computer Science, Vol. 1327, 999-1004.

Navone H D, Ceccatto H A 1994 Predicting Indian monsoon rainfall: a neural network approach, Clim Dyn 10, 305-312.

Normand C W B 1953 Monsoon seasonal forecasting, Q. J. R. Meteorol. Soc., 79, 463-473.

Parthasarathy B, K Rupa Kumar and Munot A A 1996 Homogeneous regional summer monsoon rainfall over India: Interannual variability and teleconnections, IITM Research Report, RR-70.

Platt, J. C., 1998 Fast training of support vector machines using sequential minimal optimization, In B. Schölkopf, C. J. C. Burges, and A. J. Smola (Eds.) *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge, MA.

Rumelhart D E, Hinton G E, Williams R J 1986 Learning representations by back propagating errors, Nature , 323, 533-536.

Sahai A K, Soman M K, Satyan V 2000 All Idia summer monsoon rainfall prediction using an artificial neural network, Clim Dyn 16, 291-302.

Shukla J and Mooley D A 1987 Empirical prediction of the summer monsoon rainfall over India, Mon. Wea. Rev., 115, 695-703.

Smola, A., B. Schölkopf and K. R. Müller, 1998 The connection between regularization operators and support vector kernels, *Neural Networks*, 11, 637-649.

Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, 2001 Estimating support of a high-dimensional distribution, *Neural Comput.* 13, 1443-1471.

Thapliyal V 1990 Long range prediction of summer monsoon rainfall over India: Evolution and development of new models Mausam 41,339-346.

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

Vapnik, V., S. Golowich and A. Smola, 1996 support vector method for function approximation, regression estimation and signal processing, *Adv. in Neural Inform. Proces. Syst.*, 9 (1996), 281-287.

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

Venkatesan C, Raskar S D, Tambe S S, Kulkarni B D, and Keshamurty R N 1997 Prediction of All India summer monsoon rainfall using error-back-propagation neural networks, Meteorol. Atmos. Phys., 62, 225-240.

Walker G T 1908 Correlation in seasonal variation of climate (Introduction). Mem. India Meteorol. Dept. (IMD Mem.), 20, 117-124.

Walker G T 1918 Correlation in seasonal variation of weather, Q. J. R. Meteorol. Soc., 44 223-224.

Walker G T 1923 Correlation in seasonal variation of weather- VIII. A preliminary study of world weather. Mem. India Meteorol. Dept. (IMD Mem.), 24, 75-131.

Webster P J, Magana V O, Palmer T N, Shukla J, Thomas R A, Yanai M, and Yasunari T 1998: Monsoons: Prospects, Predictability, and Prospects for prediction, J. Geophys., Res., 13, 14451-14510.

Wilks D S 1995 Statistical Methods in Atmospheric Sciences, Academic Press, New York, pp 467.

Table 1 : Variances associated with different PCs for 8 regions.

PC	All India	Homo-India	Core India	Northwest India	West Central India	Central northeast India	Northeast India	Peninsular India
1	39.4244	33.1688	32.4107	26.0293	35.2647	34.8815	43.8838	36.4829
2	39.4086	33.158	32.3917	25.9638	35.243	34.8669	43.8639	36.4349
3	7.70938	11.2841	11.1788	13.0069	10.9774	10.3192	2.41498	4.35884
4	7.68105	11.2401	11.1311	12.9422	10.9512	10.2919	2.41238	4.28064
5	1.20224	2.17143	2.33024	4.12222	1.50356	1.74815	0.995946	2.77131
6	1.15339	2.12291	2.29273	4.08843	1.47122	1.6865	0.987686	2.74544
7	0.776372	1.28549	1.46291	2.53308	0.839747	1.11773	0.977929	2.24754
8	0.569901	1.16203	1.43381	2.35886	0.791834	1.07359	0.91689	2.21936
Total	97.92533	95.59286	94.63199	91.04479	97.04266	95.98547	96.45351	91.54093

Table 2: The details of EBP models developed for the 8 regions for monthly rainfall predictions.

Region name	All-India	Homoge-nous India	Core-India	North-west India	West Central-India	Central-north east India	North-east India	Penin sular India
Number of training patterns	1428	1428	1428	1428	1428	1428	1428	1428
Number of test patterns	132	132	132	132	132	132	132	132
Learning rate	0.05	0.05	0..05	0..05	0..05	0..05	0..05	0..05
Momentum coefficient	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Number of input nodes	8	8	8	8	8	8	8	8
Number of hidden nodes in layer one	10	6	6	10	7	10	4	6
Number of hidden node in hidden layer two	12	4	4	12	5	5	3	4
Number of the output nodes	1	1	1	1	1	1	1	1

Table 3: Number of total support vectors used in 8 SVR models.

No	Regions	Number of support vectors
1	All-India	1423
2	Homogeneous	1427
3	Core-India	1427
4	North-west	1427
5	West central-India	1427
6	Central northeast India	1427
7	North-east India	1427
8	Peninsular India	1427

Table 4 : SVR parameters used for the model building.

Model	Cost-value	Gamma-value	Epsilon-value	Termination criteria
All-India	1	0.5	0.00001	0.00001
Homogeneous India	8	0.3	0.00001	0.00001
Core-India	61	0.8	0.00001	0.00001
North-west India	1	1.4	0.00001	0.00001
West central-India	10	0.2	0.00001	0.00001
Central northeast India	81	0.1	0.00001	0.00001
North-east India	71	0.3	0.00001	0.00001
Peninsular India	1	1	0.00001	0.00001

Table 5: Penalty coefficients values used in 8 aiNet models.

Model	Penalty coefficient
All-India	0.8390
Homogeneous	0.6951
Core-India	0.6499
North-west	1.4652
West Central-India	0.7040
Central northeast India	1.4717
North-east India	1.0704
Peninsular India	0.6795

Table 6 : Distribution of CVs by the three models for the actual and model simulated data for the 8 regions.

Sr.	Regions	CV (%) Actual		CV Model		
		Train	Test	EBP	Train	Test
1	All-India	105.4	100.2	aiNet	97.7	98.2
				SVR	97.4	93.8
				EBP	104.6	96.1
2	Homogeneous India	131.3	125.1	aiNet	123	120.9
				SVR	120.2	121.4
				EBP	127.1	127.1
3	Core-India	137.7	130.7	aiNet	129	129.3
				SVR	123.9	125.3
				EBP	132.7	132.9
4	North-west India	158.8	145.2	aiNet	141.5	116.7
				SVR	167.4	65.3
				EBP	107.1	60.3
5	West central-India	120.2	115.8	aiNet	112.1	112
				SVR	110.7	110.7
				EBP	118.3	117.7
6	Central northeast India	126.0	122.2	aiNet	116.5	117.2
				SVR	114.4	114.8
				EBP	122.2	123
7	North-east India	88.8	85.5	aiNet	77.6	76.5
				SVR	79.5	79.6
				EBP	86.6	86.6
8	Peninsular India	81.5	79.6	aiNet	70.4	69.4
				SVR	61.9	64.8
				EBP	73.5	75.7

Table 7 : Comparison of EBP, aiNet and SVR models in simulating monthly rainfall over 8 regions.

S. No	Regions	Model	CC		Mean				Standard Deviation			
			Train Set	Test Set	Train		Test		Train		Test	
					Actual mm	Pred mm	Actual mm	Pred mm	Actual mm	Pred mm	Actual mm	Pred mm
1	All- India	EBP	0.964	0.967	90.88	93.24	91.08	92.93	95.77	91.09	91.26	91.28
		aiNet	0.974	0.9442	90.88	94.34	91.08	97.68	95.77	91.88	91.26	91.61
		SVR	0.867	0.836	90.88	92.37	91.08	96.55	95.77	96.59	91.26	92.74
2	Homoge-neous-India	EBP	0.925	0.934	72.43	72.78	70.45	71.81	95.12	89.54	88.15	86.85
		aiNet	0.959	0.945	72.43	72.18	70.45	70.36	95.12	86.77	88.15	85.40
		SVR	0.925	0.945	72.43	69.01	70.45	68.69	95.12	87.73	88.15	87.33
3	Core-India	EBP	0.925	0.937	80.10	81.15	77.30	81.22	110.29	104.68	101.02	105.04
		aiNet	0.953	0.944	80.10	80.11	77.30	78.23	110.29	99.25	101.02	98.01
		SVR	0.928	0.929	80.10	76.04	77.30	73.02	110.29	100.90	101.02	97.04
4	North-west India	EBP	0.856	0.626	45.48	43.66	46.71	52.24	72.20	61.78	67.80	60.94
		aiNet	0.880	0.474	45.48	38.23	46.71	46.82	72.20	63.99	67.80	30.58
		SVR	0.833	0.593	45.48	40.10	46.71	56.99	72.20	42.94	67.80	34.38
5	West central-India	EBP	0.950	0.951	100.13	101.49	100.76	102.40	120.34	113.73	116.64	114.70
		aiNet	0.963	0.947	100.13	100.70	100.76	100.23	120.34	111.44	116.64	110.94
		SVR	0.941	0.941	100.13	96.88	100.76	96.92	120.34	114.58	116.64	114.03
6	Central Northeast India	EBP	0.959	0.944	90.20	95.66	86.11	93.60	113.66	111.44	105.21	109.73
		aiNet	0.969	0.938	90.20	90.40	86.11	89.41	113.66	103.40	105.21	102.61
		SVR	0.935	0.938	90.20	86.88	86.11	85.02	113.66	106.15	105.21	104.55
7	North-East India	EBP	0.934	0.913	172.19	182.35	177.06	181.89	152.85	141.52	151.41	139.21
		aiNet	0.961	0.894	172.19	172.17	177.06	172.17	152.85	136.90	151.41	137.09
		SVR	0.946	0.919	172.19	167.91	177.06	166.37	152.85	145.33	151.41	144.08
8	Peninsular India	EBP	0.872	0.882	96.38	98.84	102.69	99.31	78.53	69.63	81.74	68.94
		aiNet	0.948	0.856	96.38	96.60	102.69	96.26	78.53	59.81	81.74	62.39
		SVR	0.847	0.862	96.38	91.00	102.69	92.59	78.53	66.84	81.74	70.13

Table 8 : Regression coefficients a, b, c for EBP, SVR, AiNet models.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
All-India	a	-0.234	-0.068	0.165	0.180	0.140	0.230	-0.246	0.082	-0.367	-0.149	-0.074	0.415
	b	0.672	0.891	0.421	0.474	0.190	0.158	-0.183	0.339	0.112	0.815	0.339	0.269
	c	0.468	0.240	0.347	0.315	0.562	0.610	1.438	0.574	1.234	0.362	0.673	0.244
Homogeneous India	a	0.698	-0.534	-0.592	0.261	0.064	-0.439	-0.285	0.450	-0.372	-0.439	0.315	0.173
	b	0.384	0.975	1.177	0.180	0.000	0.399	0.204	-0.179	-0.032	0.399	0.137	0.075
	c	-0.007	0.888	0.568	0.471	0.968	1.066	1.078	0.737	1.342	1.066	0.584	0.471
Core-India	a	-0.270	-0.665	-0.696	0.010	-0.026	-0.272	-0.351	-0.603	-0.719	-0.599	-0.600	-0.485
	b	0.266	0.961	0.738	0.313	0.061	0.888	0.766	0.706	0.431	0.968	0.613	1.302
	c	1.336	1.201	1.169	0.534	0.759	0.397	0.576	0.889	1.335	0.845	1.868	-0.096
North-west India	a	-0.217	0.432	0.276	0.131	0.279	0.308	-0.542	0.166	-0.202	0.349	-0.011	0.070
	b	0.568	0.385	0.361	0.078	0.047	0.591	0.651	0.028	0.014	-0.044	-0.056	0.177
	c	0.787	0.027	-0.407	-0.053	0.102	-0.036	0.933	0.796	1.202	0.277	0.975	0.507
West central-India	a	0.383	0.198	0.705	-0.450	0.254	0.226	-0.122	-0.003	-0.160	0.030	0.215	0.250
	b	-0.085	0.648	-0.031	0.052	0.133	-0.124	0.140	0.071	-0.044	-0.100	0.007	0.044
	c	0.714	0.307	0.216	1.184	0.548	0.903	0.983	0.931	1.181	1.036	0.524	0.323
Central northeast India	a	0.077	-0.064	0.249	0.134	0.566	-0.175	-0.054	-0.063	-0.510	-0.064	0.391	0.078
	b	0.412	0.101	0.794	0.042	-0.299	0.135	-0.022	-0.085	0.385	0.101	-0.057	0.462
	c	0.409	0.945	0.057	0.590	0.486	1.013	1.061	1.097	1.045	0.945	0.452	0.303
North-east India	a	0.233	-0.061	0.040	0.396	-0.015	-0.264	0.278	-0.240	0.006	-0.018	0.324	0.145
	b	0.476	0.924	0.990	0.931	0.964	0.529	0.841	0.848	1.047	1.076	-0.236	0.950
	c	0.164	0.233	-0.029	-0.281	0.040	0.756	-0.093	0.360	-0.012	-0.086	0.316	0.018
Peninsular India	a	0.171	0.421	0.224	0.305	0.015	0.517	-0.213	0.165	0.267	-0.545	-0.603	0.071
	b	0.248	0.211	0.056	0.305	0.738	0.871	1.139	0.430	0.576	1.056	0.976	0.461
	c	0.557	-0.162	0.522	0.308	0.197	-0.396	0.097	0.325	0.130	0.547	0.622	0.378

Table 9: The distribution of SS values for different months for the test period for all the 8 regions.

		Winter			Premonsoon			Monsoon				Post monsoon			
No	Region	D	J	F	M	A	M	J	J	A	S	O	N	SS	%
1	A I	-	0.10 a	-	-	0.09 0.12 a	- 0.01a	-	0.05	0.05 0.15 a	- 0.01 e	-	.02 .06 e	4 6	33 50
2	HI	0.01 0.05 a	0.15	- 0.03 a	- 0.08s	0.10 0.17 s	-	0.07	0.04 0.10 s	0.15 0.16 e	0.02 0.10 s	0.05	0.12	9 11	75 92
3	CI	0.04	0.15	0.03	-	0.02 0.06 a	- 0.10 a	-	-	0.12	0.03 0.08 e		- 0.10 s	6 8	50 66
4	NW I	-	0.04	-	-	-	-	- 0.06 a	-	-	-	-	0.13	2	16 25
5	WCI	0.01	-	-	-	- 0.01	-	- 0.02 s	- 0.14 e	-	0.13	0.01 0.03 e	-	3 6	25 50
6	CNEI	0.01 0.06 s	-	0.01	0.27 0.31a	0.01 0.07 e	- 0.10 a	0.07	0.03 0.06 s	0.30 0.31 e	-	-	0.05 0.16 s	8 9	66 75
7	NEI	-	-	- 0.06 e	-	-	-	0.01 0.11 s	- 0.07 e	-	-	-	-	1 3	9 25
8	PI	0.07	-	-	-	0.09 0.31 a	-	0.06 0.07 e	- 0.14 s	-	0.01 0.11 e	-	0.19 0.27 e	5 6	40 50
Total		5 5	3 4	2 4	1 2	4 6	0 3	3 5	3 6	3 4	4 5	1 2	4 6	33 52	34 54
% score		62 62	37 50	25 50	12 25	50 75	0 37	37 62	37 75	37 50	50 62	12 25	50 75		

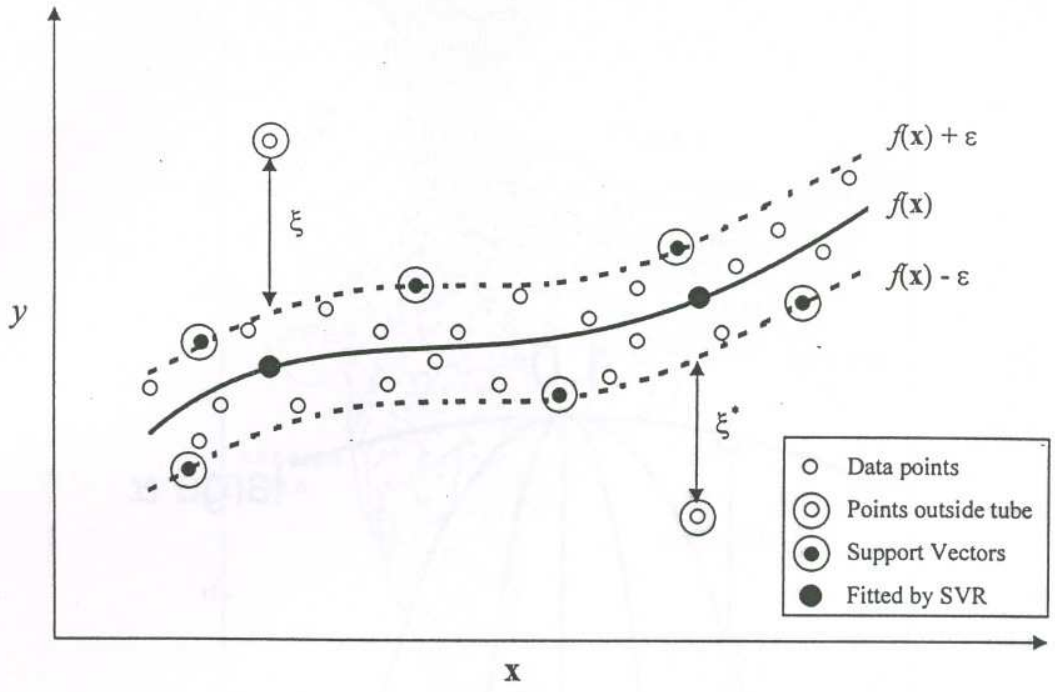


Figure 1: A schematic of support vector regression using ε-insensitive loss function.

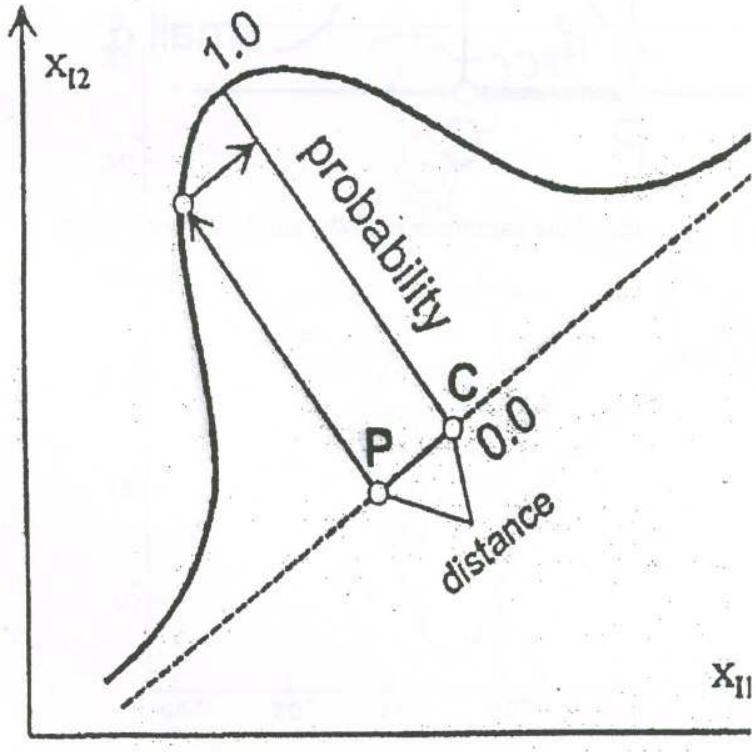


Figure 2 : An illustration of the probability assumption

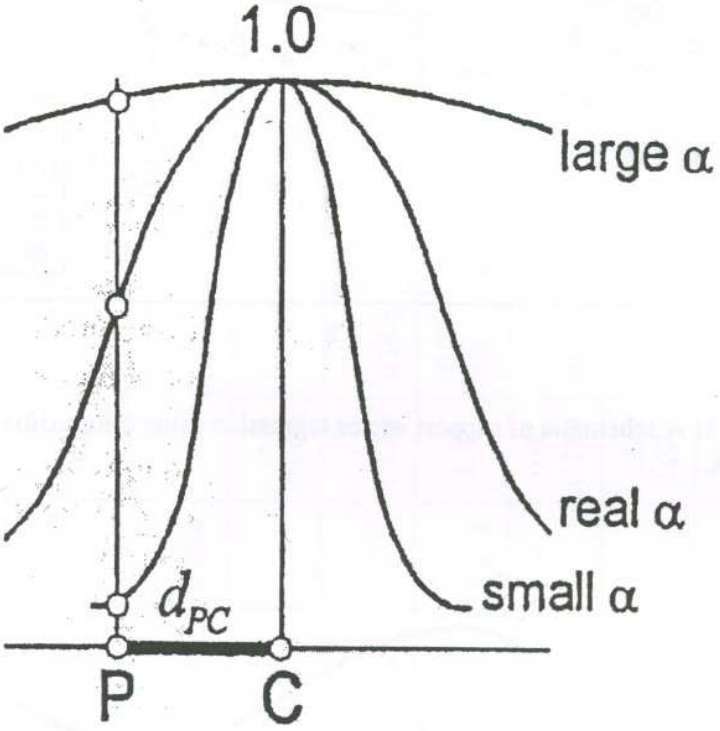


Figure 3 Influence of the parameter α on the similarity coefficient.

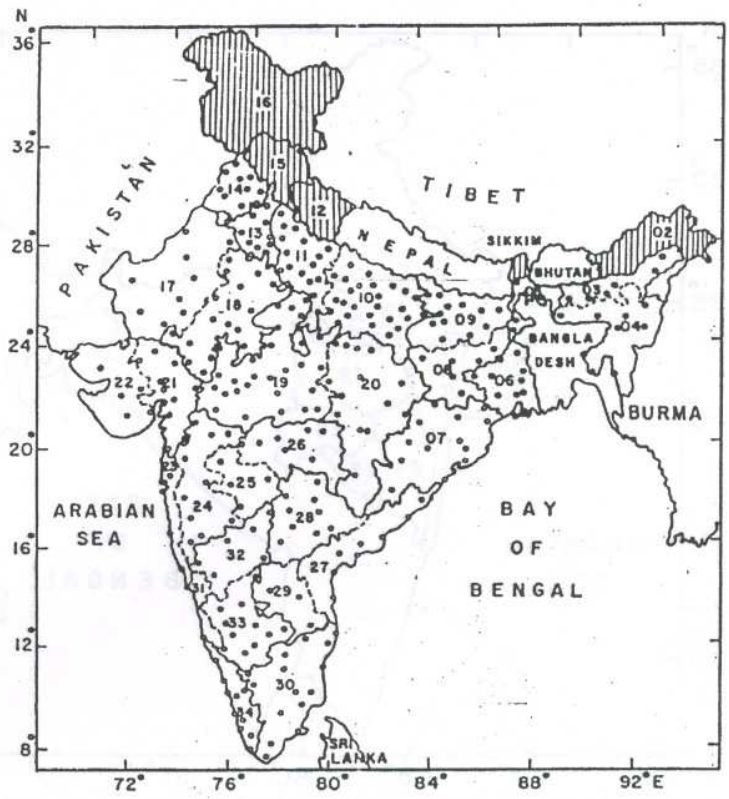


Figure 4 (a) : Meteorological subdivisions considered for All India (AI) Rainfall time series. The dots represent the locations of rain gauge stations. Hatched areas are not considered for rainfall time series

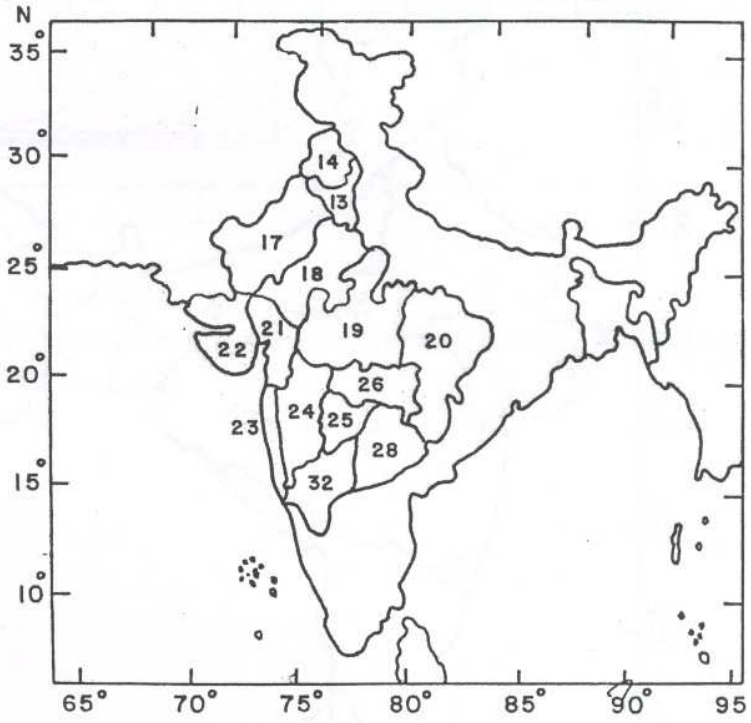


Figure 4 (b) : Meteorological subdivisions considered for Homogeneous India (HI) Rainfall time series.

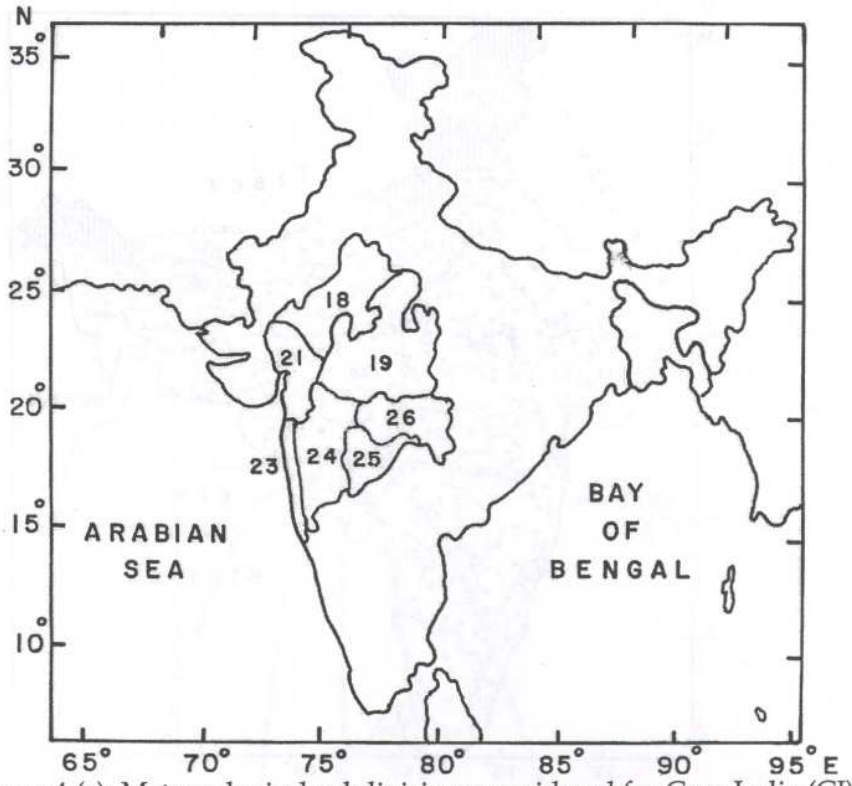


Figure 4 (c): Meteorological subdivisions considered for Core India (CI) Rainfall time series.

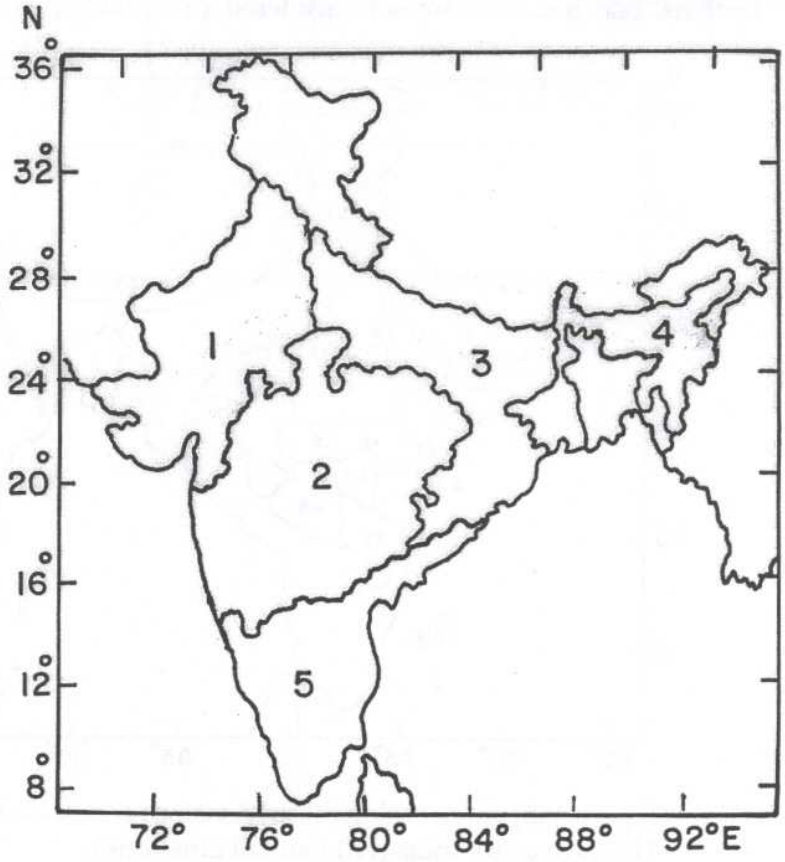


Figure 4 (d) : Five macro climatic zones of India. (1) Northwest (NWI) (2) West Central (WCI) (3) Central north east (CNEI) (4) North east (NEI) and (5) Peninsular (PI) India

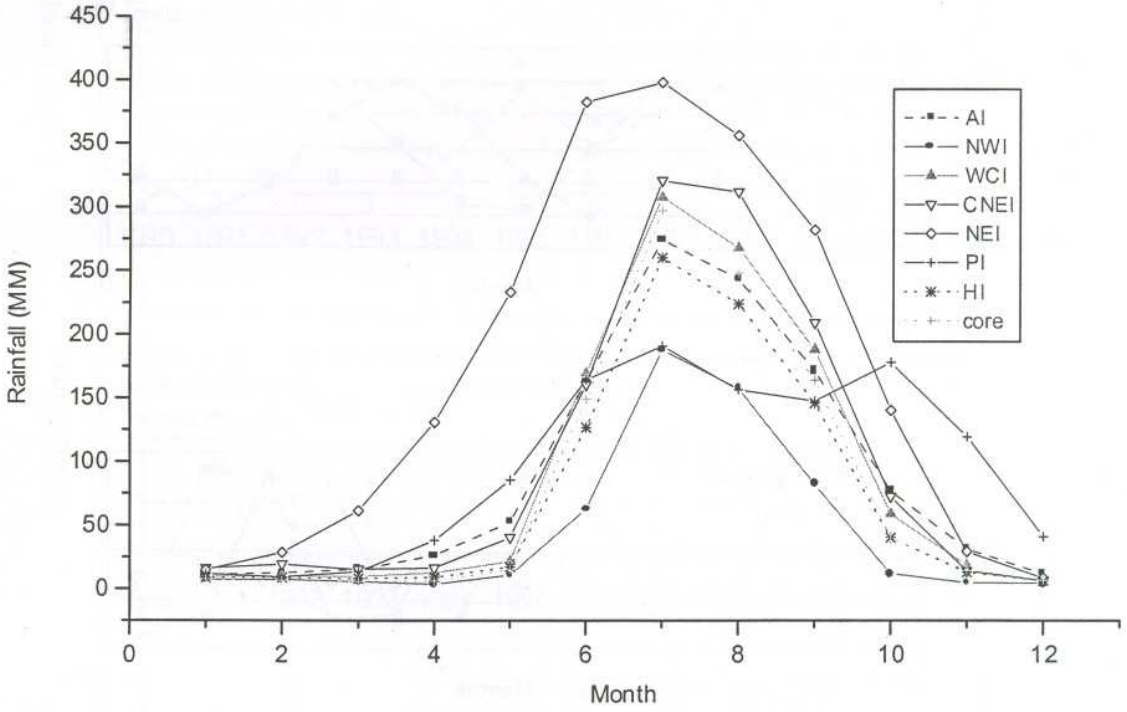
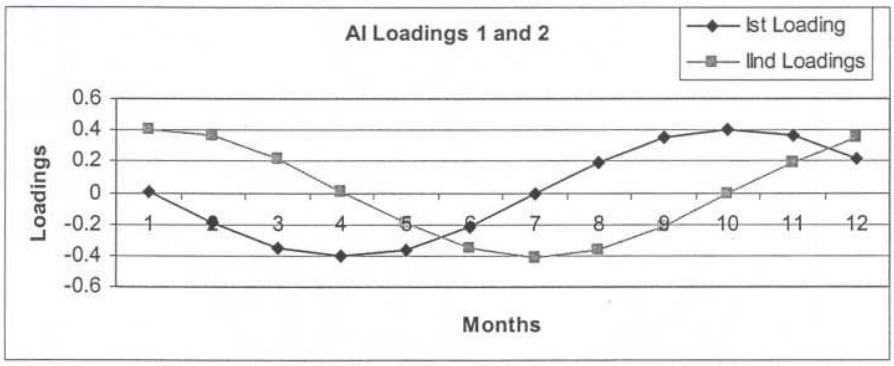
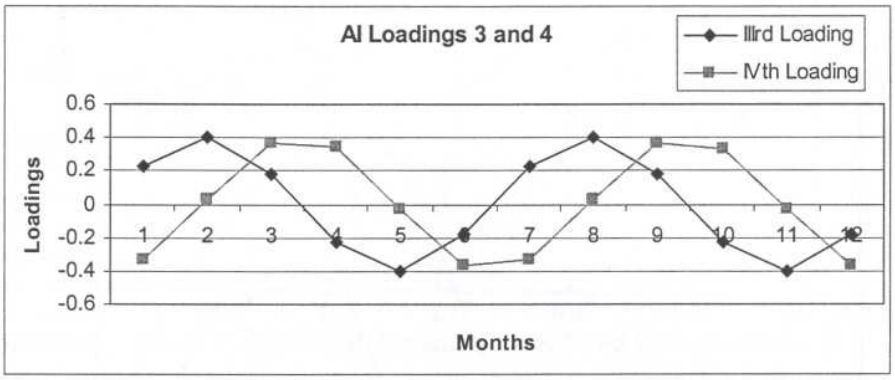
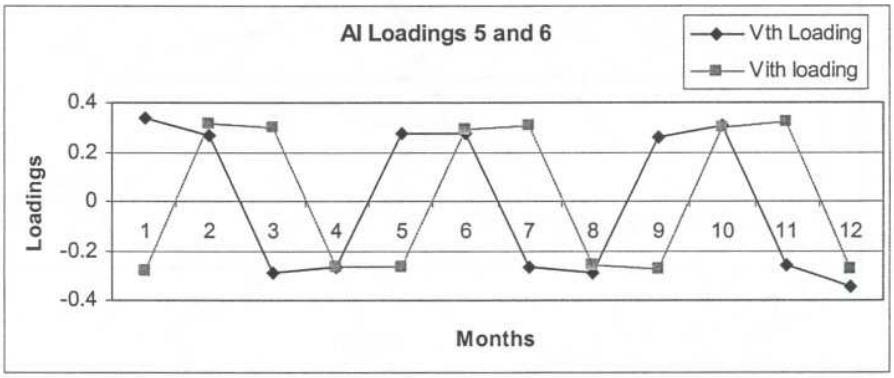


Figure 5 : Variation of mean monthly rainfall over the 8 regions.







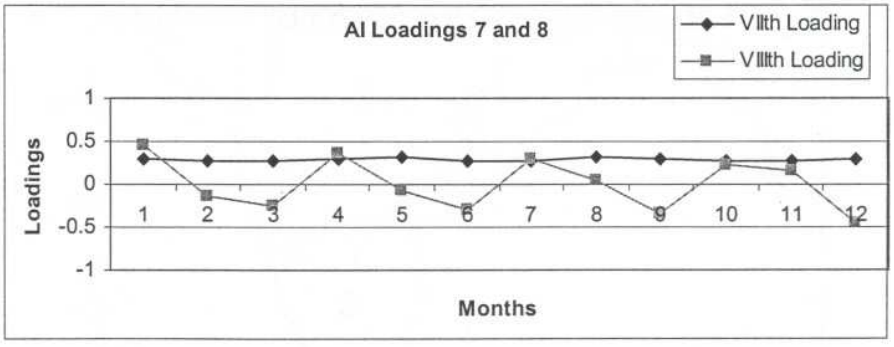


Figure 6 : First 8 eigen vectors for AI rainfall time series.

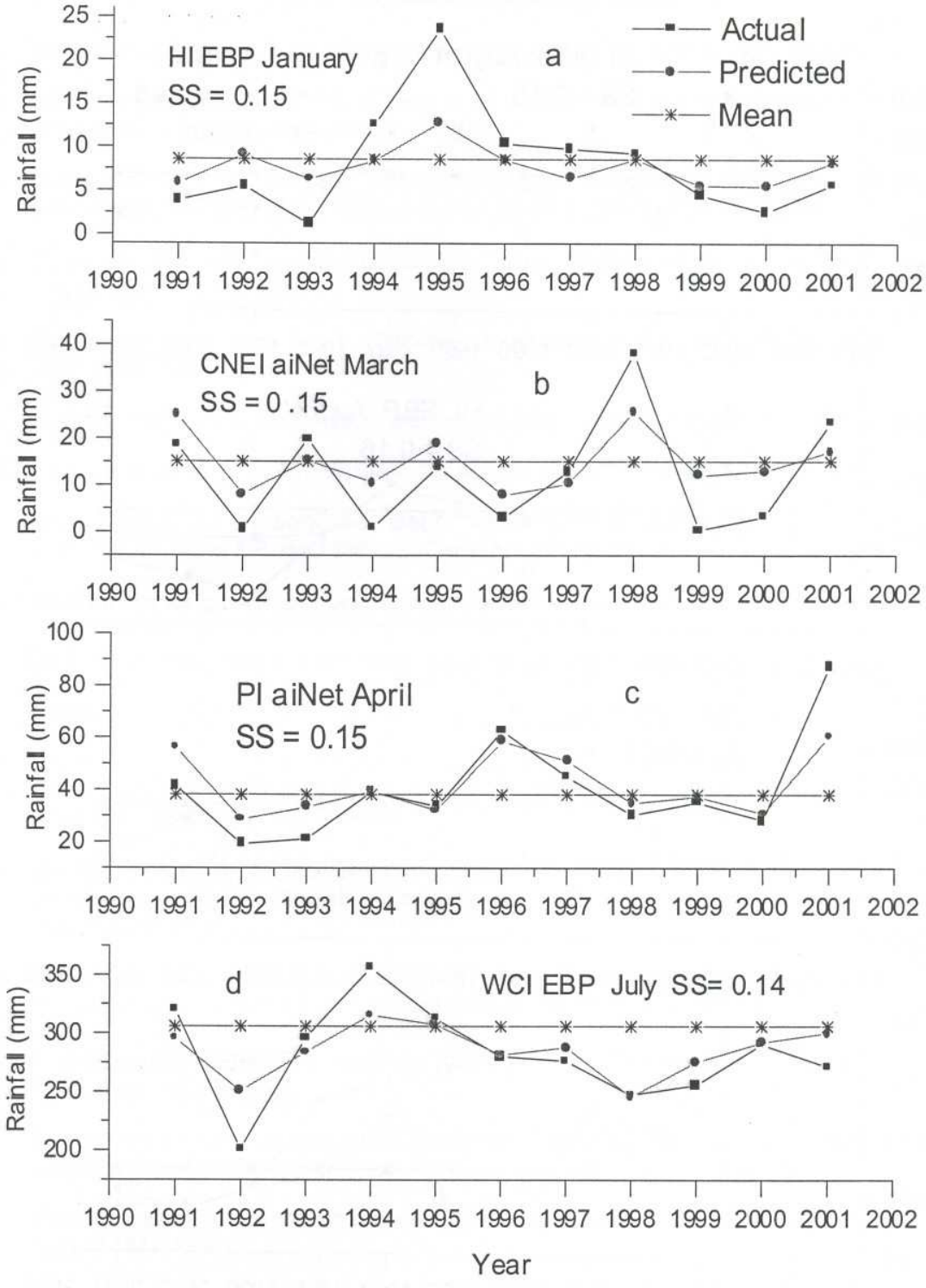


Figure 7a : Predicted and observed monthly rainfalls for some selected regions and months: (a) Homogeneous India for January by EBP (b) Central north east India for March by aiNet (c) Peninsular India for April by aiNet and (d) West central India for July by EBP. The skill scores are given for each model.

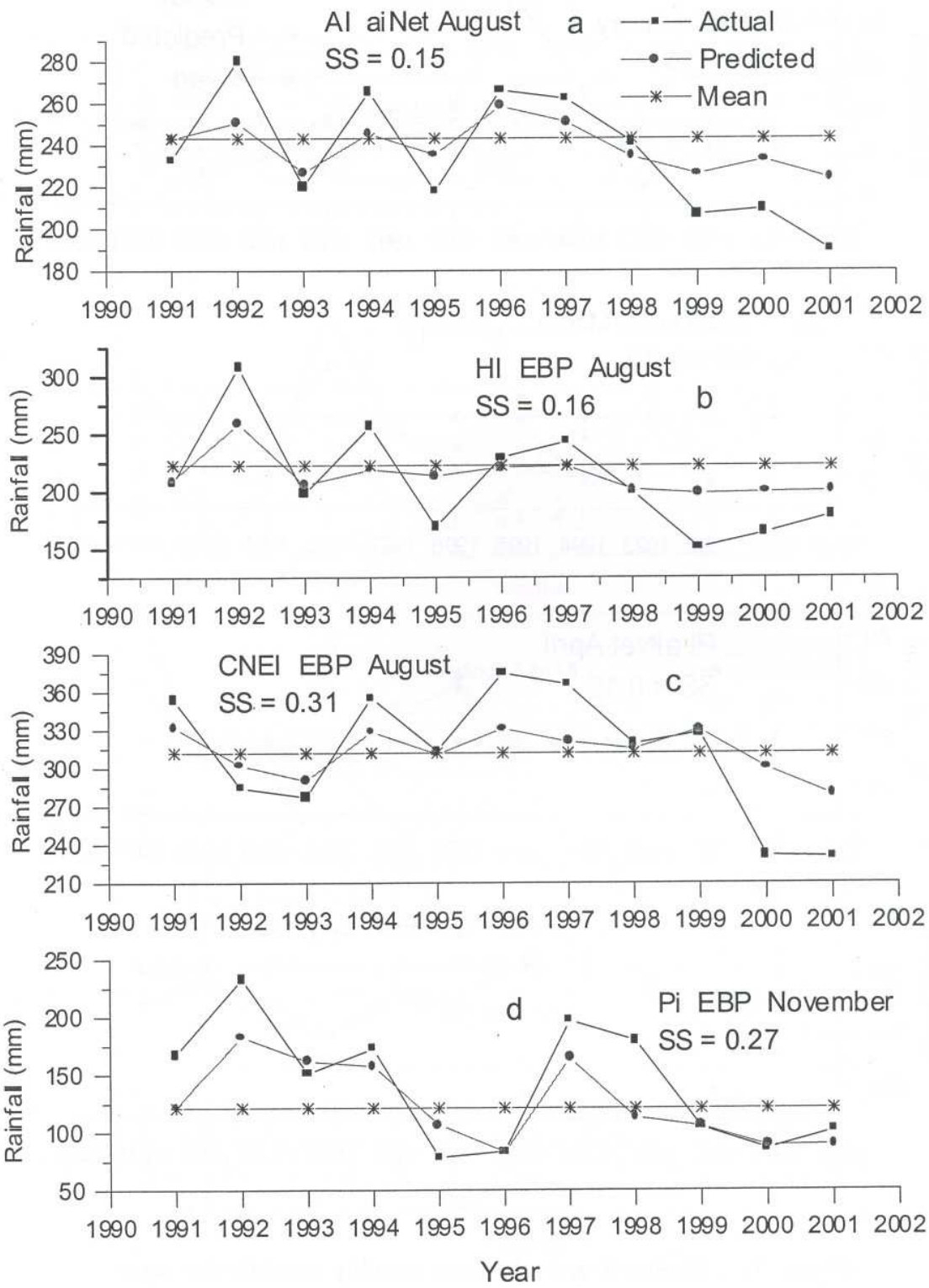


Figure 7b: Predicted and observed monthly rainfalls for some selected regions and months: (a) All India for August by aiNet, (b) Homogeneous India for August by EBP, (c) Central north east India for August by EBP and (d) Peninsular India for November by EBP. The skill scores are given for each model.

I. I. T. M. RESEARCH REPORTS

- Energetic consistency of truncated models, *Asnani G.C.*, August 1971, RR-001.
- Note on the turbulent fluxes of heat and moisture in the boundary layer over the Arabian Sea, *Sinha S.*, August 1971, RR-002.
- Simulation of the spectral characteristics of the lower atmosphere by a simple electrical model and using it for prediction, *Sinha S.*, September 1971, RR-003.
- Study of potential evapo-transpiration over Andhra Pradesh, *Rakhecha P.R.*, September 1971, RR-004.
- Climatic cycles in India-1: Rainfall, *Jagannathan P. and Parthasarathy B.*, November 1971, RR-005.
- Tibetan anticyclone and tropical easterly jet, *Raghavan K.*, September 1972, RR-006.
- Theoretical study of mountain waves in Assam, *De U.S.*, February 1973, RR-007.
- Local fallout of radioactive debris from nuclear explosion in a monsoon atmosphere, *Saha K.R. and Sinha S.*, December 1972, RR-008.
- Mechanism for growth of tropical disturbances, *Asnani G.C. and Keshavamurty R.N.*, April 1973, RR-009.
- Note on "Applicability of quasi-geostrophic barotropic model in the tropics", *Asnani G.C.*, February 1973, RR-010.
- On the behaviour of the 24-hour pressure tendency oscillations on the surface of the earth, Part-I: Frequency analysis, Part-II: Spectrum analysis for tropical stations, *Misra B.M.*, December 1973, RR-011.
- On the behaviour of the 24 hour pressure tendency oscillations on the surface of the earth, Part-III : Spectrum analysis for the extra-tropical stations, *Misra B.M.*, July 1976, RR-011A.
- Dynamical parameters derived from analytical functions representing Indian monsoon flow, *Awade S.T. and Asnani G.C.*, November 1973, RR-012.
- Meridional circulation in summer monsoon of Southeast Asia, *Asnani G.C.*, November 1973, RR-014.
- Energy conversions during weak monsoon, *Keshavamurty R.N. and Awade S.T.*, August 1974, RR-015.
- Vertical motion in the Indian summer monsoon, *Awade S.T. and Keshavamurty R.N.*, August 1974, RR-016.
- Semi-annual pressure oscillation from sea level to 100mb in the northern hemisphere, *Asnani G.C. and Verma R.K.*, August 1974, RR-017.
- Suitable tables for application of gamma probability model to rainfall, *Mooley D.A.*, November 1974, RR-018.
- Annual and semi-annual thickness oscillation in thenorthern hemisphere, *Asnani G.C. and Verma R.K.*, January 1975, RR-020.

- Spherical harmonic analysis of the normal constant pressure charts in the northern hemisphere, *Awade S.T., Asnani G.C. and Keshavamurty R.N.*, May 1978, RR-021.

- Dynamical parameters derived from analytical function representing normal July zonal flow along 87.5 °E, *Awade S.T. and Asnani G.C.*, May 1978, RR-022.

- Study of trends and periodicities in the seasonal and annual rainfall of India, *Parthasarathy B. and Dhar O.N.*, July 1975, RR-023.

- Southern hemisphere influence on Indian rainfall, *Raghavan K.., Paul D.K. and Upasani P.U.*, February 1976, RR-024.

- Climatic fluctuations over Indian region - Rainfall : A review, *Parthasarathy B. and Dhar O.N.*, May 1978, RR-025.

- Annual variation of meridional flux of sensible heat, *Verma R.K. and Asnani G.C.*, December 1978, RR-026.

- Poisson distribution and years of bad monsoon over India, *Mooley D.A and Parthasarathy B.*, April 1980, RR-027.

- On accelerating the FFT of Cooley and Tukey, *Mishra S.K.*, February 1981, RR-028.

- Wind tunnel for simulation studies of the atmospheric boundary layer, *Sivaramakrishnan S.*, February 1981, RR-029.

- Hundred years of Karnataka rainfall, *Parthasarathy B. and Mooley D.A.*, March 1981, RR-030.

- Study of the anomalous thermal and wind patterns during early summer season of 1979 over the Afro-Asian region in relation to the large-scale performance of the monsoon over India, *Verma R.K. and Sikka D.R.*, March 1981, RR-031.

- Some aspects of oceanic ITCZ and its disturbances during the onset and established phase of summer monsoon studied with Monex-79 data, *Sikka D.R., Paul D.K. and Singh S.V.*, March 1981, RR- 032.

- Modification of Palmer drought index, *Bhalme H.N. and Mooley D.A.*, March 1981, RR-033.

- Meteorological rocket payload for Menaka-II/Rohini 200 and its developmental details, *Vernekar K.G. and Brij Mohan*, April 1981, RR-034.

- Harmonic analysis of normal pentad rainfall of Indian stations, *Anathakrishnan R. and Pathan J.M.*, October 1981, RR-035.

- Pentad rainfall charts and space-time variations of rainfall over India and the adjoining areas, *Anathakrishnan R. and Pathan J.M.*, November 1981, RR-036.

- Dynamic effects of orography on the large scale motion of the atmosphere Part I : Zonal flow and elliptic barrier with maximum height of one km., *Bavadekar S.N. and Khaladkar R.M.*, January 1983, RR-037.

- Limited area five level primitive equation model, *Singh S.S.*, February 1983, RR-038.

- Developmental details of vortex and other aircraft thermometers, *Vernekar K.G., Brij Mohan and Srivastava S.*, November 1983, RR-039.

- Note on the preliminary results of integration of a five level P.E. model with westerly wind and low orography, *Bavadekar S.N., Khaladkar R.M., Bandyopadhyay A. and Seetaramayya P.*, November 1983, RR-040.

- Long-term variability of summer monsoon and climatic change, *Verma R.K., Subramaniam K. and Dugam S.S.*, December 1984, RR-041.

- Project report on multidimensional initialization for NWP models, *Sinha S.*, February 1989, RR-042.

- Numerical experiments with inclusion of orography in five level P.E. Model in pressure-coordinates for interhemispheric region, *Bavadekar S.N. and Khaladkar R.M.*, March 1989, RR-043.

- Application of a quasi-lagrangian regional model for monsoon prediction, *Singh S.S. and Bandyopadhyay A.*, July 1990, RR-044.

- High resolution UV-visible spectrometer for atmospheric studies, *Bose S., Trimbake H.N., Londhe A.L. and Jadhav D.B.*, January 1991, RR-045.

- Fortran-77algorithm for cubic spline interpolation for regular and irregular grids, *Tandon M.K.*, November 1991, RR-046.

- Fortran algorithm for 2-dimensional harmonic analysis, *Tandon M.K.*, November 1991, RR- 047.

- 500 hPa ridge and Indian summer monsoon rainfall : A detailed diagnostic study, *Krishna Kumar K., Rupa Kumar K. and Pant G.B.*, November 1991, RR-048.

- Documentation of the regional six level primitive equation model, *Singh S.S. and Vaidya S.S.*, February 1992, RR-049.

- Utilisation of magnetic tapes on ND-560 computer system, *Kripalani R.H. and Athale S.U.*, July 1992, RR-050.

- Spatial patterns of Indian summer monsoon rainfall for the period 1871-1990, *Kripalani R.H., Kulkarni A.A., Panchawagh N.V. and Singh S.V.*, August 1992, RR-051.

- FORTRAN algorithm for divergent and rotational wind fields, *Tandon M.K.*, November 1992, RR-052.

- Construction and analysis of all-India summer monsoon rainfall series for the longest instrumental period: 1813-1991, *Sontakke N.A., Pant G.B. and Singh N.*, October 1992, RR-053.

- Some aspects of solar radiation, *Tandon M.K.*, February 1993, RR-054.

- Design of a stepper motor driver circuit for use in the moving platform, *Dharmaraj T. and Vernekar K.G.*, July 1993, RR-055.

- Experimental set-up to estimate the heat budget near the land surface interface, *Vernekar K.G., Saxena S., Pillai J.S., Murthy B.S., Dharmaraj T. and Brij Mohan*, July 1993, RR-056.

- Identification of self-organized criticality in atmospheric total ozone variability, *Selvam A.M. and Radhamani M.*, July 1993, RR-057.

- Deterministic chaos and numerical weather prediction, *Selvam A.M.*, February 1994, RR-058.

- Evaluation of a limited area model forecasts, *Singh S.S., Vaidya S.S Bandyopadhyay A., Kulkarni A.A, Bawiskar S.M.,Sanjay J., Trivedi D.K. and Iyer U.*, October 1994, RR-059.

- Signatures of a universal spectrum for atmospheric interannual variability in COADS temperature time series, *Selvam A.M., Joshi R.R. and Vijayakumar R.*, October 1994, RR-060.

- Identification of self-organized criticality in the interannual veriability of global surface temperature, *Selvam A.M. and Radhamani M.*, October 1994, RR-061.

- Identification of a universal spectrum for nonlinear veriability of solar-geophysical parameters, *Selvam A.M., Kulkarni M.K., Pethkar J.S. and Vijayakumar R.*, October 1994, RR-062.

- Universal spectrum for fluxes of energetic charged particles from the earth's magnetosphere, *Selvam A.M. and Radhamani M.*, June 1995, RR-063.

- Estimation of nonlinear kinetic energy exchanges into individual triad interactions in the frequency domain by use of the cross-spectral technique, *Chakraborty D.R.*, August 1995, RR-064.

- Monthly and seasonal rainfall series for all-India homogeneous regions and meteorological subdivisions: 1871-1994, *Parthasarathy B., Munot A.A. and Kothawale D.R.*, August 1995, RR-065.

- Thermodynamics of the mixing processes in the atmospheric boundary layer over Pune during summer monsoon season, *Morwal S.B. and Parasnis S.S.*, March 1996, RR-066.

- Instrumental period rainfall series of the Indian region: A documentation, *Singh N. and Sontakke N.A.*, March 1996, RR-067.

- Some numerical experiments on roundoff-error growth in finite precision numerical computation, *Fadnavis S.*, May 1996, RR-068.

- Fractal nature of MONTBLEX time series data, *Selvam A.M. and Sapre V.V.*, May 1996, RR-069.

- Homogeneous regional summer monsoon rainfall over India: Interannual variability and teleconnections, *Parthasarathy B., Rupa Kumar K. and Munot A.A.*, May 1996, RR-070.

- Universal spectrum for sunspot number variability, *Selvam A.M. and Radhamani M.*, November 1996, RR-071.

- Development of simple reduced gravity ocean model for the study of upper north Indian ocean, *Behera S.K. and Salvekar P.S.*, November 1996, RR-072.

- Study of circadian rhythm and meteorological factors influencing acute myocardial infraction, *Selvam A.M., Sen D. and Mody S.M.S.*, April 1997, RR-073.

- Signatures of universal spectrum for atmospheric gravity waves in southern oscillation index time series, *Selvam A.M., Kulkarni M.K., Pethkar J.S. and Vijayakumar R.*, December 1997, RR-074.

- Some example of X-Y plots on Silicon Graphics, *Selvam A.M., Fadnavis S. and Gharge S.P.*, May 1998, RR-075.

- Simulation of monsoon transient disturbances in a GCM, *Ashok K., Soman M.K. and Satyan V.*,August 1998, RR-076.

- Universal spectrum for intraseasonal variability in TOGA temperature time series, *Selvam A.M., Radhamani M., Fadnavis S. and Tinmaker M.I.R.*, August 1998, RR-077.

- One dimensional model of atmospheric boundary layer, *Parasnis S.S., Kulkarni M.K., Arulraj S. and Vernekar K.G.*, February 1999, RR-078.

- Diagnostic model of the surface boundary layer - A new approach, *Sinha S.*, February 1999, RR-079.

- Computation of thermal properties of surface soil from energy balance equation using force - restore method, *Sinha S.*, February 1999, RR-080.

- Fractal nature of TOGA temperature time series, *Selvam A.M. and Sapre V.V.*, February 1999, RR-081.

- Evolution of convective boundary layer over the Deccan Plateau during summer monsoon, *Parasnis S.S.*, February 1999, RR-082.

- Self-organized criticality in daily incidence of acute myocardial infarction, *Selvam A.M., Sen D., and Mody S.M.S.*, Feburary 1999, RR-083.

- Monsoon simulation of 1991 and 1994 by GCM : Sensitivity to SST distribution, *Ashrit R.G., Mandke S.K. and Soman M.K.*, March 1999, RR-084.

- Numerical investigation on wind induced interannual variability of the north Indian Ocean SST, *Behera S.K., Salvekar P.S. and Ganer D.W.*, April 1999, RR-085.

- On step mountain eta model, *Mukhopadhyay P., Vaidya S.S., Sanjay J. and Singh S.S.*, October 1999, RR-086.

- Land surface processes experiment in the Sabarmati river basin: an overview and early results, *Vernekar K.G., Sinha S., Sadani L.K., Sivaramakrishnan S., Parasnis S.S., Brij Mohan, Saxena S., Dharamraj T., Pillai, J.S., Murthy B.S., Debaje, S.B., Patil, M.N. and Singh* A.B., November 1999, RR-087.

- Reduction of AGCM systematic error by Artificial Neural Network: A new approach for dynamical seasonal prediction of Indian summer monsoon rainfall, *Sahai A.K. and Satyan V.*, December 2000, RR-088.

- Ensemble GCM simulations of the contrasting Indian summer monsoons of the 1987 and 1988, *Mujumdar M.* and *Krishnan R.*, February 2001, RR-089.

- Aerosol measurements using lidar and radiometers at Pune during INDOEX field phases, *Maheskumar R.S., Devara P.C.S., Raj P.E., Jaya Rao Y.,Pandithurai G., Dani K.K., Saha S.K., Sonbawne S.M. and Tiwari Y.K.*, December 2001, RR-090.

- Modelling studies of the 2000 Indian summer monsoon and extended analysis, *Krishnan R., Mujumdar M., Vaidya V., Ramesh K.V. and Satyan V.*, December 2001, RR-091.

- Intercomparison of Asian summer monsoon 1997 simulated by atmospheric general circulation models, *Mandke S.K., Ramesh K.V. and Satyan V.*, December 2001, RR-092.

- Prospects of prediction of Indian summer monsoon rainfall using global SST anomalies, *Sahai A.K, Grimm A.M., Satyan V. and Pant G.B.*, April 2002, RR-093.

- Estimation of nonlinear heat and momentum transfer in the frequency domain buy the use of frequency co-spectra and cross-bispectra, *Chakraborty D.R. and Biswas M.K.*, August 2002, RR-094

- Real time simulations of surface circulations by a simple ocean model, *P Rahul Chand Reddy, Salvekar P.S., Ganer D.W. and Deo A.A.*, January 2003, RR-095

- Evidence of twin gyres in the Indian ocean : new insights using reduced gravity model by daily winds, *P Rahul Chand Reddy, Salvekar P.S., Ganer D.W. and Deo A.A.*, February 2003, RR-096

- Dynamical seasonal prediction experiments of the Indian summer monsoon, *Milind Mujumdar, R. Krishnan and V. Satyan*, June 2003, RR-097

- Thermodynamics and dynamics of the upper ocean mixed layer in the central and eastern Arabian sea, *C. Gnanaseelan, A.K. Mishra, Bijoy Thompson and P.S. Salvekar*, August 2003, RR-098

- Measurement of profiles and surface energy fluxes on the west coast of India at Vasco-Da-Gama, Goa during ARMEX 2002-03, *S. Sivaramakrishnan, B.S. Murthy, T. Dharmaraj, Cini Sukumaran and T. Rajitha Madhu Priya*, August 2003, RR-099

- Time-mean oceanic response and interannual variability in a global ocean GCM simulation, *K.V. Ramesh and R. Krishnan*, September 2003, RR-100