CHAPTER 4

Peninsular Indian Rainfall and its Association with Meteorological and Oceanic Parameters over Adjoining Oceanic Region

4.1 INTRODUCTION

The Indian summer monsoon exhibits large interannual and intraseasonal and spatial variability. Many observational and modeling studies have pointed out that the slowly varying surface boundary conditions, particularly in the winter and pre-monsoon seasons, contribute a major forcing on the interannual variability of the monsoon. Global and regional parameters representing these conditions provide the handle for seasonal prediction. Empirical modeling strategies include the identification of reliable precursors and an optimal utilization of the information contained in the data on precursors. The reasonable success achieved by the empirical approach has motivated further work on global/regional teleconnections of the Indian summer monsoon season. This resulted in a large number of predictors as well as a variety of statistical techniques. Despite the fact that atmospheric general circulation models have made quick and appreciable advances in the recent past with vastly improved representations of the Asian summer monsoon, empirical approaches still continue to have an important place in operational seasonal forecasting. In spite of the strong physical basis inherent in the empirical formulations, the relationships considered are by no means consistent in space and time, with consequent implications to their predictive skills. The notable failure of empirical models in predicting the deficient summer monsoon of 2002 is a case in point. Modeling methodologies do not appear to substantially improve
Chapter 4 A Regression Model for Peninsular Indian Rainfall

forecast skill for a given set of predictors. Further, several studies have shown that decadal variability plays a major role in the secular variation of predictand-predictor relationships. Predictability of intraseasonal variability also is a limiting factor that is proving to be a daunting task. In order to provide seasonal monsoon predictions of practical value, predictands based on sub-regional and sub-seasonal monsoon rainfall need to be given more emphasis in empirical forecasting research. While a few attempts have been made in this direction, sufficient number of predictors is not available specific to such predictands. Therefore, predictor identification needs to be pursued in a comprehensive manner using modern data sets (like reanalysis), to identify new predictors with possibly non-linear teleconnections. There have been some methodological innovations recently to handle secular variations in the teleconnections and optimize the predictive information.

There is sufficient empirical and theoretical evidence to believe that sea surface temperatures exert significant control over the atmosphere. The works of Singh (1983), Joseph and Pillai (1984), Rao and Goswami (1988), Vinayachandran and Shetye (1991), Sadhir et al (1991) have shown the significance of SST over Arabian Sea as an input parameter for the Indian monsoon rainfall. The rainfall is also related to the sea-surface temperatures (SSTs) of the Indian Ocean, Arabian Sea, and Bay of Bengal. Shukla (1975) used a numerical model and found that cold SSTs in the Arabian Sea would reduce evaporation, increase surface pressure downstream, and therefore reduce monsoon rainfall over India. Washington et al (1977) attempted to reproduce the simulations of Shukla and reported the effect of local SSTs on monsoon rainfall.

Weare (1979) used an empirical approach involving principal component analysis (PCA) and found that SSTs in the Arabian Sea and Indian Ocean over the
period 1949–1972 were negatively, but weakly, related to monsoon precipitation throughout India. Rao and Goswami (1988) examined ship records from 1900–1979 and found that warmer SSTs in the southern portion of the Arabian Sea tended to generally increase monsoon rainfall; similar findings were reported by Kumar and Sastry (1990) and Hastedrath and Grekbar (1993). SSTs in the Arabian Sea, Indian Ocean, and Bay of Bengal were positively related to rainfall in southern India and weakly related to the rainfall in other regions of the subcontinent. Many others (Shukla and Misbra 1977) have found similar results about the significant role of the SST anomalies over the Indian Ocean, the Arabian Sea and the Bay of Bengal. These studies point out the importance of the moisture budget and the amount of evaporation taking place in these seas as a determinant of the amount of precipitation taking place in different parts of the country.

Indian Ocean is different from other Oceans in many aspects. The northern boundary of Indian Ocean does not extend beyond 25° N. It is not bounded by a solid coastal eastern boundary, as that of the Atlantic and Pacific Oceans. The Indian Ocean is split into two basins - the Arabian Sea and the Bay of Bengal. Thus the Indian Ocean doesn't have the currents to transport and discharge heat to higher latitudes, as the Gulf Stream and Kuroshio do in the Atlantic and Pacific, respectively. It is relatively well connected to the western part of the Pacific Ocean characterized with its Warm Pool. The two-basin split of the ocean is a recipe for winds pattern unlike any other over the rest of the oceans where more stable trade winds patterns are observed.
Fig 4.1. Rainfall anomaly for the period 1901-2005 over peninsular India. The mean rainfall for the period is 661 mm with a standard deviation of 94 mm.

The forecast of monsoon rainfall was first made by Blanford, based on his hypothesis that varying extent and thickness of the Himalayan snows exercise a great and prolonged influence on the climate conditions and weather of the plains of northwest India (Blanford, 1884). His success in Long Range Forecast (LRF) of monsoon led to start operational LRF. Sir John Elliot, who succeeded Blanford gave forecasts based on (i) Himalayan snow cover, October to May, (ii) local peculiarities of pre-monsoon weather in India and (iii) local peculiarities over Indian Ocean and Australia (Thapliyal 1987). Expanding on the works Lackyer and Lackyer (1904), Walker (1908, 1918, 1923) initiated extensive studies of worldwide variation of weather elements such as pressure, temperature, rainfall etc., with the main aim to develop an objective method for LRF of monsoon rainfall over India. These studies led him to identify three large-scale pressure seesaw patterns; two in the Northern Hemisphere (North Atlantic
Oscillation, NAO and North Pacific Oscillation, NPO) and one in the Southern Hemisphere (Southern Oscillation, SO). While the NAO and NPO are essentially regional in nature, the SO has global-scale influences, which was later linked to the oceanic phenomenon called El Niño in the east-equatorial Pacific characterized by warming of the sea surface along the Peru coast; this led to the theory of Walker Circulation (Bjerknes, 1969). Walker was the first to introduce an objective method by including Correlation Coefficient (CC) analysis. Walker (1924) also carried out LRF over three homogenous regions of India namely Northeast India, Peninsular India and Northwest India.

Most of the studies on LRF of Indian monsoon rainfall are based on empirical or statistical techniques. These statistical techniques range from simple correlation analysis to advanced procedures such as canonical correlation analysis and artificial neural networking. But the commonly used statistical technique for LRF of monsoon rainfall is the linear regression analysis. A large number of regression models have been proposed so far (Hastenrath, 1991). The predictors for the model are chosen either subjectively or by applying objective criteria. The limitation of the subjective selection is that it may not optimize the variance explained whereas the objective selection is highly sensitive to the data window and overfit the data sample (Thapliyal 1987; Parthasarathy and Sontakke 1988; Hastenrath and Greischa, 1993). The reliability of regression models needs to be assessed by testing on as large and independent data sets, as they are sample-specific in nature.

Gowariker et al (1989) developed parametric and multiple power regression (MPR) models with 15 predictors for LRF of AISMNR, which were modified (Gowariker et al., 1991) to include 16 predictor parameters. The model indicates the likelihood of the monsoon rainfall to be excess or deficient, depending upon
the proportion of favourable/unfavorable parameters out of the total of 16 parameters. This method is highly sensitive to the nature of the predictor data set. The power regression model, successful in operational LRF during the period 1988-97, claims to account for possible non-linear interactions of different climatic forcings with the Indian monsoon system. But the predictors were identified based on their linear correlation with AISMR. The model requires rigorous statistical testing using longer and homogeneous data sets and independent verification as it failed in 1994 and 1997 as other models.

The dynamic stochastic transfer (DST) model developed for the prediction of AISMR as well as the monsoon rainfall over peninsular and northwestern India has only one predictor (Thapliyal, 1987). But the model performs with high accuracy than multiple-regression, MPR, ARIMA (Thapliyal, 1990). Shukla (1987) and Gregory (1989) suggested grouping the subdivisional rainfall, to define area averages for large homogeneous regions due to the large spatial variability of monsoon rainfall. They yielded better formulae for forecasting rainfall over homogeneous regions than when India was treated as one unit. Kumar (1994) attempted LRF of monsoon rainfall over the 29 meteorological sub-divisions in India using canonical correlation analysis (CCA) technique and found that the spatial extent and the magnitudes of skill scores are much larger than those obtained with multiple regression analysis (Prasad and Singh, 1992). The problems in LRF is also addressed (Rajeevan, 2001)

The studies on the seasonal prediction of monsoon rainfall using general circulation models (GCMs) are very few. This may be partly attributed to the lack of skill in the simulation of monsoon rainfall over the Indian subcontinent (Gadgil et al, 1992). Also, different GCM simulation gives marked differences in the simulated monsoon precipitation (WCRP, 1992). The simulation of the
Indian summer monsoon rainfall is very sensitive to the initial conditions (Palmer et al, 1992) also adds up problem to the present situation. However, strong and weak monsoon circulations based on the SST distributions over tropical Pacific and Indian Oceans could be simulated (Ju and Slingo, 1995 and Soman and Slingo, 1997). Thus, the sensitivity of the model in simulating the interannual variability in tropical circulation seems to be closer to the observed characteristics of the monsoon, while a realistic simulation of monsoon rainfall is yet to be achieved by most GCMs.

Reliable predictors for LRF are identified by analysing the relationships between Rainfall and regional/global fields of several surface/upper-air parameters. Correlation coefficients are used to identify the various forcings on the monsoon. They also exhibit notable sensitivity to the data window considered both in terms of the position and length of the window in the time domain. This leads to variations in their magnitude as well as sign, imposing some limitations on the reliability of the predictors.

### 4.1.1 Interrelationships among the predictors

Large numbers of predictors identified so far are highly interrelated and they fall into one of the categories like regional conditions, ENSO indicator, cross equatorial flow etc. The presence of such high multicollinearity among the predictors imposes the problem of redundancy and unnecessary loss of degrees of freedom when they are used in large numbers in regression-based forecast schemes (Kumar et al 1995).

Though the predictors can be classified into four different groups based on their known physical linkage with the monsoon, the forcings represented by them are not entirely exclusive to their respective groups. The forcings represented by the
various predictors can be objectively delineated and the common variance among them pooled into a set of independent principal components. However, not much work has so far been done using this approach. The results of a work by Kumar et al (1997) indicate that a single component accounts for about half of the total variance in the predictors. This component apparently represents ENSO type variability in the predictors and is highly and significantly correlated with AISMR. This clearly shows that the ENSO has a ubiquitous influence on the monsoon circulation, playing a dominant role in the LRF.

In a macro regional scale, peninsular India has got very low degrees of relation to the AISMR (Nayagam et al, 2007). A great amount of attention is given for work on regional scale forecast models for regions such as Peninsular India, northwest India and other homogenous regions (Rajeevan et al, 2000). Fig 4.1 shows the rainfall anomaly obtained over the Peninsular Indian region during the summer monsoon season for the period 1901-2005. The mean rainfall for the period is 661 mm with a standard deviation of 94 mm. The rainfall expressed in anomaly shows that the rainfall anomaly has a positive increase after 1975. Of the nineteen WET years during the period, 9 occur after 1970's. The WET years are defined as years with rainfall greater than one standard deviation. This necessitates a study in the recent years. The correlation between the AISMR and subdivisional rainfall is given in Table 4.1. It is noteworthy that all the subdivisions bear low correlation values with AISMR, which points out the requirement of a new model for the peninsular region separately. Therefore, despite attempting to have a predictive formula for India as a whole, it would be sensible to have identified some parameters, which uniquely explain the Peninsular Indian Rainfall (PIR). This work is an attempt to derive some meteorological and oceanic parameters over the adjoining oceanic region of peninsular India and to regress them into a linear model. The study is
concentrated on the recent 32 years from 1975-2006 and the region over which the rainfall is predicted is shown in Fig 4.2.

<table>
<thead>
<tr>
<th>Subdivisions</th>
<th>Correlation with AISMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal Karnataka</td>
<td>0.03</td>
</tr>
<tr>
<td>South interior Karnataka</td>
<td>0.13</td>
</tr>
<tr>
<td>North interior Karnataka</td>
<td>0.3</td>
</tr>
<tr>
<td>Kerala</td>
<td>0.09</td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>0.15</td>
</tr>
<tr>
<td>Coastal Andhra Pradesh</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 4.1. The correlation of AISMR with the subdivisional seasonal rainfall over peninsular India

Fig 4.2. Geographical map of peninsular India. Shaded area constitute the macro region, of which the rainfall data has been taken as peninsular Indian Rainfall (PIR)

4.2 DATA

Sea Surface Temperature used is, NOAA Optimum Interpolation (OI) SST V2 having 1X1° resolution in latitude and longitude. Air temperature, Zonal and
Meridional wind for 1975-2006 used for this study has been taken from NCEP. NCEP Reanalysis Derived data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/. The spatial resolution of the data is 2.5X2.5° latitude and longitude. This has 17 vertical levels. A more detailed description can be seen in Kalnay et al (1996).

Another data set used is the monthly rainfall of peninsular India obtained from the Indian Institute of Tropical Meteorology website (IITM-IMR). The primary datasets are from India Meteorological Department. The network selected for the study consists of 306 uniformly distributed stations for which rainfall data are available from 1871. The selection of the network of rain-gauge stations were based on the criteria that the network would provide one representative station per district having a reliable record for the longest possible period. The monthly (January - December) area weighted rainfall series for each of the 30 meteorological subdivisions have been prepared by assigning the district area as the weight for each rain-gauge station in that subdivision. Similarly assigning the subdivision area as the weight to each of the subdivisions in the region, area weighted monthly rainfall series are prepared for homogeneous regions of India as well as for all India, as a whole.

4.3 METHODS

4.3.1 Derivation of predictors and selection of best predictor sets

Linear regression model is a popular statistical method for meteorological prediction on various timescales like inter-annual, intra-seasonal, monthly, weekly etc. The main task of the development of the regression model includes the selection of predictors based on empirical relations between various parameters and rainfall, their careful and optimum selection in a stepwise
regression analysis, formulation of the regression equation and verification on independent samples.

The PIR was taken for the period 1975–1997, which we refer to as the training period. Spatial correlation coefficient (CC) was calculated between PIR and the set of parameters under consideration for the period. The areas that bear CC’s at 1% level of significance were identified and selected for calculating indices by taking area averages of the parameter over the respective significant area. The CC of these indices with PIR were checked for consistency for the entire period of analysis by doing 15-year sliding window correlation (Bell, 1977) Indices that have a significant CC at 5% level in all the 15-year sliding windows were retained. Thus, a total of 14 predictors were selected as the input to the stepwise regression analysis and are listed in Table 4.2. The sliding correlation for these predictors is shown in Fig 4.3.

![Figure 4.3](image)

**Fig. 4.3.** Fifteen years moving C.C. between ISMR and selected 14 predictors for the period 1975–1998. The horizontal dotted lines represent the C.C. significant at 90% level. The central year is shown in the x-axis. The details of the predictors are given in Table 4.2.
Stepwise regression analysis was employed to reduce the dimensionality of these indices \((Draper and Smith, 1981)\). Thus, the extensive basic predictor set is reduced to a candidate set by eliminating those with less influence in the variance of rainfall. The predictors thus identified are also listed in Table 4.2. The correlation maps of the selected variables with the KSMR are shown (Figure 4.4 a-d). The CC's which are significant at 1% are contoured, and the selected areas are marked (rectangles in Figure 4.4 a-d).

Formulation of regression equation from among the refined candidate predictors, various regression models were formulated with the following general form using different iteration schemes.

\[
R_j = a_0 + \sum_{i=1}^{k} a_i(X_i) + \varepsilon_j
\]  

\((4.1)\)

Where, \(R_j\) is the dependent variable (rainfall) for \(j=1\) to \(m\) time steps and \(X_i\) are the independent variables where \(i\) is the number of predictors and \(j\) analogous to the previous. \(a_0\) and \(a_i\) are model constants and \(\varepsilon_j\) the error value in that estimation.

The number of predictors that can be used in a statistical regression model has been a matter of debate. Studies like \((Wilks 1995, Delsole et al 2002)\) says that the predictors should be restricted to a small number whereas other studies have shown that eight to ten predictors are required for explaining a good amount of variation (70–75%) in the model development period \((Rajeevan et al, 2004)\). Here the stepwise regression has restricted the number of predictors to four.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Predictors</th>
<th>Level</th>
<th>Month</th>
<th>C.C. (1975-1997)</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Temperature</td>
<td>An850may</td>
<td>850 hPa</td>
<td>May</td>
<td>0.492</td>
<td>75-85E, 10-20N</td>
</tr>
<tr>
<td>Air Temperature</td>
<td>An925mar</td>
<td>925 hPa</td>
<td>March</td>
<td>-0.468</td>
<td>50-75E, 10-20N</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>Rhum1000mar</td>
<td>1000 hPa</td>
<td>March</td>
<td>0.512</td>
<td>60-75E, 0-8N</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>Rhum925mar</td>
<td>925 hPa</td>
<td>March</td>
<td>0.479</td>
<td>50-75E, 10-20N</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>Rhum600may</td>
<td>600 hPa</td>
<td>May</td>
<td>0.560</td>
<td>75-85E, 0-5N</td>
</tr>
<tr>
<td>Sensible Heat Flux</td>
<td>Shtflmar</td>
<td>Surface</td>
<td>March</td>
<td>0.514</td>
<td>55-65E, 10-20N</td>
</tr>
<tr>
<td>Specific Humidity</td>
<td>Shum925mar</td>
<td>925 hPa</td>
<td>March</td>
<td>0.534</td>
<td>55-65E, 10-18N</td>
</tr>
<tr>
<td>Sea Surface Temperature</td>
<td>SSTaug1</td>
<td>Surface</td>
<td>August</td>
<td>0.494</td>
<td>55-65E, 5S-10N</td>
</tr>
<tr>
<td>Sea Surface Temperature</td>
<td>SSTaug2</td>
<td>Surface</td>
<td>August</td>
<td>0.549</td>
<td>85-95E, 10-20N</td>
</tr>
<tr>
<td>Sea Surface Temperature</td>
<td>SSTmar</td>
<td>Surface</td>
<td>March</td>
<td>0.559</td>
<td>70-80E, 10-20S</td>
</tr>
<tr>
<td>T1000</td>
<td>T1000</td>
<td>1000 hPa</td>
<td>March</td>
<td>0.534</td>
<td>70-110E, 5-13S</td>
</tr>
<tr>
<td>Zonal Wind</td>
<td>Uwnd700feb</td>
<td>700 hPa</td>
<td>February</td>
<td>-0.459</td>
<td>60-85E, 3-10N</td>
</tr>
<tr>
<td>Zonal Wind</td>
<td>Uwnd850mar</td>
<td>850 hPa</td>
<td>March</td>
<td>0.555</td>
<td>55-70E, 10-15N</td>
</tr>
<tr>
<td>Meridional wind</td>
<td>Vwnd700jan</td>
<td>700 hPa</td>
<td>January</td>
<td>-0.579</td>
<td>65-75E, 10-20N</td>
</tr>
</tbody>
</table>

Table 4.2. List of parameters, their geographical and temporal location and correlation coefficients with PIR

The dynamics of the predictors used in this study are detailed here. The intensive NH summer insolation induces more Precipitable Water (PW) in the northern Tropics from North Africa to South and East Asia. Similarly, enhanced SH insolation in boreal winter and spring gives rise to increased PW over the
Arabian Sea, India, and the southern tropical Indian Ocean in boreal summer. Air temperature at 850 hPa (May) having positive correlation coefficient in the analysis imply the possible connection to rainfall as follows. When air temperature increases, the Planetary Boundary Layer (PBL) becomes more turbulent and the winds will accelerate. This will increase the transport of moisture into the atmosphere. The gradient of air temperature can affect the winds through thermal wind relation, which in turn affect the rainfall.

Sea Surface Temperature (March) bears a significant positive correlation coefficient with PIR. SST anomalies will lead to more evaporative and sensible heat flux anomalies, which in turn affect the winds. The gradient of SST is directly influencing the winds at upper levels through thermal wind relation.

Zonal and meridional wind at 700 hPa during February and March respectively bear a negative correlation coefficient with PIR. When the winter winds are strong (correspondingly the subtropical westerly jet is stronger) the ensuing rainfall will be less.
4.4. RESULTS

The selected indices over the Arabian Sea and adjacent seas were subjected to stepwise regression and four potential predictors for PIR were identified. The final form of the regression model is

\[ Y = -4364.514 - 33.354(x_1) + 48.035(x_2) + 135.717(x_3) - 15.876(x_4) \]  \hspace{1cm} \text{(4.2)}
where $X$ values are the predictors (Table 1) and $Y$ the predicted rainfall. The regression coefficients in the model show disproportionate values because the predictors were not expressed in standardised form. The regression coefficients do not imply relative importance, as the input parameters into the regression model are not standardized.

Multicollinearity is a measure of non-orthogonality of the predictors, i.e. there exists inter-correlation between the predictors. Thus, a multiple linear regression model lacks in its accuracy and may lead to unclear interpretation of the regression coefficients as measures of original effects (Mc Cuen, 1985). This leads to the problem of redundancy and unnecessary loss of degrees of freedom when they are used in large numbers (Kumar et al., 1995). Small data revisions would result in disproportionate effect on the calculated coefficients, i.e. the coefficients obtained will be unstable and this in turn affects the prediction. The reliability of these parameters in the prediction could be ensured by understanding the multicollinearity statistics. Variance inflation factor (VIF) is a common method used to study the multi-co-linearity (Fox, 1991). It is defined as:

\[
VIF (a_i) = \frac{1}{1-R_i^2}
\]

where $R_i^2$ is the unadjusted $R^2$ when $X_i$ is regressed against all the other explanatory variables in the model. The VIF measures how much the variance of the estimated regression coefficients are inflated compared to situations when the independent variables are uncorrelated. Values in excess of 10.0 could significantly affect the stability of the regression coefficients (Neter et al., 1990). So, the candidates having a smaller VIF only are considered so that the
parameters inhibiting interdependence is avoided to increase the credibility of
the regression output.

The VIF analysis of all the parameters that is retained in the stepwise regression
have values around 1.2 indicating an insignificant level of multicollinearity
(Table 4.3).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sig. of t-test</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Temperature</td>
<td>0.015</td>
<td>1.134</td>
</tr>
<tr>
<td>Sea Surface Temperature</td>
<td>0.001</td>
<td>1.078</td>
</tr>
<tr>
<td>Zonal Wind</td>
<td>0.020</td>
<td>1.253</td>
</tr>
<tr>
<td>Meridional Wind</td>
<td>0.011</td>
<td>1.187</td>
</tr>
</tbody>
</table>

Table 4.3. The t-test results and variance inflation factor corresponding to the variables.

Scatter matrix plot shows the relationship between each pair of variables (Fig.
4.5). Each column shows the relationship between the parameter listed in that
column with the other four parameters named in the respective rows. The
variable on the vertical axis is the variable named in that row and the variable on
the horizontal axis is the variable contained in that column. The correlation
coefficients between the dependent variable and four independent variables are
listed atop the columns. From the figure it can be seen that the PIR is positively
related to air temperature and Sea Surface temperature, but negatively correlated
to zonal and meridional wind at 700mb. The independent parameters do not
show much relation among themselves though feeble relations exist. Therefore
the predictors are independent and do not have any significant patterns. So the
interdependencies among the predictors, if any, are well below the potential to
affect the forecast.
When the $t$-test is conducted, the $P$ values indicate that the coefficients of all the predictors in regression are highly significant, with maximum significance for SST (Table 4.3). The model has a coefficient of determination (which is the proportion of variation in the dependant variable explained by the regression model) of 77.7% and a multiple correlation coefficient (correlation between the observed and predicted rainfall) of 0.88. The $F$ statistic, which tests the significance of regression model, gives $F=15.64$ and the $P$ value in the ANOVA test (Analysis of Variance) is very small. This strongly suggests that the included predictors collectively account for a significant part of the variance in the PIR.

![Scatter Matrix Plot](image)

**Fig. 4.5.** Scatter Matrix Plot of the inter relations between the dependant and independent parameters and among predictors. The predictors are Air temperature at 850 hPa for the month of May (air850may), Sea Surface Temperature for the month of March (SSTmar), Zonal wind at 700 hPa for the month of February (uwnd700feb) and Meridional wind at 700 hPa for the month of January (vwnd700jan). RFPEN is the Peninsular Indian rainfall.
The standardized departure of PIR and its forecast is shown in Fig 4.6. PIR has a standard deviation of 110.1mm, which is 16.6% of the mean rainfall (662.9 mm). The forecast of the model presented here, picks a standard deviation of 14.8% (101.1mm) of the mean rainfall. Thus both the mean (684.2mm, 662.9 mm) and the standard deviations (101.1mm, 110.1mm) of the forecast are comparable with the observed. This statistics shows that the interannual variability of PIR is well represented by this model.

![Fig 4.6: Standardized departures of PIR and its forecast](image)

During the test period the standardized departures of the actual and model fitted rainfall shows four years (2000, 2001, 2002 and 2003) with opposite sign of anomaly (Fig.4.6). Except 2002, in all these years the observed values are very close to the mean, so that a small deviation of the forecast from the observed rainfall may end up with a departure opposite to that of the observed. But, none
of these years were extreme, except 2002, and they lie within the range of one standard deviation of normal monsoon. The model showed the maximum departure for the year 2002, in which other models also failed.

The model performance was assessed by calculating CC, Root mean square Error (RMSE), Absolute Error (ABSE) and the bias (BIAS) for the (1) training period, and (2) test period. Equations used to calculate these measures are given below (Nicholls, 1984; Hastenrath, 1987).

\[
\text{RMSE} = \left[ \frac{1}{n} \sum_{y=1}^{n} (\hat{R}_y - R_y)^2 \right]^{1/2} \tag{4}
\]

\[
\text{BIAS} = \frac{1}{n} \sum_{y=1}^{n} (\hat{R}_y - R_y) \tag{5}
\]

\[
\text{ABSE} = \frac{1}{n} \sum_{y=1}^{n} |\hat{R}_y - R_y| \tag{6}
\]

Where \( \hat{R}_y \) is the model fitted rainfall, \( R_y \) is the observed rainfall and \( n \), the number of years. For the model development period high C.C of 0.88 was obtained between the observed and predicted rainfall. The RMSE was 7.6 % (50.48 mm) of observed mean rainfall, BIAS was 0 mm and ABSE was 40.2 mm. The CC has dropped to 0.67 (significant at 1% level) for the period 1975-2006. The RMSE for the test period (1998-2006) has increased to 21.5% (142.79 mm) of mean rainfall. BIAS and ABSE for the predicted rainfall are 75.59 mm and 119.18 mm respectively. Climatological predictions are also made and the RMSE, BIAS and ABSE were also computed. The values obtained for RMSE, BIAS and ABSE are 117.61 mm, -67 mm and 96.73 mm respectively.

Scatter plots of residuals with the predictors and fitted values (not shown), helps us to visualize the patterns in the residuals. Any organized pattern present in it reveals the existence of the relation between fitted and residuals. If any pattern
could be identified then the model can be further improved by extracting the relationship. No such relation is identified and therefore no relationship is found. The Durbin Watson statistic is also used to check the relationship in the residuals. It checks whether the residuals for successive observations are uncorrelated. If any significant lag one autocorrelation exists, then the residuals are not independent and the model can be further improved (Makridakis et al, 1998). It is defined as

\[ DW = \sum_{t=2}^{n}(e_t - e_{t-1})^2 / \sum_{t=1}^{n}e_t^2 \] ... (7)

where \( e_t \) is the residual at the time \( t \) and \( e_{t-1} \) that at time \( t-1 \). Its value ranges from 0 to 4. Values above two means there is some negative autocorrelation and values below 2 has a positive correlation. A value of 1.854 is obtained which suggests that the lag-1 autocorrelation is negligibly small in the residuals.

4.5 SUMMARY AND CONCLUSION

In this study a linear regression model was developed and tested for the seasonal rainfall prediction of Peninsular India using immediate regional parameters such as (i) lower tropospheric temperature during the month of May over the Peninsular India and adjoining Bay of Bengal (ii) Sea Surface Temperature during the month of March over the mid-Indian Basin over the Indian Ocean (iii) Zonal wind at 700 hPa during February and (iv) meridional wind at 700 hPa during January. The period of analysis was from 1975-2006 of which last nine years were tested against the observed rainfall. These parameters were arrived from a larger set of candidates, which had statistically significant relation with PIR. Our results show that the empirical model could explain 77.7% of the total variance in the PIR for the model development period. The model showed an RMS error of 7.6%, BIAS of 0 mm and an absolute error of 40.2 mm. The model could capture the extreme years reasonably but for the years 2001, 2002 and
2003 the model over predicted the rainfall. It is noteworthy that in 2002, a prolonged 30-day break in summer rains had lead to an extremely dry monsoon season whereas IMD models in use at the time had predicted normal rains in 2002.