3

An Analysis of Rainfall in Orissa with Sea Surface Temperature Anomaly by Similarity Search Technique

D Nagesh Kumar, Rajib Maity

ABSTRACT

It is believed that temperature is one of the key influencing factors in producing precipitation over both land and oceans. The relationship is yet to be established in terms of a mathematical model. This paper intends to address this issue. There are two main hypotheses, which control the precipitation on a global scale. Among these, the first one assumes it to be the manifestation of seasonal migration of Inter Tropical Convergence Zone (ITCZ) and the other assumes that the land ocean temperature contrast drives the monsoon. The analysis performed in this paper is based on the second hypothesis. The most influencing Sea Surface Temperature (SST) zone is identified using Similarity Search Technique affecting the rainfall over Orissa region in India. The relationship between the SST of that zone and the corresponding rainfall is found out and its further use is indicated.

INTRODUCTION AND SCOPE

Numerous studies have been carried out to understand the rainfall phenomenon. An effort has been made to identify the statistical relationship between different climatic indices and rainfall. In India, in the months of June, July, August and September (JJAS), more than 75% of annual rainfall occurs. Due to its unique character, Indian monsoon rainfall attracts many hydrologists and meteorologists to work on its behavior. A significant number of studies had been performed to assess the monsoon rainfall over India from different view points. Significant negative correlation between Indian summer monsoon rainfall and El Nino-Southern Oscillation (ENSO) was established in eighties (Rasmusson and Carpenter, 1983). Rainfall over some large-basins in India is also shown to be negatively correlated with ENSO (Rao, 1997). However, contrary to the long recognized negative correlation between the Indian summer monsoon rainfall and El Nino-Southern Oscillation, India received slightly above normal summer rainfall in 1997 (Li *et al.* 2001). This recent experience suggests that the response of monsoon to El Nino is not yet understood adequately (Gadgil *et al.* 2003). The recent identification of Dipole Mode Index (DMI) in Indian Ocean region (Saji *et al.* 1999) helps to improve understanding about rainfall over the surrounding countries including India. It shows that a good monsoon is associated with a positive phase of DMI and vice versa (Saji *et al.* 1999). This type of analysis is based on the observation in the past. The nature of the time series of individual indices as well as their spatiotemporal relationship with rainfall phenomenon is not yet fully established.

Occurrence of rainfall phenomenon is based on two different hypotheses. The monsoon occurs due to the seasonal variation of Inter Tropical Convergence Zone (ITCZ), which is defined as the zone of convergence of moist air in the lower troposphere with its strong ascent up to the upper troposphere leading to deep convective clouds. This propagates into the surrounding region and descends - causing precipitation in that region (Gadgil, 2000). ITCZ lies directly under the sun forming a narrow band of clouds spreading in the east west direction all through the year. It moves in north south direction with seasonal changes, *i.e.*, from July to September, it lies in Northern Hemisphere and from December to February, it shifts to Southern Hemisphere (Figure 1). Hence, it is the seasonal migration of ITCZ that plays an important role on monsoon (Gadgil, 2003).

According to the second hypothesis, land ocean temperature contrast is considered to be the main causal time factor behind the rainfall phenomenon.

Unequal heating of land surface and sea surface creates a temperature gradient between land and ocean, which in turn develops a pressure gradient between them. This pressure gradient drives the air to flow. If the wind comes from the ocean, it picks up moisture during its travel and on reaching the land surface, it is lifted up due to mountain barrier, low atmospheric pressure and such other causes, resulting in rainfall.

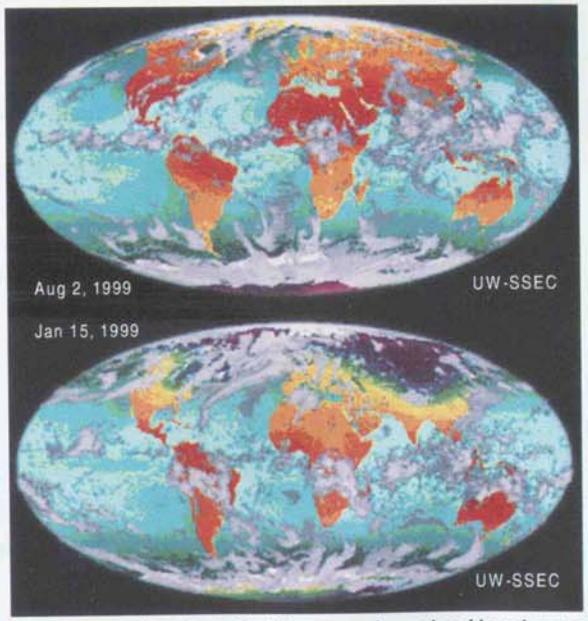

Figure 1. Seasonal migration of ITCZ as seen by a cloud band spreading in east west direction.

(*Source:* http://lilt.ilstu.edu/jrcarter/Geo211/WebPage-211/AdvSSEC1.html)

On the other hand, if wind comes from land surface, there is no scope to pick up moisture and the air remains mainly dry, failing to create rainfall. Hence temperature anomaly over land and sea surface affects both the temperature gradient and the strength of monsoon (Robock *et al.* 2003). Few recent studies also show that the basic driver of monsoon circulation is the land ocean temperature contrast (Li and Yanai, 1996; Liu and Yanai, 2001). According to Chao and Chen (2001) "...whether it (land-sea temperature contrast) really acts as the main driving force of the monsoon has not been tested in numerical experiments...role played by land-sea contrast in the monsoon is basically equivalent to that played by Sea Surface Temperature (SST) contrast. Where land-sea contrast is important for the monsoon, the monsoon can still exist if the land is replaced by ocean of sufficiently high SST.". But in such a case, the distributional pattern of rainfall may significantly change. The location of the Himalayas plays an important role on the distributional pattern of rainfall over India. Hence the relationship between the rainfall phenomenon and land ocean temperature contrast largely depends on the spatial location of the land surface.

In the Indian subcontinent, the seasonal reversal of the wind direction, that blows from the south west during one half of the year and from the north east during the other (Figure 2), is considered to be the result of such temperature contrast (http://yang.gmu.edu/~yang/nasacd/www/indian_monsoon.html).

Figure 2. Seasonal reversal of the wind direction over Indian continent
(*Source:* http://snrs.unl.edu/amet351/ogren/monsoons.html)

The climatic models are effective in very large spatial scale compared to the zonal scale, which is more useful in water resources engineering. Hence the challenge lies in extracting the climatic information from a global scale and using it in a regional scale of a watershed. For this purpose, Similarity Search Technique is used in this paper. This technique helps to identify the similarity in terms of pattern fluctuation, etc. between two time series and helps to pick the best similar pairs from a group of series.

The analysis performed in this paper is based on the second hypothesis. It is intended to find out the most influencing Sea Surface Temperature zone that creates the best possible temperature contrast series to cause rainfall over a particular region on the land surface. Different lags are tested to study the aforementioned relationship. Finally, the visual as well as statistical investigations

have been made to draw some important conclusions. Indications are also provided for further utilization of the results.

The basic theory is discussed on similarity of two time series, similarity transformation and normal form and similarity measurement. The methodology for identification of the most influencing Sea Surface Temperature zone that creates the best possible temperature contrast series to cause rainfall over Orissa in India is also discussed. Results of the analysis and conclusions based on this study are also presented.

THE BASIC THEORY OF SIMILARITY SEARCH

A time series is a sequence of numbers, each number representing a value of a particular variable at a time point. In this section, identification of similarity between two different time series will be discussed and the same will be used for the purpose of this study. Similarity Search approach will help in understanding whether patterns, fluctuations, frequency, etc. of two time series are approximately similar or not.

Similarity of Two Time Series

A general approach to measure the similarity or closeness between two time series is to calculate the Euclidean distance between the two time series considering each of the time series to be a point in n dimensional space, where n is the length of the time series. Two time series are to be considered similar if the Euclidean distance is less than some user defined threshold value (Rafiei and Mendelzon, 1997). But it may not always be easy to capture any type of similarity (Rafiei and Mendelzon, 1997). Therefore, it is important to choose an appropriate parameter to measure the closeness between two given time series.

Similarity Transformation and Normal Forms

Let us assume a time series X of length 'n' has a mean μ and standard deviation σ,

where $$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

and $$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n-1}} \quad (2)$$

The mapping of each element of the series x_i to $(a^* x_i + b)$ defines the similarity transformation $T_{a,b}$. Now a time series X is considered to be similar to a time series Y, if there exists some (a, b) such that $X = T_{a,b}(Y)$. There are a few properties of this similarity transformation $T_{a,b}$ (Golding and Kannellakis, 1995). These are:

Reflexivity: For any time series X, $X = T_{a,b}(X)$, this is also known as identity transformation.

Symmetry: If $X = T_{a,b}(Y)$ then $Y = T_{1/a,-b/a}(X) = T^{-1}{}_{a,b}(X)$. Here $T^{-1}{}_{a,b}$ is known as inverse of $T_{a,b}$.

Transitivity: If $X = T_{a,b}(Y)$ and $Y = T_{c,d}(Z)$ then $X = T_{ac,ad+b}(Z) = (T_{a,b}{}^*T_{c,d})(Z)$.

Here $(T_{a,b} * T_{c,d})$ is known as noncommutative product of $T_{(a,b)}$ and $T_{(c,d)}$.

The set of all transformed series similar to a given time series is known as a similarity class. Now a particular series is known to be normal if its $\mu = 0$ and $\sigma = 1$ and it can be shown that there exists only one normal in a similarity class. In the transformation function $T_{(a,b)}$, 'a' is known as scale factor and 'b' as shift factor. If $a = 1$, the transformation is a pure shift and if $b = 0$, it is known as pure scaling.

Similarity Measurement

The similarity between two time series is measured by similarity distance D_s and is defined between two time series X and Y as the distance between the normal forms of their respective similarity classes (Golding and Kanellakis, 1995).

$$D_S(X,Y) = D_E(X^*,Y^*) \tag{3}$$

where D_E is the Euclidean distance, defined as follows:

$$D_E(S_1,S_2) = \sqrt{\sum_{i=1}^{n}(S_1(i) - S_2(i))^2} \tag{4}$$

In hydrology, generally, discrete time series are more common. Euclidean distance is a standard metric distance in case of such discrete time series. Another point that can be noted here is that the similarity distance between any pair of series chosen from the same similarity class is always the same. A minor problem with this method is that the user has no control on the meaning of similarity other than by comparing with a threshold value. But if the intention is only to compare the different pairs of series to pick up the best pair, this method may be treated as a relative measure.

METHODOLOGY

The analysis is done for Orissa region in India (Figure 3). The monthly Sea Surface Temperature anomaly data for the entire globe is taken for this analysis. SST anomaly values are calculated as follows (http://iridl.ldeo.columbia.edu/SOURCES/.GOSTA/.atlas8/.MOHSST5/.ssta/):

$$a_{i,j} = X_{i,j} - \overline{X_i} \tag{5}$$

where $a_{i,j}$ = Anomaly value for i^{th} month and j^{th} year

$X_{i,j}$ = Actual value for i^{th} month and j^{th} year

$\overline{X}_i$ = Long term mean value for i^{th} month.

The entire globe is divided into 5° × 5° latitude by longitude grids and the anomaly data at the centre of each intersection point is obtained. Monthly rainfall data for Orissa zones are obtained for the period, 1872 to 2002 and data of maximum temperature for east coast India are obtained for the period, 1901 to 1990 (http://www.tropmet.res.in/data.html). Similarly, monthwise anomaly values of rainfall and maximum temperature are calculated from raw data series in the following way:

$$a_{i,j} = \frac{X_{i,j} - \overline{X_i}}{S_i} \tag{6}$$

where $a_{i,j}$ = Anomaly value for i^{th} month and j^{th} year.

$X_{i,j}$ = Actual value for i^{th} month and j^{th} year.

X_i = Mean value for i^{th} month.

S_i = Standard deviation for i^{th} month.

Mean and standard deviation used in the above equations are calculated based on the data length available as mentioned earlier.

The series of anomaly values for maximum temperature of east coast India are deducted from the Sea Surface Temperature anomaly series for all grid points over entire sea surface. Thus, new time series is obtained (one for each grid point on sea surface) which reflect the temperature contrast between Orissa region and that particular sea surface grid point. All these series are monthly and of duration 1901 to 1990. This new time series may contain both positive as well as negative values.

The similarity distance between the temperature contrast time series and the rainfall anomaly time series is calculated for different values of lag following the theory of similarity search. Thus for each lag case, similarity distances are obtained for each grid point on the sea surface through out the globe. Next for all the lag cases, contour maps are plotted with these values. The grid point corresponding to the minimum similarity distance is identified and considered to be the most influencing Sea Surface Temperature zone that creates the best possible temperature contrast series to cause rainfall over Orissa region.

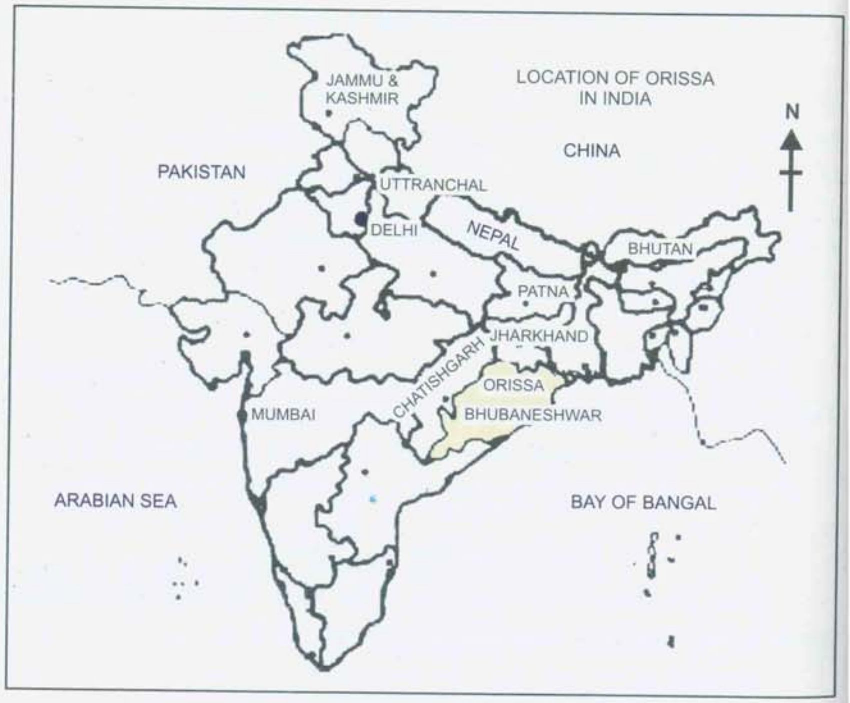

Figure 3. Location map of Orissa

(*Source:* http://www.mapsofindia.com/maps/orissa/orissalocation.htm)

RESULTS AND DISCUSSION

Similarity Search Technique is used in two different ways in this paper. First, it is used to investigate the long term relationship between the All India rainfall and Orissa rainfall. The similarity distances are calculated for monthly,

annually and JJAS rainfall series for sufficiently long time period (25 years). It is intended to investigate any discernible changes between the relationship that may be due to the climate impact because, it is experienced by both the series. The plots are shown in Figure 4. It is seen that the established relationships are more or less same over a sufficiently long period with mild periodic nature.

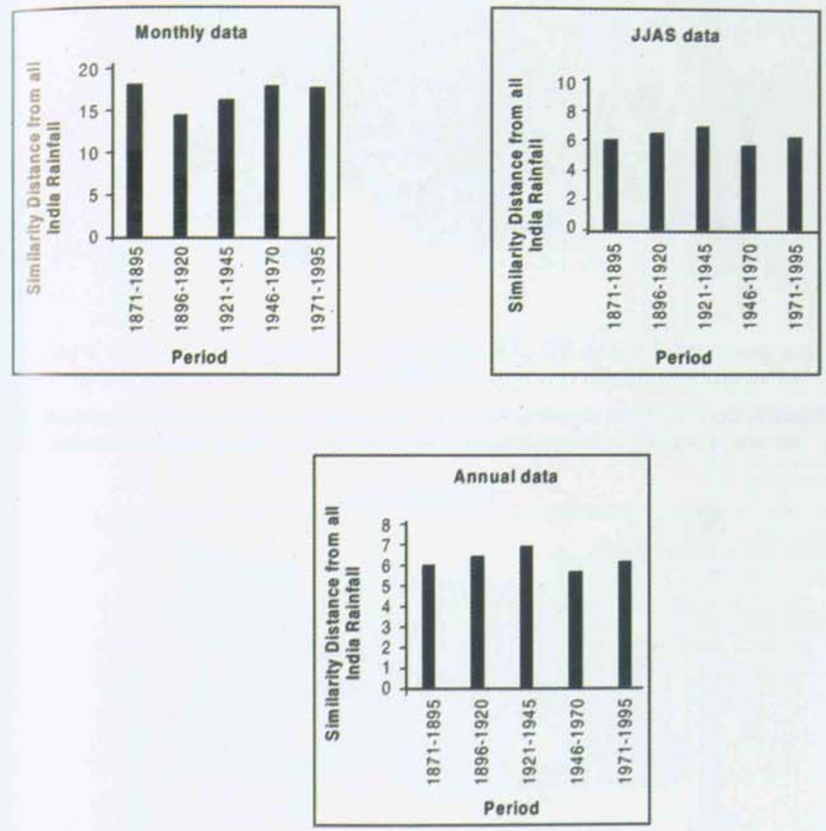

Figure 4. Graphs showing the changes in similarity distances between Orissa and All India rainfall series for monthly, annual and JJAS cases.

Secondly, the SST is used to identify the most influencing Sea Surface Temperature zone for the four different lags considered, namely, one month, two months, three months and four months. The similarity distance obtained for all cases are 43.5621, 43.9022, 43.5038 and 43.4762 respectively. The locations of the most influencing Sea Surface Temperature zone are shown in respective contour plots (Figures 5 to 8).

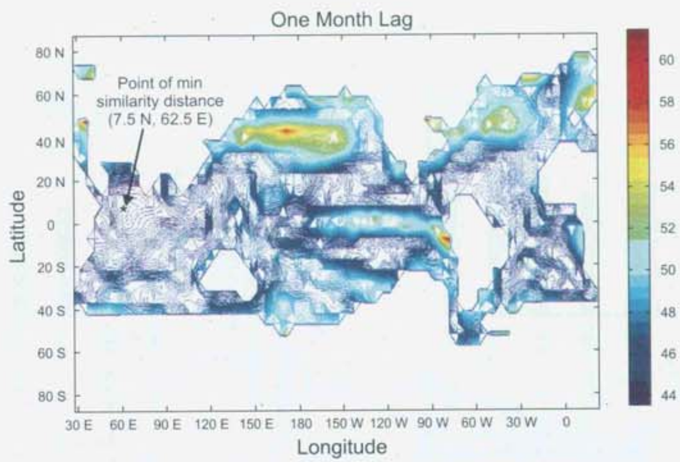

Figure 5. Contour plots showing the most influencing sea surface temperature zone for one month lag between temperature contrast and rainfall anomaly series.

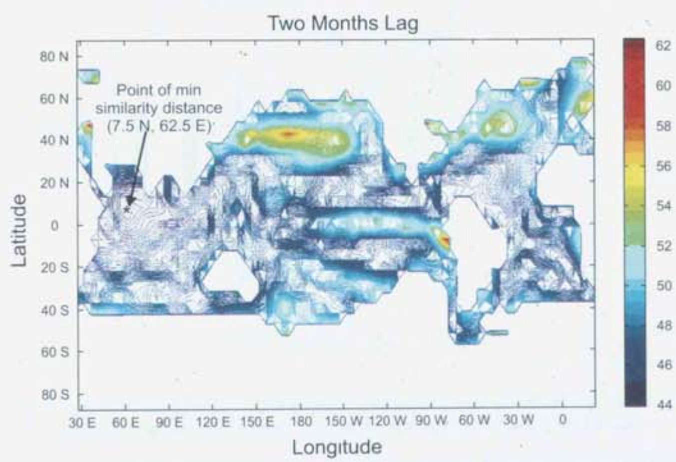

Figure 6. Contour plots showing the most influencing sea surface temperature zone for two months lag between temperature contrast and rainfall anomaly series.

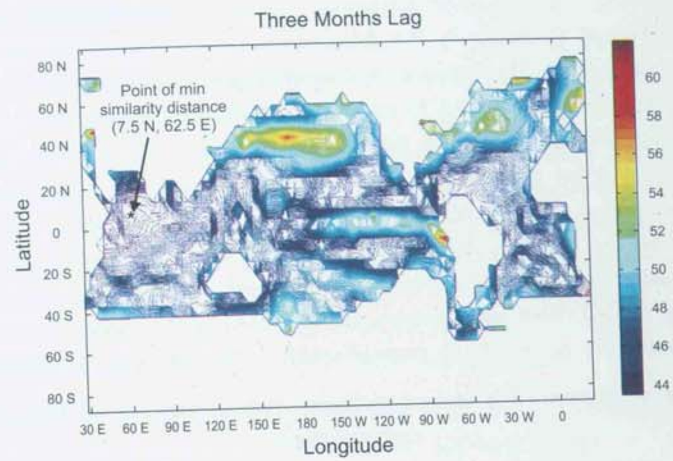

Figure 7. Contour plots showing the most influencing sea surface temperature zone for three months lag between temperature contrast and rainfall anomaly series.

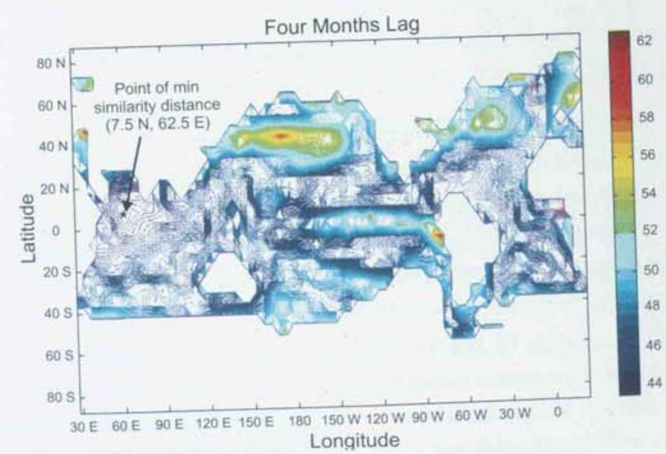

Figure 8. Contour plots showing the most influencing sea surface temperature zone for four months lag between temperature contrast and rainfall anomaly series.

It is seen that for all lags, the most influencing Sea Surface Temperature zone is located in the Indian Ocean at 7.5°N and 62.5°E. The temperature contrast series obtained from this most influencing zone can be investigated to examine the reliability of this analysis. For this purpose, the first case, *i.e.*, lag of one month is considered. The temperature contrast series is calculated from the most influencing region (7.5°N, 62.5°E) and the monthwise plot of temperature contrast and rainfall anomaly series with lag one is investigated. The similarity distance between two series for each case is tabulated in Table 1. By visual inspection of the plots (not shown here) and the similarity distances (Table 1), closeness between the two series is understood. As the two series are close to each other, temperature contrast series can be used as an exogenous input in Dynamic Linear Modeling (DLM) (West and Harrison, 1989) to forecast the rainfall series, which is not under the scope of the present study.

Table 1. Similarity distance between two series for all different combinations

	Jan. rainfall and *Dec* temp. contrast	*Feb.* rainfall and *Jan* temp. contrast	*Mar.* rainfall and *Feb* temp. contrast	*Apr.* rainfall and *Mar* Temp. contrast	*May* rainfall and *Apr* temp. contrast	*Jun.* rainfall and *May* temp. contrast	*Jul.* rainfall and *Jun* temp. contrast	*Aug.* rainfall and *Jul* temp. contrast	*Sep.* rainfall and *Aug* temp. contrast	*Oct.* rainfall and *Sep* temp. contrast	*Nov.* rainfall and *Oct* temp. contrast	*Dec.* rainfall and *Nov.* temp. contrast
Similarity Distance	14.006	13.644	13.270	12.885	10.545	11.417	14.032	12.114	11.827	11.045	12.353	13.340

CONCLUSIONS

In this study, the rainfall series of Orissa and All India are analyzed first. The relationship between these series over a long time period is investigated and nothing significant was found except a long term mild periodicity. Then by Similarity Search Technique, most influencing Sea Surface Temperature zone, that produces best similar temperature contrast series, is identified. The following conclusions are made from this study.

1. The efficacy of Similarity Search Technique to identify the similarity between two time series is shown in this study. Most influencing Sea Surface Temperature zone is identified by this technique. The theory can be applied to any other application in hydrology.
2. The temperature contrast series, obtained from the identified most influencing Sea Surface Temperature zone, and the rainfall anomaly series are very close to each other. This observation justifies the proper identification of most influencing Sea Surface Temperature zone.
3. Once the most influencing Sea Surface Temperature zone is identified, further study can be done with the data only from this zone. Thus, it helps in eliminating the huge data load and facilitates to work in small spatial scale.

Finally, it may be concluded that a firm understanding of hydrology regarding zonal precipitation and information regarding geographical location, etc. are the prerequisites for the proper application of this theory. The similar approach can be adopted to search for the most influencing zone of sea surface for other regional precipitations also.

ACKNOWLEDGEMENT

This work is partially supported by Dept. of Science and Technology, Govt. of India, through a project with reference no. ES/48/010/2003.

REFERENCES

WC Chao, B Chen. 2001. The origin of monsoon, Journal of Atmospheric Sciences, 58, pp.3497-3507.

S Gadgil. 2000. Monsoon-ocean coupling, Current Science, 78, pp.309-323.

S Gadgil. 2003. The Indian monsoon and its variability, Annual Review of Earth and Planetary Sciences, 31, pp.429-467.

S Gadgil, PN Vinayachandran, PA Francis. 2003. Drought of the Indian summer monsoon: role of clouds over Indian Ocean, Current Science, 85, pp.1713-1719.

DO Golding, PC Kannellakis. 1995. On similarity queries for time series data: constraint specification and implementation, Proceedings of 1st International Conference on the Principles and Practice of Constraint Programming, France, pp.137-153.

C Li, M Yanai. 1996. The onset and inter-annual variability of the Asian summer monsoon in relation to land-sea thermal contrast, Journal of Climate, 9, pp.358-375.

T Li, YS Zhang, CP Chang, B Wang. 2001. On the relationship between Indian ocean sea surface temperature and Asian summer monsoon, Geophysical Research Letters, 28, pp.2843-2846.

X Liu, M Yanai. 2001. Relationship between the Indian monsoon rainfall and the tropospheric temperature over the Eurasian continent, Quarterly Journal of the Royal Meteorological Society, 127, pp.909-937.

D Rafiei, A Mendelzon. 1997. Similarities based queries for time series data, Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '97), pp.13-24, Tucson, Arizona, USA.

GN Rao. 1997. Inter annual variation of monsoon rainfall in Godavari river basin-connections with the southern oscillation, Journal of Climate, 11, pp.768-771.

EM Rasmusson, TH Carpenter. 1983. The relationship between eastern equatorial pacific sea surface temperature and rainfall over India and Sri Lanka, Monthly Weather Review, 111, pp. 517-528.

A Robock, M Mu, K Vinnikov, D Robinson. 2003. Land surface conditions over Eurasia and Indian summer monsoon rainfall, Journal of Geophysical Research, 108, pp.4131.

NH Saji, BN Goswami, PN Vinayachandran, T Yamagata. 1999. A dipole mode in the tropical Indian ocean, Nature, 401, pp.360 – 363.

M West, J Harrison. 1989. Bayesian Forecasting and Dynamic Models, Spinger-Verlag, New York, USA.

http://yang.gmu.edu/~yang/nasacd/www/indian_monsoon.html

http://www.tropmet.res.in/data.html

http://www.mapsofindia.com/maps/orissa/orissalocation.htm

http://lilt.ilstu.edu/jrcarter/Geo211/WebPage-211/AdvSSEC1.html

http://snrs.unl.edu/amet351/ogren/monsoons.html

http://iridl.ldeo.columbia.edu/SOURCES/.GOSTA/.atlas8/.MOHSST5/.ssta