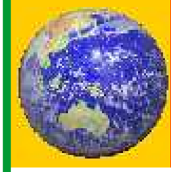
JUNE
2008



NCC
8

NCC RESEARCH REPORT

Development of a high resolution daily gridded temperature data set (1969-2005) for the Indian region

A K Srivastava,
M Rajeevan
S R Kshirsagar

NATIONAL CLIMATE CENTRE
OFFICE OF THE
ADDITIONAL DIRECTOR GENERAL OF METEOROLOGY (RESEARCH)
INDIA METEOROLOGICAL DEPARTMENT
PUNE - 411 005

Development of a High Resolution Daily Gridded Temperature Data Set (1969-2005) for the Indian Region

A K Srivastava, M Rajeevan* and S R Kshirsagar

National Climate Centre
India Meteorological Department, Pune
India 411 005

Corresponding author:

M. Rajeevan
National Climate Centre
India Meteorological Department
PUNE- 411 005 INDIA

Abstract

To meet the requirements of the research community, a high resolution daily gridded temperature data set for the Indian region was developed. For this purpose, daily temperature (maximum, minimum and mean) data of 395 quality controlled stations for the period 1969-2005 were considered. We have used a modified version of the Shepard's angular distance weighting algorithm for interpolating the station temperature data into 1^0 *Lat X* 1^0 *Long grids.*

We evaluated the temperature data set using the cross validation to estimate errors associated with the interpolation technique used and found the root mean square errors are less than 0.5^0*C. The present data set was also compared with another high resolution monthly data set. Correlations between the two data sets are more than 0.8 over most parts of the country. An analysis was also made on the data set, demonstrating its possible applications. Using the data set, mean frequency of cold and heat waves, temperature anomalies associated with the monsoon breaks have been presented. The present daily temperature data set will be made available to the research community.*

1. Introduction

One of the most significant consequences of global warming would be an increase in magnitude and frequency of extreme events like heat waves (IPCC 2007). High resolution daily temperature data are useful for analyzing such climate extremes in the past. They are also useful for validating climate model simulations and environmental modeling applications. Global scale gridded monthly temperature data set (New et al, 2000, Jones and Moberg, 2003) are available to the research community for more than ten years. The data set of Jones and Moberg (2003) forms a basis for the assessment of temperature trends by IPCC for climate assessment reports and by the World Meteorological Organization (WMO) for the annual climate statements. Regional gridded daily temperature data sets are also available for different regions including China (Feng et al. 2004), and the USA (Janowiak et al. 1999 and Piper and Stewart 1996). Recently, Caesar et al. (2006) reported the development of a new daily temperature gridded data set for the globe. The data set contained daily temperature in 2.5^0 latitude by 3.75^0 longitude grid over the period January 1946 to December 2000. The primary source of station data for this data set is the Global Historical Climatology Network-Daily (GHCND). However, this data set does not contain any data from the Indian region. Therefore, there is a need for developing a high resolution daily temperature data for regional climate applications. The present study reports the results of the development of high resolution daily temperature data set using 395 stations from India. The data set was prepared for the period 1969-2005. The details of the data set and results of the analysis based on the data set are reported here.

2. Data

India Meteorological Department (IMD) at present maintains around 550 surface observatories in the country, where daily surface air temperature observations (maximum and minimum) are taken. These data are compiled, digitized, quality controlled and archived at the National Data Centre (NDC) of IMD. However, in the archival, digitized data from only 1969 are available. The data prior to 1969 are still in manuscript form. We have considered the data for the period 1969-2005 for developing the regional gridded temperature data set. From the list of stations for which daily data are available, only those stations which have minimum 10 years of data, at least for 300 days in a year during the period 1971 to 2000, were selected for further analysis. The data were subjected to basic quality checks like rejecting values, greater than exceeding known extreme values, minimum temperature greater than maximum temperature, same temperature values for many consecutive days etc. (Gleason, 2002). Unusual high values were flagged by putting a filter which allowed values only in the range mean ± (1.76+ 0.8N) standard deviation, (Sellers and Liu (1988)), for further data analysis. The flagged values after the above checks were further examined for spatial continuity before rejection. After putting these quality checks, 395 stations were selected for the development of the gridded data set. The network of stations considered for the analysis is shown in Fig.1. The stations are well distributed over the country.

3. Methodology

We have used a modified version of the Shepard's angular distance weighting algorithm (Shepard, 1968) for the present analysis. The same interpolation technique was used by Ceasar et al (2006). Because of its flexibility, this method was also used by New et al. (2000), Kiktev et al. (2003) and Piper and Stewart (1996) for developing the gridded temperature data sets. In order to avoid biases in the gridding, daily anomalies were used instead of the absolute values. For this purpose, climatological normals of maximum and minimum temperatures for the period 1971 to 2000 were calculated for each station. The daily anomalies were calculated as the difference of each daily temperature from its daily normal values.

The present interpolation method requires an understanding of the spatial correlation structure of the station data. Therefore, the inter-station correlations were calculated to determine the distances over which observed temperature anomalies are related. For each month, for each pair of 395 stations lying within 2000 km, correlation, r, was calculated and then binned according to their separation over intervals of 100 km.The mean correlation was estimated over each 100 km interval and a two-degree polynomial function was fitted to these values. For interpolating the station data, we have to define the radius of influence. This was estimated as the Correlation Length Scale (CLS), which is defined as the distance at which the mean correlation, represented by the fitted function, fell below 1/e (Belousov et al., 1971), where e =EXP (1). The results of the estimation of CLS values for maximum temperature are shown in Fig.2.

Angular distance weighting uses two components to calculate the weighting of each station. The first component weights the station according to its distance from a grid point, with the CLS controlling the rate at which the weight decreases away from the grid point. We chose the exponential function as a representation of the observed correlation decay curves produced.

Based upon the CLS, a correlation function can be defined (Jones et al., 1997) shown in equation (1), where *x* is the distance of the station from the required grid point and x_o is the CLS appropriate to that grid point depending on its latitude.

$$r = e^{-x/x_o} \qquad \textbf{(1)}$$

We define a distance weight for a station, *i* (New et al.(2000)), in equation (2).

$$w_i = r^m \qquad \textbf{(2)}$$

As in the work of New et al (2000), we also tested different values of *m* (ranging from 1 to10) and evaluated the results based upon cross validation against withheld station data. However, we found that the cross-validation root mean square (RMS) errors were comparatively less for m = 1 and the highest value of m and tended to increase with intermediate values of *m*. Even we repeated the process up to m equal to 25, but the same results persisted. Therefore we took the values of m as 1. Following New et al. (2000), the combined angular distance weight for the *i*th station (of a total of *k* stations contributing to a grid point value), W_i is defined as:

$$Wi = w_i \left\{ 1 + \frac{\sum_k w_k \left[1 - \cos(\theta_k - \theta_i)\right]}{\sum_k w_k} \right\}, \quad i \neq k \qquad \textbf{(3)}$$

Where, the position of the *i*th station is defined in terms of its distance, x_i (equation (1)) and its angle to North θ_i, relative to the specified grid point.

To select the stations that will contribute to each grid point value, stations lying in the radius up to the CLS were selected. As was done by Piper and Stewart (1996) and New et al. (2000), we also used a variable search radius to include, respectively, the closest 4 to 10 stations to a grid point. We further, used the weighted sum of the closest 4 to 10 stations to each grid point, within the CLS distance, to estimate the grid point temperature values. If there are more than 10 stations within the CLS distance, then only the 10 closest to the grid point having maximum correlations were used.

To create the absolute temperature gridded data set, we first developed the gridded data fields of daily normals (1971-2000) using the same technique, and added to the respective gridded anomaly values.

4. Evaluation of the data set

We evaluated the temperature data set using the cross validation (Cressie, 1993) to estimate the errors associated with the interpolation technique used in the study. This was done by removing each station from the data set, and then using the interpolation technique adopted in the study to estimate the temperature anomaly time series for that station using data from the surrounding stations. Root Mean Square Errors (RMSE) on the basis of the differences between the actual station time series and the interpolated station time series were computed. Results for the annual average RMS error for maximum and minimum temperatures are shown in Fig. 3(a) and 3(b). On an average, the RMSE values are generally less than 0.5^0C over most parts of the country. Errors are relatively larger in the hilly areas of Jammu Kashmir and Uttarakhand for the obvious reason of height and data scarcity in that area.

The quality of the present data set was compared with the monthly mean temperature data set prepared by Cort Willmott & Kenji Matsuura of University of Delaware (www.cdc.noaa.gov). This data set contains monthly temperature data for 0.5^0 Lat X 0.5^0 Long resolution for the period 1951-1999. With the daily data set developed by us, monthly/annual mean temperature values for the common period 1969-1999 were prepared and correlation coefficients between the two data sets were computed. Correlation coefficients between the two sets are of the order of 0.8 at most of the grids as shown in Fig 4. a and b.

5. Analysis of the data set

The present high resolution temperature data set (1969-2005) can be used for many regional applications like analysis of extreme events (for mean and trends of heat and cold waves), studies related to monsoon variability and prediction, validation of weather and climate model simulations and other environmental applications. Here, we present some results of the analyses using the data set.

Over India, heat waves are prominent extreme temperature events occurring during the pre-monsoon season (April to June). An extreme weather event becomes disaster when society and or ecosystems are unable to cope with it effectively. The deaths due to heat wave in Orissa (a state in east coast of India) in 1998 has been widely reported as one of the rare extreme epochs over the country resulting in deaths of nearly 1300, out of which 650 were from Orissa only (De et al. 2005). Using the daily data set developed, we have prepared the climatology of heat and cold waves over the country. A heat wave is defined if the maximum temperature at a grid point is 3^0C or more than the normal temperature, consecutively for 3 days or more. Similarly, a cold wave is defined if the minimum temperature at a grid point is below the normal temperature by 3^0C or more, consecutively for 3 days of more. For heat wave events, the summer period, April to June was considered and for the cold wave events, the winter season, December to February is considered. Fig.5a and 5b show the mean frequency of cold waves and heat waves respectively over the country. Both the cold and heat waves are more frequent over the Indo-Gangetic plains of India. On an average, 5-6 heat wave events and 2-3 cold wave events occur every year over the northern parts of the country. As an example, average maximum temperature anomalies associated with a severe heat wave occurred during 1-7 June 1995 are shown in 6 a. Mean temperature anomalies were more than 4^0C over central India. Similarly, average minimum temperature anomalies associated with a severe cold wave during 10-16 February 1972 are shown in Fig.6.b. Over NW India, minimum temperatures were below normal by more than 4^0C.

During the monsoon season, there are significant intra-seasonal rainfall associations over the plains of India, known as active and break periods. Recently, Rajeevan et al (2008) have proposed criteria of identifying the active and break periods based on the standardized daily rainfall anomaly averaged over the monsoon zone. It is well known that during the break periods, rainfall activity is significantly reduced over the country except over SE and NE India. Due to reduced rainfall activity and cloudiness, temperatures rise over the country. In order to examine the temperature anomalies associated with the monsoon breaks, average temperature anomalies associated with 9 break events in July and 14 break events in August have been calculated. The results are shown in Fig. 7 a and b respectively. Over the monsoon zone, average temperature anomaly during the breaks is more than +3^{0}C in July and +1.5^0 C in August. However, temperatures are below normal over NE India and southeastern parts of the country.

Conclusions

The present daily temperature data set was developed using quality controlled 395 stations and a well proven interpolation technique. The quality of the data set was evaluated using a cross validation technique. The root mean square errors were less than 0.5^{0}C over most parts of the country. The data set has been found useful for various regional applications, like studies of extreme temperatures, validation of numerical and climate model simulations and many environmental applications. In this study, we have used the data to calculate the mean frequency of cold and heat waves over the country. However, the present data set can be used for many more applications and such case studies will be reported later. However, one negative aspect of the present data set is that it starts only from 1969. This is because the digitized data are available in the IMD archives only from 1969. The data prior to 1969 are still in manuscript form. IMD is however undertaking a data rescue project in which all the data prior to 1969 will be compiled and digitized for archival. Once the project is completed by 2009, the present gridded data set will be updated with the data prior to 1969. However, in the mean time, the present data set can be used for many applications and the data set will be made available to the research community from the National Data Centre, IMD, Pune.

Acknowledgements

We are thankful to the Director General of Meteorology, Addl. Director General of Meteorology (Research) and Dy Director General of Meteorology (Climatology) for their encouragements and kind support to complete this project.

References

Caesar, J., L. Alexandar and Russell Vose (2006), Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set,J. of Geophys. Res., 111, DO5101, doi:10.1029/2005JD006280.

Cressie, N. (1993), Statistics for Spatial Data, John Wiley, Hoboken, N.J., pp.900.

De, U.S., Dube, R.K. and Prakasa Rao, G.S. (2005), Extreme weather events over India in the last 100 years, J. Ind. Geophys. Union, 9, 173-187.

Feng, S., Q. Hu, and W. Qian (2004), Quality control of daily meteorological data in China, 1951-2000: A new dataset, Int. J. Climatol, 24, 853-870.

Gleason, B. (2002) Global daily climatology network, VI.0, Natl. Clim. Data Cent., Asheville, N.C.

IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., et a. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996pp.

Janowiak, J.E., G.D. Bell and M. Chelliah (1999), A gridded data base of daily temperature maxima and minima for the conterminous United States : 1948-1993, NCEP/CPC Atlas 6, Clim, Predict Cent., Natl. Cent. for Environ. Perdict, Camp Springs, Md.

Jones, P.D., T.J. Osborn and K.R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, J. Clim., 10, 2548-2568.

Jones, P.D. and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, J. Clim, 16, 206-223.

Kiktev, D., D.M.H. Sexton, L. Alexander and C.K. Folland (2003), Comparison of modeled and observed trends in indices of daily climate extremes, J. Clim., 16, 3560-3571.

New, M.G., M. Hulme and P.D. Jones (2000), Representing twentieth-century space-time climate variability, part II:Development of 1901-96 monthly grids of terrestrial surface climate, J. Clim., 13, 2217-2238.

Piper, S.C. and E.F. Stewart (1996), A gridded global data set of daily temperature and precipitation for terrestrial biospheric modeling, Global Biogeochem. Cycles, 10, 752-782.

Rajeevan, M., Sulochana Gadgil, and Jyoti Bhate, 2008, Active and Break Spells of the Indian monsoon, NCC Research Report No 7/2008, National Climate Centre, India Meteorological Department (available at http:/www.imdpune.gov.in)

Sellers, W.D. and Wen Liu, (1988), “Temperature patterns and trends in the upper troposphere and lower troposphere”, J. of Climate, 1, 573-581.

Shepard, D. (1968), A two-dimensional interpolation function for irregularly spaced data, paper presented at 23rd National Conference, Assoc. for Comput. Mach., New York.

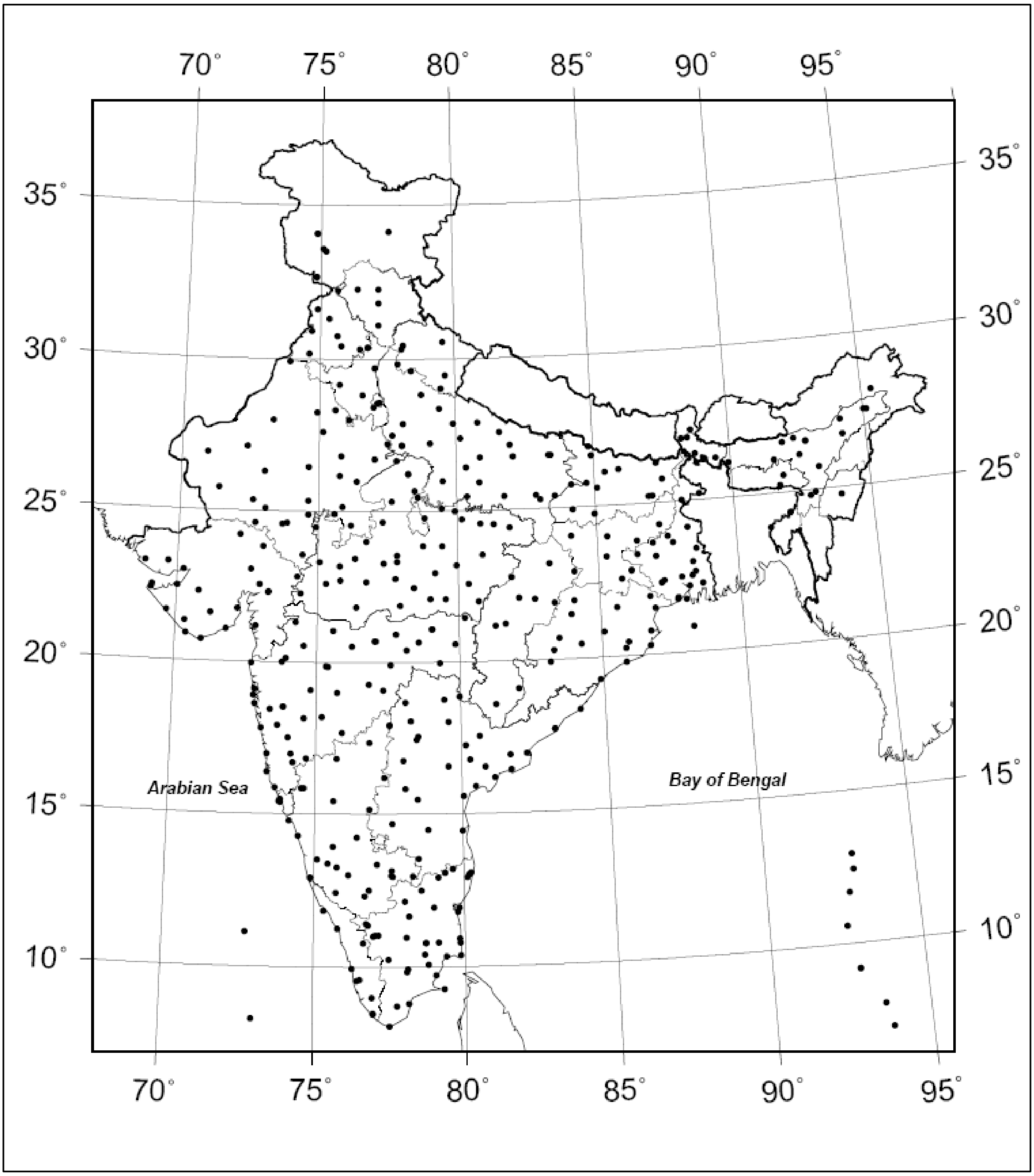


Fig: 1 : Network of 395 stations selected for developing the gridded data set

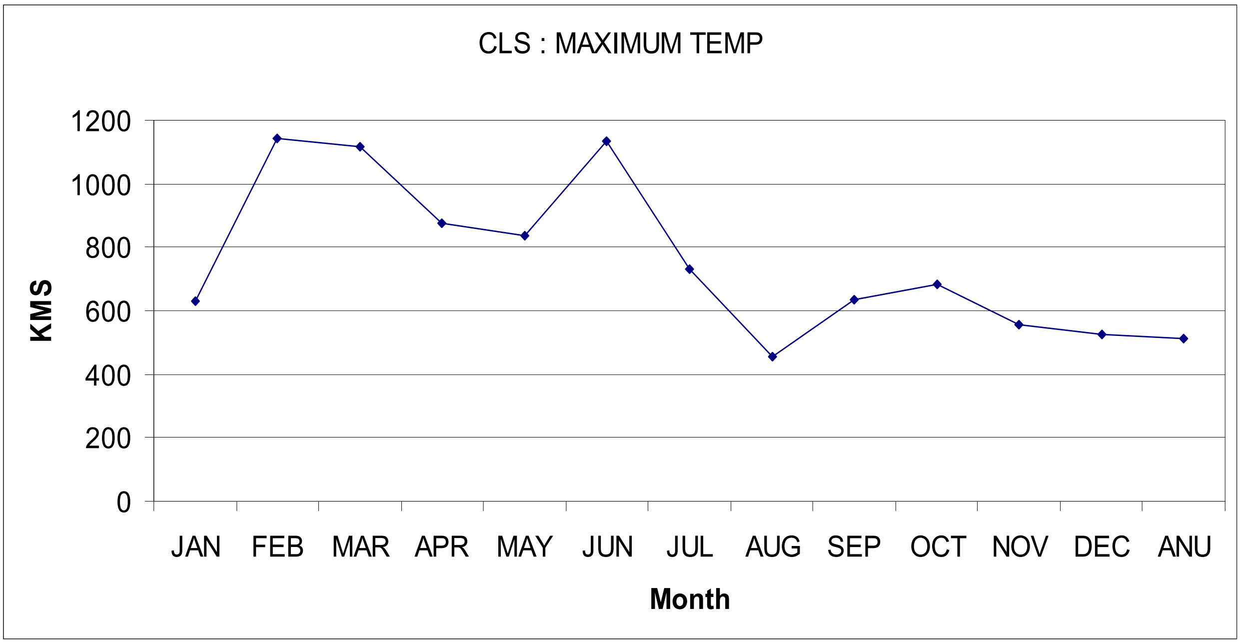
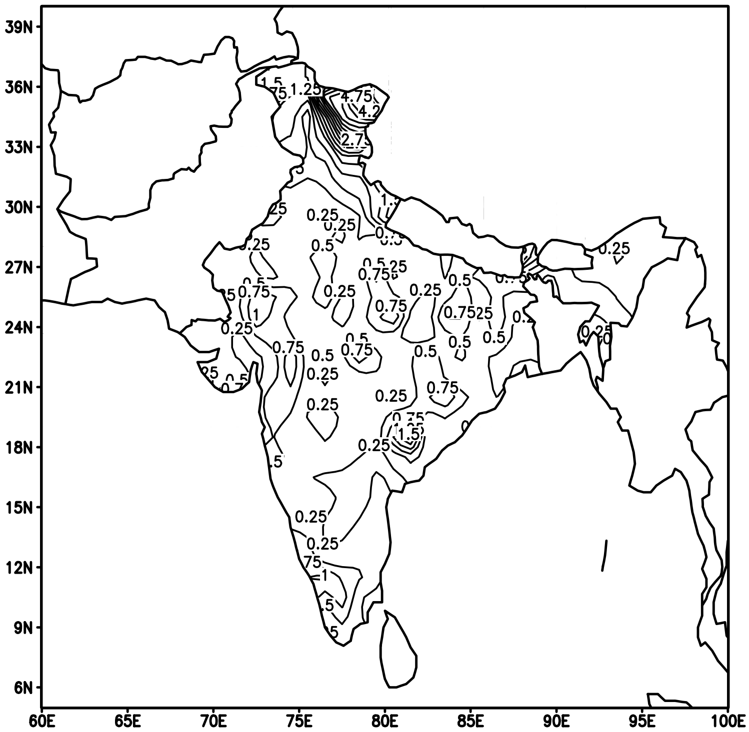


Fig: 2 : **Correlation Length Scale (CLS) for all the months for maximum temperature**

(a)



(b)

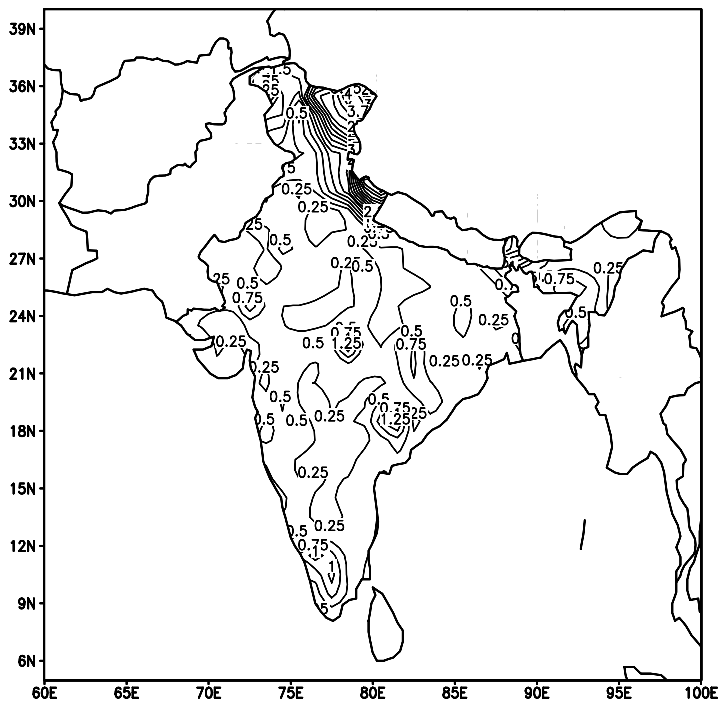


Fig: 3 : **Annually averaged RMSE for**
(a) maximum and (b) minimum temperatures

(a)

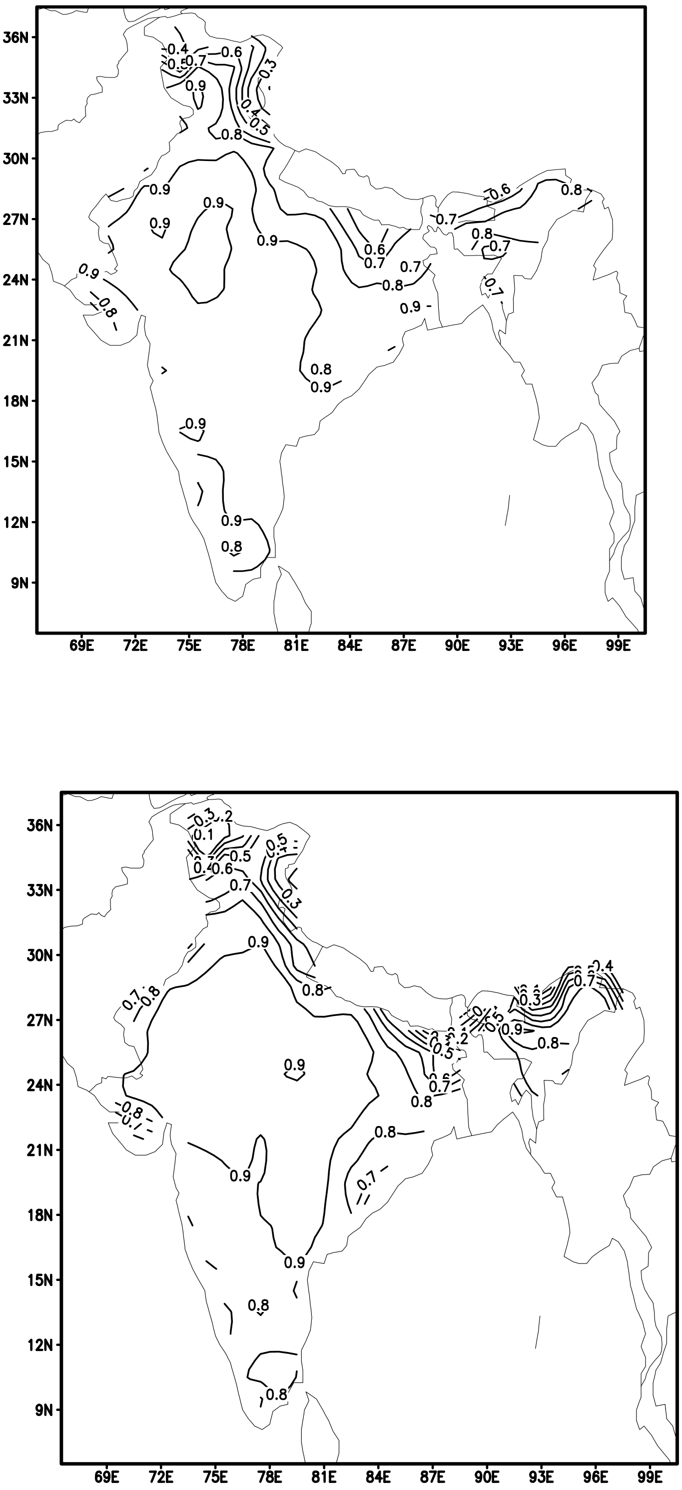


Fig.4 : Correlation coefficients between temperature data set of University of Delaware and the data set prepared in this study; a) for January and b) for July, months for 1969-1999.

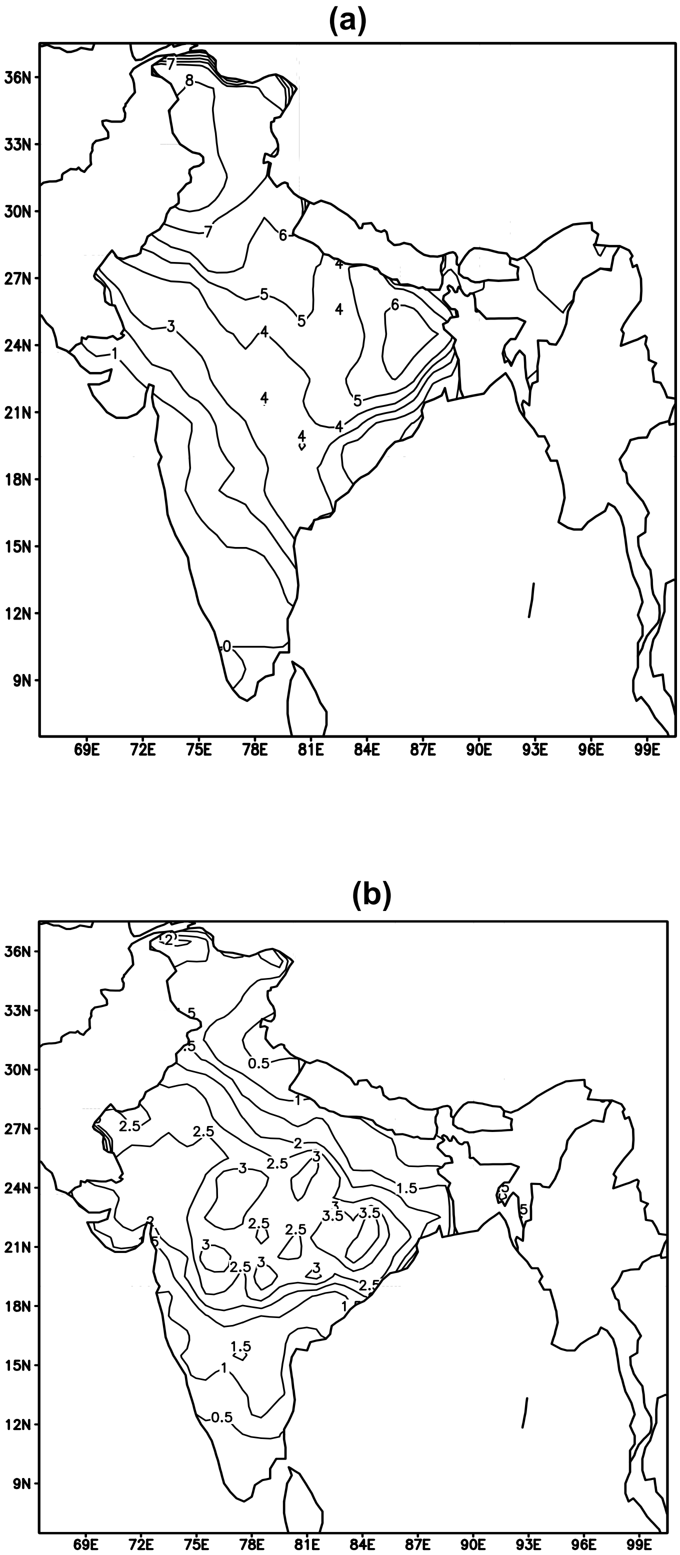


Fig.5 : Mean Frequency of a) Heat waves (April-June) and b) Cold waves (December to February) over the country. Period 1969-2005.

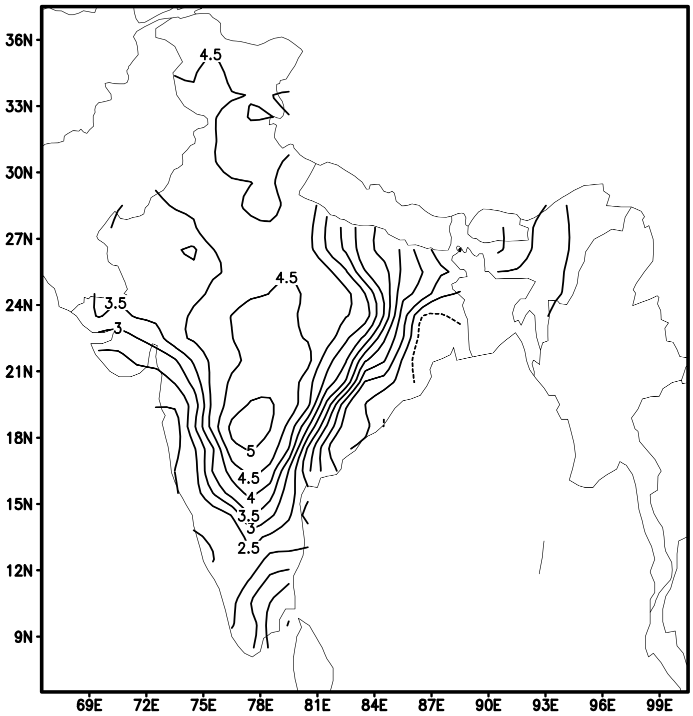


Fig 6(a) : Maximum temperature anomalies in heat wave condition during 1-7 June 1995.

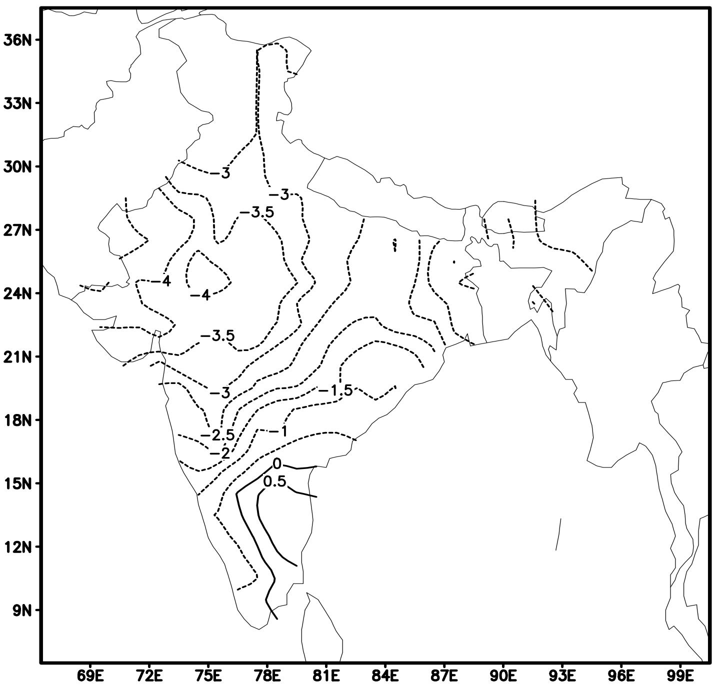
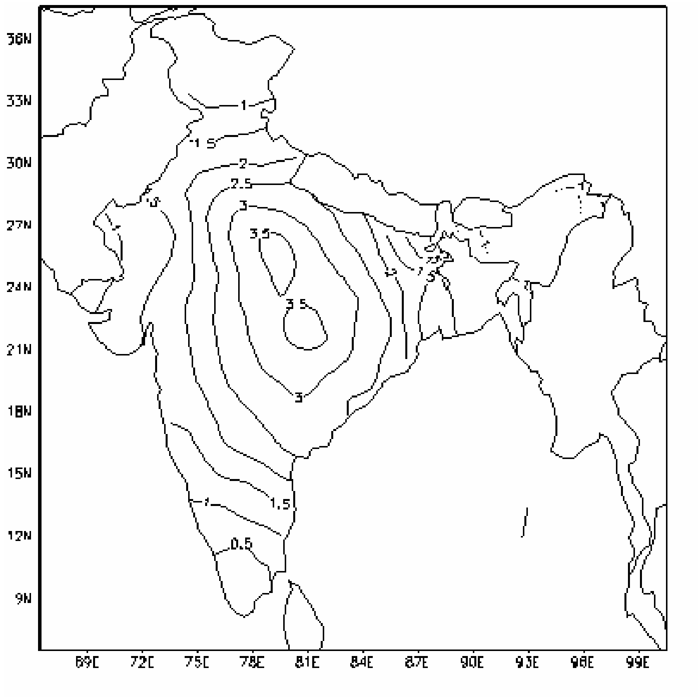


Fig 6 (b) : Minimum temperature anomalies in cold wave condition during 10-16 February 1972.

(a)



(b)

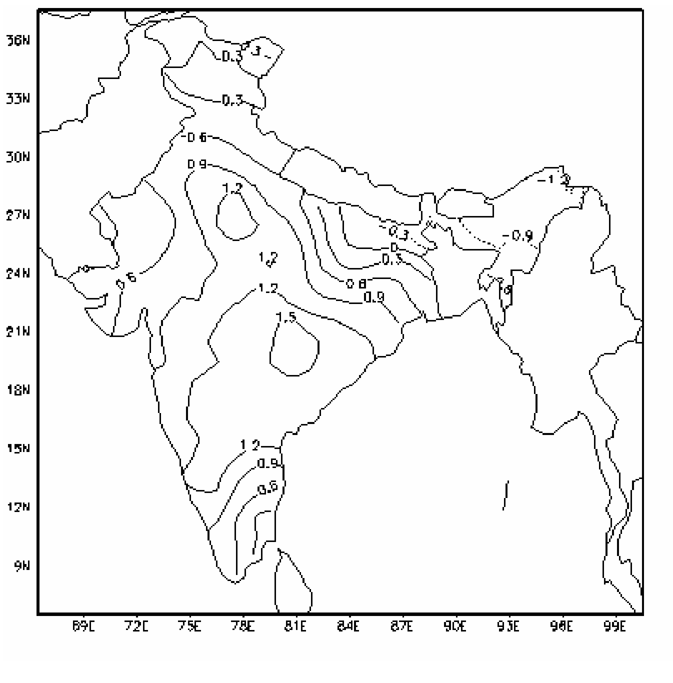


Fig. 7 : Average maximum temperature anomalies for monsoon break days occurring in, (a) July (9 occasions) and (b) August (14 occasions).