RMetS

Royal Meteorological Society

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 411 005, India

*Correspondence to:
M. Rajeevan, National
Atmospheric Research
Laboratory, SV University
Campus, P.O. Box 123,
Tirupati-517 502. India.
E-mail: rajeevan@narl.gov.in*

Abstract

A high resolution daily gridded temperature data set for the Indian region was developed using temperature data of 395 quality controlled stations for the period 1969–2005. A modified version of the Shepard's angular distance weighting algorithm was used for interpolating the station temperature data into 1° latitude × 1° longitude grids. Using the cross validation, errors were estimated and found less than 0.5 °C. The data set was also compared with another high resolution data set and found comparable. Mean frequency of cold and heat waves, temperature anomalies associated with the monsoon breaks have been presented. Copyright © 2009 Royal Meteorological Society

Keywords: Gridded temperature data; Correlation Length Scale (CLS); angular distance; heat and cold wave frequency

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 10 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 United States(Janowiak *et al.*, 1999; 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° latitude × 3.75° 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, at present, daily data from 1969 onwards only are digitized and archived at NDC. 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 1971–2000, 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 $\pm$ (1.76 + 0.8 N) standard deviation, (Sellers and Liu, 1988), for further data analysis. In fact, Sellers and Liu (1988) developed this filter for rejecting doubtful values of mean monthly data set. In this study, the filter was used for the daily data set for marking unusual high values. For 30 year period, the filter flags values beyond the range, mean $\pm$ 4.16 × SD, and it was considered that those values require further scrutiny. The flagged values after the above checks were further examined for spatial continuity. Values showing spatial continuity were accepted and only the isolated values were rejected. 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 Figure 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 Caesar *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 1971–2000 were calculated for each station. The daily anomalies
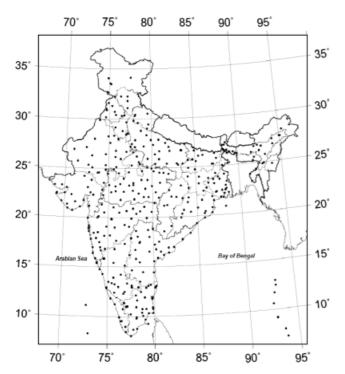

Figure 1. The network of 395 stations selected for preparing 1° × 1° gridded data.

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 interstation 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 $e \approx 2.72$. The results of the estimation of CLS values for maximum temperature are shown in Figure 2. The CLS values for each month are shown in Figure 2. It may be seen that the value of CLS reduces during April to May, but increases sharply in June and again reduces to its lowest value in August. As far as varying 'CLS' is concerned, we do not have the clear answers. Probably, 'CLS' is dependent on the homogeneity of thermal structure of the atmosphere. It may be mentioned that during April and May, extreme northern/northwestern and eastern parts of the country and Kerala receive rainfall due to air masses having different characteristics. Similarly, during August frequent break like conditions prevail due to intrusion of extra tropical westerly activities (air mass). This type of intrusions may lead to lower 'CLS'.

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.

On the basis of the CLS, a correlation function can be defined (Jones *et al*., 1997) as shown in Equation (1), where x is the distance of the station from the required grid point and x_o is the CLS appropriate
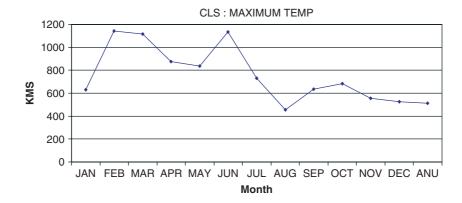

Figure 2. Correlation length scale (CLS) for all the months for maximum temperature.

to that grid point depending on its latitude.

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

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

$$w_i = r^m \quad (2)$$

As in the work of New *et al.* (2000), we also tested different values of m (ranging from 1 to 10) 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 ith station (of a total of k stations contributing to a grid point value), W_i is defined as:

$$W_i = w_i \left\{ 1 + \frac{\sum\limits_{k} w_i [1 - \cos(\theta_k - \theta_i)]}{\sum\limits_{k} w_i} \right\} \quad i \neq k$$

$$(3)$$

where, $w_i = e^{-x/x_0}$ and the position of the ith 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 ten stations within the CLS distance, then only the ten closest to the grid point having maximum correlations were used. It may be mentioned that for almost all the grid points, more than eight stations are available, except for the Jammu and Kashmir and northeastern regions.
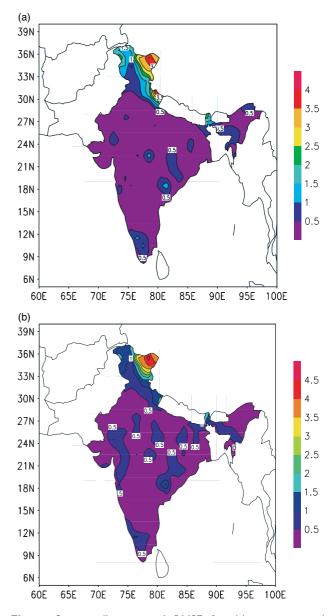
To create the absolute temperature gridded data set, we first developed the gridded data fields of daily normals (1971–2000) using the same technique, and added to the respective gridded anomaly values.
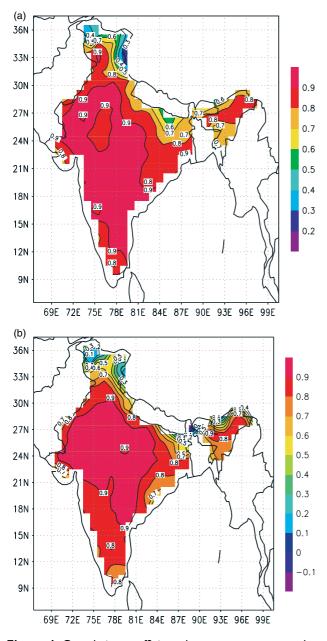
4. Evaluation of the data set

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 (RMSEs) 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 RMSE for maximum and minimum temperatures are shown in Figure 3(a) and (b). On an average, the RMSE values are generally less than 0.5 °C over most parts of the country. Errors are relatively larger in the hilly areas of Jammu and 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 and Kenji Matsuura of University of Delaware (www.cdc.noaa.gov). This data set contains

Figure 3. annually averaged RMSE for (a) maximum and (b) minimum temperatures.

Figure 4. Correlation coefficients between temperature data set of University of Delaware and the data set prepared in this study (a) for January and (b) July for the period 1969–1999.

monthly temperature data for 0.5° latitude × 0.5° longitude 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 Figure 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 (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 analysis using the temperature gridded 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 only from Orissa (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 °C 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 °C or more, consecutively for 3 days or 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. 5(a) and (b) show the mean frequency of heat waves and cold 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.
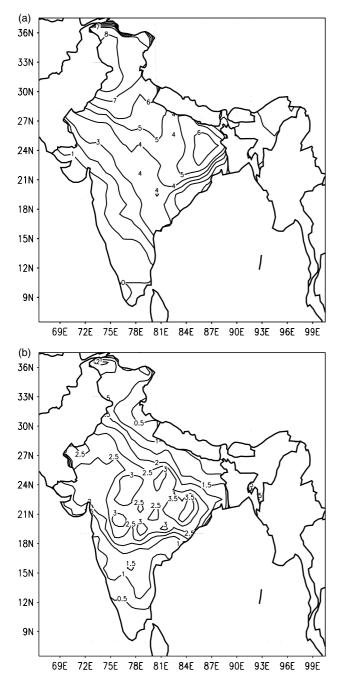
During the monsoon season, there are significant intra-seasonal rainfall variations over the plains of India, known as active and break periods. Recently, Rajeevan *et al.* (2008) have proposed criteria of identifying the active and break periods based on standardized daily rainfall averaged over the monsoon zone. It is well known that during the break periods, rainfall activity is significantly reduced over the country expect over SE and NE India. Owing to reduced rainfall activity and cloudiness, temperature 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 event in August have been calculated. The results are shown in Figure 6(a) and (b), respectively. Over the monsoon zone, average temperature anomaly during the breaks is more than +3 °C in July and +1.5 °C in August. However, temperatures are below normal over NE India and southeastern parts of the country.
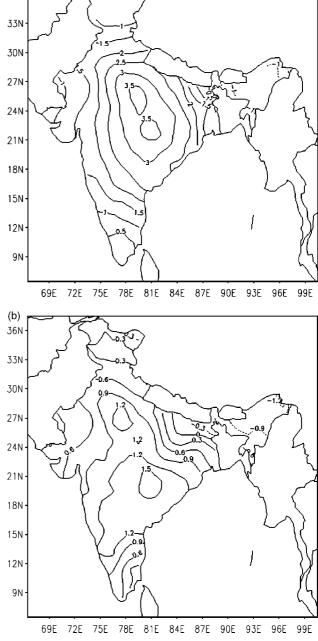
6. Conclusions

The present daily temperature data set was developed using quality controlled 395 stations and a well-proven interpolation technique. The quality of the

Figure 5. Mean frequency of (a) heat waves (April to June) over the country during the period 1969–2005. (b) Mean frequency of cold waves (December to February) over the country during the period 1969–2005.

Figure 6. Average maximum temperature anomalies for monsoon break days occurring (a) in July (9 occasions) and (b) in August (14 occasions).

data set was evaluated using a cross-validation technique. The RMSE were less than 0.5 °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, 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 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.

References

Belousov SL, Gandin LS, Mashovich SA. 1971. *Computer Processing of Meteorological Data*, Isr. Program for Sci. Transl., Jerusalem; 210.

Caesar J, Alexandar L, Vose Russell. 2006. Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded data set. *Journal of Geophysical. Research* **111**: DO5101. DOI:10.1029/2005JD006280.

Cressie N. 1993. *Statistics for Spatial Data*. John Wiley: Hoboken; 900.

De US, Dube RK, Prakasa Rao GS. 2005. Extreme weather events over India in the last 100 years. *The Journal of Indian Geophysical Union* **9**: 173–187.

Feng S, Hu Q, Qian W. 2004. Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology* **24**: 853–870.

Gleason B. 2002. *Global Daily Climatology Network*, VI.0. National Climatic Data Center: Asheville.

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 al.* (eds.). Cambridge University Press: Cambridge,New York; 996.

Janowiak JE, Bell GD, Chelliah M. 1999. *A gridded data base of daily temperature maxima and minima for the conterminous United States : 1948–1993*, NCEP/CPC Atlas 6. Climate Prediction Center, National Centers for Environmental Prediction: Camp Springs.

Jones PD, Osborn TJ, Briffa KR. 1997. Estimating sampling errors in large-scale temperature averages. *Journal of Climate* **10**: 2548–2568.

Jones PD, Moberg A. 2003. Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *Journal of Climate* **16**: 206–223.

Kiktev D, Sexton DMH, Alexander L, Folland CK. 2003. Comparison of modeled and observed trends in indices of daily climate extremes. *Journal of Climate* **16**: 3560–3571.

New MG, Hulme M, Jones PD. 2000. Representing twentieth-century space-time climate variability, part II: development of 1901–96 monthly grids of terrestrial surface climate. *Journal of Climate* **13**: 2217–2238.

Piper SC, Stewart EF. 1996. A gridded global data set of daily temperature and precipitation for terrestrial biospheric modeling. *Global Biogeochem. Cycles* **10**: 752–782.

Rajeevan M, Gadgil S, Bhate J. 2008. Active and break spells of the Indian summer monsoon. *NCC Research Report No.7/2008*. India Meteorological Department: Pune.

Sellers WD, Liu Wen. 1988. Temperature patterns and trends in the upper troposphere and lower troposphere. *Journal of Climate* **1**: 573–581.

Shepard D. 1968. A two-dimensional interpolation function for irregularly spaced data. Paper presented at *23rd National Conference*, Association for Computing Machinery: New York.