*Vol. 18, No. 2, pp. 218-223 (2018)*
**Journal of Agricultural Physics**
**ISSN 0973-032X**
**http://www.agrophysics.in**

**Research Article**

# Comparison of Gridded Weather Datasets with Point Datasets over Three Locations of India

DEBASISH CHAKRABORTY*, VINAY KUMAR SEHGAL, RAJKUMAR DHAKAR AND DEB KUMAR DAS

*Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi-110012*

**ABSTRACT**

Climate information of any place is of prime importance for understanding the dynamics of natural resources in general and agriculture in particular. But availability of good quality, long-term weather data over large region remains a challenge. Hence, in this study a comparative analysis of temperature (maximum and minimum temperature) and rainfall was conducted between the manual observatory datasets with that of the gridded datasets of India Meteorological Department (IMD) over three places of Indo-Gangetic plain of India. In this analysis the mean monthly values of the parameters as well as the daily extremes were compared. The agro-meteorological stations data used in this study are from three districts namely Ludhiana, Hisar and Kanpur over north India. Results showed that, in case of monthly based comparison, the gridded datasets could explain almost all the variability ($R^2$>0.95) in the manual observatory/point datasets while could explain about 65% to 86% variability in the rainfall amount. The higher values of rainfall were seen to be underestimated by the gridded dataset. The root mean square error (RMSE) for maximum temperature varied from 0.71°C to 0.84°C while for minimum temperature it was higher in the tune of 1.25°C to 1.36°C. For, daily values of extreme temperatures gridded data could also explain almost the whole range of variations in the extremes of station data of all the three locations as indicated by very high value of coefficient of determination ($R^2$). The nRMSE for maximum temperature varied between 2.1% to 2.5% while for minimum temperature the nRMSE ranged between 4.2% and 8.7%. Hence, it can be inferred that gridded temperature datasets may be directly used if meteorological station datasets are not available/accessible for any particular place.

**Key words:** Climate datasets, Gridded datasets, Extreme temperature, Climate change

## Introduction

Climate is the most important determining factor governing the distribution of different types of lives on earth. Agriculture is very closely dependent on the natural resources especially climate. Each and every operation of agriculture in general and crop production in particular is very snugly related to climate. Hence information about weather and climate are of utmost importance for different aspects of agriculture. Further, long–term analysis of trends for climate datasets such as temperature and rainfall data is required for study of climate change and its impacts (Piyoosh and Ghosh, 2016). But getting reliable and good quality weather data from meteorological observatories all over the country remains a challenge. Most of the observational and meteorological station data lack in spatio-temporal continuity in an area of interest for longer period (van Wart *et al*., 2015). Moreover, there is often limited and restricted access to such directly measured datasets (Hughes *et al*., 2009).

*Corresponding author,
Email: debasishagri@gmail.com

Hence, gridded weather datasets, which are prepared from interpolation of observed point datasets or satellite base rainfall products, can provide crucial weather information all over the country. Moreover, time series of gridded weather datasets are easily available from national weather agency like India Meteorological Department (IMD) along with various international agencies have made provision for open source freely accessible datasets for longer periods and these are being widely used. These datasets are spatio-temporally continuous and may even be useful for regions lacking in observational datasets or its proper accessibility. But these datasets needs to be compared with the observed point datasets from meteorological observatories for understanding the exact differences and biases which remains crucial for its application.

Several studies have been conducted to compare different meteorological datasets generated by different sources and applying different methodologies. Hughes *et al*. (2009) compared sea surface temperature (SST) data of three gridded products and time series from in-situ measurements, through calculation of annual mean anomalies and also carried out correlation analysis. Variable degree of correlation was found with different SST products and they could find greater variation (higher amplitude) in comparison to the in-situ data. Five different Chinese mean temperature series, including CRU TS3.10 and four others, were compared by Tang *et al*., (2010) on the basis of annual mean temperature anomalies, correlation and magnitude of temperature trend in each time-series. The anomalies differed widely before 1950; however, after 1951 they have shown close agreement with each other. The correlation coefficients found between any two of the five series range from 0.73 to 0.97. Wen *et al*. (2011) compared three global mean temperature data series of NCDC, HadCRUT3, and GISS. They used 10-yr averaged temperature anomalies and 100-yr warming trend of these datasets for inter-comparison and found that three series yielded similar trends of warming. While, Piyoosh and Ghosh (2016) carried out a comparative study of extreme temperature parameters from different sources by examining standardized anomalies, trends, correlation, and equivalence of datasets. Using the maximum temperature (Tmax) and minimum temperature (Tmin) for Dehradun, from two different sources of data from Climatic Research Unit (CRU) and observed data from India Meteorological Department (IMD) were used for 1901–2012. Their result showed that annual standardized anomalies of CRU data follow the similar pattern of annual standardized anomalies of IMD data. Moreover, the CRU data exhibit similar trends and are well correlated with IMD dataset. Hence, they recommended the use of high resolution CRU datasets as they are open source in nature.

In the present study, a comparative analysis of temperature (Tmax & Tmin) and rainfall was conducted between the manual observatory datasets with that of the gridded datasets of IMD over three places of Indo-Gangetic plain of India. The main objective of this study was to evaluate the gridded IMD datasets in comparison to the point manual observatory datasets for their possible usefulness over larger regions where either the observational records are lacking or it's not readily accessible.

## Material and Methods

In this study, the high resolution gridded daily temperature (1° × 1°) and gridded daily rainfall (0.25° × 0.25°) datasets developed by India Meteorological Department (IMD) was used (Srivastava *et al*., 2009; Pai *et al*., 2014). Maximum (day) and minimum (night) temperature from 395 quality controlled station was used by IMD for developing the data set (Srivastava *et al*., 2009). Interpolation of the station temperature data into 1° latitude × 1° longitude grids was carried out using a modified version of the Shepard's angular distance weighting algorithm. The developers of the dataset have compared it with high resolution datasets before successfully applying it for deriving temperature related parameters such as cold and heat waves, temperature anomalies over India (Srivastava *et al*., 2009). The errors due to interpolation in preparing the gridded data over the plains were found to be about 0.5°C at the maximum. It was relatively higher in the hilly regions of Jammu & Kashmir and Uttarakhand, mostly due to sparse sources of observational weather data availability in those zones.

The high spatial resolution (0.25° × 0.25°) daily rainfall dataset developed by IMD used daily rainfall records from 6955 rain gauge stations with varying availability periods. The number of stations used for the preparation of the new grid point data (IMD) is nearly 900 more than that used by Rajeevan *et al*., (2009) for the preparation of 0.5° × 0.5° gridded daily rainfall data. Out of these 6955 stations, 547 are IMD observatory stations, 494 are hydro-meteorology observatories and 74 are Agromet observatories. The remaining are rainfall reporting stations maintained by the State Governments. On an average, data from about 2600 stations per year were available for the preparation of daily grid point data. All the data sets were thoroughly quality checked and location verified before generating the gridded daily rainfall data sets over India using Shepard interpolation method for interpolating daily rainfall from rain gauge (sample) points to grid points. The directional effects and barriers were also included. Pai *et al*., (2014) observed that the large scale climatological and variability features of rainfall over India derived from the new data set were comparable with the existing datasets. In addition, the spatial rainfall distribution like heavy rainfall areas over the orographic regions of the west coast and northeast, low rainfall in the lee ward side of the western Ghats etc. were more realistic and better presented in new data set due to its higher spatial resolution and to the higher density of rainfall stations used for its development.

This study for comparison with the observed weather datasets of three locations viz. Ludhiana, Hisar and Kanpur The geographical co-ordinates and the duration of data corresponding to the three locations are given in table 1. The gridded temperature and rainfall dataset is available for 1951-2015 and 1901-2015, respectively. At first, the grids of temperature and rainfall datasets corresponding to the selected locations were extracted. The comparison was done in two ways, a) for the monthly values (mean for temperature and total for rainfall) and b) for extremes of temperature. In case of comparison on monthly scale we calculated the values for the last decade i.e. 2006-2015. While, in case of comparison of the extremes at first gridded data were selected for the same duration corresponding to the point location data and then from both the datasets for each day extreme was calculated. Here the 90$^{th}$ percentile was taken as the extreme value for each day. We have only considered the upper extreme of maximum and minimum temperature in this study.

All the analysis including the extraction of gridded datasets corresponding to the point datasets were carried out using open source R software and its IDE RStudio (R Core Team, 2017; R Studio Team, 2015). We have used the “Raster” package for handling all the raster datasets of temperature and rainfall datasets (Hijmans, 2015).

## Results and Discussion

The main weather variables namely maximum and minimum temperature and rainfall extracted from gridded datasets of IMD were compared with the observed datasets recorded at the agro-meteorological stations in three districts namely Ludhiana, Hisar and Kanpur over north India. In this analysis the mean monthly values of the parameters as well as the daily extremes were compared. The results of comparison of the mean

**Table 1.** The location details and duration of data for the three meteorological observatories.

| Location | Latitude | Longitude | Altitude (meter) | Data Duration |
|---|---|---|---|---|
| Ludhiana | 30°54’06.91’’ N | 75°48’27.98’’ E | 249 | 1970-2015 |
| Hisar | 29°09’00.49’’ N | 75°42’20.21’’ E | 216 | 1980-2015 |
| Kanpur | 26°29’27.79’’ N | 80°18’25.46’’ E | 131 | 1971-2015 |

monthly values of maximum and minimum temperature and monthly total rainfall over one decade (2006-2015) are presented in the figure 1. Results show that gridded data could explain almost the whole variation in the point based observation in the mean monthly maximum and minimum temperature, while for monthly rainfall there were dissimilarities across the locations. Overall gridded data could explain about 65% to 86% variability in the monthly rainfall across the locations. It can be seen that the higher monthly rainfall values were underestimated by the gridded rainfall data. In case of temperature, the root mean square error (RMSE) for maximum temperature varied from 0.71°C to 0.84°C which was about 2.3% to 2.7% of the mean (i.e. normalized root mean square error or nRMSE). In contrast, the RMSE for minimum temperature was quite higher in the tune of 1.25°C to 1.36°C with an nRMSE of 7.4% to 8.9%. Srivastava *et al*. (2009) reported that the errors due to interpolation in preparing the gridded temperature data over the plains were found to be about 0.5°C which is close to our observation for maximum temperature. The reason of difference may be either the locations or the comparison of only one station in each case. But, the results indicate that the differences in minimum temperature between the two datasets are higher as compared to that in maximum temperature.

The comparison of daily values of extreme temperatures is shown in the figure 2. Alike the case of the comparison of mean monthly values,

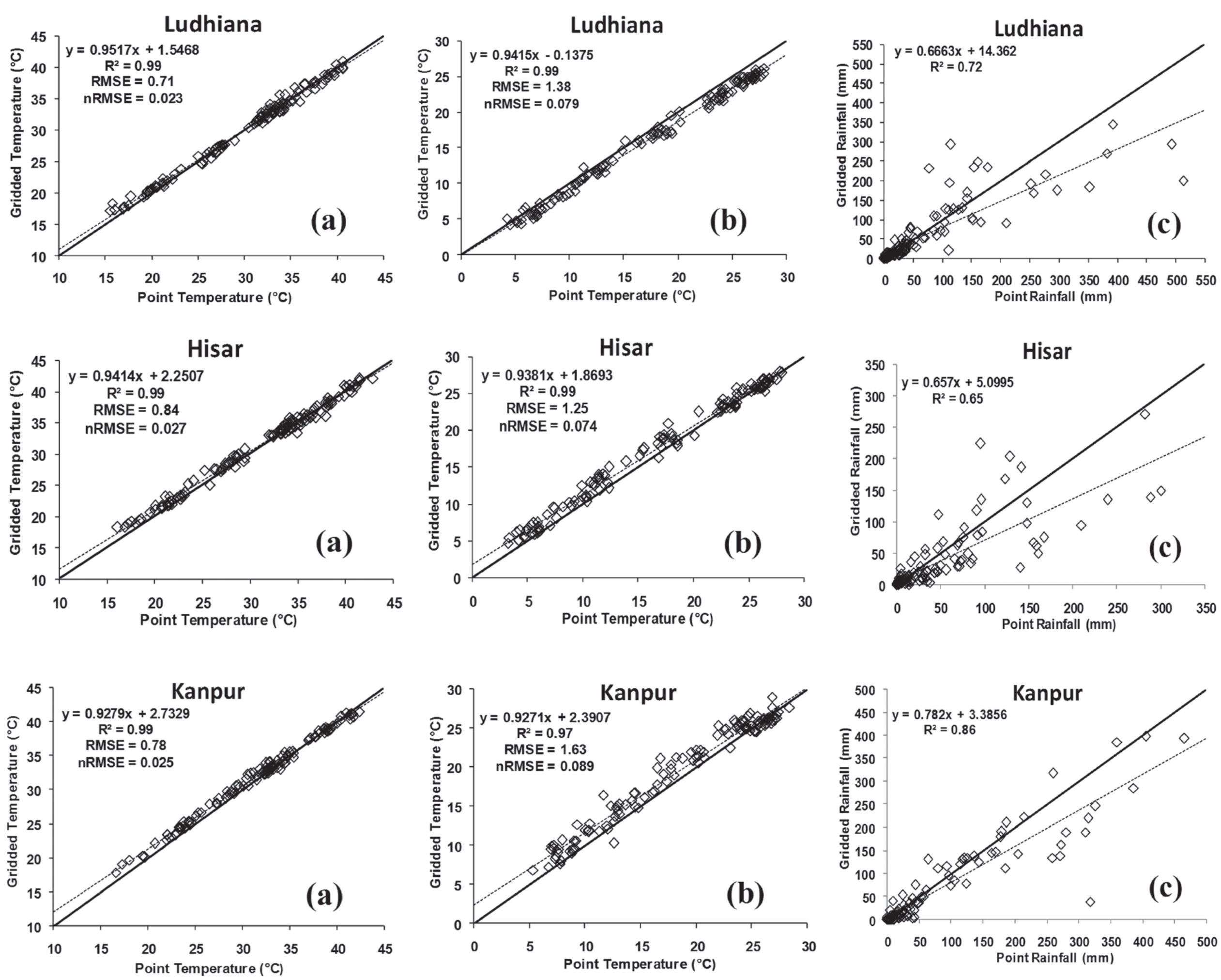


**Fig. 1.** Comparison of station (point) and gridded data of mean monthly values of maximum (a) and minimum temperature (b) and monthly total of rainfall (c) for one decade (2006-2015) over three districts of north India

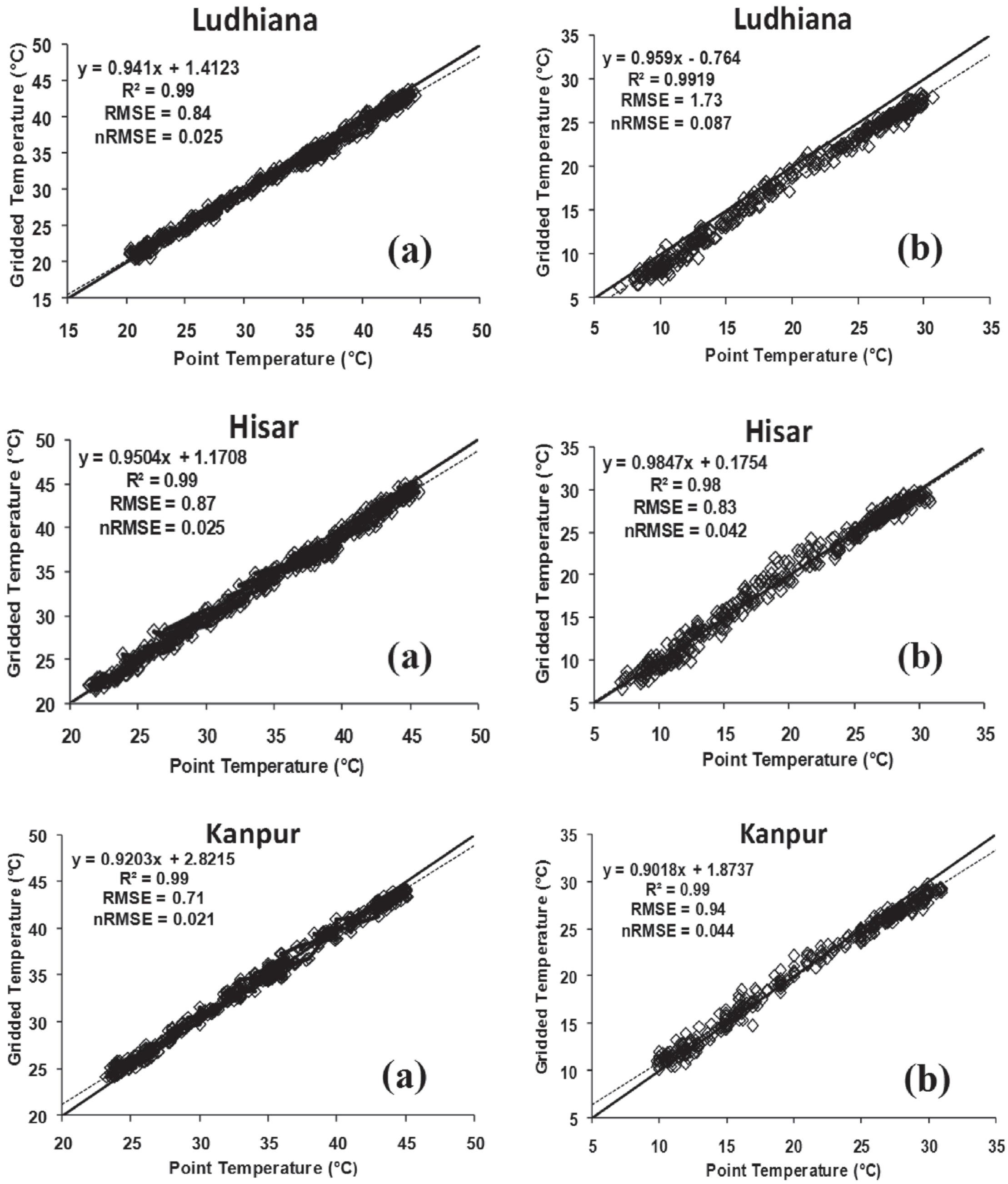


**Fig. 2.** Comparison of daily extreme values of maximum (a) and minimum temperature (b) of station (point) and gridded data over three districts of north India

gridded data could explain almost the whole range of variations in the extremes of station data of all the three locations as indicated by very high value of coefficient of determination ($R^2$). The RMSE for maximum temperature varied between 0.71°C to 0.87°C having nRMSE of 2.1% to 2.5%. In case of minimum temperature the RMSE of Hisar and Kanpur were 0.83°C and 0.94°C which were about 4.2% and 4.4% of the mean, respectively. But, for Ludhiana the RMSE of minimum temperature was about 1.73°C with nRMSE of about 8.7%, which is higher as compared to the other two locations. But it can easily be seen from the figure that for both the maximum and minimum temperature, lower values of extremes were very well picked up by the gridded data (proximity to the 1:1 line) while for higher extremes gridded data underestimated the values slightly which may be due to the interpolation which tend to smooth the peak values. Results of the study indicate that, for the maximum and minimum temperature, both the monthly mean as well as the extreme from gridded datasets were very close to that of the meteorological station datasets. Hence, gridded temperature datasets may be directly used if meteorological station datasets are not available for any particular place. While for rainfall the differences were higher for higher monthly total and also varied from location to location. The high inherent spatial variability in rainfall may be the reason behind it which needs more number of stations under each grid for proper comparison.

## Acknowledgements

First author acknowledges the fellowship provided by the Council for Scientific and Industrial Research (CSIR) during his Ph.D. programme and study leave granted by his employer, ICAR Research Complex for NEH Region, Umiam, Meghalaya. Authors acknowledge support received from IARI in-house Project Grant IARI:NRM:14:(04) and ICAR funded National Innovations in Climate Resilient Agriculture (NICRA) Project.

## References

Hijmans, R.J. 2015. *Geographic Data Analysis and Modeling* [R package raster version 2.6-7].

Hughes, S.L., Holliday, N.P., Colbourne, E., Ozhigin, V., Valdimarsson, H., Østerhus, S. and Wiltshire, K. 2009. Comparison of in situ time-series of temperature with gridded sea surface temperature datasets in the North Atlantic. *ICES J. Marine. Sci.* **66**(7): 1467-1479.

Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S. and Mukhopadhyay, B. 2014. Development of a new high spatial resolution (0.25× 0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* **65**(1), 1-18.

Piyoosh, A.K. and Ghosh, S.K. 2016. A comparative assessment of temperature data from different sources for Dehradun, Uttarakhand, India. *Journal of Meteorological Research* **30**(6): 1019-1032.

Rajeevan, M. and Bhate, J. 2009. A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies. *Curr. Sci.* **96**(4): 558-562.

Srivastava, A.K., Rajeevan, M. and Kshirsagar, S.R., 2009. Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. *Atmospheric Science Letters* **10**(4): 249-254.

Tang, G., Ding, Y., Wang, S., Ren, G., Liu, H. and Zhang, L. 2010. Comparative analysis of China surface air temperature series for the past 100 years. *Adv. Clim. Chan. Res.* **1**(1): 11-19.

Team, R. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA URL http://www. rstudio. com.

Team, R.C. 2017. R Core Team. R: a language and environment for statistical computing R Foundation for statistical computing, Vienna, Austria ISBN 3–900051–07–0.

Van Wart, J., Grassini, P., Yang, H., Claessens, L., Jarvis, A. and Cassman, K.G. 2015. Creating long-term weather data from thin air for crop simulation modeling. *Agric. For. Meteorol*. 209: 49-58.

Wen, X.Y., Tang, G.L., Wang, S.W. *et al*. 2011: Comparison of global mean temperature series. *Adv. Clim. Chan. Res.* **2**: 187-192, doi: 10.3724/SP.J. 1248.2011.00187.

Received: January 10, 2018; Accepted: July 12, 2018