

## Development of high spatial resolution weather data using daily meteorological observations over Indian region

ROHIT NAGORI and K. N. CHAUDHARI

*Space Applications Centre, ISRO, Ahmedabad – 380 015, India*

*(Received 28 January 2020, Accepted 27 May 2020)*

**e mail : rohitnagori@sac.isro.gov.in**

**सार** – फसल विकास गतिकी आदि से संबंधित अध्ययनों में उपग्रह से प्राप्त चित्रों के उपयोग को सक्षम करने के लिए सूक्ष्म स्थानिक विभेदन से प्रक्षेपित मौसम आंकड़े आवश्यक हैं। उपग्रह आंकड़े प्रतिदिन  $1 \times 1$  किमी या अधिकतम  $5 \times 5$  किमी ग्रिड में उपलब्ध हैं। मौसम आंकड़े को समान स्थानिक स्तर पर समय पर उपलब्ध कराने के लिए इस प्रक्रिया को भारत भर में स्थानिक रूप से प्रक्षेपित मौसम आंकड़े उत्पाद प्राप्त करने के लिए विकसित किया गया है। भारत मौसम विज्ञान विभाग की वेब साइट पर निश्चित समय पर उपलब्ध दैनिक मौसम आंकड़े (न्यूनतम और अधिकतम तापमान तथा वर्षा) का उपयोग इस अध्ययन में किया गया है। एक अर्ध-स्वचालित उपयोगकर्ता इंटरैक्टिव ग्राफिकल यूजर इंटरफेस (जीयूआई) विकसित किया गया है जो अनुपलब्ध आंकड़े सेटों को भरने के साथ-साथ तापमान आंकड़े सेट की गुणवत्ता की जांच करता है और साथ ही गलत आंकड़े को सही करने के लिए एक मंच प्रदान करता है जिसे पहचानने के लिए स्थानिक और कालिक असंगतता को ध्यान में रखते हुए सांख्यिकीय तरीकों का उपयोग किया गया है। तापमान पर उंचाई के प्रभाव की जानकारी प्राप्त करने के लिए पतली प्लेट्सलाइन इंटरपोलेशन तकनीक का उपयोग किया गया है जो गुणवत्ता परक मौसम आंकड़े और उंचाई की जानकारी CARTODEM आंकड़े से देता है तथा दैनिक स्थानिक रूप से प्रक्षेपित उत्पाद को चित्र के रूप में तैयार करता है। तीन अलग-अलग ऋतुओं यानी मॉनसून, ग्रीष्म और सर्दी में "जैक-नाइफ परीक्षण पद्धति" का उपयोग करके इसका वैधीकरण किया गया। सभी ऋतुओं के लिए दशकीय औसत उत्पादों की निरपेक्ष माध्य त्रुटियों में अंतर पाया गया जो अधिकतम तापमान के लिए 1.2-1.5 डिग्री सेल्सियस, न्यूनतम तापमान के लिए 1.1-1.7 डिग्री सेल्सियस, वर्षा के लिए 1.0-7.0 मिमी थी। जिसमें उच्चतर त्रुटि मॉनसून ऋतु के अधिकतम तापमान और वर्षा में तथा सर्दियों में न्यूनतम तापमान के लिए प्रेक्षित की गई। यह पाया गया कि 550 मीटर से कम उंचाई वाले स्टेशनों के लिए त्रुटियां 1 डिग्री सेल्सियस के करीब थीं, जबकि भारत के मध्य भाग में, माध्य निरपेक्ष त्रुटियां 1 डिग्री सेल्सियस से कम पाई गईं।

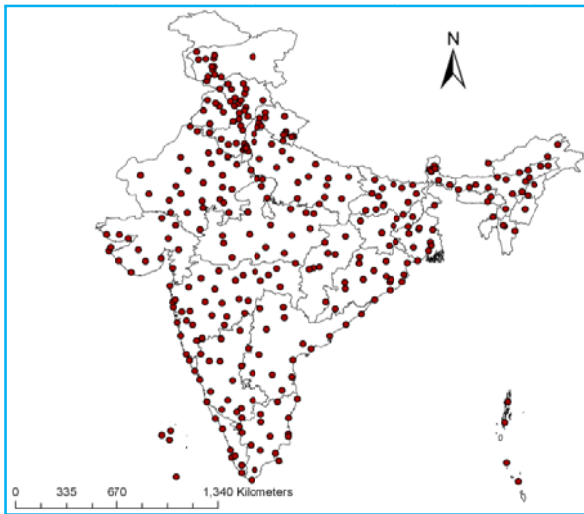
**ABSTRACT.** Finer spatial resolution interpolated weather data is essential to enable utilization of satellite-based images in studies related to crop growth dynamics, etc. Satellite data are available daily at  $1 \times 1$  km or at the most within  $5 \times 5$  km grid. To make the weather data timely available at the same spatial scale, the procedure has been developed to generate the spatially interpolated weather data product over India. Daily weather data (minimum & maximum temperature and rainfall) available at point scale on India Meteorology Department web site have been used in this study. A semi-automated user interactive Graphical User Interface (GUI) has been developed which quality checks the temperature data sets by filling the missing data sets as well as providing a platform to correct erroneous data which have been identified using statistical methods taking spatial as well as temporal incompatibility into account. Daily spatially interpolated product is generated in image form using thin plate spline interpolation technique that uses the quality checked weather data as well as elevation information from CARTODEM data in order to account for effect of elevation on temperature. The validation was performed using "Jack-knife testing method" for three different seasons *i.e.*, monsoon, summer and winter. The mean absolute errors for decadal averaged products were found to vary within 1.2-1.5 °C for maximum temperature, 1.1-1.7 °C for minimum temperature, 1.0-7.0 mm for rainfall considering all seasons with higher error observed in monsoon for maximum temperature and rainfall and in winter for minimum temperature. It was found that errors were close to 1 °C for stations with elevation less than 550 m whereas in central portion of India, mean absolute errors were found to be less than 1 °C.

**Key words** – Interpolation, Thin plate spline, Weather, Temperature and rainfall.

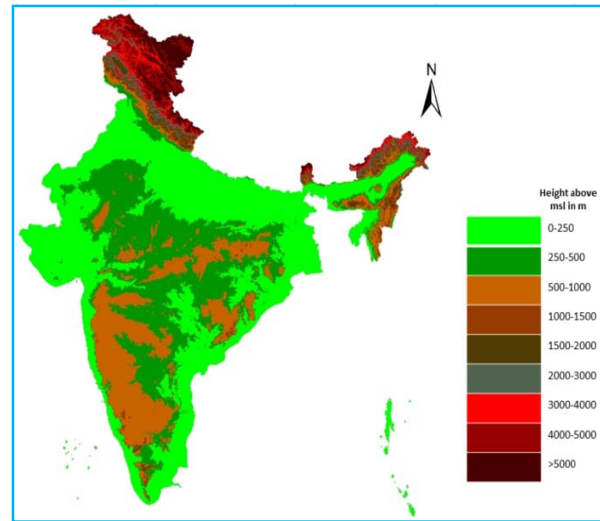
### 1. Introduction

Gridded surface climate data is important for studying the interactions of the climate with terrestrial,

hydrological and biogeochemical processes (Scholze *et al.*, 2006; Shen *et al.*, 2001). Daily spatial weather data of global solar radiation, maximum and minimum air temperatures and rainfall are much required by



**Fig. 1.** Map showing three hundred and thirty two (332) IMD ground stations stationed at various places in India



**Fig. 2.** Elevation map of India prepared using CARTODEM as utilized in our study

agricultural researchers in crop growth simulation models as important factors for spatial yield forecasting. To support Ministry of Agriculture (India), the Forecasting Agricultural output using Space, Agro-meteorology and Land based observations (FASAL) project was proposed under which a crop growth simulation model was developed which along with various other parameters requires daily meteorological observations (Chaudhari *et al.*, 2005). Various sources are available for weather data products such as satellite observations, meteorological ground stations, weather radar, etc. Satellites provide Land Surface Temperature (LST) in contrast to air temperature needed for simulation models. Indian geostationary satellites such as Kalpana-1 and INSAT-3D provide rainfall information but that is not error free as shown by Das and Singh (2014) using INSAT-3D 3 hourly IMR product ( $0.25^\circ$  spatial resolution) over Jammu and Kashmir region during the flood week of September, 2014 where large discrepancies were found in rainfall values provided by local weather stations (156.7 mm) and satellite data (60 mm). Meteorological ground stations provide point based information of weather parameters on daily basis, which form an integral part in the calibration of satellite observations too. Indian Meteorological Department (IMD) earlier provided weather data for 220 surface observatories on daily basis, which have been recently increased to 332. These point locations data need to be spatially interpolated in order to generate gridded data sets. The current availability of interpolated weather data is at very coarse resolution of about  $1^\circ \times 1^\circ$  or  $0.5^\circ \times 0.5^\circ$  for current weather or past climatic data in India (Srivastava *et al.*, 2009) while the other satellite data to be used in Crop Growth Monitoring System (CGMS) model are within  $5 \times 5$  km grid. So there

is a strong need to develop spatially gridded weather data product at high resolution in sync with satellite data.

The techniques available for spatial interpolation of point-based data include inverse-distance weighting, co-kriging, Thin plate splines, etc. (Myers, 1994). In our study, we have few stations to produce interpolated product over vast region of India and that too non-uniformly distributed. IDW fails or produces extreme errors in such conditions. There is non-availability of *a priori* information too whereas co-kriging results rely heavily on a previously well-chosen semi-variogram (Collins & Bolstad, 1996). The thin plate spline (TPS) interpolation on the other hand generates infinitely differentiable smooth surface and balances the fitted surface' smoothness and fidelity of data by minimizing the generalized cross-validation (Craven & Wahba, 1979). Numerous studies have been carried out in recent past for generating weather products from point observations using TPS interpolation (ex: Hong *et al.*, 2005). As elevation plays an important role in temperature of a place and can't be excluded from interpolation of point-based temperature observations (Dodson & Marks, 1997), it must be accounted for a country like India with heterogeneous terrain. Inclusion of temperature data requires dealing with missing observations, unreasonable or spatially incompatible readings, spurious values, etc. (Kotsiantis *et al.*, 2006) beforehand. Dealing with unreasonable or spatially incompatible readings due to human interventions is a major challenge for the datasets used in the study.

The first objective of the study is to fill the gaps in  $T_{\max}$  and  $T_{\min}$  data set available from IMD's stations. The

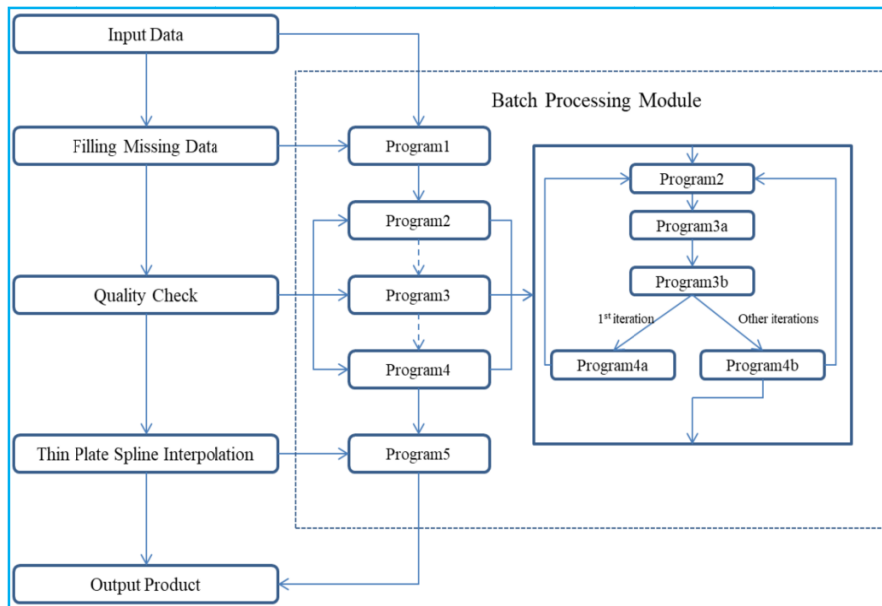


Fig. 3. Schematic of the process employed for interpolated product generation

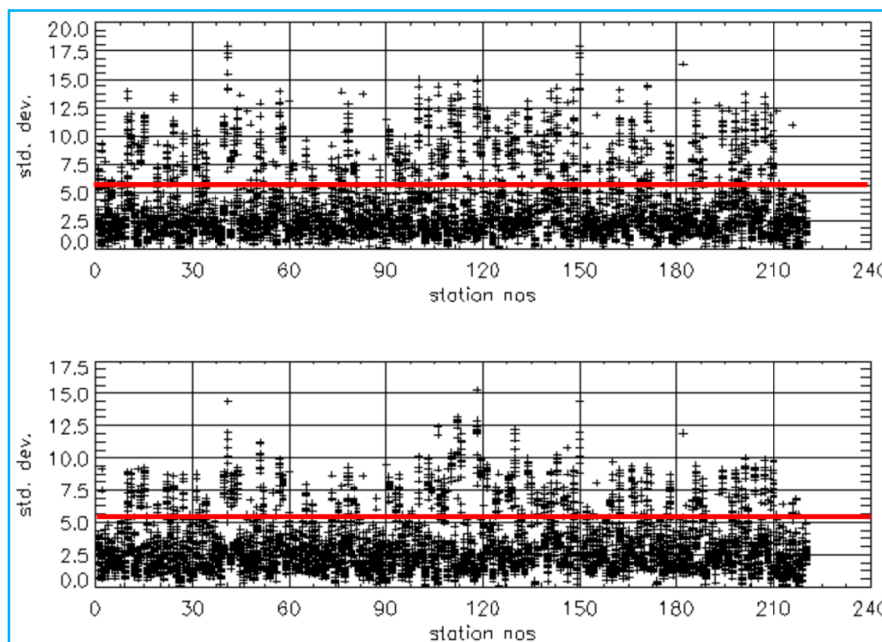


Fig. 4. Standard deviation (Graph) for each station in its 1° neighborhood for every day for  $T_{max}$  (above) and  $T_{min}$  (below) for all the stations. The data is of February, 2015 from IMD stations in study region and shows high  $\sigma$  values representing erroneous stations

second objective is to develop a semi-automated model using statistical parameters in order to find erroneous temperature data and then providing a platform to correct those errors there itself. The third objective is to generate the gridded data product for  $T_{max}$ ,  $T_{min}$  and Rain over Indian mainland using TPS method and subsequent

generation of solar radiation product. The fourth objective of the study is to validate the product generated, thus stating its reliability and qualitatively analysing the deviations in interpolated product from the actual observed values. The last objective is to develop a Graphical User Interface (GUI) to facilitate easier user interaction.

**TABLE 1**  
**Month-wise lapse rate magnitude for Indian mainland for both  $T_{\max}$  and  $T_{\min}$**

Month	Lapse Rate $T_{\max}$	Lapse Rate $T_{\min}$	Month	Lapse Rate $T_{\max}$	Lapse Rate $T_{\min}$
January	6.7	5.5	July	4.9	5.3
February	6.6	5.5	August	5.1	5.2
March	6.6	5.7	September	6.0	5.6
April	6.4	5.9	October	6.0	5.7
May	6.0	5.8	November	6.1	5.2
June	5.2	5.7	December	6.1	5.0

## 2. Methodology

### 2.1. Study area

The study region (Fig. 1) is Indian mainland with largely variable climatic conditions throughout. The region enclosed in the longitude range  $67^{\circ} 4' 45.75''$  E -  $97^{\circ} 51' 31.81''$  E and the latitude range  $7^{\circ} 7' 34.61''$  -  $37^{\circ} 35' 40.33''$  has been considered and divided in the grids of size  $5 \times 5$  km (masking out grids outside Indian region as -99). The total numbers of samples and lines are 594 and 673 respectively.

### 2.2. Data

IMD currently provides daily data for 332 surface observatories as shown in Fig. 1. For our study, three weather data parameters  $T_{\max}$ ,  $T_{\min}$  and rain, measured from these stations, have been considered. The data has been taken from Space Applications Centre intranet data archive website "IMD Weather Database" storing weather data as published by IMD on daily basis ([www.imd.gov.in](http://www.imd.gov.in)). The missing values in data are represented as -999. The CARTODEM data from <http://bhuvan.nrsc.gov.in> was used to generate median elevation data for each  $5 \times 5$  km grid in the study region (Fig. 2).

The station information file (as by IMD) is available too with latitude, longitude and elevation values for each station. Tripathy *et al.* (2008) generated solar radiation coefficients using empirical methods such as Angstrom-PreScott method, Supit method, Hargreave's formula and Bristow Campbell method for 20 stations around India in order to estimate amount of solar radiation received which has been utilized for radiation estimation part. A separate linear lapse rate table (Table 1) for minimum and maximum temperatures published by Rao and Prasad (1971) has also been utilized in our study. Interactive data

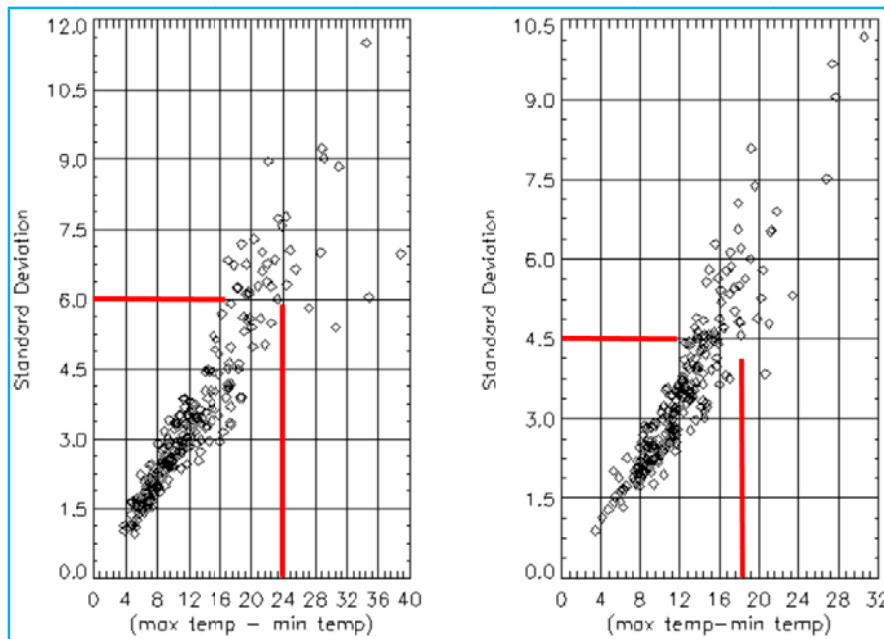
learning (IDL) programming platform supporting ENVI features has been used as the programming base for algorithm implementation.

### 2.3. Weather data cleaning

The schematic of the whole process being employed from input of weather data to the generation of interpolated product as the output has been shown in Fig. 3. There are primarily two steps to screen the weather data (temperature), *i.e.*, replacement of missing values or erroneous values. To handle missing data, methods available are ignoring instances with unknown feature values, most common feature value, mean substitution, regression, hot deck imputation, etc. (Kotsiantis *et al.*, 2006). The missing values are replaced using linear interpolation in temporal domain as temperature of a particular station does not vary drastically across consecutive days. In data set of our study, there are some stations that produce data irregularly which led us to undertake certain constraints while filling gaps. Let ' $n$ ' be the total number of stations, then ' $f$ ' represents the fraction of ' $n$ ' that produces data on frequent basis, gaps need to be filled only for a day where data available for less than fraction ' $f$ ' of total numbers of stations ' $n$ '. Also if a particular station reports data for more than fraction ' $t$ ' of total number of days ' $d$ ', then only it qualifies for gap filling as less than that represents a station producing irregular data. For our study, ' $f$ ' and ' $t$ ' are 0.84 and 0.6 respectively. For second objective, we took two primary conditions accounting for spatial and temporal variability in temperature data which has been brought down to mean sea level (m.s.l.) using lapse rate as shown in equation 1:

$$T(H_1) = T(H_2) + \text{L.R.} \times (H_2 - H_1) \quad (1)$$

where,  $T(H_1)$  and  $T(H_2)$  are temperature values at  $H_1$  (lower elevation, m.s.l.) and  $H_2$  (higher elevation) respectively and L.R. is Lapse Rate.



**Fig. 5.** Graph between  $\sigma$  and difference between the minimum and maximum temperature registered in February, 2015 for  $T_{\max}$  (left) and  $T_{\min}$  (right) for all the stations. The starting of winged structure marks the threshold

Around each station, a circle of certain radius is constructed and standard deviation ( $\sigma$ ) is calculated for the temperature data of the stations enclosed within that circle. A threshold on that  $\sigma$  (Fig. 4) helps us to identify the stations which either have erroneous value themselves or in their neighbourhood. For our study, the radius of circle is chosen as  $1^\circ$  ( $\sim 110$  km) so that circle around 85 percent of stations encompass at least 2 stations for calculation of statistical parameters. Also, the available IMD interpolated product using IDW (Srivastava *et al.*, 2009), which crucially depends on distances between stations, is of spatial resolution  $1^\circ$  making this radius favourable.

In second condition involving temporal variability, a winged structure is seen in the graph (Fig. 5) between the  $\sigma$  and the difference between the maximum and minimum temperature of that particular temperature data set ( $T_{\max}$  or  $T_{\min}$ ) plotted for each station. If a station has largely variable temperature, then  $\sigma$  will be higher in comparison to difference between maximum and minimum temperature. If a station has similar temperatures except some erroneous data, the difference between maximum and minimum temperature will be high but  $\sigma$  will be low. So, the starting point of the imparted wing like structure is taken as threshold for  $\sigma$  and difference between maximum and minimum temperature.

The thresholds to be provided by user are tolerant and should be error inclusive in nature. The stations having  $T_{\min}$  greater than  $T_{\max}$  are immediately identified as error and reported for correction. The intersection of the conditions 1 and 2 points out error after which many other filters are applied such as threshold on the maximum change in temperature between consecutive days and maximum deviation that can occur in data from the median value of temperature for a data set for a station. Hence, data having largest probability of error is presented to user for correction along with ancillary information such as latitude, longitude, elevation of the erroneous station,  $T_{\max}$ ,  $T_{\min}$  and rainfall values for preceding, present and successive day, the graph showing the temperature variation ( $T_{\max}$  and  $T_{\min}$ ) at that station and the nearest station. Other options such as ignoring the error data or replacing the error value using linear regression are also made available to user but for higher accuracy, manual method is recommended as most of the errors are due to human negligence such as swapping of  $T_{\min}$  and  $T_{\max}$  values, error in entering the first digit of data, etc. which can be rectified easily with careful and experienced deductions. For ex: If a station shows  $T_{\max}$  for previous, current and consecutive days as  $42^\circ\text{C}$ ,  $21^\circ\text{C}$ ,  $39^\circ\text{C}$  with current day value highlighted as erroneous and no significant variation in  $T_{\min}$  is observed. Also, if there is no trace of rainfall and no significant variation in temperatures of nearby stations, the careful deduction of

correct  $T_{\max}$  value for current day will be 41 °C which will be authentic in comparison to values estimated automatically or ignoring the instances (means losing data from already sparse dataset). The process is repeated iteratively till all the data is spatially and temporally compatible.

#### 2.4. Thin plate spline interpolation

The thin plate spline interpolation can be visualized as thin metal plate in which deflections are introduced in the direction perpendicular to plate at the positions where we have point based information. These deflections manifest the whole surface accordingly (combined gradient introduced by the deflection from all the point based data) thus giving interpolated value at each and every position (Hutchinson, 1993).

In  $n$  dimensions, the idea of thin-plate splines as presented in Meinguet (1979) is to choose a function  $f(x)$  that exactly interpolates the data points  $(x_i; y_i)$ , say,  $y_i = f(x_i)$  and that minimizes the bending energy,

$$E(f) = \int_{\mathbb{R}^n} |D^2 f|^2 dX \quad (2)$$

where,  $D^2 f$  is the matrix of second-order partial derivatives of  $f$  and  $|D^2 f|^2$  is the sum of squares of the matrix entries. The infinitesimal element of hyper-volume is  $dX = dx_1 \dots dx_n$ , where  $x_i$  is the component of  $X$  and  $\mathbb{R}^n$  denotes integral over  $n$ -dimensional real space. It is also possible to formulate the problem with a smoothing parameter for regularization (Wahba, 1990). A function  $f$  is chosen that does not necessarily exactly interpolate all the data points but that does minimize

$$E(f) = \sum_{i=1}^m |f(x_i) - y_i|^2 + \lambda \int_{\mathbb{R}^n} |D^2 f|^2 dX \quad (3)$$

The smoothing parameter is  $\lambda > 0$  and is chosen *a priori*. ' $m$ ' represents data points.

ANUSPLIN software by Hutchinson (2013) is used in large numbers of meteorological studies. Dodson & Marks (1997) in their study of interpolating daily maximum and minimum air temperature over a large mountainous region investigated methods to account for the effect of elevation on temperature and found linear lapse rate adjustment (LLRA, using constant lapse rate) as preferable. In their study, it emerged as the best method in terms of accuracy, speed and ease of implementation. We also implemented LLRA in our study in contrast to other studies where elevation is used as variable that has

smoothing effect over the nearby plain surfaces having low elevation. In India, only 18-22 among 45 solar radiation recording stations report solar radiation on regular basis. Hence, gridded product was developed by interpolating the solar radiation estimated at all stations available in our data using the regression coefficients of the nearest solar radiation recording station as calculated and discussed in detail by Tripathy *et al.* (2008). Due to availability of  $T_{\max}$  and  $T_{\min}$  data only, regression coefficients derived by Hargreaves (Hargreaves *et al.*, 1985) and Bristow-Campbell (Bristow & Campbell, 1984) method has to be used. Hargreaves model performed better at most of the stations but for stations where radiation has non-linear behaviour with temperature, Bristow-Campbell method was used. Das & Pujari (1993) method has been used for calculating solar radiation where equations from the book of Burman and Poshop (1994) have been utilized.

#### 2.5. Error determination

One of the crucial step is to validate the product generated, thus stating its reliability and qualitatively analyzing the deviations in interpolated product from the actual observed values. For a vast country like India with small numbers of stations in our study, "jack-knife testing" method or leave one out method has been chosen for validation. The method involves removing one observation point ( $j$ ) at a time from the whole data set to estimate a value of parameter at  $j$  from the remaining data points (Tomczak, 1998) and thus eliminate bias from interpolation. Interpolated temperature product suffer from errors due to various factors such as insufficiency as well as non-uniform distribution of the stations, locations of meteorological stations to be biased toward lower elevations (Robeson, 1995), the abrupt weather variability at high elevation stations the smoothed interpolated product fails to predict, the position of the station at the boundary latitudes and longitudes of the interpolated product as runaway in interpolation may happen at boundary and the inherent error of the TPS method itself. Bacchi & Kottegoda (1995) mentioned that estimating the gridded distribution rainfall from point based estimates depends upon the existing spatial relationships (mostly statistical) of the point based values, *i.e.*, two stations few kilometres apart often share similar weather conditions. But, in lot of cases, a station close to a rainfall experiencing station may not experience rain giving rise to large errors. A large number of factors may contribute to errors while estimating rainfall from interpolation methods such as low number density of stations, large variability in rainfall (when rainy season is taken in study), large topographic variability (presence of mountain ranges), particular direction being followed by water bearing winds and not an isotropic distribution.

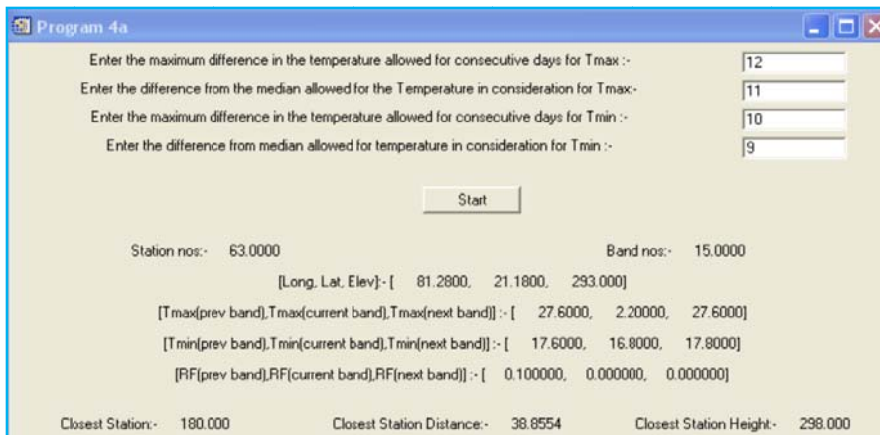


Fig. 6. Software (Snap) pointing out inaccurate data with all the ancillary information provided for deducing the correct value

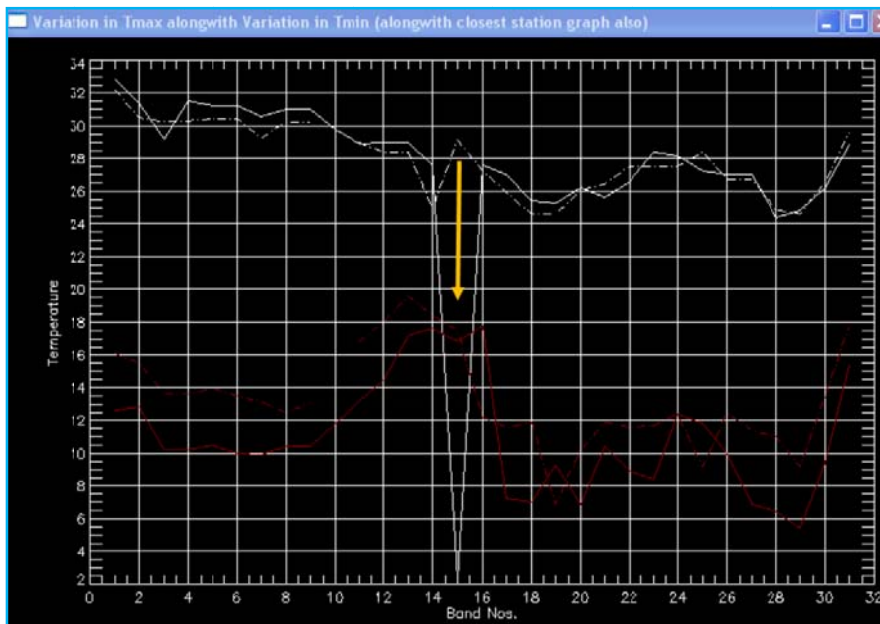
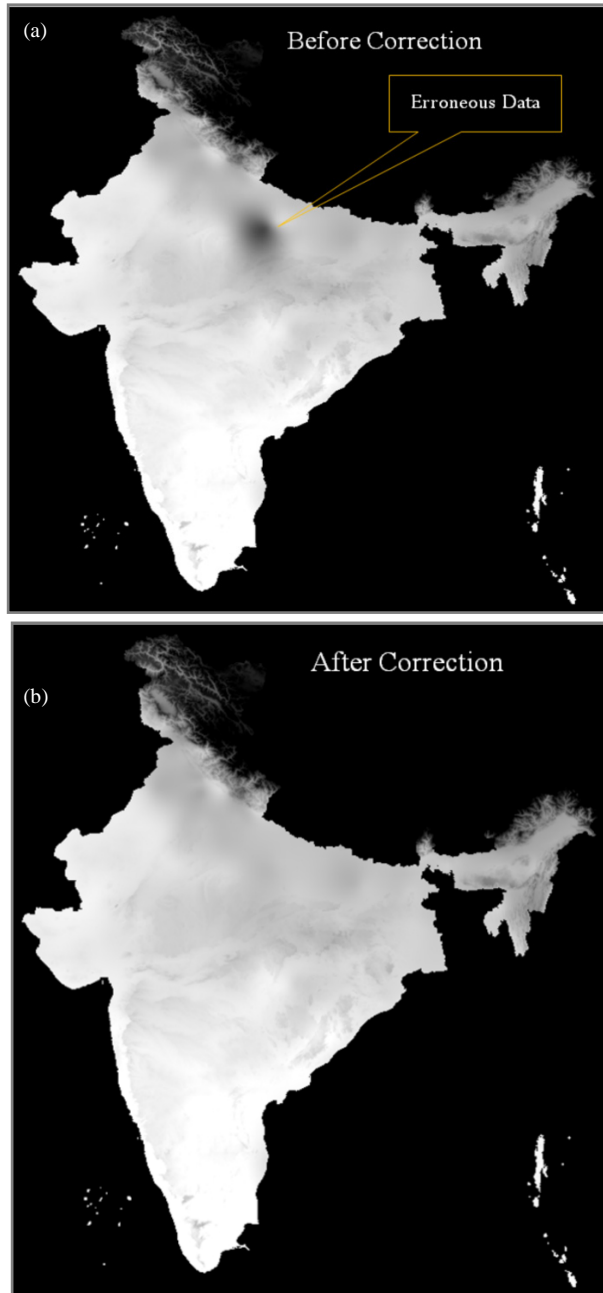


Fig. 7. Software (Snap) showing variation of temperature [both  $T_{max}$  (white) and  $T_{min}$  (red)] for a particular station (solid line) along with its closest station (dashed line). The abrupt incline confirms the erroneous data

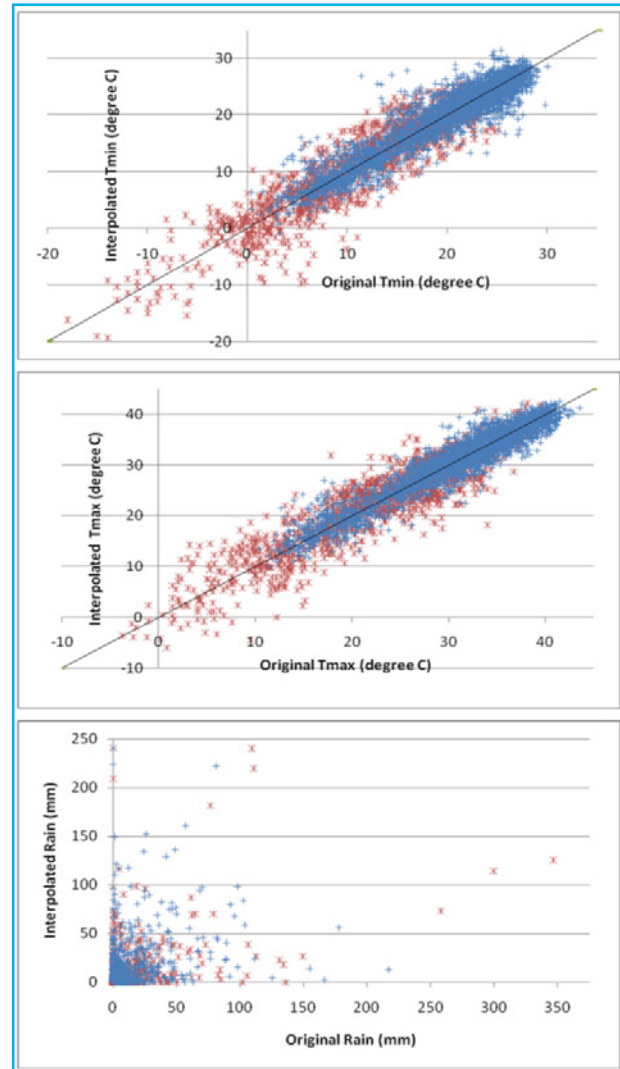
Data set was prepared for error analysis using 10 days' data each from 11<sup>th</sup> to 20<sup>th</sup> of month of January, April and August of year 2014 having 220 stations (Set 1). It was done keeping in mind the unbiased representation of seasons (winter, summer and monsoon respectively) prevailing in Indian region and also to understand seasonal effects on errors produced. In agrometeorological studies, temporal mean product is mostly utilized hence we have analyzed errors for two types of products decadal-averaged and pent-averaged. To understand the dependency of error on elevation, different

thresholds on elevation were tried with the increment of 50 m in order to determine the boundary to separate the low elevated regions from high elevated regions such that at that boundary limit, lowest and highest root mean square (rms) errors appear for low and high elevated regions respectively. The error analysis has also been performed separately for the central portion of the interpolated product enclosed in a square (LX : 74°, LY : 20°, RX : 83°, RY : 28°) to understand the influence of the location of the station in the interpolated product. Further the availability of data over 332 stations in the



**Figs. 8(a&b).** Interpolated product in image form (a) Interpolated  $T_{\max}$  product of 1 March, 2014 without quality check, dark dot represents error as  $T_{\max}$  value is 2.4 °C at Kanpur and (b) Interpolated product generated after quality check and correcting error

successive phase led us to compare the error generated with different number of stations. Here also 10 days' data each of winter, summer and monsoon were taken but due to non-availability of data, December, 2015 and March, 2016 were taken for representation of winter and summer respectively (Set 2). In Set 2, original 220 stations were filtered out and interpolated to generate product and errors



**Fig. 9.** Graph between Interpolated  $T_{\max}$  and original  $T_{\max}$  (upper left), Interpolated  $T_{\min}$  and Original  $T_{\min}$  (upper right) and Interpolated rainfall and original rainfall (lower left) for Set 1 data set with regions having elevation greater than 550 m represented with 'x' symbol (red) and with less than 550 m as '+' symbol (blue)

were listed. Similarly, whole data of 332 stations was interpolated to generate spatial product and errors were calculated for original 220 stations in order to compare. The mean bias (MB), mean absolute bias (MAB) and rms error (RMSE) have been reported for error analysis.

$$MB = \frac{\sum_{i=1}^n (W_{\text{ori}} - W_{\text{int}})_i}{n} \quad (4)$$

$$MAB = \frac{\sum_{i=1}^n |W_{\text{ori}} - W_{\text{int}}|_i}{n} \quad (5)$$



**TABLE 2**  
**Mean Bias (MB), Mean Absolute Bias (MAB) and Root Mean Square Error (RMSE) shown for decadal-averaged product for total data set as well as all the seasons**

Decadal average 2014												
	All Data			(Elevation <= 550 m)			(Elevation > 550 m)			20° N ≤ Lat.* ≤ 28° N 74° E ≤ Long.* ≤ 83° E		
	MB	MAB	RMSE	MB	MAB	RMSE	MB	MAB	RMSE	MB	MAB	RMSE
<b>January</b>												
$T_{max}$	0.07	1.24	1.78	0.24	1.07	1.42	0.39	1.70	2.50	0.02	0.95	1.30
$T_{min}$	0.06	1.55	2.18	0.02	1.28	1.76	0.17	2.29	3.03	0.03	0.88	1.18
Rain	0.94	1.51	4.98	0.92	1.53	4.45	0.99	1.48	6.20	0.02	0.79	1.41
<b>April</b>												
$T_{max}$	0.04	1.33	1.77	0.01	1.08	1.42	0.13	1.99	2.48	0.14	0.80	1.06
$T_{min}$	0.07	1.64	2.32	0.00	1.42	2.01	0.26	2.23	2.99	0.10	1.26	1.53
Rain	0.32	1.22	2.08	0.44	0.95	1.60	0.01	1.94	3.02	0.01	0.92	1.98
<b>August</b>												
$T_{max}$	0.06	1.47	2.07	0.16	1.03	1.36	0.23	2.66	3.31	0.09	1.09	1.37
$T_{min}$	0.05	1.26	1.80	0.10	1.06	1.43	0.11	1.78	2.54	0.05	1.35	1.56
Rain	1.12	6.38	13.31	1.74	5.14	9.81	0.57	9.73	19.87	0.05	2.83	4.56
<b>All seasons</b>												
$T_{max}$	0.01	1.35	1.88	0.02	1.06	1.40	0.10	2.11	2.79	0.09	0.94	1.25
$T_{min}$	0.06	1.48	2.11	0.04	1.25	1.75	0.10	2.10	2.86	0.02	1.17	1.44
Rain	0.79	3.04	8.30	1.03	2.54	6.29	0.13	4.40	12.17	0.33	1.55	3.04

\*Lat. is latitude, \*Long. is longitude

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |W_{ori} - W_{int}|_i^2}{n}} \quad (6)$$

$W$  represents the weather parameter, int and ori means interpolated and original data value;  $n$  means number of data values respectively.

### 3. Results and discussion

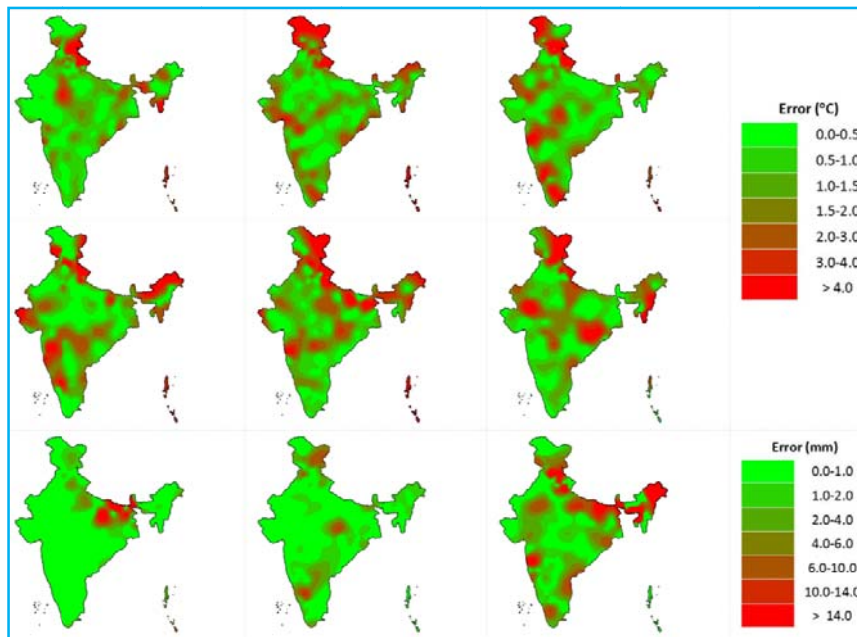
#### 3.1. Product generated and effect of quality check on output product

Fig. 6 shows the snapshot of standalone software showing the ancillary information required to correct the erroneous data. Fig. 7 shows the variation of temperature [ $T_{max}$  (white) and  $T_{min}$  (red)] of error generating station as well as variation at nearby station clearly pointing out the day of erroneous data. On the basis of information provided, user can ignore the instance thus replacing the value by -999, keep the same value or replace the value by either averaging or hot-deck computation method. The

text file created as output contains the parameter values associated with each grid for all days against respective grid code. Figs. 8 (a&b) show the output interpolated product created in the image form for visualization purpose. The spatially incompatible value appears as dark blotch in black-white image. The two images show the interpolated spatial product generated with and without using the quality check program.

#### 3.2. Validation of interpolated product

The graph (Fig. 9) shows the comparison between the original  $T_{max}$  versus interpolated  $T_{max}$ , original  $T_{min}$  versus interpolated  $T_{min}$  and original rain versus interpolated rain for whole Set1 data set. It can be observed that stations on elevation less than 550 m are found to have close match between the interpolated and original values of temperatures whereas those at higher elevation have large differences between the same. The interpolated rainfall product has large errors due to no significant match between the interpolated and original values at the stations.



**Fig. 10.** Spatial distribution of absolute errors for  $T_{\max}$ ,  $T_{\min}$  and rain (row wise) for three different seasons (column wise) of winter (January), summer (April) and monsoon (August) of Set 1 (2014)

It is observed that pent-averaged product has greater variability in comparison to decadal averaged product which is the direct consequence of averaging out of the values thus nullifying the effect of deviations. It is also observed that the mean absolute bias and rms errors are less when actual station height is taken into consideration instead of median height of grid. It is expected as it uses lapse rate corresponding to that station thus consideration of local factor into account. In further analysis, we will be emphasizing on decadal-averaged product and values associated with actual station elevation. From Table 2 for Set1 (2014), mean absolute errors were found to be in range of 1.24-1.47 °C for  $T_{\max}$ , 1.26-1.64 °C for  $T_{\min}$ , 1.22-6.38 mm for rainfall. Errors around 1 °C are noticed for  $T_{\max}$  for stations having altitude less than 550 m. Table 2 clearly brings out the influence of elevation and higher errors at higher elevation as expected. Fig. 10 showing the “seasonal” spatial distribution of absolute errors also gives a visualization of large errors noticed in high elevation regions. Low errors are noticed in central portion of interpolated product as expected. Table 2 shows that mean absolute errors less than 1 °C are noticed for  $T_{\max}$  in central portion.

When Set 2 (2015-16) was analyzed for old 220 stations with respect to interpolated product developed out of total 332 stations (data 1) and filtered 220 stations (data 2), it is clearly noticed from Fig. 11 that the error reduced in the central portion and the regions where new stations were introduced along with increase in

error in some newer high elevated boundary areas where earlier no stations lay. It shows complete lowering of error in some of the regions where density of stations increased. Mean Absolute Errors were found to be in range of 1.22-1.45 °C for  $T_{\max}$ , 1.06-1.67 °C for  $T_{\min}$ , 1.2-7.0 mm for rainfall. Mean bias errors important to agrometeorological studies were always found to be lower than 0.2 °C for temperatures and 1.4 mm for rainfall. While studying seasonal variations, it is noticed for all the datasets that  $T_{\max}$  had high and least errors in monsoon and winter respectively.  $T_{\min}$  and rainfall followed an inverse relation in errors such that when  $T_{\min}$  had highest errors, least errors were registered for rainfall data. For Set 1 data,  $T_{\min}$  had high and least errors for summer and monsoon respectively whereas for Set 2 datasets,  $T_{\min}$  had high and least errors for winter and monsoon respectively.

Pan *et al.* (2004) applied GCV (Generalized Cross Validation) embodied in Spline to monthly mean temperature interpolation of China where the predictive errors of interpolation varied between 0.87 and 1.40 °C. Hong *et al.* (2005) used spline interpolation techniques to develop a gridded climate database for China at a resolution of 0.01° in latitude and longitude employing digital elevation model (DEM) too using ANUSPLIN software. The largest interpolation errors were 0.57 °C and 0.83 °C for monthly maximum and minimum temperature surfaces respectively. Knight *et al.* (2005) evaluated the suitability of three interpolation methods in terms of accuracy for small urban catchments

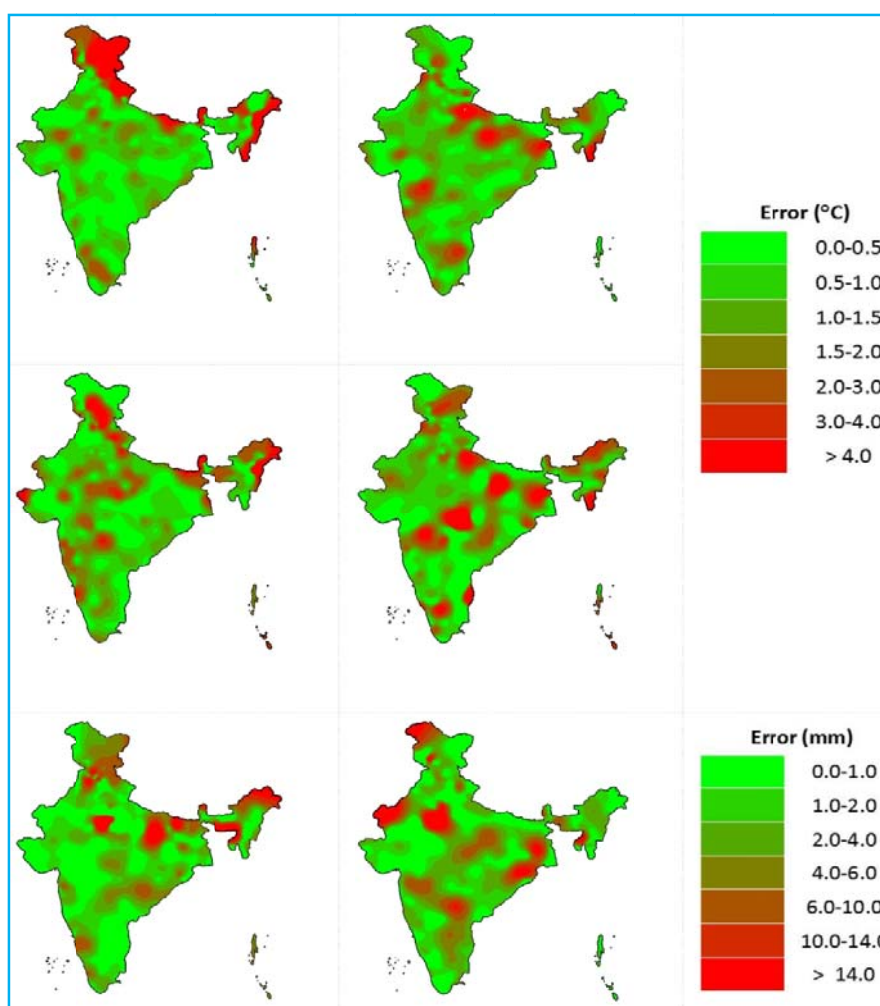


Fig. 11. Spatial distribution of absolute errors for the interpolated product developed using number of stations as 332 (left, data 1) and 220 (right, data 2) respectively

(Bridgewater Creek catchment in south Brisbane) and used ANUSPLIN software for carrying out spline interpolation. The rms error reported was 22.5 mm for 8 storm events and 28.8 mm for consistent set of 4 storm events used for comparison. In our validation product, we used decadal and pent-averaged product with high spatial resolution which would lead to larger errors than above mentioned studies.

#### 4. Conclusions

The high resolution spatially gridded weather data products generated by developed software have proved crucial for crops growth simulation model under FASAL project and are much required in agro-meteorological and other studies as numerous stations are being continually installed all over the globe. The semi-automated software

effectively identifies the erroneous temperature values and provide an easy to interact platform for correction before generating interpolated product. For decadal averaged products, mean absolute errors were found to vary within 1.2-1.5 °C for maximum temperature, 1.1-1.7 °C for minimum temperature, 1.0-7.0 mm for rainfall with higher errors being observed in monsoon for maximum temperature and rainfall and in winter for minimum temperature. Mean absolute error of less than 1 °C is noticed over central region of India for the same product. Increase in density of ground stations will lead to higher accuracy (validation done in middle portion of India supports this argument). In future, efforts will be made towards making the product fully automated where the error values can be estimated directly by using regression methods on previous available data sets and thus an attempt to generate weather data products on daily basis.

### Acknowledgement

Authors gratefully acknowledge Director, SAC; Deputy Director, EPSA, SAC; Dr. Bimal Kumar Bhattacharya, Head, AED/SAC and Dr. A. S. Arya, Head, PSD/SAC for their valuable suggestions and encouragement. Shri Shashikant Sharma, Group Head, VRG/SAC is acknowledged for providing daily downloaded IMD weather data on priority basis. The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

### References

- Bacchi, B. and Kottegoda, N. T., 1995, "Identification and calibration of spatial correlation patterns of rainfall", *Journal of Hydrology*, **165**, 311-348. [https://doi.org/10.1016/0022-1694\(94\)02590-8](https://doi.org/10.1016/0022-1694(94)02590-8).
- Bristow, K. and Campbell, G. S., 1984, "On the relation between incoming solar radiation and daily maximum and minimum temperature", *Agric. For Meteorol.*, **31**, 159-166.
- Burman, R. D. and Poshop, L. O., 1994, "Evaporation, Evapotranspiration and Climatic Data ; Developments in Atmospheric Science, Amsterdam", 22 Elsevier Science B.V., p278.
- Chaudhari, K. N., Sarkar, C., Mallick, K., Dadhwal, V. K. and Parihar, J. S., 2005, "Development and demonstration of crop growth monitoring system (CGMS) for India - An Interim Report", *Scientific Report, SAC/RSMET/CGMS/SR/01/ February 2005*.
- Collins, F. C. and Bolstad, P. V., 1996, "A Comparison of Spatial Interpolation Techniques in Temperature Estimation", Proceedings of the 3rd International Conference/Workshop on Integrating GIS and Environmental Modeling, National Centre for Geographic Information and Analysis, Santa Barbara, Santa Fe, NM; Santa Barbara, CA.
- Craven, P. and Wahba, G., 1979, "Smoothing Noisy Data with Spline Functions- Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation", *Numer. Math*, **31**, 377-403. <https://doi.org/10.1007/BF01404567>.
- Das, A. and Singh, C., 2014, "Analysis of the atmospheric parameters during recent floods in Jammu and Kashmir using satellite data : A case study", *ISPRS*, p403.
- Das, H. P. and Pujari, A. D., 1993, "A simple approach towards estimating solar irradiance", *Mausam*, **44**, 239-242.
- Dodson, R. and Marks, D., 1997, "Daily air temperature interpolated at high spatial resolution over a large mountainous region", *Climate Research*, **8**, 1-20. <https://doi.org/10.3354/cr008001>.
- Hargreaves, G. L., Hargreaves, G. H. and Riley, J. P., 1985, "Irrigation water requirement for Senegal River Basin". *J. Irrig and Drainage Engg. ASCE*, **111**, 265-275. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1985\)111:3\(265\)](https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265)).
- Hong, Y., Nix, H. A., Hutchinson, M. F. and Booth, T. H., 2005, "Spatial Interpolation of monthly mean climate data for China", *International Journal for Climatology*, **25**, 1369-1379. <https://doi.org/10.1002/joc.1187>.
- Hutchinson, M. F., 1993, "On thin plate splines and kriging. In : M.E. Tarter and M.D.", Lock (eds), *Computing and Science in Statistics*, Vol. **25**. Interface Foundation of North America, University of California, Berkeley, 55-62.
- Hutchinson, M. F., 2013, "ANUSPLIN Version 4.4 User Guide. Fenner School of Environment and Society", The Australian National University, Canberra.
- Knight, Y., Yu, B., Jenkins, G. and Morris, K., 2005, "Comparing rainfall interpolation techniques for small subtropical urban catchments", In A. Z. and R. M. Argent (Ed.), Presented at the MODSIM05, Melbourne.
- Kotsiantis, S., Kostoulas, A., Lykoudis, S., Argiriou, A. and Menagias, K., 2006, "Filling missing temperature values in weather data banks", *IET Conference Publications*, **1**, 327-334. <https://doi.org/10.1049/cp:20060659>.
- Meinguet J., 1979, "Multivariate interpolation at arbitrary points made simple", *Journal of Applied Mathematics and Physics (ZAMP)*, **30**, 292-304. <https://doi.org/10.1007/BF01601941>.
- Myers, D. E., 1994, "Spatial interpolation: an overview", *Geoderma*, **62**, 17-28. [https://doi.org/10.1016/0016-7061\(94\)90025-6](https://doi.org/10.1016/0016-7061(94)90025-6).
- Pan, Y. Z., Gong, D., Deng, L., Li, J. and Gao, J., 2004, "Smart distance searching based and DEM informed interpolation of surface air temperature in China", *Acta Geographica Sinica*, **59**, 366-374. <https://doi.org/10.11821/xb200403006>.
- Rao, K. N. and Prasad, B., 1971, "Analysis of Variation of Temperature with Latitude, Longitude and Height in India", *Meteorological Monograph, Climatology/No. 1*.
- Robeson, S. M., 1995, "Resampling of network-induced variability in estimates of terrestrial air temperature change", *Climate Change*, **29**, 213-229. <https://doi.org/10.1007/BF01094017>.
- Scholze, M., Knorr, W., Arnell, N. W. and Prentice, I. C., 2006, "A climate-change risk analysis for world ecosystems", Proceedings of the National Academy of Sciences, **103**, 13116-13120. <https://doi.org/10.1073/pnas.0601816103>.
- Shen, S. S., Dzikowski, P., Li, G. and Griffith, D., 2001, "Interpolation of 1961-97 daily temperature and precipitation data onto Alberta polygons of Ecodistrict and Soil Landscapes of Canada", *Journal of Applied Meteorology*, **40**, 2162-2177. [https://doi.org/10.1175/1520-0450\(2001\)040<2162: IODTAP>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<2162: IODTAP>2.0.CO;2).
- Srivastava A. K., Rajeevan M. and Kshirsagar, S. R., 2009, "Development of high resolution daily gridded temperature data set (1969-2005) for the Indian region", *Atmospheric Science Letters*, **10**, 249-254. <https://doi.org/10.1002/asl.232>.
- Tomczak, M., 1998, "Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW) - Cross-validation / Jackknife approach", *Journal of Geographic Information and Decision Analysis*, **2**, 2, 18-30.
- Tripathy, R., Chaudhari, K. N. and Patel, N. K., 2008, "Evaluation of different methods to estimate incoming solar radiation", *Journal of Agrometeorology*, **10**, 174-182.
- Wahba, G., 1990, "Spline models for observational data", CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), p180. <https://doi.org/10.1137/1.9781611970128>.