scientific reports

OPEN

Identifying appropriate prediction models for estimating hourly temperature over diverse agro-ecological regions of India

Santanu Kumar Bal, V. P. Pramod, V. M. Sandeep, N. Manikandan✉, M. A. Sarath Chandran, A. V. M. Subba Rao, P. Vijaya Kumar, M. Vanaja & V. K. Singh

The present study tests the accuracy of four models in estimating the hourly air temperatures in different agroecological regions of the country during two major crop seasons, *kharif* and *rabi*, by taking daily maximum and minimum temperatures as input. These methods that are being used in different crop growth simulation models were selected from the literature. To adjust the biases of estimated hourly temperature, three bias correction methods (Linear regression, Linear scaling and Quantile mapping) were used. When compared with the observed data, the estimated hourly temperature, after bias correction, is reasonably close to the observed during both *kharif* and *rabi* seasons. The bias-corrected Soygro model exhibited its good performance at 14 locations, followed by the WAVE model and Temperature models at 8 and 6 locations, respectively during the *kharif* season. In the case of *rabi* season, the bias-corrected Temperature model appears to be accurate at more locations (21), followed by WAVE and Soygro models at 4 and 2 locations, respectively. The pooled data analysis showed the least error between estimated (uncorrected and bias-corrected) and observed hourly temperature from 04 to 08 h during *kharif* season while it was 03 to 08 h during the *rabi* season. The results of the present study indicated that Soygro and Temperature models estimated hourly temperature with better accuracy at a majority of the locations situated in the agroecological regions representing different climates and soil types. Though the WAVE model worked well at some of the locations, estimation by the PL model was not up to the mark in both *kharif* and *rabi* seasons. Hence, Soygro and Temperature models can be used to estimate hourly temperature data during both *kharif* and *rabi* seasons, after the bias correction by the Linear Regression method. We believe that the application of the study would facilitate the usage of hourly temperature data instead of daily data which in turn improves the precision in predicting phenological events and bud dormancy breaks, chilling hour requirement etc.

Temperature is a critical meteorological parameter for crop development and function both in the short and long term[1]. High or low temperature, even for a short period, affects crop growth, especially in temperature-sensitive crops like wheat. Very importantly, plants' response to changes in day and night temperature (Thermoperiodism) affects the production of various enzymes and plant growth chemicals[2]. Further, an increase in temperature increases the water vapour in the air exponentially at saturation, raising the vapour pressure deficit (VPD) causing an enhancement in water loss from plants, causing a substantial reduction in vegetation productivity and yield[3]. Therefore, shifts in long-term mean annual temperature and extreme temperature events will likely have significant impacts on crop production.

Biological development rates are linearly related to temperature. It is important to include the variations in temperature in agricultural models which describe the crop phenology and development based on heat accumulation[4]. Although daily mean temperature determines the growth and development of crop plants, the day and night cycle/diurnal pattern of temperature is more important as it affects many plant morphological characteristics like leaf, shoot orientation, plant height and internodes length etc.[5] and disrupts the nutrient balance[6]. Therefore, analysing the response of crop plants to hourly temperature is more crucial than daily data.

ICAR-Central Research Institute for Dryland Agriculture, Hyderabad, Telangana 500059, India. ✉email: manikandan.
narayanan@icar.gov.in

Many studies revealed the need for sub-daily/hourly maximum and minimum temperatures to monitor the crop phenological events[7,8], dormancy breaks of temperate fruits[9,10], chilling hour requirement for apple[11–14], forecasting of plant disease transmission[15] and prediction of frost occurrence[16].

Generally, meteorological observatories record daily maximum and minimum temperatures and hourly temperature data is available at a few places[17]. The period for missing observational data can be replaced with data generated by weather generators[18]. At locations, where Automatic Weather Stations (AWS) are installed even record erroneous and uneven hourly/sub-hourly weather data due to malfunctioning of sensors/unknown disruptions[19]. Moreover, the higher cost of installation and periodic maintenance of AWS makes it difficult to increase its network, especially in developing/third-world countries.

As there is a paucity of hourly temperature data, often daily maximum and minimum temperature data are used to estimate the diurnal temperature curves[20]. The diurnal temperature curves have been in a variety of ways that vary from simple curve-fitting models based upon sine curves[21–27] to more sophisticated techniques utilizing Fourier analysis[28] and complex energy balance models[29–31]. Even though modelling approaches are different[32–35], most models are based on temperature. Employing statistical tools to enhance the performance of these predictive models/equations is a general practice in the research arena. Prior work with regard to the estimation of hourly temperature values based on the equations in widely used crop simulation models gave unsatisfactory results. This led to the use of bias correction technique to improve the model prediction ability in the present study.

With this background, the present study explores the accuracy of different models for estimating hourly temperatures from daily maximum and minimum temperatures at places representing various climatic and soil types. Three bias correction methods to improve the accuracy and usability of the estimated hourly temperature data was also taken up in this study.

Results and discussion

The estimated hourly temperature from daily maximum and minimum temperature data using four different models (WAVE, PL, Soygro and Temperature) indicated appreciable error between estimated and observed data. To increase the accuracy of estimated hourly temperature data three bias correction methods (Linear Regression, Linear Scaling and Quantile Mapping) were used and the results are furnished below.

Evaluation of bias correction methods. The heatmaps of performance metrics in terms of three efficiency criteria and two different measures for each of the 12 bias-corrected models at each location during *kharif* (2012 and 2013) and *rabi* (2012–13 and 2013–14) are presented in Figs. 1, 2, 3, 4 and 5. During the *kharif*
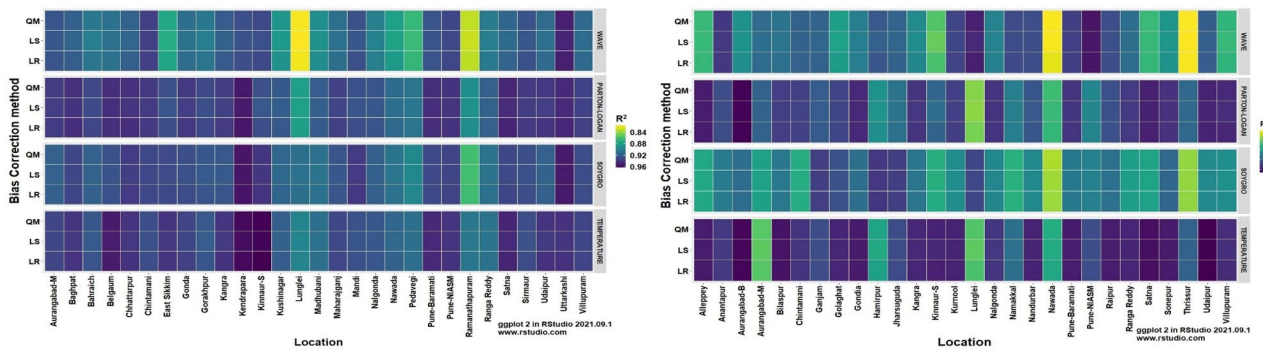

Figure 1. Performance of bias-corrected models *w.r.t* R^2 across the study locations during *kharif* (left) and *rabi* (right).
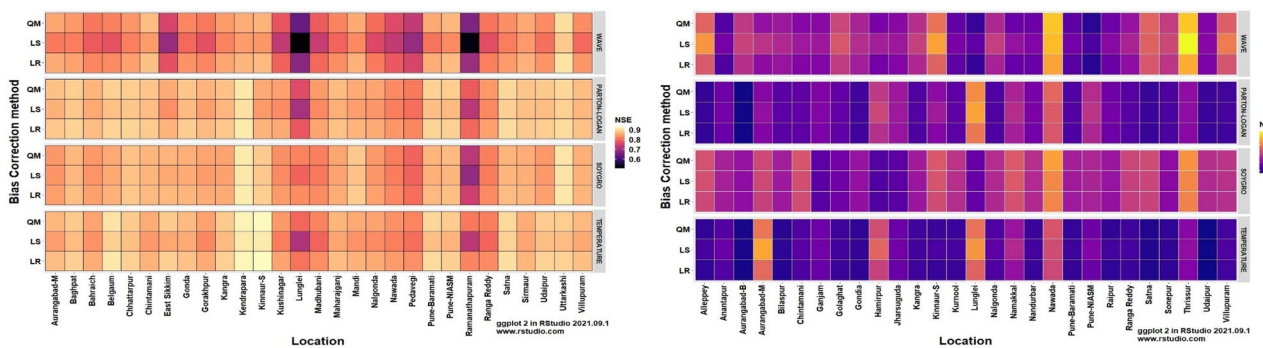

Figure 2. Performance of bias-corrected models *w.r.t.* NSE across the study locations during *kharif* (left) and *rabi* (right).
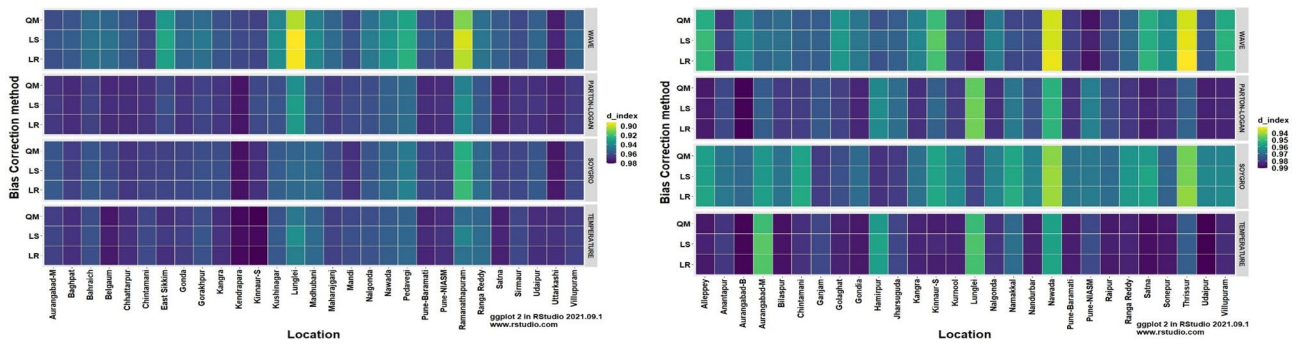
Figure 3. Performance of bias-corrected models *w.r.t* D-index across the study locations during *kharif* (left) and *rabi* (right).
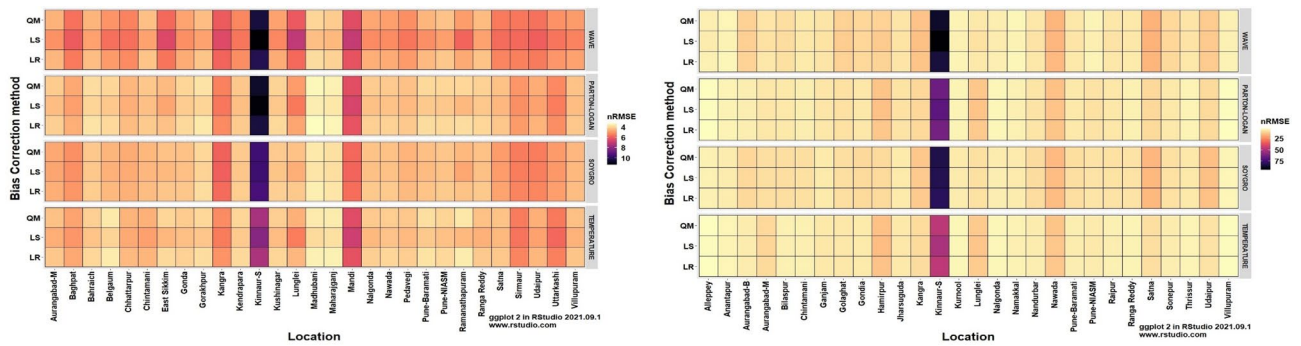

Figure 4. Performance of bias-corrected models *w.r.t* nRMSE across the study locations during *kharif* (left) and *rabi* (right).
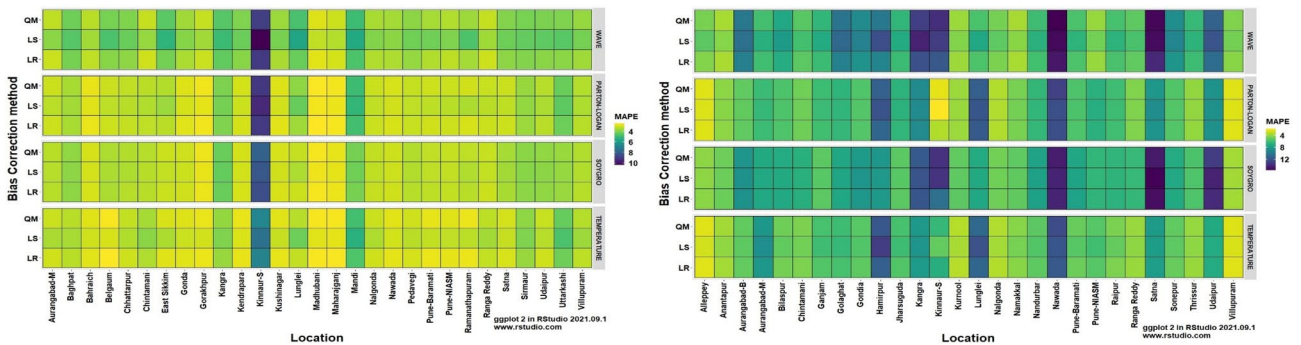

Figure 5. Performance of bias-corrected models *w.r.t* MAPE across the study locations during *kharif* (left) and *rabi* (right).

season, the bias-corrected Soygro model exhibits higher R^2, NSE and D-index values at 14 locations, followed by the WAVE model and Temperature models at 8 and 6 locations, respectively. The results of Reicosky et al.[36] indicated that the Soygro model was equally robust as the WAVE model in estimating hourly temperature from daily maximum and minimum temperature in most circumstances. Across these 28 locations, bias correction using Linear Regression was found to be more appropriate in estimating the hourly temperature. In the case of difference measures (nRMSE), Soygro model with Linear Regression as biascorrection appears to be most accurate in estimating the hourly temperature at 14 locations, followed by the WAVE and Temperature models at 8 and 6 locations, respectively, which is the same as that of the index measures. MAPE value for bias corrected (Linear Regression) Temperature model was less at 12 locations, followed by the Soygro and WAVE models at 9 and 7 locations, respectively. Table 1 shows the best of the bias-corrected models in terms of efficiency criteria and difference measures across each of the study locations.

Variation among the models in different methods of error analysis was not consistent, especially with respect to MAPE. Therefore, a greater number of efficiency criteria and different measures are chosen for error analysis to find out an appropriate bias-corrected model for estimating the hourly temperature at any particular location.

During the *rabi* season, the Temperature model with LR bias-correction was found to be accurate at more locations (20), followed by WAVE and Soygro models at 5 and 2 locations, respectively with respect to both

3

Locations	Efficiency criteria			Difference measures		
	R^2	NSE	D-Index	nRMSE	MAPE	Mode
Bahraich, Gonda, Gorakhpur, Kushinagar Madhubani, Pedavegi, Ram-anathapuram and Ranga Reddy	S-LR	S-LR	S-LR	S-LR	S-LR	S-LR
Baghpat, Chhattarpur, Kendrapara, Nawada, Udaipur and Villupuram	S-LR	S-LR	S-LR	S-LR	T-LR	S-LR
East Sikkim, Kangra, Maharajganj, Nalgonda and Satna	T-LR	T-LR	T-LR	T-LR	T-LR	T-LR
Kinnaur	T-LR	T-LR	T-LR	T-LR	S-LR	T-LR
Aurangabad-M, Belgaum, Chintamani, Kinnaur, Lunglei, Mandi, Pune and Uttarkashi	W-LR	W-LR	W-LR	W-LR	W-LR	W-LR
Sirmaur	W-LR	W-LR	W-LR	W-LR	T-LR	W-LR

Table 1. Best of bias-corrected models for *kharif* season in terms of efficiency criteria and difference measures across each of the study locations. S, Soygro; T, Temperature; W, Wave; LR, Linear Regression.

indices (R^2, D-index and NSE) and difference measures (nRMSE and MAPE). The study of Cesaraccio et al.[7] also revealed that the estimated hourly temperature data by Temperature model was superior over PL, WAVE and Wilkerson model[37]. Baker et al.[38] also reported that all three methods (WAVE, PL and Linear model) performed equally well in estimating hourly temperature data from daily maximum and minimum temperature data. However, the above paper reported that all three methods performed well during summer when compared to winter. The best of the bias-corrected models in terms of efficiency criteria and different measures across the study locations during *rabi* season is shown in Table 2. It is to be noted that, LR was found to be more effective in reducing the bias-correction of the appropriate model at each location. We used the LR criterion with an assumption that there are no measurement errors in the observed data, as hourly data was collected from AWS.

Comparison of mean hourly temperature. To compare the appropriate method, with the observed data, before and after bias correction a line plot of mean hourly temperature for both *kharif* and *rabi* seasons was plotted and presented in Fig. S1 and Fig. S2 (Supplementary file), respectively at each location. The estimated hourly temperature with and without bias correction shows a clear difference in estimating the hourly temperature, during both *kharif* and *rabi* seasons. The LR bias-correction method explicitly shows the improvement in estimating the hourly temperature across each location. The estimated hourly temperature, after bias correction, is reasonably close to the observed values during both *kharif* and *rabi* seasons (Figs. 1, 2, 3, 4, 5). McDonnel et al.[39] found LR method to be the most appropriate bias correction method for both air and soil temperatures, which improved the air and soil temperature forecasts given by European Centre for Medium-Range Weather Forecasts for Ireland. The magnitude of the errors with both uncorrected and bias-corrected models seems to vary throughout the 24 h at each location and is more distinctly noted by plotting the average hourly error, *i.e.* the difference between the estimated temperature, with and without bias-correction, and the observed temperatures in Fig. 6. The pooled data analysis showed that, during the *kharif* season, the difference between estimated (uncorrected and bias-corrected) and observed error is smallest during 04 to 08 h as most of the models take Tmin as input, whereas during the *rabi* season it is observed from 03 to 08 h in the morning. Errors with the uncorrected models at other times of the day are as large as −1.2 °C during *kharif*, and −1.4 °C during *rabi*. Reicosky et al.[36] also observed the least variation between estimated and observed hourly temperature when minimum temperature was recorded and the highest was observed during other times of the day. Further, they reported that maximum temperature didn't appear to affect the accuracy of the models employed to estimate hourly temperature from daily data. After bias correction, the magnitude of error (difference in model output and observed) got reduced to ~0 °C in *kharif*, whereas still underestimated by 0.2 °C in *rabi*. This could be attributed to the low diurnal variation in temperature that occurs during *kharif* season due to cloudy/overcast conditions in most of the day. Whereas, *rabi* is the driest season of the year during which mostly clear days and

Locations	Efficiency criteria			Difference measures		
	R^2	NSE	D-index	nRMSE	MAPE	Mode
Hamirpur	S-LR	S-LR	S-LR	S-LR	S-LR	S-LR
Nawada	S-LR	S-LR	S-LR	S-LR	T-LR	S-LR
Anantapur, Aurangabad-M, Lunglei and Namakkal	W-LR	W-LR	W-LR	W-LR	W-LR	W-LR
Alleppey, Aurangabad-B, Bilaspur, Chintamani, Golaghat, Gondia, Kangra, Kurnool, Nalgonda, Nandurbar, Pune-Baramati, Ranga Reddy, Raipur, Satna, Sonepur, Thrissur, Udaipur and Villupuram	T-LR	T-LR	T-LR	T-LR	T-LR	T-LR
Ganjam and Jharsuguda	T-LR	T-LR	T-LR	T-LR	S-LR	T-LR
Kinnaur	T-LR	T-LR	T-LR	T-LR	PL-LS	T-LR

Table 2. Best of the bias-corrected models for *rabi* season terms of efficiency criteria and difference measures across each of the study locations. S, Soygro; T, Temperature; W, Wave; LR, Linear Regression.
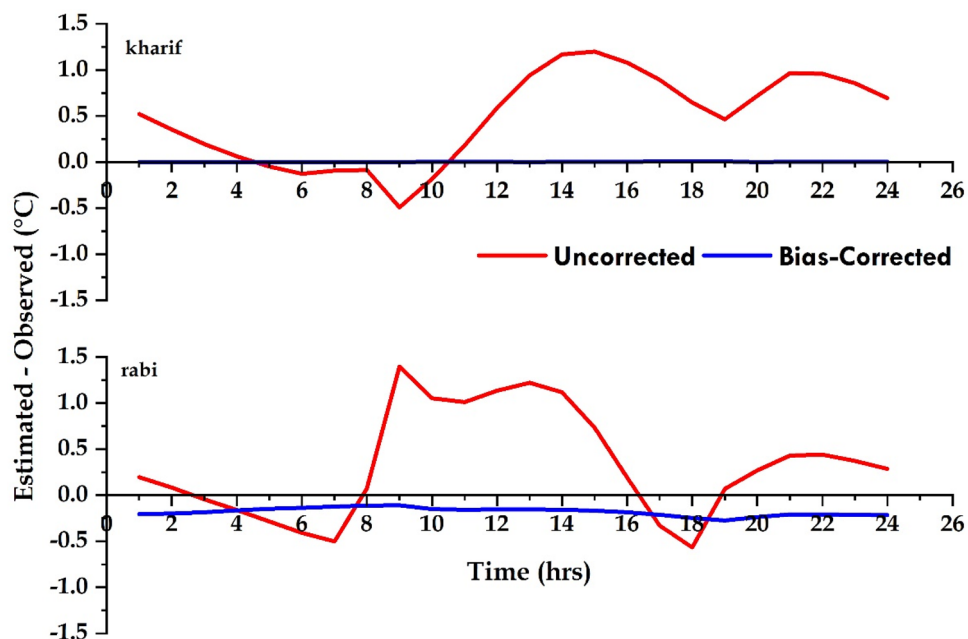
Figure 6. Average hourly temperature error calculated by comparing estimated and observed hourly temperature during *kharif* and *rabi* seasons.

nights are observed. Reicosky et al.[36] also found that predicted hourly temperature data from daily maximum and minimum temperatures fitted well during cloud-free days than cloudy days.

Suitability of models in various climatic and soil types during *kharif* and *rabi* seasons. *Climate type.* During *kharif* season, Soygro model was found to be performing well in estimating the hourly temperature in hot sub-humid (7 locations) and hot semi-arid (7 locations) climates. However, in locations representing warm per-humid/sub-humid to hot semi-arid climates WAVE and Temperature models performed better (Table 3). These observations indicate a mixed performance of the models in different climates. *Kharif* is the main rainy season of the year and prevailing cloudy conditions/overcast sky during the season leads to low diurnal fluctuation in temperature[40]. An overcast sky disturb/interfere with the incoming radiant energy during day hours and especially energy loss during night hours in the form of longwave radiation. This leads to a change in the diurnal pattern of air temperature i.e. reduction in the difference between daytime and nighttime. It is understood from the detail of the respective models that Soygro and Temperature models have divided 24 h of a day into three parts while WAVE and PL models divided the day into two parts (details in Materials and Methods section). Soygro model could capture well the change in the diurnal pattern of temperature, as two out of the three equations in the model consider data during midnight to sunrise hours and sunset to midnight hours. This could be the reason for the better performance of Soygro model in many locations when compared to WAVE and Temperature models.

Climate type	Kharif season (28 locations)				Rabi season (27 locations)			
	WAVE model	PL model	Soygro model	Temperature model	WAVE model	PL model	Soygro model	Temperature model
Warm Per-humid	1	–	–	–	1	–	–	–
Hot humid and sub-humid	–	–	–	1	–	–	–	2
Warm sub-humid	3	–	–	2	–	–	1	2
Hot sub-humid	–	–	7	2	–	–	1	9
Hot semi-arid	4	–	7	1	2	–	–	8
Hot arid	–	–	–	–	1	–	–	–

Table 3. Performance of models in predicting hourly temperature in different climates during *kharif* and *rabi* seasons.

During *rabi* season, the Temperature model exhibited better performance in 21 out of 27 locations falling under different AERs from hot per-humid & humid to hot semi-arid climates. It is observed that Soygro model worked well only in two locations representing hot and warm sub-humid climates. WAVE model was found to be best-suited for one per-humid, one arid and two hot semi-arid locations. *Rabi* season is the driest season of the year characterised by low atmospheric humidity, cloud-free sky and a higher diurnal temperature range[40]. The Temperature model, comprising two equations dealing with daylight hours and one equation for night hours, probably captured the hourly march of temperature well during the *rabi* season. It was also observed that the PL model could not perform better either during *kharif* or *rabi* seasons in any of the locations. One out of two equations of the PL model, which deals with night hours, assumes temperature decrement from sunset to sunrise at an exponential rate. Probably, this assumption can be attributed to the poor performance of PL model across locations.

Soil type.   Soil colour and texture play a predominant role in governing the diurnal pattern of soil temperature, which in turn affects the atmospheric temperature through latent and sensible heat fluxes[41]. The performance of models in estimating the hourly temperature for locations having different soil types was looked at. The results indicated that Soygro model performed better in locations having alluvium-derived soils, especially during *kharif* season. However, Temperature and WAVE models did not show any specific pattern for soil types. During *rabi* season, the hourly air temperature was well predicted by the Temperature model in all soil types. The Temperature model worked well in 11 out of 27 locations represented by red, black, and lateritic soils (Table 4). Performance of the model was also good at 6 locations having alluvium soils and 2 locations each represented by red loamy and brown forest podzolic soils. At the same time, the WAVE model performed well in red, black, lateritic and red loamy soils, while Soygro model performed well in only two locations represented by alluvium and brown forest soils. The study showed that Soygro and Temperature models worked well in most of the locations during *kharif* and *rabi* seasons, respectively when compared to the WAVE model. However, the PL model's performance was not up to the mark, compared to the other three models in any of the soil types. It is understood from the results that Soygro and Temperature models which have three different equations to deal with the day and night hours helped for precise estimation of hourly data by these models.

As mentioned earlier, the Soygro model performed well in the locations with alluvium soils, especially during *kharif* season and not during the *rabi* season. In general, the average soil moisture status of the soil is higher during the *kharif* season than the *rabi* season and vice versa. During the *kharif* season, when soil moisture nears field capacity, it takes more time to heat up the soil and also the atmosphere later as the specific heat of wet sandy soil is almost two times higher (0.4 cal g^{-1} deg^{-1}) than dry sandy soil (0.2 cal g^{-1} deg^{-1})[42]. During the *rabi* season, as soil is dry, it absorbs and releases the heat energy in less time and also the diurnal soil temperature range would be greater than the *kharif* season. Though Soygro and Temperature models use three equations, Soygro model consists of only one equation to deal with the daylight period while the Temperature model has two equations for the daylight period. This could be the reason for the better performance of the Temperature model during *rabi* season in most of the locations having different soil types.

Conclusions

Four models viz*,* WAVE, PL, Soygro and Temperature model were used in estimating the hourly temperature at selected locations in different AERs of the country during two crop seasons (*kharif* and *rabi*). When compared with the hourly observed data, the bias-corrected Soygro model exhibits higher values *w.r.t* three efficiency criteria (R^2, NSE and D-index) at 14 locations, followed by the WAVE model and Temperature models in 8 and 6 locations, respectively during the *kharif* season. In the case of the *rabi* season, bias corrected Temperature model appears to be accurate at more locations (20), followed by WAVE and Soygro models at 5 and 2 locations, respectively. Bias correction was found to be more effective in reducing the error of the appropriate model at each location with Linear Regression irrespective of the season. The magnitude of errors with both uncorrected and bias-corrected models seems to be changing over the 24-h period of the day across each of the locations. The performance of four different models over different agroecological regions of the country revealed the better performance of the Soygro model during the *kharif* season at locations representing hot sub-humid and semi-arid climates. Whereas, the Temperature model outperformed the other model during the *rabi* season at most of the locations ranging from hot per-humid to hot semi-arid climates. Estimating the accuracy of hourly

Soil type	Kharif season (28 locations)				Rabi season (27 locations)			
	WAVE model	PL model	Soygro model	Temperature model	WAVE model	PL model	Soygro model	Temperature model
Alluvium derived	–	–	11	2	–	–	1	6
Red, black and lateritic	4	–	2	2	3	–	–	11
Red loamy	1	–	1	–	1	–	–	2
Brown forest podzolic	3	–	–	2	–	–	1	2

Table 4. Performance of models in predicting hourly temperature in different soil types during *kharif* and *rabi* seasons.

temperature in different soil types indicated that the Soygro model worked well during the *kharif* season at the locations having alluvium-derived soils. At the same time, during the *rabi* season, the performance of the Temperature model was better than other models in the agro ecological regions dominated by red, black and lateritic soils. The smallest difference between estimated (uncorrected and bias-corrected) and observed was noticed from 04 to 08 h during the *kharif* season when most of the models assume Tmin as input, while it was from 03 to 08 h in the morning during the *rabi* season. The results of the study strongly suggested that hourly temperature during the *kharif* and *rabi* seasons can be estimated using the Soygro and Temperature models, respectively after the bias correction by the Linear Regression method, as these two models estimated hourly temperature with better accuracy at a majority of the locations representing different climate and soil types. Nevertheless, the WAVE model also can be used to estimate the hourly temperature data at some of the study locations. However, the PL model did not perform satisfactorily at any of the locations. Hourly temperature data in lieu of daily temperature (maximum and minimum) data would enhance the precision in predicting phenological and other biological events.

Data and methods

Study area.   Observed maximum and minimum temperatures, at daily and hourly scales, were collected from the AWS of the AICRPAM-NICRA network. National Innovations in Climate Resilient Agriculture (NICRA), a network project of the Indian Council of Agricultural Research (ICAR) aims to enhance the resilience of Indian agriculture to climate change and climate vulnerability through strategic research and technology demonstration in the 100 most vulnerable districts of the country. The climatic variability is being assessed by the All India Coordinated Research Project on Agrometeorology (AICRPAM) by installing AWS in each district under the NICRA project[43]. The AWS records 6 meteorological parameters viz., temperature (maximum and minimum), relative humidity (maximum and minimum), wind speed, wind direction, rainfall and solar direction. Depending on the availability of continuous hourly temperature (maximum and minimum) data over the study period, 44 locations were selected for the present study (Fig. 7 and Table S1 of the Supplementary file). Out of 44 locations, the hourly temperature was estimated at 28 locations during the *kharif* season (2012 and 2013) and at 27 locations during the *rabi* season (2012–13 and 2013–14).

Methods of estimating hourly temperature.   Four different methods for estimating hourly temperature were tested at different agro-ecological regions (AERs) of the country during two major crop seasons viz., the *kharif* (15-Jun to 15-Oct) and *rabi* (15-Oct to 15-Mar), by inputting the daily maximum and minimum temperature. These methods were taken from different crop growth simulation models. Therefore, in the present study, these methods are known by the name of their respective crop growth models or the sub-routines in which they were employed. The details of each method are discussed hereunder.

WAVE model.   The model was first introduced by De Wit et al.[44] and it was included in the subroutine WAVE in ROOTSIMU V4.0[45]. The model uses two different relations; formula one for estimating the hourly temperature from the time of minimum temperature (sunrise hour) to the time of maximum temperature (1400 h) and another for estimating from the time of maximum temperature to the time of the minimum temperature of the next day. A cosine function is used in the model for both periods and is given by the following Eqs. (1) and (2).

For Sunrise Hour ≤ H ≤ 1400 h

$$T_{(H)} = T_{ave} - T_{amp} * (Cos\ [\pi * (H\text{-Sunrise Hour})/(14\text{-Sunrise Hour})]) \tag{1}$$

For 1400 h < H < Sunrise Hour of next day

$$T_{(H)} = T_{ave} + T_{amp} * (Cos\ [\pi * (H')/(10 + Sunrise Hour)]) \tag{2}$$

where $T_{(H)}$ is the temperature at hour H, $T_{ave} = \frac{T_{max} + T_{min}}{2}$; $T_{amp} = \frac{T_{max} - T_{min}}{2}$; T_{max} and T_{min} are the maximum and minimum temperature, respectively. H' = H-14 if H > 1400 h and H' = H + 10 if H < Sunrise Hour. For estimating the hourly temperature past 1400 h, T_{min} of the next day is considered. In the WAVE model, no site-specific calibration is required.

Parton and Logan (PL) model.   The model developed by Parton and Logan[46] uses two different equations, like in the WAVE model, for estimating the hourly air temperature during daylight hours and night-time hours. Parton-Logan's model considers the sunrise hour to sunset hour as daylight hours and the time between the sunset hour and sunrise hour of the next day as night-time hours. The model utilizes a truncated sine wave at daylight hours (day-time) and an exponential decrease in temperature at night-time for estimating the hourly temperature. The temperature at any given hour ($T_{(H)}$) during the daylight hours is given by Eq. (3).

$$T_{(H)} = T_{min} + (T_{max} - T_{min}) * Sin\left[\frac{\pi}{2} * \frac{(H - H_{min})}{(H_{max} - H_{min})}\right] \tag{3}$$

where T_{max} and T_{min} are the maximum and minimum temperature, respectively and H_{max} and H_{min} are the hour of maximum and minimum temperatures, respectively. In the present study, H_{max} is considered as 4 h before sunset hour (i.e. $H_{max} = H_{ss}$-4 where H_{ss} is the sunset hour) and sunrise hour is considered as the H_{min}. The temperature at any given hour ($T_{(H)}$) during the night-time hours is given by Eq. (4).
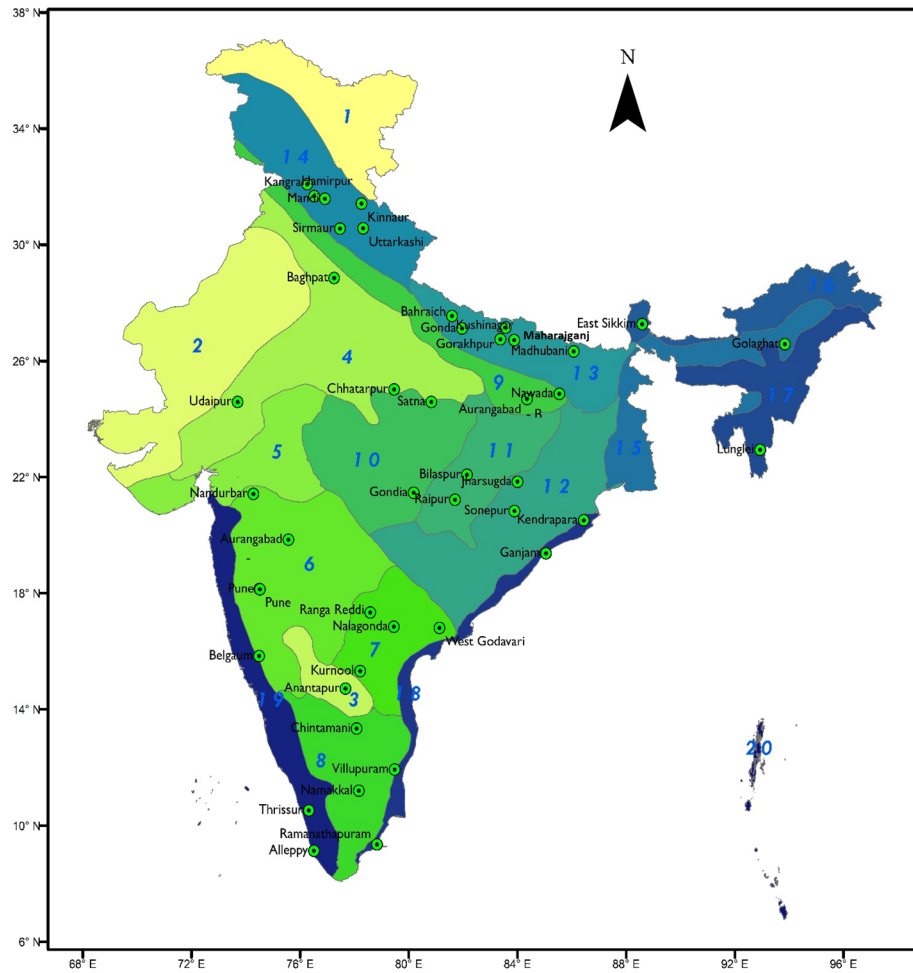
1 - Western Himalayas, cold arid ecoregions, with shallow skeletal soils
2 - Western Plain, hot arid ecoregions with desert & saline soils
3 - Deccan Plateau, hot arid ecoregions with red & black soils
4 - Northern Plain & Central Highlands, hot semi-arid ecoregions with alluvium derived soils
5 - Central High lands, hot semi-arid ecoregions with medium & deep black soils
6 - Deccan Plateau, hot semi-arid ecoregions with shallow & medium black soils
7 - Deccan Plateau & Eastern Ghats, hot semi-arid ecoregions with red & medium black soils
8 - Eastern Ghats, Tamil Nadu Uplands & Deccan Plateau, hot semi-arid ecoregions with red loamy soils
9 - Northern Plain, hot sub-humid ecoregions with alluvium derived soils
10- Central High lands, hot sub-humid ecoregions with black & red soils
11- Eastern Plateau, hot sub-humid ecoregions with red & yellow soils
12- Eastern Plateau & Eastern Ghats, hot sub-humid ecoregions with red & lateritic soils
13- Eastern Plain, hot sub-humid ecoregions with alluvium derived soils
14- Western Himalayas, warm sub-humid ecoregions with brown forest & podzolic soils
15- Bengal & Assam Plain, hot sub-humid to humid ecoregions with alluvium derived soils
16- Eastern Himalayas, warm per-humid ecoregions with brown and red hill soils
17- North Eastern Hills, warm per-humid ecoregions with red and lateritic soils
18- Eastern Coastal Plain, hot sub-humid to semi-arid ecoregions with coastal alluvium derived soils
19- Western Ghats & Coastal Plain, hot humid-per-humid ecoregions with red, lateritic & alluvium derived soils
20- Islands of Andaman-Nicobar & Lakshadweep hot humid to per-humid island ecoregion with red loamy & sandy soils

Figure 7. Study locations over different AER's of the country *(Numbers within each polygon represent different AERs)* (map generated using ArcMap 10.3; URL: https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview).

$$T_{(H)} = T_{min} + (T_{ss} - T_{min}) * e^{-b*\frac{N}{L}} \tag{4}$$

where T_{ss} is sunset hour temperature which can be determined using the daytime hour temperature relation, b is the empirical constant (dimensionless), N is the time since sunset (s) and L is the night length (s). In the present study, the empirical constant 'b' is taken as 2.20 across all the locations, as the parameter did not appear to be strongly site-specific[36].

Soygro model. The model introduced by Wilkerson et al.[37] estimates the hourly temperature by dividing the day into three segments viz., midnight to sunrise + 2 h; daylight hours and sunset to midnight. For each segment, the model uses three different Eqs. (5) to (14), which are given below.

For midnight to sunrise + 2 h

$$TAU = (SET_{n-1} - RISE_{n-1} - 2)/(SET_{n-1} - RISE_{n-1}) \tag{5}$$

$$TLIN = TMIN_{n-1} + (TMAX_{n-1} - TMIN_{n-1}) * Sin(TAU) \tag{6}$$

$$SLOPE = (TLIN - TMIN_n)/(24 - SET_{n-1} + RISE_n + 2) \tag{7}$$

$$T_{(H)} = TLIN - SLOPE(H + 24 - SET_{n-1}) \tag{8}$$

For sunset to midnight

$$TAU = (SET_n - RISE_n - 2)/(SET_n - RISE_n) \tag{9}$$

$$TLIN = TMIN_n + (TMAX_n - TMIN_n) * Sin(TAU) \tag{10}$$

$$SLOPE = (TLIN - TMIN_{n+1})/(24 - SET_n + RISE_{n+1} + 2) \tag{11}$$

$$T_{(H)} = TLIN - SLOPE(H - SET_n) \tag{12}$$

For daylight hours

$$TAU = (H - RISE_n - 2)/(SET_n - RISE_n) \tag{13}$$

$$T_{(H)} = TMIN_n + (TMAX_n - TMIN_n) * Sin(TAU) \tag{14}$$

where TAU, TLIN, SLOPE are temporary variables in calculations, $T_{(H)}$ is the temperature at hour H, n is the current day of the year (1 to 365), RISE is the sunrise hour and SET is the sunset hour. The method assumes a change from night to day temperature at sunrise + 2 h, and the night temperatures are linear with time. It is to be noted that, like the WAVE and Parton-Logan models, the Soygro model also does not consider any site-specific constants.

Temperature model. The temperature model presented by Cesaraccio et al.[7] is an empirical model for estimating the hourly mean temperature. Like the Soygro model, the temperature model also divides the day into three time periods i.e. from the sunrise hour (H_n) to the time of maximum temperature (H_x), from H_x to the sunset hour (H_o) and from H_o to the sunrise hour for the next day (H_p). The model uses two sine-wave functions in the daylight and a square-root decrease in temperature at night. The sunrise and sunset hours are determined as a function of site latitude and the day of the year, whereas the hour of maximum temperature is assumed to be 4 h before sunset ($H_x = H_o - 4$). H_p is calculated as $H_p = H_n + 24$. T_n and T_x are the minimum and maximum temperature of the current day which occurs at H_n and H_x, respectively while the temperature during the sunset hour (T_o) is determined using the below-mentioned Eq. (15).

$$T_o = T_x - c(T_x - T_p) \tag{15}$$

where T_p is the minimum temperature of the following day and c is the empirical constant, which is obtained by fitting the equation to the observed hourly data set. For a given T_n, T_x, T_o and T_p, the Temperature model estimates the temperature at any given hour ($T_{(H)}$) for the three-time periods with the help of respect to three different Eqs. (16) to (18).

For $H_n < H \leq H_x$

$$T_{(H)} = T_n + \alpha * Sin\left[\left(\frac{H - H_n}{H_x - H_n}\right)\frac{\pi}{2}\right] \tag{16}$$

For $H_x < H < H_o$

$$T_{(H)} = T_o + R * Sin\left[\frac{\pi}{2} + \left(\frac{H - H_x}{4}\right)\frac{\pi}{2}\right] \tag{17}$$

For $H_o < H \leq H_p$

$$T_{(H)} = T_o + b\sqrt{H - H_o} \tag{18}$$

where $\alpha = T_x - T_n$; $R = T_x - T_o$ and $b = \frac{T_p - T_o}{\sqrt{H_p - H_o}}$.

Bias correction methods. Researchers used different bias correction methods to reduce the errors and improve the accuracy of prediction significantly in different parts of the world[47–50]. In the present study also, to

reduce the biases and improve the accuracy of estimated hourly temperature, three bias correction methods were used and the details of each are discussed below.

Linear regression. Linear Regression (LR) is simply a line of best fit through the standard regression plot of the observed against the estimated. Since temperature is well described by a normal distribution, a linear fit is sufficient. The relative relationship of the values being compared is given by the slope value, whereas the intercept indicates the lead or lag between the data. A slope of one and a y-intercept of zero show a perfect fit between simulated and observed data[51] as mentioned in Eq. (19).

$$T_{cor,m,d,h} = m_{m(h)} * T_{est,m,d,h} + c_{m(h)} \qquad (19)$$

where $T_{cor,m,d,h}$ and $T_{est,m,d,h}$ are the bias-corrected and estimated temperature (using either of the four models), respectively for the 'hth' hour of 'dth' day and 'mth' month, $m_{m(h)}$ and $c_{m(h)}$ are the slope and intercepts of 'hth' hour of 'mth' month, respectively.

Linear scaling. The Linear Scaling (LS) method aims to perfectly match the monthly mean of corrected values with that of observed values[52]. It operates with monthly correction values worked out as the difference between observed and simulated data. In the present study, the bias correction for the estimated hourly temperature is given by Eq. (20).

$$T_{cor,m,d,h} = T_{est,m,d,h} + \mu\left(T_{obs,(m(h))}\right) - \mu\left(T_{est,m(h)}\right) \qquad (20)$$

where $\mu\left(T_{obs,m(h)}\right)$ and $\mu\left(T_{est,m(h)}\right)$ are the mean value of observed and estimated temperature for the 'hth' hour of 'mth' month.

Quantile mapping. Among the different bias correction methods, Quantile Mapping (QM) is considered the most useful and popular. By applying a transfer function, QM corrects the systematic bias in the simulated data so that it matches the distribution of the observational dataset[53,54]. For the estimated hourly temperature, the QM bias correction is given by Eq. (21).

$$T_{cor,m,d,h} = CDF^{-1}_{obs,m}\left(CDF_{est,m}\left(T_{est,m,d,h}\right)\right) \qquad (21)$$

where $T_{cor,m,d,h}$ and $T_{est,m,d,h}$ are the bias-corrected and estimated temperature (using either of the four models), respectively for the 'hth' hour of 'dth' day and 'mth' month; $CDF^{-1}_{obs,m}$ is inverse CDF of observed data set in the validation period; $CDF_{est,m}$ is CDF of estimated data set in the validation period.

Methods of error analysis. After applying bias correction techniques (LR, LS and QM), the accuracy of various models (WAVE, Parton-Logan, Soygro and Temperature) in estimating the hourly temperature was tested by comparing them with the observed values at each location. The "goodness of fit" of each model was assessed using three efficiency criteria (coefficient of determination, coefficient of efficiency and index of agreement) and two different measures (normalized root mean square error and mean absolute percentage error). The bias-corrected model having more efficiency criteria and fewer difference measures is chosen as the appropriate model for estimating the hourly temperature at a particular location. The details of each efficiency criterion and different measures are given below.

Coefficient of determination. The coefficient of determination (R^2) is the square of Pearson's product-moment correlation coefficient and describes the proportion of the total variance in the observed data that can be explained by the model. R^2 value ranges from 0 to 1, with higher values indicating better agreement and the formula of R^2 is given in Eq. (22).

$$R^2 = \left[\frac{\sum_{i=1}^{n}\left(O_i - O_{avg}\right)\left(S_i - S_{avg}\right)}{\left[\sum_{i=1}^{n}\left(O_i - O_{avg}\right)^2\right]^{0.5}\left[\sum_{i=1}^{n}\left(S_i - S_{avg}\right)^2\right]^{0.5}}\right]^2 \qquad (22)$$

where O_i and S_i and are the observed and bias-corrected estimated values, O_{avg} and S_{avg} are the average values of observed and bias-corrected estimated values, respectively and n is the total number of observations.

Coefficient of efficiency. As correlation-based measures are more sensitive to outliers, the coefficient of determination measure leads to a bias towards the extreme events if employed in model evaluation. Nash and Sutcliffe[55] defined the coefficient of efficiency (NSE) as an improvement over the coefficient of determination for model evaluation purposes. Physically, it is the ratio of the mean square error to the variance in the observed data, subtracted from unity (Eq. 23).

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(O_i - S_i\right)^2}{\sum_{i=1}^{n}\left(O_i - O_{avg}\right)^2} \qquad (23)$$

The value of NSE ranges from $-\infty$ to 1, with higher values indicates perfect simulation.

Index of agreement.   In order to overcome the insensitivity of correlation-based measures to differences in the observed and model-simulated means, Willmott[56] introduced the index of agreement (D-index), given by Eq. (24).

$$D\text{-index} = 1 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} \left(\left| S_i - O_{avg} \right| + \left| O_i - O_{avg} \right| \right)^2} \tag{24}$$

Like R^2, the value of D-index also varies from 0 to 1. The value of 1 indicates better agreement between model simulated and observed values.

Normalized root mean square error.   Root mean square error (RMSE) is increasingly being used in the comparison and evaluation of simulation models and considered as the paramount measure of accuracy. It is defined in the Eq. (25).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(S_i - O_i)^2}{n}} \tag{25}$$

In terms of percentage error, root mean square error is given by normalized RMSE (nRMSE), which is given by the Eq. (26).

$$nRMSE = \left(\frac{RMSE}{O_{avg}} \right) * 100 \tag{26}$$

Mean absolute percentage error.   Mean Absolute Percentage Error (MAPE) is a measure of difference between two continuous variables, which is given by the Eq. (27).

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{S_i - O_i}{O_i} \right| \times 100}{n} \tag{27}$$

MAPE is less sensitive to extreme values and therefore, it is intuitively more appealing than RMSE. However, for evaluating the performance of model both RMSE and MAPE can be used as many researchers employed the both indices[50,57].

Data availability
The data used for the study are available from the corresponding author on reasonable request.

References
1. Moore, C. E. *et al.* The effect of increasing temperature on crop photosynthesis: From enzymes to ecosystems. *J. Exp. Bot.* **72**(8), 2822–2844. https://doi.org/10.1093/jxb/erab090 (2021).
2. Patel, D. & Franklin, K. A. Temperature-regulation of plant architecture. *Plant Signal. Behav.* **47**(7), 577–579 (2009).
3. Konings, A. G., Williams, A. P. & Gentine, P. Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. *Nat. Geosci.* **10**, 284–288 (2017).
4. VijayaKumar, P. *et al.* Algorithms for weather-based management decisions in major rainfed crops of India: Validation using data from multi-location field experiments. *Agronomy J.* **113**, 1816–1830. https://doi.org/10.1002/agj2.20518 (2021).
5. Myster, J. & Moe, R. Effect of diurnal temperature alternations on plant morphology in some greenhouse crops—A mini review. *Sci. Hortic.* **62**(4), 205–215 (1995).
6. Inthichack, P., Nishimura, Y. & Fukumoto, Y. Diurnal temperature alternations on plant growth and mineral absorption in eggplant, sweet pepper and tomato. *Hortic. Environ. Biotechnol.* **54**, 37–43 (2013).
7. Cesaraccio, C., Spano, D., Duce, P. & Snyder, R. L. An improved model for determining degree-day values from daily temperature data. *Int. J. Biometeorol.* **45**, 161–169 (2001).
8. Bal, S. K. *et al.* Water demand of maize is projected to decrease under near-future climate in India. *Sustainability* https://doi.org/10.3390/su14031419 (2022).
9. Anderson, J. L., Richardson, E. A. & Kesner, C. D. Validation of chill unit and flower bud phenology models for "Montmorency" sour cherry. *Acta Hortic.* **184**, 71–78 (1986).
10. Erez, A., Fishman, S., Linsley-Noakes, G. C. & Allan, P. The dynamic model for rest completion in peach buds. *Acta Hortic.* **276**, 165–174 (1990).
11. Basannagari, B. & Kala, C. P. Climate change and apple farming in Indian Himalayas: A study of local perceptions and responses. *PLoS ONE* **8**(10), e77976. https://doi.org/10.1371/journal.pone.007796 (2013).
12. Hartta, Y. S. Climate change and apple productions in Himachal Pradesh: A study of last two decades. *Acad. Discourse* **3**(1), 75–81 (2014).
13. Pramanick, K. *et al.* Role of changing climate on chilling unit accumulation and yield for apple (*Malus* x *Domestica* Borkh) cultivation at Shimla, Himachal Pradesh, India. *Int. J. Trop. Agric.* **33**(2), 1039–1044 (2015).
14. Sahu, N. *et al.* Why apple orchards are shifting to the higher altitudes of the Himalayas?. *PLoS ONE* **15**(7), e0235041. https://doi.org/10.1371/journal.pone.0235041 (2020).
15. Kang, W. S. *et al.* A web-based information system for plant disease forecast based on weather data at high spatial resolution. *Plant Pathol. J.* **26**(1), 37–48 (2010).
16. Bal, S. K. *et al.* Developing frost prediction models using multivariate statistical techniques for two diverse locations of Northern India. *Theor. Appl. Climatol.* **146**, 1097–1110. https://doi.org/10.1007/s00704-021-03786-8 (2021).

17. Critchfield, H. J. *General Climatology* 4th edn, 453 (PHI Learning Pvt. Ltd, 2013).
18. Bal, S. K., Choudhury, B. U., Sood, A., Jalota, S. K. & Singh, H. Evaluation of climgen model to generate weather parameters in different climatic situations in Punjab. *J. Agrometeorol.* **10**(1), 73–79 (2008).
19. Shoaib, T. A. & Rasool, S. N. Correcting real time automatic weather stations data through quality checks and analysis. *Vayu Mandal.* **41**, 69–76 (2015).
20. Snyder, R. L., Spano, D., Cesaraccio, C. & Duce, P. Determining degree-day threshold from field observations. *Int. J. Biometeorol.* **42**, 177–182 (1999).
21. Allen, J. C. A modified sine wave method for calculating degree days. *Environ. Entomol.* **5**(3), 388–396 (1976).
22. Hansen, J. E. & Driscoll, D. M. A mathematical model for the generation of hourly temperatures. *J. Appl. Meteor.* **16**(9), 935–948 (1977).
23. Floyd, R. B. & Braddock, R. D. A simple method for fitting average diurnal temperature curves. *Agric. For. Meteorol.* **32**(2), 107–119 (1984).
24. Wann, M., Yen, D. & Gold, H. J. Evaluation and calibration of three models for daily cycle of air temperature. *Agric. For. Meteorol.* **34**(2–3), 121–128 (1985).
25. De Gaetano, A. & Knapp, W. W. Standardization of weekly growing degree day accumulations based on differences in temperature observation and method. *Agric. For. Meteorol.* **66**(1–2), 1–19 (1993).
26. Yin, X., Kropff, M. J., McLaren, G. & Visperas, R. M. A non-linear model for crop development as a function of temperature. *Agric. For. Meteorol.* **77**, 1–16 (1995).
27. Roltsch, J. W., Zalom, F. G., Strawn, A. J., Strand, J. F. & Pitcairn, M. J. Evaluation of several degree day estimation methods in California climates. *Int. J. Biometeorol.* **42**, 169–176 (1999).
28. Carson, J. E. Analysis of soil and air temperatures by Fourier techniques. *J. Geophys. Res.* **68**(8), 2217–2232 (1963).
29. Carson, J. E. & Moses, H. The annual and diurnal heat exchange cycle in upper layers of soil. *J. Appl. Meteorol.* **2**(3), 397–406 (1963).
30. Brown, G. W. Predicting temperatures of small streams. *Water Resour. Res.* **5**(1), 68–75 (1969).
31. Lemon, E., Steward, D. W. & Shawcraft, R. W. The sun's works in a cornfield. *Science* **174**(4007), 371–378 (1971).
32. Galán, C., Fuillerat, J. M., Comtois, P. & Dominguez-Vilches, E. Bioclimatic factors affecting daily Cupressaceae flowering in southwest Spain. *Int. J. Biometeorol.* **41**, 95–100 (1998).
33. Hänninen, H. Modeling bud dormancy release in trees from cool and temperate regions. *Acta. For. Fenn.* **213**, 47p (1990).
34. Kramer, K. Selecting a model to predict the onset of growth of *Fagus sylvatica*. *J. Appl. Ecol.* **31**(1), 172–181 (1994).
35. Maak, K. & von Storch, H. Statistical downscaling of monthly mean air temperature to the beginning of flowering of Galanthus-nivalis L. in Northern Germany. *Int. J. Biometeorol.* **41**, 5–12 (1997).
36. Reicosky, L. J., Winkelman, J. M., Baker, J. M. & Baker, D. G. Accuracy of hourly air temperatures calculated from daily minima and maxima. *Agric. For. Meteorol.* **46**(3), 193–209 (1989).
37. Wilkerson, G. G., Jones, J. W., Boote, K. J., Ingram, K. T. & Mishoe, J. W. Modeling soybean growth for crop management. *Trans. ASAE* **26**(1), 63–73 (1983).
38. Baker, J. M., Reicosky, D. C. & Baker, D. G. Estimating the time dependence of air temperature using daily maxima and minima: A comparison of three methods. *J. Atmos. Ocean. Technol.* **5**(6), 736–742 (1988).
39. McDonnell, J. *et al.* Verification and bias correction of ECMWF forecasts for Irish weather stations to evaluate their potential usefulness in grass growth modelling. *Meteorol. Appl.* **25**(2), 292–301 (2018).
40. Samui, R. P. & John, G. A simple approach in assessing the impact of weather on rice yield in two rice growing seasons at Pattambi, Kerala. *Mausam* **54**(2), 477–482 (2003).
41. Pan, H. L. & Mahrt, L. Interaction between soil hydrology and boundary-layer development. *Bound. Layer Meteorol.* **38**, 185–202 (1987).
42. Rosenberg, N. J. *Micro Climate: The Biological Environment* 315 (Wiley-Interscience Publication, 1974).
43. VijayaKumar, P. *et al. Network of Automatic Weather Stations: An AICRPAM-NICRA Initiative* 40 (ICAR-Central Research Institute for Dryland Agriculture, 2018).
44. De Wit, C. T., Goudriaan, J. & VanLaar, H. H. *Simulation of Assimilation, Respiration and Transpiration of Crops* 148 (Centre for Agricultural Publishing and Documentation, 1978).
45. Hoogenboom, G., Huck, M. G., ROOTSIMU v. 40. *A Dynamic Simulation of Root Growth, Water Uptake, and Biomass Partitioning in a Soft-Plant-Atmosphere Continuum: Update and Documentation* 109th edn, 83 (Alabama Agr. Exp. Stn., 1986).
46. Parton, W. J. & Logan, J. A. A model for diurnal variation in soil and air temperature. *Agric. Meteorol.* **23**, 205–216 (1981).
47. Piani, C., Haerter, J. O. & Coppola, E. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* **99**, 187–192 (2010).
48. Wilcke, R. A. I., Mendlik, T. & Gobiet, A. Multi-variable error correction of regional climate models. *Clim. Change* **120**, 871–887. https://doi.org/10.1007/s10584-013-0845-x (2013).
49. Fang, G. H., Yang, J., Chen, Y. N. & Zammit, C. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrol. Earth Syst. Sci.* **19**(6), 2547–2559. https://doi.org/10.5194/hess-19-2547-2015 (2015).
50. Rao, A. V. M. S. *et al.* Evaluating area-specific adaptation strategies for maize under future climates of India. *Sci. Total Environ.* **836**, 155511. https://doi.org/10.1016/j.scitotenv.2022.155511 (2022).
51. Moriasi, D. N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50**(3), 885–900 (2007).
52. Lenderink, G., Buishand, A. & van Deursen, W. Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. *Hydrol. Earth Syst. Sci.* **11**(3), 1145–1159 (2007).
53. Déqué, M. *et al.* An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections. *Clim. Change* **81**, 53–70 (2007).
54. Block, P. J., Souza-Filho, F. A., Sun, L. & Kwon, H. H. A stream-flow forecasting framework using multiple climate and hydrological models. *J. Am. Water Resour. Assoc.* **45**(4), 828–843 (2009).
55. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models-I. A discussion of principles. *J. Hydrol.* **10**(3), 282–290 (1970).
56. Willmott, C. J. On the validation of models. *Phys. Geogr.* **2**(2), 184–194 (1981).
57. Legates, D. R. & McCabe, G. J. Jr. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**(1), 233–241 (1999).

Acknowledgements

Author contributions

S.K.B.: Conceptualization, Methodology, Investigation, Formal analysis, Writing, Reviewing, Editing V.P.P.: Data curation, Formal analysis, Investigation, Writing V.M.S.: Reviewing & Editing N.M.: Investigation, Formal analysis, Writing, Reviewing & Editing MAS: Writing, Reviewing & Editing A.V.M.S.: Methodology, Investigation, Reviewing and Editing P.V.K.: Investigation, Reviewing & Editing M.V.: Investigation, Reviewing & Editing V.K.S.: Reviewing, Editing and Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-34194-9.

Correspondence and requests for materials should be addressed to N.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.