Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 25 (4) : 571-576 (December - 2023)
https://doi.org/10.54386/jam.v25i4.2250
https://journal.agrimetassociation.org/index.php/jam

Research Paper

ARIMA approach for temperature and rainfall time series prediction in Punjab

K. K. GILL, KAVITA BHATT, BALJEET KAUR* and SANDEEP SINGH SANDHU

Department of Climate Change and Agricultural Meteorology, Punjab Agricultural University, Ludhiana -141004, Punjab, India
**Corresponding author email: bchahal57@gmail.com*

ABSTRACT

The present study aims to explore the effectiveness of Seasonal Autoregressive Integrated Moving Average (SARIMA) models in forecasting meteorological time series data exhibiting seasonal patterns. We compared the performance of SARIMA models with different configurations and evaluate their forecasting accuracy using real-world meteorological datasets for three different agroclimatic zones of Punjab (sub mountainous region, central region and south west region) was analyzed to forecast mean monthly maximum air temperature, minimum air temperature and rainfall. The weather data was used from 1984-2022 for sub-mountainous zone (Ballowal Saunkhri), 1970-2022 for Central zone (Ludhiana) and 1977-2022 for south west zone (Bathinda). The results provide insights into the suitability and limitations of SARIMA models for meteorological forecasting and offer practical recommendations for practitioners and researchers in the field. The goodness of fit was tested against residuals using Ljung-Box test. The accuracy of the model was tested using Mean Absolute Error (MAE) and root square mean error (RMSE). The model achieved Mean Absolute Errors (MAE) ranging from 0.61 to 0.78 for maximum temperature, 0.74 to 0.49 for minimum temperature, and 32.12 to 45.44 for rainfall, with lower MAE values indicating higher predictive accuracy. The fitted model was able to capture dynamics of the temperature time series and produce a sensible forecast. However, the model was unable to forecast rainfall series efficiently.

Keywords: ARIMA, time series analysis, forecast, SARIMA, temperature, rainfall

Climate dynamics are predominantly governed by fluctuations in temperature and rainfall. While temperature rise can yield mixed impacts on crop yields, it generally results in reduced quality and productivity, notably affecting cereal and food grain crops (Richard *et al.,* 2017). Concurrent elevation in maximum and minimum air temperatures can disrupt vital phenological stages, thereby influencing growth and yield (Bokhari *et al*., 2017). The far-reaching repercussions of extreme weather events necessitate accurate weather prediction, underpinning sound environmental policies and effective communication among stakeholders (Pathak *et al*., 2022). Time series analysis plays a pivotal role in climate change investigations (Balibey and Serpil, 2015), serving as a cornerstone for formulating strategies against weather adversities such as heat stress, cold stress, floods, and droughts.

Forecasting meteorological variables by extrapolating historical time series data is paramount for agrophysical modeling (Krzyszczak *et al*., 2017). Among various forecasting techniques, Auto Regressive Integrated Moving Average (ARIMA) models have gained prominence in recent years, with Auto Regression Integrated Moving Average (ARIMA) and Seasonal Auto Regression Integrated Moving Average (SARIMA) models emerging as optimal choices to comprehend climate variations such as temperature and precipitation (Machekposhti *et al*., 2018). In the context of Punjab, Singh and Sharma (2017) employed ARIMA models to forecast temperature and rainfall patterns, demonstrating accurate short-term predictions. Likewise, Kaur and Verma (2019) utilized ARIMA models to anticipate temperature variations in Punjab, underscoring the significance of incorporating autoregressive and moving average components to capture temperature fluctuations.

Although limited studies have delved into seasonal ARIMA modeling for time series analysis in Punjab, this study leverages seasonal ARIMA models to forecast temperature and rainfall time series, aiming to identify the optimal fitting models.

MATERIALS AND METHODS

Study area and data source

The climate of Punjab is warm and temperate. The

Article info - DOI: *https://doi.org/10.54386/jam.v25i4.2250*
Received: 15 June 2023; Accepted: 9 October 2023; Published online : 30 November, 2023

geographical location of Ballowal Saunkhri is 31° 65''N, 76° 26' E lies above 355 m above mean sea level, Ludhiana is 30°54'' N, 75°48' E lies above 247 m above sea level and Bathinda is 30°17'' N, 74°58' E lies above 211 m above sea level. The average annual temperature of Ballowal Saunkhri is 23.1 °C and total annual rainfall is about 1045 mm, whereas, for Ludhiana the average annual temperature is 23.2 °C and total annual rainfall is about 761 mm and for Bathinda the average annual temperature and total annual rainfall are 24.1 °C and 521.3 mm, respectively. The meteorological data for maximum & minimum temperature and rainfall for three different agroclimatic zones were used for the analysis. The time span of meteorological data used for study is different as per the availability of data from the respective sources. The data was recorded from the meteorological observatories installed at Ballowal Saunkhri, Ludhiana and Bathinda and the data taken was for 39 years (1984-2022) for Ballowal Saunkhri, 53 years (1970-2022) for Ludhiana and 46 years (1977-2022) for Bathinda.

Stationarity of data

The first step in building the model was to work out whether there is any stationarity in the observed data. For this Augmented Dickey Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Mann-Kendall was used to check the stationarity of the weather data. Also, the visual inspection of the ACF (autocorrelation function) and PACF (partial autocorrelation function) indicated whether the series was stationary or not.

ARIMA model description

The fundamental aspect of the ARIMA model involves combining autoregressive (AR) and moving average (MA) polynomials into a complex polynomial formulation (Box *et al.*, 2016), as depicted in equation (1)

$$y_t = \mu + \sum_{i=1}^{p}\left(\sigma y_{t-1}\right) + \sum_{i=1}^{q}\left(\theta \varepsilon_{t-1}\right) + \varepsilon_t \qquad (1)$$

The ARIMA (p, d, q) model is then employed across all data points within the time series dataset. Here's a breakdown of the notation: μ represents the mean value of the time series data; p stands for the number of autoregressive lags; σ corresponds to the autoregressive coefficients (AR); q denotes the number of lags in the moving average process; ϴ represents the moving average coefficients (MA); Ɛ symbolizes the white noise present in the time series data; and finally, d indicates the number of differences calculated according to equation (2).

$$\Delta y_i = y_i - y_{i-1} \qquad (2)$$

Determining the values of the ARIMA parameters (p, d, q) for both AR and MA involves analyzing the behavior of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). These functions play a crucial role in estimating the parameters that are utilized to forecast data through the application of the ARIMA model.

Calibration and validation

Calibration was done to assess the models for quality and accuracy of prediction. Calibration procedures were carried out to strengthen the model's performance. From the historical observed data, 34 years of data for Ballowal Saunkhri (1984-2017), 48 years for Ludhiana (1970-2017) and 41 years for Bathinda (1977-2017) were used for calibration of the model, while the remaining weather data was used to validate the model (2018–2022) for all the locations. The models were evaluated and validated using the root mean square error (RMSE), mean absolute error (MAE) and AICC.

RESULTS AND DISCUSSION

The monthly weather data was used to compute the descriptive statistics of temperature. The statistical parameters of maximum and minimum temperature and rainfall, i.e., minimum (Min.), maximum (Max.), mean (M), standard deviation (SD), coefficient of skewness (CS) and coefficient of kurtosis (CK) were calculated to describe the characteristics of temperature over the Ballowal Saunkhri, Ludhiana and Bathinda of Punjab (Table 1). The value of Min. for maximum temperature was ranged from 14.6 (Ludhiana) to 15.9 (Ballowal Saunkhri), whereas, for minimum temperature it ranged from 1.3 (Bathinda) to 3.4 (Ludhiana). Statistical parameter Max. of maximum temperature and minimum temperature was recorded highest for Bathinda and lowest for Ballowal Saunkhri, whereas, the value of Max. for rainfall ranged between 103.3 (Bathinda) to 769.6 (Ballowal Saunkhri). SD of maximum temperature varies between 6.2 to 7.2 for minimum temperature was 7.3 to 8.2 and for rainfall it ranged between 16.6 to 120.8.To assess the adherence of the annual rainfall data to a normal distribution, the skewness and kurtosis were calculated. A kurtosis value of zero characterizes the standard normal distribution. Positive kurtosis reflects a distribution with a sharper peak, while negative kurtosis signifies a more spread-out distribution. The value of Kurtosis for Ballowal Saunkhri was -0.9 for maximum temperature, 4.6 for minimum temperature and -1.4 for rainfall, for Ludhiana it was -1.1 for maximum temperature, -1.5 for minimum temperature and 7.5 for rainfall and for Bathinda it was -1.0 for maximum temperature, -1.4 for minimum temperature and 10.8 for rainfall. Skewness value between -0.5 to 0.5 indicates that the data is symmetrical. The value of Skewness for maximum temperature was ranged between -0.35 (Ballowal) to -1.4 (Ludhiana and Bathinda), for minimum temperature it ranged between -0.1 (Ludhiana and Bathinda) to -0.2 (Ballowal Saunkhri) and for rainfall between 2.1 (Ballowal Saunkhri) to 3.05 (Bathinda).

Stationary time series is an essential step to develop and test SARIMA model. Therefore, analysis of the data was done to test and confirm the stationarity of the time series using ADF, KPSS and Mann-Kendall test. Data indicated that all the test data were stationary (Table 2). However, the rainfall data showed seasonality after decomposing the time series which was removed. Augmented Dickey-Fuller test showed that the maximum and minimum temperature and rainfall for all three zones (Ballowal Saunkhri, Ludhiana and Bathinda) were stationary at 5% significance level. It ranged between -20.44 to -22.62 for maximum temperature, -13.42 to -26.37 for minimum temperature and -11.99 to -15.32 for rainfall data. KPSS test showed maximum temperature ranged between 0.01 to 0.04, minimum temperature between 0.01 to 0.12 and rainfall between 0.01 to 0.06, whereas, Man-Kendall test ranged between

Table 1: Descriptive statistics of temperature and rainfall series in Punjab

Statistics	Ballowal Saunkhri			Ludhiana			Bathinda		
	Max. temp.	Min. temp.	Rainfall	Max. temp.	Min. temp.	Rainfall	Max. temp.	Min. temp.	Rainfall
Min.	15.9	2.2	0.0	14.6	3.4	0.0	15.2	1.3	0.0
Max.	41.3	27.3	769.6	42.5	28.5	668.6	45.1	38.3	103.3
Mean	29.9	16.3	86.1	29.7	16.8	63.4	31.2	16.9	9.4
Standard deviation	6.2	7.3	120.8	6.8	7.7	96.5	7.2	8.2	16.6
Kurtosis	-0.9	4.6	-1.4	-1.1	-1.5	7.5	-1.0	-1.4	10.8
Skewness	-0.35	-0.2	2.1	-0.4	-0.1	2.5	-0.4	-0.1	3.05
Range	25.4	25.1	769.6	27.9	25.1	668.6	29.9	37.0	103.3
Coefficient of variation (CV)	20.8	24.3	39.5	23.0	25.8	51.6	22.9	28.1	87.3

Table 2: Stationarity test of temperature and rainfall time series for Ballowal Saunkhri, Ludhiana, Bathinda and of Punjab

Test	Location	*p*-value			Interpretation
		Max. temp.	Min. temp.	Rainfall	
Augmented Dickey-Fuller	Ballowal Saunkhri	-20.44*	-13.54*	-12.70*	Stationary
	Ludhiana	-22.62*	-26.37*	-15.32*	Stationary
	Bathinda	-21.26*	-13.42*	-11.99*	Stationary
Kwaitkowski-Phillips-Schmidt-Shin	Ballowal Saunkhri	0.04	0.02	0.06	Stationary
	Ludhiana	0.01	0.01	0.05	Stationary
	Bathinda	0.04	0.12	0.01	Stationary
Mann-Kendall	Ballowal Saunkhri	0.26	0.59	0.18	Stationary
	Ludhiana	0.91	0.69	0.33	
	Bathinda	0.33	0.53	0.98	Stationary

*significant at p=0.05

0.26 to 0.91 for maximum temperature, 0.53 to 0.69 for minimum temperature and 0.18 to 0.33 for rainfall.

Further, stationarity was confirmed by the autocorrelation plots (Fig.1-3) which indicated rapid decay of spikes toward zero. Also, AR and MA components of ARIMA were achieved using the ACF (Auto correlation function) and PACF (Partial Auto Correlation Function) plots. The model was calibrated using the training data (1984-2017, 1970-2017 and 1977-2017 for Ballowal Saunkhri, Ludhiana and Bathinda, respectively). The best fit models with lowest AIC selected for all the three weather parameters. Starting from SARIMA (1,0,1)(2,0,0)12, final models given in Table 4 by, at each step, were selected dropping the term with the highest AICC value associated to it and re-estimating the remaining parameters until all the lowest AICC values for all estimated parameters were achieved. The same models were used to validate using test data (2018-2022). Then the residuals of all the models were diagnosed for the goodness of fit using the Ljung-Box test (Table 3). The results of the test were insignificant for SARIMA models for maximum and minimum temperature and rainfall.

Simulated data on different weather parameters i.e. maximum temperature, minimum temperature and rainfall was compared with observed data for different districts of Punjab (Fig. 1). For Ludhiana, the model predicted the values of maximum temperature, minimum temperature and rainfall as 30.2 ^{0}C, 16.3 ^{0}C and 56.5 mm corresponding to the observed values of 29.8 ^{0}C, 18.0 ^{0}C and 65.8 mm, respectively. For Ludhiana, the R^2 value for all the weather parameters ranged from 85 to 98 % and RMSE ranged from 1.64 to 39.85. The predicted values showed very close proximity with the observed values and these trends showed that the model was able to simulate the temperatures and rainfall reasonably well for Ludhiana region, where, the maximum temperature overestimated while the minimum temperature and rainfall are underestimated.

The model satisfactorily simulated both the temperatures and rainfall for Bathinda region (Fig. 1). Both the temperatures and rainfall simulated by the model correspond reasonably well with those actually observed for the region. The model overestimated the minimum temperature (17.6 ^{0}C) and rainfall (18.8 mm) as compared to observed 17.3 ^{0}C and 15.6 mm, respectively. Whereas, maximum temperature was underestimated by the model i.e. 30.2 ^{0}C corresponding to 30.9 ^{0}C. The R^2 value ranged from 69 % to 82 %.

The comparison of observed and predicted values for maximum temperature, minimum temperature and rainfall revealed both over-estimation and under-estimation by the model (Fig. 1). The observed as well as simulated values were underestimated for minimum temperature and rainfall while overestimated for maximum temperature. Overall, the simulated values deviated by 0.006%, 0.0411 % and 0.150% for maximum temperature,

Table 3: Statistical value of Ljung-Box test for residuals

Weather parameter/Region	Maximum temperature	Minimum temperature	Rainfall	Interpretation
Ballowal	0.11	0.52	0.10	
Ludhiana	0.18	0.19	0.64	Does not show lack of fit
Bathinda	0.22	0.06	0.29	

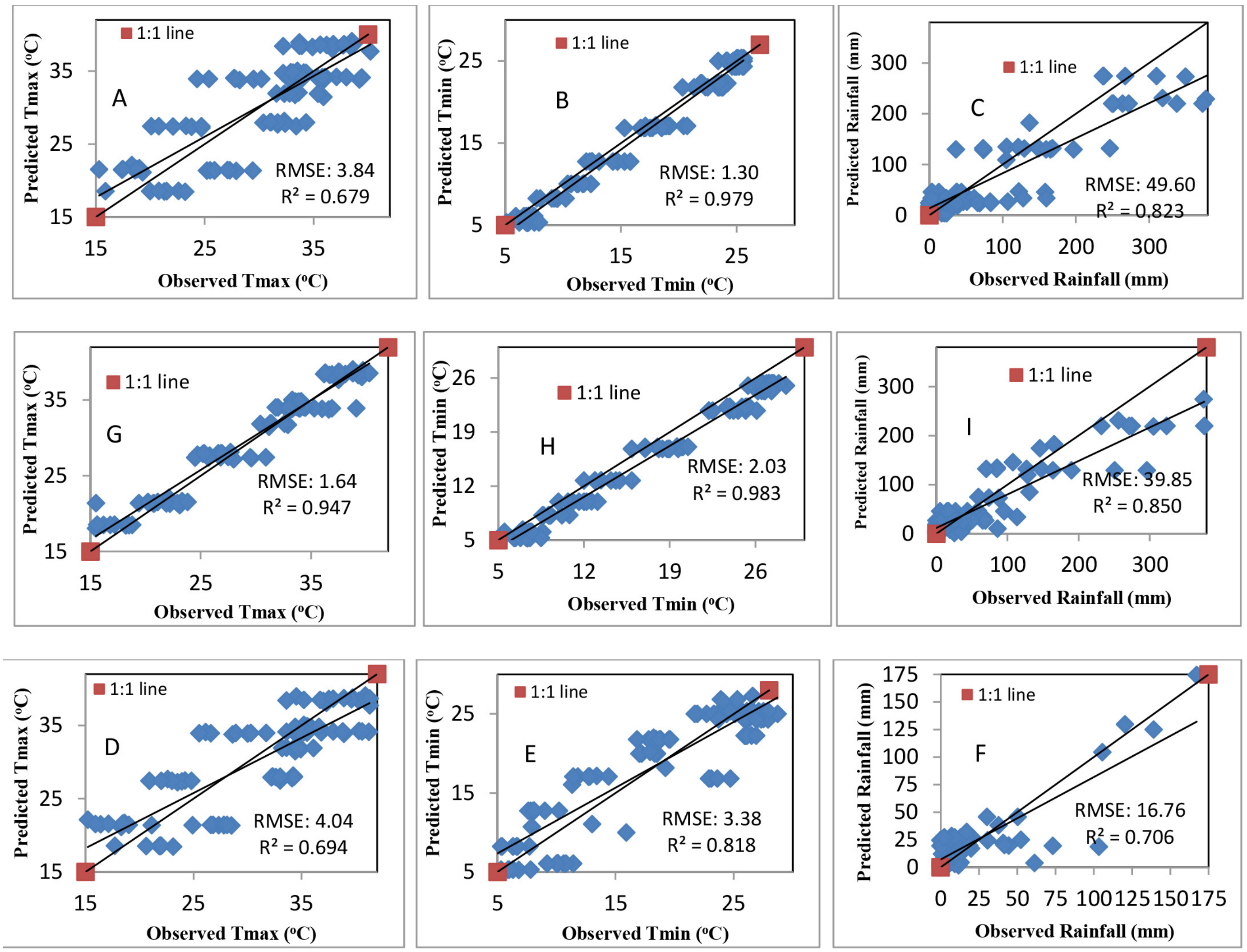

Fig.1: Validation of ARIMA model for maximum temperature, minimum temperature and rainfall for Ballowal Saunkhri (A, B, C), Ludhiana (G, H, I) and Bathinda (C, D, F) districts of Punjab

minimum temperature and rainfall, respectively under Ballowal region. The validation results obtained with the model for all the regions demonstrated satisfactory predictions of weather parameters with ARIMA approach.

Performance efficiency of the models

The model performance was evaluated using AICC, MAE and RMSE for maximum and minimum temperature and rainfall for all the three zones of Punjab. The values for the respective models are presented in Table 4 and are the evidence of model adequacy. For maximum temperature, the AICC ranged between 142.2 (Bathinda) to 142.9 (Ballowal Saunkhri), for minimum temperature 109.8 (Ballowal Saunkhri) to 191.8 (Ludhiana) and for rainfall it ranged between 431.8 (Bathinda) to 489.2 (Ludhiana). The MAE was 0.61 (Ballowal Saunkhri) to 0.78 (Bathinda) for maximum temperature, for minimum temperature 0.74 (Ballowal Saunkhri) to 0.49 (Ludhiana and Bathinda) and for rainfall it ranged between 32.12 (Ballowal Saunkhri) to 45.44 (Ludhiana). Closer the MAE is to zero, more accurate is the model. Therefore, the model was more accurate for maximum and minimum temperature than rainfall which may be due to the complex seasonality of rainfall.

Finally, the simulations of monthly maximum and minimum temperature for next ten years (2023-2032) are shown in Fig. 2. The shaded part represents the 95 % confidence interval. The predicted values are well fitted in the confidence interval, meaning that maximum and minimum air temperatures are predicted to be

Table 4: Best autoregressive integrated moving average models and goodness of fit statistics

Region	Weather element	Best model	AICC	MAE	RMSE
Ballowal Saunkhri	Maximum temperature	(1,0,1) (2,1,0) 12	142.4	0.61	1.88
	Minimum temperature	(1,0,2) (2,1,0) 12	109.8	0.47	1.64
	Rainfall	(1,0,2) (2,1,0) 12	445.1	32.12	1.45
Ludhiana	Maximum temperature	(2,0,2) (1,1,0) 12	142.9	0.76	1.72
	Minimum temperature	(1,0,0) (2,0,0) 12	191.8	0.49	1.69
	Rainfall	(3,1,3) (3,2,2) 12	489.2	45.44	1.50
Bathinda	Maximum temperature	(1,0,1) (1,1,1) 12	142.2	0.78	1.41
	Minimum temperature	(1,0,1) (4,1,1) 12	114.5	0.49	1.78
	Rainfall	(1,0,2) (2,1,0) 12	431.8	41.51	1.71

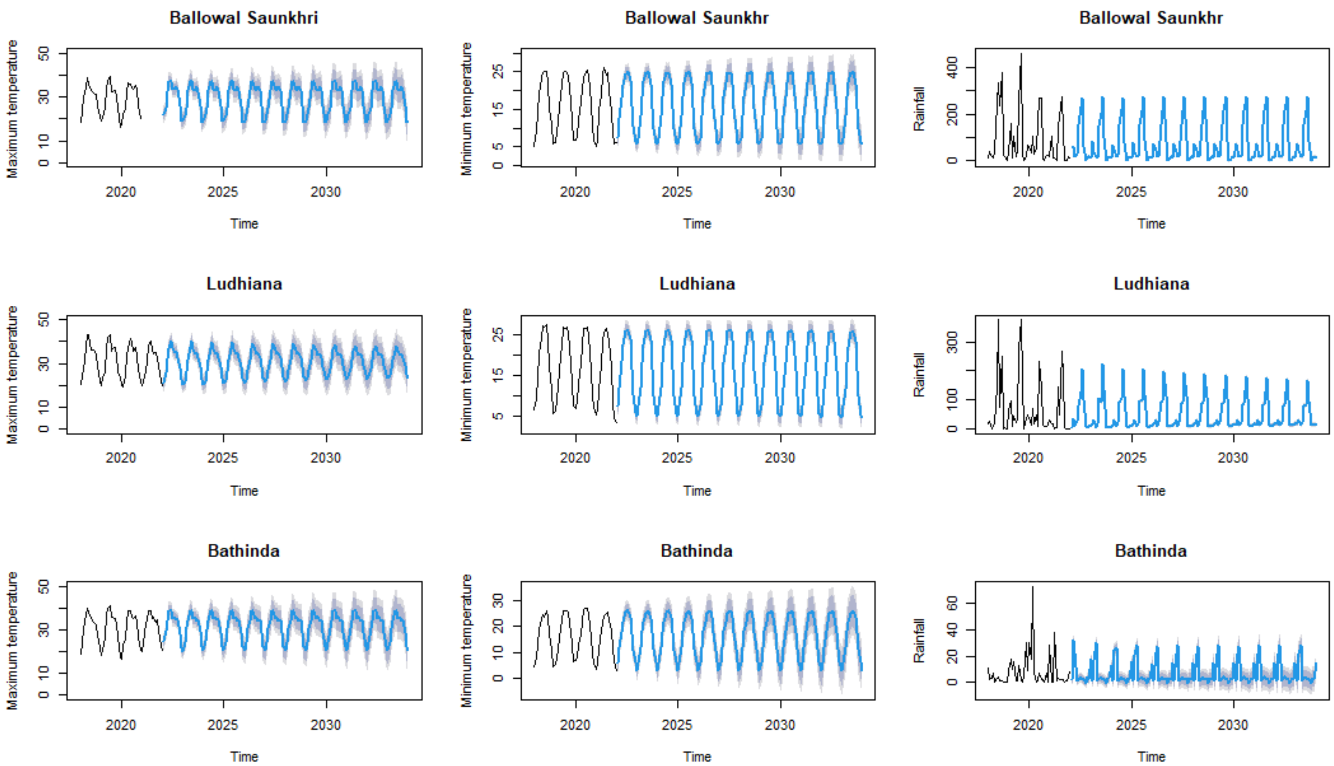

Fig. 2: Observed and forecast series of maximum and minimum air temperature and rainfall for Punjab

stable with same pattern during upcoming 10 years. The model over-predicted the minimum air temperature for Ludhiana, Bathinda and Ballowal Saunkhri. However, the maximum air temperature and rainfall were under-predicted by the seasonal ARIMA models for all the three districts. The results were in agreement with the earlier findings of Gorantiwar *et al.,* (2011) and Kumar *et al.,* (2013).

Limitation of the study

The model's effectiveness in predicting rainfall for all three Punjab regions was compromised, likely due to the presence of intricate and overlapping seasonal patterns. Rahman and Dutta (2020) highlighted the limitation of ARIMA models in comprehensively capturing the intricate seasonal dynamics, leading to constrained predictive accuracy. Furthermore, the intricate non-linear relationships characterizing rainfall patterns, influenced by various factors, posed a challenge for ARIMA models with their inherent linear nature, limiting their capacity to adequately model these intricate dynamics (Hossain and Rahman, 2016).

CONCLUSIONS

In the present study, it has been concluded that the SARIMA models can efficiently forecast the long time series. The goodness of fit confirms the adequacy of the models. The RMSE was between 1.41 and 1.88. The maximum temperature exhibited Mean Absolute Errors (MAE) ranging from 0.61 (Ballowal Saunkhri) to 0.78 (Bathinda). For minimum temperature, MAE ranged from 0.74 (Ballowal Saunkhri) to 0.49 (Ludhiana and Bathinda), while rainfall predictions had MAE ranging between 32.12 (Ballowal Saunkhri) and 45.44 (Ludhiana), with lower values indicating higher model accuracy. Although the chosen models cannot predict the exact temperatures, they can only provide the information that helps policy makers to create better strategies for agriculture and to set up priorities for coping against upcoming weather changes. Such forecasting tools will play a pivotal role in helping societies prepare for and adapt to changing weather conditions. As such, continued research and development in this field is not just beneficial—it's crucial.

ACKNOWLEDGEMENT

The present work was carried out under the Strategic Research Component of Gramin Krishi Mausam Sewa (GKMS). The authors would like to thank India Meteorological Department and Punjab Agricultural University, Ludhiana for providing weather data and their support to carry out this research work.

Funding: The authors declare no specific funding for this project.

Conflict of interest: The authors declare no conflict of interest related to this article.

Data availability: Data can be provided on request

Author's contribution: **K. K. Gill**: Conceptualization, writing, reviewing; **K. Bhatt:** Data collection, analysis and interpretation; **B. Kaur**: Data collection, analysis and interpretation; **S. S. Sandhu**: Writing, reviewing.

***Disclaimer*:** The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Balibey, M. and Serpil, T. (2015). A Time series approach for precipitation in Turkey. *GU. J. Sci.* 28(4): 549–559

Bokhari, S. A. A., Rasul, G., Ruane, A. C., Hoogenboom, G., Ahmad, A. (2017). The Past and Future Changes in Climate of the Rice-Wheat Cropping Zone in Punjab, Pakistan. *Pak. J. Meteorol.,*13(26): 9-23

Box, G.E., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. (2016). Time series analysis: Forecasting and control (4th ed.). John Wiley & Sons.

Gorantiwar, S. D., Meshram, D.T. and Mittal, H. K. (2011). Seasonal ARIMA model for generation and forecasting evapotranspiration of Solapur district of Maharashtra; *J. Agrometeorol*., 13(2): 119-122. https://doi.org/10.54386/jam.v13i2.1354

Hossain, F., and Rahman, M. A. (2016). Rainfall prediction using ARIMA: A case study of the southwestern region of Bangladesh. *J. Earth Sci. Clim. Change., 7*(2): 330.

Kaur, K., and Verma, N. (2019). Forecasting of daily temperature of Ludhiana city using ARIMA model. *Int. J. Res. Elect. Computer Eng*., 7(2): 122-125.

Krzyszczak, J., Baranowski, P., Hoffmann, H., Zubik, M. and Sławiński, C. (2017). Analysis of Climate Dynamics Across a European Transect Using a Multifractal Method, In: Advances in Time Series Analysis and Forecasting (Eds I. Rojas, H. Pomares, O. Valenzuela). Selected Contributions from ITISE 2016. Springer Int. Publishing, Cham., doi:10.1007/978-3-319-55789-2_8

Kumar, A.J., M. Muralidhar, M. Jayanthi, And M. Kumaran. (2013). Trend analysis of weather data in shrimp farming areas of Nagapattinam district of Tamil Nadu. *J. Agrometeorol.,* 15(2): 129–134. https://doi.org/10.54386/jam.v15i2.1459.

Machekposhti, H. K., Sedghi, H., Telvari, A. and Babazadeh, H. (2018). Modelling Climate Variables of Rivers Basin Using Time Series Analysis (Case Study: Karkheh River Basin at Iran). *Civil Eng. J.,* 4(1): 78–92

Pathak, M., Slade, R., Shukla, P.R., Skea, J., Pichs-Madruga, R., and Ürge-Vorsatz, D. (2022). Technical Summary. In: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926.002

Rahman, M. M., and Dutta, D. (2020). Prediction of Monthly Rainfall Using ARIMA Model for Chittagong Region of Bangladesh. *American J. Water Sci. and Eng., 6*(4): 79-84.

Richard, M., Adams, Brian, H. H., Stephanie, L., and Neil, L. (2017). Effects of global climate change on agriculture: an interpretative review. *Clim. Res.,*11:19–30

Singh, P., and Sharma, D. (2017). Application of ARIMA model for forecasting weather parameters: A case study of Patiala district (Punjab). *Int. J. Applied Res.,* 3(4): 44-47.