Journal of Agrometeorology
ISSN : 0972-1665
Vol. No. 24 (1) : 42-49 (March 2022)
https://journal.agrimetassociation.org/index.php/jam

Research Paper

Forecasting mean monthly maximum and minimum air temperature of Jalandhar district of Punjab, India using seasonal ARIMA model

BALJEET KAUR[1*], NAVNEET KAUR[2] , K.K. GILL[3], JAGJEEWAN SINGH[4], S.C. BHAN[5] and SUNAYAN SAHA[6]

[1]*PAU-Krishi Vigyan Kendra, Jalandhar, Punjab, India*
[2]*PAU-Regional Research Station, Ballowal Saunkhri, Punjab, India*
[3]*Punjab Agricultural University, Ludhiana, Punjab, India*
[4]*PAU-Krishi Vigyan Kendra, Barnala, Punjab, India*
[5]*AAS Division, India Met Deptt, New Delhi, India*
[6]*ICAR-CPRI, Jalandhar, Punjab, India*
***Corresponding author email :** bchahal57@gmail.com*

ABSTRACT

The long-term air temperature data series from 1971-2019 was analyzed and used for forecasting mean monthly air temperature at the district level. The Augmented Dickey-Fuller test, Kwiatkowski-Phillips-Schmidt-Shin test, and Mann-Kendall test were employed to test the stationarity and trend of the time series. The mean monthly maximum air temperature did not show any significant variation while an increasing trend of 0.04°C per annum was observed in mean monthly minimum air temperature, which was detrended. Box-Jenkins autoregressive integrated moving–averages were used to forecast the forthcoming 5 years (2020-2024) air temperature in the district Jalandhar of Punjab. The goodness of fit was tested against residuals, the autocorrelation function, and the histogram. The fitted model was able to capture dynamics of the time series data and produce a sensible forecast.

Keywords : Time series, forecast, Box-Jenkins, air temperature, model

The air temperature at the Earth's surface is one of the most important environmental factors. It is a weather parameter that directly influences the yield and production of crops. All the biological and chemical processes taking place in the soil are affected by air temperature. Deviation of air temperatures from critical limits negatively affects the biochemical processes in cells, limiting the growth of plants and causing death of plants. Increase in temperature is found to cut back yields and quality of many crops, most importantly cereal and food grains (Richard *et al*., 2017) by affecting the growth and development of some critical phenological stages (Bokhari *et al.,* 2017). It also has some direct consequences on animal productivity. Increased thermal stress reduced animal eating and grazing activity (Mader and Davis, 2004) and can cause reductions in productivity and fertility. Almost every aspect of environment is affected by the air temperature variations, which makes it necessary to forecast the upcoming air temperature to adopt suitable mitigation and adaptation strategies. Sound environmental policies are based on modeling the variations of surface air temperature and making dependable forecasts (Romilly, 2005), paving the way for smooth communication among environmental researchers, policy makers, and the general public (IPCC, 2007). In a broader context, air temperature affects many other environmental factors in complex ways. Moreover, air temperature is a critical input parameter in many eco-environmental models in the fields of crop growth simulation (Bechini *et al.*, 2006), agro-ecological zoning (Ye *et al.*, 2008), and food security assessment (Ye *et al.*, 2012), for example.

Historical air temperature records have been used many a time for the statistical modeling of air temperature variations (Rahmstorf *et al.*, 2007). Among several methods available for the time series analysis and forecasting, ARIMA models are widely used by various scientists in recent time, especially last decade. Auto Regression Integrated Moving Average (ARIMA) and Seasonal

Article info - DOI : *https://doi.org/10.54386/jam.v24i1.777*
Received: 5 April 2021; Accepted: 6 January 2022; Published online: 11 February 2022

Auto Regression Integrated Moving Average (SARIMA) were found to be the best fit model to understand the climatic variables (precipitation and air temperature) (Gorantiwar *et al.,* 2011; Kumar *et al.,* 2013; Dwivedi *et al.*, 2017).

A few studies have been done in time series of air temperature using ARIMA in Punjab. Therefore, in this paper, we study the time series of mean monthly maximum and minimum air temperature to examine the statistical properties and to develop the predictive model to forecast the mean monthly maximum and minimum air temperature upto five years ahead using seasonal ARIMA model.

MATERIALS AND METHODS

Study area and data source

The geographical location of Jalandhar is 31°19'32'' N 75°34.75' E. It is classified as central plain zone of Punjab, India. It lies on 242m above mean sea level. The climate of Jalandhar is warm and temperate. The average annual air temperature is 23.9 °C and rainfall is about 769 mm per year. In this study, we obtained air temperature data for 49 years (1971-2019) from Regional Potato Research Institute, Jalandhar and India Meteorological Department for the Jalandhar district of Punjab, India. Fig. 1 gives the sketch of the actual mean monthly maximum and minimum air temperature (1971-2019) data.

Normality of data

In this paper, we used quantile-quantile plot (Q-Q) plot and Shapiro-Wilk normality test to check the normality of the weather data. A Q-Q plot is employed to compare a data set with a distribution, and consists of a scatter diagram of the data set $\{x_1, \ldots, x_n\}$ in ascending order with the values $\{F^{-1}(1/2n), F^{-1}(3/2n), \ldots, F^{-1}(1-1/2n)\}$.

The Shapiro-Wilk test calculates a W statistic that tests whether or not a random sample, x1, x2,…,xn comes from (specifically) a normal distribution. Small values of W indicate departure from normality and percentage points for the W statistic, obtained via Monte Carlo simulations. The test has best results in comparison studies with other goodness of fit tests. The W statistic is calculated as follows:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Where, the x(i) are the ordered sample values (x(1) is the smallest) and the ai are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution.

Stationary of data

The first step in building the model was to establish whether there is any stationarity in the observed data. A stationary time series has constant statistical properties (mean, variance and covariance) over the time which is the only accurate population estimate (Cowpertwait and Metcalfe, 2009). Moreover, the stationary time series is easily predicted and it is a serious assumption made by many models while making predictions from the past values (Nua, 2014). In this study, we used Augmented Dickey Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Mann-Kendall test (Huang *et al.,* 2016) to check the stationarity of the weather data.

ARIMA description

ARIMA model is a valuable approach for predicting the futures of 'time series' variable, which assumes stationarity of the time series. It is consists of three models using auto-regressive, integrated, moving-average (ARIMA) models for time series data. An ARIMA (*p*, *d*, *q*) model can account for temporal dependence in many ways. Firstly, the time series is *d*-differenced to render it stationary. If $d = 0$, the observations are predicted directly, and if $d = 1$, the observations need to be differenced to make it stationary. Secondly, the time dependence of the stationary process {*Xt*} is modelled by including *p* auto-regressive models. The equation for *p* is:

$$X_t = \mathrm{c} + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t, \tag{1}$$

where: c is the constant, *φ* is the parameter of the model, *xt* is the value that observed at *t* and *εt* stands for random error. Thirdly, *q* are moving-average terms, in addition to any time-varying covariates. It takes the observation of previous errors. The equation for *q* is:

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{2}$$

where: *θi* is the parameter of the model, *εt* is the error term. Finally, by combining these three models, we get the ARIMA model. Thus, the general form of the ARIMA models is given by:

$$Y_t = \mathrm{c} + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}, \tag{3}$$

where: *Yt* is a stationary stochastic process, c is the constant, *εt* is the error or white noise disturbance term, *φi* means auto-regression coefficient and *θj* is the moving average coefficient. For a seasonal time series, these steps can be repeated according to the period of the cycle, whatever time interval. Usually, ARIMA models are described using the backward operator *B* defined as:

$$B\mathrm{k}(X\mathrm{t}) = X\mathrm{t\text{-}k} \; t>k;\ t,\ k \in N, \tag{4}$$

where: *k* is the index denoting how many times backward operator *B* is applied to time series *Xt* characterized by time interval *t*, and *N* is the total number of time intervals. By employing the following notation:

Table 1: Descriptive statistics of mean monthly maximum and minimum temperature

Statistics	Maximum temp (^{0}C)	Minimum temp (^{0}C)
Mean	29.6	16.5
Standard Deviation	6.9	7.8
Kurtosis	-0.9	-1.4
Skewness	-0.3	-0.2
Range	29.0	27.1

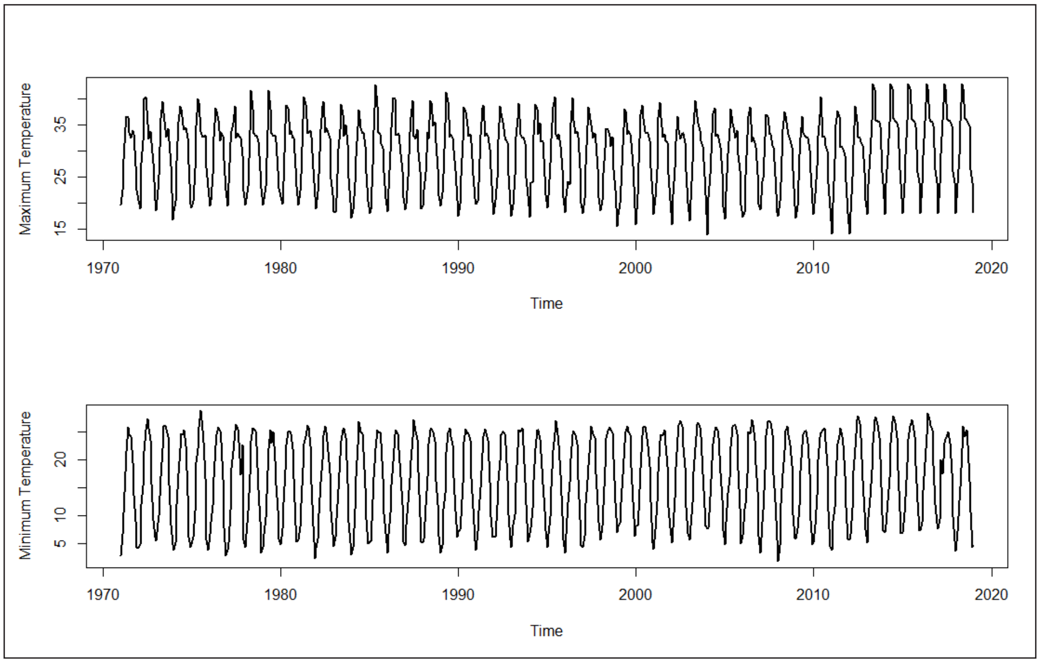

Fig 1: Time series plots of mean monthly maximum and minimum temp (**^{0}C)**

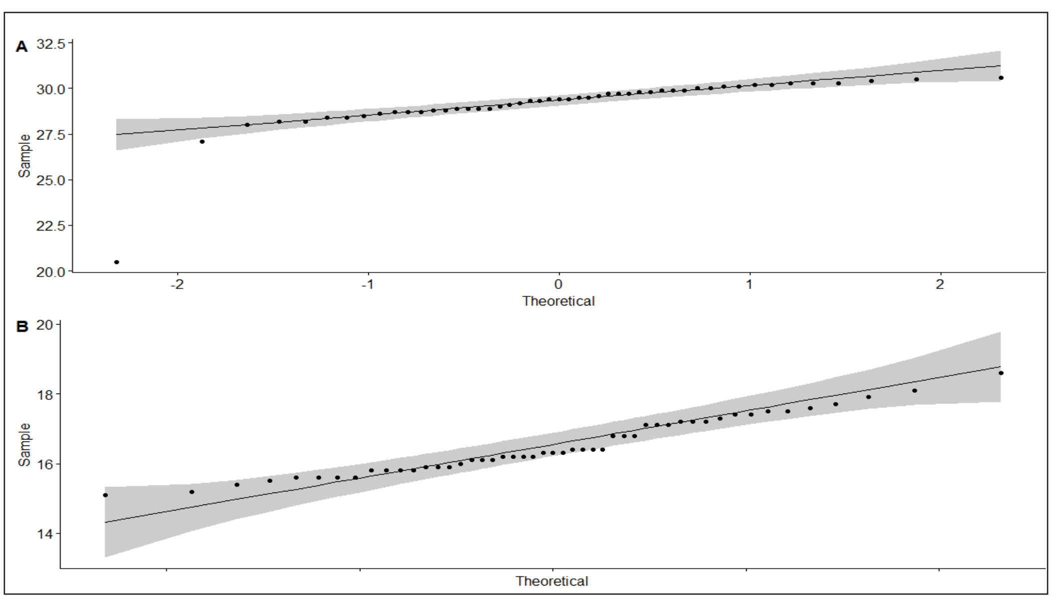

Fig 2: Q-Q plots of mean monthly maximum and minimum temp (**^{0}C)**

Table 2: Stationarity test results

Test	*p*-value (Tmax)	*p*-value (Tmin)	Interpretation
Augmented Dickey-Fuller	-12.602	-12.283	Stationary
Kwaitkowski-Phillips-Schmidt-Shin	0.078	0.158	Stationary

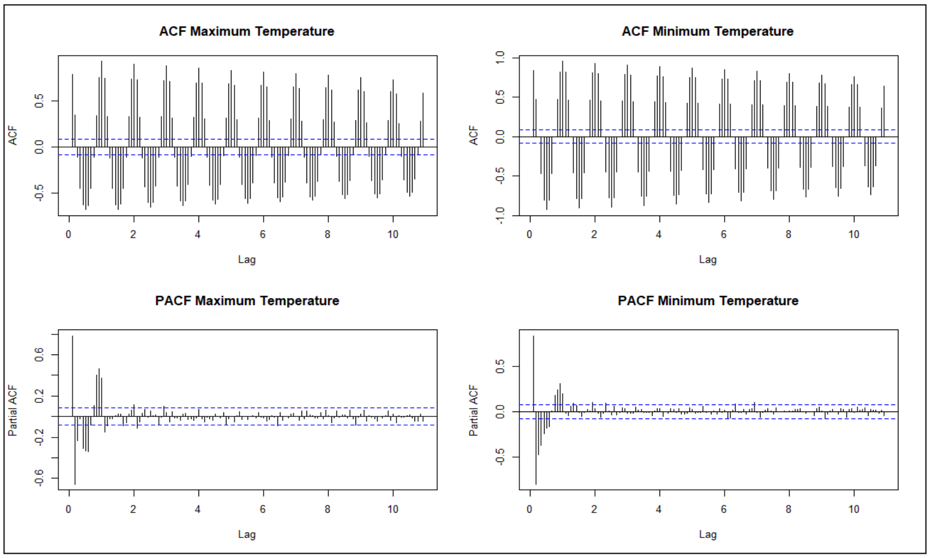

Fig 3: ACF (top) and PACF (bottom) plot of mean monthly maximum and minimum temp (**^{0}C**)

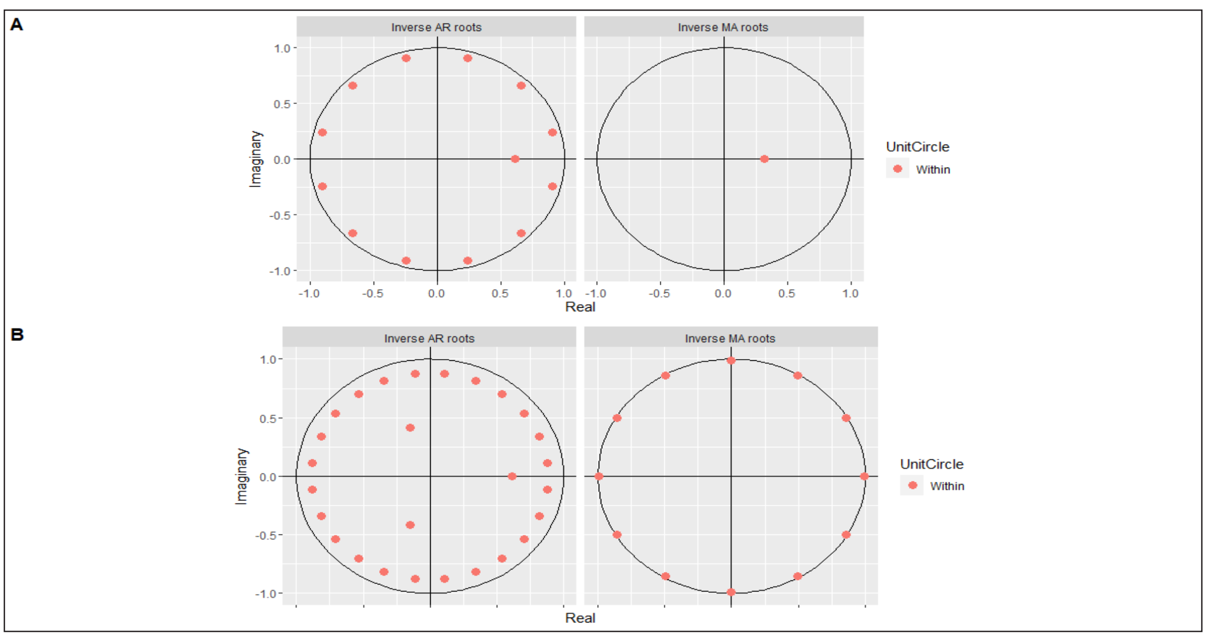

Fig 4: Inverse roots of seasonal ARIMA model for mean monthly maximum temp (A) and minimum temp (B)

$$\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i, \quad \phi_p \neq 0, \tag{5}$$

$$\theta(z) = 1 - \sum_{i=1}^{q} \theta_i z^q, \quad \theta_q \neq 0, \tag{6}$$

Eq. (1) can be written, respectively, as:

$$\phi(B)(1-B)^d Y_t = c + \theta(B)\varepsilon_t. \tag{7}$$

The seasonal ARIMA (*p, d, q*) (*P, D, Q*)*m* process noted also as SARIMA (*p, d, q*) (*P, D, Q*)*m* is given by:

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d Y_t = c + \Theta(B^m)\theta(B)\varepsilon_t, \tag{8}$$

Table 3: Ljung-Box test of residuals

Weather parameter	Maximum temp (^{0}C)	Minimum temp (^{0}C)	Interpretation
p-value	6.48	0.21	Does not show lack of fit

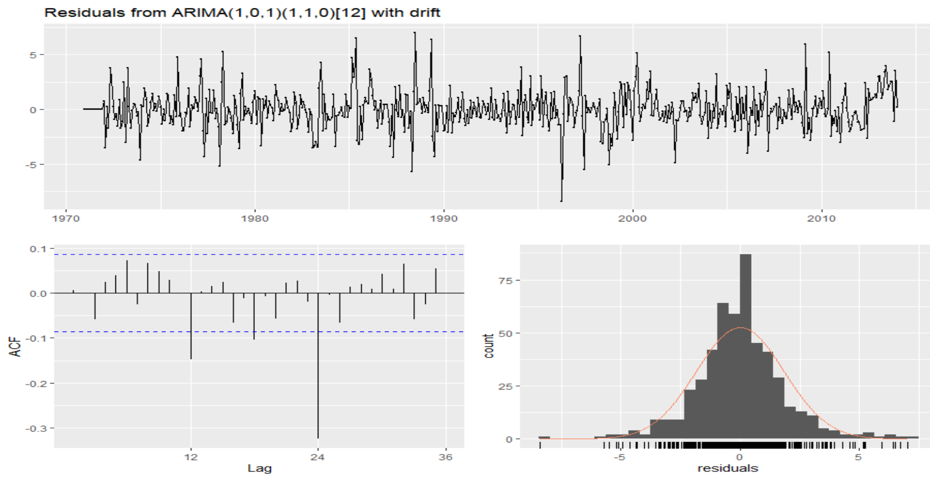

Fig 5: Residuals plots of mean monthly maximum temperature

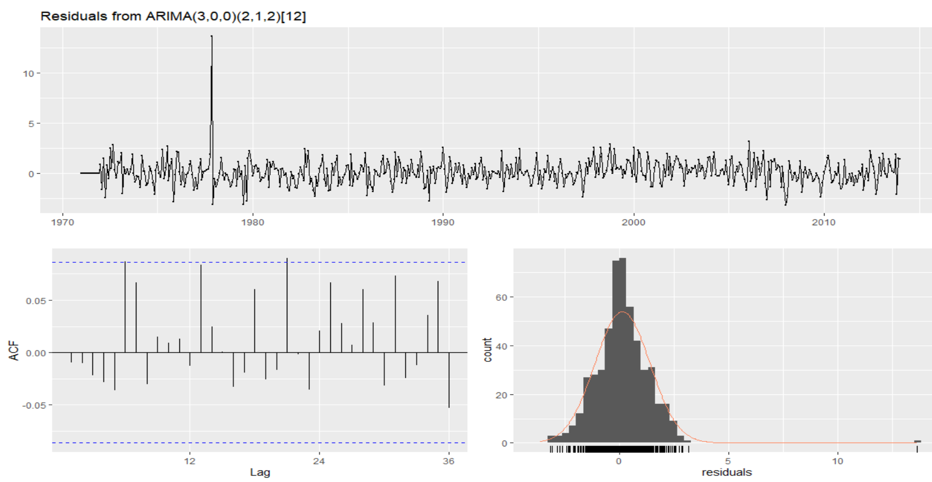

Fig 6: Residual plots of mean monthly minimum temperature

where: *m* is the seasonal period, $\Phi(z)$ and $\Theta(z)$ are polynomials of orders *P* and *Q,* respectively, each containing no roots inside the unit circle. If $c \neq 0$, there is an implied polynomial of order $d + D$ in the forecast function (Brockwell and Davis, 1991).

Model diagnostic

To evaluate the model in order to select the best model for each category of data, the Akaike information criteria corrected (AICC) criteria for selecting the best model was used. AICC was established by AKAIKE (Akaike, 1974) to choose the best model among the class of plausible models. The models with the lowest value of AICC was selected as the most suitable model and used for the forecasting. The equation governing the above mentioned criteria is:

$$AICC = -2 \ln \text{Likelihood}(\hat{\varphi}, \hat{\theta}, \hat{\sigma}^2) + 2(p + q + 1)$$

Where, φ = a class of autoregressive parameters, ˆ θ = a class of moving average parameters, σˆ 2 = variance of white noise, n = number of observations, p = order of the autoregressive component, q = order of the moving average component.

Model estimation

The analysis of model is required for the best fit and can be achieved by detecting the ACF and the PACF of data. Autoarima command was used in R software and it chose model having the lowest AICC value was chosen as the best for Jalandhar station of Punjab.

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

Best model	Goodness of fit statistics	
	Maximum temp	Minimum temp
	(1,0,1) (1,1,0)12	(3,0,0) (2,1,2)12
AICC	59.22	191.81
ME	0.96	0.94
MAPE	-1.32	-1.90
RMSE	1.40	1.69

Calibration and validation

Calibration was done to assess the models for quality and accuracy of prediction. Calibration procedures were carried out to strengthen the models performance. Part of the observed field data, that is data from the meteorological station (from 1971 to 2014) was used for calibration of the model, while the remaining data were used to validate the model (2014–2019).

Performance evaluation criteria

The models were evaluated and validated using the following performance criteria: root mean square error (*RMSE*), sum of squares error (*SSE*), mean square error (*MSE*) and mean absolute percent error (*MAPE*).

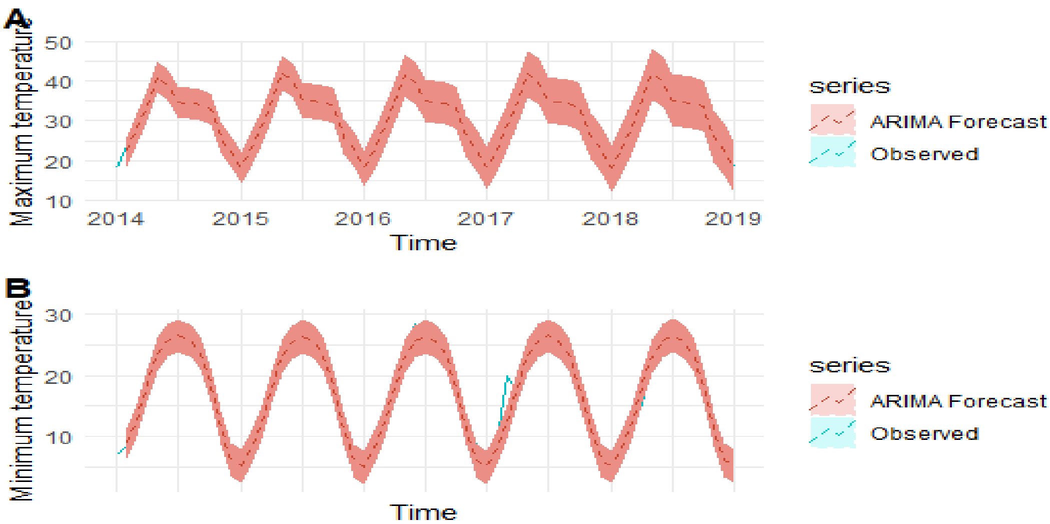

Fig 7: Validation of seasonal ARIMA model for mean monthly maximum temp and minimum temp

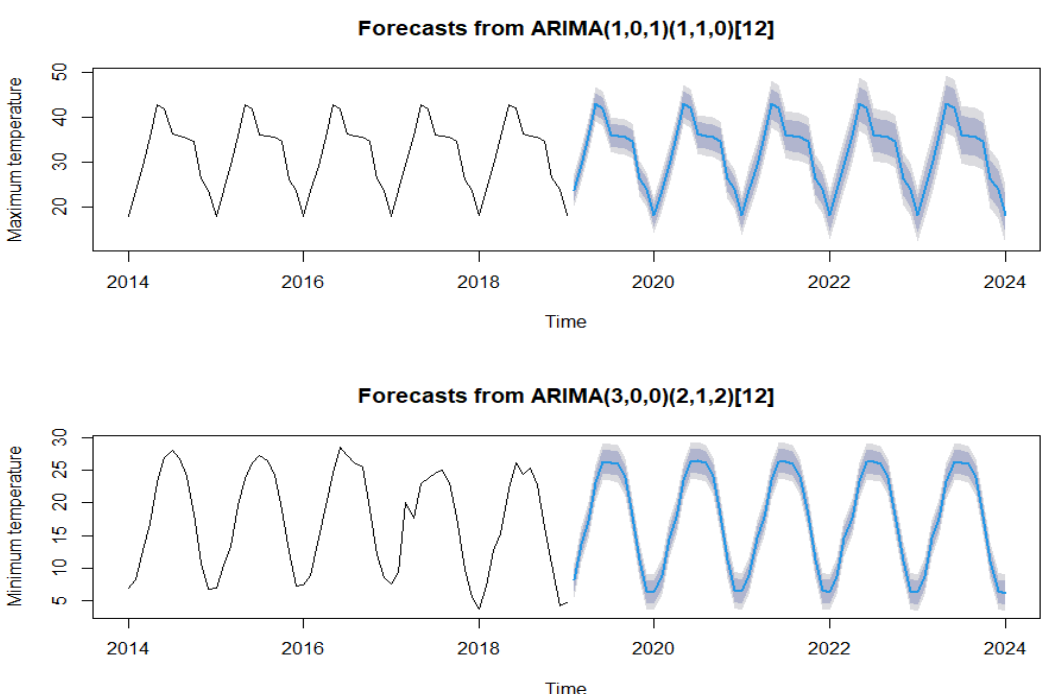

Fig 8: Observed and predicted time series of mean monthly maximum and minimum temp (**^{0}C**)

RESULTS AND DISCUSSION

The mean monthly maximum and minimum air temperature series were used to compute the descriptive statistics of air temperature. The statistical parameters of air temperature, i.e., mean (M), standard deviation (S), coefficient of skewness (CS) and coefficient of kurtosis (CK) were calculated to describe the characteristics of air temperature over the district Jalandhar (Tables 1). Fig. 2 illustrated that the data was normally distributed as all the points fall within the confidence level (grey shaded area).

Stationary time series is imperative to develop and test ARIMA model. Therefore, we analyzed the data to test and confirm the stationarity of the time series using ADF and KPSS test. Both the test confirmed that the data was stationary (Table 2). However, the data showed seasonal trend which was removed first order differencing. Further, stationarity was confirmed by the autocorrelation plots (Fig. 3) due to rapid decay of spikes toward zero. Also, AR and MA components of ARIMA were achieved using the ACF and PACF plots. The model was calibrated using the training data (1971-2014). The auto.arima function of R-software automatically chose the best fit model with lowest AICC. It selected seasonal ARIMA (1,0,1) (1,1,0)12 for maximum air temperature and (3,0,0) (2,1,2)12 for minimum air temperature. The same models were used to validate using test data (2014-2019). We used inverse AR and MA roots to test the stationarity and invertibility of the fitted model. Fig. 4 confirmed that the AR and MA roots are stationary and invertible as the roots of ACF and PACF lied with the circle. Hence, models could be considered to be valid.

Then we diagnosed the residuals of both models using the Ljung-Box test (Table 2). Residuals are the difference between observed and the forecast data. The results of the test were insignificant for ARIMA models of maximum and minimum air temperature. As most of the spikes are close to zero, the residuals therefore, were uncorrelated, producing explicit coverage of the prediction intervals. This implied that for seasonal ARIMA model for mean monthly maximum air temperature, the residuals resembled random white noise as seen from the autocorrelation plots in Fig. 5. However, in Fig. 6, one spike is higher which might be due to the large difference between observed and fitted value. All the other residuals are close to zero. Hence, the residuals of seasonal ARIMA model for mean monthly minimum air temperature also exhibit random white noise.

The time plot and histogram of the residuals showed almost constant variance (Fig 5 and 6). Residuals are the difference of observed and fitted values. Time plot confirm that the residuals have no trend or pattern. It implicates that the forecast value is much close to the observe data, indicating the performance efficiency of the model. Both the models would be useful to forecast the mean monthly maximum and minimum air temperature data. The evidence of normality is ratified by the histogram of the residuals which are almost bell shape.

Performance efficiency of the models

The models performance was evaluated using model efficiency (ME), mean absolute percentages error (MAPE) and root mean square error (RMSE) for both mean monthly maximum and minimum air temperature. The values for the respective models are presented in Table 3 and are the evidence of model adequacy. The resulting time series models were validated with test data set for the period 2014-2019 to test the forecast accuracy of the models. Fig. 7 revealed the excellent agreement between observed data and the seasonal ARIMA model. The forecast data followed the observed mean monthly maximum and minimum air temperature data.

Finally, the simulations of mean monthly maximum and minimum air temperature for next five years (2020-2024) are shown in Fig. 8. The shaded part represents the 95% confidence interval. The predicted values are well fitted in the confidence interval, meaning that both the air temperatures are predicted to be stable with same pattern during upcoming 5 years. The results obtained were in agreement with the earlier findings of Babazadeh and Shamsnia (2014) and Abebe (2020).

CONCLUSION

In the present study, we have concluded that the seasonal ARIMA models can efficiently forecast the long time series. The goodness of fit confirms the adequacy of the models. The maximum and minimum air temperature time series fitted by model revealed that it is possible to predict the mean monthly maximum and minimum air temperature on the basis of historical data. Both the air temperatures for the next five years seem to be insignificantly higher from that of the reference period 1971-2019. Although the chosen models cannot predict the exact air temperatures, they can give us the information that can help policy makers to create better strategies for agriculture and to set up priorities for coping against serious climate change impacts.

Conflict of Interest Statement*:* The author(s) declare(s) that there is no conflict of interest.

Disclaimer: The contents, opinions and views expressed in the research article published in 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

Abebe, T. H. (2020). Time series analysis of monthly average air temperature and rainfall using seasonal arima model (in case of ambo area, Ethiopia), *Int. J. Theoretical. Applied. Maths*.,6(5): 76-87.

Akaike H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr*. 19, 716–723.

Babazadeh, H., and S. A., Shamsnia. (2014). Modeling climate variables using time series analysis in arid and semi arid regions. *African J. Agril. Res.,*9(26): 2018–27.

Bechini, L., S. Bocchi, T. Maggiore, and R. Confalonieri. (2006).

Parameterization of a crop growth and development simulation model at sub-model components level. An example for winter wheat (*Triticum aestivum* L.). *Env. Modelling and Software.*,21**:** 1042–54.

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. M. and Jenkins G.E.P., (1976). "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco., 1976.

Brockwell P.J. and Davis R.A. (1991). Time Series: Theory and Methods. 2nd edition. Springer-Verlag, New York.

Cowpertwait, P. S. and Metcalfe, A. (2009). *Introductory Time Series with R*. New York, NY: Springer-Verlag.

Dwivedi, D. K., Sharma, G. R. and Wandre, S. S. (2017). Forecasting mean air temperature using SARIMA Model for Junagadh City of Gujarat; *Ind. J. Academic. Sci. Res*., 7(4): 183–194.

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

Huang, Y.F., Mirzaei, M. and Yap, W.K. (2016). Flood analysis in Langat River Basin using stochatic model; *Inter. J. Geomat.,* 11: 2796–2803.

IPCC, (2007). Technical summary. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds., Cambridge University Press, Cambridge, 74pp.

Kumar, J. A., Murlidhar, M., Jayanthi, M. and Kumaran, M. (2013). Trend analysis of weather data in shrimp farming areas of Nagapattinam district of Tamil Nadu; *J. Agrometeorol.,* 15(2): 129-134

Mader, T.L. and Davis, M. S. (2004) Effect of management strategies on reducing heat stress of feedlot cattle: feed and water intake. *J. Animal. Sci.*, 82**:** 3077-3087.

Nua, R. (2014). *Statistical Forecasting*, [Online Lecture Notes]. Available online at: http://people.duke.edu/~rnau/411home.htm

Rahmstorf, S., A. Cazenave, J. A. Church, J. E. Hansen, R. F. Keeling, D. E. Parker, and R. C. J. Somerville, (2007) Recent climate observations compared to projections. *Sci.*, 316,

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.

Romilly, P., (2005) Time series modeling of global mean air temperature for managerial decision-making. *J. Env. Mgmnt.*,76: 61–70.

Ye, L., and Coauthors (2012). Climate change impact on China food security in 2050. *Agron. For Sus. Dev.*doi: 10.1007/s13593-012-0102-0.

Ye, L., H. Tang, J. Zhu, A. Verdoodt, and E. Van Ranst.(2008). Spatial patterns and effects of soil organic carbon on grain productivity assessment in China. *Soil Use Mangt.*, 24:80–91.