The Pharma Innovation Journal 2023; 12(8): 715-720

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(8): 715-720

www.thepharmajournal.com
Received: 15-05-2023
Accepted: 18-06-2023

MK Manjunatha
Ph.D., Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Shrikant
Ph.D., Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Rahul Patil
Assistant Professor Dept of Soil and Water, CAE, UAS, Raichur, Karnataka, India

GV Srinivasa Reddy
Associate Professor Dept of Irrigation and Drainage Engineering, CAE, UAS, Raichur, Karnataka, India

BS Polisgowdar
Professor Dept of Irrigation and Drainage Engineering, CAE, UAS, Raichur, Karnataka, India

Corresponding Author:
MK Manjunatha
Ph.D., Scholar, Department of Soil and Water Engineering, CAE, UAS, Raichur, Karnataka, India

Seasonal Arima models used to forecast potential evapotranspiration (PET) for Koppal district

MK Manjunatha, Shrikant, Rahul Patil, GV Srinivasa Reddy and BS Polisgowdar

Abstract
Accurate prediction of potential evapotranspiration (PET) is crucial for effective irrigation management. This article presents a study that utilized seasonal autoregressive integrated moving average (SARIMA) models to forecast PET in Koppal district. The study used maximum and minimum temperature data (°C) along with the Thornthwaite method to estimate PET. Selection of the SARIMA models was based on the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values. The study's findings revealed that the SARIMA models were capable of providing accurate PET forecasts up to one month in advance. Among the different stations, the models for Koppal and Kushtagi demonstrated particular promise. The implications of these findings suggest that employing SARIMA models can significantly enhance irrigation planning and command area management practices in Koppal district, thereby leading to improved water resource management in the region. For each station, specific SARIMA models were selected, namely ARIMA(1,0,1) $(2,1,0)_{12}$ for Koppal, ARIMA(1,0,1) $(2,1,0)_{12}$ for Gangavathi, ARIMA (1,0,1) $(2,1,0)_{12}$ for Kushtagi, and ARIMA (2,0,2) $(2,1,0)_{12}$ for Yelburga. All four models displayed superior results, offering reliable forecasts up to one month ahead. These models have the potential to elevate irrigation planning and command area management practices, leading to more effective water resource management.

Keywords: ACF, PACF, SARIMA, PET and ARIMA

Introduction
Evapotranspiration (ET) represents a crucial element of the hydrologic cycle, encompassing the transfer of water from the Earth's surface to the atmosphere (Asadi *et al.*, 2013) [2]. On a global scale, ET accounts for approximately 60% of the annual precipitation received by the Earth's surface. Its importance extends to various applications, including crop production, water resource management, and environmental evaluation (Aruna *et al.*, 2017) [1]. The adequate supply of water to fulfill the evapotranspiration needs of agricultural crops directly influences the quality and quantity of their yield. Consequently, ET data has gained increasing importance in both irrigation practices and water resources management. Multiple hydrological parameters, such as temperature, relative humidity, solar radiation, and wind speed, regulate the process of ET.

Stochastic models are designed to account for time-dependent variations and incorporate random effects associated with the evapotranspiration (ET) process. These models are fitted to hydrological data or time series, specifically ET series, enabling seamless integration of on-farm systems with the main system and facilitating real-time operation of irrigation systems. To effectively design and plan irrigation systems, it is crucial to develop synthetic or forecast data sets. In this context, one of the most effective models for forecasting time series data is the autoregressive integrated moving average (ARIMA) model. ARIMA models define the forecast of a variable as a linear combination of its past states and previous forecast errors. This makes the ARIMA method a powerful tool for time series modeling and forecasting, applicable to a wide range of time series data. ARIMA models have been extensively employed in hydrological time series modeling in various studies. For instance, Popale and Gorantiwar (2014) [12] utilized ARIMA models to predict rainfall in the Rahuri region of India. Gorantiwar and Patil (2009) [15] analyzed evapotranspiration in the Rahuri region using ARIMA models. Hamdi *et al.* (2008) [9] developed seasonal ARIMA models for the Jordan Valley. Asadi *et al.* (2013) [2] employed ARIMA models to forecast evapotranspiration in humid and semi-humid regions. Salas *et al.* (1980) [13] provided a detailed discussion on time series modeling.

Understanding evapotranspiration is importance for a variety of activities, encompassing watershed management, meteorological and hydrological modeling, as well as water management in irrigated agriculture (Dutta *et al.,* 2016) [5]. Evapotranspiration plays a significant role in determining crop water requirements (CWR). As CWR constitutes over 95% of total ET, comprehending its behavior based on historical data becomes essential for improved water resource management. The primary aim of this study is to create a time series model that can effectively analyze and forecast potential evapotranspiration in the Koppal district.

Materials and Method

Koppal district was formed on August 25, 1997, when it was bifurcated from Raichur district. It is located in the northern part of Karnataka and spans a geographical area of 5,559 sq km. The district comprises four taluks: Koppal, Yelburga, Gangavathi, and Kustagi. Known to be a backward district in the region, Koppal lies between 15°09' to 16° 01' north latitude and 75° 46' to 76° 48' east longitude, situated in the Hyderabad Karnataka region. According to the 2001 census, the district is home to a population of 1,196,089 people, with a population density of 215 people per square kilometer. The area is located in the Tungabhadra sub-basin of the Krishna basin, and the Tungabhadra River flows along its southern boundary in a north-easterly direction. Koppal District borders Bagalkot District to the north, Raichur District to the east, Bellary District to the south, and Gadag District to the west.

The potential evapotranspiration (PET) is calculated using the Thornthwaite method

Potential evapotranspiration $(PET) = 16K \left(\frac{10T}{I}\right)^m$

Where,
T - monthly mean temperature (°C),
I - heat index calculated as the sum of 12 month index values, and m is a coefficient dependent on I.

$$m = 6.75 \times 10^{-7} \cdot I^3 - 7.71 \times 10^{-7} \cdot I^2 + 1.79 \times 10^{-2} \cdot I + 0.492$$

The correction coefficient K is a function of the latitude and month.

Auto correlation test (Box Ljung analysis)

The Box-Ljung Test's null hypothesis (H_0) asserts that our model exhibits a good fit, meaning there is no lack of fit. The alternative hypothesis (H_a) posits that the model does display a lack of fit. When the test yields a significant p-value, it rejects the null hypothesis, indicating that the time series is indeed autocorrelated.

Stationary test (Dickey fuller test)

A time series is considered stationary in the weak sense when its statistical properties, such as means and variance, remain constant over time. Conversely, if the computed p-values exceed 0.05, the time series is categorized as non-stationary. Stationary time series are needed to fit stochastic models.

Description of the stochastic model

Stochastic models, often referred to as time series models, find extensive applications in mathematical, economic, and engineering contexts to study time series data. Time series modeling techniques have demonstrated their efficacy as systematic analytical tools, enabling the simulation and prediction of the behavior of unpredictable hydrological systems. Additionally, they enable the assessment of forecast accuracy, as demonstrated by (Mishra and Desai in 2005) [11].

Autoregressive integrated moving average model

By combining autoregressive (AR) models and moving average (MA) models, we create a distinct and efficient class of time series models known as autoregressive integrated moving average (ARMA) models. Within an ARMA model, the current value of the time series is expressed as a linear combination of 'p' lagged values, a weighted sum of the past 'q' deviations from the mean, and a random parameter. This approach offers a powerful way to capture and explain the patterns and characteristics present in the time series data.

ARIMA models are primarily applied to time series that exhibit stationarity. Nonetheless, they can also be employed for non-stationary data sets by applying differencing to the series. The ARIMA methodology, developed by Box and Jenkins in 1976 [4] emphasizes the analysis of time series' stochastic properties independently, rather than constructing single or simultaneous equation models. This approach provides a valuable tool for understanding and modeling various types of time series data, whether they display stationarity or non-stationarity.

Autoregressive integrated moving average models facilitate the representation of each variable as a linear combination of its own past values and stochastic error terms. The general non-seasonal Autoregressive integrated moving average models combines an autoregressive (AR) component of order 'p' and a moving average (MA) component of order 'q', operating on the 'd^{th}' difference of the time series 'zt'. Consequently, an ARIMA model is characterized by three parameters: 'p', 'd', and 'q', all of which can take non-negative integer values (Mishra and Desai, 2005) [11].

The general non-seasonal ARIMA model can be expressed as follows:

$$\phi(B)\, \nabla^d z_t = \theta(B)\, a_t$$

The non-seasonal AR operator of order 'p' can be represented as ϕ(B), where ϕ is a polynomial of oder p is written as:

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p)$$

The non-seasonal MA operator of order 'q' is denoted as θ(B), where θ is a polynomial of order q is written as:

$$\theta(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots \theta_q B^q)$$

Seasonal Autoregressive integrated moving average model models

Cyclic behavior is frequently observed in many time series. In hydrologic time series, this cyclic behavior often occurs annually due to the earth's rotation around the sun. These types of series are considered cyclically non-stationary. To analyze such series, it is essential to remove the deterministic cyclic effects. One common approach for modeling the remaining stochastic part of the series is the Autoregressive Integrated Moving Average (ARIMA) method. Gorantiwar *et al.* (2011) [6] explained that after eliminating the deterministic cyclic effects, the ARIMA approach can be employed to develop a linear model for the stochastic component of the time series. To address seasonality in time series, Box *et al.* (1994) [3] proposed a standardized version of the ARIMA model, known as Seasonal Autoregressive Integrated Moving Average (SARIMA) models. These models are particularly useful because they can handle non-stationarity in both the seasonal and non-seasonal components of the data (p, d, q) represents the non-seasonal part of the model, (P, D, Q)s represents the seasonal part of the model. In summary, SARIMA models offer a powerful tool for analyzing time series data with cyclic behavior and non-stationarity in both seasonal and non-seasonal components.

$$\emptyset\_p\ (B)\ \Phi\_p\ (B^\wedge s)\ \nabla\_s^\wedge d\ Z\_t = \theta\_p\ (B)\varphi\_Q\ \theta(B^\wedge s)a\_t$$

In a SARIMA model, the parameters are as follows:
Where, p: The order of the non-seasonal autoregression (AR) component, d: The number of regular differences needed to make the series stationary (non-seasonal differencing), q: The order of the non-seasonal moving average (MA) component, P: order of seasonal autoregression, D: The number of seasonal differences needed to make the series stationary (seasonal differencing), Q: The order of the seasonal moving average (MA) component, S: The length of the season, which indicates the number of time steps in a seasonal cycle.

Model recognizing
The next step involves identifying the suitable ARIMA model that best represents the behavior of the time series. This can be achieved by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) (Mishra and Desai, 2005; Hsin-Fu Yeh and Hsin-Li Hsu, 2019) [11, 10]. By analyzing the ACF and PACF, we can determine the appropriate order of the ARIMA model and gain insights into the types of models that may be applicable. Ultimately, the final model selection is based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC).
The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are utilized for model ranking, with the model having the lowest criterion value considered the most suitable. The mathematical expressions for AIC and SBC are as follows.

The AIC and SBC can be expressed mathematically as follows
AIC = -2 * log(L) + 2 * k
SBC = -2 * log(L) + k * log(n)

Where, in the expressions, "k" represents the number of parameters in the model, "L" denotes the likelihood function of the ARIMA model, and "n" stands for the number of observations.

Estimation of parameters
Once the suitable model has been identified, the subsequent step involves estimating the model parameters. Maximum likelihood estimation, a statistical technique that maximizes the likelihood of the data given the model, is used for this purpose. After the parameter estimation, the AR and MA parameters are subjected to statistical significance testing to determine whether they hold substantial statistical value or not.

Checking of Diagnostic
Diagnosing the ARIMA model is a crucial and final step in the model development process. This step involves carefully evaluating the chosen model's appropriateness by inspecting various diagnostic statistics and residual plots to assess whether the residuals exhibit characteristics of white noise. In our study, we utilized the residual autocorrelation function (RACF) to determine whether the residuals conform to the properties of white noise.

Data forecasting
The most suitable models identified from historical data were employed to predict potential evapotranspiration (PET). The essential statistical properties of both observed and predicted data were computed and examined to assess whether the predicted data retained the fundamental statistical characteristics of the observed PET series. Correlation coefficients (R), root mean square error (RMSE), and mean absolute error (MAE) were calculated to evaluate the agreement between the observed and predicted data.

Input Dataset and software
The time series of temperature data set (Max & Min) was taken from the Main Agriculture Research Station (MARS) Gangavathi. The data set were from 1990-2020, out of which 1990-2018 was used for the development of the model and the 2019-2020 was used for the validation purpose. The estimation of Potential evapotranspiration was estimated using MS Excel and SARIMA models were developed in the R studio.

Results and Discussion
Before developing the model, two prerequisite tests were conducted: stationarity and autocorrelation. The Box test was used for the autocorrelation test, and the results are presented in Table 1. The outcomes indicated significant autocorrelation in the data for Koppal, Gangavathi, Kushtagi, and Yelburga, with test statistics of 185.94 (p-value: 0.01), 184.88 (p-value: 0.01), 180.42 (p-value: 0.01), and 183.12 (p-value: 0.01) respectively at a 5% level of significance. Subsequently, the Augmented Dickey-Fuller (ADF) test was performed to assess stationarity. The data was found to be non-stationary, leading to the application of seasonal differencing on the data sets (Table 2).
The initial step in constructing a Box-Jenkins ARIMA model involves identifying the appropriate model. This process includes exploring different orders of the autoregressive (AR) and moving average (MA) parameters, denoted as p and q, respectively. The goal is to select the combination that maximizes the log-likelihood while minimizing the Akaike information criterion (AIC) and Bayesian information criteria (BIC) values. The obtained results for the Koppal, Gangavathi, Kushtagi, and Yelburga stations are presented in Tables 3 and 4. To determine the model, the autocorrelation function (ACF) and partial autocorrelation function (PACF) were plotted (Fig. 1 and Fig. 2). It was found that the data exhibited seasonality, leading to the selection of seasonal ARIMA models with seasonal differencing. The best models chosen for the four stations were as follows: Koppal: ARIMA (1, 0, 1) $(2, 1, 0)_{12}$ with a maximum likelihood value of -1532.07, Gangavathi: ARIMA(1, 0, 1) $(2, 1, 0)_{12}$ with a maximum likelihood value of -1566.40, Kushtagi: ARIMA(1, 0, 1) $(2, 1, 0)_{12}$ with a maximum likelihood value of -1532.32,

Yelburga: ARIMA (2, 0, 2) $(2, 1, 0)_{12}$ with a maximum likelihood value of -1644.31. The estimated parameters for the different stations are presented in Table 4. Additionally, the residuals were obtained by differencing the original series with the fitted series. Table 5 demonstrates that the residuals were found to exhibit characteristics of white noise.

After developing models for four taluks, forecasting was carried out at different lead times (1-6 months). The results in Table 6 indicate that initially, for all stations, the forecast was observed to be good at a 1-month lead time with correlation coefficients of 0.90, 0.96, 0.92, and 0.89 for Koppal, Gangavathi, Kushtagi, and Yelburga, respectively. The root mean square error (RMSE) and mean absolute error (MAE) were found to be at their lowest at 1-month lead time and increased as the lead time increased. These stochastic models were found to be suitable for forecasting up to a 1-month lead time. However, it can be noticed from Table 6 that as the lead time increases, the error rate increases significantly. It can be concluded that Seasonal ARIMA models are well-suited for forecasting at a 1-month lead time for potential evapotranspiration forecasting under the Koppal region. To validate the forecasted data, basic statistical properties were compared with the observed data for the 1-month lead time. The comparison was performed using t-test for means and F-test for standard deviation (Haan 1977)[8], as shown in Table 7. The tcal values related to means were within the t-critical table values (±1.71 for two-tailed at a 5% significance level), indicating that there is no significant difference between the mean values of observed and predicted data. Similarly, the Fcal values for standard deviation were smaller than the F-critical values at a 5% significance level. Hence, the results demonstrate that the predicted data preserves the basic statistical properties of the observed series.

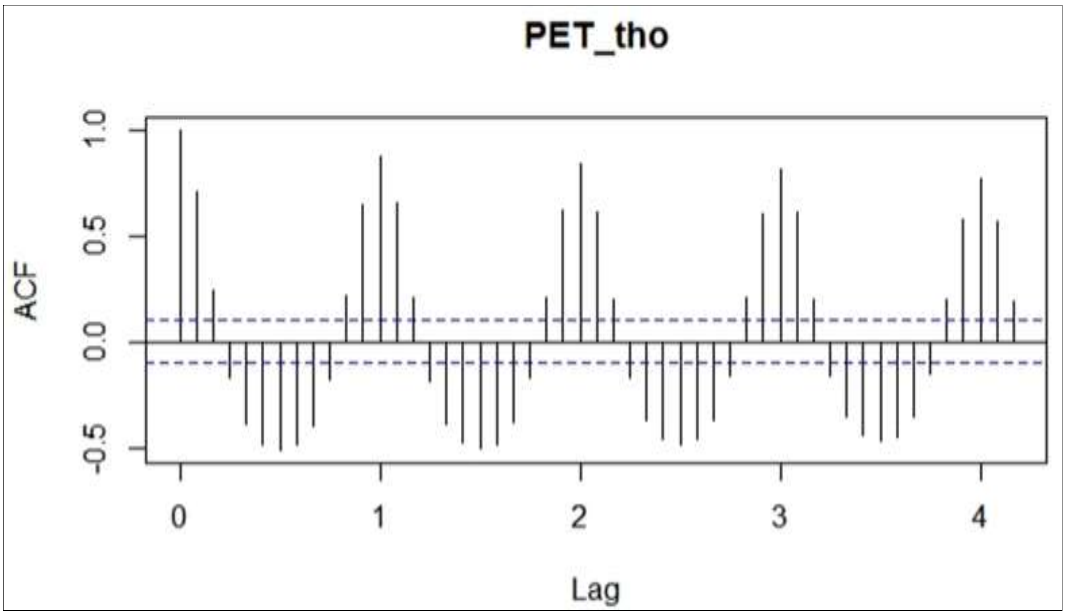

Fig 1: Autocorrelation function plot of PET time series for Koppal Station

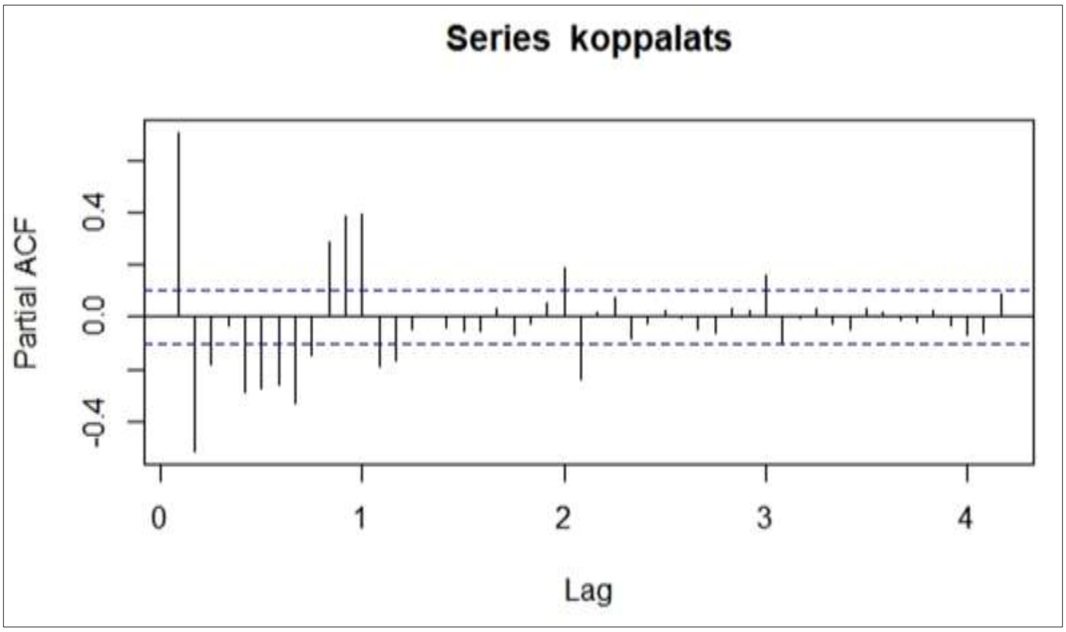

Fig 2: Partial autocorrelation function plot of PET time series for koppal Station

Table 1: Results of the autocorrelation tests for different stations.

Station	Chi-Square	Lag order	P-value
Koppal	185.94	1.0	<0.001
Gangavathi	184.88	1.0	<0.001
Kushtagi	180.42	1.0	<0.001
Yelburga	183.12	1.0	<0.001

Table 2: Stationary tests conducted for different stations

Station	Dickey fuller	Lag order	P-value
Koppal	-16.961	7.0	0.01
Gangavathi	-17.267	7.0	0.01
Kushtagi	-16.632	7.0	0.01
Yelburga	-16.306	7.0	0.01

Table 3: Log likelihood, AIC, and BIC values of the ARIMA models for different stations.

Station	Model	Log-Likelihood	AIC	BIC
Koppal	SARIMA $(1,0,1)(2,1,0)_{12}$	-1532.07	3074.14	3093.4
Gangavathi	SARIMA $(1,0,1)(2,1,0)_{12}$	-1566.4	3142.79	3162.05
Kushtagi	SARIMA $((1,0,1)(2,1,0))_{12}$	-1532.32	3074.64	3093.9
Yelburga	SARIMA $(2,0,2)(2,1,0)_{12}$	-1644.31	3304.62	3335.44

Table 4: Parameter estimations of SARIMA models obtained using the maximum likelihood method for different stations.

Station	Model	Parameter	Estimate	S.E.	Z- value	P- value
Koppal	SARIMA $(1,0,1)(2,1,0)_{12}$	ARI	0.738	0.080	-1.811	<0.001
		MA1	-0.401	0.110	3.324	< 0.001
		SAR1	-0.842	0.049	0.791	< 0.001
		SAR2	-0.406	0.050	-4.096	< 0.001
Gangavathi	SARIMA $(1,0,1)(2,1,0)_{12}$	AR1	0.703	0.089	7.897	< 0.001
		MA1	-0.360	0.119	-3.026	< 0.001
		SAR1	-0.844	0.049	-16.987	< 0.001
		SAR2	-0.413	0.049	-8.315	< 0.001
Kushtagi	SARIMA $(1,0,1)(2,1,0)_{12}$	AR1	0.738	0.080	9.207	< 0.001
		MA1	-0.401	0.110	-3.630	< 0.001
		SAR1	-0.842	0.049	-16.918	< 0.001
		SAR2	-0.406	0.050	-8.133	< 0.001
Yelburga	SARIMA $(2,0,2)(2,1,0)_{12}$	AR1	1.052	0.883	1.191	< 0.001
		AR2	-0.214	0.641	-0.333	< 0.001
		MA1	-0.753	0.882	-0.853	< 0.001
		MA2	0.0948	0.393	0.241	< 0.001
		SAR1	-0.820	0.051	-16.056	< 0.001
		SAR2	-0.391	0.116	-7.666	< 0.001

Table 5: Auto correlation check for residuals of seasonal ARIMA model at different station

Station	Chi-Square	Lag order	P-value
Koppal	0.01	1.0	0.91
Gangavathi	0.007	1.0	0.93
Kushtagi	0.01	1.0	0.90
Yelburga	0.0003	1.0	0.98

Table 6: Performance measures of the seasonal ARIMA model for different stations

Station	Model	Performance measuers	
Koppal	ARIMA $(1,0,1)(2,1,0)_{12}$	RMSE	15.44
		MAPE	7.60
		R	0.90
Gangavathi	ARIMA $(1,0,1)(2,1,0)_{12}$	RMSE	19.49
		MAPE	7.02
		R	0.96
Kushtagi	ARIMA $(1,0,1)(2,1,0)_{12}$	RMSE	15.46
		MAPE	7.62
		R	0.92
Yelburga	ARIMA $(2,0,2)(2,1,0)_{12}$	RMSE	28.93
		MAPE	8.14
		R	0.89

Table 7: Comparison statistic prosperities of the observed and predicted data

Station	Mean observed	Mean forecasted	Decision (t<1.71)	Observed variance	Forecasted variance	Decision (f<4.05)
Koppal	151.54	150.53	0.83	1392.85	994.21	0.32
Gangavathi	159.47	152.99	0.26	5814.43	4236.48	0.30
Kushtagi	152.00	150.66	0.82	4579.89	3472.66	0.33
Yelburga	178.75	172.18	0.45	8433.50	6118.59	0.30

Conclusion

The seasonal ARIMA models were able to forecast potential evapotranspiration accurately up to 12 months ahead, with the best results at the Koppal station. The R, RMSE, and MAPE values for Koppal were 0.90, 15.44, and 7.60, respectively. The seasonal ARIMA models were able to accurately forecast potential evapotranspiration up to one month ahead for all stations, with the least error. The fundamental statistical analysis further revealed that the disparity between the observed and forecasted means was not significant. This implies that the models hold promise for accurately forecasting potential evapotranspiration throughout the study period. Accurate predictions of evapotranspiration are crucial for reliable project planning, design, and efficient operation of irrigation systems.

References

1. Aruna, KT, Satish Kumar U, Ayyanagowdar MS, Srinivasa Reddy GV, Shanwad UK. Estimation of crop coefficient (Kc) values for groundnut crop with evaluated crop evapotranspiration at different moisture deficit levels under agro climatic condition of Raichur, Karnataka. Green Farming. 2017;8(5):1161-1164.
2. Asadi A, Vahdat SF, Sarraf A. The forecasting of potential evapotranspiration using time series analysis in humid and semi humid regions. American Journal of Engineering Research. 2013;2(1):296-302.
3. Box GE, Jenkins GM, Reinsel GC. Time Series Analysis,

Forecasting and Control, (3rd Edition), (Englewood Cliþs, NJ, Prentice-Hall); c1994.
4. Box GE, Jenkins GM. Time series analysis forecasting and control San Francisco. Calif: Holden-Day, San Francisco; c1976.
5. Dutta B, Smith WN, Grant BB, Pattey E, Desjardins RL, Li C. Model development in DNDC for the prediction of evapotranspiration and water use in temperate field cropping systems. Environmental Modelling & Software. 2016;8(1):9-25.
6. Gorantiwar SD, Meshram DT, Mittal HK. Seasonal ARIMA model for generation and forecasting evapotranspirtion of Solapur district of Maharashtra. Journal of Agro meteorology. 2011;13(2):119-22.
7. Gorantiwar SD, Patil PD. Stochastic modelling of crop evapotranspiration for Rahuri region. International Journal of Agricultural Engineering. 2009;2(1):140-145.
8. Haan CT. "Statistical methods in hydrology", Iowa State Press, Iowa; c1977.
9. Hamdi MR, Bdour AN, Tarawneh ZS. Developing reference crop evapotranspiration time series simulation model using Class a Pan: A case study for the Jordan Valley Jordan. Jordan J Earth Environ. Sci. 2008;1(1):33-44.
10. Hsin-Fu Yeh, Hsin-Li Hsu. Stochastic model for drought forecasting in the Southern Taiwan basin. Water. 2019;11(10):1-15.
11. Mishra AK, Desai VR. Drought forecasting using stochastic models. Stochastic environmental research and risk assessment. 2005;19(5):326-339.
12. Popale PG, Gorantiwar SD. Stochastic generation and forecasting of weekly rainfall for Rahuri Region. International Journal of Innovative Research in Science, Engineering and Technology. 2014;3(4):185-96.
13. Salas JD. Applied modeling of hydrologic time series. Water Resources Publication; Highlands Ranch Colorado, USA; c1980. p. 1-482.