

FORECASTING POTENTIAL EVAPOTRANSPIRATION FOR KALABURAGI DISTRICT USING SEASONAL ARIMA MODEL

ABSTRACT

The prediction of potential evapotranspiration (PET) is quite an important task for reliable management of irrigation systems. This article is based on the models which try to mimic the actual occurrence of the potential evapotranspiration in the future days for a Kalaburagi district. In this study, the potential evapotranspiration was estimated with the help of maximum and minimum temperature ($^{\circ}\text{C}$) data using a Thornthwaite method and the prediction was carried out using the Seasonal Autoregressive Moving Average method (SARIMA). The models were developed based upon autocorrelation function (ACF) and partial autocorrelation function (PACF). Furthermore, the model with the least Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values were selected. The models selected for different stations were SARIMA (1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,0)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂ and SARIMA(1,0,1)(2,1,0)₁₂ for Kalaburagi, Chincholi, Sedam, Chittapur, Aland, Jewargi and Afzalpur respectively. The results showed that, the models developed for Jewargi and Chincholi were found to be quite promising compared to the other two stations. All four models were found to be producing better results. The models will provide significant potential in improving the decision making in irrigation planning and command area management practices for better management of water resources.

Keywords: ACF, PACF, SARIMA, PET

INTRODUCTION

Evapotranspiration (ET) is usually the largest component of the hydrologic cycle, given that most precipitation that falls on land is returned to the atmosphere (Asadiet *al.*, 2013). Globally, ET consumes about 60 per cent of the annual precipitation that falls over the earth's

surface. The ET quantifications are used for many purposes including crop production, management of water resources and environmental assessment (Aruna *et al.*, 2017). It is a major component for water balance in the field and needs to be quantified accurately. The amount of water supplied to meet the needs of the agricultural crops for evapotranspiration dictates the quality and quantity of production in a field. The ET data for agricultural crops has become increasingly important in irrigation as well as in water resources management. The process of ET is majorly regulated by many hydrological parameters such as temperature, relative humidity, solar radiation, wind speed and other climatological parameters.

The stochastic models are based on the time dependent variation and consider random effects involved in the ET process. The stochastic linear models are fitted to hydrological data are time series such as evapotranspiration series: enabling the integration of an on-farm system with the main system and facilitating the real-time operation of an irrigation system. It is quite important to develop a synthetic or forecast data set in order to design or plan any irrigation systems; in this context, an autoregressive integrated moving average (ARIMA) model is considered as one of the best model in forecasting the time series dataset. In this model, the forecast of a variable is defined as a linear combination of the previous state of variable and previous forecast error. The ARIMA method is a strong time series modeling and forecasting technique which has versatility to include characteristic of time series. In past, ARIMA models have been widely used to model hydrological time series (Popale and Gorantiwar, 2014). Popale and Gorantiwar (2014) used ARIMA model for prediction of rainfall of Rahuri region, India. Gorantiwar and Patil (2009) have carried out analysis of evapotranspiration for Rahuri region, India. Hamdi *et al.* (2008) developed seasonal ARIMA model for the Jordan valley. Asadi *et al.* (2014) forecasts evapotranspiration for humid and semi-humid region. Salas *et al.* (1980) discussed in detail about time series modelling.

Knowledge of evapotranspiration is important for watershed management activities in meteorological and hydrological modelling, particularly water management in irrigated agriculture (Dutta *et al.*, 2016). The evapotranspiration plays a major role in crop water requirement (CWR) of any crop. As CWR accounts for more than 95% of the ET, it is quite important to understand its behaviour based on the historical data for better management of water resources, in order to understand and solve the irrigation problems. The objective of this study is

to establish a time series model to analyse and forecast potential evapotranspiration for the Kalaburagi district.

MATERIALS AND METHODOLOGY

Kalaburagi is an administrative district in the Indian state of Karnataka and is the largest city in the region of North Karnataka (Kalyana-Karnataka). This district is situated in north Karnataka between 76°04' and 77°42' east longitude and 17°12' and 17°46' north latitude, covering an area of 10,951 km². This district is bounded on the west by Vijayapura district of Karnataka and Solapur district of Maharashtra state; on the north side by Bidar district and Osmanabad district of Maharashtra state; on the south by Yadgir district and on the east by Sangareddy and Vikarabad districts of Telangana state. The entire district is on the Deccan Plateau and the elevation ranges from 300 to 750 m above MSL. The two main rivers, Krishna and Bhima flow through the district.

Thornthwaite method (Potential evapotranspiration)

The potential evapotranspiration is calculated by:

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

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

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

K is a correction coefficient computed as a function of the latitude and month(Shadabet *al.*, 2020).

Auto correlation test (Box Ljung test)

The null hypothesis of the Box Ljung Test, H_0 , is that the model does not show lack of fit (or in simple terms—the model is just fine). The alternate hypothesis, H_a , is just that the model does show a lack of fit. A significant p-value in this test rejects the null hypothesis that the time series isn't an auto correlated (Malik *et al.*, 2021).

Stationary test (Dickey fuller test)

A time series is said to be stationary (in the weak sense) if its statistical properties do not vary with time (means and variance). If the computed p values are greater than 0.05, the series is said to be non-stationary. The time series need to be in stationary form in order to fit to stochastic models (Patil *et al.*, 2022).

Description of the stochastic models

The stochastic models, also referred to as time series models, were used for the study of time series in mathematical, economic and engineering applications. The time series modeling techniques have been shown to provide a systematic analytical tool to simulate and predict the behavior of unpredictable hydrological systems and to measure the predicted accuracy of the forecasts (Mishra and Desai, 2005).

ARIMA models

Autoregressive (AR) models can be considered in conjunction with moving average (MA) models to create a specific and effective class of time series models called autoregressive integrated moving average (ARIMA) models. In an ARIMA model, the present value of the time series is explained as a linear aggregate of 'p' lagged values and a weighted sum of 'q' former deviations plus a random parameter.

An ARIMA models are generally used for a time series which are stationary in nature. However, these models can be used in non-stationary data set by differencing the series. Box and Jenkins (1976) developed a new forecasting tool, known as the ARIMA methodology, that focus on analyzing the stochastic characteristics of time series on its own rather than constructing single or simultaneous equation models.

ARIMA models allow stating each variable by its own lagged values and stochastic error terms. The general non-seasonal ARIMA model is AR to order p and MA to order q and operates on d^{th} difference of the time series z_t ; thus a model of the ARIMA family is classified by three parameters (p, d, q) that can have zero or positive integral values (Mishra and Desai, 2005).

The general non-seasonal ARIMA model may be written as:

$$\phi(B)\nabla_z^d = \theta(B)a_t$$

Where, $\theta(B)$ are polynomials of order p and q respectively. The non-seasonal AR operator of order p can be written as:

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

and non-seasonal MA operator of order q is written as:

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

Seasonal ARIMA models

Many time series features are cyclic. Quite frequently, such characteristics are on an annual period in hydrologic time series mainly due to earth's rotation around the sun. Such types of series are cyclically non-stationary. After removing the determinist cyclic effects from a series, the ARIMA approach may be applied to obtain a linear model for the stochastic part of the series (Gorantiwaret *al.* 2011). Box *et al.* (1994) standardized the ARIMA model to address seasonality and defined a general multiplicative seasonal ARIMA model commonly referred to as SARIMA models. An inherent advantage of the SARIMA family of models is that, the description of time series requires few model parameters, which exhibit non-stationarity both in season and throughout. In general, the SARIMA model described as ARIMA $(p, d, q) (P, D, Q)_s$, where (p, d, q) is the non-seasonal part of the model and $(P, D, Q)_s$ is the seasonal part of the model, which is mentioned below :

$$\phi_p(B)\phi_p(B^s)\nabla^d\nabla_s^D Z_t = \theta_p(B)\theta_Q(B^s)a_t$$

where p is the order of non-seasonal auto-regression, d the number of regular differencing, q the order of non-seasonal MA, P the order of seasonal auto-regression, D the number of seasonal differencing, Q the order of seasonal MA, s is the length of season, seasonal AR parameter of order P , seasonal MA parameter of order Q .

a. Model identification

This step is to identify the possible ARIMA model which represents the time series behavior. The series behavior was investigated based upon the behavior of autocorrelation function (ACF) and partial autocorrelation function (PACF) (Mishra and Desai, 2005; Hsin-Fu Yeh and Hsin-Li Hsu, 2019). The ACF and PACF were used to support in determining the order of the model. The information given by ACF and PACF is useful in suggesting the type of models that may be constructed. The final model was then selected using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Prasad *et al.*, 2019).

These criteria help to rank models (the models with the lowest criterion value being the best). The AIC and SBC take the mathematical form as shown below:

$$\text{AIC} = -2 \log(L) + 2k \text{ and } \text{SBC} = -2 \log(L) + k \ln(n).$$

Where k is number of parameters in the model; L is the likelihood function of the ARIMA model; and n is the number of observations.

b. Parameter estimation

The estimation of model parameters was achieved after identification of the appropriate model as an essential step. The model estimate values for the AR and MA parts were calculated using Maximum likelihood. The AR and MA parameters were tested to make sure that, they are statistically significant or not.

c. Diagnostic checking

Diagnosis of the ARIMA model is a crucial part of model development and the last step. It involves checking the appropriateness of the model chosen. Several diagnostic statistics and residual plots are examined to see whether or not the residuals are correlated to white noise. In this study, we obtained the residual ACF function (RACF) to determine whether residuals are white noise.

d. Drought forecasting

The prediction of potential evapotranspiration was done using the best fit models from historical data. Basic statistical properties of the observed and predicted data were computed and tested whether the predicted data preserve the basic statistical properties of the observed PETseries. The correlation coefficients (R), RMSE and MAE were observed between the observed and predicted data.

Input Dataset and software

The time series of temperature data set (Max and Min) were taken from the Zonal Agricultural Research Station (ZARS)Kalaburagi. The data set were from 1990-2021, out of which 1990-2019 was used for the development of the model and the 2020-2021 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 software.

RESULTS AND DISCUSSION

Development of model was done with prerequisite tests namely stationary and autocorrelation tests. The autocorrelation test was carried out using box test and corresponding probability levels are presented in Table 1. The results revealed that, the test statistic for box.test with a Chi square and P values were 187.17(0.01), 185.64(0.01), 172.77 (0.01), 182.4(0.01), 187.26(0.01), 99.88(0.01) and 189.64(0.01) for Kalaburagi, Chincholi, Sedam, Chittapur, Aland, Jewargi and Afzalpur respectively, were observed to be significant at 5 % level of significance reflecting autocorrelation in data. On the other hand, adf.test was carried out to check whether the data is stationary or not. The data was observed to have seasonality thereby seasonal differencing was done (Table 2) to the data sets.

Table 1: Auto correlation test for different stations of Kalaburgi district

Station	Chi-Square	Lag order	P-value
Kalaburagi	187.17	1	<0.001
Chincholi	185.64	1	<0.001
Sedam	172.77	1	<0.001
Chittapur	182.4	1	<0.001
Aland	187.26	1	<0.001
Jewargi	99.883	1	<0.001
Afzalpur	189.64	1	<0.001

Table 2: Stationarity test for different stations of Kalaburgi district

Station	Dickey fuller	Lag order	P-value
Kalaburagi	-18.022	7	0.01
Chincholi	-17.777	7	0.01
Sedam	-17.193	7	0.01
Chittapur	-17.491	7	0.01
Aland	-18.028	7	0.01
Jewargi	-4.8282	7	0.01
Afzalpur	-18.195	7	0.01

The principal step in Box-Jenkins ARIMA model building is identification of the model. Different orders of Autoregressive (AR) and Moving Average (MA) parameters p and q were considered and combination of the order which yields maximum log-likelihood and lowest values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC)

were considered as best model. The results pertaining to Kalaburagi, Chincholi, Sedam, Chittapur, Aland, Jewargi and Afzalpur stations regarding model development are presented in Table 3 & 4. The ACF and PACF were plotted (Fig. 1 and 2) to determine the model. The data have seasonality; therefore seasonal ARIMA models were selected with a seasonal differencing as shown in Table 4. The best selected models for different stations were: SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,0)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂, SARIMA(1,0,1)(2,1,0)₁₂ and SARIMA(1,0,1)(2,1,0)₁₂ with an maximum likelihood values of -1532.32, -1542.42, -1732.01, -1552.32, -1512.62, -1542.24 and -1544.46 respectively for Kalaburagi, Chincholi, Sedam, Chittapur, Aland, Jewargi and Afzalpur. The parameters estimated for different stations are presented in Table 4. In addition, the residuals were obtained by differencing original series with the fitted series and residuals were found (Table 5) to be as white noise.

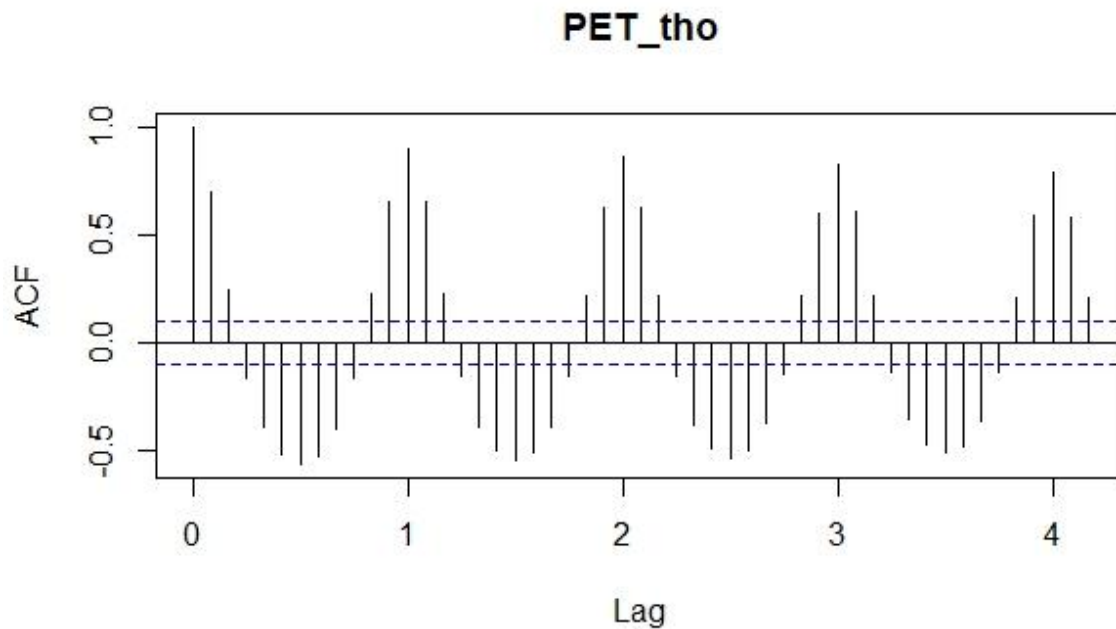


Fig. 1: Autocorrelation function plot of PET time series for Kalaburagi Station

Series kats

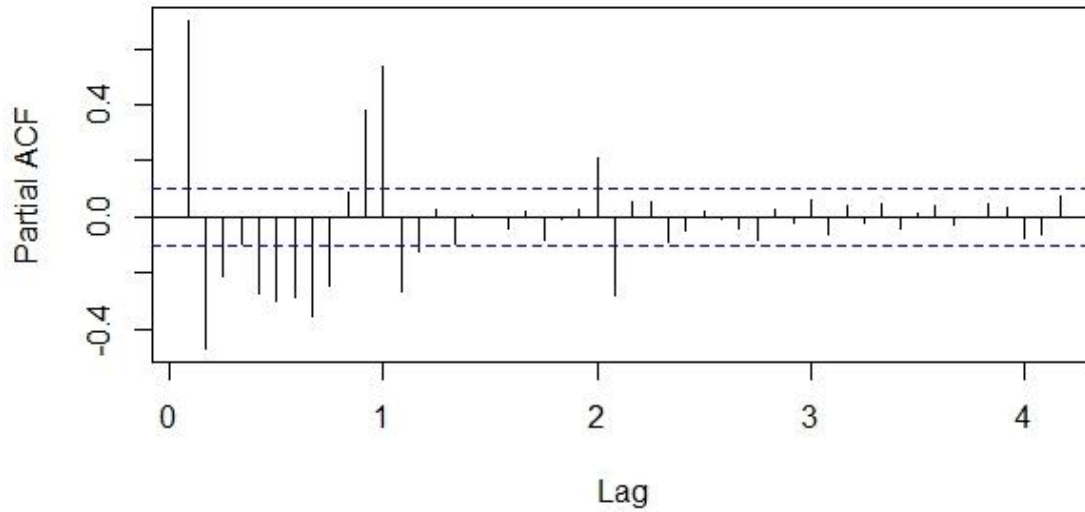


Fig.

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

Table 3: Log likelihood AIC and BIC values of ARIMA model for different stations

Stations	Model	Log-Likelihood	AIC	BIC
Kalaburagi	SARIMA (1,0,1)(2,1,0) ₁₂	-1532.32	3074.64	3093.9
Chincholi	SARIMA(1,0,1)(2,1,0) ₁₂	-1542.42	3024.82	3019.5
Sedam	SARIMA(1,0,0)(2,1,0) ₁₂	-1732.01	3474.02	3488.45
Chittapur	SARIMA(1,0,1)(2,1,0) ₁₂	-1552.32	3064.64	3083.9
Aland	SARIMA (1,0,1)(2,1,0) ₁₂	-1512.62	3094.64	3098.9
Jewargi	SARIMA (1,0,1)(2,1,0) ₁₂	-1542.24	3054.64	3053.9
Afzalpur	SARIMA (1,0,1)(2,1,0) ₁₂	-1544.46	3094.64	3148.6

Table 4: Parameter estimation of SARIMA by maximum likelihood method for different stations

Station	Model	Parameters	Estimate	S.E.	Z value	P-value
Kalaburagi	SARIMA (2,0,2)(1,1,2) ₁₂	Ar ₁	0.73	0.080	9.207	<0.001
		Ma ₁	-0.401	0.110	-3.630	<0.001
		Sar ₁	-0.84	0.049	-16.919	<0.001
		Sar ₂	-0.40	0.050	-8.133	<0.001
Chincholi	SARIMA (1,0,1)(2,1,0) ₁₂	Ar ₁	0.74	0.082	8.263	<0.001
		Ma ₁	-0.412	0.116	-5.654	<0.001

		Sar ₁	-0.88	0.045	-18.324	<0.001
		Sar ₂	-0.49	0.052	-9.124	<0.001
Sedam	SARIMA (1,0,1)(1,1,2) ₁₂	Ar ₁	0.314	0.050	6.245	<0.001
		Sar ₁	-0.752	0.052	-14.289	<0.001
		Sar ₂	-0.306	0.052	-5.789	<0.001
Chittapur	SARIMA (1,0,1)(2,1,0) ₁₂	Ar ₁	0.728	0.088	9.408	<0.001
		Ma ₁	-0.409	0.114	-3.840	<0.001
		Sar ₁	-0.862	0.042	-16.918	<0.001
		Sar ₂	-0.400	0.062	-8.152	<0.001
Aland	SARIMA (2,0,2)(1,1,2) ₁₂	Ar ₁	0.732	0.084	9.407	<0.001
		Ma ₁	-0.410	0.118	-3.830	<0.001
		Sar ₁	-0.842	0.042	-16.812	<0.001
		Sar ₂	-0.406	0.054	-8.203	<0.001
Jewargi	SARIMA (1,0,1)(2,1,0) ₁₂	Ar ₁	0.74	0.081	9.203	<0.001
		Ma ₁	-0.420	0.122	-3.633	<0.001
		Sar ₁	-0.862	0.044	-16.84	<0.001
		Sar ₂	-0.410	0.051	-8.234	<0.001
Afzalpur	SARIMA (2,0,2)(1,1,2) ₁₂	Ar ₁	0.82	0.086	9.308	<0.001
		Ma ₁	-0.54	0.212	-3.744	<0.001
		Sar ₁	-0.842	0.056	-16.984	<0.001
		Sar ₂	-0.48	0.058	-8.458	<0.001

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

Station	Chi-Square	Lag order	P-value
Kalaburagi	0.01	1	0.919
Chincholi	0.04	1	0.82
Sedam	0.330	1	0.565
Chittapur	2.27	1	0.13
Aland	0.0103	1	0.919
Jewargi	0.06	1	0.921
Afzalpur	0.08	1	0.924

After the development of models for all the seven taluks (selected stations) of Kalaburgi district, the forecasting part was carried out and the results are represented in Table 6. The results reveal that, initially for all stations the forecast was observed to be good with a correlation coefficient of 0.96, 0.95, 0.90, 0.88, 0.89, 0.92 and 0.86 for Kalaburagi, Chincholi, Sedam, Chittapur, Aland, Jewargi and Afzalpur respectively. Basic statistical properties were compared between observed and forecasted data using t-test for the means and F-test for standard deviation (Haan, 1977) and are shown in Table 7. Since t_{cal} values related to means were between t-critical table values (± 1.71) for two tailed at 5% significance level, the data shows that, there is no

significant difference between the mean values of observed and predicted data. Similarly, the F_{cal} values of standard deviation were smaller than the F-critical values at 5% significance level. Thus, the results show that, the predicted data preserves the basic statistical properties of the observed series.

Table 6: Performance measure of Seasonal ARIMA models at different stations

Station	Model	Performance measures	
Kalaburagi	SARIMA (2,0,2)(1,1,2) ₁₂	RMSE	22.61
		MAPE	13.72
		R	0.96
Chincholi	SARIMA (1,0,1)(2,1,0) ₁₂	RMSE	21.61
		MAPE	15.52
		R	0.95
Sedam	SARIMA (1,0,1)(1,1,2) ₁₂	RMSE	23.45
		MAPE	11.25
		R	0.90
Chittapur	SARIMA (1,0,1)(2,1,0) ₁₂	RMSE	28.54
		MAPE	19.45
		R	0.88
Aland	SARIMA (2,0,2)(1,1,2) ₁₂	RMSE	29.52
		MAPE	9.85
		R	0.89
Jewargi	SARIMA (1,0,1)(2,1,0) ₁₂	RMSE	20.15
		MAPE	14.96
		R	0.92
Afzalpur	SARIMA (2,0,2)(1,1,2) ₁₂	RMSE	30.12
		MAPE	18.92
		R	0.86

Table 7: Comparison of statistical properties of the observed and predicted data

Stations	Mean observed	Mean forecasted	Decision (t<1.79)	Observed variance	Forecast variance	Decision (f < 0.74)
Kalaburagi	151.59	142.86	0.19	4580.23	6143.2	0.31
Chincholi	142.36	112.24	0.17	3245.26	3878.69	0.25
Sedam	136.74	122.25	0.006	3008.58	3152.63	0.29
Chittapur	136.54	152.36	-0.045	2254.69	2296.69	0.21
Aland	161.23	112.02	0.17	2985.63	3001.98	0.31
Jewargi	159.85	98.85	-0.017	4012.96	4112.65	0.30
Afzalpur	165.66	107.25	0.15	2160.52	2560.32	0.24

CONCLUSION

The Seasonal ARIMA models revealed that, the models have an ability to forecast up to 12 month lead time with a higher accuracy over all the stations. Out of all the stations, Seasonal ARIMA model provided excellent results at Jewargi station with an R, RMSE and MAPE values of 0.92, 20.15 and 14.96 respectively. The seasonal ARIMA models for different stations were found to forecast potential evapotranspiration accurately up to one ahead with the least error. Similarly, for the basic statistical analysis, the difference between the observed and forecasted mean were found to be non-significant, which in turn reveal that, the models are found to be quite promising in forecasting the potential evapotranspiration over the study area. The prediction of evapotranspiration guarantees reliable project planning, design and operation of irrigation systems.

REFERENCES

- 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.
- Asadi, A. Vahdat, S. F. and 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
- Box GE, Jenkins GM, Reinsel GC. Time Series Analysis, Forecasting and Control, 3rd Edn (Englewood Cliffs, NJ, Prentice-Hall). 1994.

Box GE, Jenkins GM. Time series analysis: Forecasting and control San Francisco.Calif: Holden-Day. 1976.

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.

Gorantiwar SD, Meshram DT, Mittal HK. Seasonal ARIMA model for generation and forecasting evapotranspiration of Solapur district of Maharashtra.Journal of Agrometeorology. 2011; 13(2):119-22.

Gorantiwar SD, Patil PD. Stochastic modelling of crop evapotranspiration for Rahuriregion.International Journal of Agricultural Engineering. 2009;2(1):140-145.

Haan CT, Statistical methods in hydrology. Iowa State Press, Iowa. 1977.

Hamdi MR, Bdour AN, TarawnehZS.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.

Hsin-Fu Yeh, Hsin-Li Hsu, Stochastic model for drought forecasting in the Southern Taiwan basin.Water. 2019; 11(10): 1-15.

Malik A, Tikhamarine Y, Sammen SS, Abba SI, Shahid S. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms.Environmental Science and Pollution Research. 2021; 28(29): 39139-39158.

Mishra AK, Desai VR.Drought forecasting using stochastic models. Stochastic environmental research and risk assessment. 2005;19(5):326-39.

Mohan S, Arumugam N. Forecasting weekly reference crop evapotranspiration series. Hydrological sciences journal. 1995; 40(6):689-702.

Patil R, Nagaraj DM, Polisgowdar BS, RATHOD S. Forecasting potential evapotranspiration for Raichur district using seasonal ARIMA model.Mausam. 2022; 73(2): 433-440.

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.

Prasad G, Vuyyuru UR, Gupta MD. Agriculture commodity arrival prediction using remote sensing data: insights and beyond. arXiv preprint arXiv:1906.07573. 2019 .

Salas JD. Applied modeling of hydrologic time series.Water Resources Publication; 1980.

Shadab A, Ahmad S, Said S. Spatial forecasting of solar radiation using ARIMA model. Remote Sensing Applications: Society and Environment. 2020; 20(3): 100427.

UNDER PEER REVIEW