Eco. Env. & Cons. 29 (November Suppl. Issue) : 2023; pp. (S154-S160)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2023.v29i06s.025

Forecasting Potential Evapotranspiration for Yadgir District Karnataka, India using Seasonal Arima Model

Mallikarjun Reddy[1], Basavareddy[2] and Rahul Patil[1], Pragna Guguloth[2] and Premkumara[2]

[1]Department of Agricultural Engineering, UAS, Raichur, India
[2]Department of Soil and Water Engineering, CAE, UAS, Raichur, India

(Received 3 June, 2023; Accepted 18 July, 2023)

ABSTRACT

The prediction of potential evapotranspiration (PET) is quite important task for reliable management of irrigation systems. This article is generally based on the models which try to mimic the actual occurrence of the Potential evapotranspiration in the future days for Yadgir district. In this study the potential evapotranspiration was estimated with the help of max and min temperature (°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 ARIMA(2,0,2) $(2,1,0)_{12}$, ARIMA(2,0,2)$(2,1,0)_{12}$, ARIMA (1,0,1)$(2,1,0)_{12}$, and ARIMA(2,0,2)$(2,1,0)_{12}$, for Yadgir, Gurmitkal, Shahapur and Shorapur respectively. Furthermore, the results showed that the models developed for Gurmitkaland Shahapur were found to be quite promising compared to the other two stations. All four models were found to be producing better results. The models provided significant potential in improving the decision making in irrigation planning and command area management practices for better management of water resources.

Key words: *Evapotranspiration, Seasonal Arima model, Yadgir*

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 (Asadi *et al*., 2013). Globally, ET consumes about 60 per cent of the annual precipitation that falls over the earth's surface ET quantification is 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 and wind speed etc.

The stochastic models are based on the time dependent variation and consider random effects involved in the ET process. For two main reasons, sto-

([1]Assistant Professor [2]Ph. D. Scholar) chastic linear models are fitted to hydrological data or 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 autoregressive integrated moving average (ARIMA) model is considered as one of the best models 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) did analysis of evapotranspiration of 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 so it is quite important to understand its behaviour based on the historical data for better management of water resource, in order to understand and solve the irrigation problem the research was carried out. The objective of this study is to establish a time series model to analyse and forecast reference crop evapotranspiration for the Yadgir district.

Materials and Methodology

Yadgir is an administrative district in the Indian state of Karnataka. Yadgir District was carved out from the erstwhile Kalburgi district as the 30th district of Karnataka on 31st December 2009. Located in the North east part of the state surrounded by Kalburgi in the North, Raichur in the South, Vijayapur in the West and the state of Telengana in the East. Yadgir District is the 2nd smallest district in the state.Yadgir district is spread over an area of 5270 sq. Km constituting 8.46 percent area of Karnataka State. Geographical location of Yadgir is 16°202 to 17°452 North latitude and 76°42 to 77°422 East Longitude. The region is generally hot and temperature of this region is approximately 45 ^{0}C (max) and 22 ^{0}C (min). Yadgir has been blessed by the incessant flowing of two main rivers, "Krishna' and 'Bhima'.

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.

Auto correlation test (Box Ljung test)

The null hypothesis of the Box Ljung Test, H_0, is that our 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 auto correlated.

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 compute 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.

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 (ARMA) models. In an ARMA 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 analysing 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

$$\emptyset(B)\nabla^d_{z_t} = \theta(B)a_t$$

Where θ (B) are polynomials of order p and q, respectively. Non-seasonal AR operator of order p is written as

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 – \phi_p BP)$$

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

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

Seasonal ARIMA models

Many time series features the cyclic. Quite frequently such characteristics are on an annual period in hydrologic time series mainly due to earth's rotation around the sun. Such type 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 (Gorantiwar *et 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:

$$\phi_p(B)\Phi_p(B^s)\Delta^d\Delta_s{}^D Z_t = \theta_p(B)\varphi_Q(B^s)a_t$$

where *p* is the order of non-seasonal autoregression, *d* the number of regular differencing, *q* the order of nonseasonal MA, *P* the order of seasonal autoregression, *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.

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).

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.

AIC= -2 log (L) + 2k

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.

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.

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.

Drought forecasting

The prediction of Potential evapotranspiration was done for 1-6 month lead time using the best fit models from historical data. Basic statistical properties of the observed and predicted data for 1-6 month lead time was computed and tested whether the predicted data preserve the basic statistical properties of the observed PET series. The predictions are calculated for different lead time. For instance, a 1-month lead time prediction means that during January 2017, the prediction for February 2017 is computed. The correlation coefficients (R), RMSE and MAE were observed between the observed and predicted data for 1 to 6 month lead times.

Input Dataset and software

The time series of temperature data set (Max and Min) was taken from the Main Agriculture Research Station (MARS) Raichur. The data set were from 1984-2018, out of which 1984-2016 was used for the development of the model and the 2017-2018 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

Development of model was done with prerequisite tests namely Stationary and autocorrelation test. 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 237.74(0.01), 265.15(0.01), 241.37(0.01)and 239.91(0.01) for Yadgir, Gurmitkal, Shahapur and Shorapur 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 to the data sets Table 2.

Table 1. Auto correlation test for different station

Station	Chi-Square	Lag order	P-value
Yadgir	237.74	1	<0.001
Gurmitkal	265.15	1	<0.001
Shahapur	241.37	1	<0.001

Table 2. Stationarity test for different station

Station	Dickey fuller	Lag order	P-value
Yadgir	-18.991	7	0.01
Gurmitkal	-18.771	7	0.01
Shahapur	-19.451	7	0.01
Shorapur	-19.854	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 are considered and combination of the order which yields maximum log-likelihood and lowest values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are considered as best model. The results pertaining to Yadgir, Gurmitkal, Shahapur and Shorapurstations 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 were observed have a seasonality thereby seasonal ARIMA models were selected with a seasonal differencing as shown in Table 4. The best selected models for different stations were ARIMA $(2,0,2)(2,1,0)_{12}$, ARIMA $(2,0,2)(2,1,0)_{12,}$ ARIMA $(1,0,1)(2,1,0)_{12}$ andARIMA $(2,0,2)(2,1,0)_{12}$ with an maximum likelihood values of -1644.31, -1651.42, -1592.08 and -1665.82 respectively for Yadgir,

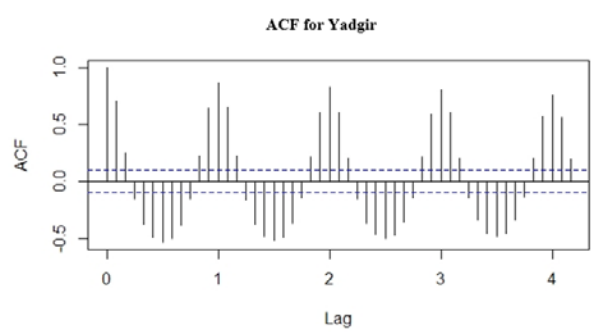

Fig. 1. Autocorrelation function plot of PET time series for Yadgir Station

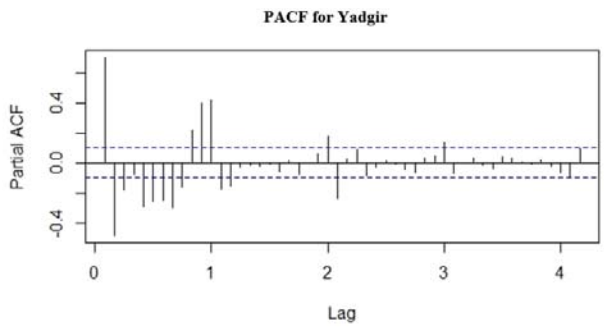

Fig. 2. Partial autocorrelation function plot of PET time series for Yadgir Station

Gurmitkal, Shahapur and Shorapur respectively. 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 to be white noise as presented in Table 5.

After the development of models for 4 taluks the forecasting part was carried out at different lead time (1-6 months) and the results Table 6 reveal that initially for all stations the forecast was observed to be good at 1 lead time with a correlation coefficient of 0.90, 0.92, 0.86 and 0.91 for Yadgir, Gurmitkal, Shahapur and Shorapurrespectively. The RMSE and MAE were observed to be least at 1 leads and increases as the lead time increase, these stochastic models were found to suitable to forecast up to 1 lead time. A view at the Table 6 can be easily noticed that as the lead time increases the error rate showed increase tremendously. It can be easily concluded that Seasonal ARIMA models suited well for forecasting at 1-month lead time for Potential evapotranspiration forecasting under Yadgir Region. Basic statistical properties are compared between observed and forecasted data for 1-month lead time, using t-test for the means and F-test for standard deviation (Haan, 1977), shown in Table 7. Since t_{cal} values related to means were between t-critical table

Table 3. Log likelihood AIC and BIC values of ARIMA model for different station

Stations	Model	Log-Likelihood	AIC	BIC
Yadgir	$SARIMA(2,0,2)(2,1,0)_{12}$	-1644.31	3304.62	3335.44
Gurmitkal	$SARIMA(2,0,2)(2,1,0)_{12}$	-1651.42	3311.22	3336.12
Shahapur	$SARIMA(1,0,1)(2,1,0)_{12}$	-1592.08	3301.16	3334.05
Shorapur	$SARIMA(2,0,2)(2,1,0)_{12}$	-1665.82	3322.11	3335.28

Table 4. Parameter estimation of SARIMA by maximum likelihood method for different station

Station	Model	Parameters	Estimate	S.E.	Z value	P-value
Yadgir	$SARIMA(2,0,2)(2,1,0)_{12}$	AR1	-0.0748	0.134	-1.812	0.065
		AR2	0.2834	0.116	5.171	< 0.001
		MA1	0.2541	0.171	3.320	< 0.001
		MA2	-0.386	0.151	-2.469	0.012
		SAR1	0.251	0.321	0.788	0.421
		SAR2	-1.271	0.313	-4.099	< 0.001
Gurmitkal	$SARIMA(2,0,2)(2,1,0)_{12}$	AR1	0.641	0.118	5.739	< 0.001
		AR2	-0.282	0.146	-1.952	0.051
		MA1	-0.749	0.047	-15.788	< 0.001
		SAR2	-0.369	0.041	-7.786	< 0.001
Shahapur	$SARIMA(1,0,1)(2,1,0)_{12}$	AR1	0.671	0.116	5.860	< 0.001
		AR2	-0.373	0.149	-2.536	0.01
		MA1	0.107	0.314	0.333	0.76
		SAR2	-1.161	0.296	-3.883	< 0.001
		AR1	0.270	0.291	0.869	0.391
Shorapur	$SARIMA(2,0,2)(2,1,0)_{12}$	AR1	0.628	0.126	5.113	< 0.001
		AR2	-0.312	0.148	-2.079	0.029
		MA1	0.126	0.313	0.399	0.689
		SAR2	-1.171	0.293	-3.929	< 0.001
		AR1	0.261	0.291	0.905	0.37

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

Station	Chi-Square	Lag order	P-value
Yadgir	2.15	1	0.15
Gurmitkal	0.05	1	0.84
Shahapur	0.02	1	0.93
Shorapur	2.31	1	0.14

values (±1.71 for two tailed at a 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 a 5% significance level. Thus, the results show that predicted data preserves the basic statistical properties of the observed series.

Conclusion

The Seasonal ARIMA models revealed that the models have an ability to forecast up to 1-month lead month with a higher accuracy over all the stations. Of the all stations, Seasonal ARIMA model provided excellent results at Shahpur station with an RMSE, MAPEandMAE, values of 10.75, 5.23 and 7.70 respectively. The seasonal ARIMA models for different station are found to forecasting potential evapotranspiration accurately up to one ahead with a 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 period. The prediction of evapotranspiration guarantees reliable project planning, design and operating of irrigation systems.

References

Aruna, K.T. Satish Kumar, U. Ayyanagowdar, M.S. Srinivasa Reddy, G.V. and Shanwad, U.K. 2017. Estimation of crop coefficient (Kc) values for groundnut crop with evaluated crop evapotranspiration at different moisture deficit levels under agro

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

Station	Model	Performance measures	Lead time 1	2	3	4	5	6
Yadgir	$SARIMA(2,0,2)(2,1,0)_{12}$	RMSE	17.51	24.88	25.36	24.90	23.22	22.23
		MAPE	23.36	14.61	14.31	13.81	15.20	13.96
		MAE	13.71	21.17	20.63	19.88	20.17	17.41
		R	0.90	0.84	0.83	0.82	0.80	0.80
Gurmitkal	$SARIMA(2,0,2)(2,1,0)_{12}$	RMSE	10.75	18.10	19.69	18.53	17.59	17.45
		MAPE	9.31	16.61	18.15	17.71	17.35	16.91
		MAE	8.96	15.89	16.65	15.92	15.10	13.83
		R	0.92	0.84	0.83	0.82	0.83	0.81
Shahapur	$SARIMA(1,0,1)(2,1,0)_{12}$	RMSE	10.57	20.27	39.59	52.20	59.65	61.51
		MAPE	5.23	10.50	22.45	29.16	33.45	37.80
		MAE	7.70	15.77	32.81	43.15	48.93	51.73
		R	0.86	0.776	0.71	0.73	0.67	0.55
Shorapur	$SARIMA(2,0,2)(2,1,0)_{12}$	RMSE	20.78	22.9	53.22	61.48	64.08	66.21
		MAPE	10.52	40.35	30.68	35.91	41.03	45.02
		MAE	15.62	33.57	44.61	51.32	54.91	56.35
		R	0.91	0.82	0.81	0.82	0.79	0.78

Table 7. Comparison of statistic properties of the observed and predicted data

Stations	Mean observed	Mean forecasted	Decision (t<1.71)	Observed variance	Forecast variance	Decision (f < 4.05)
Yadgir	138.87	135.20	1.04	1393.86	995.22	0.15
Gurmitkal	98.77	97.81	0.19	641.61	548.638	0.0009
Shahapur	138.76	138.56	0.009	1090.29	1399.84	0.0004
Shorapur	139.67	139.99	-0.075	1260.67	1549.44	0.0008

climatic condition of Raichur, Karnataka. *Green Farming.* 8 (5) : 1161-1164.

Asadi, A. Vahdat, S. F. and Sarraf, A. 2013. The forecasting of potential evapotranspiration using time series analysis in humid and semi humid regions. *American Journal of Engineering Research*. 2(1): 296–302

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis Forecasting And Control.* Holden-Day, San Francisco.

Box, G.E.P. Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis, Forecasting and Control,* (3rd Edition), Englewood Cliffs, N.J. Prentice Hall.

Dutta, B. Smith, W.N. Grant, B.B. Pattey, E. Desjardins, R.L. and Li, C. 2016. Model development in DNDC for the prediction of evapotranspiration and water use in temperate field cropping systems. *Environmental Modelling & Software.* 80(1): 9-25.

Gorantiwar, S.D. Meshram, D.T. and Mittal, H.K. 2011. Seasonal ARIMA model for generation and forecasting evapotranspirtion of Solapur district of Maharashtra. *Journal of Agrometeorology.* 12 (2) : 119-122.

Gorantiwar, S.D. and Patil, P.D. 2009. Stochastic modelling for crop evapotranspiration for Rahuri Region. *International Journal of Agricultural Engineering.* 2(1) : 140–145.

Haan, C.T. 1977. *Statistical Methods in Hydrology.* Iowa State Press, Iowa.

Hamdi, M.R. Bdour, A.N. and Tarawneh, A.N. 2008. Developing reference crop evapotranspiration time series simulation model using class a pan: A case study for the Jordan Valley, Jordan. *Journal of Earth and Environmental Sciences.* 1(1) : 33–44.

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

Mishra, A.K. and Desai, V.R. 2005. Drought forecasting using stochastic models. *J. Stoch. Environ. Res. Risk Assess*. 19(1): 326–339.

Mohan, S. and Arumugam, N. 1995. Forecasting weekly reference crop evapotranspiration series. *Hydrological Sciences Journal.* 40(6): 689-702.

Popale, P.G. and Gorantiwar, S.D. 2014. Stochastic generation and forecasting of weekly rainfall for Rahuri Region. *International Journal of Innovative Research in Science, Engineering and Technology.* 3(4): 185–196.

Salas, J.D. Delleure, J.W. Yevjevich, V. and Lane, W.L. 1980. *Applied Modeling of Hydrologic Time Series*. Water Res. Publ., Colorado, USA.