Forecasting potential evapotranspiration using seasonal ARIMA model for Northern Telangana Zone

Abstract:

Potential evapotranspiration is a major component in crop water consumption management plans. Consequently, forecasting of potential evapotranspiration is the keystone of any effective water resources management plans in a semi-arid environment. The estimation of daily evapotranspiration was carried out using the Thronthwaite method while the forecasting of the potential evapotranspiration was carried out using Seasonal Auto Regressive Integrated Moving Average (SARIMA). Time series analysis of evapotranspiration data set showed a seasonality behaviour and thus Seasonal ARIMA model with the least Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values were selected. Seasonal Arima model selected for the districts Adilabad, Jagtial, Karimnagar, KumuramBheem, Nirmal, and Peddapalli was $(2,0,2)(2,1,0)_{12}$ and for Nizamabad district $(2,0,2)(1,1,0)_{12}$. Basic statistical properties are used to compare observed and forecasted data which shown that that there is no significant difference between the mean values of observed and predicted data at a 5% significance level. Hence the developed model was optimum to forecast the evapotranspiration over the study area and sustain the forecasting accuracy.

Keywords: SARIMA, Potential Evapotranspiration, Forecasting, Model,

Introduction:

Evapotranspiration (ET) represents the combination of evaporation and transpiration, where evaporation is vaporization from the soil surface or water surface and transpiration is plant water absorption from the root zone [Nolz 2016]. Both precipitation and ET represent the climate of a region and are used as a decision support tool for water management in agriculture. While contributing to the surface energy balance, ET quantifies the water requirement for efficient water management [Afzaal *et al.* 2020, Bogawski and Bednorz 2014]. Not only in irrigation assessments but also in the the accurate modelling of river basin hydrology, estimation of local ET is one of the essential tasks [Karimi and Bastiaanssen 2015]. Krishna (2018) highlighted that the accurate estimation of ET is important because understanding and quantifying the processes governing ET clarifies the uncertainties in the behaviour of the hydrologic cycle with the changing climate. Since ET is a critical factor in water balance from plot scale to a global scale, well-grounded ET estimations are required to regulate the components of the irrigation system: the size of canals and dams, and the capacity of pumps [Kharrou 2021].

Irrigation is one of the most important backward linkages to Agriculture. Competition for water, high pumping costs, complexities of water storage and delivery, and concerns for the environment are among the factors that drive an interest in improving the water use efficiency and the operation of large irrigation systems. For purposes of timely and efficient water application, agricultural managers have long relied on evapotranspiration (ET) measurements or estimations. Therefore, an accurate assessment of ET is prerequisite to improving water management practices.

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 the crop water requirement (CWR) of any crop. Determining the crop water requirement using evapotranspiration is considered one of the main planning needs for water resources management. In much long-term planning, it is necessary to outline the future state of rainfall and dry and wet periods for the region. For this purpose, the prediction of drought and the estimation of its characteristics are of great importance in water resources management (Raziei et al. 2007). One of the ET forecasting methods is the use of time series analysis, which has been rapidly developed for predictive and analytical issues since the 1970s. Therefore, due to the nature of the hydrological events, if the correct selection of the model and correct calculations are made, time series can be particularly consistent with the hydrologic data (Salas 1993). In many time series, there is a consistent correlation between observations, which is a hallmark of the autoregressive integrated moving average model (ARIMA) and seasonal ARIMA (SARIMA) models (Toufani et al. 2011).

One definition of a time series is that of a collection of quantitative observations that are evenly spaced in time and measured successively. Time series are analysed in order to understand the underlying structure and function that produce the observations. Understanding the theoretical explanation of a time series analysis allows a mathematical model to be developed to explain the data trend in such a way that monitoring, simulation, prediction, assessment and management can occur (Box and Jenkins 1976; Brockwell and Davis 2002; Box et al. 2015; Sentas and Psilovikos 2010).

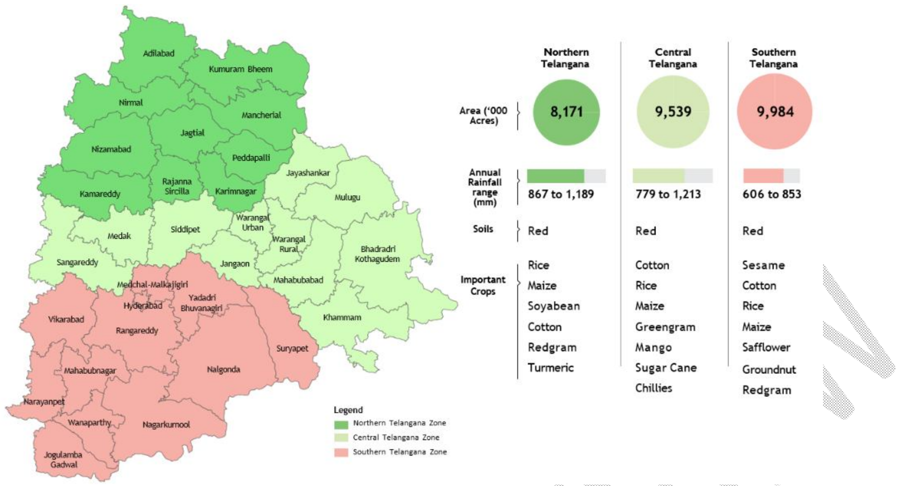
Hydrological processes are complicated; they are influenced by not only deterministic, but also stochastic factors (Wang et al., 2007). The deterministic change in a hydrological process is always accompanied by the stochastic change. Generally speaking, determinism includes periodicity, tendency, and abrupt change. A strict deterministic hydrological process is rare. Stationary time series has been widely used in hydrological data assimilation and prediction to tackle the stochastic factors in hydrological processes. Some researchers have used ARIMA model for the analysis of hydrological process, most of the studies (Ahmad et al., 2001; Lehmann and Rode, 2001; Oi and Zhen, 2006) neglected stationary test and the influence from inter-monthly variation within a year.

The state of Telangana is the young eststate in the Indian Union.Telangana is ranked 12^{th} in the country in terms of population and ranked 11^{th} in terms of area. The region is majorly drained by Godavari and Krishna rivers with 79% and 69% catchment areas respec tively.Telangana state is marked by erratic and uneven rainfall and mostly depends on monsoons. The Government of Telangana has adopted a comprehensive irrigation development strategy to provide irrigation facilities to about 125 lakh acres of land across the state. The government has also taken up several measures and adopted strategies for the completion of pending irrigation projects, on a fast-track. The atmospheric temperature is projected to increase with climate change, and it provides more energy leading to more evaporation. Therefore, improvements of water-use efficiency and sustainable water management in agriculture must be based on the accurate estimation of ET. Taking these facts into account, the current study is to predict evapotranspiration values over Northern Telangana Zone through identification or patterns in correlated data trends and seasonal variation and to assess the accuracy of the forecasting model.

Materials and methods:

The Northern Telangana semi-arid zone lies between 17° 42' and 19° 84' N Latitude and 77° 38' and 81° 16' E Longitude. This zone includes the districts of Adilabad, KumuramBheem, Manchiriala, Nirmal, Karimnagar, Jagtial, Peddapalli, Rajanna Siricilla, Nizamabad and Kamareddy with Regional Agricultural Research Station, Jagtial as Regional headquarters. The annual average rainfall is 900 to 1150 mm mostly from the south-west monsoon. The maximum and minimum temperature during south-west monsoon ranges from 32°C to 37°C and 21°C to 25°C respectively. Red soils are predominant in the zone which includes chalkas, red sandy, deep red loamy and very deep black cotton soils are also seen in some parts of the zone. This zone has a total geographical area of 35.5 lakh ha. The climate is typically tropical rainy. The net sown area is 2.21 m. ha. of which 0.67 m. ha. is irrigated representing 30.3 per cent of the net sown area. Cropping intensity is 110 per cent. Wells are the main source of irrigation followed by canals. Important crops grown are Rice, maize, soybean, cotton, redgram and turmeric.

Image 1:



Source: ACRC, Prof.Jayashankar Telangana State Agriculture University (PJTSAU)

ARIMA models:

A hydrological time series {y_t, t = 1, 2, ..., n} could be either stationary or non-stationary. Given that there are essentially no strictly deterministic hydrological processes in nature, the analysis of hydrological data by means of non-stationary time series is of importance, among which ARIMA model is one of the available choices. 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:

$$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p\right)$$

And non-seasonal MA operator of order q is written

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q\right)$$

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 a type of series is 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 a 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 :

$$\emptyset_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D Z_t = \theta_p(B)\varphi_Q(B^s)a_t$$

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* is the number of seasonal differencing, *Q* is the order of seasonal MA, *s* is the length of season, seasonal AR parameter of order P, seasonal MA parameter of order Q.

Implementation of the ARIMA model

The procedure of estimating ARIMA model involves the following steps:

1. Stationary identification: The input time series for an ARIMA model needs to be stationary, i.e., the time series should have a constant mean, variance, and auto correlation through time. Therefore, the stationarity of the data series needs to be identified first. If not, the non-stationary time series is then required to be stationarized. Although the stationary test, such as the unit root test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used to identify whether a time series is stationary, plotting approaches based on scatter diagrams, autocorrelation function diagrams, and partial correlation function diagrams are also often used. The latter approach can usually not only provide

information on the testing time series is stationary, but it can also indicate the order of the differencing which is needed to stationarize the time series.

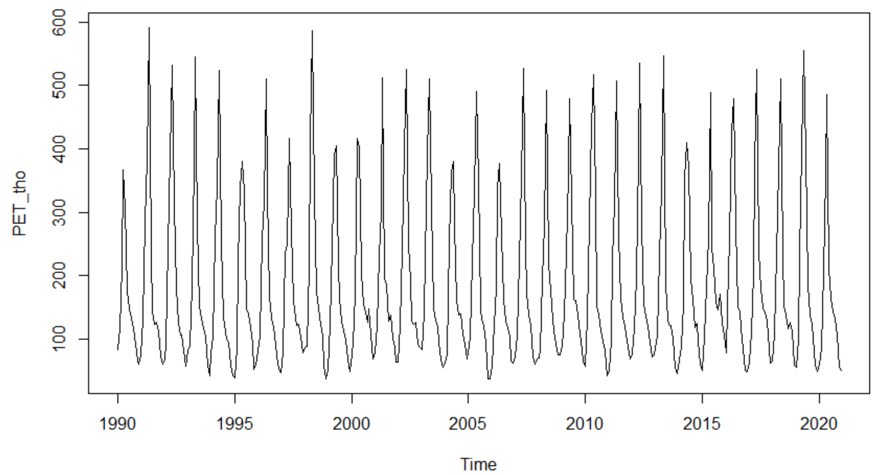


Fig 1: Line plot of differenced potential evapotranspiration data of first order (d=1)

In this paper, we identify the stationarity of a time series from the autocorrelation function diagram and partial correlation function diagram. If a time series is identified as non-stationary, differencing is usually made to stationarize the time series. In the differencing method, the correct amount of differencing is normally the lowest order of differencing that yields a time series, which fluctuates around a well-defined mean value, and whose autocorrelation function (ACF) plot decays fairly rapidly to zero, either from above or below. The time series is often transformed for stabilizing its variance through proper transformation, e.g., logarithmic transformation. Although logarithmic transformation is commonly used to stabilize the variance of a time series rather than directly stationarize a time series, the reduction in the variance of a time series is usually helpful to reduce the order of difference in order to make it stationary.

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.

2. Identification of the order of ARIMA model: After a time series has been stationarized, the next step is to determine the order terms of its ARIMA model, i.e., the order of differencing, d for non-stationary time series, the order of auto-regression, p, the order of moving average, q, and the seasonal terms if the data series show seasonality. While one could just try some different combinations of terms and see what works best strictly, the more systematic and common way is to tentatively identify the orders of the ARIMA

model by looking at the autocorrelation function (ACF) and partial autocorrelation (PACF) plots of the stationarized time series. The ACF plot is merely a bar chart of the coefficients of correlation between a time series and lags of itself, and the PACF plot presents a plot of the partial correlation coefficients between the series and lags of itself. The detailed guidelines for identifying ARIMA model parameters based on ACF and PACF, can be found elsewhere, e.g., Pankratz (1983) and Shumway and Stoffer (2005). It should be noted that, to be strict, the ARIMA model built in this step is actually an ARMA model with if the time series is stationary, which is in fact a special case of ARIMA model with d = 0.

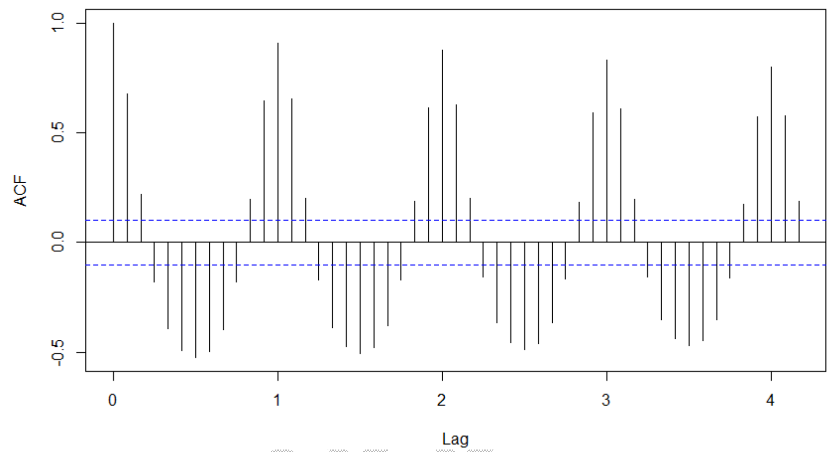


Fig 2: Autocorrelation function plot of PET time series for Raichur Station

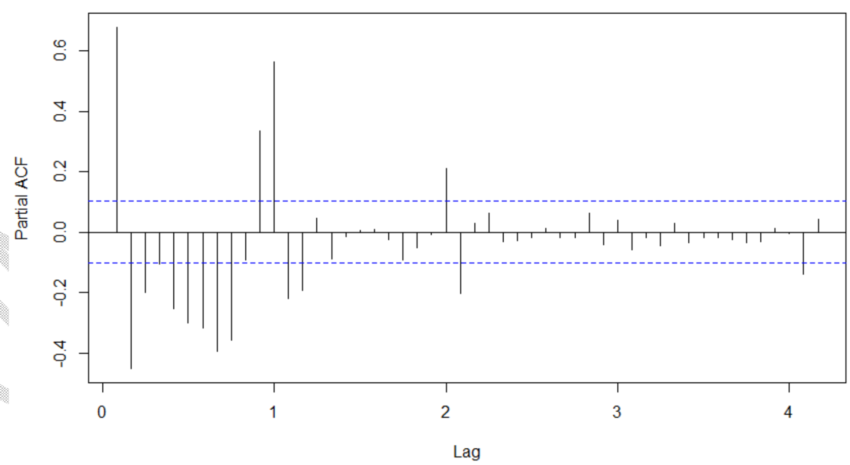


Fig 3: Partial autocorrelation function plotof PET time series for Raichur Station

3. Estimation of ARIMA model parameters: while least square methods (linear or nonlinear) are often used for the parameter estimation, we use the maximum likelihood method (Mcleod and Sales, 1983; Melard, 1984) in this paper. A t test is also performed to test the statistical significance. 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.

4. White noise test for residual sequence: it is necessary to evaluate the established ARIMA model with estimated parameters before using it to make forecasting. We use white noise test here. If the residual sequence is not a white noise, some useful information has not been extracted and the model needs to be further tuned. 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.
5. ET forecasting: The prediction of Potential evapotranspiration was done time using the best fit models from historical data. Basic statistical properties of the observed and predicted data was computed and tested whether the predicted data preserve the basic statistical properties of the observed PET series. 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) was taken from NASA POWER-Prediction of Worldwide Energy Resources and for Jagtial District, data collected from meteorological station, Regional Agriculture Research Station (RARS) Jagtial. The data set were from 1990-2020, out of which 1990-2019 was used for the development of the model and the data set for 2019-2020 was used for the validation purpose. The Potential evapotranspiration was estimated using Thornthwaite method and ARIMA models were developed in the R studio.

Thornthwaite method (Potential evapotranspiration estimation)

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.

RESULTS AND DISCUSSION

For any given time series data set there is at least one assumed systematic pattern embedded in the data. The most common patterns are trends and seasonality; trends are generally either linear or quadratic. To find out trends and/or moving averages, regression analysis is often used. Seasonality is a trend that repeats itself systematically over time (Box and Jenkins 1976; Vandaele 1983).

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 174.24, 171.36, 176.55, 169.84, 173.69, 176.62 and 170.28 and P values <0.001were for Adilabad, Jagtial, Karimnagar, KumuramBheem, Nirmal, Nizamabad and Peddapalli 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 a seasonality thereby seasonal differencing was done to the data sets (Table 2).

Table 1: Auto correlation test for different districts of NTZ

STATION	Chi-Square	Lag order	P-Value
Adilabad	174.24	1	<0.001
Jagtial	171.36	1	<0.001
Karimnagar	176.55	1	<0.001
KumuramBheem	169.84	1	<0.001
Nirmal	173.69	1	<0.001
Nizamabad	176.62	1	<0.001
Peddapalli	170.28	1	<0.001

Table 2: Stationery test for different districts of NTZ

STATION	Dickey fuller	Lag order	P-value
Adilabad	-19.367	7	0.01
Jagtial	-19.132	7	0.01
Karimnagar	-18.725	7	0.01
KumuramBheem	-19.256	7	0.01
Nirmal	-18.965	7	0.01
Nizamabad	-18.73	7	0.01
Peddapalli	-18.758	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 Adilabad, Jagtial, Karimnagar, KumuramBheem, Nirmal, Nizamabad and Peddapalli districts regarding model development are presented in Tables 3&4. The ACF and PACF were plotted (Figs. 2 and 3) to determine the model, the data were observed to 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), ARIMA(2,0,2)(2,1,0), ARIMA(2,0,2)(2,1,0), ARIMA(2,0,2)(2,1,0), ARIMA(2,0,2)(2,1,0), ARIMA(2,0,2)(1,1,0) and ARIMA(2,0,2)(2,1,0) with an maximum likelihood values of -1753.77, -1760.99, -1722.34, -1793.98, -1735.35, -1722.63 and -1812.89 respectively for Adilabad, Jagtial, Karimnagar, KumuramBheem, Nirmal, Nizamabad and Peddapalli. The parameters estimated for different districts 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.

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

STATION	model	Log-Likelihood	AIC	BIC
Adilabad	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1753.77	3523.54	3554.36
Jagtial	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1760.99	3537.97	3568.79
Karimnagar	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1722.34	3458.68	3485.64
KumuramBheem	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1793.98	3603.96	3634.77
Nirmal	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1735.35	3486.69	3517.51
Nizamabad	ARIMA(2,0,2)(1,1,0)$_{[12]}$	-1722.63	3459.27	3486.23
Peddapalli	ARIMA(2,0,2)(2,1,0)$_{[12]}$	-1812.89	3639.79	3666.75

After the development of models for 7 districts, the forecasting part was carried out at different and the results (Table 6) reveal that for all stations the forecast was observed to be good with a correlation coefficient of 0.883, 0.838, 0.813, 0.865, 0.847, 0.813 and 0.806 for Adilabad, Jagtial, Karimnagar, KumuramBheem, Nirmal, Nizamabad and Peddapalli districts. The RMSE and MAE were observed to be least and hence 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 Seasonal ARIMA models suited well for forecasting Potential evapotranspiration under Northern Telangana zone. 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.

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

STATION	model	Parameters	Estimate	S.E.	Z value	P-value
Adilabad	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.079653	0.464948	-0.1713	0.864
		AR2	0.24941	0.349428	0.7138	0.475
		MA1	0.256002	0.47104	0.5435	0.587
		MA2	-0.07307	0.302982	-0.2412	0.809
		SAR1	-0.697556	0.056636	-12.3164	< 0.001
		SMA1	-0.272944	0.056214	-4.8555	< 0.001
		SMA2	0.050321	0.122127	0.412	0.680
Jagtial	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.0791	0.3581	-0.2209	0.825
		AR2	0.2732	0.2852	0.9579	0.338
		MA1	0.2537	0.3651	0.6947	0.487
		MA2	-0.0806	0.26	-0.3101	0.757
		SAR1	-0.6878	0.0568	-12.1141	< 0.001
		SMA1	-0.2394	0.0565	-4.2392	< 0.001
		SMA2	0.0527	0.1301	0.4048	0.686
Karimnagar	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.072463	0.278068	-0.2606	0.794
		AR2	0.188861	0.237323	0.7958	0.426
		MA1	0.329784	0.27958	1.1796	0.238
		MA2	0.070375	0.204407	0.3443	0.731
		SAR1	-0.563659	0.047201	-11.9416	< 0.001
		SMA2	0.073067	0.155137	0.471	0.638
Kumurambheem	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.127469	0.367393	-0.347	0.729
		AR2	0.247821	0.307589	0.8057	0.420
		MA1	0.263579	0.374477	0.7039	0.482
		MA2	-0.073922	0.293495	-0.2519	0.801
		SAR1	-0.67717	0.05713	-11.8531	< 0.001
		SMA1	-0.247445	0.056703	-4.3639	< 0.001
		SMA2	0.045271	0.133255	0.3397	0.734
Nirmal	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.0102837	0.4648525	-0.0221	0.982
		AR2	0.1916833	0.3503454	0.5471	0.584
		MA1	0.2279929	0.467351	0.4878	0.626
		MA2	-0.0027449	0.2841486	-0.0097	0.992
		SAR1	-0.7142443	0.0562602	-12.6954	< 0.001
		SMA1	-0.2738084	0.055936	-4.895	< 0.001
		SMA2	0.0674401	0.1205473	0.5594	0.576
Nizamabad	ARIMA $(2,0,2)(1,1,0)_{[12]}$	AR1	-0.072633	0.278162	-0.2611	0.794
		AR2	0.188288	0.237448	0.793	0.428
		MA1	0.329702	0.279661	1.1789	0.238
		MA2	0.070771	0.20454	0.346	0.729
		SAR1	-0.563714	0.047203	-11.9424	< 0.001
		SMA2	0.073186	0.155168	0.4717	0.637
Peddapalli	ARIMA $(2,0,2)(2,1,0)_{[12]}$	AR1	-0.058909	0.301447	-0.1954	0.845
		AR2	0.376635	0.24794	1.5191	0.129
		MA1	0.214592	0.31331	0.6849	0.493
		MA2	-0.191379	0.241965	-0.7909	0.429
		SAR1	-0.66275	0.056576	-11.7142	< 0.001
		SAR2	-0.216286	0.056216	-3.8474	< 0.001

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

STATION	Chi-Square	Lag order	P-value
Adilabad	1.66E-06	1	0.999
Jagtial	3.57E-05	1	0.9952
Karimnagar	2.55E-06	1	0.9987
KumuramBheem	3.60E-06	1	0.9985
Nirmal	1.16E-05	1	0.9973
Nizamabad	2.45E-06	1	0.9967
Peddapalli	0.0016663	1	0.9674

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

STATION	Performance measures	1-LEAD TIME
Adilabad	RMSE	36.40649
	MAPE	14.89317
	MAE	23.38337
	R	0.8827
Jagtial	RMSE	37.18512
	MAPE	14.50108
	MAE	23.79058
	R	0.8379
Karimnagar	RMSE	33.333
	MAPE	13.7328
	MAE	21.34267
	R	0.8127
Kumurambheem	RMSE	40.89183
	MAPE	15.19015
	MAE	25.578
	R	0.8654
Nirmal	RMSE	34.51507
	MAPE	13.96106
	MAE	22.29738
	R	0.8472
Nizamabad	RMSE	33.36109
	MAPE	13.7337
	MAE	21.35144
	R	0.8131
Peddapalli	RMSE	43.19527
	MAPE	15.1979
	MAE	26.82257
	R	0.8061

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)
Adilabad	157.62	189.46	0.457	10142.03	22187.32	0.105
Jagtial	156.29	189.10	0.453	9219.09	20328.94	0.103
Karimnagar	145.53	180.87	0.474	6860.23	14470.09	0.116
KumuramBheem	165.54	196.11	0.457	11616.63	25411.04	0.105
Nirmal	151.03	189.25	0.462	8415.17	18218.94	0.108
Nizamabad	145.58	180.96	0.474	6881.91	14522.76	0.116
Peddapalli	169.74	198.51	0.504	11211.69	22223.83	0.136

Since t_{cal} values related to means were between $t_{critical}$ and table 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. Hence, we can conclude that the selected ARIMA(2,0,2)(2,1,0) and ARIMA(2,0,2)(1,1,0)seem to provide an adequate predictive model for evapotranspiration.Combining the use of Remote Sensing data to estimate the evapotranspiration and the use of Seasonal ARIMA model provides the keystone of advance and rational water resources management in arid ecosystems to be agreed with similar results conducted by Landeraset al. (2010) and Patil et al. (2022).

Conclusion:

From the trend shown in both estimation and the forecasting of potential evapotranspiration values, the Seasonal ARIMA models have an ability to forecast potential evapotranspirationwith anoptimum accuracy over all the districts of NTZ. From the basic statistical analysis conducted in the study, it is revealed that the difference between the observed and forecasted mean were found to be non-significant. Since the trends in the potential evapotranspiration estimation were replicated trends in the forecasted potential evapotranspiration, hence forecasting of evapotranspiration would be a powerful tool for attribution studies. The prediction of potential evapotranspiration using SARIMA model hence guarantees reliable project planning, design and operating of irrigation systems.

Declarations:

Disclaimer: The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References:

Afzaal, H., Farooque, A.A., Abbas, F., Acharya, B. and Esau, T., 2020. Precision irrigation strategies for sustainable water budgeting of potato crop in Prince Edward Island. *Sustainability*, *12*(6), p.2419.

Ahmad, S., Khan, I.H. and Parida, B.P., 2001. Performance of stochastic approaches for forecasting river water quality. *Water research*, *35*(18), pp.4261-4266.

Bogawski, P. and Bednorz, E., 2014. Comparison and validation of selected evapotranspiration models for conditions in Poland (Central Europe). *Water Resources Management, 28*(14), pp.5021-5038.

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G.E. and Jenkins, G.M., 1976. Time series analysis: Forecasting and control San Francisco. *Calif: Holden-Day*.

Brockwell, P.J. and Davis, R.A. eds., 2002. *Introduction to time series and forecasting*. New York, NY: Springer New York.

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*, pp.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*, *13*(2), pp.119-122.

Haan, C. T., 1977, “Statistical methods in hydrology”, Iowa State Press, Iowa.

Karimi, P. and Bastiaanssen, W.G., 2015. Spatial evapotranspiration, rainfall and land use data in water accounting–Part 1: Review of the accuracy of the remote sensing data. *Hydrology and Earth System Sciences*, *19*(1), pp.507-532.

Kharrou, M.H., Simonneaux, V., Er-Raki, S., Le Page, M., Khabba, S. and Chehbouni, A., 2021. Assessing irrigation water use with remote sensing-based soil water balance at an irrigation scheme level in a semi-arid region of Morocco. *Remote Sensing*, *13*(6), p.1133.

Krishna, P.R., 2019. Evapotranspiration and agriculture—A review. *Agric. Rev*, *40*, pp.1-11.

Martí, P., Gasque, M. and Royuela, Á., 2010. Discussion of “Forecasting Weekly Evapotranspiration with ARIMA and Artificial Neural Network Models” by G. Landeras, A. Ortiz-Barredo, and JJ López. *Journal of Irrigation and Drainage Engineering*, *136*(6), pp.435-438.

Lehmann, A. and Rode, M., 2001. Long-term behaviour and cross-correlation water quality analysis of the river Elbe, Germany. *Water Research*, *35*(9), pp.2153-2160.

McLeod, A.I. and Sales, P.H., 1983. Algorithm AS 191: an algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Applied Statistics*, pp.211-223.

Melard, G., 1984. Algorithm AS 197: a fast algorithm for the exact likelihood of autoregressive-moving average models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *33*(1), pp.104-114.

Mishra, A.K. and Desai, V.R., 2005. Drought forecasting using stochastic models. *Stochastic environmental research and risk assessment*, *19*(5), pp.326-339.

Nolz, R., 2016. A review on the quantification of soil water balance components as a basis for agricultural water management with a focus on weighing lysimeters and soil water sensors/Ein Überblick über die Ermittlung von Wasserhaushaltsgrößen als Basis für die landeskulturelle Wasserwirtschaft mit Fokus auf Lysimeter und Bodenwassersensoren. *Die Bodenkultur: Journal of Land Management, Food and Environment*, *67*(3), pp.133-144.

Pankratz, A., 2009. *Forecasting with univariate Box-Jenkins models: Concepts and cases*. John Wiley & Sons.

PATIL, R., Nagaraj, D.M., Polisgowdar, B.S. and RATHOD, S., 2022. Forecasting potential evapotranspiration for Raichur district using seasonal ARIMA model. *Mausam*, *73*(2), pp.433-440.

Oi, W. and Zheng, M., 2006. Winters and ARIMA model analyses of the lake level of salt Lake Zabuye, Tibetan Plateau. *Journal of Lake Sciences*, *18*(1), pp.21-28.

Raziei, T., Daneshkar Arasteh, P., Akhtari, R. and Saghafian, B., 2007. Investigation of meteorological droughts in the Sistan and Balouchestan province, Using the standardized precipitation index and Markov chain model. *Iran-Water Resources Research*, *3*(1), pp.25-35.

Salas, J.D., 1993. Analysis and modelling of hydrological time series. *Handbook of hydrology*, *19*.

Sentas, A. and Psilovikos, A., 2010, June. Comparison of ARIMA and transfer function (TF) models in water temperature simulation in dam–lake Thesaurus, eastern Macedonia, Greece. In *Proceedings of the International Symposium: Environmental Hydraulics, Athens, Greece* (Vol. 2, pp. 929-934).

Shumway, R.H. and Stoffer, D.S., 2005. Time Series Analysis and Its Applications Springer-Verlag New York. *Inc., Secaucus, NJ, USA.*

Toufani, P., Mosaedi, A. and FAKHERI, F.A., 2011. Prediction of precipitation applying wavelet network model (case study: Zarringol station, Golestan province, Iran).

Vandaele, V.V., 1983. Applied Time Series and Box-Jenkins Models Academic press. *Nevv York.*

Hongrui, W., Letian, Y., Changming, L., Chi, Y. and Peng, L., 2007. Problems in wavelet analysis of hydrologic series and some suggestions on improvement. *Progress in Natural Science*, *17*(1), pp.80-86.