International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10841-10847

Forecasting Evapotranspiration for Irrigation Scheduling using Neural Networks and ARIMA

Vardaan Kishore
Student, Department of Computer Science and Engineering,
SRM University, SRM Nagar, Potheri, Kattankulathur Chennai, India.

Orcid Id: 0000-0001-5660-9681

M. Pushpalatha
Professor, Department of Computer Science and Engineering,
SRM university, SRM Nagar, Potheri, Kattankulathur Chennai, India.

Orcid Id: 0000-0003-2313-4274

Abstract

The prediction of evapotranspiration is necessary for a reliable management of irrigation systems. This paper is based on models used for the prediction of reference crop evapotranspiration in the area of Kanchipuram district,Tamil Nadu, INDIA. Two models Neural Networks and ARIMA have been presented based on their extensive use in time-series forecasting. Based on the results of model comparisons it is concluded that ARIMA and ANN provide a reasonably accurate prediction of reference crop evapotranspiration. The obtained results show that the ARIMA model is a very effective and reliable prediction model for short term forecasts. This can be helpful in saving excess water wasted in irrigation.

Keywords: ARIMA, reference crop evapotranspiration, Artificial neural networks, Time series forecasting.

INTRODUCTION

Irrigation scheduling involves determining both the timing of irrigation and the quantity of water to apply. It is an essen- tial daily management practice for a farm manager growing irrigated crops. Proper timing of irrigation can be done by monitoring the soil water content or monitoring the crop in the field. Plant stress responses provide the most direct measure of identifying the plant demand for water. However, it should be noted that while plant stress indicators provide a direct measure of when water is required, they do not provide a direct volumetric measure of the volume of water required to be applied [1].Irrigation scheduling using Evapotranspiration information is like a checkbook account- ing procedure. Evapotranspiration the amount of crop water withdrawal that must be balanced against water deposits of rainfall and irrigation. The water balance must be kept within the limits of crop stress as determined by the field condition, irrigation capacity, and crop variety. Through the scheduling procedure, the amount of water application required and the time of application can be determined [2]. This is particularly important for the agricultural areas where the evapotranspiration prediction guarantees a reliable project planning, design and operating of an irrigation systems. Reference crop evapotranspiration is the rate of evapotranspiration from an extended surface of 8 to 15 cm tall, green cover of uniform height, actively growing, completely shading the ground and not short of water [3]. The data provided in [4](the monthly values of reference crop evapotranspiration) were divided into month wise.

Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. This modeling approach is particularly useful when little knowledge is available on the underlying data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables. One of the most important and widely used time series models is the autoregressive integrated moving average (ARIMA) model. The popularity of the ARIMA model is due to its statistical properties as well as the well- known Box–Jenkins methodology [5] in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models [6]. Although ARIMA models are quite flexible in that they can represent several different types of time series, i.e., pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARIMA) series, their major limitation is the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore, no nonlinear patterns can be captured by the ARIMA model. The approximation of linear models to complex real-world problem is not always satisfactory. Recently, artificial neural networks (ANNs) have been extensively studied and used in time series forecasting. [7] presented a recent review in this area. The major advantage of neural networks is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data.

In this paper the two models are compared by training them on a data set from 1901-1996 and the tested against the years 1997-2002.Both models show good results. These models proposed can be helpful in a way where the excessive water which is wasted in irrigation purposed can be controlled and yet be sufficient for a good crop yield.

PREVIOUS WORK

The ARIMA or Box-Jenkins methodology has been used in a study, the monthly maximum of the 24-h average time-series data of ambient air quality—sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and suspended particulate matter (SPM) con- centration monitored at the six National Ambient Air Quality Monitoring (NAAQM) stations in Delhi, was analyzed using this modeling approach [8]. [9] used ARIMA model for forecasting rainfall of Rahuri Region, India. 10 also used the artificial neural network and as well as ARIMA models to forecast rainfall. ANNs have been used to forecast or estimate monthly Evatranspiration based on the historical Evapotranspi- ration data [11] . [12] used the artificial neural network and Auto-Regression (AR) models to the river flow forecasting problem. A comparative study of both ANN and the AR conventional model networks indicated that the artificial neural networks performed better than the AR model. [13] forecasts evapotranspiration for humid and semi-humid region using ARIMA. [14]forecasted remotely sensed daily ETo of Nile Delta Region, Egypt. The methods proposed in the above papers use ARIMA and ANN effectively while calculating the evapotranspiration or other data, but the models used especially the ANN are complex and because various other factors are also taken into account it can become a complex and time consuming process. In this paper we use the ARIMA and a simple ANN-NAR(Non-linear Auto Regressive) model with a feedback loop to compute the future values to an extended period of time.

METHODOLOGY

A. The ARIMA model

ARIMA is an important method used in various time series prediction models. ARIMA is made up of two models, An autoregressive model of order p which is AR(p) and a moving average model of order q which is MA(q) and if we include a difference term d which is used to achieve stationarity, this integrated model is represented as ARIMA(p,d,q). The equation of a simple ARIMA model is as follows:

$$y_t = \theta_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots\cdots + \Phi_p y_{t-p}$$
$$+ \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots\cdots - \theta_p \varepsilon_{t-q}$$

One central task of the ARIMA model building is to determine the appropriate model order (p,q). The basic idea of model identification is that if a time series is generated from an ARIMA process, it should have some theoretical autocorrelation properties. By matching the empirical autocorrelation patterns with the theoretical ones, it is often possible to identify one or several potential models for the given time series. Box and Jenkins [Box and Jenkins, 1990] proposed to use the autocorrelation function and the partial autocorrelation function of the sample data as the basic tools to identify the order of the ARIMA model. In the identification step, data transformation is often needed to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. A stationary time series has the property that its statistical characteristics such as the mean and the autocorrelation structure are constant over time. Once a tentative model is specified, estimation of the model parameters is straightforward. The parameters are estimated such that an overall measure of errors is minimized. The last step of model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors, t, are satisfied. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the model to the historical data. If the model is not adequate, a new model should be identified, which is again followed by the steps of parameter estimation and model verification. The final model can then be used for prediction purposes.

B. The ANN-NAR model

A number of ANN models and training algorithms have been proposed for various applications. However, in this paper one problem being tackled, ET time series prediction; where ET output is forecasted using previous output data Multilayer feed forward network has been used in . It consists of an input layer, hidden layer and an output layer. A large number of hidden neurons may deteriorate the performance of the network as it requires huge storage memory for network variables, which in turn complicates the training. However, if very few numbers of neurons are used in the hidden layer, the network will not be able to adjust the weights and biases properly during training, resulting in over fitting. Over fitting makes the network excessively complex, generating random error and providing very poor classification. The Lavenberg marquardt (LM) back propagation training algorithm has been followed in both the cases. It is a combination of gradient descent algorithm (GDA) and Gauss- newton iteration which does not require the calculation of hessian matrix(involving second order partial derivatives) [15].The LM algorithm tries to minimize the MSE i.e. the performance function to be minimized is form of a sum of squares.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 10841-10847

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2$$

The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T$$

After the neural network is trained, it can forecast data based on the historical input presented to it.

RESULTS

A. ANN-NAR model:

It predicts the future values of the time series using the previously known values of the series. In this case no external input is fed along with the target data. The equation defining

$$y(t) = f(y(t-1), y(t-2), ...y(t-n_y))$$

here y(t) is predicted using the previous values of Potential Evapotranspiration.

The neural network model used has the following parameters as described in table 1:

Table 1: Parameters Used In Training The Model

Parameters	Values
Number of previous data used in model	4
Number of input delays	2
Number of feedback delays	1
Number of neurons in hidden layer	10
% Division of used data for training/validation/testing	50/25/25

These parameters were trained on the Levenberg-Marquardt (LM) algorithm. It adjusts the bias values according to the target time series data. The network was trained for 100 epochs but the network converged at 10 epoch and an MSE of 0.00401

Fig4 shows actual and predicted data plotted together respectively. The time series prediction neural network is trained with an open loop, where instead of feeding the predicted data as the next input to the model, an actual value is fed from the available data in order to ensure better accuracy, but such a model can only do one step prediction and hence the loop is closed for doing multiple time series prediction like in this particular case. Regression value gives an estimate of the relationship between the actual data and our model output. For a highly accurate model, the data points should lie completely over the 45 degree fit line and the R value should be equal to 1 The R value of the model ranged around 0.09 for NAR model.

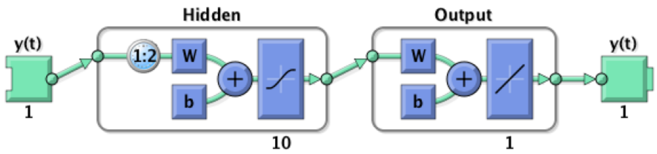

Figure 1: NAR Network

Figure 2 shows the regression values and plot for training validation and testing, as the data is taken month wise for the prediction purpose the values and closely related and this results in it being an average fit.

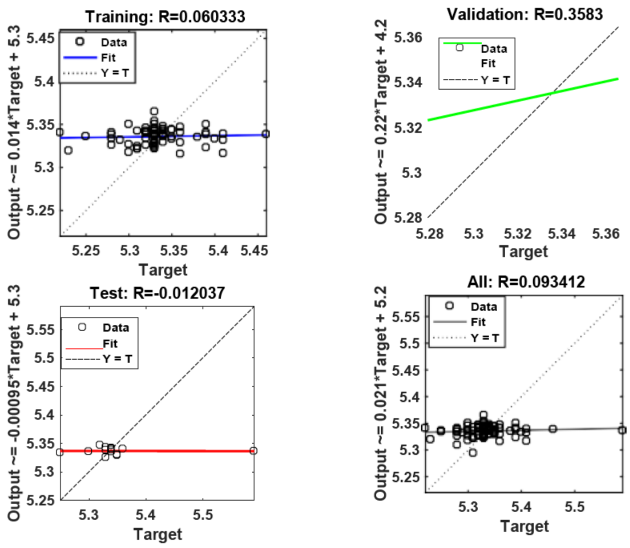

Figure 2: Regression values for training, validation and testing-NAR model

Figure 3 below shows the performance of the Neural Network architecture an as we can see the MSE performance, it has a decreasing slope with every epoch and this indicates that the network is converging.

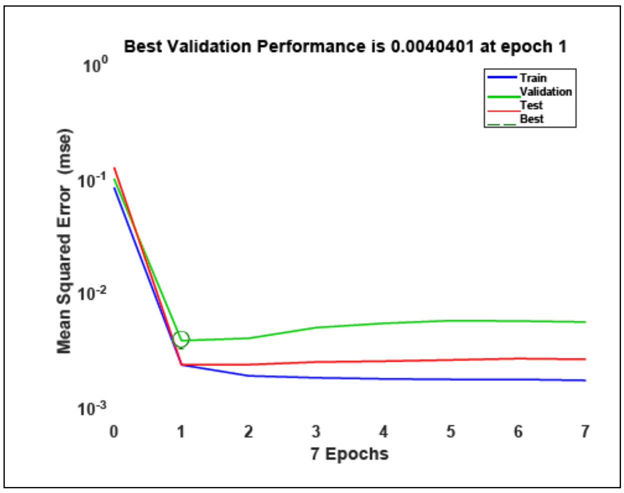

Figure 3: Performance of Model after every epoch

Figure 4 below shows the Original data along with the Fit generated by Neural Network architecture, it is noted that the fit is moderately accurate with respect to the original data of the Reference Evapotranspiration.

MULTI STEP AHEAD PREDICTION:

To perform multi step ahead prediction we need to close the loop as shown in the figure below. The closed loop works by sending the predicted output back to the input to help adjust the values of the weights in the network for better prediction of the time-series.

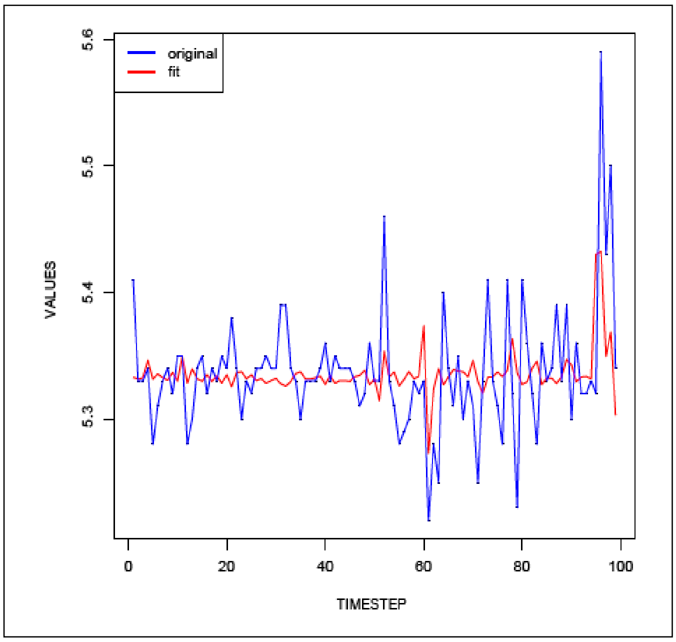

Figure 4: Original and Generated data on same graph

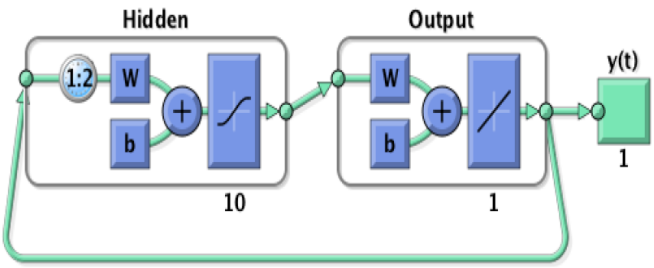

Figure 5: Closed NAR network for Multi-Step Prediction

The predicted values of the next 5 time-steps from year 1997-2002 are show in table number 2:

Table 2: Predicted Values For the Next 5 Years.

YEAR	Values
1997	5.2675
1998	5.3940
1999	5.3489
2000	5.3205
2001	5.3405
2002	5.3394

B. ARIMA Model

TEST FOR STATIONARITY: AUGMENTED DICKEY-FULLER (ADF) TEST:

Our null hypothesis (H0) in the test is that the time series data is non-stationary while alternative hypothesis (Ha) is that the series is stationary. The hypothesis then is tested by applying the ADF test to the time series data. The ADF test result, as obtained upon application, is shown below:

Dickey-Fuller = -5.5395, Lag order = 3, p-value = 0.01

We, therefore, fail to accept the H0 and hence can conclude that the alternative hypothesis is true i.e. the series is stationary in its mean and variance. Thus, there is no need for differencing the time series and we adopt d = 0 for our ARIMA(p,d,q) model. This test enables us to go further in steps for ARIMA model development i.e. to find suitable values of p in AR and q in MA in our model. For that, we need to examine the correlogram and partial correlogram of the stationary time series. .

Correlogram and Partial correlogram :

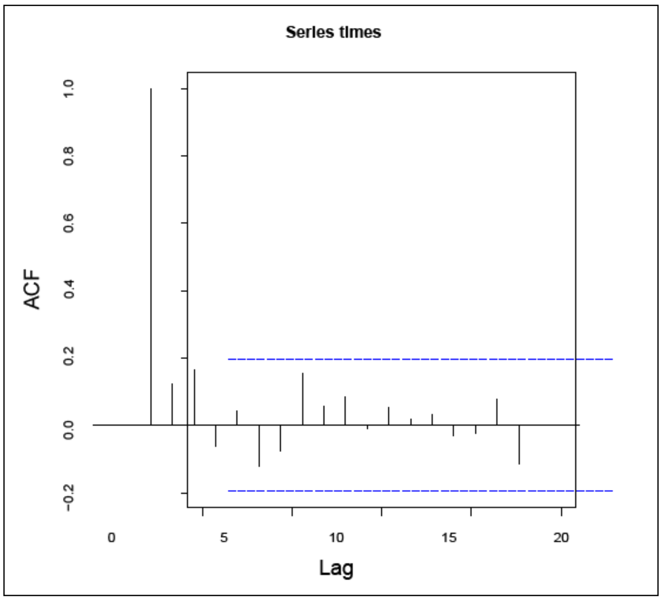

Figure 6: ACF plot

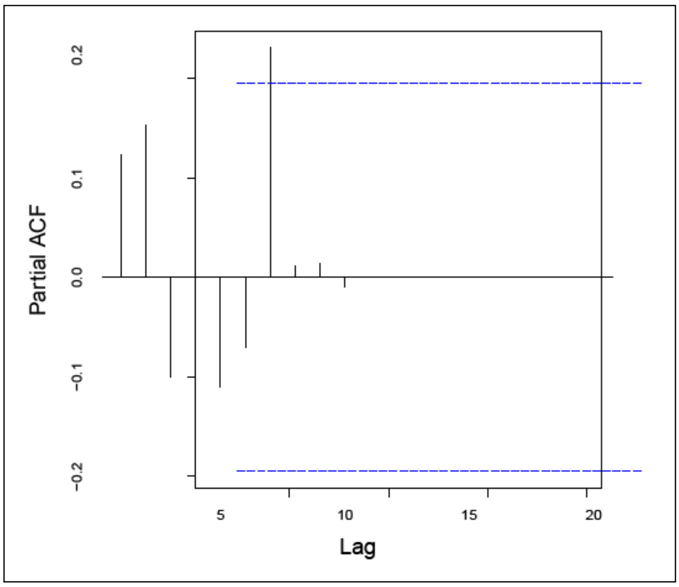

Figure 7: PACF plot

Since the correlogram (ACF) tailing off to zero after lag 2 and the partial correlogram (PACF) tailing off to zero after lag 2 (omitting the outlier), we can define the following possible ARIMA (auto regressive moving average) models for the time series data of Evapotranspiration in Kanchipuram:

1. An ARIMA(2,0) model i.e. autoregressive model of order p=2 since the partial autocorrelation is zero after lag 2 and the autocorrelation is zero.
2. An ARIMA(0,2) model i.e. moving average model of order q=3 since the autocorrelation is zero after lag 3 and the partial autocorrelation is zero.
3. An ARIMA(p,q) model i.e. a mix model with p and q both greater than 0 since autocorrelation and partial autocorrelation both tail off to zero.

Since ARIMA(2,0) has 2 parameters in it, ARIMA(0,2) has 2 parameters in it and ARIMA(p,q) has at least 2 parameters in it, therefore, by using principle of parsimony, the models ARIMA(2,0) and ARIMA(p,q) are the best candidate models for further step. In the next step, we have to device the best ARIMA model using the ARIMA(2,0) model (with p=2 & q=0), ARIMA(p,q) mixed model (with p & q both greater than 0), and order of differencing .

Therefore, based upon the conditions, we can have only following three tentative ARIMA(p,d,q) models: ARIMA(2,0,0), ARIMA(2,0,1), and ARIMA(2,0,2) .

To select as the best suitable model for forecasting out of three above, we will choose the one with lowest BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) values. Following Table 3 summarizes the output of each of the fitted ARIMA model in our time series:

Table 3: Comparison of Different ARIMA Models

ARIMA model	$\sigma 2$	Log Likelihood	AIC	BIC
(2,0,0)	0.002447	160.3	312.6	302.1359
(2,0,1)	0.002406	161.12	312.24	299.1679
(2,0,2)	0.002317	162.8	313.59	297.9023

From the above table we can see that the best model would be ARIMA(2,0,2) and hence this model can be the best predictive model for making forecasts for future values of our time series data.

Forecasting using selected ARIMA model :

We now will fit the chosen ARIMA(2,0,2) model to forecast for the future values of our time series. Following Table 4 shows the forecast for the next 5 years with 80%, 95% and 99.5% (low and high) prediction intervals

Figure 8 below shows the plot for 5 years forecast of the Reference Evapotranspiration by fitting ARIMA(2,0,2) model to our time series data.

Table 4: 5-Year Future Forecast of Reference Evapotranspiration Of Arima

Prediction	Forecast	Low 80	High 80	Low 95	High 95
1997	5.339368	5.2811	5.3975	5.2504	5.4283
1998	5.308109	5.2499	5.3662	5.2191	5.3970
1999	5.326886	5.2684	5.3853	5.2374	5.4163
2000	5.334929	5.2764	5.3934	5.2454	5.4244
2001	5.335005	5.2764	5.3935	5.2455	5.4244
2002	5.334181	5.2756	5.3926	5.2447	5.4236

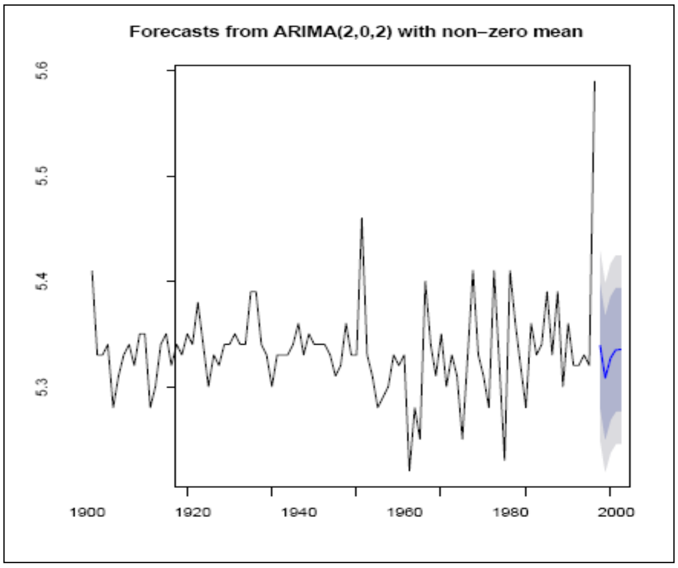

Figure 8: Forecast from ARIMA(2,0,2)

COMPARISON OF THE TWO MODELS

The results of the two models are compared over a 5 year time period . The model was trained on data from 1901-1996 and it was tested by prediciting the values from 1997-2002, the predicted results were compared against the data set and the accuracy measure for each yaer was presented in Table number 4.

Table 5: Accuracy of Each Forecasted Year for ANN vs ARIMA

Year	Original Data	Predicted values		Accuracy	
Values	Values	ANN	ARIMA	ANN	ARIMA
1997	5.43	5.26	5.33	97%	98.2%
1998	5.5	5.39	5.30	98%	96.4%
1999	5.34	5.34	5.32	100%	99.7%
2000	5.43	5.32	5.33	98%	98.2%
2001	5.22	5.34	5.33	97.8%	98%
2002	5.21	5.33	5.33	97.75%	97.7%

Tabel 5 shows the Overall Error measure comparisons of the two models, as we can see both models have a similar MAE value suggesting that they are both useful in prediction of the time series data.

Table 6: Overall Error Measures Of The Two Models

MEASURE(One-Step Ahead)	MAE
ARIMA	0.0317
ANN	0.0282

CONCLUSION AND FUTURE WORK

The prediction of evapotranspiration guarantees reliable project planning, design and operating of irrigation systems. This paper presents the use of ANN and ARIMA models for predicting reference crop evapotranspiration in the area of Kanchipuram, India. The ANN model bases the prediction on the appropriate previous values of reference crop evapo-transpiration, ARIMA model was also presented. According to the comparison results it can be said that seasonal ARIMA models guarantee the most reliable prediction of reference crop evapotranspiration for one step ahead and ANN can be used for long term forecasts. Future work in the area of forecasting can be done using the new state of the art machine learning models such as deep learning and reinforcement learning which can model the data in a better way and can give accurate predictions.

REFERENCES

[1] Jamal Abubaker. Irrigation scheduling for efficient water use in dry climates. 09 2017.

[2] Danny H. Rogers. Using evapotranspiration reports for furrow irrigation scheduling. Cooperative Extension Service, Kansas State University, 2013.

[3] Dirk Raes Martin Smith Richard G. Allen, Luis S. Pereira. Crop evapotranspiration - Guidelines for computing crop water requirements. Food and Agriculture Organization of the United Nations, 1998.

[4] Indian water portal. Indian water portal.

[5] George Edward Pelham Box and Gwilym Jenkins. Time Series Analysis, Forecasting and Control. Holden-Day, Incorporated, 1990.

[6] Ed. McKenzie. General exponential smoothing and the equivalent arma process. Journal of Forecasting, 3(3):333–344, 1984.

[7] Peter Zhang, Eddy Patuwo, and Michael Hu. Forecasting with artificial neural networks: The state of the art. 14:35–62, 03 1998.

[8] Pragati Sharma, Avinash Chandra, and S. C. Kaushik. Forecasts using box–jenkins models for the ambient air quality data of delhi city. Environmental Monitoring and Assessment, 157(1):105–112, Oct 2009.

[9] P. G. Popale and S.D. Gorantiwar. Stochastic generation and forecasting of weekly rainfall for rahuri region. International Journal of Innovative Research in Science,Engineering and Technology, 3, April, 2013.

[10] Elena Toth, A Brath, and Alberto Montanari. Comparison of short-term rainfall prediction models for real-time flood forecasting. 239:132–147, 12 2000.

[11] S. A Tahir. Estimating potential evaporation using artificial neural network. 10th ICID Afro-Asian Regional Conference on Drainage, 2 A, 1998.

[12] Khaled Al Khnaifes Abdel karim M. Baareh, Alaa F. Sheta. Forecasting river flow in the usa: A comparison between auto-regression and neural network nonparametric models. Journal of Computer Science, 2006.

[13] Amirpooya sarraf Arash Asadi, Seyed Farnood Vahdat. The forecasting of potential evapotranspiration using time series analysis in humid and semi humid regions. American Journal of Engineering Research, 2, issue-12:296–302, 2013.

[14] Aris Psilovikos and Mohamed Elhag. Forecasting of remotely sensed daily evapotranspiration data over nile delta region, egypt. Water Resources Management, 27(12):4115–4130, Sep 2013.

[15] M. T. Hagan and M. B. Menhaj. Training feedforward networks with the marquardt algorithm. IEEE Transactions on Neural Networks, 5(6):989–993, Nov 1994.