Predicting Irrigated Water Demands Based on Analyzing Time Series

Nghiem Van Tinh
Thai Nguyen University of Technology – Thai Nguyen University, Thai Nguyen, Viet Nam
nghiemvantinh@tnut.edu.vn

Bui Thi Thi
Thai Nguyen University of Technology – Thai Nguyen University, Thai Nguyen, Viet Nam
buithithi@tnut.edu.vn

Abstract* — *Water is one of the most critical resources, and its sustainable development and management is essential to any society. Current increases in demand, particularly for urban and industrial use coupled with climate change-induced uncertainty over supply, means that water resources are closer to sustainability limits than ever before and in many areas limits have already been exceeded. The need for good demand forecasting becomes important for effective water resources management. This study proposes and experimentally evaluates univariate time series models that predict the value of reference evapotranspiration, a metric of the water loss from crop to the environment. Reference evapotranspiration plays an essential role in irrigation management since it can be used to reduce the amount of water that will not be absorbed by the crop. The experiments performed under the meteorological dataset generated by a weather station. Moreover, the results indicated that the method is a viable and lower cost solution for predicting ET0, since only a variable needs to be observed.

Keywords—Time series analysis; Predicting Irrigation; ARIMA; SARIMA.

I. INTRODUCTION

Water is vital to every civilization and to the development process. Despite increasingly efficient water use, freshwater resources are becoming scarce, both in quality and in quantity, in many developed and developing countries. This fact can be attributed to population growth, industrialization, growing agricultural demand, poor water management practices and climatic variations. Efficient management of existing water resources has become an increasingly important aspect of water policy in the Brazil. The importance of efficient water use and management is supported by rapidly growing water demand and constant and/or decreasing supplies of water in the many parts of the Brazil. Besides that, population growth and changes in climate directly impact on worldwide food security. One of the primary objectives of agricultural research is to find improved ways to produce food. According to [1], 72% of freshwater is consumed in irrigation, in Brazil. It is estimated that a massive portion of this amount is wasted due to poorly executed irrigation and lack of control from farmers about the exact amount of water to use in irrigation process. Evapotranspiration value (RT_0) plays a key role in support to decision making in irrigation management, which is the simultaneous occurrence of evaporation and transpiration processes in a crop, measured in millimeters per a unit of time. We use the following equation to compute it: $RT_m = K_c \times ET_0$, where K_c is the crop coefficient *c*, given at INMET website[1], ET_0 is the reference crop evapotranspiration, which corresponds to the evapotranspiration rate of a grass surface. The value of RT_0 is very relevant to management and scaling in irrigation since it gives the information of how much water the crop loses to the environment [1]. The traditional *Penman - Monteith* method [2] used to compute ET_0 is complex and does not tolerate the unavailability of some of its variables, which makes its use unfeasible. The paper [3] proposes a Machine Learning based approach to forecast RT_0 based on Linear Regression [4] and M5P [5]. Despite the good results obtained in both techniques, they are multivariate models, which means that it requires a weather station with many sensors to capture all the required variables, and there is no guarantee that models will fit, as well as in the absence of some variables.

Experiments performed by [6] with univariate time series model demonstrated the Autoregressive Integrated Moving Average (ARIMA) [7] model as a promising technique to achieve good accuracy performance in the forecast of financial time series. ARIMA model aims at describing the correlations in the data with each other. An improvement over ARIMA is Seasonal ARIMA (SARIMA) [7], which takes into account the seasonality of dataset and was successfully used in short-term forecast [8]. In this paper, we use both approaches in our experiments. The key contributions of this paper are: (i) offer an accurate and lower cost solution to estimate ET_0, since only a variable needs to be monitored; (ii) compare the performance of ARIMA, SARIMA, Linear Regression and M5P with respect to mini- mization achieved in the error rates in prediction; and (iii) release the dataset used in this work, for research and possible improvements by the scientific community.

The rest of this paper is organized as follows: Section 2 explains the proposed approach. The steps necessary to accomplish our goals are presented in Section 3. Section 4 compares the experiment results of the proposed method with its counterparts and its analysis. Finally, Section 5 offers some conclusions and proposes future developments.

II. Time Series Forecasting

A time series (TS) is a series of data records indexed by dates. A time series model supposes that a series Zt could be defined as Zt = Tt + St + αt, being T the tendency, S the seasonality and α the white noise, at a moment t [9]. Most of the TS models work on the assumption that the T S is stationary, i.e., its statistical properties such as mean and standard deviation remain constant over time. Due to many real time series being non-stationary, statisticians had figured out ways to make TS stationary [7].

In particular, differencing operator (∇) is a simple and efficient operator to transform a non-stationary *T S* to stationary. It is defined by the equation: $\nabla Z_t = Z_t - Z_{t-1}$, where Z is a *T S* at a moment t [9]. In other words, we take the difference of the observation at a particular instant t with that at the previous instant $t-1$. The ARIMA model takes three hyper-parameters *p, d, q*, which capture the key elements of the model, which are: (i) Autoregression (*AR*), a regression model that uses the relationship between an observation and a number (*p*) of lagged observations; (ii) Integrated (*I*), the number (*d*) of differentiation required to obtain stationarity; (iii) Moving Average (*MA*), an approach that takes into accounts the dependency between observations and the residual error terms when a moving average model is used for the lagged observations (*q*) [7, 8].

The SARIMA model incorporates both seasonal and non-seasonal factor in a *T S* data, its signature is *SARIM A*(*p, d, q*)×(*P, D, Q*)*S*, where *p* and *P* are the non-seasonal and seasonal AR order; *d* and *D* are the non - seasonal and seasonal differencing; q and Q are the non - seasonal and seasonal MA order; and S is the time span of repeating seasonal pattern, respectively [8].

III. Material and Methodology

A. Data source

The climatic data were collected by a weather station, in the period from January, 1st to November, 29th of 2017 in the city of Quixada´, Ceara´, Brazil. The original dataset contains 7941 hourly records, and it is composed of the features described in Table 1. This dataset is available in https://github.com/Dieinison/ProjectET0/blob/master/dataset.csv.

Table I. Samples of dataset

Date	Atmospheric pressure		Air Temperature			Relative huidity			Solar radiation		Temperature		Precipitation	Wind Speed	ET0
	Max	Min	Max	Min	Mean	Max	Min	Mean	Total	Mean	Max.	Min.			
29-11-17	620.5	599.7	21.4	19.6	32	55.2	45.3	50.1	1610	12.7	21.4	19.6	0.0	1.58	0.095
28-11-17	620.2	599.7	21.7	19.4	32	52.3	41.9	46.9	1638	11.9	21.7	19.4	0.0	1.73	0.109
27-11-17	620.4	599.6	20.9	19.1	34	45.8	39.7	42.3	1620	19	20.9	19.1	0.0	2.10	0.147
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

We aggregated the original hourly data on a daily basis. Furthermore, we detected outliers observations through Proximity-Based Outlier Detection technique [10] and remove them. The tuples contain the values including Precipitation larger than or equal to 60, Minimum temperature smaller than or equal to 0 and Minimum relative humidity smaller than or equal to 20 which were removed. At the end of this procedure, 333 tuples remained.

B. Prediction models

To create the prediction models, we split the dataset into 80% for training and 20% for testing. Each algorithm produced its particular model using the attributes taken as input. Thus, we generated four distinct models, Linear Regression and M5P were created from all the attributes of the dataset, ARIMA and SARIMA models were generated only with ET_0. These models and their comparisons are presented in Section 4. For purposes of comparisons between the models generated, we used the same dataset (given by weather station from UFC Quixada´). We performed the prediction models by applying the Linear Regression and M5P algorithms, both implemented in the WEKA[2] tool. In order to forecast through ARIMA and SARIMA, we perform the Box-Jenkins methodology [7], defined as: (i) identification of the model, i.e., finding the appropriate orders for *p, d, q, P, D, Q, S*; (ii) estimation of the unknown parameters; (iii) validation of the model; and (iv) forecast future outcomes based on the known data.

C. Models Evaluations

To evaluate both techniques, the Mean Absolute Percentage Error (MAPE) are calculated as the evaluation metrics of the performance, defined by equation (1) as follow:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i-\hat{F_i}}{F_i}\right| * 100\% \qquad (1)$$

where i is the sample index, n is the total number of observations, F_i is the expected attribute value and $\hat{F_i}$ is the value output by the algorithm used [4]. Both metrics can range from 0 to ∞. They are negatively-oriented scores, which means lower values are better.

IV. Experiments and Results

As stated earlier, these experiments used a real dataset with observations collected from a weather station located in Campus UFC Quixada´, in Brazil. Initially, we generated the Machine Learning-based approaches, through WEKA tool. Due to lack of space, we do not present in this paper our Linear Regression and M5P prediction models. They are available in http://bit.ly/result_linear regression and http://bit.ly/result_m5p, respectively. With the view to generate time series models, we checked stationarity by plotting rolling average and rolling standard deviation as shown in Fig.1. The evaluated mean and standard deviation show significant instability over time, suggesting the data is non- stationary. Another technique to evaluate the non-stationary is the Dickey-Fuller (DF) test. The DF is a unit root test that evaluates the strength of trend in a time series component [11]. The output for DF test is shown in Table 3. As we can see, the DF Statistic is higher than the critical values, so this series is non-stationary. Therefore we can approach this with ARIMA models.

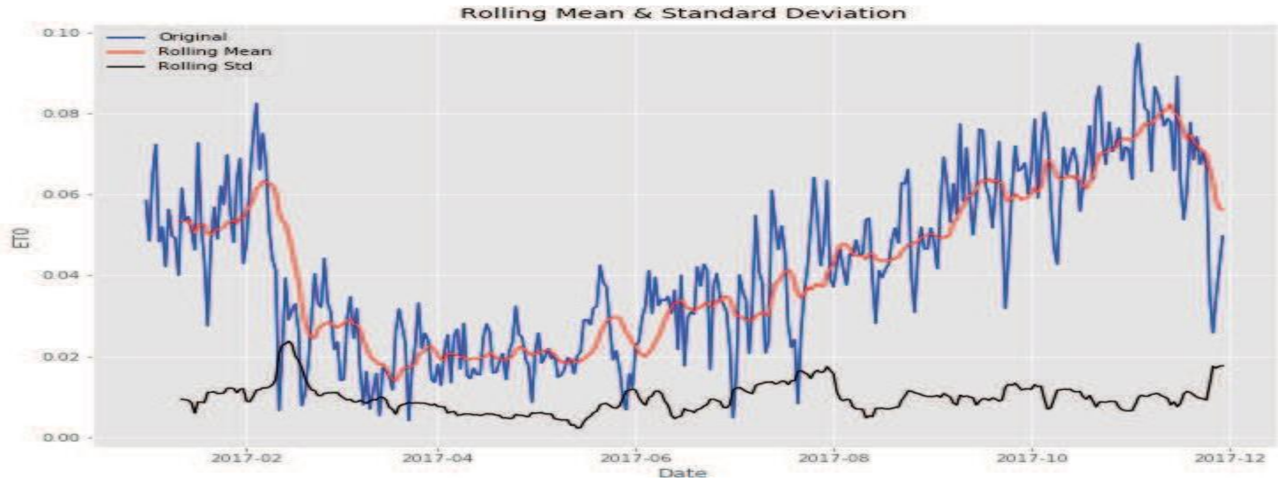

Fig. 1. Original ET_0

TABLE 2: RESULTS OF DF TEST

DF Statistic	-1.598421
Critical Value 1%	-3.398686
Critical Value 5%	-2.900503
Critical Value 10%	-2.610112

In order to obtain the optimal hyper-parameters for ARIMA and SARIMA models, we used a function, called *auto arima*, from Pyramid[3], an API under MIT License that provide an systematic approach to find the best hyper-parameters, based on a given information criteria, which in this case will be the Corrected Akaike Information Criterion (AIC_c) as recommended in [9]. This criterion includes a penalty term to discourage the fitting of too many parameters, i.e., the fitted model with the smaller value of AIC_c will be the best choice [11, 12]. Tables 3 and 4 present the parameters output by auto arima function for ARIMA and SARIMA models, respectively.

TABLE 3. ARIMA PARAMETERS

Parameter	Value
Difference order *d*	1
MA order *q*	1
AR order *p*	1

TABLE 4. SARIMA PARAMETERS

Parameter	Value
Seasonal AR order *P*	1
Seasonal difference *D*	1
Seasonal MA order *Q*	2
S	12
Difference order *d*	1
MA order *q*	1
AR order *p*	1

Table 5 shows MAPE generated from models. As we can see, the univariate ARIMA and SARIMA models presented error values very low as it is close to zero. A value of MAPE equal to zero would that the estimator is predicting observations with perfect accuracy. Besides, in Table 6, we showed statistical properties of our label variable, ET0, thus, as errors rates (MAPE) are less than the standard deviation, our results indeed show a good accuracy [13].

The results show an outperformance of multivariate model M5P, under MAPE metrics, over univariate time series models. Nevertheless, univariate time series models show us that these models indeed fit well the data, since there were small differences between predictions and expected values. Regarding *TS* models, ARIMA outperformed SARIMA in both metrics, indicating that our data is better fitted by a non- seasonal model.

TABLE 5. METRICS COMPARISONS BE- TWEEN TECHNIQUES

Model	**MAPE**
ARIMA	1.69%
Linear Regression	0.58%
M5P	0.57%
SARIMA	1.98%

TABLE 6. MEAN AND STANDARD DEVIATION OF OBSERVED ET_0

Statistic	**Value**
Mean	0.0439
Standard deviation	0.0453

Due to the costs of owning a weather station with many sensors, capture all the variables required for multivariate models might not be affordable for low-income farmers. In contrast, the results show us that an ARIMA model is an affordable solution for predicting ET_0 since only a variable needs to be monitored, with no need of multiples sensors.

V. CONCLUSION

The objective of this paper was to forecast the value of reference evapotranspiration, a metric of the water loss from crop to the environment based on evaluating univariate time series models. This study compares the accuracy of univariate ARIMA and SARIMA models with multivariate Machine Learning-based algorithms, Linear Regression and M5P. The results show that M5P outperform the other techniques. Despite that, this paper advocates the benefits of applying univariate time series algorithms to predict ET_0, since these models presented small differences between predictions and expected values, i.e., good accuracy. Besides, *TS* models might be an affordable solution for low-income farmers, since only a variable needs to be monitored. For future works, we aim at improving and validating our proposed models for other datasets and compare with deep learning based methods.

REFERENCES

[1] H. F. Thiago,M.D.C. Wagner and A.F. Marcus, “Atlas Irigacão: Uso da águana agricultura irrigada. Agência Nacional de Águas Brasília , DF, Brasil, 2017.

[2] R.G. Allen, L.S. Pereira, D. Raes, M. Smith, et al. Crop evapotranspiration- guidelines for computing crop water requirementsfao irrigation and drainage paper 56. *FAO, Rome*, 300(9):D05109, 1998.

[3] Caminha, H., Silva, T., Rocha, A., and Lima, S. Estimating reference evapotran- spiration using data mining prediction models and feature selection. *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017)*, 1:272–279, 2017.

[4] James, G., Witten, D., Hastie, T., and Tibshirani, R. An introduction to statistical learning, volume 112. Springer, New York, USA, 2013.

[5] Wang, Y. and Witten, I. H. Induction of model trees for predicting continuous classes, 1996.

[6] Siami-Namini, S. and Namin, A. S. Forecasting economics and financial time series: Arima vs. lstm. arXiv preprint arXiv:1803.06386, 2018.

[7] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. Time series analysis: forecasting and control. John Wiley & Sons, New Jersey, USA, 2015.

[8] Tseng, F.-M. and Tzeng, G.-H. A fuzzy seasonal arima model for forecasting, 2002.

[9] Brockwell, P. J. and Davis, R. A. *Introduction to time series and forecasting*, 2016.

[10] Tan, P.-N. et al. *Introduction to data mining*. Pearson Education India, India, 2006.

[11] P. Avishek, P. P. *Practical Time Series Analysis: Master Time Series Data Pro- cessing, Visualization, and Modeling using Python*. Packt, Birmingham, UK, 2017.

[12] Smith, T. G. *Pyramid: ARIMA estimators for Python*. MIT, USA, 2017.

[13] Legates, D. R. and McCabe, G. J. Evaluating the use of goodness-of-fit mea- sures in hydrologic and hydroclimatic model validation. *Water resources research*, 35(1):233–241, 1999.