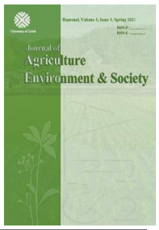
Agriculture, Environment & Society (2021) Volume 1, Issue 1, P. 27-30

Agriculture, Environment & Society

Journal homepage: www.aes.uoz.ac.ir



Evaluation of five data mining algorithms in predicting monthly potential evapotranspiration (case study: Shiraz)

Seyedhassan Mirhashemi *[a], Ali Asghar Bour [b]

[a] M.Sc Graduate of Water Engineering, Faculty of agriculture, University of Zanjan, Zanjan, Iran
[b] Master of Hydraulic Structures, graduated from Islamic Azad University, Science and Research Branch, Tehran, Iran

ARTICLE INFO

Article history:
Received: 1 August 2021
Accepted: 23 January 2022
Available online: 1 February 2022

Keywords:
Air temperature
CART algorithm
Sunshine hours
Wind speed
Weather station

ABSTRACT

Data mining algorithms were used in this study to predict Shiraz's monthly potential evapotranspiration. The CART (Classification and Regression Trees), M5P, K-star, M5Rules, and REP-Tree (Reduced Error Pruning Tree) algorithms were used to predict potential evapotranspiration. Meteorological data from the Shiraz weather station from 2001 to 2016 were used in this study. The CART algorithm performed better in estimating monthly averages, according to statistical indicators. The maximum amount of potential evapotranspiration was reached when the sunshine hours exceeded 9.5 hours and the wind speed exceeded 0.3 meters per second, according to the results. When there was less than 9.5 hours of sunshine and the air temperature was less than 2 °C, the potential evapotranspiration rate was the lowest. The sensitivity analysis revealed that the parameters of sunshine hours, air temperature, wind speed, and relative humidity had a positive effect on the CART algorithm's performance in estimating monthly evapotranspiration.

1. Introduction

The actual evapotranspiration of the crop is considered equal to the water requirement of the crop, which is necessary to estimate in all irrigation projects. For this purpose, several computational methods have been proposed to estimate the potential evapotranspiration of grass as a reference crop in different parts of the world. After calculating the potential evapotranspiration using the crop coefficient, the reference crop evapotranspiration (ET_o) is calculated and considered as the water requirement of the plant (Fooladvand, 2010).

Samadianfard and Panahi (2018) estimated the reference crop evapotranspiration of Tabriz city by both Thornthwaite and Hargreaves and data mining methods, including Support Vector Regression (SVR) and M5 Tree models. They concluded that data mining methods have a better ability to estimate reference crop evapotranspiration. Sameti et al. (2013) estimated the reference crop evapotranspiration of Shiraz and Kermanshah cities by Penman-Monteith and Hargreaves-Samani and compared their results with that of the M5 Tree model. They concluded that the M5 Tree model performs as well as the abovementioned methods.

Due to the occurrence of the phenomenon of climate change, the use of indicators such as precipitation and potential evapotranspiration is inevitable (Asadi Zarch et al., 2017). So far, many studies have been conducted to accurately estimate evapotranspiration, which has led to a variety of methods for estimating evapotranspiration (Martí et al., 2015). The Penman- Monteith method is a physical algorithm that incorporates aerodynamic and thermodynamic aspects (Asadi Zarch et al., 2017). In Northeast India, by comparing the methods of reference evapotranspiration and examining different equations, the results obtained from the algorithms based on solar radiation and air temperature are similar to the results of the FAO Penman-Monteith algorithm (Pandey et al., 2016).

Among all the methods, the FAO method introduced a kind of standard in the interpretation and use of various expressions, such as potential evapotranspiration and evapotranspiration of the reference crop. The FAO has suggested that the hypothetical reference surface should be considered as the reference surface, just like the vast

* Corresponding author.
E-mail address: Hassan.mirhashemi@yahoo.com (S.H. Mirhashemi)
http://dx.doi.org/10.22034/aes.2021.144191

surface of green grass with uniform height, active growth, complete shading of the ground and sufficient water for evapotranspiration (Allen et al., 1998). In the FAO approach, the surface characteristics that affect evapotranspiration are clearly quantified (Itenfisu et al., 2003). The rate of evapotranspiration from the reference surface, without water shortage, is called the evapotranspiration of the reference crop and is indicated by ETO (Allen et al., 1998).

Data mining is the process of selecting, identifying, and algorithmically processing large amounts of data. Another definition is the process of selecting, exploring, and algorithmically mining large amounts of data, to uncover hidden relationships and achieve results that are clearly beneficial to the database's owner (Giudici, 2003). Data mining algorithms can automatically develop these equations from information contained in the data set (Crows, 1999). Tree algorithms have been evaluated by Mirhashemi and Panahi (2014b) to predict potential evapotranspiration and air temperature. It has been found that the performance of tree algorithms in predicting potential evapotranspiration has been acceptable. The performance of the M5P algorithm for estimation of the potential evapotranspiration and air temperature at Rasht meteorological station is evaluated. The results showed that the M5P algorithm has a better performance in predicting potential evapotranspiration (Mirhashemi and Panahi, 2014a). In a study conducted by Mirhashemi and Panahi (2014b) at Arak Meteorological Station, the performance of the CART algorithm in predicting potential evapotranspiration and air temperature was evaluated.

Conducting research to determine the water requirements of different crops during the growing season is necessary in order to prevent water wastage with proper planning and also to use appropriate management methods for the future. One of these issues that can be applied in line with the mentioned goals for the future is the prediction of ETo for the future in order to enable better use of available water resources with proper planning. Application of data mining algorithms is a suitable tool for various predictions. It was found that the performance of the M5P algorithm in predicting potential evapotranspiration is better than that of the minimum and maximum air temperatures of Sari meteorological station (Mirhashemi and Panahi, 2014a). Therefore, data mining algorithms can predict ETo to make a good prediction for the required water resources in the future. The purpose of this study is to evaluate the performance of tree algorithms in predicting monthly potential evapotranspiration in Shiraz.

2. Materials and methods

2.1. Study area

Shiraz is located in the central part of Fars province, at an altitude of 1486 meters above mean sea level in the mountainous region of Zagros and has a temperate climate. The city is bounded on the west by Mount Drak, on the north by the mountains of Bamo, Sabzpooshan, Chehel Magham and Baba Koohi (from the Zagros Mountains).

2.2. Data used

Monthly meteorological data of forty-six-year period from 1960 to 2005 of the Shiraz synoptic weather station" were used as inputs data, which includes: mean air temperature, sunshine hours, dew point, relative humidity, mean wind speed and saturation vapor pressure deficit. Many scientists have studied the Penman-Monteith equation to estimate ETO (Allen et al., 1998).

The Penman-Monteith method was ranked as the best method for all climatic conditions. Application of the Penman-Monteith equation FAO-56 requires data on sunshine, Wind speed, Air temperature, vapor pressure, and relative humidity, but all these input variables are not readily available in every location.

In developed countries, all climatic variables encountered in application of the FAO - 56 Penman-Monteith method are collected and can be used when needed, but in developing countries this is not the case

To predict the potential evapotranspiration of the next month, the six abovementioned variables were used. That was considered on a monthly basis, month after month, as the input data and the monthly potential evapotranspiration the next month as the output data. CART, M5P, K-star, M5Rules, and REP-Tree were used in this study to predict "potential evapotranspiration of the next month,", therefore 75% of the data were used as the prediction algorithm and 25% as the test algorithm.

3. Results

The potential evapotranspiration of each month was calculated based on the potential evapotranspiration of the preceding month.

Values of monthly potential evapotranspiration are estimated by using the five algorithms and then compared with monthly potential evapotranspiration calculated by the Penman-Monteith equation. For the selection of the best algorithm for calculation of the monthly potential evapotranspiration, three statistical indices consisting of correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE) were used for comparing the values calculated by data mining and the Penman-Monteith equation. As can be seen in Table 1, the CART algorithm, with a correlation coefficient of 0.92, RMSE of 1.02, and MAE of 1.45, is the best algorithm for estimation of the monthly potential evapotranspiration.

Table 1. Comparison of five algorithms of data mining with three statistical indices.

Algorithms	R	MAE	RMSE
K-star	0.75	3.12	3.45
CART	0.92	1.45	1.02
M5Rules	0.78	2.95	3.06
REP-Tree	0.84	3.41	3.20
M5P	0.52	4.18	4.91

The diagram in Figure 1 shows the linear diagram of the algorithm in predicting potential evapotranspiration.

According to the tree diagram, it is divided into two main branches. The dividing factor is the two main branches of the sunshine hours. According to the diagram, the maximum amount of potential evapotranspiration was reached when the sunshine hours were more than 9.5 hours and the wind speed was more than 0.3 meters per second. Increasing sunshine hours and wind speeds have a significant effect on increasing potential evapotranspiration. The lowest potential evapotranspiration rate was when the sunshine was less than 9.5 hours and the air temperature was less than 2 ° C. Air temperatures of less than 2 ° C have a great effect on reducing potential evapotranspiration. The prediction of this branch probably occurred in the autumn and winter seasons.

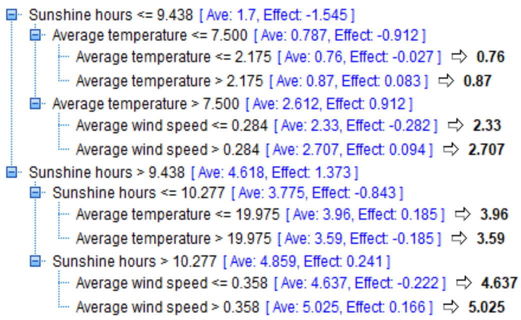


Figure 1. Linear diagram of the algorithm in predicting potential evapotranspiration

Figure 2 shows the importance of meteorological factors in predicting potential evapotranspiration by the CART algorithm. According to the figure, the factors of sunshine hours, air temperature, wind speed, and relative humidity have the greatest effect on predicting potential evapotranspiration.

3.1. Sensitivity analysis

To determine the most important factor for algorithming, "mean monthly potential evapotranspiration for the next month" via the CART algorithm was compared by changing the input data and using the statistical parameters. Which contains the "correlation coefficient", "root mean square error" and "mean absolute error" when compared to the third row, which includes five meteorological parameters, has the maximum "correlation coefficient" and minimum "square root error" and mean "absolute error".

As a result, five parameters were used to have the greatest impact on the performance of the CART tree algorithm.

The mentioned six parameters are listed in the first row, including the mean monthly air temperature, sunshine hours, dew point, relative humidity, wind speed, and saturation vapor pressure deficit, and the five parameters are listed in the last four rows, including sunshine hours, air temperature, wind speed, and relative humidity, which have the positive impact on the CART algorithm's proper functioning in estimating mean monthly evapotranspiration for the next month.

In Table 2, different combinations of input data were used to evaluate the sensitivity analysis of meteorological parameters.

In the first row, all meteorological parameters were used. In the next rows, the number of meteorological parameters decreased. Finally, according to the two statistical tests, the fifth row was selected as the best combination of input parameters. The fifth row of parameters, as in Figure 2, are sunshine hours, air temperature, wind speed and relative humidity.

These have a positive impact on the CART algorithm's proper performance in estimating average monthly evapotranspiration for the next month.

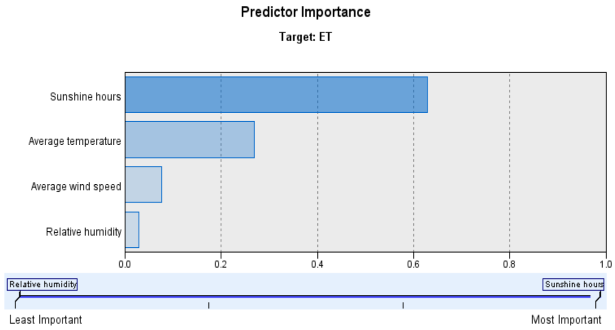


Figure 2. Importance of meteorological parameters in predicting potential evapotranspiration by CART algorithm.

The results of this study are similar to the results of Mirhashemi and Panahi (2014a) in predicting potential evapotranspiration and air temperature, in Rasht and Arak cities. They applied M5P and CART algorithms for the evaluation of the models. Also the results are in agreement with the results of Sameti et al. (2013) and Samadianfard and Panahi (2018).

It must be noted that in the current study, five data mining models were applied for evaluating different methods, while Sameti et al. (2013) have applied one model.

Table 2. The combination of input parameters to estimate the monthly potential evapotranspiration for the next month, using CART algorithm.

combination of input parameters*	R	MAE	RMSE
T, n, U, RH, T_d, sd	0.92	1.45	1.02
T, n, U, RH ,sd	0.92	1.50	1.89
T, n, RH, T_d, sd	0.93	1.27	1.01
T, n, U, sd	0.91	2.20	2.89
T, n, U, RH	0.95	1.1	0.81
T, n, U	0.90	2.31	3.04
T, n	0.89	3.41	3.17
T,U	0.88	3.82	3.51
n, U	0.86	3.89	3.80

*The parameters "dew point" (ºC), "relative humidity" (percent), "sunshine hours" (h), "saturation vapor pressure deficit" (hPa.), "wind speed" (m/s), "mean monthly air temperature '(ºC), are shown as T_d, RH, n, sd, U, and T respectively.

4. Conclusions

From this study, it can be concluded that the CART algorithm is a suitable model for estimation of the monthly potential evapotranspiration of the next month using meteorological elements. The CART algorithm can be used to estimate potential evapotranspiration in a variety of stations that are deficient in recorded meteorological elements. Furthermore, in some areas where there is a shortage of meteorological data, using data mining algorithms is appropriate.

According to the results, by using the CART algorithm it was found that sunshine hours have the greatest effect on predicting potential evapotranspiration. After sunshine hours, air temperature, wind speed and relative humidity have the greatest effect on the prediction of potential evapotranspiration, respectively.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith M., 1998. Crop evapotranspiration, Irrigation and Drainage Paper No. 56. Food and Agriculture Organization: Rome, Italy.

Asadi Zarch, M.A., Sivakumar, B., Malekinezhad, H., Sharma, A., 2017. Future aridity under conditions of global climate change. Journal of Hydrology, 554, 451-469.

Crows Corporation, T., 1999. Introduction to data mining and knowledge discovery, third ed., Postmac, MD. Available at: www.twocrows.com, (April 29, 2000)

Fooladvand, H., 2010. Monthly prediction of reference crop evapotranspiration in Fars province. Iranian Journal of Soil and Water Science, 20 (4):157-169.

Giudici, P., 2003. Applied data mining: statistical methods for business and industry. Wiley, London . pp. 364.

Itenfisu, D., Elliott, R.L., Allen R.G., Walter I.A., 2003. Comparison of reference evapotranspiration calculations as part of the ASCE standardization effort. Journal of Irrigation and Drainage Engineering, ASCE 129, 440-448

Martí, P., González-Altozano, P., López-Urrea, R., Mancha, L.A., Shiri, J., 2015. Modeling reference evapotranspiration with calculated targets. Assessment and implications. Agricultural Water Management, 149, pp.81-90.

Mirhashemi, S.H., Panahi, M., 2014a. Evaluation of CART model to estimate potential evapotranspiration and relative humidity (Case Study: Arak station). The first national conference on agriculture and sustainable natural resources. 30 January, 2014. Mehr Arvand Higher Education Institute.

Mirhashemi, S.H., Panahi, M., 2014b. Evaluation of M5P model in predicting average temperature of air and monthly potential evapotranspiration for the next month in Rasht synoptic station. The First National Conference of Water Use Optimization. 5-6 March, 2014. Gorgan University of Agricultural Sciences and Natural Resources.

Pandey, P.K., Dabral, P.P., Pandey, V., 2016. Evaluation of reference evapotranspiration methods for the northeastern region of India. International Soil and Water Conservation Research, 4, 52-63.

Samadianfard, S., Panahi, S., 2018. Estimating daily reference evapotranspiration using data mining methods of support vector regression and m5 tree model. Journal of Watershed Management Research, 9(18), 157-167, (In Persian).

Sameti, M., Ghahraman, N., Ghorbani, Kh., 2013. Application of M5 model in estimating reference crop evapotranspiration in Shiraz and Kermanshah. Journal of Water Researches in Agriculture, 27(3), 289-299. (In Persian).