Development of random forest model as decision support tool in water resources management of Ogun headwater catchments

O. O. Aiyelokun¹ · O. A. Agbede¹

Received: 21 April 2021 / Accepted: 21 June 2021 / Published online: 30 June 2021
© The Author(s) 2021

Abstract

Water resources cannot be effectively managed unless potential evapotranspiration is determined with high accuracy at headwater catchments. The study presents the most suitable feature combinations for building a reliable potential evapotranspiration (PET) model in the headwater catchments of Ogun River Basin, Southwest Nigeria. Using rainfall (R), wind speed (U_2), sunshine hour (S), relative humidity (Rh), minimum temperature (T_{min}) and maximum temperature (T_{max}) as input features, a Random Forest (RF) model was developed to predict PET. Although the model yielded satisfactory results, it was subjected to the minimal depth and percentage increase in mean square error (%IncMSE). This was done to reduce the input features and to increase model accuracy. Thereafter various combinations of important input features were examined in order to establish the best combinations required to yield optimum results. The study revealed that although T_{max} (%IncMSE of 652.09, p value < 0.05) and Rh (%IncMSE of 254.36, p value < 0.05) were the most important predictors of PET, a more reliable RF model was achieved when S and U_2 were combined with them. Consequently, this study presents RF with a combination of four parameters (T_{max} , Rh , S and U_2) as an excellent computational technique for the prediction of PET in headwater catchments.

Keywords Headwater catchments · Random forest · Potential evapotranspiration · Water resources management

Introduction

There are inherent connections among headwaters, landscape systems and downstream waters, and these causal relationships influence water supply, as well as the transport and fate of water and solute in watersheds (Alexander et al. 2007). According to Richardson (2020), headwater catchments constitute the originating streams of all river basins at their initiation stages. They can make up as much as 80% of the total stream length, and they can be very susceptible to pollutants, landscape dynamics and climate anomalies due to their small size and water volume (OHI 2014; Richardson 2020).

Potential evapotranspiration (PET) is an important component of the hydrological system, because it combines changes in many other climate parameters including temperature, radiation, humidity and wind speed (Zou

et al. 2017). It is expedient to quantify evapotranspiration at river basin level for water resources management; this is because it directly influences the hydrologic water balance, which further informs the design of irrigation and drainage canals, reservoir operation, potentials for rain-fed agricultural production and crop water requirements of the basin (Dinpashoh 2006). Therefore, it becomes apparent to develop robust decision support tools for predicting PET in headwater catchments in order to manage water resources effectively.

Potential evapotranspiration represents water loss, water unavailable for exploitation by man and hence it plays an important role in water resource management and hydrological cycle because it is the main source of atmospheric moisture. At global and continental scales, evapotranspiration is the second largest component of the terrestrial water budget after precipitation (Glenn et al. 2007). Evapotranspiration transfers a large volume of water from soil and vegetation into the atmosphere (Anderson et al. 2012). About 60% of the global annual land precipitation is lost to it, while water loss from vegetation constitutes about 80% of terrestrial evapotranspiration (Glenn et al. 2007). It also serves

✉ O. O. Aiyelokun
aiyelokuntobi@gmail.com

¹ Department of Civil Engineering, University of Ibadan, Ibadan, Nigeria

as a form of sink in groundwater models since they remove water from the groundwater system. Therefore, efforts to ensure water resources sustainability cannot succeed without a proper understanding of the amount of water lost to evapotranspiration. Since it is very difficult to quantify, the derivation of numerical models for its predictions is a crucial step in water resources management.

Numerical models are adapted to mimic systems in order to predict their response to changes, exclusive of the quantification of the processes and physical characteristics of those systems. Since potential evapotranspiration is contingent on several interconnected weather variables, such as temperature, atmospheric water content, radiation, and wind speed, it is challenging for linear models to efficiently convey all its associated physical processes (Yassin et al. 2016; Yin et al. 2016). In this sense, the effective method to handle such nonlinear interaction between climate (independent) variables and potential evapotranspiration (dependent/response variable) are the utilizations of artificial intelligence and data-driven models (Wu et al. 2020). One data-driven algorithm that has recently gained popularity in the field of water resources management is the Random Forest (RF).

The RF algorithm was developed by Breiman (2001), and it is classified as an algorithm of learning that include some of the most popular aggregation systems. RF algorithms combine multiple poor learners, these learners being decision trees (Scornet 2017). A series of separate trees are constructed to represent a single data set by adding randomness into the original method of tree construction. RF forecast is then calculated as the average of all the tree forecasts (Scornet 2017). As these random forests seem to be capable of detecting related characteristics even in noisy settings, they are very versatile to handle complex feature spaces, as a result, the algorithm has been found useful in different fields such as ecology, bioinformatics, chemoinformatics and object tracking and recognition to mention a few (Scornet 2017).

RF has many advantages which include its fantastic lenience for outliers and noise, its difficulty in generating an over-fitting phenomenon, the ability to overcome the limitations of the “black-box” concept and its advantages of evaluating important features (Wu et al. 2020). With respect to approximating water resources, RF has been used by various researchers worldwide, for instance, Wu et al. (2020) evaluated the potentiality of RF in modeling potential evapotranspiration of an arid oasis region in China using some selected meteorological variable as input, while Wang et al. (2015) recommended the RF model for evaluating flood hazard risk in Dongjiang River Basin of China, both studies showed that RF was able to achieve satisfactory results. Additionally, Feng et al. (2017) compared RF and generalized regression neural networks (GRNN) models for approximating potential evapotranspiration estimates in

southwest China, and discovered that RF model performed better than GRNN model.

RF model has demonstrated substantial potential in many studies. However, the utilization of the RF model for predicting PET of headwater catchments is relatively new, especially for river basins in Nigeria. Recently, various researchers have endeavored to assess the most important input feature combinations for optimizing the prediction of PET (Wu et al., 2020; Tao et al., 2018; Mattar 2018; Khoshravesh et al., 2015). For instance, temperature and relative humidity were reported to be the most important factors for predicting PET by Huo et al. (2012), while the utilization of the only temperature in predicting PET was discouraged by Wu et al. (2020). In many cases, the important input features are used in future model development, while the less important input features are disregarded. Therefore, researchers fail to ascertain the impact of adding or removing input features on model performance. As a result, the present study endeavors to close this gap in knowledge, by evaluating the most suitable feature combinations for building a reliable potential evapotranspiration (PET) model in the headwater catchments of Ogun River Basin. As well as to investigate the impact of the addition and deletion of input features on model performance. The objectives of the study are 1) to develop RF model for predicting PET in the study area, 2) to evaluate the performance of the developed RF model, 3) to evaluate the importance of climate parameters in the predicting PET of the study area, and 4) to investigate the impact of adapting or disregarding input features (based on their level of importance) on model performance.

Materials and methods

Description of study area

Ogun River Basin is a major river basin in the southwestern region of Nigeria. It is located within latitudes 6° 33' N and 8° 58' N; and longitudes 2° 28' E and 4° 8' E and encompasses a spatial extent of approximately 23,700 square kilometers (Oke et al., 2015). River Ogun takes its source from the Iganran Hills at an altitude of about 503 m east of Saki and flows for about 410 km before discharging into the Lagos Lagoon (Fig. 1a). In terms of physiography the basin area is distinguished as the upper-central basin to the north and lower basin to the south of Abeokuta. It is characterized by a tropical climate, two maxima rainfall pattern (Adediji and Ajibade 2008) and high varying temperature both spatial and temporally. The study area is the northern headwater catchment of Ogun River Basin (Fig. 1b), which is made up of a network of headwater sub-basins of the Ogun, Okifi and Oyan Rivers

Fig. 1 Map of Ogun River Basin area showing the major tributaries (a) and Northern Headwater Catchment (b)

. As shown in the map (Fig. 1b), this area is largely composed of 1st, 2nd and 3rd order streams.

Data acquisition

Meteorological data were obtained from the Nigeria Meteorological Agency (NIMET) for Saki station (Fig. 1b), which is the main meteorological station located in Ogun headwater catchments. The meteorological parameters obtained at these stations included maximum and minimum temperature, relative humidity, sunshine duration and wind speed at two meters, which were required for the computation of potential evapotranspiration. The time series of the obtained meteorological data span over a period of 33 years (1984–2016). No missing data were found in the time series.

FAO penman–monteith (FAO P-M) method

The study adopts the FAO P-M method, which has been recommended as the main standard scheme for the estimating evapotranspiration. This method was chosen because its essential data are available, and because of its robustness. It is relatively complete in theory, considers a wide-range of various climatic factors affecting evapotranspiration, and its accuracy and reliability have been confirmed in several regions around the globe (Xin-e et al. 2015). The FAO P-M method is represented by Eq. 1:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \left[\frac{900}{(T_{\text{mean}} + 273)} \right] U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where ET_0 is reference evapotranspiration (mm), Δ is the slope of the saturated vapor pressure ($\text{kPa}/^\circ\text{C}$), R_n is the net radiation at the crop surface ($\text{MJ}/(\text{m}^2\text{d})$), G is the soil

heat flux density ($\text{MJ}/(\text{m}^2\text{d})$), T mean is the mean monthly air temperature at a height of 2 m ($^\circ\text{C}$), u_2 is wind speed at a height of 2 m (m/s), $e_s - e_a$ is the saturation vapor pressure deficit (kPa), and g is the psychrometric constant ($\text{kPa}/^\circ\text{C}$).

Random forest (RF) method

RF is based on Classification and Regression Trees (CART) model. The nucleus of the RF is centered on the extraction of a repeated random K sample from a unique training sample set N for generating a new set of training samples through the bootstrap resampling method. Thereafter K decision trees are generated and the bootstrap sample collection forms a random forest.

With respect to a classification problem, a classified outcome of novel data depends on the quantity of votes obtained by classification tree votes, and in the case of a regression problem, all the averages of the predictive value of decision trees are regarded as final prediction outcomes (Fig. 2). The description of the variant, parameters, and feature importance capabilities of RF and its functioning procedure has been detailed in the work Boulesteix et al. (2012), Scornet (2017) and Wu et al. (2020).

Since the estimated potential evapotranspiration are continuous values, a supervised regression framework was considered in this study, which entails the development of an optimized model for predicting potential evapotranspiration in the study area. Where the computed monthly potential evapotranspiration (PET) represents the dependent variable while rainfall (R), wind speed (U), sunshine hour (S), relative humidity (Rh), minimum temperature ($T_{\min}$) and maximum temperature ($T_{\max}$) are the predictors as shown in Eq. 2:

$$\text{PET} = f(R, U_2, S, \text{Rh}, T_{\min}, T_{\max}) \quad (2)$$

The model was developed using the "random forest" package written in R statistical software (<http://www.R-project.org> version 3.6). The random forest was calibrated using 70% of the dataset, while the remaining 30% served as the validation set. It is widely speculated that 'ntree', which signifies quantity of trees in the forest, and 'mtry' which

signifies the number of distinct descriptors verified based on respective partition, are important parameters for calibrating RF models (Li et al. 2016; Scornet 2017; Rakhee et al. 2020; Aiyelokun et al. 2021). A ntree of 350, mtry of 5 were found to be adequate for building the model used in this study.

Evaluation of model performance

The performance of the model was assessed using mean absolute relative error (MEA), coefficient of determination (R^2) and Nash–Sutcliffe efficiency coefficient (NSCE). A low MEA value, and R^2 and NSCE value close to 1 are evidence of a good model. The mathematical descriptions of MEA, R^2 , and NSCE are listed as follows:

$$\text{MAE} = \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)}{n} / \tilde{y}_i \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n [(y_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}})]^2}{\left[\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\tilde{y}_i - \bar{\tilde{y}})^2 \right]} \quad (4)$$

$$\text{NSCE} = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where y_i is the estimated PET, $\tilde{y}_i$ is the predicted PET, while $\bar{y}$ and $\bar{\tilde{y}}$ indicate the average observed and predicted PET, respectively. The terms 'training' and 'testing' phases used for the data intelligent model also means calibration and validation of physically based hydrological model (Li et al. 2016). These evaluation criteria were chosen because of their wild application in hydrological model evaluation.

Important feature selection methods of random forest

The random forest has great potential in dealing with data that has outliers and noise, this makes it efficient as an important feature assessment tool (Jaiswal and Samikannu 2017). The important feature section capability of RF puts it in the frontline compared to numerous algorithms used for generating insights during data-driven model development.

Fig. 2 Workflow diagram of a random forest (Adapted from Wu et al., 2020)

The study used two important feature selection methods of RF which include minimal depth (Ehrlinger 2015; Ishwaran et al. 2010) and percentage increase in mean square error (%IncMSE) (Zhang et al. 2018).

Minimum depth distribution

The minimal depth distribution is an important technique for ranking the predictors used for constructing a RF model. The minimum depth method is based on the assumption that the nodes of predictor variables with high influence on the response variable are frequently split close to the root node (tree trunks) (Ehrlinger, 2015). The relative distance of a node to the trunk of the tree determines the numeric attributes which the node takes, while the root node takes zero (Ehrlinger 2015). The minimum depth evaluates the essential risk factors by taking an average of depth of all first split of individual variable against all the trees in the forest, such that the important variables have shorter distance to the root (Ehrlinger 2015). In summary, the minimal depth of a variable is a proxy measure of predictive strength of the variable. The lesser the minimal depth, the more influence the

variable has in cataloging observations, and consequently on the forest accuracy.

Percentage increase in mean square error (%IncMSE)

Percentage increase in Mean Square Error (%IncMSE) is the most efficient and revealing method for identifying important variables from RF. The important features in prediction have the highest percent increase in MSE after the data have been permuted (Zhang et al 2018). The level of significance of the %IncMSE was further assessed based on the permutation of the response variable. This involves the development of a RF model to measure the feature importance, thereafter the response variable is permuted n number of times, while new RF models are being developed for each n step, as the p values of each predictor (feature) is estimated. P value of less than 0.05 is considered to be more likely to significantly affect the model predictive performance.

The summary of the steps used in conducting the study based on the methodology is presented in Fig. 3.

Fig. 3 General methodological framework

Result and discussion

Performance evaluation of RF model

The importance of determining PET with reliable accuracy cannot be overstated in water resources management, especially in water-scarce regions (Nourani et al. 2020). This study endeavors to establish a highly reliable RF model for predicting PET which is an important variable in the management of water resources. The functioning of the developed RF model was evaluated using evaluation criteria of MAE, R^2 and NSCE as presented in Table 1. It could be observed that there was a small average absolute difference between the observed and predicted PET based on the calibration and validation MAE, while a coefficient of determination of 0.98 and 0.92, respectively, for the calibration and validation phase is an indication of a good fit between the modeled and the observed PET. Furthermore, the NSEC of the model is very close to 1 at both calibration and validation phase, indicating that RF model is adequate for predicting PET in the study area.

This result corroborates various other studies that have been conducted to evaluate the performance of RF in predicting PET. For instance, Rakhee (2020) observed that RF forest performed better than logistic regression and discriminate analysis in predicting evapotranspiration. Wu et al. (2020) developed RF models with different input

combinations that were able to achieve satisfactory results at calibration and validation phases. In their study, the R^2 of the models ranged from 0.962 to 0.996 at the calibration phase and 0.909–0.990 at validation phase. Similarly, Feng et al. (2017) established that RF model with complete data input constituting of maximum temperature, minimum temperature, solar radiation, wind speed, relative humidity and extraterrestrial radiation performed better (NSCE of 0.958 ~ 0.990) than RF models with incomplete input data (maximum, minimum air temperature and extraterrestrial radiation) with NSCE ranging from 0.862 to 0.951. Generally, since the R^2 and NSCE at both calibration and validation phases (Table 1) are greater than 0.70, it can be concluded that the results of the RF model is acceptable.

The agreement between the observed and the predicted PET at the calibration and validation phase was examined by combining scatter and time series plots as shown, respectively, in Fig. 4 and Fig. 5. Both combinations of plots show that there is a close agreement between observed and predicted PET; for instance, the predicted PET follows the patterns of the observed PET very closely. It is obvious that the model underestimated extreme low and high PET in some months at both calibration and validation phase. In the study area, PET are usually high in part of the dry seasons between January and March and are usually low in part of the wet season between July and September (Egwuonwu et al. 2012; Ashaolu and Iroye 2018). This implies RF underestimated PET mostly in the wettest and driest months as could be seen in Figs. 4 and 5. A similar case was reported by Feng et al. (2017), who observed that RF models of different input combinations underestimated PET between March and August and overestimated PET between September and December. This is an indication that RF may be sensitive to extreme values of PET and their predictor, in agreement with the study of Wu et al. (2020), which concluded that models may find it difficult to explain minimum and maximum PET perfectly.

Table 1 Summary of performance of random forest models

Evaluation criteria	Calibration	Validation
Mean absolute error (MAE)	3.04	7.17
Coefficient of determination (R^2)	0.98	0.92
Nash–sutcliffe efficiency (NSCE)	0.98	0.91

Fig. 4 Summary of agreement between observed and predicted PET at calibration phase based on scatter and time series plot

Fig. 5 Summary of agreement between observed and predicted PET at validation phase based on scatter and time series plot

Table 2 Statistical evaluation of important climate parameters in predicting PET

Features	%IncMSE	<i>P</i> value
Rainfall (mm)	64.47	1.00
Wind speed (m ² /s)	185.24	1.00
Sshine duration (Hr)	119.51	0.98
Relative humidity (%)	254.36	0.02
Minimum temperature (C°)	60.52	1.00
Maximum temperature (C°)	652.09	0.02

Evaluation of features importance in predicting PET

The evaluation of important climate parameters in predicting the PET of Ogun Headwater catchments was assessed using the minimum depth plot (Fig. 5) and the analysis of the increase in MSE with their level of significance (Table 2). Based on Fig. 6, maximum temperature and relative humidity have the lowest mean minimum depth of 1.04 and 1.42, respectively, with less than 100 trees. Sunshine duration and wind speed have a mean minimum depth of 1.80 and 2.06, respectively, while minimum temperature and rainfall had the highest mean minimum depth of trees of 2.15 and 2.49, respectively. This is further established in Table 2, which shows the climate variables were important in the prediction of PET in the order of $T_{\max} > Rh > U_2 > S > R > T_{\min}$ based on their values of IncMSE%. This implies that $T_{\max}$, Rh , U_2 and S are important variables in predicting PET of the study area. However, since the *p* values of $T_{\max}$ and Rh were lesser than 0.05, then they could be considered to be the most important variables for predicting PET in Ogun headwater catchments.

Similar results have been obtained in previous studies. For instance, Huo et al. (2012) revealed that temperature and relative humidity were the most important factors affecting PET, while Wu et al. (2020) have also confirmed that using temperature only is not sufficient for predicting PET, their study established that in support of previous closely related studies; sunshine hour, wind speed and relative humidity play a significant role in reducing model errors when predicting PET. In contrast to previous studies (Wu et al. 2020; Tao et al. 2018; Mattar 2018; Khoshravesh et al. 2015), the present study discovers that minimum temperature is not important in determining PET. While rainfall was also found to be unimportant in PET prediction, which could have been the reason why rainfall has been ignored by many researchers when constructing input features for building PET models.

Having established that minimum temperature and rainfall were not important features for model development; different combinations of the important features were thereafter used to build RF models. This was done to examine the impact of these combinations on the reliability of predicting PET in the study area. As shown in Table 3, RF model developed using the most important variables ($T_{\max}$ and Rh with *p* value < 0.05) performed less in comparison with the results of the initial combination presented in Table 2. This shows that the utilization of features with high IncMSE% and statistical significance (*p* value < 0.05) may not achieve more accurate model. As a result, statistically non-significant features should not be disregarded without further evaluation when building RF models. Generally, the RF model that combines features with high IncMSE% and low minimum depth (Fig. 6) achieved the overall best performance as shown in Fig. 7.

Table 3 Performance of RF models based on different combinations of importance feature

Input Combinations	Calibration			Validation		
	MAE	R^2	NSCE	MAE	R^2	NSCE
$T_{\max}$, Rh	5.89	0.94	0.94	11.43	0.69	0.69
$T_{\max}$, U_2	3.42	0.98	0.98	7.84	0.89	0.89
$T_{\max}$, Rh, S, U_2	2.52	0.99	0.99	6.29	0.93	0.93
$T_{\max}$, Rh, S	4.30	0.96	0.96	11.88	0.70	0.69
$T_{\max}$, Rh, U_2	3.35	0.98	0.98	8.22	0.91	0.90

Fig. 6 Distribution of minimum depth and mean of maximum temperature ($T_{\max}$), relative humidity (Rh), sunshine duration (S), wind speed (U_2), minimum temperature ($T_{\min}$) and rainfall

Therefore, it could be inferred from the foregoing analysis that model combinations that have wind speed and sunshine hour in combination with maximum temperature as input features, achieved the least generalization error (MAE of 2.52 and 6.29, respectively, at calibration and validation phase) and highest R^2 and NSCE values. This is in agreement with Martí et al. (2015) who revealed that high percentage of variations in PET is determined by temperature and solar radiation, and Feng et al. (2014) who unraveled that sunshine hour is important when developing models to predict PET.

Fig. 7 Comparison of model errors based on different combination of important features

Conclusion

Headwater catchments play important roles in the sustainability of water resource processes of a river basin. In this study, an attempt was made to present the most suitable feature combinations for building a reliable PET model in the headwater catchments of Ogun River Basin, Southwest Nigeria. It was established that RF with a combination of four parameters (maximum temperature, relative humidity, sunshine duration and wind speed) can successfully model PET in the study area with high accuracy. However, in contrast to previous studies where different combinations of model inputs were assessed to ascertain best combinations. The present study evaluated the performance of different models based on the combinations of already established important features, taking cognizance of the effects of adapting or disregarding input features of climate parameters (with respect to their level of importance). This approach was shown to yield satisfactory results for the study area. Based on the results presented in this study, RF is offered as a decision support tool that can be used to accurately predict PET in headwater catchments and data-limited areas. This method can be applied to inform decision-making in water resources management of the study area and other parts of the world.

Funding No specific funding was received by the authors for this work.

Declarations

Conflict of interest There are no conflicts of interest to declare.

Ethical approval The manuscript includes no issue related to ethical compliance.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adediji A, Ajibade LT (2008) The change detection of major dams in Osun State, Nigeria using remote sensing (RS) and GIS techniques. *J Geogr Reg Plan* 1(6):110–115
- Aiyelokun O, Ogunsanwo G, Aiyelokun O et al (2021) Effectiveness of improved bootstrap aggregation (IBA) technique in mapping hydropower to climate variables. *Int J Energy Water Resour*. <https://doi.org/10.1007/s42108-020-00105-1>
- Alexander RB, Boyer EW, Smith RA, Schwarz GE, Moore RB (2007) The role of headwater streams in downstream water quality 1. *JAWRA J Am Water Resour Assoc* 43(1):41–59
- Anderson MC, Allen RG, Morse A, Kustas WP (2012) Use of landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sens Environ* 122:50–65
- Ashaolu E, Iroye K (2018) Rainfall and potential evapotranspiration patterns and their effects on climatic water balance in the Western Lithoral Hydrological Zone of Nigeria. *Ruhuna J Sci* 9(2):92. <https://doi.org/10.4038/rjs.v9i2.45>
- Boulesteix A, Janitza S, Kruppa J, König I (2012) Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip Rev: Data Min Knowl Discov* 2(6):493–507. <https://doi.org/10.1002/widm.1072>
- Breiman L (2001) Random forests. *Mach Learn* 45(5–32):2001
- Dinpashoh Y (2006) Study of reference crop evapotranspiration in I.R. of Iran. *Agric Water Manag* 84:123–129
- Egwuonwu CC, Okafor VC, Ezeanya NC, Nzediegwu C, Okorafor OO (2012) A comparison of the reliability of six evapotranspiration computing models for Abeokuta in South Western Nigeria. *Greener J Phys Sci* 2(4):064–069
- Ehrlinger J (2015). Randomforests: Visually exploring a random forest for regression. *arXiv: 1501.07196 [stat.CO]*
- Feng Y, Cui NB, Gong DZ, Zhang QW, Zhao L (2017) Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modeling. *Agric Water Manag* 193:163–173
- Glenn EP, Huete AR, Nagler PL, Hirschboek K, Brown P (2007) Integrating remote sensing and ground methods to estimate evapotranspiration. *Crit Rev Plant Sci* 26:139–168
- Ishwaran H, Kogalur U, Gorodeski E, Minn A, Lauer M (2010) High-dimensional variable selection for survival data. *J Am Stat Assoc* 105(489): 205–217. Retrieved April 20, 2021, from <http://www.jstor.org/stable/29747021>
- Jaiswal JK, Samikannu R (2017) Application of random forest algorithm on feature subset selection and classification and regression. In: 2017 world congress on computing and communication technologies (WCCCT), Tiruchirappalli, India, (pp. 65–68). <https://doi.org/10.1109/WCCCT.2016>
- Khoshravesh M, Sefidkouhi MAG, Valipour M (2015) Estimation of reference evapotranspiration using multivariate fractional polynomial, bayesian regression, and robust regression models in three arid environments. *Appl Water Sci* 7:1911–1922
- Li J, Monroe W, Ritter A, Galley M, Gao J, Jurafsky D (2016) Deep reinforcement learning for dialogue generation. Proceedings of the 2016 conference on empirical methods in natural language processing, pages 1192–1202, Austin, Texas, November 1–5, 2016. c 2016 Association for Computational Linguistics.
- Martí P, González-Altozano P, López-Urrea R, Mancha LA, Shiri J (2015) Modeling reference evapotranspiration with calculated targets. *Assess Implic Agric Water Manag* 149:81–90
- Mattar MA (2018) Using gene expression programming in monthly reference evapotranspiration modeling: a case study in Egypt. *Agric Water Manag* 198:28–38
- Nourani V, Elkiran G, Abdullahi J (2020) Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2019.124434>
- Oke M, Martins O, Idowu O, Aiyelokun O (2015) comparative analysis of empirical formulae used in groundwater recharge in Ogun–Oshun River Basins. *Ife J Sci* 17(1):53–63
- Ontario headwaters institutes (OHI) (2014). The importance of headwaters to watershed health (p. 3). Ontario: OHI.
- Rakhee R, Singh A, Mittal M, Kumar A (2020) Qualitative analysis of random forests for evaporation prediction in Indian Regions. *Indian J Agric Sci* 90(6):1140–1144
- Richardson JS (2020) Headwater streams. In: Goldstein MI, DellaSala DA (eds) *Encyclopedia of the World's Biomes*. Elsevier, Amsterdam, Netherlands
- Scornet E (2017) Tuning parameters in random forests. *ESAIM: Proc Surv* 60:144–162. <https://doi.org/10.1051/proc/201760144>
- Tao H, Diop L, Bodian A, Djaman K, Ndiaye PM, Yaseen ZM (2018) Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso. *Agric Water Manag* 208:140–151
- Wang ZL, Lai CJ, Chen XH, Yang B, Zhao SW, Bai XY (2015) Flood hazard risk assessment model based on random forest. *J Hydrol* 527:1130–1141
- Wu M, Feng Q, Wen X, Deo R, Yin Z, Yang L, Sheng D (2020) Random forest predictive model development with uncertainty analysis capability for the estimation of evapotranspiration in an arid oasis region. *Hydrol Res* 51(4):648–665. <https://doi.org/10.2166/nh.2020.012>
- Yassin M, Alazba A, Mattar M (2016) Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agric Water Manag* 163:110–124. <https://doi.org/10.1016/j.agwat.2015.09.009>
- Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *J Hydrol* 561:918–929
- Zou M, Niu J, Kang S et al (2017) The contribution of human agricultural activities to increasing evapotranspiration is significantly greater than climate change effect over Heihe agricultural region. *Sci Rep* 7:8805. <https://doi.org/10.1038/s41598-017-08952-5>