DATA-DRIVEN APPROACHES FOR IMPROVED EVAPOTRANSPIRATION MODELLING WITH LIMITED DATA

Barbora Považanová*[,1], Milan Čistý[1]

*barbora.povazanova@stuba.sk
[1]Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 811 07 Bratislava, Slovakia

Abstract
This study uses data-driven methods to estimate FAO Penman-Monteith Reference Evapotranspiration (ETo) using only temperature data. Reference evapotranspiration, as an important variable for estimating actual evapotranspiration, is crucial in various water management tasks. However, some data for the Penman-Monteith equation is often unavailable. Thus, the need to use alternative methods emerges. The research shows DDM's effectiveness particularly when feature engineering was used. The study tested standard equations (Hargreaves Samani) and a proposed CatBOOST model with feature engineering to model ETo. The CatBOOST model achieved a higher R2 of 0.94 than the standard equations' R2 of 0.86. This result underscores DDM’s potential to refine evapotranspiration modelling for wide applications in water resource management, irrigation, and agriculture.

Keywords
Reference evapotranspiration, data-driven methods, minimal input data

1 INTRODUCTION

Evaporation of water plays a key role in the Earth's climate system and is critical to the planet's hydrological cycle. Water loss to the atmosphere is a complex process containing different types of evaporation and the evaporation of water from vegetation called transpiration. This compound process is collectively known as evapotranspiration. The accurate estimation of actual evapotranspiration, particularly in specific crop or landscape scenarios, is crucial for various applications, such as assessing the water balance in watersheds, drought threat analysis, crop water requirement estimation, and irrigation scheduling [1], [2]. Accurate measurement and comprehension of evapotranspiration processes are essential for the effective management of water resources and for addressing the challenges posed by changing climate patterns and agricultural demands.

Reference evapotranspiration (*ETo*) represents a fundamental variable used to estimate actual evapotranspiration. The actual evapotranspiration (*Eta*) can be calculated as the reference evapotranspiration multiplied by the crop coefficient (*Kc*):

$$ETa = Kc \cdot ETo \tag{1}$$

where *ETa* is actual evapotranspiration (mm day^{-1}), *ETo* reference evapotranspiration (mm day^{-1}), *Kc* crop coefficient (-).

The Kc, or crop coefficient, is specific to a particular crop and is typically established through empirical methods. Kc values encapsulate the cumulative influences of factors such as alterations in leaf area, plant height, crop attributes, the crop development stage, planting dates, canopy coverage density and extent, canopy resistance, soil conditions, as well as management practices. Each crop possesses a unique set of *Kc* values, allowing for the prediction of distinct water consumption across various crops and their growth stages [3]. Fig. 1 provides an illustration of a *Kc* curve concerning days or weeks following planting, illustrating the *Kc* values for initial, developmental, mid-season, and end-season phases.

The construction of the crop coefficient curve requires only three key parameters: kc_{ini}, kc_{mid}, and kc_{end}. These values are available in published tables specific to various plant species. Reference evapotranspiration is specific to a given location and time, so it must always be calculated, not looked up in tables; therefore, this work focuses on this quantity. When applying the reference evapotranspiration calculation methods discussed in this study, along with these crop coefficient curve parameters (Eq. 1), it becomes feasible to compute actual

DOI 10.13164/juniorstav.2024.24104

evapotranspiration of a particular crop in locations where the input data necessary for other calculation methods is limited, thus enabling a more efficient estimation process.

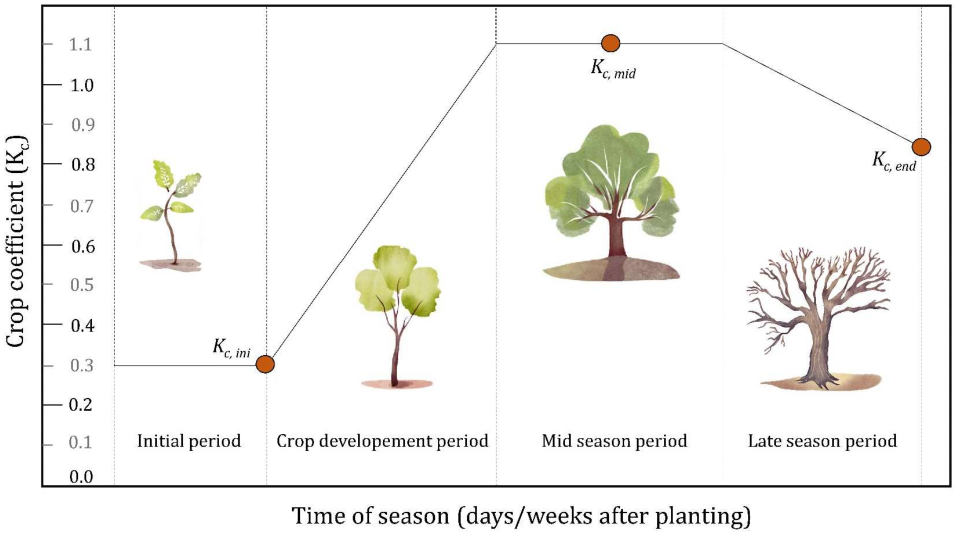

Fig. 1 Changing values of crop coefficient across different plant development stages.

The Food and Agriculture Organization (FAO) offers a widely employed method for *ETo* calculation, utilizing meteorological variables including temperature, humidity, wind speed, and solar radiation [4]. However, the availability of these meteorological data varies across regions, posing challenges to accurate ETo estimation. This issue is further complicated by the spatially inhomogeneous nature of evapotranspiration processes and various other factors. This problem can be effectively addressed through the application of data-driven techniques, also referred to interchangeably as machine learning methods within this study.

Data-driven (DDM) or (interchangeably) machine learning (ML) algorithms, renowned for their capacity to model complex processes with limited physical or mathematical description, have emerged as a promising solution. DDM algorithms are suitable for identifying dependencies and correlations in the provided data. This study employs DDM to address the regression task, aiming to enhance the precision of *ETo* calculations while minimising requirements on historically measured input data.

A growing body of research has explored the use of DDM to approximate ETo using fewer climatic variables as the conventional FAO methodology prescribes. Studies, such as that by Dimitriadou et al. [5], have applied multiple linear regression models to compute *ETo* with fewer climatic data, demonstrating effective input variable combinations. Ferreira and colleagues [6] calculated ETo by juxtaposing the FAO56-PM formula against random forest (RF) (ANN), multivariate adaptive regression splines (MARS), and extreme gradient boosting (XGBoost). Their findings indicated that integrating the soft computing models with the FAO56-PM equation for ETo estimation yielded results comparable to utilizing each model independently. Notably, deep learning and tree-based models have been leveraged for *ETo* estimation and other hydrological parameters in various studies, evaluated, for example, in Goyal et al.'s recent review [7].

Despite the growing body of literature on DDM techniques for *ETo* estimation, several questions remain. Theaim of this paper is to address two of them:

- Comparing various standard models of ETo necessitate minimal measured data with ML algorithms. The accomplished comparative analysis aims to clarify the advantages of DDM or ML methods in relation to established empirical equations.
- The advantages of feature engineering within DDM models, a facet that has received limited systematic exploration, were examined. Feature engineering involves transforming raw data into more sophisticated inputs, enhancing DDM's performance.

In the following sections, a comprehensive description and validation of the proposed method for ETo estimation using DDM is presented. The paper encompasses details about the case study, including the period of data used in the study measurements. Subsequently, the Methodology section elaborates on the approach to DDM techniques. The outcomes of our simulation studies, along with contrasts to existing ETo estimation techniques, are outlined in the Results section and elaborated upon in the Discussion segment.

DOI 10.13164/juniorstav.2024.24104

2 DATA AND STUDY AREA

The data used in this paper originates from the Košice station located in the eastern part of Slovakia, positioned at a longitude of 21.2167 and a latitude of 48.6667, with an elevation of 230 meters above sea level (Fig. 2).

Fig. 2 Study area.

All data were sourced from the ECA&D database [8], where they are freely available. In this study, data from 1951 to 2020 was used. The variable under consideration in the modelling process was the reference evapotranspiration of the Košice station. Data used for the calculation of the FAO reference evapotranspiration included maximal and minimal daily temperature, relative humidity, cloud cover, global radiation, sunshine hours, and wind speed. Clear sky solar radiation was calculated using the R package FAO56 [9], as described in the Methodology section. A dataset comprising a total of 25,260 days with available data was utilized. The dataset was partitioned into training (also known as calibration) and testing sets, with a split ratio of 70:30.

3 METHODOLOGY

In this chapter, the authors provide a concise overview of the chosen empirical equations and data-driven techniques used in the paper. Readers are directed to the cited literature for detailed descriptions if they consider it necessary.

Reference evapotranspiration by FAO Penman-Monteith formula

As a benchmark method in this work, we utilize the FAO Penman-Monteith method, which is widely accepted by both scientific and practical communities. The equation for the FAO method, described in detail in [4], is as follows:

$$ET_o = \frac{o.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot \frac{900}{T + 273} \cdot u_2 \cdot (es - ea)}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)} \tag{2}$$

where *ETo* is reference evapotranspiration (mm day^{-1}), Δ slope vapour pressure curve (kPa °C^{-1}), R_n net radiation at the crop surface (MJ m^{-2} day^{-1}), *G* soil heat flux density (MJ m^{-2} d-1), γ psychrometric constant (kPa °C^{-1}), *T* mean daily temperature in °C, u_2 wind speed at 2m (km h^{-1}), *es* saturation vapour pressure (kPa), and *ea* mean actual vapour pressure (kPa).

Given the article's main goal – achieving precise ETo estimation with less measured data – just the models using fewer climatic quantities than the FAO equation [2] were examined. The selection of these methods was based on the initial accuracy assessments and findings from scientific literature [10].

DOI 10.13164/juniorstav.2024.24104

For this article, four empirical equations were selected: one temperature-based method, the Blaney-Criddle method [11] and three radiation-based methods: the Hargreaves-Samani method [12], the McGuinness-Bordne method [13] and the Makkink equation [14].

Data driven methods (DDM)

The choice of DDM models evolved from more straightforward to more intricate models to determine the need for advanced approaches in this context. Well-known linear regression and its regularized variant, LASSO [15], were employed as simpler methods. Furthermore, the study integrated three diverse ML algorithm categories: random forest [16] as a representation of tree-based methods, support vector regression [17] for kernel techniques, and CatBOOST [18] for boosting methodologies. Ensemble stacking regression [19] was utilized to enhance the depth of the analysis, standing as the most advanced ML technique explored in this research.

Feature engineering

In data-driven modelling (DDM), input data can be represented as straightforward raw variables, e.g., temperature or wind speed. Yet, beyond these basic representations, the input data can be meticulously transformed into "features." These features are intricate derivatives of the original raw data, crafted to furnish a nuanced and refined description of the underlying problem. Such transformations frequently culminate in heightened performance metrics for data-driven models. In the present study, this facet was explored by proposing some strategies pertinent to Feature Engineering (FE), shedding light on its significance in optimizing modelling outcomes.

Firstly, the pivotal raw variable for ETo computation, temperature, transforms to an approximate evapotranspiration value using the Hargraeves-Samani and McGuinness-Bordne equations. The temperature is also harnessed to compute the saturation vapour pressure, employing the Magnus-Tetens formula [20] without requiring additional variables. Additionally, a clear-sky solar radiation (Rso) is incorporated. Notably, Rso is also an integral component in the computation of the original FAO Penman-Monteith method [4], but it is based solely on latitude and elevation, i.e., no additional measured data are required. The authors deem the inclusion of this variable to be a substantial and noteworthy addition to their research, and as a result, they delve into a more comprehensive exploration of the calculation of clear sky solar radiation in the subsequent subsection.

All the supplementary variables pertain exclusively to climatic data derived from temperature measurements and some easily obtainable information. All equations can be found in the referenced literature [4] and are easy to understand; only more complicated computation of the clear sky radiation is described below. A categorical variable for the month was also included in the calculation to account for seasonal variations in estimating evapotranspiration.

Clear sky solar radiation

Clear sky solar radiation refers to the amount of solar radiation that reaches the Earth's surface on a cloudless or perfectly clear day. It is an essential parameter in various fields, such as agriculture, meteorology, and solar energy, as it serves as a reference for understanding the potential solar energy available under ideal, clear sky conditions.

The FAO56 methodology [4] provides a way to estimate clear sky solar radiation, (ET0). As can be seen from Equation 3, the *Rso* calculation includes variables that are available in most places (altitude, latitude). Relative inverse distance and solar declination use only Julian as an input variables and sunset hour angle uses latitude. The advantage of this variable is therefore its complete independence from the measured climatic variables. Another advantage is its adaptability for different areas, as it takes into account latitude and altitude. For this reason, clear-sky solar radiation as an input feature was used in machine learning models.

The first step in calculation of *Rso* is to first calculate extraterrestrial radiation (Ra) (3) and then the clear-sky solar radiation (*Rso*) (4)

$$Ra = \frac{12 \cdot 60}{\pi} \cdot Gsc \cdot dr \cdot (\cos(\varphi) \cdot \cos(\delta) \cdot \sin(ws) + ws \sin(\varphi) - \sin(\delta) \tag{3}$$

$$Rso = (0{,}75 + 2 \cdot (10^{-5}) \cdot z)\; Ra \tag{4}$$

Where *Rso* represents clear sky solar radiation (in $MJ/m^2 \cdot day^{-1}$), z denotes altitude above sea level (in meters), Ra stands for extraterrestrial radiation on a horizontal surface (in $MJ/m^2 \cdot day^{-1}$), Gsc is the solar constant (0.0820 $MJ/m^2 \cdot min^{-1}$), *dr* signifies relative inverse distance Earth-Sun (in radians), φ represents latitude (in radians), δ represents solar declination (in radians), *ws* stands for sunset hour angle (in radians).

4 RESULTS

The authors juxtaposed daily reference evapotranspiration values derived from the FAO Penman-Monteith method, which serves as the benchmark approach, against those computed through:

- Empirical equations including Blaney-Criddle, Hargreaves-Samani, Makkink, and McGuinness-Bordne.
- Data-driven machine learning techniques encompassing multiple linear regression (MLR), LASSO for regularized form of linear regression, ranger aka random forest model (RF), CatBOOST, support vector regression (SVR), and stacking model (autoML). These methodologies were implemented using diverse input combinations for the sake of comprehensive evaluation.

The comparison of results was evaluated based on three statistical characteristics: Percent Bias (PBIAS), Coefficient of Determination (R-squared or R2), and Root Mean Square Error (RMSE). These are common goodness-of-fit tests used in various fields, such as statistics, hydrology, and environmental modelling.

Estimation of ETo using FAO PM and empirical equations

The study conducted a comparison between ETo values determined using the widely accepted FAO Penman-Monteith method, regarded as a standard approach, and evapotranspiration computations obtained according to various authors. These sources include models like Blaney-Criddle, Hargreaves-Samani, Makkink, and McGuinness-Bordne equations, as well as statistical and DDM models such as multiple linear regression (MLR), LASSO, support vector machine (SVR), CatBOOST, ranger, which is effective form of random forest (RF), and stacking models. Different combinations of input climate variables were employed with these techniques.

Regarding the computation of *ETo* by the FAO methodology and the mentioned equations, the FAO Penman-Monteith equation was considered the reference method for calculating *ETo* in this research. It was utilized to determine reference evapotranspiration for the Košice station on a daily basis from January 3, 1951, to April 30, 2020. These calculations were performed using the R programming language and the FAO56 and Evapotranspiration packages [20], [21]. The study's outcomes were then compared to those obtained from the FAO Penman-Monteith method, and the results are presented in Tab. 1. In addition to the goodness of fit statistics, the accuracy of the equations was also evaluated based on the number of negative values (negat. count) that are considered incorrect in the case of evapotranspiration. The optimal outcome was attained through the application of the Hargreaves-Samani equation.

Tab. 1 Statistical evaluation of *ETo* computed by empirical equations (*Tmin* – minimal daily temperature, *Tmax* – maximal daily temperature, *n* – sunshine hours).

Equation	Variables	Negat. count	RMSE	PBIAS %	R2
Hargreaves-Samani	*Tmax, Tmin*	4	0.73	3.4	0.86
Blaney-Criddle	*Tmax, Tmin*	4	1.3	-29.8	0.84
Makkink	*Tmax, Tmin, n*	12	1.04	51	0.96
McGuinness-Bordne	*Tmax, Tmin*	1311	0.86	-6	0.86

Estimation of ETo using data-driven methods

The use of features, as opposed to originally measured data, represents a critical issue in enhancing the performance of machine learning models, yet it is frequently overlooked. This study assessed the impact of different feature utilization strategies on DDM's models. Tab. 2 presents various DDM models that use only minimal and maximal daily temperatures, along with one feature drawn only from temperature. These derived data, referred to as features, encompass the transformation of temperature to rough evapotranspiration using the McGuiness-Bordne (*MGB*) and Hargreaves-Samani (*HS*) equations, as well as clear-sky solar radiation (*Rso*), and saturation vapor pressure (*es*). Furthermore, this study introduced a derived feature – a categorical variable identifying months – that is easily obtainable. Models displayed in Tab. 2 use temperature again as the only measured climate variable, identically to the empirical models presented in Tab. 1.

Tab. 2 Statistical evaluation of *ETo* by data-driven models (*Tmax* – maximal daily temperature, *Tmin* – minimal daily temperature, *HS* – evapotranspiration by the Hargreaves-Samani equation, *MGB* – evapotranspiration by McGuiness-Bordne equation, *es* – saturation vapour pressure, *Rso* – clear-sky solar radiation).

Model no.	Inputs	model	RMSE	PBIAS %	R2
1	*Tmax, Tmin*	MLR	0.963	0.500	0.750
		LASSO	0.935	-0.400	0.760
		SVR	0.915	-2.100	0.770
		RF	0.915	-0.500	0.770
		CatBOOST	0.910	-0.600	0.770
		autoML	0.910	0.000	0.770
2	*Tmax, Tmin, month*	MLR	0.606	0.300	0.900
		LASSO	0.522	-0.100	0.930
		SVR	0.518	-1.300	0.930
		RF	1.557	-0.400	0.790
		CatBOOST	0.522	-0.100	0.930
		autoML	0.520	-0.700	0.930
3	*Tmax, Tmin, MGB*	MLR	0.604	-0.100	0.900
		LASSO	0.526	-0.200	0.930
		SVR	0.478	-0.600	0.940
		RF	0.484	-0.200	0.940
		CatBOOST	0.484	-0.100	0.940
		autoML	0.481	1.200	0.940
4	*Tmax, Tmin, HS*	MLR	0.704	-0.500	0.870
		LASSO	0.636	-0.500	0.890
		SVR	0.471	-0.600	0.940
		RF	0.492	-0.100	0.940
		CatBOOST	0.485	-0.100	0.940
		autoML	0.478	-0.100	0.940
5	*Tmax, Tmin, es*	MLR	0.942	-0.400	0.760
		LASSO	0.922	-0.700	0.770
		SVR	0.920	-2.100	0.780
		RF	0.919	-0.600	0.770
		CatBOOST	0.915	-0.600	0.780
		autoML	0.915	-1.200	0.780
6	*Tmax, Tmin, Rso*	MLR	0.612	0.700	0.900
		LASSO	0.473	-0.200	0.940
		SVR	0.471	-0.600	0.940
		RF	0.475	-0.100	0.940
		CatBOOST	0.469	-0.200	0.940
		autoML	0.467	0.300	0.940

As indicated in Tab. 2, which presents the primary findings of this study, the inclusion of features (FE) significantly enhances the model's performance when compared to the results in Tab. 1. In other words, models that use FE demonstrate a notable improvement in computing *ETo* compared to standard equations or machine learning (ML) techniques that do not incorporate FE. The most substantial enhancement is observed, especially when clear sky solar radiation (*Rso*) is added (model No. 6). This trend aligns with the results obtained from empirical methods, where the Hargreaves-Samani equation, which also employs a comparable supplemental variable, extraterrestrial radiation Ra, yields excellent results.

Furthermore, the introduction of a simple categorical variable, "month" (model No. 2), substantially enhances results compared to model No. 1, which relies solely on raw temperature data. Additionally, favourable results are achieved by including approximate evapotranspiration values computed by the empirical equations *MGB* (model

No. 3) and *HS* (model No. 4). The least significant gain in precision of the model is observed by the addition of saturation vapor pressure (*es*).

In summary, incorporating features consistently yields better results compared to using raw climate temperature data. The evaluation of different DDM models by statistical indicators also implies that autoML and CatBOOST models consistently deliver superior results.

5 DISCUSSION

The results of this study demonstrate the potential benefits of data-driven approaches in evapotranspiration modelling, particularly in situations where the use of the established FAO Penman-Monteith formula is limited by data availability or quality. Among the empirical equations, the Hargreaves-Samani model offers the most satisfactory results, which can be ascribed to the fact that extraterrestrial radiation is used in its calculation.

The data-driven methods outperformed the traditional empirical equations [16], [17] in estimating reference evapotranspiration, highlighting the importance of alternative methods development when input data is absent [7]. The study also showed that incorporating additional features, such as categorical variables, can significantly enhance the performance of ML models.

The introduction of a simple categorical feature, "month," substantially improved the results compared to other models that relied solely on raw temperature data. This finding is consistent with previous studies that have shown the importance of incorporating additional features to improve evapotranspiration modelling [15]. Furthermore, the study demonstrates the importance of thoughtful feature selection in improving model accuracy. The least significant improvement was observed when adding saturation vapor pressure (*es*) as a feature. The best results were obtained after introducing the variable *Rso*, clear sky solar radiation, calculated from altitude, latitude, and Julian day.

This suggests that not all features are equally useful in improving model performance, and that features should be selected carefully, respecting the calculated variable. Accurate estimates of reference evapotranspiration are crucial for various water management tasks, such as irrigation design or crop water requirements. The improved accuracy of data-driven models has the potential to enhance the efficiency and sustainability of water resource management and agriculture. However, it is important to note that data-driven approaches are not without limitations. One of the main drawbacks of data-driven modelling is the risk of overfitting, where the model is too closely fitted to the training data and performs poorly on new data. To mitigate this risk, it is important to carefully evaluate model performance on independent validation data and to use appropriate regularization techniques.

In summary, the results of this study provide evidence that data-driven approaches, specifically machine learning and data-driven methods, can improve the accuracy of reference evapotranspiration modelling. Incorporating additional features, such as categorical variables, can significantly enhance model performance. However, feature selection and engineering should be done thoughtfully, and appropriate regularization techniques should be used to mitigate the risk of overfitting. These findings have important implications for water resource management, irrigation, and agriculture, where accurate estimates of reference evapotranspiration are crucial for decision-making.

6 CONCLUSIONS

This research explores the application of data-driven methods, to enhance the computation of reference evapotranspiration (*ETo*) using the FAO Penman-Monteith method. The key innovation lies in the deliberate emphasis on temperature data, a readily accessible climate variable., This emphasis minimizes data input demands while strengthening it with engineered variables or features. To evaluate the effectiveness of this approach, the study aimed to surpass conventional methods like Hargreaves-Samani or Blaney-Criddle equations. Multiple data-driven algorithms were employed for comparison. The results revealed CatBOOST and autoML as the most promising models, particularly when integrated with feature engineering in the ML modelling process. A noteworthy finding of this research was the importance of clear-sky solar radiation (*Rso*) as a valuable variable. *Rso* could be easily calculated using only latitude, altitude, and Julian day, making it adaptable across diverse geographical regions. This study underscores the significant impact of feature engineering on the performance of ML models, an aspect that had not been comprehensively explored in previous studies on this topic. While the models proposed by the authors demonstrated promise using data from a single climatic station, future research should verify their applicability across a wider range of climatic conditions. Furthermore, calibration of the models with data from more than one climatic station may lead to a model with broader applicability. In conclusion, presented study undescores the importance of incorporating feature engineering into DDM models to achieve precise ETo estimation. The prospect of improved predictions, decreased data input demands, and enhanced efficiency offer considerable promise for optimizing various water management tasks, irrigation planning, and other decision-making in the agricultural sector.

Acknowledgements

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and by the Slovak Academy of Sciences, Grant No. 1/0067/23 and also by the Slovak Research and Development Agency under Contract No. APVV-19-0383 "Natural and technical measures oriented to water retention in sub-mountain watersheds of Slovakia".

References

[1] MEHTA, R. and PANDEY, V. Reference evapotranspiration (ETo) and crop water requirement (ETc) of wheat and maize in Gujarat. *Journal of Agrometeorology* [online]. 2015, 17(1), pp. 107-113. ISSN 0972-1665. Available at: https://doi.org/10.54386/jam.v17i1.984

[2] SILVA-JÚNIOR, R. O., SOUZA-FILHO, P. W. M., SALOMÃO, G. N., TAVARES, A. L., SANTOS, J. F. et al. Response of Water Balance Components to *Changes in Soil Use and Vegetation Cover Over Three Decades in the Eastern Amazon. Frontiers* in Water [online]. 2021, 3. ISSN 2624-9375. Available at: https://doi.org/10.3389/frwa.2021.749507

[3] JORGENSEN, S. E. and FATH, B. D. Encyclopedia of ecology. Elsevier BV, 2008. ISBN 0444520333.

[4] ALLEN, R. G, PEREIRA, L. S, RAES, D. and SMITH, M. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome. 1998, vol. 300, no. 9: A05109

[5] DIMITRIADOU, S. and NIKOLAKOPOULOS, K. G. Multiple Linear Regression Models with Limited Data for the Prediction of Reference Evapotranspiration of the Peloponnese, Greece. *Hydrology* [online]. 2022, 9(7). ISSN 2306-5338. Available at: https://doi.org/10.3390/hydrology9070124

[6] FERREIRA, L. B., DA CUNHA, F. F. and FERNANDES FILHO, E. I. Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil. *Agricultural Water Management* [online]. 2022, 259: A107281. ISSN 03783774. Available at: https://doi.org/10.1016/j.agwat.2021.107281

[7] GOYAL, P., KUMAR, S. and SHARDA, R. A review of the Artificial Intelligence (AI) based techniques for estimating reference evapotranspiration: Current trends and future perspectives. *Computers and Electronics in Agriculture* [online]. 2023, 209: A107836. ISSN 01681699. Available at: https://doi.org/10.1016/j.compag.2023.107836

[8] KLEIN TANK, A. M. G., WIJNGAARD, J. B., KÖNNEN, G. P., BÖHM, R., DEMARÉE, G. et al. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* [online]. 2002, 22(12), pp. 1441–1453. ISSN 08998418. Available at: https://doi.org/10.1002/joc.773

[9] AHANI, A. and MOUSAVI NADOUSHANI, S. S. FAO56: Evapotranspiration Based on FAO Penman-Monteith Equation: R package version 0.1.0. [online]. 2021. [Accessed 21.10.2023] Available at: https://CRAN.R-project.org/package=FAO56

[10] YEH, H. F. Comparison of Evapotranspiration Methods Under Limited Data. Current Perspective to Predict Actual Evapotranspiration [online]. InTech, 2017. ISBN 978-953-51-3173-1. Available at: https://doi.org/10.5772/intechopen.68495

[11] BLANEY, H. F. and CRIDDLE, W. D. Determining water requirements in irrigated areas from climatological and irrigation data. United States Department of Agriculture, Soil Conservation Service, Washington 25, D. C. 1950, pp. 1–48

[12] HARGREAVES, G. H. and SAMANI, Z. A. Reference Crop Evapotranspiration from Temperature. Applied Engineering in Agriculture [online]. 1985, 1(2), pp. 96–99. ISSN 1943-7838. Available at: https://doi.org/10.13031/2013.26773

[13] MCGUINNESS, J. L. and BORDNE, E. F. A comparison of lysimeter-derived potential evapotranspiration with computed values. US Department of Agriculture, 1972

[14] MAKKINK, G.F. Testing the Penman formula by means of lysimeters. Journal of the Institution of Water Engineers. 1957, vol. 11, pp. 277–288

[15] FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R., NARASIMHAN, B., TAY, K. et al. Glmnet: Lasso and elastic-net regularized generalized linear models: R package version. 2009, 1(4), pp. 1–24

[16] BREIMAN, L. Random forests. Machine learning. 2001, vol. 45, pp. 5–32. ISSN 0885-6125

[17] AWAD, M. and KHANNA, R. Support Vector Regression. Efficient Learning Machines [online]. Berkeley, CA: Apress, 2015, pp. 67-80. ISBN 978-1-4302-5989-3. Available at:

https://doi.org/10.1007/978-1-4302-5990-9_4
[18] DOROGUSH, A. V., ERSHOV, V. and GULIN, A. CatBoost: gradient boosting with categorical features support. ArXiv preprint arXiv [online]. 2018. Available at: https://doi.org/https://doi.org/10.48550/arXiv.1810.11363.
[19] H2O.AI. H2O Documentation: H2O for R Users. [online]. 2022. Available at: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html#h2o-package-for-r. [cit. 2023-06-07].
[20] TETENS, O. Uber einige meteorologische begriffe [About some meteorological terms]. Z. Geophys. 1930, vol. 6, pp. 297–309.
[21] GUO, D., WESTRA, S. and MAIER, H. R. An R package for modelling actual, potential and reference evapotranspiration. Environmental Modelling & Software [online]. 2016, 78, pp. 216-224. ISSN 13648152. Available at: https://doi.org/10.1016/j.envsoft.2015.12.019.