



Evaluation of Random Forest model to calculate potential Evapotranspiration using limited meteorological data (study area: Ardabil Plain)

Javanshir Azizi Mobaser [✉] | Ali Rasoulzadeh | Amin Akbari Majd

1. Corresponding Author, Department of Water Engineering, Faculty of Agriculture and Natural Resources, Water Management Research Center, University of Mohaghegh Ardabili, Ardabil 56199-11367, Iran. E-mail: ja_mobaser@uma.ac.ir
2. Department of Water Engineering, Faculty of Agriculture and Natural Resources, Water Management Research Center, University of Mohaghegh Ardabili, Ardabil 56199-11367, Iran. E-mail: rasoulzadeh@uma.ac.ir
3. Phd Student of Water Resources, Department of Water Science and Engineering, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: amin.a.m.ce@gmail.com

Article Info	ABSTRACT
<p>Article type: Research Article</p> <p>Article history:</p> <p>Received: Oct. 29, 2024</p> <p>Revised: Dec. 5, 2024</p> <p>Accepted: Dec. 25, 2024</p> <p>Published online: April. 2025</p> <p>Keywords: Machine Learning, Multiple Linear Regression, Random Forest, Reference Evapotranspiration.</p>	<p>As the global demand for water resources increases, the reduction in water loss, including Evapotranspiration, becomes more obvious. Although many models have been developed to predict evapotranspiration, no universally accepted model for all climate regions has been established. Several soft computational models have been created to circumvent the constraints of empirical models and accurately predict ET. Soft computing models typically require less data and are applicable across various climatic zones. This study aimed to analyze how well two Random Forest models and Multiple Linear Regression could predict ETo in the Ardabil plain region. Meteorological data from the Iranian Meteorological Organization were used to calculate the reference evapotranspiration from 2014 to 2016. In constructing the model, data from 4 meteorological stations were combined to generate a random time series, while the fifth station was reserved for evaluating the models. The assessment metrics used comprised RMSE, R2, and NSE. The RF model achieved higher accuracy with R2, NSE, and RMSE values of 0.74, 0.743, and 8.20 mm, respectively, compared to the MLR model. The results demonstrate that random forest models are reliable tools for forecasting ETo with minimal climate data. In general, using the results of this study and other similar research, we conclude that RF and MLR models simulate potential evapotranspiration with acceptable accuracy but are sensitive to the number of input parameters.</p>
<p>Cite this article: Azizi Mobaser, J., Rasoulzadeh, A., & Akbari Majd, A. (2025). Evaluation of random forest model to calculate potential evapotranspiration using limited meteorological data (study area: Ardabil Plain), <i>Iranian Journal of Soil and Water Research</i>, 56 (2), 545-569. https://doi.org/10.22059/ijswr.2024.384545.669825</p> <p>© The Author(s). Publisher: The University of Tehran Press.</p> <p>DOI: https://doi.org/10.22059/ijswr.2024.384545.669825</p>	





EXTENDED ABSTRACT

Introduction

Water is necessary for environmental protection and meeting various human needs. Water is used for agricultural, domestic, energy generation, recreational, and industrial purposes (directly) and irrigation of green areas (indirectly). Furthermore, proper ET assessment is critical for controlling agricultural water requirements, monitoring and efficiently using water supplies, and forecasting droughts or similar situations. Despite using multiple empirical models for ET prediction, no consensus exists on the suitability of applying all the proposed models across diverse climate zones. Various soft computing models have been developed to overcome the restrictions associated with empirical models and appropriately estimate ET.

Method and Materials

The measurements were conducted in the Ardabil Plain region and its surrounding basin. In this research, model inputs, which included meteorological data, were sourced from the Iranian Meteorological Organization's website between 2006 and 2023. The output and target parameters, which included reference evapotranspiration data, were computed every 10 days (first, second, and third decade of each month) using the FAO Penman-Monteith method (FAO56). This research uses two models, multiple linear regression and random forest, to estimate reference evapotranspiration in the Ardabil plain. In this research, the primary approach was to assess the precision of the random forest model in estimating reference evapotranspiration using observational data, which was initially compared in the correlation analysis of 14 meteorological parameters with reference evapotranspiration data. By examining the correlations between the primary parameters, the parameters that displayed a very low correlation with the reference evapotranspiration were removed from the input data list. Finally, the models and the remaining meteorological data with a mixed-random time series were used to compute the reference evapotranspiration. Evaluation statistics such as the Nash-Sutcliffe efficiency coefficient (NSE), R^2 , and mean square error were used in this study.

Results

From 2006 to 2023, meteorological data were first calculated at five meteorological stations: Nair, Sarein, Ardabil, Ardabil Airport, and Nemin. Using data from four meteorological stations—Nair, Sarein, Ardabil, and Ardabil Airport—a model was developed to understand and establish a relationship between input parameters and reference evaporation and transpiration during the statistical period. Finally, the models were evaluated using data from the Namin station, which had the shortest statistical period and the least complete data among the stations. The final input characteristics were the minimum, maximum, and average temperatures, evaporation, maximum wind direction, daily daylight hours, and soil temperature. A random mix of data from four stations that created a time series with a period of 1600 units was used to train or construct the models in the first stage. The obtained Namin station was then used to evaluate the final model.

The RF model was more accurate in calculating the reference evapotranspiration from the same data, according to the obtained R^2 , NSE, and RMSE values. The present study also demonstrated that, when considering accuracy and stability, hybrid learning models, such as random forest and random tree models, can be reliable machine learning models for predicting ET using limited climate datasets for different climatic zones of the Ardabil Plain.

Conclusions

Although using all seven parameters produced the best results, the results showed that the parameters of soil temperature, maximum temperature, average temperature, minimum temperature, wind speed direction, hours of radiation, and evaporation made up the best-suggested combination. Temperature parameters are more important when estimating reference evapotranspiration using the Penman-Monteith technique, according to the constant coefficients obtained from the MLR model at the model's input. Additionally, the RF model outperformed the MLR model regarding prediction accuracy among the models included in the study.

Author Contributions

A.A.M.: Writing – original draft, Formal analysis, Conceptualization, Data curation, Methodology, Validation, Writing – review & editing. J.A.M.: Writing – review & editing. A.R.: Writing – review & editing.

Data Availability Statement

Not applicable

Acknowledgments

The authors would like to thank all participants of the present study.

Ethical considerations

The Ethics Committee of the Mohaghegh Ardabili University approved the study. The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

ارزیابی مدل جنگل تصادفی برای محاسبه تبخیر و تعرق مرجع با استفاده از داده‌های هواشناسی محدود در منطقه مطالعاتی دشت اردبیل

جوانشیر عزیزی مباصر^۱ | علی رسولزاده^۲ | امین اکبری مجد^۳۱. نویسنده مسئول، گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران. رایانامه: ja_mobaser@uma.ac.ir۲. گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران. رایانامه: rasoulzadeh@uma.ac.ir۳. گروه علوم و مهندسی آب دانشکده محقق اردبیلی، اردبیل، ایران. رایانامه: amin.a.m.ce@gmail.com

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	به دلیل افزایش فشار جهانی بر در دسترس بودن منابع آب، توجه به تلفات آب بیش‌تر از قبل مشخص می‌گردد، تبخیر و تعرق (ET) به عنوان بخش مهمی از تلفات بوده و تخمین آن برای بررسی تغییرات آب و هوا، جلوگیری از آبیاری ناکارآمد و استفاده مناسب از منابع آب، حیاتی است. علی‌رغم مدل‌های تجربی فراوان برای پیش‌بینی ET، هنوز هیچ اجماع جهانی در مورد استفاده از یک مدل تجربی مشخص وجود ندارد. مدل‌های محاسبات نرم به دلیل نیاز به داده‌های کم‌تر، برای جلوگیری از محدودیت مدل‌های تجربی و برای برآورد دقیق‌تر ET توسعه داده شده‌اند. در این تحقیق برای تخمین تبخیر و تعرق مرجع با داده‌های هواشناسی در حدفصل سال‌های ۱۳۸۵ تا ۱۴۰۱ و در دشت اردبیل کارایی دو مدل جنگل تصادفی (RF) و رگرسیون خطی چندگانه (MLR) ارزیابی شد. برای ساخت مدل، از ترکیب داده‌های ۴ ایستگاه هواشناسی استفاده و از ایستگاه پنجم برای ارزیابی نهایی مدل‌ها استفاده شد. آماره های ارزیابی شامل R^2 ، NSE و RMSE بود. نتایج به‌دست آمده برای مدل RF به ترتیب برابر بود با ۰/۷۴۳، ۰/۷۴ و ۸/۲۰ میلی‌متر که در مقایسه با نتایج مدل MLR از دقت بالاتری برخوردار بود. مطالعه حاضر نشان داد که مدل‌های جنگل تصادفی می‌تواند یک مدل مطمئن با در نظر گرفتن دقت و ثبات، برای پیش‌بینی ET و با استفاده از مجموعه داده‌های محدود باشد. به طور کلی، با استفاده از نتایج این تحقیق می‌توان گفت که مدل‌های RF، تبخیر و تعرق مرجع را در مناطقی با داده‌های محدود با دقت قابل قبولی شبیه‌سازی می‌کند.
واژه‌های کلیدی: جنگل تصادفی، رگرسیون خطی چندگانه، تبخیر و تعرق مرجع، یادگیری ماشین.	
تاریخ دریافت: ۱۴۰۲/۸/۸	
تاریخ بازنگری: ۱۴۰۳/۹/۱۵	
تاریخ پذیرش: ۱۴۰۳/۱۰/۵	
تاریخ انتشار: اردیبهشت ۱۴۰۴	

استناد: عزیزی مباصر، جوانشیر؛ رسولزاده، علی؛ و اکبری مجد، امین (۱۴۰۴). ارزیابی مدل جنگل تصادفی برای محاسبه تبخیر و تعرق مرجع با استفاده از داده‌های هواشناسی محدود در منطقه مطالعاتی دشت اردبیل، *مجله تحقیقات آب و خاک ایران*، ۵۶ (۲)، ۵۴۵-۵۶۹.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2024.384545.669825>

© نویسندگان.



مقدمه

آب یکی از حیاتی‌ترین منابع برای حفظ محیط‌زیست و رفع بسیاری از نیازهای انسان است. انسان‌ها آب را از طریق برداشت آب برای مصارف کشاورزی، خانگی، تولید انرژی، تفریحی و صنعتی (مستقیم) و همچنین از طریق آبیاری فضای سبز (غیرمستقیم) استفاده می‌کنند (Kuriqi et al., 2021). مقدار آب مصرفی برای مقاصد کشاورزی، بیش‌ترین سهم آب مصرفی توسط انسان و نزدیک به ۷۰ درصد از مصرف منابع آبی را به خود اختصاص می‌دهد (Rost et al., 2008). انتظار می‌رود که رشد تصاعدی جمعیت در دهه‌های آینده و تغییر در رژیم غذایی، تقاضای جهانی غذا را افزایش داده و در نتیجه فشار بیش‌تری بر تقاضای آب در مصارف کشاورزی وارد کند (Adnan et al., 2020). علاوه بر این، حجم قابل توجهی از منابع آبی به دلیل دما و سرعت بالای باد از منابع آبی تلف می‌شود. تبخیر و تعرق (ET) شامل تلفات آب از سطح زمین به اتمسفر است که توسط فرآیندهای ترکیبی تبخیر از منابع آب‌های سطحی، خاک لخت و سطوح گیاهی و غیره شکل می‌گیرد (فخار و کاویانی، ۱۴۰۰; Li et al., 2009; Zheng et al., 2022). تغییرات آب و هوایی بر چرخه آب تاثیر می‌گذارد و باعث می‌گردد تا میزان تاثیر پارامترهای مختلف و موثر در تعادل آب، از منطقه‌ای به منطقه‌ای دیگر تغییر کند. بنابراین، ET یک فرآیند مهم از چرخه هیدرولوژیکی است که تعادل آب و انرژی سطح زمین را به هم مرتبط می‌کند. به دلیل افزایش فشار جهانی بر در دسترس بودن منابع آب به سبب مصارف مختلف و تغییرات اقلیمی، اهمیت توجه به کاهش تلفات آب با استفاده از سامانه‌های آبیاری کارآمد بیش‌تر مشخص می‌گردد (DehghaniSanij et al., 2004). فرآیند تبخیر و تعرق یک فرآیند ناملموس بوده و اندازه‌گیری آن دشوار است که باید با اندازه‌گیری مستقیم یا تخمین با استفاده از مدل‌های ریاضی تعیین شود (Alizamir et al., 2020; El-Kenawy et al., 2022). برآورد دقیق ET برای مطالعه تغییرات آب و هوا و اثرات زیست‌محیطی آن، جلوگیری از آبیاری ناکارآمد و استفاده مناسب از منابع آب در مصارف کشاورزی ضروری و حیاتی است. این موضوع در مناطقی از جمله محدوده دشت اردبیل که با محدودیت‌های منابع آب درگیر است، اهمیت بیشتری خواهد داشت (جعفرزاده و جعفرزاده، ۱۳۹۴; Kord & Asghari-moghaddam, 2014).

علاوه بر این، برآورد دقیق ET ارزش کاربردی زیادی در مدیریت نیازهای آب کشاورزی، نظارت و استفاده موثر از منابع آب و پیش‌بینی خشکسالی و یا غیره را دارد (عزیزی و فرجی، ۱۴۰۳; Zhao et al., 2013; Malik et al., 2022). تبخیر و تعرق مرجع برای محاسبه شاخص خشکسالی (شاخص تبخیر و تعرق بارش استاندارد) استفاده می‌شود که در ارزیابی خشکسالی ضروری است. با وجود اهمیت نیاز به برآورد دقیق ET، به دلیل نیاز به هزینه و منابع بالا برای اندازه‌گیری‌های مستقیم، برآورد آن هنوز یک فرآیند بسیار چالش برانگیز و پیچیده است. مهندسان و دانشمندان تلاش زیادی برای توسعه مدل‌هایی برای تخمین ET با دقت و هزینه مناسب برای مکان‌ها و مناطق آب و هوایی مختلف انجام داده‌اند. این مدل‌ها از طریق داده‌های ورودی، روابط عملکردی، مقیاس مکانی، مقیاس زمانی، درجه پیچیدگی و قابلیت کاربرد، مانند برنامه‌ریزی فعالیت‌های کشاورزی و تجزیه و تحلیل بیلان آب متمایز می‌شوند (عزیزی، ۱۳۹۲; Chia, 2024; Huang, & Koo, 2022; Monteiro et al., 2021; Mostafa et al., 2023; Veysi et al., 2024).

در میان اولین مدل‌های تجربی تعریف شده برای ET، مدل پنمن از معروف‌ترین آن‌ها است که مانیتث آن را بهبود بخشید، که بعد از آن زمان، این مدل، مدل پنمن-مانیتث^۲ نامیده می‌شود و متداول‌ترین مدل در سراسر جهان است (محمدی و همکاران، ۱۳۹۹; سلطانی و همکاران، ۱۳۹۸; Alizamir et al., 2020). پس از پنمن مانیتث، بسیاری از مدل‌های مبتنی بر فیزیک مساله، مانند Blaney-Criddle, Hargreaves, Linacre, Kharrufa, Ravazzani و مدل‌های تجربی مبتنی بر انتقال جرم مانند Dalton, Trabert, Brockamp, مورد استفاده واقع شدند. به طور کلی، مدل‌های مبتنی بر دما بیش‌ترین استفاده را نسبت به سایر مدل‌ها دارند زیرا به داده‌های ورودی کم‌تری نیاز دارند. با این وجود، دقت به دست آمده توسط آن مدل‌ها بسیار کم‌تر از مدل‌های تجربی است. بنابراین، مدل‌های وابسته به دما، نیاز به کالیبراسیون دقیق براساس مشاهدات محلی دارند (Aryalekshmi et al., 2021; Gharehbaghi et al., 2024). از طرف دیگر، مدل‌های مبتنی بر انتقال جرم، بدون توجه به منطقه آب و هوایی، دقت بالاتری از تخمین ET را تضمین می‌کنند. با این حال، آن‌ها به داده‌های ورودی بیش‌تری نیاز دارند که در برخی موارد، به دلایل مختلف، یافتن همه انواع اطلاعات مورد نیاز، غیرممکن است (Monteiro et al., 2021).

علی‌رغم مدل‌های تجربی فراوان برای پیش‌بینی ET، هنوز هیچ اجماع جهانی در مورد مناسب بودن استفاده از یک مدل پیشنهادی،

1 Evapotranspiration

2 Penman-Monteith

برای مناطق مختلف آب و هوایی وجود ندارد (سلطانی و قمرنیا، ۱۳۹۷؛ شیرمحمدی و همکاران، ۱۳۹۹). بنابراین، این مدل‌ها، به ویژه زمانی که در مناطق نیمه خشک با داده‌های آب و هوایی محدود اعمال می‌شوند، قبل از تخمین ET نیاز به کالیبراسیون محلی دقیق دارند (Alizamir et al., 2020; DehghaniSanij et al., 2004). مدل‌های مختلف محاسبات نرم برای جلوگیری از محدودیت‌های مرتبط با مدل‌های تجربی و برآورد ET با دقت معقول توسعه داده شده‌اند. در محاسبات نرم، به طور کلی مدل‌ها به داده‌های ورودی کم‌تری نیاز دارند و می‌توانند در مناطق مختلف آب و هوایی استفاده شوند. (مزیدی و همکاران، ۱۴۰۲؛ Gavili et al., 2018; Gocić et al., 2015; Kaya et al., 2021; Saremi, 2015) سه مدل محاسبات نرم را با پنج مدل تجربی برای تخمین تبخیر و تعرق مرجع (ET_o) در یک منطقه نیمه خشک مقایسه کردند. یافته‌های آن‌ها نشان داد که همه مدل‌های محاسبات نرم آزمایش شده، از مدل‌های تجربی بهتر عمل کردند. در میان مدل‌های محاسباتی نرم، شبکه عصبی مصنوعی (ANN^۱) نتایج بهتری نسبت به سیستم استنتاج عصبی فازی تطبیقی (ANFIS^۲) و برنامه‌ریزی بیان ژن (GEP^۳) ارائه داد. (Fan et al., 2019) عملکرد یک مدل محاسباتی نرم مبتنی بر درخت تصمیم، به نام ماشین تقویت گرادیان نور (LightGBM^۴) در مقایسه با درخت مدل درختی (M5 Tree) و جنگل‌های تصادفی (RF^۵) و همچنین چهار روش تجربی ارزیابی کردند. مدل‌ها، یعنی هارگریوز-سامانی، طبری، مکینک و تراپرت با استفاده از ترکیب‌های مختلف داده‌های آب و هوایی روزانه مورد ارزیابی قرار گرفت، آنها به این نتیجه رسیدند که مدل پیشنهادی، یعنی LightGBM از نظر دقت نتایج خوبی را ارائه می‌دهد و می‌تواند به عنوان یک مدل جایگزینی برای تخمین روزانه ET_o، به ویژه زمانی که داده‌های طولانی هواشناسی در دسترس نیست، استفاده شود. (Shamshirband et al., 2016) مقدار ET_o را با استفاده از ترکیبی از الگوریتم جستجوی فاخته (CSA^۶) با ANN و ANFIS تخمین زدند و نتایج را با دو مدل تجربی مقایسه کردند. آن‌ها به این نتیجه رسیدند که هر دو مدل محاسبات نرم ترکیبی بهتر از مدل‌های تجربی عمل می‌کند. (Aghelpour et al., 2022) تبخیر و تعرق محصول را با استفاده از پرسپترون چند لایه (MLP)، تابع پایه شعاعی (RBF^۷)، شبکه عصبی رگرسیون تعمیم یافته (GRNN^۸) و روش گروهی پردازش داده‌ها (GMDH^۹) برآورد کردند. یافته‌های آن‌ها نشان داد که مدل GMDH بهتر از سایر مدل‌های محاسباتی نرم و تجربی عمل می‌کند. (Mokari et al., 2022) از چهار مدل یادگیری ماشین (ML) یعنی ماشین یادگیری افراطی (ELM^{۱۰})، برنامه‌ریزی ژنتیکی (GP)، جنگل تصادفی (RF) و رگرسیون برداری پشتیبان (SVR^{۱۱}) برای تخمین ET_o روزانه با داده‌های اقلیمی محدود، در مناطق مختلف آب و هوایی در نیومکزیکو با استفاده از سناریوهای سناریوهای مختلف در ترکیب ورودی داده‌ها استفاده کردند. یافته‌های آن‌ها نشان داد که SVR و ELM نسبت به سایر مدل‌های محاسباتی نرم برای همه سناریوهای ورودی بهتر عمل می‌کنند. (Ferreira et al., 2022) مقدار ET_o را با مقایسه معادله FAO56-PM با جنگل تصادفی (RF-ANN) خطوط رگرسیون تطبیقی چند متغیره (MARS^{۱۲}) و افزایش گرادیان شدید (XGBoost^{۱۳}) به این نتیجه رسیدند که ترکیب مدل‌های محاسباتی نرم با معادله FAO56-PM برای تخمین ET_o عملکردی مشابه با استفاده از آن‌ها به صورت جداگانه دارد. (Sharma et al., 2022b) مقدار ET_o را با استفاده از کانولوشن-حافظه طولانی کوتاه مدت (Conv-LSTM) و شبکه عصبی پیچشی-حافظه طولانی کوتاه مدت (CNN-LSTM) تخمین زدند و آن‌ها را با سایر مدل‌های تجربی مانند هارگریوز، مکینک و ریچی با در نظر گرفتن ترکیب‌های ورودی مختلف داده‌های آب و هوایی مقایسه کردند. داده‌ها برای یافتن حداقل پارامترهای مورد نیاز برای تخمین ET_o با دقت معقول نیز مورد ارزیابی قرار گرفت. یافته‌های آن‌ها نشان می‌دهد که CNN-LSTM^{۱۴} و Conv-LSTM^{۱۵} از مدل‌های تجربی عملکرد بهتری دارند.

- 1 Artificial Neural Network
- 2 Adaptive network-based fuzzy inference system
- 3 Gene Expression Programming
- 4 Light Gradient-Boosting Machine
- 5 Random Forest
- 6 Contextualized Semantic Analysis
- 7 Radial Basis Functions
- 8 Generalized Regression Neural Network
- 9 Group Method of Data Handling
- 10 Extreme Learning Machine
- 11 support vector machine
- 12 Multivariate Adaptive Regression Splines
- 13 Extreme Gradient Boosting
- 14 Convolutional Neural Network (CNN) and Long Short-Term Memory network (LSTM)
- 15 Convolutional LSTM Network

به طور مشابه (Sharma et al., ۲۰۲۲a; Zhangkang Shu et al., ۲۰۲۲) دریافت‌اند که مدل پیشنهادی آن‌ها، می‌تواند ETO را با دقت قابل قبولی با استفاده از داده‌های کم‌تری نسبت به مدل‌های تجربی تخمین بزند. بنابراین، آن‌ها پیشنهاد کردند که مدل آن‌ها می‌تواند به عنوان یک مدل جایگزین برای برآورد ETO با هزینه کم‌تر از نظر داده‌های مورد نیاز و زمان محاسبات، استفاده شود. (Chi a, Huang, Koo, et al., ۲۰۲۲) از CNN-1D، حافظه کوتاه مدت (LSTM) و GRU (ساختار جعبه سیاه) برای تخمین ETO ماهانه در یک منطقه آب و هوای مرطوب استفاده کردند. آن‌ها پیشنهاد دادند که مدل‌های LSTM و GRU اگر به جای ساختار جعبه سیاه ساده، به صورت هیبریدی طراحی شوند، عملکرد بهتری خواهند داشت. این نشان می‌دهد که مدل‌های محاسبات نرم بهتر از مدل‌های تجربی در تخمین ETO عمل می‌کنند. همچنین، در میان مدل‌های محاسباتی نرم، مدل‌های ساختارهای ترکیبی عملکرد بهتری در زمان محاسبات و کاهش خطا دارند (El bel tagi et al., ۲۰۲۲; Zhao et al., ۲۰۲۳).

در حوضه دشت اردبیل پنج ایستگاه هواشناسی وجود دارد که سه ایستگاه با نام‌های اردبیل، فرودگاه و نمین در داخل محدوده دشت و دو ایستگاه با نام‌های نیر و سرعین در سمت جنوب غربی دشت و خارج از محدوده آن واقع شده‌اند. در میان آن‌ها به جز ایستگاه اردبیل، سایر ایستگاه‌ها به طور متوسط از سال ۱۳۸۵ یا به بعد راه‌اندازی شده‌اند که نشان از عدم وجود داده‌های هواشناسی بلندمدت در این منطقه دارد. همچنین موقعیت مکانی ایستگاه‌ها در این محدوده به صورتی است که هیچ اطلاعاتی آماری در خصوص داده‌های هواشناسی در نواحی شمالی و جنوبی دشت ارائه نمی‌دهند؛ از طرفی همانگونه که بیان شد برای محاسبه تبخیر و تعرق، به عنوان بخش مهمی از تلفات منابع آب به‌ویژه در منطقه مورد مطالعه که با محدودیت‌های شدید منابع آب درگیر است، مدل قابل اطمینانی نیاز است تا باتوجه به محدودیت‌های مرتبط با مدل‌های تجربی در مواجهه با داده‌های مشاهداتی محدود و ناقص، آن مدل با حفظ پایداری در مقابل تغییرات داده‌ها، به عنوان مدل مناسب و قابل اطمینان برای محاسبه تبخیر و تعرق باشد. به همین سبب در این مطالعه تلاش شد برای تخمین ETO در منطقه دشت اردبیل کارایی دو مدل جنگل تصادفی و رگرسیون خطی چندگانه بررسی شود.

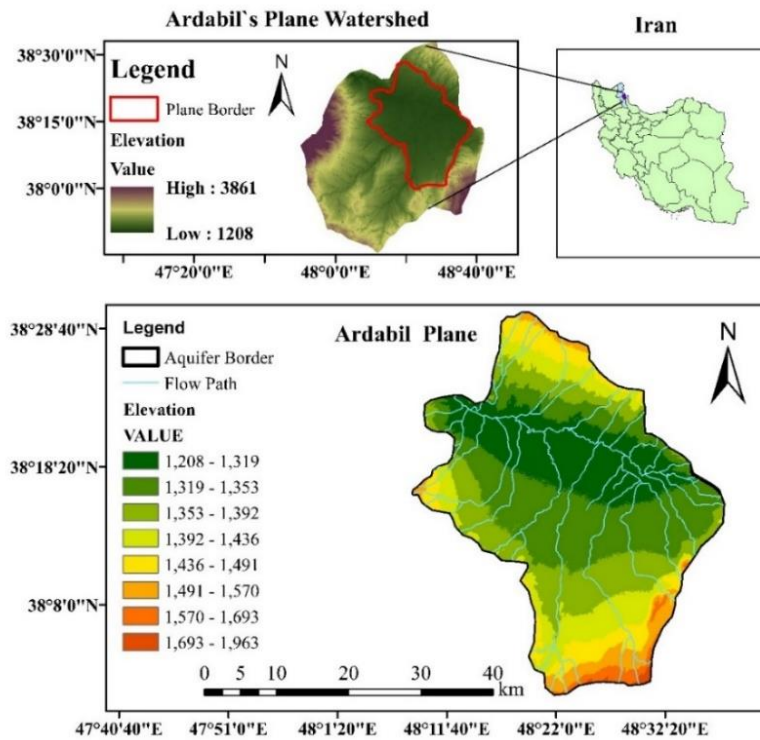
مواد و روش‌ها

منطقه مورد مطالعه

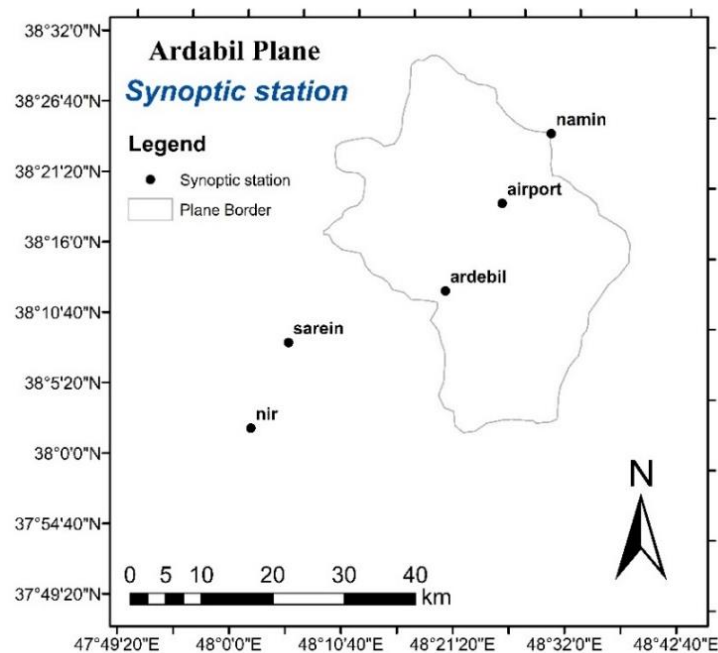
این تحقیق در محدوده‌ی دشت اردبیل و حوضه مجاور آن انجام شد. دشت اردبیل در منطقه‌ی مرکزی استان اردبیل و شمال غرب ایران قرار دارد و با مساحتی در حدود ۱۲۹۱ کیلومتر مربع، در محدوده‌ی مختصات $38^{\circ}00'50''$ تا $38^{\circ}31'10''$ شمالی و $48^{\circ}09'00''$ تا $48^{\circ}39'00''$ شرقی واقع شده است. این منطقه دارای شرایط اقلیمی نیمه خشک سرد است که متوسط بارش سالانه‌ی آن در ایستگاه سینوپتیک اردبیل حدود ۲۸۰ میلی‌متر ثبت شده است. در این دشت، اردیبهشت، پربارش‌ترین ماه و میانگین دمای سالانه ۹/۹ درجه سانتی‌گراد است. حداقل دمای مطلق این دشت، $-33/8$ درجه‌ی سانتی‌گراد و متوسط تعداد روزهای یخبندان در آن، ۱۳۰ روز در سال ثبت شده است (شکل ۱).

داده‌های مورد استفاده

در این مطالعه و در حدفاصل سال‌های ۱۳۸۵ تا ۱۴۰۲، ورودی‌های مدل، شامل داده‌های هواشناسی، از سایت سازمان هواشناسی ایران دریافت شد. خروجی و پارامتر هدف، شامل داده‌های مربوط به تبخیر و تعرق مرجع (ETO) بود که به روش پنمن ماتیت فائو (FAO56) محاسبه شد (عبدزادگوهری و همکاران، ۱۴۰۲؛ کریمی و همکاران، ۱۴۰۳؛ طباطبایی و همکاران، ۱۴۰۱). با این حال مطابق توضیحات قبلی، چالش‌هایی مانند عدم دقت اندازه‌گیری و یا عدم ثبت داده‌ها در بازه‌های کوتاهی از دوره مورد مطالعه، در طول فرآیند جمع‌آوری داده‌ها وجود داشت که برای کاهش تاثیر مشکل نقص داده‌ها، که در برخی ایستگاه‌ها به صورت روزانه مشاهده می‌گردید، و همچنین به منظور کاهش حجم داده‌ها، مقادیر پارامترها در بازه‌های ده روزه (دهه اول، دوم و سوم هر ماه) محاسبه شد. به جز استانداردسازی، داده‌کاوی دیگری به سبب انتخاب نوع مدل‌ها و ویژگی ابر روی پارامترها انجام نشد. برای این منظور از داده‌های مجموع پنج ایستگاه هواشناسی به نام‌های ایستگاه هواشناسی نیر، سرعین، اردبیل، فرودگاه اردبیل و نمین که در منطقه و یا نزدیک به محدوده دشت بودند استفاده شد (شکل ۲).



شکل ۱. منطقه مورد مطالعه (دشت اردبیل)



شکل ۲. موقعیت ایستگاه‌های هواشناسی

داده‌های هواشناسی دریافتی شامل ۱۴ پارامتر به شرح حداقل دما (t_{min})، حداکثر دما (t_{max})، میانگین دما (t_m)، بارش روزانه ($rrr24$)، تبخیر روزانه (evt)، حداکثر سرعت باد (ff_{max})، میانگین سرعت باد (ffm)، جهت حداکثر باد (dd_{max})، میزان تابش ۲۴ ساعته ($radglo24$)، ساعت آفتابی روز ($sshn$)، میانگین دمای خاک ($tsoil_m$)، حداکثر رطوبت ($umax$)، حداقل رطوبت ($umin$) و میانگین رطوبت (um) بود.

روش مطالعه

تعبیر تبخیر و تعرق مرجع (ETO) برای اولین بار توسط جنسن (Jensen, 1968)، برای تکمیل و توضیح مفهوم تبخیر و تعرق مرجع ارائه شد، سپس در نشریه فائو تعریف آن به صورت کامل و به عنوان یک مرجع ارائه گردید (Allen et al., 1998). تبخیر و تعرق مرجع مهم‌ترین بخش در تخمین مقدار نیاز آبی گیاهان و طراحی سامانه‌های آبیاری است. یکی از مهم‌ترین چالش‌ها برای برآورد دقیق تبخیر و تعرق مرجع، عدم وجود و یا اغلب کمبود داده‌های هواشناسی در کشور است (Barideh et al., 2022). بیشتر روش‌های مورد استفاده بر مبنای اندازه‌گیری نقطه‌ای پارامترها بنا شده است، با این وجود برآورد هر پارامتری در مقیاس نقطه‌ای منتهی به حصول نتایج قابل قبول در مقیاس منطقه‌ای نمی‌شود (Kalluri et al., 2003).

در این مطالعه برای برآورد میزان تبخیر و تعرق مرجع در دشت اردبیل از دو مدل رگرسیون خطی چندگانه (MLR) و جنگل تصادفی (RF) استفاده شد. دلیل انتخاب این مدل‌ها استفاده گسترده از آن‌ها در مدل‌های پیش‌بینی هیدرولوژیکی بود (Z. Shu et al., 2022); مدل MLR به سبب سادگی به معنی راحتی در تفسیر نتایج، محاسبات کارآمد و ارایه بینش‌هایی قابل درک در مورد اهمیت نسبی متغیرها، برای مدل‌سازی تبخیر و تعرق به‌ویژه در مناطقی که دارای محدودیت داده‌های هواشناسی است مورد استفاده قرار گرفته است (Da Silva et al., 2016; Dimitriadou & Nikolakopoulos, 2022; Duhan et al., 2023; Koç & Erkan Can, 2023). همچنین مدل RF نیز به سبب قابلیت انطباق با محدودیت داده‌ها، به‌طور موثر برای مناطقی با داده‌های محدود مناسب بوده و در تعاملات پیچیده بین متغیرهای آب و هوایی مدلی قابل اعتماد است که با توانایی که دارد می‌تواند در تصمیم‌گیری برای شرایط عدم قطعیت کمک‌کننده باشد (Abed et al., 2022; Amani & Shafizadeh-Moghadam, 2023; Ao et al., 2018; T R et al., 2023; Wu et al., 2020;) (Zhu, 2020).

با این حال روش کلی به کار رفته در این تحقیق، به منظور بررسی دقت مدل جنگل تصادفی در برآورد تبخیر و تعرق مرجع با استفاده از داده‌های مشاهداتی محدود، بدون انجام داده‌کاوی، با توجه به قابلیت‌های مدل جنگل تصادفی بود. شرح کلی مراحل مدل‌سازی به این صورت است که، در مرحله اول همبستگی ۱۴ پارامتر هواشناسی که از سازمان هواشناسی ایران دریافت شده بود با داده‌های تبخیر و تعرق مرجع که با روش پنمن مانیتیت (فائو ۵۶) محاسبه شده بود و به عنوان پارامتر هدف مدنظر بود، مقایسه گردید، سپس برای کاهش حجم داده‌ها، پارامترها با محاسبه مقادیر داده‌های هواشناسی، به بازه‌های ۱۰ روزه (دهه اول، دوم و سوم هر ماه) تقسیم‌بندی شدند. با بررسی همبستگی‌ها از بین پارامترهای اولیه، پارامترهایی که همبستگی بسیار پایینی با تبخیر تعرق مرجع داشتند از لیست داده‌های ورودی حذف شدند. با انتخاب پارامترهای ورودی نهایی، یک سری‌زمانی ترکیبی و تصادفی به طول ۱۶۰۰ واحد، بدون رعایت ترتیب زمانی و مکانی، با استفاده از کل داده‌های چهار ایستگاه منتخب برای ساخت و آموزش مدل (ایستگاه‌های اردبیل، نیر، سرعین و فرودگاه) ساخته شد. ساخت این سری‌زمانی علاوه بر کاهش نیاز مدل‌ها به داده‌کاوی، باعث پایداری مدل در مقابل داده‌های جدید می‌شود (Akbari Majd et al., 2024). در نهایت با استفاده از ۷۰ درصد داده‌های سری‌زمانی فوق در مرحله آموزش و بعد از آن ۱۵ درصد برای اعتبارسنجی و تست، مدل‌ها ساخته شدند. ارزیابی نهایی با استفاده از آمار ایستگاه نمین که در مرحله قبل از آن‌ها استفاده نشده بود انجام گردید.

رگرسیون خطی چندگانه

یک روش آماری به نام تحلیل رگرسیون چندگانه یا MLR، مجموعه‌ای از برآوردهای پارامتر را برای پیش‌بینی نتیجه یک متغیر وابسته را ترکیب می‌کند. از رگرسیون‌های خطی چندگانه برای نشان دادن ارتباط خطی بین متغیرهای مستقل و وابسته استفاده می‌شود. روش‌های مختلفی از تحلیل رگرسیون وجود دارد که از بین آن‌ها استفاده از رگرسیون خطی بسیار معمول‌تر است. به‌طور کلی دو نوع رگرسیون خطی وجود دارد، رگرسیون خطی ساده و رگرسیون خطی چندگانه که روش‌های مختلفی نیز برای محاسبه خطا و حداقل کردن آن وجود دارد. رابطه تعریف شده در حالت کلی به صورت زیر است:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon \quad (۱) \text{ رابطه}$$

که برای n داده مشاهداتی Y_i متغیر وابسته، x_i متغیر مستقل، β_0 ضریب ثابت، β_p ضریب متغیر مستقل، ε خطای مدل است. معیار ارزیابی که در مدل رگرسیون خطی چندگانه استفاده می‌شود، حداقل کردن مجموع مربعات خطا است. چون که میانگین

مقدارهای خطا صفر در نظر گرفته شده، پس زمانی مجموع مربعات خطا، حداقل ممکن را خواهند داشت که توزیع داده‌ها نرمال باشند. در نتیجه، نرمال بودن داده‌های مربوط به متغیر وابسته یا باقیمانده‌ها یکی از فرضیات مهم برای مدل رگرسیونی خطی است (fattahi dolatabadi et al., 2018; Gokcekus et al., 2023; Krzywinski & Altman, 2015; Nolan et al., 2015; Piña-Monarez & Ortiz-Yañez, 2015).

مدل جنگل تصادفی (RF)

جنگل تصادفی یک طبقه‌بندی‌کننده (رگرسیون) خاص است که توسط مجموعه‌ای از درخت‌های تصمیم که به‌عنوان بردارهای تصادفی مستقل و به‌طور یکسان توزیع شده‌اند، تشکیل می‌شود که توسط بریمن (Breiman, 2001) پیشنهاد شده است. بریمن مدل‌های Bagging را به RF ارتقا داد. مدل RF از دو پارامتر استفاده می‌کند: تعداد درختان رگرسیون (ntree) برای رشد، و تعداد ویژگی‌هایی که به‌طور تصادفی در هر تقسیم (mtry) نمونه‌برداری شده‌اند (Breiman, 2001; Dong et al., 2020). در نتیجه این ویژگی، هنگامی که همه درختان آموزش داده شدند، بدون اعمال تغییر (هرس)، هر مورد جدید توسط همه درختان تجزیه و تحلیل می‌شود و پیش‌بینی نهایی با میانگین نتایج به‌دست می‌آید. آموزش الگوریتم پایه در مدل RF مانند مدل Bagging موازی است. مدل RF پیش‌بینی را با دقت بالا، بدون برازش بیش از حد، برای مدل‌سازی ویژگی‌های هیدرولوژیکی ارائه می‌دهد (Breiman, 2001; Ruiz-Álvarez et al., 2021; Wang et al., 2017). هنگامی که مدل آموزش داده شد، پیش‌بینی را می‌توان به‌صورت زیر به‌دست آورد (Ruiz-Álvarez et al., 2021):

$$\hat{Y}_i = \frac{1}{M} \sum_{m=1}^M T_m(f_t) \quad \text{رابطه ۲}$$

که در آن M تعداد درخت‌ها، T_m درخت تصمیم منفرد و f_t بردار پیش‌بینی‌کننده است. در مطالعه حاضر از مدل RF در تخمین ETo با استفاده از متغیرهای آب و هوایی استفاده شده است.

معیارهای ارزیابی

آماره‌های ارزیابی مورد استفاده در این مطالعه شامل میانگین مربع خطا (MSE) (Hyndman & Koehler, 2006)، R^2 (Draper, 1960; Steel & Torrie, 1990; Glantz, 1990; 1998)، ضریب کارایی نش-ساتکلیف (NSE) (Nash & Sutcliffe, 1970) است.

$$RMSE = \sqrt{MSE}; \quad MSE = \frac{1}{n} \sum_{i=1}^n (Y_o - Y_m)^2 \quad \text{رابطه ۲}$$

$$R^2 = 1 - \frac{RSS}{TSS}; \quad (RSS: \sum_{i=1}^n (Y_o - Y_m)^2, \quad TSS: \sum_{i=1}^n (Y_o - \bar{Y}_o)^2) \quad \text{رابطه ۳}$$

$$NSE = 1 - \frac{\sum_{t=1}^T (Y_o - Y_m)^2}{\sum_{t=1}^T (Y_o - \bar{Y}_o)^2} \quad \text{رابطه ۴}$$

در تمامی روابط: \bar{Y}_o میانگین داده‌ای مشاهده‌ای، Y_m خروجی مدل، Y_o داده مشاهده‌ای، n : تعداد داده‌ها، COV : کوواریانس، σ_{Y_o} & σ_{Y_m} : انحراف معیار پارامترهای است.

نتایج و بحث

جمع‌آوری داده‌ها

در وهله اول داده‌های هواشناسی دریافتی شامل ۱۴ پارامتر به شرح حداقل دما (tmin)، حداکثر دما (tmax)، میانگین دما (tm)، بارش روزانه (rr24)، تبخیر روزانه (evt)، حداکثر سرعت باد (ff_max)، میانگین سرعت باد (ffm)، جهت حداکثر باد (dd_max)، میزان تابش ۲۴ ساعته (radglo24)، ساعت آفتابی روز (sshn)، میانگین دمای خاک (tsoil_m)، حداکثر رطوبت (umax)، حداقل رطوبت (umin) و میانگین رطوبت (um) و تبخیر و تعرق مرجع ETo، در محل ۵ ایستگاه هواشناسی شامل ایستگاه‌های نیر، سرعین، اردبیل، فرودگاه اردبیل و نمین در حد فاصل سال‌های ۱۳۸۵ تا ۱۴۰۱ محاسبه شد.

جدول ۱. آمار توصیفی ایستگاه هواشناسی اردبیل

متغیرها	تعداد	میانگین	واریانس	حداقل	میان	حداکثر	چولگی	کشیدگی
tmin	۶۱۲	۴/۲۸	۴۷/۵۱	-۲۳/۸۸	۵/۲۷	۱۶/۸۰	--/۵۲	--/۲۶
tmax	۶۱۲	۱۶/۶۰	۷۳/۴۲	-۷/۹۴	۱۷/۸۳	۳۲/۹۸	--/۲۹	--/۹۹
tm	۶۱۲	۱۰/۴۴	۵۸/۰۷	-۱۵/۹۱	۱۱/۴۲	۲۳/۲۹	--/۴۰	--/۷۶
rrr ^{۲۴}	۶۱۲	۰/۷۰	۰/۸۶	۰/۰۰	۰/۳۲	۶/۰۵	۲/۰۵	۵/۱۴
evt	۳۸۸	۵/۷۳	۶/۹۸	۰/۴۵	۵/۷۳	۲۰/۰۰	۰/۸۹	۳/۳۴
ff_max	۶۱۲	۹/۴۷	۳/۹۶	۱/۷۰	۹/۴۸	۱۹/۷۸	۰/۳۸	۲/۳۷
ffm	۶۱۲	۳/۷۵	۱/۹۲	۰/۳۵	۳/۵۴	۹/۹۹	۱/۰۴	۱/۷۶
dd_max	۶۱۲	۱۱۹/۴۱	۳۰۳۰/۲۴	۳۱/۰۰	۱۲۱/۰۰	۲۸۲/۰۰	۰/۱۵	--/۸۸
radglo ^{۲۴}	۴۷۲	۱۷۰۹/۸۰	۵۹۷۵۵۰/۶۰	۳۸۱/۷۰	۱۶۱۶/۳۰	۵۸۲۶/۸۰	۱/۲۰	۳/۴۴
sshn	۶۱۲	۷/۰۸	۵/۹۸	۰/۴۸	۶/۸۱	۱۲/۶۱	۰/۱۹	--/۶۵
tsoil_m	۶۱۲	۲/۶۶	۴۴/۶۷	-۲۶/۳۰	۳/۳۰	۱۵/۵۰	--/۵۱	--/۱۶
umax	۶۱۲	۸۹/۰۷	۵۴/۳۷	۶۰/۵۰	۹۰/۴۶	۱۰۰/۰۰	-۱/۰۳	۰/۸۴
umin	۶۱۲	۵۲/۰۶	۱۶۲/۵۷	۲۲/۱۰	۵۱/۲۰	۸۷/۵۰	۰/۱۷	--/۴۹
um	۶۱۲	۷۱/۹۶	۱۰۲/۳۶	۴۳/۲۶	۷۲/۴۱	۹۵/۱۹	--/۳۱	--/۳۲
ETO	۶۱۲	۲۹/۰۵	۲۵۴/۱۳	۵/۵۵	۲۶/۲۱	۶۵/۸۹	۰/۲۸	-۱/۲۱

جدول ۲. آمار توصیفی ایستگاه هواشناسی نمین

متغیرها	تعداد	میانگین	واریانس	حداقل	میان	حداکثر	چولگی	کشیدگی
tmin	۵۲۸	۵/۸۹	۵۰/۸۹	-۱۵/۰۶	۶/۱۴	۱۹/۴۶	--/۲۲	--/۹۸
tmax	۵۲۸	۱۵/۲۹	۶۱/۹۵	-۶/۶۸	۱۶/۵۸	۳۵/۴۷	--/۳۰	--/۹۱
tm	۵۲۸	۱۰/۵۸	۵۵/۶۷	-۱۰/۴۳	۱۱/۵۰	۲۷/۵۸	--/۲۷	--/۹۶
rrr ^{۲۴}	۵۲۶	۰/۷۱	۰/۸۶	۰/۰۰	۰/۴۰	۵/۹۰	۲/۱۹	۵/۸۴
evt	۲۷۳	۵/۹۸	۶/۵۲	۰/۰۰	۵/۹۰	۱۲/۶۸	۰/۱۵	--/۶۱
ff_max	۵۲۸	۱۰/۲۳	۵/۸۷	۳/۶۰	۱۰/۴۰	۱۷/۹۰	--/۱۱	۰/۰۸
ffm	۵۲۸	۵/۶۹	۳/۰۴	۱/۱۶	۵/۶۲	۱۰/۶۸	۰/۱۶	--/۲۸
dd_max	۵۲۸	۱۳۰/۱۴	۲۵۳۸/۳۲	۶۳/۶۴	۱۲۳/۰۰	۲۶۲/۵۰	۰/۵۱	--/۹۲
radglo ^{۲۴}	۱	۲۲۲۲/۰۰	*	۲۲۲۲/۰۰	۲۲۲۲/۰۰	۲۲۲۲/۰۰	*	*
sshn	۳۴۳	۶/۶۹	۵/۰۱	۰/۹۳	۶/۵۷	۱۲/۱۹	۰/۰۷	--/۳۸
tsoil_m	۵۲۸	۴/۹۲	۶۰/۵۹	-۱۷/۴۰	۵/۰۰	۲۶/۴۴	۰/۱۲	--/۳۷
umax	۵۲۸	۷۹/۵۸	۱۴۷/۳۱	۳۹/۰۰	۸۱/۵۰	۱۰۰/۰۰	--/۷۴	۰/۳۰
umin	۵۲۸	۵۱/۵۰	۱۹۲/۹۳	۱۷/۹۰	۵۰/۹۱	۸۹/۴۰	۰/۱۹	--/۴۸
um	۵۲۸	۶۴/۸۰	۱۷۱/۸۱	۳۱/۲۰	۶۶/۱۱	۹۴/۶۶	--/۲۱	--/۳۶
ETO	۵۴۸	۲۷/۷۷	۲۵۸/۷۲	۵/۴۴	۲۴/۲۶	۸۱/۷۱	۰/۵۷	--/۵۵

جدول ۳. آمار توصیفی ایستگاه هواشناسی نیر

متغیرها	تعداد	میانگین	واریانس	حداقل	میان	حداکثر	چولگی	کشیدگی
tmin	۵۰۶	۳/۹۱	۳۹/۵۲	-۱۴/۳۸	۴/۴۲	۱۶/۷۴	--/۳۳	--/۸۲
tmax	۵۰۶	۱۶/۳۵	۸۳/۴۸	-۲/۰۹	۱۶/۸۸	۳۲/۸۶	--/۱۲	-۱/۲۱
tm	۵۰۶	۱۰/۱۴	۵۸/۲۳	-۸/۲۵	۱۰/۵۸	۲۴/۵۵	--/۲۱	-۱/۱۲
rrr ^{۲۴}	۵۰۶	۰/۹۸	۱/۴۴	۰/۰۰	۰/۵۲	۶/۹۲	۱/۸۵	۴/۰۸
evt	۳۱۴	۵/۲۰	۴/۴۰	۰/۰۰	۵/۱۴	۹/۵۷	--/۱۰	--/۷۱
ff_max	۵۰۶	۹/۶۵	۵/۹۴	۳/۷۰	۹/۲۰	۱۹/۳۰	۰/۹۴	۱/۳۳
ffm	۵۰۶	۵/۱۳	۲/۶۱	۱/۲۶	۵/۰۰	۱۰/۷۷	۰/۶۸	۰/۴۴
dd_max	۵۰۶	۱۴۳/۳۰	۴۱۳۲/۲۹	۳۵/۸۳	۱۵۰/۸۷	۳۶۰/۰۰	۰/۰۷	--/۶۷
radglo ^{۲۴}	۶۰	۱۸۶۷/۰۰	۷۷۵۰۰۳/۰۰	۱۶/۰۰	۱۹۲۶/۰۰	۳۲۶۱/۰۰	--/۲۷	--/۸۵
sshn	۳۰۰	۶/۹۶	۵/۲۴	۰/۹۵	۶/۸۲	۱۲/۵۱	۰/۱۴	--/۵۱
tsoil_m	۵۰۶	۱/۱۹	۴۱/۵۰	-۱۷/۸۰	۱/۴۷	۱۴/۹۰	--/۳۵	--/۶۶
umax	۵۰۶	۸۰/۲۳	۹۵/۹۴	۴۴/۲۰	۸۱/۷۴	۹۹/۰۰	--/۵۱	۰/۱۰



umin	۵۰۶	۴۴/۵۳	۱۶۵/۷۲	۱۶/۵۰	۴۳/۵۰	۸۹/۰۰	۰/۲۶	-۰/۲۷
um	۵۰۶	۶۰/۲۷	۱۳۲/۶۶	۲۸/۷۶	۶۰/۲۰	۹۲/۹۲	-۰/۰۴	-۰/۱۸
ETO	۵۰۶	۲۸/۲۸	۱۹۵/۵۹	۵/۳۰	۲۷/۱۴	۷۱/۳۵	۰/۴۰	-۰/۷۴

جدول ۴. آمار توصیفی ایستگاه هواشناسی سرعین

متغیرها	تعداد	میانگین	واریانس	حداقل	میان	حداکثر	چولگی	کشیدگی
tmin	۶۰۹	۴/۹۹	۵۵/۴۹	-۱۶/۱۰	۵/۶۸	۱۸/۶۴	-۰/۲۶	-۰/۹۸
tmax	۶۰۹	۱۵/۶۳	۷۴/۸۳	-۶/۲۸	۱۶/۰۶	۳۱/۸۰	-۰/۱۳	-۱/۱۶
tm	۶۰۹	۱۰/۳۱	۶۴/۱۷	-۱۰/۴۲	۱۰/۸۰	۲۴/۷۱	-۰/۱۹	-۱/۱۰
rrr ^{۲۴}	۶۰۹	۱/۰۷	۱/۵۷	-/۰۰	-/۰۰	۷/۰۸	۱/۷۷	۳/۷۰
evt	۳۷۷	۴/۷۸	۵/۲۹	-/۰۰	۴/۹۰	۲۰/۱۰	۰/۷۴	۴/۱۲
ff_max	۶۰۹	۷/۳۵	۵/۱۹	۲/۰۰	۷/۱۷	۲۰/۱۰	۰/۹۱	۳/۱۰
ffm	۶۰۹	۳/۱۸	۱/۷۳	-/۴۸	۳/۰۸	۹/۳۴	۰/۶۷	۱/۰۷
dd_max	۶۰۹	۱۳۷/۱۹	۲۷۱۴/۹۸	۲۵/۰۰	۱۲۵/۴۵	۲۵۵/۰۰	۰/۱۳	-۰/۸۶
radglo ^{۲۴}	۶۷	۱۸۷۵/۰۰	۱۰۴۲۵۷۱/۰۰	۲۴۹/۰۰	۱۶۲۶/۰۰	۵۷۰۳/۰۰	۱/۴۶	۳/۰۹
sshn	۴۰۳	۶/۸۲	۴/۸۱	-/۰۹	۶/۶۳	۱۲/۵۱	۰/۱۶	-۰/۱۳
tsoil_m	۶۰۹	۲/۶۶	۵۴/۸۴	-۱۷/۸۰	۳/۳۰	۱۶/۳۰	-۰/۲۶	-۰/۹۵
umax	۶۰۹	۷۶/۱۰	۱۵۲/۱۶	۳۴/۵۰	۷۷/۷۸	۱۰۰/۰۰	-۰/۵۴	۰/۰۶
umin	۶۰۹	۴۴/۸۰	۱۶۶/۵۹	۱۲/۸۰	۴۴/۳۳	۸۷/۲۰	۰/۲۱	-۰/۲۴
um	۶۰۹	۵۹/۲۱	۱۶۳/۲۷	۲۲/۰۶	۵۹/۷۰	۹۲/۹۴	-۰/۱۴	-۰/۲۹
ETO	۶۰۹	۲۶/۵۳	۱۸۵/۶۱	۵/۶۱	۲۵/۶۹	۶۲/۸۸	۰/۳۱	-۱/۱۳

جدول ۵. آمار توصیفی ایستگاه هواشناسی فرودگاه اردبیل

متغیرها	تعداد	میانگین	واریانس	حداقل	میان	حداکثر	چولگی	کشیدگی
tmin	۶۰۹	۲/۳۹	۵۶/۹۸	-۲۶/۲۲	۳/۷۰	۱۵/۷۷	-۰/۶۲	-۰/۱۴
tmax	۶۰۹	۱۶/۳۱	۶۷/۸۷	-۶/۴۶	۱۷/۷۴	۳۱/۷۰	-۰/۳۶	-۰/۹۰
tm	۶۰۹	۹/۳۵	۵۹/۸۳	-۱۶/۳۴	۱۰/۵۸	۲۲/۰۸	-۰/۵۰	-۰/۶۲
rrr ^{۲۴}	۶۰۹	۰/۶۲	۰/۷۱	-/۰۰	-/۲۹	۵/۸۹	۲/۲۸	۶/۵۵
evt	۳۶۸	۵/۶۵	۷/۱۰	-/۰۰	۵/۵۵	۱۲/۹۷	۰/۱۰	-۰/۷۵
ff_max	۶۰۹	۶/۸۵	۶/۰۸	-/۸۵	۶/۵۲	۱۲/۹۳	۰/۳۰	-۰/۶۵
ffm	۶۰۹	۱۰/۳۲	۳/۶۶	۲/۳۰	۱۰/۶۰	۱۵/۶۰	-۰/۷۲	۰/۸۰
dd_max	۶۰۹	۵/۰۶	۱/۵۸	-/۲۶	۵/۱۳	۸/۶۴	-۰/۲۸	۰/۴۰
radglo ^{۲۴}	۶۰۹	۱۲۲/۸۷	۲۴۱۵/۸۷	۵۸/۰۰	۱۱۶/۶۷	۲۵۷/۰۰	۰/۴۲	-۰/۹۸
sshn	۲	۰/۰۰	۰/۰۰	-/۰۰	-/۰۰	-/۰۰	*	*
tsoil_m	۶۰۸	۰/۷۶	۵۴/۱۴	-۲۶/۱۰	۱/۸۷	۱۴/۵۰	-۰/۵۷	-۰/۲۷
umax	۶۰۹	۹۰/۹۴	۴۹/۸۴	۶۲/۰۰	۹۲/۹۰	۱۰۰/۰۰	-۱/۱۴	۰/۹۸
umin	۶۰۹	۵۰/۱۰	۱۶۳/۱۳	۲۱/۲۰	۴۹/۲۰	۸۷/۵۰	۰/۲۵	-۰/۲۸
um	۶۰۹	۷۱/۸۶	۱۱۹/۱۲	۳۸/۳۷	۷۲/۷۹	۹۶/۳۹	-۰/۳۵	-۰/۳۰
ETO	۶۰۹	۲۹/۶۸	۲۹۰/۴۶	۵/۲۷	۲۶/۹۰	۷۷/۵۸	۰/۴۱	-۰/۸۷

همان‌گونه که از جدول شماره (۱) تا (۵) قابل ملاحظه است، طول دوره آماری در ایستگاه‌های مختلف، متفاوت از یکدیگر هستند که دلیل آن بیش‌تر به سال تاسیس ایستگاه مربوط می‌شود. همچنین در بین داده‌های هر ایستگاه، دوره آماری برخی پارامترها مانند رطوبت یا میزان تابش ۲۴ ساعته نیز متناسب با طول دوره آماری با سایر پارامترها نیست. به همین دلیل و به سبب ساخت یک مدل که توانایی درک و ایجاد ارتباط، مابین پارامترهای ورودی و تبخیر و تعرق مرجع را داشته باشد، از ترکیب داده‌های ۴ ایستگاه هواشناسی شامل ایستگاه‌های نیر، سرعین، اردبیل و فرودگاه اردبیل استفاده شد. از داده ایستگاه نیر که کم‌ترین طول دوره آماری و بیش‌ترین داده ناقص را در بین سایر ایستگاه‌ها داشت، به‌عنوان داده‌هایی که با آن ارزیابی نهایی مدل‌ها انجام شد، استفاده گردید.

انتخاب داده‌های نهایی

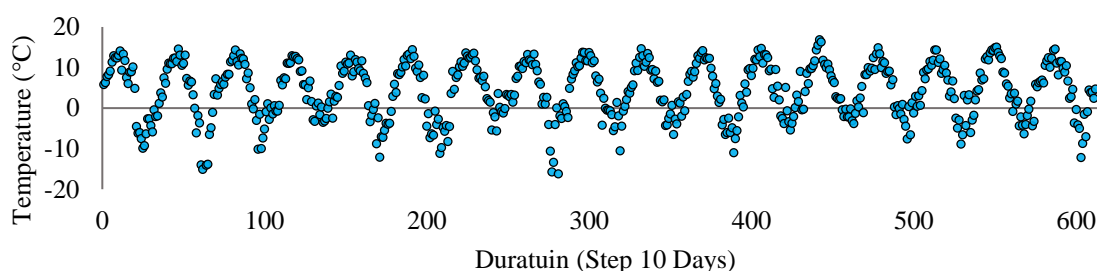
در این مرحله و به منظور انتخاب ترکیب نهایی پارامترهای ورودی، همبستگی مابین پارامترهای ورودی و تبخیر و تعرق مرجع محاسبه شد و با توجه به نتیجه به دست آمده، پارامترهایی که کمترین همبستگی را با تبخیر و تعرق مرجع داشتند، از بین ورودی‌ها حذف شد. برای اختصار تنها نتیجه همبستگی مابین داده‌های ایستگاه هواشناسی اردبیل در جدول (۶) ارائه شده است.

جدول ۶. همبستگی تبخیر و تعرق مرجع با پارامترهای هواشناسی

ffm	ff_max	evt	۲۴rrr	tm	tmax	tmin	
-.۳۱	-.۰۱	۰/۴۹	-.۲۲	۰/۸۶	۰/۸۴	۰/۸۵	ETO
um	umin	umax	tsoil_m	sshn	radglo۲۴	dd_max	
-.۱۶	-.۲۹	۰/۰۰	۰/۸۵	۰/۶۶	۰/۵۴	-.۷۸	ETO

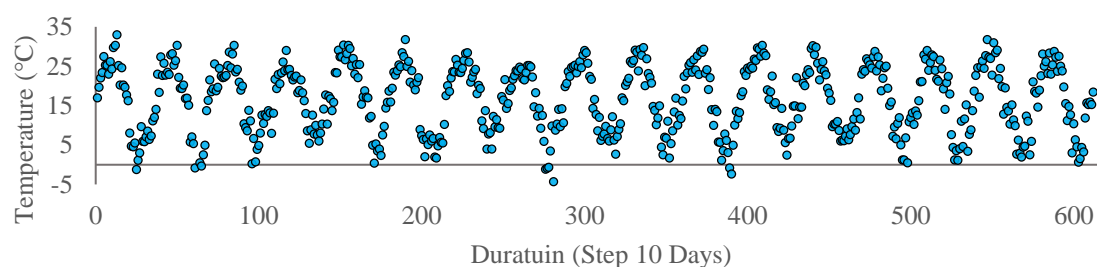
با توجه به نتیجه به دست آمده از بین ۱۴ پارامتر اولیه، ۷ پارامتر شامل دمای مینیمم، ماکزیمم و میانگین، تبخیر، جهت حداکثر باد، ساعت آفتابی روز و دمای خاک که به صورت پرننگ در جدول (۶) نمایش داده شده است، به عنوان پارامترهای ورودی نهایی انتخاب شدند (شکل ۳). پارامتر میزان تابش ۲۴ ساعته با توجه به جدول (۱) تا (۵) به دلیل تعداد مشاهدات بسیار ناقص از بین ورودی‌ها حذف شد. همچنین علیرغم همبستگی بالا مابین دمای مینیمم، ماکزیمم و میانگین با یکدیگر، به سبب تغییرات دمایی موجود در سطح دشت که در جدول (۱) تا (۵) نیز قابل مشاهده است و برای تشخیص الگو و طبقه‌بندی (ویژگی مدل جنگل تصادفی) مابین ورودی‌ها و تابع هدف و در نتیجه در مناطق مختلف محدوده مطالعاتی، از هر سه آن‌ها در داده‌های ورودی استفاده شد.

Temperature (min)

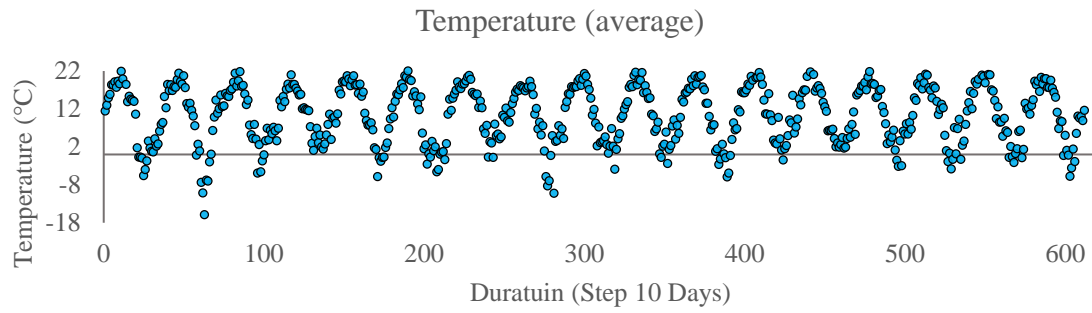


شکل ۱,۳. سری زمانی دمای حداقل

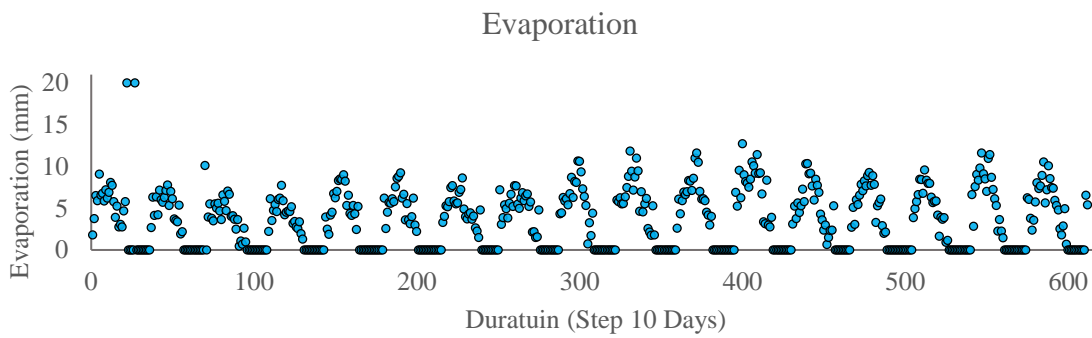
Temperature (max)



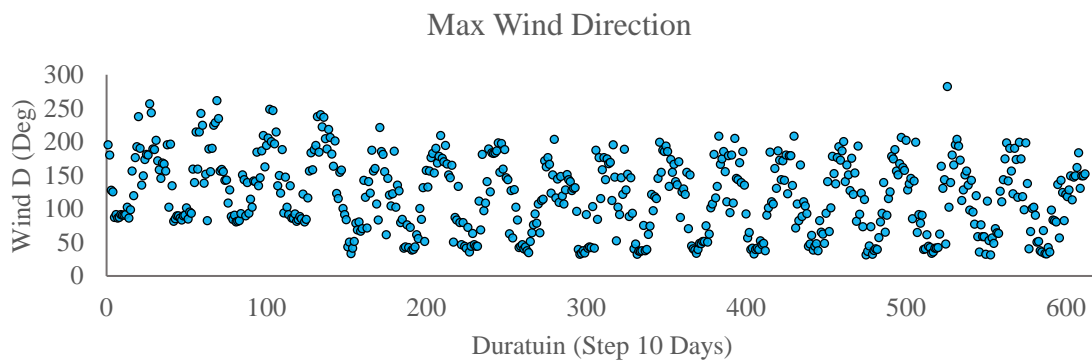
شکل ۲,۳. سری زمانی دمای ماکزیمم



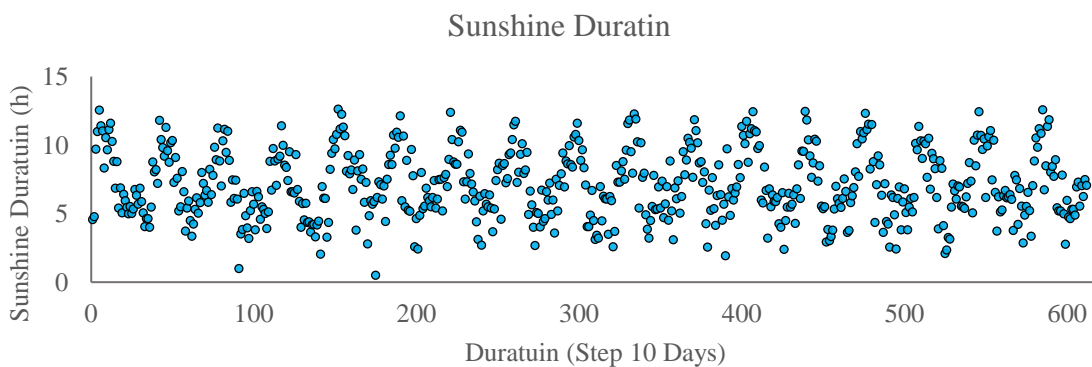
شکل ۳,۳. سری زمانی دمای متوسط



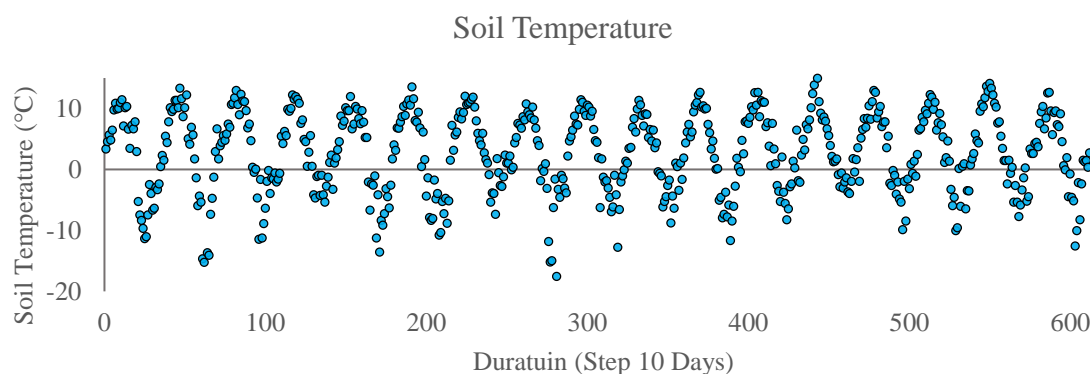
شکل ۴,۳. سری زمانی تبخیر در محل ایستگاه



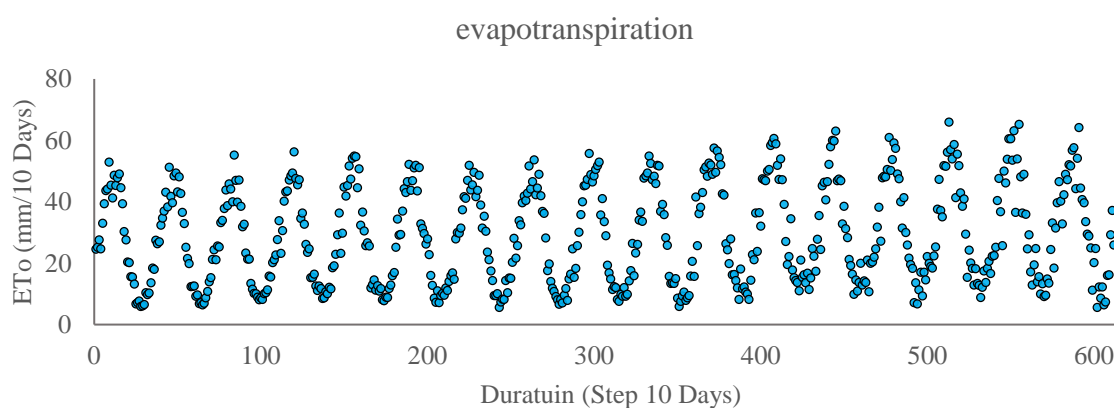
شکل ۵,۳. سری زمانی حداکثر جهت باد



شکل ۶,۳. سری زمانی تعداد ساعت آفتابی



شکل ۳، ۷. سری زمانی دمای خاک



شکل ۳، ۸. سری زمانی تبخیر و تعرق مرجع

شکل ۳. سری زمانی پارامترهای نهایی برای مدل (ایستگاه هواشناسی اردبیل)

در بخش اول مدل‌ها با استفاده از ترکیب تصادفی داده‌های ۴ ایستگاه که یک سری زمانی با طول دوره ۱۶۰۰ واحدی را تشکیل می‌داد، آموزش و یا ساخته شد. سپس ارزیابی نهایی مدل‌ها، با استفاده از داده‌های ایستگاه نمین انجام گرفت. در انتهای بررسی دقت مدل‌ها در هر بخش، ترکیب داده‌های ورودی (جدول ۷) در دقت تخمین تابع هدف براساس آماره R^2 بررسی شد تا تاثیر وجود پارامترهای مختلف بر دقت مدل‌ها نیز مشخص گردد؛ ترتیب حذف داده‌ها براساس همبستگی کمتر به بیش‌تر در داده‌های ورودی با تابع هدف بود.

جدول ۷. ضرایب ثابت معادله رگرسیون خطی چند گانه

ترکیب پارامترها	شماره ترکیب
tm, tmax, tsoill_m, tmin, dd_max, sshn, evt	Comb ۱
tm, tmax, tsoill_m, tmin, dd_max, sshn	Comb ۲
tm, tmax, tsoill_m, tmin, dd_max	Comb ۳
tm, tmax, tsoill_m, tmin	Comb ۴
tm, tmax, tsoill_m	Comb ۵
tm, tmax	Comb ۶
tm	Comb ۷

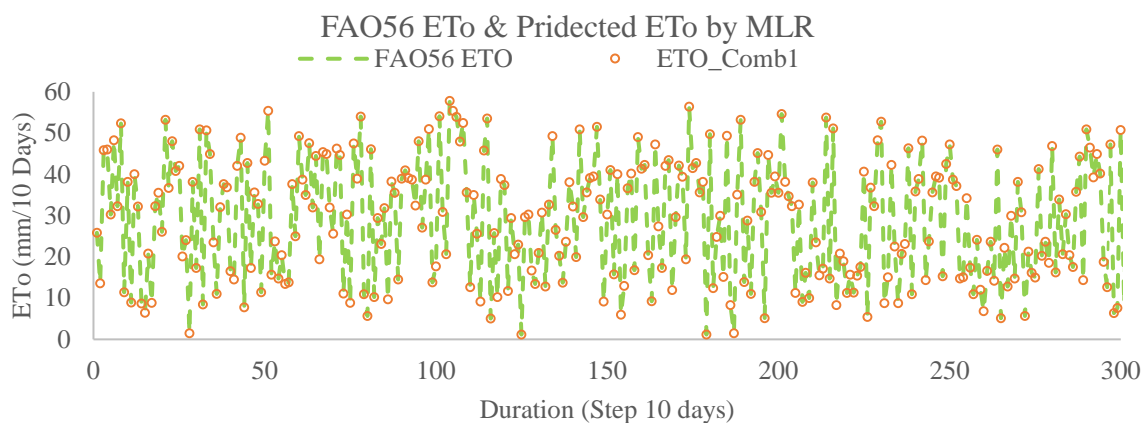
برای محاسبه ضرایب و در نتیجه تشکیل معادله رگرسیون خطی چندگانه از نرم‌افزارهای Excel و Minitab استفاده شد که ضرایب پارامترهای ورودی به شرح جدول (۸) و نتیجه آن به مطابق شکل (۴) به دست آمد؛ با این توضیح که برای امکان مقایسه بهتر، تنها ۳۰۰



داده اول در شکل (۴) ترسیم شد.

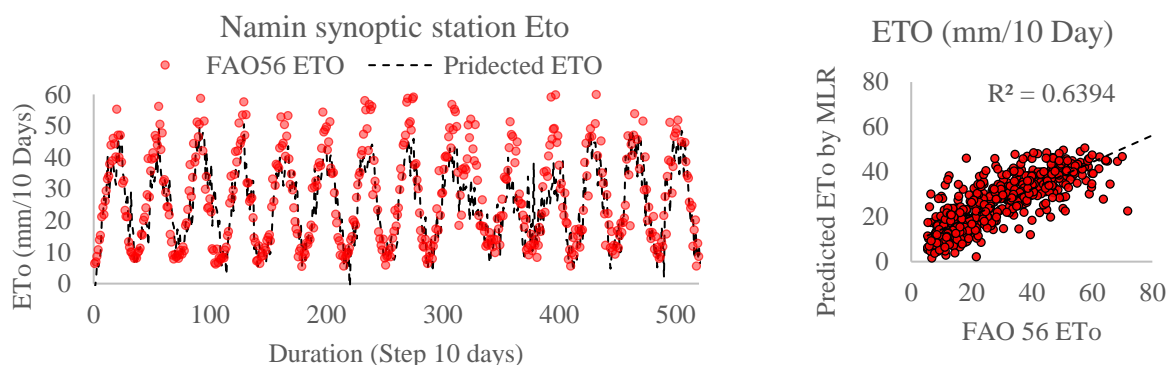
جدول ۸. ضرایب ثابت معادله رگرسیون خطی چند گانه

Upper ۹۵/۰%	Lower ۹۵/۰%	Upper ۹۵%	Lower ۹۵%	P-value	t Stat	خطای استاندارد	ضرایب	پارامترها
							.	Intercept
-۵/۷۰	-۲۸/۸۸	-۵/۷۰	-۲۸/۸۸	۰/۰۰	-۲/۹۳	۵/۹۱	-۱۷/۲۹	tm
۱۵/۹۶	۴/۳۷	۱۵/۹۶	۴/۳۷	۰/۰۰	۳/۴۴	۲/۹۵	۱۰/۱۶	tmax
۰/۵۳	-۰/۳۰	۰/۵۳	-۰/۳۰	۰/۶۰	۰/۵۲	۰/۲۱	۰/۱۱	tsoil_m
۱۳/۷۳	۲/۱۶	۱۳/۷۳	۲/۱۶	۰/۰۱	۲/۶۹	۲/۹۵	۷/۹۴	tmin
۰/۰۰	-۰/۰۱	۰/۰۰	-۰/۰۱	۰/۴۴	-۰/۷۷	۰/۰۰	۰/۰۰	dd_max
۰/۵۰	۰/۳۰	۰/۵۰	۰/۳۰	۰/۰۰	۷/۸۶	۰/۰۵	۰/۴۰	sshn
۱/۵۸	۱/۰۷	۱/۵۸	۱/۰۷	۰/۰۰	۱۰/۲۲	۰/۱۳	۱/۳۲	evt



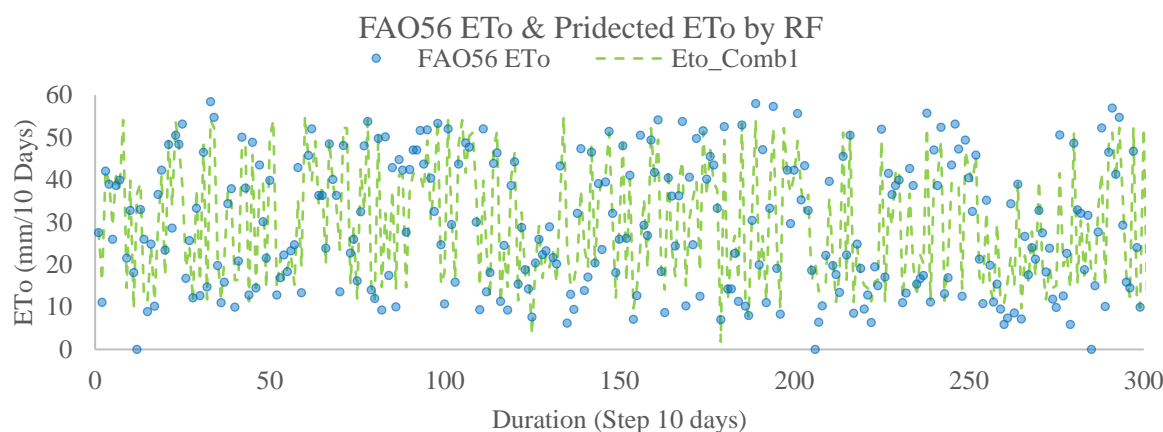
شکل ۴. تبخیر و تعرق مرجع و برآورد شده برای سری زمانی ترکیبی در روش MLR

برای بررسی دقت مدل از آماره‌های ارزیابی به شرح جدول (۹) استفاده شد و پس از نهایی شدن ساختار مدل، از آن برای برآورد تبخیر و تعرق مرجع ایستگاه نمین استفاده شد که نتایج آن در شکل (۵) و جدول (۹) ارائه شده است.



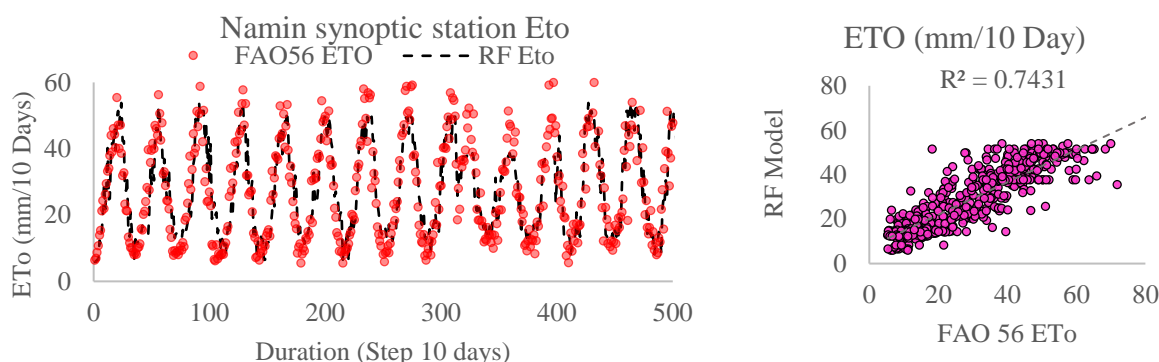
شکل ۵. تبخیر و تعرق مرجع و برآورد شده برای ایستگاه نمین در روش MLR

در مرحله بعد برای ساخت مدل جنگل تصادفی از نرم‌افزار Matlab استفاده شد. ساختار نهایی مدل شامل ۵ درخت تصمیم‌گیری با ۵۰ شاخه بود که از تفکیک ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای تست برای سری زمانی ترکیبی و تصادفی استفاده شد. نتیجه آن به مطابق شکل (۶) به دست آمد؛ با این توضیح که برای امکان مقایسه بهتر، تنها ۳۰۰ داده اول ترسیم شده است.



شکل ۶. تبخیر و تعرق مرجع و برآورد شده برای سری زمانی ترکیبی در روش RF

از مدل ساخته شده برای برآورد تبخیر و تعرق مرجع ایستگاه نمین استفاده شد که نتایج آن در شکل (۷) و جدول (۹) ارائه شده است.



شکل ۷. تبخیر و تعرق مرجع و برآورد شده برای ایستگاه نمین در روش RF

برای بررسی دقت مدل از آماره‌های ارزیابی به شرح جدول (۹) استفاده شد.

جدول ۹. دقت مدل‌ها در ارزیابی برآورد تبخیر و تعرق مرجع

داده‌ها	مدل	RMSE	NSE	R ²
داده‌های ساخت مدل	MLR	۸/۶۹	۰/۷۱۵	۰/۷۲
	RF	۷/۲۱	۰/۸۰۳	۰/۸۰
ایستگاه نمین	MLR	۱۰/۸۸۱	۰/۶۰۳	۰/۶۴
	RF	۸/۲۰	۰/۷۴۳	۰/۷۴

مطابق نتایج به دست آمده در جدول (۹) مقادیر حاصل از آماره R^2 و NSE و RMSE نشان می‌دهد که مدل RF از دقت بالاتری در برآورد تبخیر و تعرق مرجع با داده‌های یکسان و مشترک برخوردار است. که این نتیجه مشابه مطالعات برخی محققین مانند (Feng et al., 2017; Rahimikhoob, 2016; Shiri, 2018) بود و این نتایج نشان داد که مدل‌های یادگیری ماشین در پیش‌بینی ETo موفق‌تر هستند.

در تحقیقی مشابه (Pagano et al., 2023) از جنگل تصادفی (RF) برای پیش‌بینی تبخیر و تعرق واقعی روزانه استفاده کردند که بهترین عملکرد، با توجه به مقادیر (RMSE) و (R) به ترتیب برابر ۰/۳۹ میلی‌متر در روز و ۰/۸۴، توسط روش جنگل تصادفی به دست آمده است. (Hou et al., 2023) عملکرد مدل هیبریدی RF-SVR-PSO ابتکاری را در برابر یک مدل SVR مستقل، یک مدل شبکه عصبی پس انتشار (BPNN) و یک مدل RF را جهت مقایسه با مقادیر ETC که از معادله Penman-Monteith محاسبه شده بود، استفاده کردند. با توجه به آماره‌های استفاده شده نتایج تحقیق این محققین در مقایسه با تحقیق حاضر بهتر بود که می‌تواند تأثیر مثبت ضریب گیاهی بر نتایج آن‌ها باشد، زیرا در آن تحقیق تبخیر و تعرق واقعی محاسبه شده، در حالی که در این تحقیق تبخیر و تعرق مرجع محاسبه

شده است. در تحقیقی دیگر (Song et al., 2023) بر اساس داده‌های ماهواره‌ای Landsat 8 OIL_TIRS و داده‌های تصویر سنجش از دور، بر روی رودخانه یانگ تسه برای دوره ۲۰۱۶-۲۰۲۱، همراه با داده‌های خاک و هواشناسی منطقه، از یک مدل رگرسیون خطی چندگانه (MLR) و یک مدل (ELM) برای تحلیل تبخیر و تعرق، به عنوان پارامتر هدف، استفاده کردند. در این تحقیق به مراتب نتایج رضایت‌بخش‌تری بدست آمد که از جمله عواملی که باعث برتری نتایج این محققین شد، استفاده آن‌ها از اطلاعاتی از قبیل تابش خورشیدی خالص، شار گرمای خاک، ضریب پوشش گیاهی جزئی بود که تأثیر مثبت و قابل توجهی بر ET داشتند.

(Z. Shu et al., 2022) در مطالعه‌ای، یک روش جدید برای برآورد ETO در مناطقی با داده‌های محدود را پیشنهاد کردند. آن‌ها چهار مدل تجربی با نیازهای داده‌ای متفاوت شامل آلبرشت، هارگریوز-سامانی، پرستلی-تیلور و پنمن را به کار بردند و پارامترها را با مدل پنمن-مانتیت بهینه‌سازی کردند که از دو مدل یادگیری ماشین جنگل تصادفی (RF) و رگرسیون خطی چندگانه (MLR) برای ایجاد منطقه‌بندی پارامتر مدل تجربی استفاده شد. نتایج مطالعه آن‌ها نشان داد که براساس روش منطقه‌بندی با یادگیری ماشین RF، برآورد ETO در مناطق فاقد داده می‌تواند به طور کارآمد بهبود یابد که تأیید کننده‌ی نتایج این تحقیق است. (Gharehbaghi et al., 2024) مطالعه‌ای تبخیر و تعرق پتانسیل را با استفاده از سه فرمول تجربی هارگریوز و سامانی، پرستلی و تیلور و سازمان جهانی هواشناسی (WMO) برآورد کردند، که با توجه آماره‌های مورد استفاده، نتایج این تحقیق که از دو مدل یادگیری ماشین جنگل تصادفی (RF) و رگرسیون خطی چندگانه (MLR) بدست آمده در مقایسه با مدل‌های تجربی مناسب‌تر بود البته ممکن است این اختلاف به علت تفاوت اقلیم باشد نه به علت مدل‌های استفاده شده ولی حساسیت آن‌ها به تعداد پارامترهای ورودی بیش‌تر از مدل‌های تجربی بود که نتایج این تحقیق را برای مدل‌های هوشمند (با توجه به جدول ۱۰) تأیید می‌کند. (Liu et al., 2024) نیز در مطالعه‌ای از داده‌های سنجش از دور دمای سطح زمین (LST) و از مدل یادگیری ماشین (جنگل تصادفی) برای ساخت یک مدل ETO سنجش از دور استفاده کردند. این مدل در ۵۶۷ ایستگاه هواشناسی در چین با R^2 برابر با ۰/۹۷، RMSE برابر ۰/۴۰ و MAPE برابر با ۰/۱۱ در مقایسه با FAO-PM عملکرد بسیار خوبی دارد، که نشان دهنده کارایی این مدل‌هاست.

باتوجه به نتایج جدول (۱۰) مشاهده شد به طور نسبی با حذف داده‌ها به ترتیب از ترکیب شماره یک تا هفت، مقدار آماره R^2 در مدل RF و MLR کاهش می‌یابد، که باتوجه به این موضوع، بهترین ترکیب در این مدل‌ها، ترکیب شماره یک است.

جدول ۱۰. دقت مدل‌ها برای ترکیب مختلف داده‌های ورودی

R^2		
مدل‌ها	RF	MLR
FAO۵۶ ETO	۱/۰۰۰۰	۱/۰۰۰۰
ETO_Comb۱	۰/۸۹۶۵	۰/۸۴۷۲
ETO_Comb۲	۰/۸۷۸۱	۰/۸۳۵۹
ETO_Comb۳	۰/۸۷۸۳	۰/۸۲۶۶
ETO_Comb۴	۰/۸۶۷۷	۰/۸۲۲۹
ETO_Comb۵	۰/۸۶۸۳	۰/۸۲۲۰
ETO_Comb۶	۰/۸۶۷۲	۰/۸۲۱۲
ETO_Comb۷	۰/۸۵۹۷	۰/۸۱۵۸

در خصوص نتایج این تحقیق می‌توان گفت که در هر دو مدل (RF و MLR)، شاخص RMSE دارای مقدار بالایی داشت که توصیه شده برای افزایش بیش‌تر همپوشانی مابین تابع هدف و خروجی مدل از روش‌های داده کاوی شامل عوامل مختلفی اعم از وجود روند و یا حذف داده‌های پرت است (Chatfield & Xing, 2019; Coucke & Soumali, 2017; Heckert et al., 2002; Kalteh & Hjorth, 2008; Little et al., 2013; Martí et al., 2024; Organization, 2009) استفاده کرد.

برای محاسبه آنالیز حساسیت می‌توان از مقایسه بین خطای استاندارد و مقادیر پارامترهای برآورد شده در جدول (۸) استفاده کرد (Rasoulzadeh & Ghoorabjiri, 2014; Rasoulzadeh & Yaghoubi, 2014)؛ در این صورت مدل به تبخیر و دمای ماکزیمم بیشترین حساسیت و به دمای خاک کمترین حساسیت را داشت. اما اشاره به این نکته هم می‌توان داشت که با توجه به در دسترس بودن بخشی از پارامترهای مورد نیاز در یک منطقه، باتوجه به نتایج جدول (۱۰)، می‌توان از ترکیب پارامترهای متفاوتی برای مدل استفاده کرد.

در مطالعات مشابه درخصوص ترکیب پارامترهای ورودی (Saremi, 2015) در مطالعه خود به بررسی میزان حساسیت ETo نسبت به پارامترهای اقلیمی با استفاده از شبکه عصبی پرداخته‌اند. نتایج نشان داد که افزایش تعداد اطلاعات در لایه ورودی لزوماً منجر به بهبود نتایج مدل‌های هوشمند نمی‌شود. در شرایط کمبود داده‌های هواشناسی دو پارامتر دمای حداکثر و سرعت باد به عنوان ورودی شبکه برآوردهای معقولی را در برداشت، (سلطانی و همکاران، ۱۳۹۸) در تحقیقی دقت معادله فائو پنمن-مانتیت برای برآورد تبخیر و تعرق مرجع را در اقلیم‌های مختلف و در شرایطی که داده‌های هواشناسی محدود باشند، بررسی کردند. آن‌ها نشان دادند حداقل داده لازم برای برآورد دقیق تبخیر و تعرق مرجع به روش فائو پنمن-مانتیت در اقلیم‌های مرطوب و گرم و نیمه خشک، داده‌های حداکثر و حداقل دما و سرعت باد محلی می‌باشد و در اقلیم بسیار خشک حداقل داده لازم، داده‌های حداکثر و حداقل دمای روزانه و سرعت باد روزانه می‌باشند.

بر اساس نتایج این تحقیق و سایرین، اگر چه برآورد ETo از طریق مدل‌های یادگیری ماشین در بسیاری از موارد می‌تواند بسیار دقیق‌تر از هر مدل تجربی باشد، اما نکته بسیار مهم این است که مدل‌های تجربی معمولاً در مقابل تغییرات جزئی داده‌ها و یا اضافه شدن داده‌ها پایدار هستند، در حالی که مدل‌های یادگیری ماشین می‌توانند به طور قابل توجهی کم‌ترین دقت پیش‌بینی‌کننده را هنگامی که یک مجموعه داده جدید به مدل اضافه می‌شود، ایجاد کنند و ناپایدار باشند (Dong et al., 2020; Hassan et al., 2017). چراکه بسیاری از مدل‌های داده‌مبنا، معمولاً یک مجموعه داده منفرد را برای آموزش و یک مجموعه داده واحد را برای آزمایش در نظر می‌گیرند، که ممکن است این موضوع باعث اشتباه در یادگیری مدل شده یا نتیجه نسبتاً خوب و دقیقی ارائه کند، زیرا دقت عملکرد فقط برای یک الگوی خاص از مجموعه داده‌های اصلی بیان می‌شود و شاید آن مجموع داده منتخب، الگوی مناسبی را برای محدوده مطالعاتی را در نظر نگیرد (Martí et al., 2014; Shiri, 2018; Shiri et al., 2014). پس در نظر گرفتن یک استراتژی اعتبارسنجی برای امکان ارزیابی عملکرد معتبر مدل، کافی است. چراکه عملکرد هر مدل بر اساس الگوهای داده و مناطق آب و هوایی است. چراکه، یک مدل همسان، عملکردهای متفاوتی را در مطالعات مختلف تحت شرایط آب و هوایی مختلف نشان می‌دهد (Fan et al., 2018; Wu & Fan, 2019)، که در این صورت استفاده از ترکیب داده‌های ورودی از نقاط مختلف منطقه مطالعاتی می‌تواند در ثبات و پایداری مدل کمک کننده باشد (Akbari et al., 2024). به همین منظور در این تحقیق یک سری زمانی ترکیبی و تصادفی بدون رعایت ترتیب زمانی و مکانی با استفاده از کل داده‌های چهار ایستگاه منتخب برای ساخت و آموزش مدل تشکیل شد که این سری زمانی علاوه بر کاهش نیاز مدل‌ها به داده‌کاوی باعث پایداری مدل در مقابل داده‌های جدید نیز می‌شود.

در بین مدل‌های مورد مطالعه در این تحقیق، مدل RF به عنوان مدل برتر نسبت به MLR بود. نکته مهم و قابل توجه دیگر این بود که مطالعه حاضر نشان داد که مدل‌های یادگیری ترکیبی، مانند جنگل تصادفی و درخت تصادفی می‌تواند یک مدل یادگیری ماشین مطمئن‌تر در نظر گرفتن دقت و ثبات، برای پیش‌بینی ETo با استفاده از مجموعه داده‌های آب و هوایی محدود در مناطق مختلف آب و هوایی دشت اردبیل باشد و ممکن است برای اقلیم‌های مشابه در سراسر جهان نیز قابلیت استفاده داشته باشد. با این حال، تحقیقات آینده برای ارزیابی عملکرد مدل‌های RF برای دوره‌های مختلف، به عنوان مثال، مقیاس‌های ساعتی یا ماهانه یا سایر مناطق مشابه جهان با آب و هوای متفاوت مورد نیاز است.

نتیجه‌گیری

در این تحقیق کارایی مدل RF برای تخمین دقیق ETo، با ترکیبی از ورودی پارامترهای آب و هوایی در محل پنج ایستگاه هواشناسی بررسی شد. منطقه مورد مطالعه حوضه دشت اردبیل و دوره آماری مورد بررسی بین سال‌های ۱۳۸۵-۱۴۰۱ با بازه‌ی ده روزه در نظر گرفته شد. مطابق نتایج به‌دست آمده از بین متغیرهای هواشناسی در دسترس و قابل دریافت از سازمان هواشناسی، هفت ترکیب از پارامترها، می‌تواند دقت رضایت‌بخشی برای تخمین ETo ارائه دهد، اما به سبب در نظر گرفتن الگوهای متفاوت آب و هوایی در سطح منطقه مورد مطالعه و جلوگیری از عدم‌ثبات مدل‌ها در صورت تغییر داده‌ها، بهترین ترکیب پیشنهادی، در هر دو مدل مورد مطالعه ترکیب پارامترهای دمای خاک، دمای ماکزیمم، دمای متوسط، دمای مینیمم، جهت سرعت باد، ساعت تابش و تبخیر بود، علیرغم اینکه با وجود بهترین نتیجه در استفاده از هر هفت پارامتر در ورودی مدل‌ها، با در نظر گرفتن ضرایب ثابت به دست آمده در مدل MLR و آنالیز حساسیت می‌توان نتیجه گرفت پارامترهای دمایی هوا از درجه اهمیت بیش‌تری در برآورد تبخیر و تعرق مرجع با روش پنمن-مانتیت و مدل‌های مورد استفاده دارد. همچنین در بین مدل‌های مورد استفاده در مطالعه، مدل RF دقت پیش‌بینی بالاتری را نسبت به MLR نشان داد (مقدار بالاتر R^2 ، NSE و مقدار کم‌تر RMSE)، اما با در نظر گرفتن مقدار بالای RMSE در هر دو مدل، برای افزایش تطابق بهتر خروجی مدل‌ها و ETo،



استفاده از داده‌های پارامترهای ورودی پیشنهاد شد. در نهایت نتایج این مطالعه برای محققانی که عملکرد مدل‌های مختلف برای محاسبه تبخیر و تعرق مرجع را شبیه‌سازی یا ارزیابی می‌کنند مفید خواهد بود. همچنین، برای سیاست‌گذاران و دست‌اندرکاران برای مدیریت منابع آب کشاورزی در دشت اردبیل که با مشکل محدودیت داده‌های هواشناسی مواجه هستند و یا نیاز به بررسی تبخیر و تعرق مرجع تحت تغییرات اقلیمی را دارند، قابل توسعه و استفاده خواهد بود.

تشکر و قدردانی

نویسندگان از گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی که موجبات تسهیل انجام این تحقیق را فراهم نموده‌اند، تشکر و قدردانی می‌نمایند.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

- جعفرزاده، جعفر و جعفر، بهنام. (۱۳۹۴). ارائه راهکارهای پیشنهادی جهت بهره‌برداری بهینه از منابع آب موجود دشت اردبیل اولین کنگره بین‌المللی زمین، فضا و انرژی پاک.
- سلطانی، اسعد، میرلطیفی، سیدمجید و دهقانی سانجی، حسین. (۱۳۹۸). برآورد تبخیر-تعرق مرجع با استفاده از داده‌های محدود هواشناسی در شرایط اقلیمی مختلف. آب و خاک، ۲۶.
- سلطانی، نرگس و قمرنیا، هوشنگ. (۱۳۹۷). ررسی کارآیی روش‌های تجربی برآورد تبخیر-تعرق مرجع (بر پایه تشتک تبخیر) در اقلیم‌های مختلف (مطالعه موردی ایران). تحقیقات منابع آب ایران، ۱۴(۴)، ۱۷۰-۱۸۳.
- شیرمحمدی، زهرا، صابری، سید فرهاد و کوهی، منصوره. (۱۳۹۹). ارزیابی روش‌های مختلف محاسبه تبخیر و تعرق پتانسیل گیاه مرجع در مقیاس سالانه در شمال شرق ایران پژوهشنامه مدیریت حوزه آبخیز، ۱۱(۲۲)، ۲۰۹-۱۹۹.
- طباطبائی، سید محمدرضا، سهرابی، تیمور، محب‌الحجه، علیرضا و هورفر، عبدالحسین. (۱۴۰۱). بررسی و توسعه معادله تبخیر-تعرق هارگریوز - سامانی برای شرایط رطوبتی مختلف. تحقیقات آب و خاک ایران، ۵۳(۱۲)، ۲۸۶۳.
- عبدزادگوهری، علی، نیک اختر، امیر، ابراهیمی پاک، نیازعلی و تافته، آرش. (۱۴۰۲). استفاده از سامانه نیاز آب در تعیین مقدار آب مصرفی سویا بر اساس حل معکوس توابع عملکرد در شرایط مختلف آبیاری. پژوهش آب در کشاورزی، ۳۷(۲)، ۱۵۹-۱۶۹.
- عزیزی مبصر، جوانشیر. (۱۳۹۲). مقایسه روش‌های مختلف تبخیر و تعرق پتانسیل در دشت اردبیل. دومین کنگره ملی کشاورزی ارگانیک و مرسوم. عزیزی مبصر، جوانشیر و فرجی عموقین، عرفان. (۱۴۰۳). تعیین نیاز آبی الگوی کشت غالب دشت اردبیل بر اساس آمارهای به‌روز هواشناسی. مدل سازی و مدیریت آب و خاک، ۴(۱)، ۲۸۵-۲۹۸.
- فخار، محدثه السادات و کاویانی، عباس. (۱۴۰۰). مقایسه مفاهیم تبخیر-تعرق پتانسیل و مرجع با استفاده از داده‌های لایسیمیتری در استان قزوین. محیط زیست و مهندسی آب، ۷(۴)، ۶۶۸-۶۸۲.
- کریمی، سیده رقیه، نصرالهی، علی حیدر و ایرانشاهی، معین. (۱۴۰۳). بررسی اثرات تغییر اقلیم بر تبخیر و تعرق مرجع بر اساس سناریوهای گزارش ششم تغییر اقلیم (مطالعه موردی: ایستگاه خرم‌آباد). تحقیقات آب و خاک ایران، ۵۴(۱۱)، ۱۷۷۷-۱۷۷۷.
- محمدی، اکبر، نجفی مود، محمدحسین، خاشعی سیوکی، عباس و شهیدی، علی. (۱۳۹۹). تعیین ضرایب گیاهی گل نرگس به روش لایسیمیتری در دشت بیرجند. پژوهش آب در کشاورزی، ۳۴(۴)، ۵۹۱-۶۰۱.
- مزیدی، مریم، حسام، موسی، قربانی، خلیل و کمکی، چوقی بایرام. (۱۴۰۲). ارزیابی تخمین تنش آبی پنبه با استفاده از تصاویر چند طیفی ماهواره‌ای، مبتنی بر مدل درخت M5. پژوهش آب در کشاورزی، ۳۷(۴)، ۳۸۵-۴۰۰.

REFERENCES

- Abdzad Gohari, A., Nik Akhtar, A., Ebrahimipak, N., & Tafteh, A. (2023). Using NIAZAB System to Determine Soybean Water Use Based on the Inverse Solution of the Production Functions under Different Irrigation Conditions. *Journal of Water Research in Agriculture*, 37(2), 159-169. <http://10.22092/jwra.2023.361550.978>. (In Persian).
- Abed, M., Imteaz, M. A., Ahmed, A. N., & Huang, Y. F. (2022). Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms. *Sci Rep*, 12(1), 13132. <http://10.1038/s41598-022-17263-3>.

- Adnan, R. M., Liang, Z., Heddarn, S., Zounemat-Kermani, M., Kisi, O., & Li, B. (2020). Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *Journal of Hydrology*, 586, 124371. <https://doi.org/10.1016/j.jhydrol.2019.124371>.
- Aghelpour, P., Bahrami-Pichaghchi, H., & Karimpour, F. (2022). Estimating daily rice crop evapotranspiration in limited climatic data and utilizing the soft computing algorithms MLP, RBF, GRNN, and GMDH. *Complexity*, 2022(1), 4534822. <https://doi.org/10.1155/2022/4534822>.
- Akbari Majd, A., Azizi Mobaser, J., Rasoulzadeh, A., Hasanpour Kashani, M., & Kisi, O. (2024). Enhancing the accuracy of metaheuristic neural networks in predicting underground water levels using meteorological data and remote sensing: A case study of Ardabil Plain, Iran. *Ain Shams Engineering Journal*, 103061. <https://doi.org/10.1016/j.asej.2024.103061>.
- Alizamir, M., Kisi, O., Muhammad Adnan, R., & Kuriqi, A. (2020). Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophysica*, 68, 1113-1126. <https://doi.org/10.1007/s11600-020-00446-9>.
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56 (Vol. 300).
- Amani, S., & Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. *Agricultural water management*, 284, 108324. <https://doi.org/10.1016/j.agwat.2023.108324>.
- Ao, Y., Li, H., Zhu, L., Ali, S., & Zhongguo, Y. (2018). The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *Journal of Petroleum Science and Engineering*, 174. <https://10.1016/j.petrol.2018.11.067>.
- Aryalekshmi, B., Biradar, R. C., Chandrasekar, K., & Ahamed, J. M. (2021). Analysis of various surface energy balance models for evapotranspiration estimation using satellite data. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3), 1119-1126. <https://doi.org/10.1016/j.ejrs.2021.11.007>.
- Azizi Mobaser, J. (2013). *Comparison of different potential evapotranspiration methods in Ardabil plain* The Second National Congress on Organic and Conventional Agriculture, <https://civilica.com/doc/271805>. <https://civilica.com/doc/271805>. (In Persian).
- Azizi Mobaser, J., & Faraji Amogein, E. (2024). Determining the water requirement of the crop pattern of Ardabil Plain based on up-to-date meteorological statistics. *Water and Soil Management and Modelling*, 4(1), 285-298. <https://10.22098/mmws.2023.12358.1231>. (In Persian).
- Barideh, R., Veysi, S., Ebrahimipak, N., & Davatgar, N. (2022). The challenge of reference evapotranspiration between the WaPOR data set and geostatistical methods. *Irrigation and Drainage*, 71(5), 1268-1279. <https://doi.org/10.1002/ird.2738>.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5-32. <https://10.1023/A:1010933404324>.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R*. Chapman and hall/CRC.
- Chia, M. Y., Huang, Y. F., & Koo, C. H. (2022). Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agricultural water management*, 261, 107343. <https://doi.org/10.1016/j.agwat.2021.107343>.
- Chia, M. Y., Huang, Y. F., Koo, C. H., Ng, J. L., Ahmed, A. N., & El-Shafie, A. (2022). Long-term forecasting of monthly mean reference evapotranspiration using deep neural network: A comparison of training strategies and approaches. *Applied Soft Computing*, 126, 109221. <https://doi.org/10.1016/j.asoc.2022.109221>.
- Coucke, W., & Soumali, M. (2017). Demystifying EQA statistics and reports. *Biochemia Medica*, 27, 37-48. <https://10.11613/BM.2017.006>.
- Da Silva, H. J., Santos, M., Cabral Júnior, J., & Spyrides, M. H. (2016). Modeling of reference evapotranspiration by multiple linear regression. *Journal of Hyperspectral Remote Sensing*, 6. <https://10.5935/2237-2202.20160005>.
- DehghaniSanij, H., Yamamoto, T., & Rasiah, V. (2004). Assessment of evapotranspiration estimation models for use in semi-arid environments. *Agricultural water management*, 64(2), 91-106. [https://doi.org/10.1016/S0378-3774\(03\)00200-2](https://doi.org/10.1016/S0378-3774(03)00200-2).
- Dimitriadou, S., & Nikolakopoulos, K. G. (2022). Multiple Linear Regression Models with Limited Data for the Prediction of Reference Evapotranspiration of the Peloponnese, Greece. *Hydrology*, 9(7), 124. <https://doi.org/10.3390/hydrology9070124>.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241-258. <https://10.1007/s11704-019-8208-z>.



- Drapear, N. (1998). *Applied regression analysis*. McGraw-Hill. Inc.
- Duhan, D., Singh, M. C., Singh, D., Satpute, S., Singh, S., & Prasad, V. (2023). Modeling reference evapotranspiration using machine learning and remote sensing techniques for semi-arid subtropical climate of Indian Punjab. *Journal of Water and Climate Change*, 14(7), 2227-2243. <https://doi.org/10.2166/wcc.2023.003>.
- El-Kenawy, E.-S. M., Zerouali, B., Bailek, N., Bouchouich, K., Hassan, M. A., Almorox, J.,...Ibrahim, A. (2022). Improved weighted ensemble learning for predicting the daily reference evapotranspiration under the semi-arid climate conditions. *Environmental Science and Pollution Research*, 29(54), 81279-81299. <https://doi.org/10.1007/s11356-022-21410-8>.
- Elbeltagi, A., Raza, A., Hu, Y., Al-Ansari, N., Kushwaha, N., Srivastava, A.,...Zubair, M. (2022). Data intelligence and hybrid metaheuristic algorithms-based estimation of reference evapotranspiration. *Applied Water Science*, 12(7), 152. <https://10.1007/s13201-022-01667-7>.
- Fakhar, M. S., & Kaviani, A. (2021). Comparison of the Concepts of Potential and Reference Evapotranspiration using Lysimetric Data in Qazvin Province. *Environment and Water Engineering*, 7(4), 668-682. <https://10.22034/jewe.2021.279059.1535>. (In Persian).
- Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural water management*, 225, 105758. <https://doi.org/10.1016/j.agwat.2019.105758>.
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Xiukang, W.,...Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225-241. <https://10.1016/j.agrformet.2018.08.019>.
- Fattahi Dolatabadi, K., babazadeh, H., najafi, P., & sedghi, h. (2018). A Model for Irrigation Scheduling Using the Difference between Air and Leaf Temperature of Corn. *Journal of Water Research in Agriculture*, 32(2), 305-320. 10.22092/jwra.2018.116972.
- Feng, Y., Peng, Y., Cui, N., Gong, D., & Zhang, K. (2017). Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Computers and Electronics in Agriculture*, 136, 71-78. <https://doi.org/10.1016/j.compag.2017.01.027>.
- Ferreira, L. B., da Cunha, F. F., & Fernandes Filho, E. I. (2022). Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil. *Agricultural water management*, 259, 107281. <https://doi.org/10.1016/j.agwat.2021.107281>.
- Gavili, S., Sanikhani, H., Kisi, O., & Mahmoudi, M. H. (2018). Evaluation of several soft computing methods in monthly evapotranspiration modelling. *Meteorological Applications*, 25(1), 128-138. <https://10.1002/met.1676>.
- Gharehbaghi, A., Afaridegan, E., Kaya, B., & Adhami, M. (2024). Calibration and evaluation of various reference evapotranspiration estimation methods in a humid subtropical climate: A case study in Samsun Province, Türkiye. *Physics and Chemistry of the Earth Parts A/B/C*, 136. <https://10.1016/j.pce.2024.103734>.
- Glantz, M. H. (1990). On the interactions between climate and society. *Population and Development Review*, 16, 179-200. <https://10.1007/s11707-008-0045-6>.
- Gocić, M., Motamedi, S., Shamshirband, S., Petković, D., Ch, S., Hashim, R., & Arif, M. (2015). Soft computing approaches for forecasting reference evapotranspiration. *Computers and Electronics in Agriculture*, 113, 164-173. <https://doi.org/10.1016/j.compag.2015.02.010>.
- Gokcekus, H., Kassem, Y., & Woyea, L. T. (2023). A Prediction of Rainfall of Haifa Using MLR and ARIMA Models. *International Journal of Engineering and Applied Physics*, 3(1), 612-624. <https://ijeap.org/ijeap/article/view/109>.
- Hassan, M. A., Khalil, A., Kaseb, S., & Kassem, M. A. (2017). Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Applied Energy*, 203, 897-916. <https://doi.org/10.1016/j.apenergy.2017.06.104>.
- Heckert, N., Filliben, J., Croarkin, C., Hembree, B., Guthrie, W., Tobias, P., & Prinz, J. (2002). *Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods*. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD.
- Hou, W., Yin, G., Gu, J., & Ma, N. (2023). Estimation of Spring Maize Evapotranspiration in Semi-Arid Regions of Northeast China Using Machine Learning: An Improved SVR Model Based on PSO and RF Algorithms. *Water*, 15(8), 1503. <https://doi.org/10.3390/w15081503>.

- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- Jafarzadeh, J., & Jafarzadeh, B. (2017). *Providing suggested solutions for optimal utilization of existing water resources in the Ardabil Plain* The1st International Congress on Earth, Space & Clean Energy, <https://civilica.com/doc/456620/>. (In Persian).
- Jensen, M. E. (1968). *Water consumption by agricultural plants (Chapter 1)*. <https://eprints.nwisrl.ars.usda.gov/id/eprint/742>
- Kalluri, S., Gilruth, P., & Bergman, R. (2003). The potential of remote sensing data for decision makers at the state, local and tribal level: experiences from NASA's Synergy program. *Environmental Science & Policy*, 6(6), 487-500. <https://doi.org/10.1016/j.envsci.2003.08.002>.
- Kalteh, A. M., & Hjorth, P. (2009). Imputation of missing values in precipitation-runoff process database. *Hydrology Research*, 40. <https://10.2166/nh.2009.001>.
- Karimi, S. R., Nasrolahi, A., & Iranshahi, M. (2024). Investigating the effects of climate change on reference evapotranspiration based on the SSP scenarios. *Iranian Journal of Soil and Water Research*, 54(11), 1759-1777. <https://10.22059/ijswr.2023.358697.669493>. (In Persian).
- Kaya, Y. Z., Zelenakova, M., Üneş, F., Demirci, M., Hlavata, H., & Mesaros, P. (2021). Estimation of daily evapotranspiration in Košice City (Slovakia) using several soft computing techniques. *Theoretical and Applied Climatology*, 144, 287-298. <https://doi.org/10.1007/s00704-021-03525-z>.
- Koç, D. L., & Erkan Can, M. (2023). Reference evapotranspiration estimate with missing climatic data and multiple linear regression models. *PeerJ*, 11. <https://10.7717/peerj.15252>.
- Kord, M., & Asghari-moghaddam, A. (2014). Evaluation of drinking water quality of Ardabil plain aquifer by cokriging and fuzzy logic. *Journal of Water and Soil Conservation*, 21(5), 225-240. [https://20.1001.1.23222069.1393.21.5.12.5.\(2322-2069\)](https://20.1001.1.23222069.1393.21.5.12.5.(2322-2069)).
- Krzywinski, M., & Altman, N. (2015). Multiple linear regression: when multiple variables are associated with a response, the interpretation of a prediction equation is seldom simple [Report]. *Nature Methods*, 12, 1103+. <https://doi.org/10.1038/nmeth.3665>.
- Kuriqi, A., Pinheiro, A. N., Sordo-Ward, A., Bejarano, M. D., & Garrote, L. (2021). Ecological impacts of run-of-river hydropower plants—Current status and future prospects on the brink of energy transition. *Renewable and Sustainable Energy Reviews*, 142, 110833. <https://doi.org/10.1016/j.rser.2021.110833>.
- Li, Z.-L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B.,...Zhang, X. (2009). A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors*, 9(05), 3801-3853. <https://doi.org/10.3390/s90503801>.
- Little, T., Thompson, B., Coxe, S., Woods, C., von Eye, A., Buskirk, T.,...Wang, L. (2013). The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis. <https://10.1093/oxfordhb/9780199934898.001.0001>.
- Liu, D., Wang, Z., Wang, L., Chen, J., Li, C., & Shi, Y. (2024). Improved remote sensing reference evapotranspiration estimation using simple satellite data and machine learning. *Science of The Total Environment*, 947, 174480. <https://doi.org/10.1016/j.scitotenv.2024.174480>.
- Malik, A., Saggi, M. K., Rehman, S., Sajjad, H., Inyurt, S., Bhatia, A. S.,...Yaseen, Z. M. (2022). Deep learning versus gradient boosting machine for pan evaporation prediction. *Engineering Applications of Computational Fluid Mechanics*, 16(1), 570-587. <https://>
- 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, 81-90. <https://doi.org/10.1016/j.agwat.2014.10.028>.
- Martí, P., López-Urrea, R., Mancha, L. A., González-Altozano, P., & Román, A. (2024). Seasonal assessment of the grass reference evapotranspiration estimation from limited inputs using different calibrating time windows and lysimeter benchmarks. *Agricultural water management*, 300, 108903. <https://doi.org/10.1016/j.agwat.2024.108903>.
- Mazidi, M., Hesam, M., Ghorbani, K., & Komaki, C. B. (2024). Evaluation of Cotton Water Stress Estimation Using Multispectral Satellite Images Based on M5 Tree Model. *Journal of Water Research in Agriculture*, 37(4), 385-400. <https://10.22092/jwra.2023.363398.1005>. (In Persian).
- Mohammadi, A., NajafiMood, M. H., khashei, A., & Shahidi, A. (2021). Determination of Crop Coefficient of Narcissus Flower by Lysimeter method in Birjand Plain. *Journal of Water Research in Agriculture*, 34(4), 591-601. <https://10.22092/jwra.2021.123654>. (In Persian).
- Mokari, E., DuBois, D., Samani, Z., Mohebzadeh, H., & Djaman, K. (2022). Estimation of daily reference evapotranspiration with limited climatic data using machine learning approaches across different climate



- zones in New Mexico. *Theoretical and Applied Climatology*, 147, 575-587. <https://doi.org/10.1007/s00704-021-03855-y>.
- Monteiro, A. F. M., Martins, F. B., Torres, R. R., de Almeida, V. H. M., Abreu, M. C., & Mattos, E. V. (2021). Intercomparison and uncertainty assessment of methods for estimating evapotranspiration using a high-resolution gridded weather dataset over Brazil. *Theoretical and Applied Climatology*, 146(1), 583-597. <https://doi.org/10.1007/s00704-021-03747-1>.
- Mostafa, R. R., Kisi, O., Adnan, R. M., Sadeghifar, T., & Kuriqi, A. (2023). Modeling Potential Evapotranspiration by Improved Machine Learning Methods Using Limited Climatic Data. *Water*, 15(3), 486. <https://doi.org/10.3390/w15030486>.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282-290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Nolan, B. T., Fienen, M. N., & Lorenz, D. L. (2015). A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology*, 531, 902-911. <https://doi.org/10.1016/j.jhydrol.2015.10.025>.
- Organization, W. M. (2008). *Manual on Low-flow Estimation and Prediction*. World Meteorological Organization. <https://books.google.com/books?id=A5YiQwAACAAJ>
- Pagano, A., Amato, F., Ippolito, M., De Caro, D., Croce, D., Motisi, A.,...Tinnirello, I. (2023). Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation. *Ecological Informatics*, 76, 102133. <https://doi.org/10.1016/j.ecoinf.2023.102133>.
- Piña-Monarez, M. R., & Ortiz-Yañez, J. F. (2015). Weibull and lognormal Taguchi analysis using multiple linear regression. *Reliability Engineering & System Safety*, 144, 244-253. <https://doi.org/10.1016/j.ress.2015.08.004>.
- Rahimikhoob, A. (2016). Comparison of M5 Model Tree and Artificial Neural Network's Methodologies in Modelling Daily Reference Evapotranspiration from NOAA Satellite Images. *Water resources management*, 30(9), 3063-3075. <https://doi.org/10.1007/s11269-016-1331-9>.
- Rasoulzadeh, A., & Ghoorabjiri, M. (2014). Comparing hydraulic properties of different forest floors. *Hydrological Processes*, 28. <https://doi.org/10.1002/hyp.10006>.
- Rasoulzadeh, A., & Yaghoubi, A. (2014). Inverse modeling approach for determining soil hydraulic properties as affected by application of cattle manure. *International Journal of Agricultural and Biological Engineering*, 7, 27-35. <https://doi.org/10.3965/j.ijabe.20140702.004>.
- Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J., & Schaphoff, S. (2008). Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, 44(9). <https://doi.org/10.1029/2007WR006331>.
- Ruiz-Álvarez, M., Gomariz-Castillo, F., & Alonso-Sarría, F. (2021). Evapotranspiration Response to Climate Change in Semi-Arid Areas: Using Random Forest as Multi-Model Ensemble Method. *Water*, 13(2), 222. <https://doi.org/10.3390/w13020222>.
- Saremi, M. (2015). Determination of Effective Parameters in Estimating Reference Crop Evapotranspiration Using Artificial Neural Networks (Case study: Lorestan province). *Iranian Journal of Irrigation & Drainage*, 9(4), 614-623, Article (In Persian). https://idj.iaid.ir/article_55085.html?lang=en. (2008-7942).
- Shamshirband, S., Amirmojahedi, M., Gocić, M., Akib, S., Petković, D., Piri, J., & Trajkovic, S. (2016). Estimation of reference evapotranspiration using neural networks and cuckoo search algorithm. *Journal of Irrigation and Drainage Engineering*, 142(2), 04015044. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000949](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000949).
- Sharma, G., Singh, A., & Jain, S. (2022a). DeepEvap: Deep reinforcement learning based ensemble approach for estimating reference evapotranspiration. *Applied Soft Computing*, 125, 109113. <https://doi.org/10.1016/j.asoc.2022.109113>.
- Sharma, G., Singh, A., & Jain, S. (2022b). A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data. *Neural Computing and Applications*, 1-20. <https://doi.org/10.1007/s00521-021-06661-9>.
- Shiri, J. (2018). Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. *Journal of Hydrology*, 561, 737-750. <https://doi.org/10.1016/j.jhydrol.2018.04.042>.
- Shiri, J., Sadraddini, A., Nazemi, A., Kisi, O., Landaras, G., fakheri fard, A., & Marti, P. (2014). Generalizability of Gene Expression Programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *Journal of Hydrology*, 508, 1-11.

<https://10.1016/j.jhydrol.2013.10.034>.

- Shu, Z., Zhou, Y., Zhang, J., Jin, J., Wang, L., Cui, N.,...Chen, X. (2022). Parameter regionalization based on machine learning optimizes the estimation of reference evapotranspiration in data deficient area. *Sci Total Environ*, 844, 157034. <https://10.1016/j.scitotenv.2022.157034>.
- Shu, Z., Zhou, Y., Zhang, J., Jin, J., Wang, L., Cui, N.,...Chen, X. (2022). Parameter regionalization based on machine learning optimizes the estimation of reference evapotranspiration in data deficient area. *Science of The Total Environment*, 844, 157034. <https://doi.org/10.1016/j.scitotenv.2022.157034>.
- Soltani, A., Mirlatifi, S. M., & Dehghanisanij, H. (2012). Estimating Reference Evapotranspiration Using Limited Weather Data under Different Climatic Conditions. *Water and Soil*, 26(1), -. <https://10.22067/jsw.v0i0.13637>. (In Persian).
- Song, E., Zhu, X., Shao, G., Tian, L., Zhou, Y., Jiang, A., & Lu, J. (2023). Multi-Temporal Remote Sensing Inversion of Evapotranspiration in the Lower Yangtze River Based on Landsat 8 Remote Sensing Data and Analysis of Driving Factors. *Remote Sensing*, 15(11), 2887. <https://doi.org/10.3390/rs15112887>.
- Soultani, N. (2018). Evaluating efficiency of empirical estimation reference evapotranspiration (Pan based method) in different climate conditions (Case Study of Iran). *Iran-Water Resources Research*, 14(4), 170-183. <https://doi.org/10.22092/jwra.2021.123654>. (In Persian).
- Steel, R. G. D., & Torrie, J. H. (1960). *Principles and procedures of statistics*. <https://doi.org/10.2134/agronj1961.00021962005300050002x>
- TR, J., Reddy, N. V. S., & Acharya, U. D. (2023). Modeling Daily Reference Evapotranspiration from Climate Variables: Assessment of Bagging and Boosting Regression Approaches. *Water resources management*, 37(3), 1013-1032. <https://10.1007/s11269-022-03399-4>.
- Tabatabaei, S. M., Sohrabi, T., Mohebalhojeh, A., & Hoorfar, A. (2023). Evaluation and Development of Hargreaves-Samani Evapotranspiration Equation Based on Humidity Conditions. *Iranian Journal of Soil and Water Research*, 53(12), 2863-2873. <https://10.22059/ijswr.2023.351792.669405>. (In Persian).
- Veysi, S., Nouri, M., & Jabbari, A. (2024). Reference evapotranspiration estimation using reanalysis and WaPOR products in dryland Croplands. *Heliyon*, 10(4), e26531. <https://doi.org/10.1016/j.heliyon.2024.e26531>.
- Wang, L., Kisi, O., Hu, B., Bilal, M., Zounemat-Kermani, M., & Li, H. (2017). Evaporation modelling using different machine learning techniques. *International Journal of Climatology*, 37(S1), 1076-1092. <https://doi.org/10.1002/joc.5064>.
- Wu, L., & Fan, J. (2019). Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLOS ONE*, 14(5), e0217520. <https://10.1371/journal.pone.0217520>.
- Wu, M., Feng, Q., Wen, X., Deo, R. C., 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. *Hydrology Research*, 51(4), 648-665. <https://10.2166/nh.2020.012>.
- Shirmohamadi, Z., Saberli, F., & Kouhi, M. (2020). Evaluation of Different Methods of Calculating the Potential Evapotranspiration at the Annual Timescale in the Northeast of Iran [Research]. *Watershed Manage Res*, 11(22), 199-209. <https://10.52547/jwmr.11.22.199>. (In Persian).
- Zhao, L., Xia, J., Xu, C.-y., Wang, Z., Sobkowiak, L., & Long, C. (2013). Evapotranspiration estimation methods in hydrological models. *Journal of Geographical Sciences*, 23, 359-369. <https://doi.org/10.1007/s11442-013-1015-9>.
- Zhao, L., Zhao, X., Li, Y., Shi, Y., Zhou, H., Li, X.,...Xing, X. (2023). Applicability of hybrid bionic optimization models with kernel-based extreme learning machine algorithm for predicting daily reference evapotranspiration: a case study in arid and semiarid regions, China. *Environmental Science and Pollution Research*, 30(9), 22396-22412. <https://10.1007/s11356-022-23786-z>.
- Zheng, C., Jia, L., & Hu, G. (2022). Global land surface evapotranspiration monitoring by ETMonitor model driven by multi-source satellite earth observations. *Journal of Hydrology*, 613, 128444. <https://doi.org/10.1016/j.jhydrol.2022.128444>.
- Zhu, T. (2020). Analysis on the Applicability of the Random Forest. *Journal of Physics: Conference Series*, 1607, 012123. <https://10.1088/1742-6596/1607/1/012123>.