



Research Paper

Comparative Analysis of Estimating Monthly Reference Evapotranspiration Using Kernel and Tree-Based Data Mining Models Versus Empirical Methods

Sahar Javidan¹, Mohammad Taghi Sattari^{2*} and Ahmad Mehrabi³

¹M.Sc. Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

²Assoc. Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

³Expert in the Office of Water Resources Studies, Qazvin Regional Water Company, Qazvin, Iran

Article information

Received: November 28, 2022

Revised: February 26, 2022

Accepted: February 27, 2022

Keywords:

Data Mining
Evapotranspiration
Modified Makkink
Random Tree

*Corresponding author:
mtsattar@tabrizu.ac.ir



Abstract

Because direct measurement of evapotranspiration is costly and time-consuming, researchers have turned to the estimation of evapotranspiration via indirect approaches. The aim of this study is to investigate the capability of kernel-based, tree-based, bagging-based data-driven, and empirical models to estimate reference evapotranspiration. For this purpose, data related to meteorological parameters such as average temperature, hours of sunshine, maximum and minimum temperature, wind speed, precipitation, and relative humidity were collected over a period of 39 years. A correlation matrix, relief algorithm, and trial and error based on the author's own experience were used to select input scenarios. The performance of these methods was evaluated using correlation coefficient (R^2), root mean square error (RMSE), scattering index (SI), Nash Sutcliffe (NS), and Wilmot indexes (WI). Based on the results, scenario 13 includes maximum temperature and monthly time index based on the relief algorithm was selected as the best scenario, also on the other hand the random tree model with $R=0.99$, $RMSE=0.04$ mm/day, and $SI=0.01$ was selected as the superior method. Thus, the maximum temperature was defined as the efficient meteorological parameter for the reference evapotranspiration modeling.

© Authors, Published by **Environment and Water Engineering** journal. This is an open-access article distributed under the CC BY (license <http://creativecommons.org/licenses/by/4.0/>).



Introduction

Evapotranspiration is one of the important elements of the hydrological cycle and is an indispensable part of irrigation planning. As a result of the correct estimation of evapotranspiration, an effective irrigation planning can be made. In arid and semi-arid

countries like Iran, estimating the amount of evapotranspiration correctly in a period of water shortage attracts the attention of scientists. Accurate calculation of plant water requirements is a basic prerequisite for effective planning and operation of irrigation systems. There are several methods for calculating evapotranspiration,



which are divided into direct and indirect categories. It is not possible to use the direct measurement method (lysimeter) in all areas due to its cost and time. One of the most common indirect methods is to use the FAO-Penman-Monteith equation, which provides an acceptable estimate of evapotranspiration values in the absence of a lysimeter. However, this method has its practical limitations due to the need for many meteorological parameters. For this reason, researchers have turned into data mining methods. In this context, the aim of the study is to estimate the amount of evapotranspiration with the help of data-based models and to compare the performance of different methods. It is also the determination of the meteorological parameters that are effective on the evapotranspiration in the region during the estimation process.

Material and Methods

To estimate the monthly evapotranspiration rate of Qazvin synoptic station, meteorological data including maximum, minimum, and average temperature (T_{max} , T_{min} , T_{mean}), monthly precipitation (P), hours of sunshine (n), relative humidity (RH), and wind speed (U_{10}) were used in the present study for a 39-year statistical period (1361-1399). Sixteen different input combinations of meteorological data based on three known preprocessing algorithms including 1) correlation matrix, 2) relief algorithm, and 3) trial and error based on the author's own experience were considered. 80% of the data were considered for calibration and 20% for validation. The Gaussian Process Regression (GPR), Support Vector Regression (SVR), Random Forest (RF), Random Tree (RT), M5P decision tree, Reduced Error Pruning Tree (REPT), Priestley Taylor, Modified Makkink, and McGuinness Bordne experimental methods were applied. The performances of these methods were evaluated and compared with each other using correlation coefficient (R^2), root mean square error (RMSE), scattering index (SI), Nash Sutcliffe (NS), and Wilmot indexes (WI). Weka software was used in the data-based modeling process for monthly evapotranspiration estimation. Taylor and Violin diagrams were also used for the visual evaluation of the model results. A violin plot is a method of plotting numeric data and it is similar to a box plot.

Results

In this study, different input scenarios are considered for monthly evapotranspiration

estimation. In these scenarios, input combinations consisting of different meteorological parameters are considered. The aim of different scenarios was to determine the effective parameters in the estimation process. Naturally, scenarios with minimum input and more accuracy are important for implementation. Scenario 9 including average temperature, wind speed, hours of sunshine, and monthly time index, Scenario 12 including average temperature and monthly time index, and scenario 13 including maximum temperature and monthly time index produced better results in all methods compared to the other scenarios applied. The SVR, M5P, and RT methods had higher accuracy among scenarios 9, scenario 12, and scenario 13, respectively. Scenario 9 may have relatively few applications due to the rather difficulty in measuring wind speed and sundial parameters, and if it was easy to measure these two parameters and other meteorological parameters, methods such as Penman-Monteith FAO could have been used as an accurate and standard method. The purpose of using data mining techniques is that they are easy to use and employ fewer input parameters when calculating ET_0 . Therefore, scenarios 12 and 13 with high accuracy and lower parameters (including temperature and monthly time index) were considered as applicable scenarios. The modified Making method showed better performance among the three experimental methods studied (Fig 1).

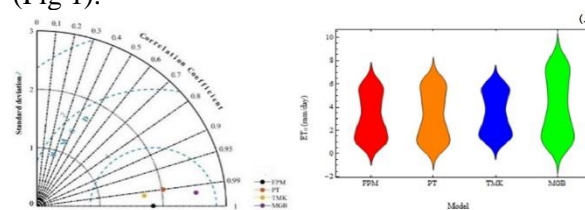


Fig. 1 Taylor diagram and violin diagram experimental

Overall, according to the results provided in the tables and figures, it was concluded that T_{max} had the greatest effect on estimating monthly reference evapotranspiration, and the performance of the random tree model as a tree-based method was better compared to that of core-based and experimental methods. The McGuinness-Bordne method resulted in the least accuracy among the experimental methods investigated. Not only are not experimental methods accurate enough, but they are also time-consuming and difficult to calculate when there are a large number of parameters. On the other

hand, data-based methods provide more acceptable results in the case of a very small number of parameters, and this has increased the popularity of data-based techniques. Moreover, with the increase of measurement facilities and having a long period of meteorological data in the future, it is predicted that these data-based methods will attract more attention from national and international researchers. When the RT method was considered as the basic algorithm of the classification method, the performance of the model improved in scenarios 9 and 12.

Conclusions

Although the direct measurement of the evapotranspiration amount is accurate, it is a difficult and costly process. In this case, it is more advantageous and easier to estimate the evapotranspiration amount relatively accurately with different models using meteorological data. Among the 16 scenarios studied, scenario 9 (including average temperature, wind speed, hours of sunshine, and monthly time index), scenario 12 (including average temperature and monthly time index), and scenario 13 (including maximum temperature and monthly time index), that were selected based on Relief algorithm,

performed quite well. According to the results obtained from the evaluation criteria, the performance of the group of tree-based models (random tree) was better in comparison with that of the group of kernel-based models and experimental methods. Among three experimental methods studied, the modified Making and McGuinness-Bordne methods represented the best performance and the worst performance, respectively. Since the results of this study depend entirely on regional meteorological variables, generalization cannot be made and this issue can be considered as a disadvantage of the study. For this reason, it is recommended to conduct similar studies in different climates of Iran and to select appropriate models and input parameters for each region.

Data Availability

The data can be sent by corresponding author by email based on the request.

Conflicts of interest

The authors of this paper declared no conflict of interest regarding the authorship or publication of this article.



ISSN: 2476-3683

محیط‌زیست و مهندسی آب

Homepage: www.jewe.ir

مقاله پژوهشی

تحلیل مقایسه‌ای روش‌های داده‌کاوی مبتنی بر هسته و درخت در مقابل روش‌های تجربی در تخمین تبخیر-تعرق مرجع ماهانه

سحر جاویدان^۱، محمدتقی ستاری^{۲*} و احمد مهرابی^۳

^۱دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

^۲دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

^۳کارشناس دفتر مطالعات منابع آب، شرکت آب منطقه‌ای استان قزوین، قزوین، ایران

اطلاعات مقاله

چکیده

تاریخ دریافت: [۱۴۰۰/۰۹/۰۷]

تاریخ بازنگری: [۱۴۰۰/۱۲/۰۷]

تاریخ پذیرش: [۱۴۰۰/۱۲/۰۸]

واژه‌های کلیدی:

تبخیر-تعرق مرجع

داده‌کاوی

درخت تصادفی

مکینگ اصلاح‌شده

*نویسنده مسئول:

mtsattar@tabrizu.ac.ir



هزینه‌بر و زمان‌بر بودن اندازه‌گیری مستقیم تبخیر-تعرق باعث شده تا پژوهشگران جهت پیش‌بینی تبخیر-تعرق به استفاده از روش‌های غیرمستقیم روی آورند. هدف پژوهش حاضر بررسی توانایی روش‌های داده‌مبنای مبتنی بر هسته، مبتنی بر درخت، روش دسته‌بندی و روش‌های تجربی در برآورد میزان تبخیر-تعرق مرجع می‌باشد. بدین منظور، داده‌های مربوط به پارامترهای هواشناسی از جمله دمای میانگین، تعداد ساعات آفتابی، حداکثر و حداقل دما، سرعت باد، بارش و رطوبت نسبی در بازه زمانی ۳۹ yr گردآوری شد. ماتریس همبستگی، الگوریتم رلیف و دانش و تجربه نویسندگان همراه با سعی و خطا مبنای انتخاب سناریوهای ورودی بودند. عملکرد روش‌های مذکور با معیارهای ضریب همبستگی (R^2)، ریشه میانگین مربعات خطا (RMSE)، شاخص پراکندگی (SI)، نش ساتکلیف (NS) و ویلموت (WI) مورد ارزیابی قرار گرفت. بررسی نتایج نشان داد از بین کلیه سناریوها، سناریو ۱۳ شامل ترکیب دمای بیشینه و شاخص زمانی ماهانه مبتنی بر الگوریتم رلیف به‌عنوان سناریو برتر و از سوئی دیگر مدل درخت تصادفی با $R = 0.99$ ، $RMSE = 0.04$ mm/day و $SI = 0.01$ به‌عنوان روش برتر انتخاب شد. بدین ترتیب حداکثر دما به‌عنوان مهم‌ترین پارامتر هواشناسی تأثیرگذار در مدل‌سازی تبخیر-تعرق مرجع شناسایی گردید.

۱- مقدمه

مستقیم (لایسیمتر) به‌دلیل هزینه‌بر بودن و زمان‌بر بودن در همه مناطق امکان‌پذیر نمی‌باشد (Khoshhal et al. 2015).

یکی از رایج‌ترین روش‌های غیرمستقیم، استفاده از معادله فائو-پنمن-مانتیت است که در فقدان لایسیمتر، تخمین قابل قبولی از مقادیر تبخیر-تعرق ارائه می‌نماید، ولی به دلیل نیاز

تبخیر-تعرق یکی از مهم‌ترین اجزای چرخه هیدرولوژیکی است که تخمین درست آن در طراحی و مدیریت سامانه‌های آبیاری به‌خصوص در مناطق خشک و نیمه‌خشک از اهمیت زیادی برخوردار است. روش‌های متعددی برای محاسبه تبخیر-تعرق وجود دارد که به دو دسته مستقیم و غیرمستقیم تقسیم می‌شوند. استفاده از روش اندازه‌گیری



بوستینگ^۹ و جنگل تصادفی^{۱۰}، مدل درخت گرادیان بوستینگ را برای برآورد تبخیر-تعرق در منطقه دشت سیستان توصیه کرده و دمای هوا را به‌عنوان مهمترین متغیر معرفی کردند. در پژوهش Nasrolahi et al. (2020) مقایسه عملکرد مدل‌ها نشان از برتری مدل ماشین بردار پشتیبان نسبت به شبکه بیزین داشت. Samadianfard et al. (2020) نشان دادند که سناریو چهارم (ترکیبی از پارامترهای هواشناسی شامل دما، رطوبت نسبی، سرعت باد، ساعات آفتابی) مدل GPR^{۱۱} با شاخص پراکندگی در محدوده ۰/۱۳۲ تا ۰/۱۷۹ در ایستگاه‌های آستارا، بندر انزلی، رشت، رامسر، نوشهر و ساری، سناریو ۴ مدل SVR^{۱۲} با شاخص پراکندگی ۰/۱۱۶ تا ۰/۱۲۰ در ایستگاه‌های بندر ترکمن و گنبدکاووس و روش هارگریوز-سامانی با شاخص پراکندگی ۰/۵۰۹ در ایستگاه گرگان برآوردهای به‌مراتب دقیق‌تری از تبخیر-تعرق گیاه مرجع داشتند.

نتایج پژوهش Sattari et al. (2020) حاکی از آن است که مدل جنگل تصادفی با ضریب همبستگی ۰/۹۹۳ بهتر از سایر مدل‌های مبتنی بر درخت است. علاوه بر این، مدل GRU^{۱۳} عملکرد خوبی را نسبت به مدل‌های دیگر داشته و حداکثر دما در تخمین ET₀ نسبت به سایر متغیرها مؤثرتر است. Sattari et al. (2021) دریافتند در روش‌های رگرسیون فرآیند گاوسی و رگرسیون بردار پشتیبان مبتنی بر هسته، هسته جهانی مبتنی بر تابع Pearson VII موفق‌ترین بوده است. Ayaz et al. (2021) دو مدل LSTM^{۱۴} و SVR را به‌عنوان قوی‌ترین مدل‌های یادگیری ماشین برای تخمین ET₀ با حداقل داده‌های آب و هوایی معرفی کردند. مطالعات آن‌ها نشان داد که حتی یک ترکیب سه پارامتری (مقادیر دما، سرعت باد و رطوبت نسبی) یا ترکیب دو پارامتری (دما و رطوبت نسبی، دما و سرعت باد) نیز می‌تواند در برآورد ET₀ امیدوارکننده باشد. هدف پژوهش حاضر برآورد میزان تبخیر-تعرق مرجع ماهانه در ایستگاه سینوپتیک قزوین براساس روش‌های رگرسیون فرآیند گاوسی، رگرسیون بردار پشتیبان، جنگل تصادفی، درخت تصادفی، درخت تصمیم M5P و کاهش خطای هرس درخت و روش‌های تجربی

به وجود پارامترهای زیاد هواشناسی دارای محدودیت کاربردی است (Hashemi et al. 2017). روش‌های داده-کاوی و مدل‌های مبتنی بر داده، اطلاعات و الگوهای پنهان درون داده‌ها را عموماً براساس الگوریتم‌های یادگیری کشف می‌کنند. در سال‌های اخیر فن مبتنی بر داده توجه زیادی را به خود جلب کرده و به‌عنوان یک ابزار جدید در حال ظهور برای تجزیه و تحلیل پدیده‌های عمدتاً غیرخطی شناخته شده است (Hamoud et al. 2018).

در زمینه برآورد تبخیر-تعرق، پژوهش‌های بسیاری در سطح ملی و بین‌المللی انجام پذیرفته است. نتایج حاصل از محاسبات Hozhabr et al. (2014) نشان داد که شبکه‌های عصبی مصنوعی نسبت به تمامی روش‌های مورد استفاده دارای عملکرد مناسب‌تری بوده و در میان روش‌های کلاسیک، مدل تورک در اولویت قرار دارد. Moradi and Ziaean (2019) مدل‌های بلانی کریدل^۱ و فائو-پنمن-مانتیت^۲ را به‌عنوان مناسب‌ترین مدل در مقایسه با مقادیر اندازه‌گیری شده از لایسیمتر برای هرمزگان و مناطقی با شرایط اقلیمی مشابه پیشنهاد کردند. Piri and Pouzan (2019) نشان دادند که روش ماکینگ اصلاح‌شده^۳ در اقلیم سرد و نیمه‌خشک، سرد و معتدل، گرم و خشک، سرد و خشک و گرم و مرطوب برآورد بهتر و نزدیک‌تری به روش فائو-پنمن-مانتیت داشته است. Siasar and Honar (2019) مدل جنگل تصادفی با ضریب همبستگی (R=0.983) و کمترین میانگین قدر مطلق خطا (MAE=0.798)^۴ را به‌عنوان مدلی با بیشترین دقت برای برآورد تبخیر-تعرق در منطقه دشت سیستان توصیه نمودند. Wu and Fun (2019) روش SVM^۵ را برای پیش‌بینی ET₀ روزانه در سراسر چین با توجه به دقت بالا و پایداری بسیار توصیه کردند، در حالی که KNEA^۶ و MARS^۷ نیز مدل‌های قدرتمند امیدوارکننده‌ای بودند. Siasar et al. (2020) از بین مدل‌های خطی تعمیم‌یافته^۸، درخت گرادیان

¹ Blaney Criddle

² FAO Penman Monteith

³ Modified Makkink

⁴ Mean Absolute Error

⁵ Support Vector Machine

⁶ Kernel-based Nonlinear Extension of Arps decline model

⁷ Multivariate Adaptive Regression Spline

⁸ Generalized Linear Model

⁹ Boosting Gradient Tree

¹⁰ Random Forest

¹¹ Gaussian Process Regression

¹² Support Vector Regression

¹³ Gated Recurrent Unit

¹⁴ Long Short-Term Memory

در پژوهش حاضر برای برآورد میزان تبخیر-تعرق مرجع ماهانه در دوره آماری ۳۹ ساله (۱۳۶۱-۱۳۹۹) از داده‌های متغیرهای هواشناسی شامل حداکثر، حداقل و میانگین دما (T_{mean} , T_{min} , T_{max})، بارش ماهانه (P)، ساعات آفتابی (n)، متوسط رطوبت (RH) و سرعت باد (U_{10}) استفاده شد. مشخصات آماری متغیرهای مورد استفاده در جدول (۱) آورده شده است

در جدول (۱) نشان‌دهنده انحراف معیار است که برای تعیین پراکندگی داده‌ها به کار می‌رود و CV ضریب تغییرات است که از نسبت انحراف معیار به میانگین به دست می‌آید. با استفاده از ماتریس همبستگی، روش رلیف^۱ و سعی و خطای متعدد مبتنی بر تجارب پیشین نویسندگان مقاله، انواع ترکیب‌های ورودی مشخص شد (جدول ۲).

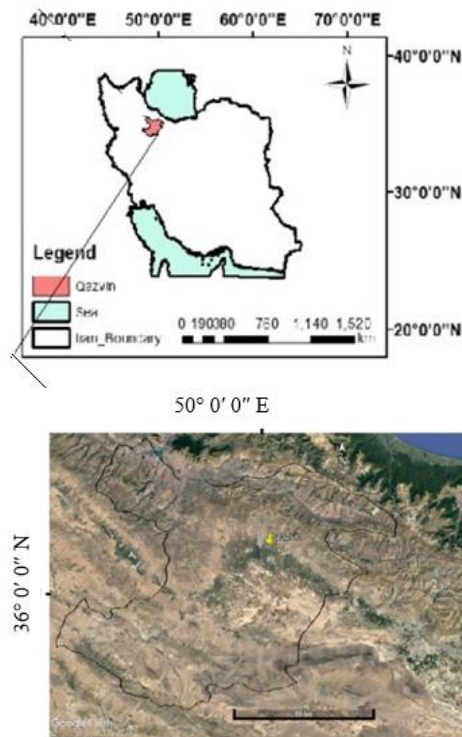
بررسی ماتریس همبستگی نشان داد که تبخیر-تعرق مرجع ماهانه به ترتیب با دمای میانگین، تعداد ساعات آفتابی، حداکثر دما، حداقل دما، سرعت باد، بارش، شاخص زمانی ماهانه و رطوبت نسبی بیش‌ترین همبستگی را داشته است. رلیف یک الگوریتم انتخاب ویژگی برای کاهش ابعاد مسئله است. نکات قوت این الگوریتم ساده بودن اصول و عدم پیچیدگی آن، قابل حل بودن با توابع چندجمله‌ای مرتبه پایین، قابل استفاده بودن برای داده‌های پیوسته و نیاز به تعداد کم داده‌های آموزشی است. سناریوهای به دست آمده از ماتریس همبستگی و روش رلیف در نرم‌افزار وکا^۲ انجام شد. از بین داده‌های موجود، ۸۰٪ برای واسنجی و مابقی ۲۰٪ برای صحت‌سنجی در نظر گرفته شده است. برای تمامی روش‌های داده مبنای نرم‌افزار وکا استفاده شد.

پرستلی تیلور، ماکینگ اصلاح شده و مک گاینس بردنه^۱ و مقایسه تطبیقی بین این روش‌ها بود. همچنین به منظور بررسی دقیق‌تر عملکرد روش‌های داده‌مبنای فوق‌الذکر و در راستای بهبود نتایج مدل‌سازی به بررسی عملکرد روش دسته‌بندی^۲ به‌عنوان روشی ترکیبی^۳ نیز پرداخته شد. از دیگر اهداف این پژوهش شناسایی مؤثرترین پارامترهای هواشناسی در برآورد تبخیر-تعرق می‌باشد

۲- مواد و روش‌ها

۲-۱- محدوده مطالعاتی و داده‌های موردنیاز

استان قزوین از شمال به استان‌های مازندران و گیلان، از جنوب به استان مرکزی، از شرق به استان البرز و از غرب به استان‌های همدان و زنجان محدود می‌شود. ایستگاه سینوپتیک قزوین در طول جغرافیایی $59^{\circ} 2' 50''$ ، عرض جغرافیایی $36^{\circ} 15' 0''$ و ارتفاع ۱۲۷۹ m از سطح دریا قرار گرفته است. دمای میانگین (T_{mean}) از $6^{\circ}C$ تا $29^{\circ}C$ و بارش ماهانه از صفر تا ۱۴۹ mm متغیر است. موقعیت مکانی ایستگاه مورد مطالعه در شکل (۱) نشان داده شده است.



شکل ۱- موقعیت مکانی ایستگاه سینوپتیک قزوین
Fig. 1 Location of Qazvin Synoptic Station

¹Relief
²Weka

¹McGuinness-Bordne
²Bagging
³Ensemble

جدول ۱- مشخصات آماری پارامترهای مورد استفاده در دوره آماری ۳۹ ساله (۱۳۹۹-۱۳۶۱)

Table 1 Statistical characteristics of implemented parameters

Station	Variable	mean	max	min	S _d	CV
Qazvin	T _{min} (°C)	7.3	33.8	-10.9	7.8	1.1
	T _{max} (°C)	21.6	38.8	-2.9	11.0	0.5
	T _{mean} (°C)	14.5	29.0	-5.7	9.2	0.6
	RH (%)	53.9	85.5	31.8	13.1	0.2
	N	8.2	12.6	1.9	2.7	0.3
	U ₁₀ (m/s)	1.7	3.2	0.5	0.5	0.3
	P (mm)	26.8	148.9	0.0	29.3	1.1

جدول ۲- ترکیب پارامترهای مورد استفاده در سناریوهای مختلف

Table 2 Combination of parameters used in different scenarios

Scenario Number	Input Parameters	Scenario selection method
1	T _{mean}	Correlation Matrix
2	T _{mean} , N	Correlation Matrix
3	T _{mean} , N, T _{max}	Correlation Matrix
4	T _{mean} , N, T _{max} , T _{min}	Correlation Matrix
5	T _{mean} , N, T _{max} , T _{min} , U	Correlation Matrix
6	T _{mean} , N, T _{max} , T _{min} , U, P	Correlation Matrix
7	T _{mean} , N, T _{max} , T _{min} , U, P, MTI	Correlation Matrix
8	T _{mean} , N, T _{max} , T _{min} , U, P, MTI, RH	Correlation Matrix
9	T _{mean} , MTI, N, U	Relief
10	MTI	Relief
11	T _{max}	Relief
12	MTI, T _{mean}	Relief
13	MTI, T _{max}	Relief
14	N	Trial and Error
15	T _{mean} , T _{min} , T _{max}	Trial and Error
16	T _{mean} , N, U	Trial and Error

MTI: Monthly time index

۲-۲ روش‌های مورد استفاده

۲-۲-۱- روش فائو- پنمن- مانتیث

روش فائو-پنمن-مانتیث، تکامل یافته روش ترکیبی پنمن-مانتیث ۱۹۶۵ است که توسط Allen et al. (1998) برای محاسبه تبخیر-تعرق در نشریه ۵۶ سازمان فائو ارائه شده است. در پژوهش حاضر از این روش به عنوان روش استاندارد و دقیق محاسبه تبخیر-تعرق طبق رابطه (۱) استفاده شد.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \left(\frac{900}{T+273} \right) u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

که، ET₀ تبخیر-تعرق گیاه مرجع (mm/day)، R_n تابش خالص در سطح پوشش گیاهی (MJ/m/day)، G شار گرما به داخل خاک (MJ/m/day)، T متوسط دمای هوا در ارتفاع دو متری از سطح زمین (°C)، u₂ سرعت باد در ارتفاع دو متری از سطح زمین (m/s)، e_a فشار بخار واقعی (KPa)، e_s فشار بخار اشباع هوا (KPa)، e_s-e_a کمبود فشار بخار اشباع (KPa)، Δ شیب منحنی فشار بخار (KPa/°C) و λ ضریب رطوبتی (KPa/°C) است.

۲-۲-۲ روش‌های تجربی

در پژوهش حاضر برای محاسبه تبخیر-تعرق مرجع از ۳ معادله تجربی پرستلی تیلور Priestley and Taylor (1972)، ماکینگ اصلاح شده Hansen (1984) و مک گاینس بردنه Oudina (2005) طبق رابطه‌های (۲) تا (۴) استفاده شد.

$$ET_0 = \frac{1.26\Delta}{\Delta + \gamma} \times (R_n - G) \quad (2)$$

$$ET_0 = 0.7 \frac{\Delta}{\Delta + \gamma} \times \frac{R_s}{\lambda} \quad (3)$$

$$ET_0 = \frac{1}{\lambda} \times R_a \times \left(\frac{T+5}{68} \right) \quad (4)$$

که، λ گرمای نهان تبخیر (MJ/Kg)، R_s کل تابش خورشیدی (MJ/m².d)، T دمای میانگین (°C) و R_a تابش برون زمینی (MJ/m².d) است.

۲-۲-۳ روش رگرسیون فرآیند گاوسی

این روش تعمیم یافته توزیع گاوسی است. مدل‌های رگرسیون فرآیند گاوسی قادر به درک توزیع پیش‌بینی متناظر با ورودی آزمون هستند. فرآیندهای گاوسی همان قدر که



همبستگی کم‌تری بین درختان جامعه به دست می‌آید (et Suchetana al. 2017). در روش جنگل تصادفی، مقدار موارد ترجیحی برای پردازش، در صورت انجام پیش‌بینی دسته‌ای برابر ۱۰۰ و تعداد درختان ۱۰۰ در نظر گرفته شده است.

۲-۲-۶- روش درخت تصادفی

روش درخت تصادفی می‌تواند با مشکلات طبقه‌بندی و رگرسیون مقابله کند. طبقه‌بندی‌کننده درختان تصادفی، بردار ویژگی ورودی را می‌گیرد، آن را با هر درخت در جنگل طبقه‌بندی می‌کند و برچسب کلاسی را که اکثریت "رای" را دریافت کرده است، برمی‌گزیند. در مورد رگرسیون، پاسخ طبقه‌بندی‌کننده، میانگین پاسخ‌ها به تمام درختان جنگل است (Ajayram et al. 2021). در روش درخت تصادفی، حداقل نسبت واریانس در تمام داده‌هایی که باید در یک گره وجود داشته باشد تا تقسیم درختان رگرسیون انجام شود برابر ۰/۰۰۱ در نظر گرفته شده است. حداقل وزن کل نمونه‌ها در یک برگ برابر یک و مقدار موارد ترجیحی برای پردازش، در صورت انجام پیش‌بینی دسته‌ای برابر ۱۰۰ در نظر گرفته شده است.

۲-۲-۷- مدل درختی MSP

الگوریتم MSP بازسازی منطقی و توسعه‌یافته از M5 است که توسط (Wang and Witten 1997) معرفی شده است. معیار تقسیم در یک گره بر اساس انتخاب انحراف معیار مقادیر خروجی که به آن گره می‌رسند به‌عنوان معیار خطا است. با آزمودن هر صفت (پارامتر) در گره، کاهش مورد انتظار در خطا محاسبه می‌شود. کاهش انحراف معیار با رابطه (۶) محاسبه می‌شود (Wang and Witten 1997).

$$SDR = \frac{m}{|T|} \times \beta(i) \times \left[sd(T) - \sum_{j \in (L,R)} \frac{|T_j|}{|T|} \times sd(T_j) \right] \quad (6)$$

کم‌تر از درصد مجموعه آموزشی باشد، آن گره تقسیم نمی‌شود. اگر گسترش درخت در حین ساخت متوقف شود، این فرآیند را پیش اجرا می‌نامند. اما در روش پس از هرس، درخت تا زمانی که تمام برگ‌ها خالص شده و خطایی در مجموعه آموزشی وجود نداشته باشد، منبسط می‌شود (Chen et al. 2019). در روش REPT^۱ حداقل وزن کل نمونه‌ها در یک برگ برابر ۲، مقدار داده‌های مورد استفاده

ساده‌اند، فراگیر نیز هستند. هر مشاهده y را می‌توان از طریق مدل نویز گاوسی به یک تابع اصلی (رابطه ۵) ارتباط داد.

$$y = f(x) + N(0, \sigma_f^2) \quad (5)$$

که، $N(0, \sigma_f^2)$ به‌عنوان نویز تابع نرمال با میانگین صفر و واریانس σ_f^2 است. در این روش پس از انجام سعی و خطاهای متعدد مشخص گردید، نتیجه بهینه زمانی حاصل می‌شود که تابع کرنل، Puk در نظر گرفته شود.

۲-۲-۴- روش رگرسیون SVM

روش رگرسیون SVM رویکرد یادگیری ماشین در زمینه پژوهش‌های داده‌محور است. مدل‌های SVM به دو بخش تقسیم می‌شوند: (۱) مدل‌های طبقه‌بندی‌کننده ماشین بردار پشتیبان، (۲) مدل رگرسیون بردار پشتیبان (SVR). مدل SVM برای حل طبقه‌بندی داده‌ها در کلاس‌های مختلف و مدل SVR برای پیش‌بینی استفاده می‌شود (Demirci 2019). در روش SVR نیز مانند روش GPR انتخاب کرنل مناسب با سعی و خطا در نرم‌افزار وکا انجام شد. بهترین کرنل در این روش، Puk و پولی کرنل در نظر گرفته شد.

۲-۲-۵- روش جنگل تصادفی

الگوریتم جنگل تصادفی یک الگوریتم طبقه‌بندی نظارت شده است. بین تعداد درختان الگوریتم و نتایجی که می‌توان به دست آورد، رابطه مستقیم وجود دارد. با افزایش تعداد درختان، نتیجه قطعی گرفته می‌شود. نقطه ضعف روش جنگل تصادفی این است که برخلاف روش درخت طبقه‌بندی، این روش درختی را به‌عنوان خروجی نمی‌دهد. مزیت انتخاب برآوردگرهای تصادفی در این روش این است که مدل حاصل از دقت بیش‌تری برخوردار است، زیرا که، SDR کاهش انحراف معیار، T نشان‌دهنده سری نمونه‌هایی که به گره می‌رسند، m تعداد نمونه‌هایی که برای این صفت مقادیر گمشده ندارند، $\beta(i)$ یک عامل اصلاحی، T_L و T_R مجموعه‌هایی هستند که از تقسیم بر روی این صفت به‌وجود می‌آیند.

۲-۲-۸- روش کاهش خطای هرس درخت

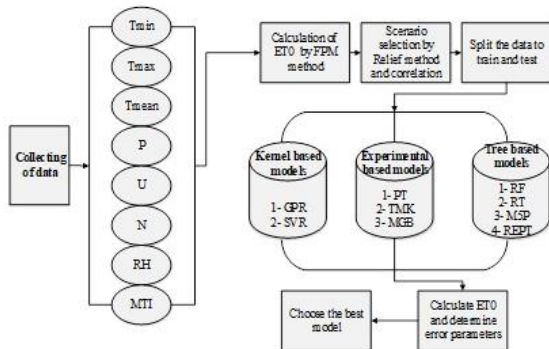
دو روش برای هرس درختان تصمیم وجود دارد: پیش‌هرس و پس‌هرس. وقتی تعداد نمونه‌هایی که به یک گره می‌رسد

¹ Reduced Error Pruning Tree

$$NS = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (10)$$

$$WI = \left| 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (|y_i - \bar{x}| + |x_i - \bar{x}|)^2} \right|, 0 \leq WI \leq 1 \quad (11)$$

که، y_i تبخیر-تعرق گیاه مرجع تخمین زده شده، x_i تبخیر-تعرق گیاه مرجع محاسبه شده با روش فائو-پنمن-مانتیت و N تعداد داده‌ها است. با استفاده از نرم‌افزار Visio 2016 روندنمای مراحل انجام پژوهش ترسیم و در شکل (۲) نشان داده شده است.



شکل ۲- شمای کلی مراحل انجام پژوهش

Fig. 2 Outline of the research processes

۳- یافته‌ها و بحث

در این بخش، نتایج به دست آمده از ۱۶ سناریوی ورودی مورد استفاده در روش‌های مختلف داده‌کاوی مبتنی بر هسته و مبتنی بر درخت در جدول (۳) نشان داده شده است. با توجه به جدول (۳) سناریوهای ۹، ۱۲ و ۱۳ در تمام روش‌ها نتایج بهتری نسبت به سایر سناریوهای مورد استفاده داشته است. از بین سناریوهای ۹، روش SVR، سناریو ۱۲، روش M5P و سناریو ۱۳، روش RT دقت بالاتری را داشته‌اند. سناریو ۹ به دلیل وجود پارامترهای سرعت باد و ساعات آفتابی که اساساً اندازه‌گیری آن‌ها نسبتاً سخت هست، می‌تواند کاربرد نسبتاً کمتری داشته باشد. زیرا اساساً در صورت راحتی اندازه‌گیری این دو پارامتر و سایر پارامترهای هواشناسی روش‌هایی مثل پنمن مانیتیت فائو به‌عنوان روشی دقیق و استاندارد قابل استفاده بود. هدف از به کار بردن روش‌های داده‌کاوی، آسان بودن استفاده از آن‌ها و به کار بردن پارامترهای کمتر به‌عنوان ورودی مدل برای محاسبه ET_0 است. بنابراین سناریو ۱۲ و ۱۳ با داشتن دقت بالا و پارامترهای کمتر (شامل دما و شاخص زمانی ماهانه) به‌عنوان سناریوهای کاربردی در نظر گرفته شدند.

برای هرس، برابر ۳ و مقدار موارد ترجیحی برای پردازش، در صورت انجام پیش‌بینی دسته‌ای برابر ۱۰۰ در نظر گرفته شده است.

۲-۲-۹- روش دسته‌بندی

روش دسته‌بندی یکی از روش‌های گروهی و یا ترکیبی در یادگیری ماشینی است. آموزش الگوریتم پایه در روش دسته‌بندی به‌صورت موازی انجام می‌شود. در نهایت برای انتخاب مدل نهایی یک رأی‌گیری در بین همه مدل‌ها انجام می‌شود و هر مدلی که بیشترین میزان رأی را بیاورد، به‌عنوان مدل نهایی انتخاب می‌شود. استفاده از این روش واریانس را کاهش داده و از افتادن در بیش‌برازش جلوگیری می‌کند (Bue et al. 2016). در روش دسته‌بندی، مقدار موارد ترجیحی برای پردازش، در صورت انجام پیش‌بینی دسته‌ای برابر ۱۰۰، تعداد تکرارهایی که باید انجام شود برابر ۱۰ و اندازه هر کیسه به‌عنوان درصدی از اندازه مجموعه آموزشی برابر ۱۰۰ در نظر گرفته شده است و تمام روش‌های استفاده شده در این پژوهش به‌عنوان الگوریتم پایه در نظر گرفته شده است.

۲-۳- معیارهای ارزیابی مدل‌ها

در این پژوهش یکی از روش‌های انتخاب ترکیب سناریوهای ورودی مدل‌ها الگوریتم رلیف بود. برای مقایسه مقادیر ET_0 محاسبه شده با مدل‌های مورد بررسی با مقادیر روش فائو-پنمن-مانتیت از شاخص‌های ضریب همبستگی (R)، ریشه میانگین مربعات خطا^۱ (RMSE)، شاخص پراکندگی^۲ (SI)، نش ساتکلیف^۳ (NS) و ضریب ویلموت^۴ (WI) استفاده شد. بهترین مدل، مربوط به مدلی با ضرایب R، NS و WI بیش-تر و RMSE و SI کمتر بود. فرمول‌های آماره‌های فوق به-ترتیب در روابط (۷) تا (۱۱) ارائه گردیده است:

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (8)$$

$$SI = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2}}{\bar{x}} \quad (9)$$

¹Root Mean Square Error

²Scatter Index

³Nash Sutcliffe

⁴Wilmot Index

جدول ۳- معیارهای ارزیابی روش‌های مبتنی بر هسته و درخت براساس سناریوهای تعریف شده

Table 3 Evaluation criteria according to kernel and tree based methods in defined scenarios

Scenario	Method														
	GPR					SVR					RF				
	R	RMSE	SI	NS	WI	R	RMSE	SI	NS	WI	R	RMSE	SI	NS	WI
1	0.95	0.59	0.19	0.90	0.97	0.95	0.64	0.20	0.88	0.97	0.91	0.76	0.24	0.83	0.95
2	0.94	0.61	0.19	0.89	0.97	0.95	0.58	0.18	0.90	0.97	0.94	0.66	0.21	0.87	0.96
3	0.95	0.60	0.19	0.89	0.97	0.95	0.57	0.18	0.90	0.97	0.95	0.60	0.19	0.89	0.97
4	0.94	0.68	0.21	0.86	0.96	0.94	0.68	0.21	0.86	0.96	0.94	0.66	0.21	0.87	0.96
5	0.95	0.60	0.19	0.99	0.97	0.95	0.63	0.20	0.88	0.96	0.94	0.64	0.20	0.88	0.96
6	0.96	0.53	0.17	0.92	0.98	0.97	0.50	0.16	0.93	0.98	0.95	0.58	0.18	0.90	0.97
7	0.99	0.30	0.10	0.97	0.99	0.99	0.17	0.05	0.99	1.00	0.99	0.24	0.08	0.98	1.00
8	0.99	0.31	0.10	0.97	0.99	0.99	0.20	0.06	0.99	1.00	0.99	0.23	0.07	0.98	1.00
9	0.99	0.20	0.20	0.99	1.00	0.99	0.11	0.03	1.00	1.00	0.99	0.20	0.06	0.99	1.00
10	0.99	0.29	0.09	0.98	0.99	0.99	0.25	0.08	0.98	1.00	0.99	0.25	0.08	0.98	1.00
11	0.93	0.66	0.21	0.87	0.97	0.93	0.73	0.23	0.84	0.96	0.98	0.33	0.10	0.99	0.99
12	0.99	0.17	0.06	0.99	1.00	0.99	0.17	0.05	0.99	1.00	0.99	0.18	0.06	0.99	1.00
13	0.99	0.29	0.09	0.98	0.99	0.99	0.25	0.08	0.98	1.00	0.99	0.08	0.03	1.00	1.00
14	0.93	0.77	0.24	0.82	0.95	0.93	0.76	0.24	0.83	0.95	0.89	0.91	0.29	0.76	0.93
15	0.94	0.62	0.20	0.89	0.97	0.94	0.63	0.20	0.88	0.97	0.94	0.65	0.20	0.88	0.97
16	0.96	0.56	0.17	0.91	0.97	0.96	0.57	0.18	0.90	0.97	0.95	0.60	0.19	0.89	0.97
Scenario	M5P					RT					REPT				
	R	RMSE	SI	NS	WI	R	RMSE	SI	NS	WI	R	RMSE	SI	NS	WI
	1	0.95	0.59	0.18	0.90	0.97	0.87	0.91	0.29	0.75	0.93	0.94	0.61	0.19	0.89
2	0.95	0.62	0.19	0.89	0.97	0.90	0.83	0.26	0.80	0.95	0.95	0.63	0.20	0.88	0.97
3	0.95	0.61	0.19	0.89	0.97	0.92	0.78	0.24	0.82	0.95	0.95	0.62	0.19	0.89	0.97
4	0.91	0.78	0.25	0.82	0.95	0.89	0.89	0.28	0.76	0.94	0.94	0.65	0.21	0.87	0.97
5	0.82	1.19	0.37	0.58	0.90	0.92	0.73	0.23	0.84	0.96	0.93	0.68	0.21	0.86	0.96
6	0.88	0.90	0.28	0.76	0.94	0.91	0.77	0.24	0.82	0.95	0.93	0.68	0.21	0.87	0.96
7	0.98	0.39	0.12	0.96	0.99	0.97	0.45	0.14	0.94	0.99	0.99	0.26	0.08	0.98	1.00
8	0.99	0.26	0.08	0.98	1.00	0.98	0.41	0.13	0.95	0.99	0.99	0.24	0.08	0.98	1.00
9	0.99	0.15	0.05	0.99	1.00	0.99	0.25	0.08	0.98	1.00	0.99	0.25	0.08	0.98	1.00
10	0.99	0.34	0.11	0.97	0.99	0.99	0.25	0.08	0.98	1.00	0.99	0.25	0.08	0.98	1.00
11	0.94	0.63	0.20	0.88	0.97	0.99	0.20	0.06	0.99	1.00	0.95	0.61	0.19	0.89	0.97
12	0.99	0.16	0.05	0.99	1.00	0.99	0.21	0.07	0.99	1.00	0.99	0.18	0.06	0.99	1.00
13	0.99	0.25	0.08	0.98	1.00	0.99	0.04	0.01	1.00	1.00	0.99	0.17	0.05	0.99	1.00
14	0.92	0.78	0.25	0.82	0.94	0.86	1.04	0.33	0.68	0.91	0.90	0.85	0.27	0.79	0.94
15	0.88	0.97	0.31	0.72	0.93	0.91	0.81	0.25	0.81	0.95	0.94	0.64	0.20	0.88	0.97
16	0.96	0.55	0.17	0.91	0.97	0.92	0.74	0.23	0.84	0.95	0.94	0.65	0.88	0.88	0.96

معیارهای ارزیابی روش‌های تجربی نیز در جدول (۴) نشان داده شده است. نتایج جدول (۴) نیز نشان داد که از ۳ روش تجربی مورد مطالعه روش ماکینگ اصلاح شده دارای عملکرد بهتری بود. (Piri and Pouzan (2019) نیز روش ماکینگ اصلاح شده را برای اقلیم‌های سرد و نیمه خشک به عنوان روش برتر معرفی کردند.

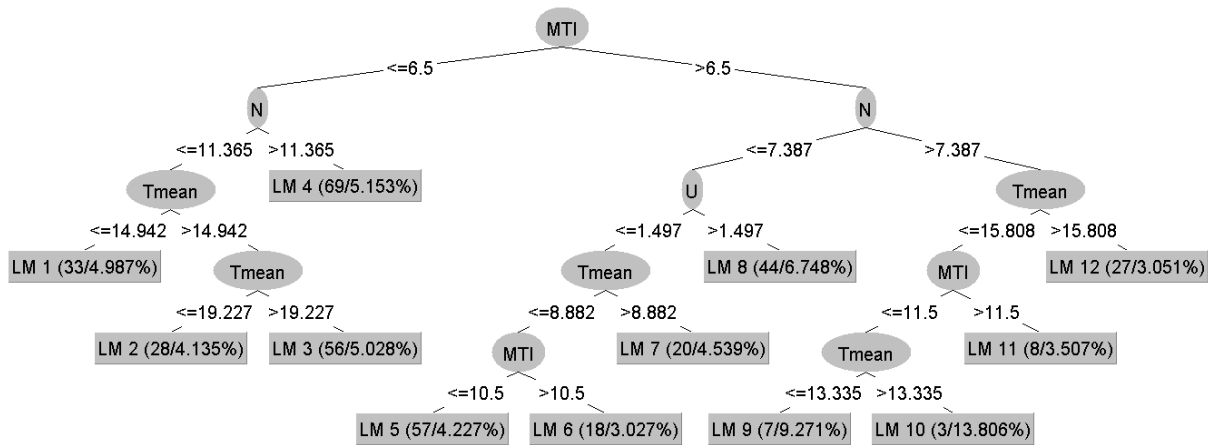
جدول ۴- معیارهای ارزیابی روش‌های تجربی

Table 4 Evaluation criteria of experimental methods

Method	Evaluation Metrics				
	R	RMSE (mm/day)	SI	NS	WI
PT	0.990	0.330	0.104	0.968	0.993
TMK	0.995	0.350	0.110	0.964	0.990
MGB	0.996	1.084	0.340	0.652	0.940

فرم درختی سناریو برتر در روش M5P در شکل (۳) نشان داده شده است. با توجه به شکل (۳) مدل درختی M5P برای پیش‌بینی ET₀ برای شرایط مختلف و دامنه تغییرات پارامترهای هواشناسی ۱۲ معادله ارائه نموده است که چهار معادله به نیمه اول سال و هشت معادله به نیمه دوم سال

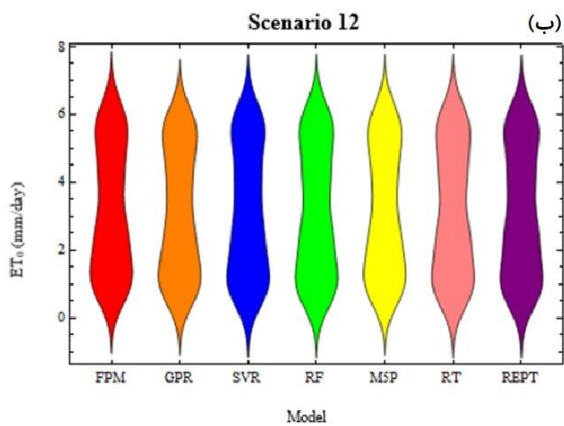
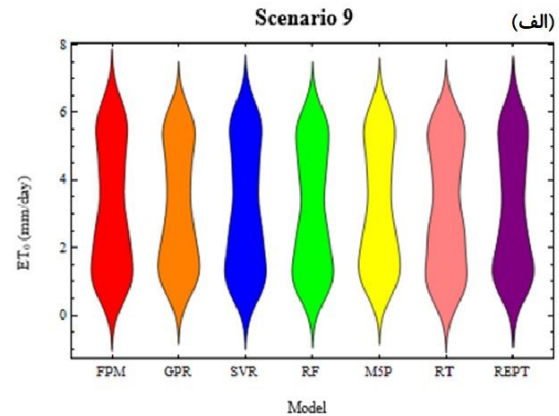
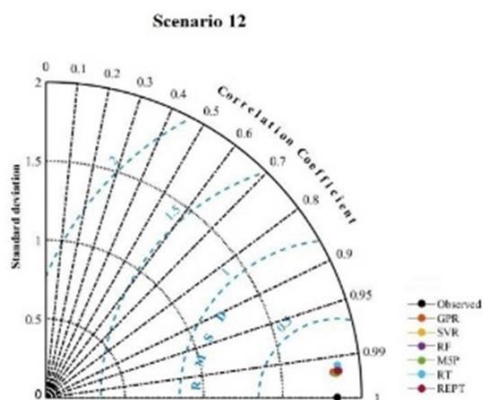
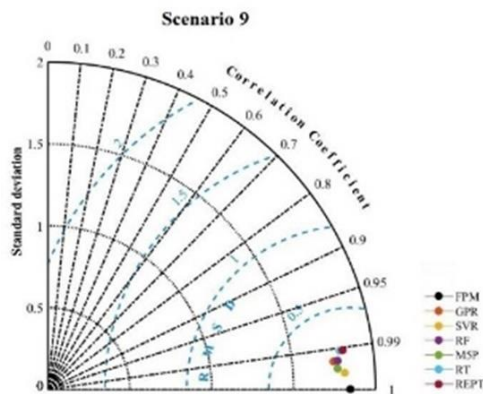
سرعت آفتابی، سرعت باد و شاخص زمانی ماهانه بیشترین تأثیر در تبخیر-تعرق مرجع را داشته‌اند.

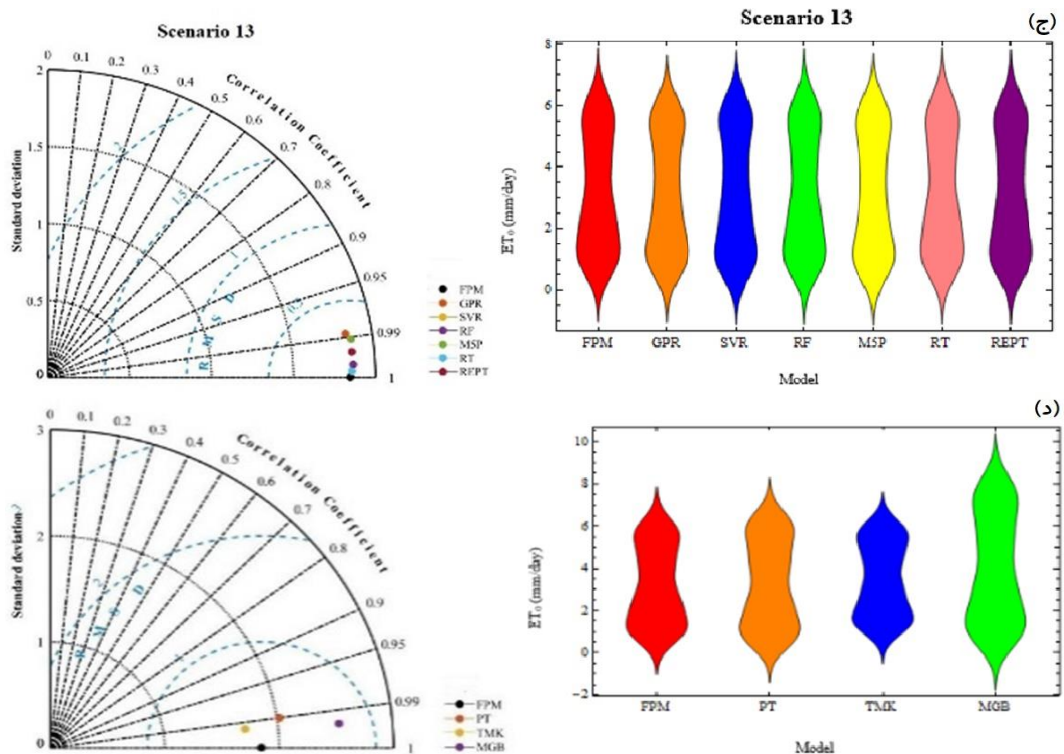


شکل ۳- فرم درخت سناریو برتر روش M5P
Fig. 3 Tree Visualized of Top Scenario on M5P Method

انحراف استاندارد داده‌های برآورد شده از مدل، نزدیک به هم بوده و نتایج قابل قبولی را ارائه کرده است. نمودار ویلونی نیز بیانگر نزدیک و قابل قبول بودن نتایج مدل‌ها به هم در سه سناریو معرفی شده است.

برای نمایش بهتر عملکرد مدل‌ها، دیاگرام تیلور و نمودار ویلونی برای سناریوهای برتر (۹، ۱۲ و ۱۳) و روش‌های تجربی در شکل (۴) ترسیم شده است. با توجه به شکل (۴) بهترین عملکرد، مربوط به سناریو ۱۳ مدل درخت تصادفی بود. زیرا مقدار انحراف استاندارد داده‌های مشاهداتی و





شکل ۴- دیاگرام تیلور و نمودار ویلونی الف- سناریو ۹، ب- سناریو ۱۲، ج- سناریو ۱۳، و د- روش‌های تجربی
 Fig. 4 Taylor diagram and violin diagram a) Scenario 9, b) Scenario 12, c) Scenario 13, and d) experimental

جدول ۵- مشخصات آماری تبخیر-تعرق مرجع مشاهداتی و تخمینی از مدل‌ها بر حسب میلی‌متر بر روز برای بخش صحت سنجی

Table 5 Statistical properties of observed and estimated ET₀ (mm/day) values for the test phase

Method	Scenario	Minimum	Maximum	Mean	Std. Deviation
GPR	9	0.728	5.904	3.180	1.751
	12	0.659	5.917	3.160	1.843
	13	0.756	5.921	3.158	1.837
SVR	9	0.754	6.071	3.165	1.816
	12	0.661	6.043	3.175	1.873
	13	0.681	6.076	3.154	1.869
RF	9	0.744	5.957	3.115	1.776
	12	0.660	6.113	3.176	1.878
	13	0.698	6.353	3.173	1.864
MSP	9	0.713	6.137	3.191	1.771
	12	0.681	6.149	3.186	1.862
	13	0.618	5.997	3.154	1.866
RT	9	0.681	5.926	3.140	1.799
	12	0.649	6.128	3.174	1.873
	13	0.682	6.462	3.177	1.858
REPT	9	0.757	6.001	3.149	1.813
	12	0.595	6.105	3.175	1.870
	13	0.662	6.201	3.164	1.861
PT		0.579	6.480	3.269	2.022
TMK		1.098	6.100	3.450	1.710
MGB		0.582	8.262	4.003	2.534
FPM		0.733	6.462	3.186	1.847

سخت می‌باشد. این در حالی است که روش‌های داده‌مبنا با تعداد پارامتر بسیار کم نتایج قابل قبول تری ارائه می‌کنند و این موضوع باعث افزایش محبوبیت روش‌های داده‌مبنا می‌گردد. از طرفی دیگر با افزایش امکانات اندازه‌گیری و داشتن دوره طولانی مدتی از داده‌های هواشناسی در آینده پیش‌بینی می‌شود که این روش‌های داده‌مبنا همچنان در کانون توجه پژوهشگران داخلی و خارجی باشد. تمام روش‌های استفاده شده در پژوهش حاضر به‌عنوان الگوریتم پایه روش دسته‌بندی در نظر گرفته شد تا کارایی این روش سنجیده شود. نتایج R و RMSE روش مذکور برای سناریوهای برتر در جدول (۶) آورده شد.

جدول ۶- معیارهای آماری روش ترکیبی Bagging برای سناریوهای برتر

Table 6 Statistical criteria of the combined bagging method for superior scenarios

Scenario	Method					
	GPR		SVR		RF	
	R ²	RMSE	R ²	RMSE	R ²	RMSE
9	0.995	0.204	0.998	0.109	0.995	0.208
12	0.996	0.175	0.996	0.166	0.996	0.173
13	0.987	0.293	0.991	0.250	0.998	0.115
Scenario	M5P		RT		REPT	
	R ²	RMSE	R ²	RMSE	R ²	RMSE
	9	0.998	0.137	0.995	0.202	0.990
12	0.996	0.162	0.995	0.189	0.996	0.167
13	0.991	0.249	0.999	0.086	0.996	0.167

منظور ۱۶ ترکیب ورودی با پارامترهای هواشناسی مختلف در نظر گرفته شد. نتایج نشان داد:

۱- از بین ۱۶ سناریو مورد مطالعه، سناریوهای ۹ (شامل دمای میانگین، سرعت باد، تعداد ساعات آفتابی و شاخص زمانی ماهانه)، ۱۲ (شامل دمای میانگین و شاخص زمانی ماهانه) و ۱۳ (شامل حداکثر دما و شاخص زمانی ماهانه) که مبتنی بر الگوریتم رلیف انتخاب شده بودند، عملکرد مناسبی داشتند.

۲- با توجه به مقادیر معیارهای ارزیابی، عملکرد گروه مدل‌های مبتنی بر درخت (درخت تصادفی) بهتر از عملکرد گروه مدل‌های مبتنی بر هسته و همچنین روش‌های تجربی بود.

۳- از بین سه روش تجربی مورد مطالعه روش ماکینگ اصلاح شده بهترین و روش مک گاینس برده بدترین عملکرد را داشتند.

مقادیر مشخصات آماری تبخیر-تعرق مرجع بخش صحت‌سنجی برای داده‌های مشاهداتی و روش‌های داده-مبنا در جدول (۵) آورده شده است. به‌طور کلی با توجه به نتایج جداول و شکل‌ها نتیجه گرفته شد که T_{max} بیش‌ترین تأثیر را در برآورد تبخیر-تعرق مرجع ماهانه داشته است و عملکرد مدل درخت تصادفی به‌عنوان روش مبتنی بر درخت بهتر از عملکرد روش‌های مبتنی بر هسته و روش‌های تجربی بوده است. روش مک گاینس برده نیز به‌عنوان روشی با کمترین دقت در بین روش‌های تجربی مورد مطالعه معرفی شد. روش‌های تجربی ضمن اینکه از دقت کافی برخوردار نیستند، محاسبه آن‌ها با تعداد پارامترهای زیاد زمان بر و

با توجه به جدول (۶) روش ترکیبی دسته‌بندی در همه موارد نتیجه بهتری نداشت. وقتی روش RT به‌عنوان الگوریتم پایه روش دسته‌بندی در نظر گرفته شد در سناریوهای ۹ و ۱۲ عملکرد مدل بهبود یافت. نتایج تحقیق Wu et al. (2019) نشان داد که زمانی که در یک ایستگاه فقط داده دما وجود داشته باشد، مدل‌های MARS و SVM نسبت به سایر مدل‌ها برتری دارند. Salam and Towfiqul Islam (2020) نشان دادند که در مقایسه با سایر مدل‌های پیشرفته، مدل RT برای پیش‌بینی ET_0 در تمام ترکیب‌های ورودی و سپس RF, Bagging, RS و SVM عملکرد بهتری داشته‌اند. بررسی دقیق‌تر نتایج آن‌ها نشان داد که سرعت باد (U_2) و تابش خورشیدی (R_s) متغیرهای تأثیرگذار بر ET_0 هستند. در پژوهش حاضر نیز حداکثر دما به‌عنوان مؤثرترین پارامتر در برآورد تبخیر-تعرق معرفی شد.

۴- نتیجه‌گیری

در پژوهش حاضر برای برآورد میزان تبخیر-تعرق از روش‌های مختلف داده‌مبنا و تجربی استفاده شد. بدین

با اقلیم‌های مختلف مورد بررسی قرار گرفته و برای هر اقلیم، بهترین روش مشخص گردد.

دسترسی به داده‌ها

داده‌ها حسب درخواست، از طرف نویسنده مسئول از طریق ایمیل قابل ارسال می باشد.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که، هیچ‌گونه تضاد منافی در رابطه با نویسندگی و یا انتشار این مقاله ندارند.

۴- به کمک پارامتر هواشناسی حداکثر دمای ماهانه می‌توان با دقت بسیار بالایی مقدار تبخیر-تعرق مرجع را در منطقه قزوین پیش‌بینی نمود.

با توجه به دو فاکتور سادگی و کاربردی بودن استفاده از روش‌های درختی به‌منظور برآورد تبخیر-تعرق مرجع، در اقلیم‌هایی مشابه با منطقه مورد مطالعه این روش‌ها پیشنهاد می‌گردند. عیب اساسی روش‌های مورد استفاده در این تحقیق این است که قابل تعمیم به کلیه مناطق با اقلیم‌های متفاوت نیستند. با توجه به نتایج مذکور، پیشنهاد می‌گردد روش‌های مورد استفاده در مطالعه حاضر درحوضه‌های آبریز

References

- Ajayram, K. A., Jegadeeshwaran, R., Sakthivel, G., Sivakumar, R. and Patange, A. D. (2021). Condition monitoring of carbide and non-carbide coated tool insert using decision tree and random tree – A statistical learning. Mater. Today: Proceedings., DOI: 10.1016/j.matpr.2021.02.065.
- Allen, R.G., Pereira, L. S., Raes, D. and Smith, M. (1998). Crop evapotranspiration (Guidelines for computing crop water requirements). FAO irrigation and drainage Paper No. 56. Food and Agricultural Organization of the United Nations, Rome, 300p.
- Ayaz, A., Rajesh, M., Kumar, S. and Rehana, S. (2021). Estimation of reference evapotranspiration using machine learning models with limited data. AIMS Geosci., 7(3), 268–290. DOI: 10.3934/geosci.2021016.
- Bui, D. T., Ho, T. C., Pradhan, B., Pham, B. T., Nhu, V. H. and Revhaug, I. (2016). GIS-based modeling of rainfall-induced landslides using data mining based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environ. Earth Sci., 75(14), 1101. DOI : 10.1007/s12665-016-5919-4.
- Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y. and Wang, W. (2019). Flood susceptibility modeling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. J. Hydrol., 864-873. DOI: 10.1016/j.jhydrol.2019.05.089.
- Demirci, M. (2019). Prediction of precipitation flow relationship using support vector machines and M5 decision tree methods. DUMF Eng. J., 10: 1113–1124.
- Hamoud, A., Hashim, A. S. and Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. Int. J. Interact. Multimed. Artif. Intell., 5- 26. DOI: 10.9781/ijimai.2018.02.004.
- Hansen, S. (1984). Estimation of potential and actual evapotranspiration. Nordic Hydrol., 15, 205–212. DOI: 10.2166/nh.1984.0017.
- Hashemi, E., Ahmadpari, H. and Kohneh, K. (2017). Comparison of different methods of estimating potential evapotranspiration with FAO Penman-Mantith method (Case study: Sepidan region). Nivar, 41, 13-22. DOI: 10.30467/nivar.2017.51886. [In Persian]
- Hozhabr, H., Moazed, H. and Shokrikhoochak, S. (2014). Estimation of reference evapotranspiration (ET_o) using empirical models, artificial neural network modeling and their comparison with lysimeter data in Urmia Kahrizi Station. Journal of Irrigation and Water Engineering., 15, 13-25 [In Persian].
- Khoshhal J., Zareh H. and Joshani A. (2015). Different methods for estimating reference evapotranspiration by FAO evaporation pan method in the east and southeast of the country. Quart. J. Nat. Geogr., 8(28), 1-16. [In Persian]
- Moradi, A., Ziaean, A. H. (2019). Evaluation of eleven models for reference crop evapotranspiration estimation in Haji Abad region of Hormozgan. Iran. J. Irrig. Drain., 6, 1623-1637 [In Persian].



- Nasrolahi, A., Sabzevari, Y., Sharifipour, M. and Shahinejad. (2020). Evaluation of bayesian network and support vector machine models in estimation of reference evapotranspiration (Case Study: Khorramabad). *Iran. J. Irrig. Drain.*, 2, 522-534 [In Persian].
- Oudina L, Hervieua F, Michela C, Perrina C, Andre´assiana V, Anctilb, F. and Loumagnea, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modeling. *Journal of Hydrology.*, 303(1–4), 290-306. DOI: 10.1016/j.jhydrol.2004.08.026.
- Piri, H. and Pouzan, M. T. (2019). Evaluation of 24 evapotranspiration models of reference plants in different climates of Iran. *Echohydrol.*, 6, 611-622 [In Persian].
- Priestley, C. and Taylor, R. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Month. Weather Rev.*, 100(2), 81- 92.
- Salam, R. and Towfiqul Islam, A. R. (2020). Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *Journal Hydrology*, DOI: 10.1016/j.jhydrol.2020.125241.
- Samadianfard, S., Salarifar, M., Javidan, S. and Mikaeili, F. (2020). Estimation of daily reference evapotranspiration in humid climates using data-driven methods of gaussian process regression, support vector regression and random forest. *Environ. Water Eng.*, 4, 360-373. DOI: 10.22034/jewe.2020.241690.1394 [In Persian].
- Sattari, M. T., Apaydin, H. and Shamshirband, S. (2020). Performance evaluation of deep learning-based gated recurrent units (GRUs) and tree-based models for estimating ETo by using limited meteorological variables. *Mathemat.*, 8. DOI:10.3390/math8060972.
- Sattari, M. T., Apaydin, H., Shamshirband, S., mosavi, A. and Prasad, R. (2021). Comparative analysis of kernel-based versus ANN and deep learning methods in monthly reference evapotranspiration estimation. *Hydrol. Earth Syst. Sci.*, 25, 603-618. DOI: 10.5194/hess-25-603-2021.
- Siasar, H. and Honar, T. (2019). Application of support vector machine, CHAID and random forest models, in estimated daily reference evapotranspiration in northern Sistan and Baluchestan Province. *Iran. J. Irrig. Drain.*, 2, 378-388 [In Persian].
- Siasar, H., Honar, T. and Abdolahipour, M. (2020). Comparing of generalized linear models, random forest and gradient boosting trees in estimation of reference crop evapotranspiration (Case study: the Sistan Plain). *J. Water Soil Sci.*, 23, 395-410 [In Persian].
- Suchetana, B., Rajagopalan, B. and Silverstein, J. (2017). Assessment of wastewater treatment facility compliance with decreasing ammonia discharge limits using a regression tree model. *Sci. Total. Environ.*, 598, 249–257. DOI: 10.1016/j.scitotenv.2017.03.236.
- Wang, Y. and Witten, I. H. (1997). Inducing model trees for continuous classes. in *Proceedings of the Ninth European Conference on Machine Learning*. Prague, Czech Republic: Springer, 128-137.
- Wu, L. and Fun, 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. DOI: 10.1371/journal.pone.0217520.
- Wu, L., Peng, Y., Fan, J. and Wang, Y. (2019). Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on crossstation and synthetic data. *Hydrol. Res.*, 50(6), 1730-1750. DOI: 10.2166/nh.2019.060.

How to cite this paper:

Javidan, S., Sattari, M. T. and Mehrabi, A. (2022). Comparative analysis of estimating monthly reference evapotranspiration using kernel and tree-based data mining models versus empirical methods. *Environ. Water Eng.*, 8(4), 908–922. DOI: 10.22034/jewe.2022.317241.1686

