

Research Article

Prediction of Daily Evapotranspiration Using the Strategy of Combining Tree Models with Empirical Hargreaves Method

M Sharafi¹, E Abdi², M Mohebbiyan³, S Samadianfard^{4*}

Received: March 14, 2023

Accepted: September 3, 2023

Revised: August 19, 2023

Published online: June 21, 2024

1- PhD candidate, Water Engineering Dept., Faculty of Agriculture, Urmia University, Urmia, Iran

2- MSc student, Water Engineering Dept., Faculty of Agriculture, University of Tabriz, Iran

3- BSc student, Water Engineering Dept., Faculty of Agriculture, University of Tabriz, Iran

4- Assoc. Prof., Water Engineering Dept., Faculty of Agriculture, University of Tabriz, Iran

*Corresponding Author, Email: s.samadian@tabrizu.ac.ir

Abstract

Background and Objectives

The constant need to increase agricultural production, along with more and more frequent drought events in our country, requires a more accurate assessment of irrigation consumption and thus a more accurate estimate of actual evapotranspiration. Prediction of water consumption over agricultural areas is important for agricultural water resources planning, management, and regulation. It leads to the establishment of a sustainable water balance, mitigates the impacts of water scarcity, as well as prevents the overusing and wasting of precious water resources. As evapotranspiration is a major consumptive use of irrigation water and rainwater on agricultural lands, improvements of water use efficiency and sustainable water management in agriculture must be based on the accurate estimation of ET. Irrigated agriculture is expected to produce more crops with less water consumption in the future. Therefore, accurate forecasting of water demand along with sustainable management and more efficient methods to meet the growing demand under scarce water resources is necessary. The models used to predict evapotranspiration should be used in different regions with different climates to evaluate their performance. Therefore, in this research, tree models and Hargreaves were used in Yazd and West Azerbaijan provinces, which have different climates, in order to evaluate the performance of the models used.

Methodology

In recent years, water management issues have been addressed using models derived from artificial intelligence research. The use of combined models has made significant progress in recent years. Combined models are able to perform processing in a short period of time and at the same time with high accuracy. Using these models, the main challenging aspects are represented by the selection of the best possible algorithm, the selection of suitable representative variables and the availability of suitable data sets. Therefore, in this study, the ability of tree models (M5P and RF) with Hargreaves model (Hs) for estimating the daily evapotranspiration of Urmia and Yazd stations was investigated during the period of 2000-2021. The noteworthy point is that in the combined tree-Hargreaves model, the used tree models were used as input to the Hargreaves model. The combined model has been used for the first time in this research and the use of this model can predict daily evapotranspiration with appropriate accuracy. Also, the model's results have been evaluated using 5 evaluation criteria: coefficient of determination, root mean square error, Nash-Sutcliffe coefficient, and Wilmot's agreement index.



Findings

The best scenario in all the models used was the model whose input included the parameters of minimum temperature, maximum temperature, relative humidity, wind speed, and sunshine hours. The results obtained at the Yazd station showed that the fifth scenario of the M5P-Hs model and then the fifth scenario of the RF-Hs model, respectively with an error of (mm day^{-1}) 0.24 and (mm day^{-1}) 0.26 had the highest accuracy. In the Urmia station, the fifth scenario of the RF-Hs model with an error of (mm day^{-1}) of 0.27 had the best performance, and after that, the fifth scenario of the RF-Hs model with an error of (mm day^{-1}) of 0.33 had good accuracy. The comparison between the results of the Urmia and Yazd stations shows that the coefficient of determination and the root mean square error for the Urmia station was higher than the Yazd station in all scenarios. On the other hand, with the increase of the data range, the coefficient of determination index will have a higher value and the root mean square error index will have a lower value. Therefore, the reason for the Decreasing of determination coefficient and the error rate in Yazd station can be attributed to the limited range of parameters used. Finally, tree models have increased the accuracy of the Hargreaves model in this research.

Conclusion

Finally, the RF, M5P, RF-Hs and M5P-Hs models were able to predict daily evapotranspiration values in the shortest time and with the highest accuracy. However, the results showed that the lower the model inputs, the weaker the model prediction. The results of this research showed that the combination of tree models with Hargreaves model is able to predict daily evapotranspiration values with high accuracy compared to individual models. The results of this research showed that the wind speed parameter is one of the most important meteorological parameters needed in estimating daily evapotranspiration, so adding this parameter results in the highest accuracy in all models. Also, due to the important role of wind speed in predicting daily evapotranspiration values and the unavailability of the maximum wind speed parameter in this research, it is recommended to use the maximum wind speed parameter as one of the model inputs for further studies.

Keywords: Agriculture, Evapotranspiration, Hargreaves, Prediction, Tree model

مقاله پژوهشی

پیش‌بینی مقادیر تبخیر-تعرق روزانه با استفاده از راهبرد ترکیب مدل‌های درختی با روش تجربی هارگریوز

میلاذ شرفی^۱، عرفان عبدی^۲، مهدی محبیان^۳، سعید صمدیان فرد^{۴*}

تاریخ دریافت: ۱۴۰۱/۱۲/۲۳

تاریخ پذیرش: ۱۴۰۲/۰۶/۱۲

تاریخ ویرایش: ۱۴۰۲/۰۵/۲۸

تاریخ انتشار آنلاین: ۱۴۰۳/۰۴/۰۱

۱- دانشجوی دکتری، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه

۲- دانشجوی کارشناسی ارشد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۳- دانشجوی کارشناسی، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۴- دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

*مسئول مکاتبات، پست الکترونیکی: s.amadian@tabrizu.ac.ir

چکیده

نیاز دائمی به افزایش تولیدات کشاورزی، همراه با رویدادهای خشکسالی بیشتر و مکرر در کشور، مستلزم ارزیابی دقیق‌تری از نیازهای آبیاری و در نتیجه برآورد دقیق‌تر تبخیر و تعرق واقعی است. در سال‌های اخیر، چندین موضوع مدیریت آب با استفاده از مدل‌های به‌دست آمده از تحقیقات هوش مصنوعی مورد توجه قرار گرفته است. هنگام استفاده از این مدل‌ها، جنبه‌های چالش‌برانگیز اصلی با انتخاب بهترین الگوریتم ممکن، انتخاب متغیرهای معرف مناسب و در دسترس بودن مجموعه داده‌های مناسب نشان داده می‌شوند. بنابراین، در این مطالعه توانایی مدل‌های درختی (M5P) و RF با مدل هارگریوز (Hs) در برآورد تبخیر-تعرق روزانه در ایستگاه‌های ارومیه و یزد، طی دوره ۲۰۲۱-۲۰۰۰ با استفاده از چهار معیار آماری مورد ارزیابی قرار گرفت. در تمام مدل‌های بکار گرفته شده، سناریوی برتر مدلی بود که ورودی آن شامل پارامترهای حداقل دما، حداکثر دما، رطوبت نسبی، سرعت باد و ساعات آفتابی بود. نتایج به‌دست آمده در ایستگاه یزد نشان داد که سناریو پنجم مدل M5P-Hs و سپس سناریو پنجم مدل RF-Hs به ترتیب با داشتن خطای (mm day^{-1}) ۰/۲۴ و ۰/۲۶ بالاترین دقت را داشتند. در ایستگاه ارومیه نیز سناریو پنجم مدل RF-Hs با داشتن خطای (mm day^{-1}) ۰/۲۷ بهترین عملکرد را داشته و پس از آن سناریو پنجم مدل RF-Hs با داشتن خطای (mm day^{-1}) ۰/۳۳ دقت مناسبی را ارائه داد. نتایج حاصل از این پژوهش نشان داد که پارامتر سرعت باد از مهم‌ترین پارامترهای هواشناسی مورد نیاز در برآورد تبخیر-تعرق روزانه می‌باشد، به‌طوری‌که افزودن این پارامتر سبب کاهش خطا و بهبود دقت پیش‌بینی مدل‌ها گردید.

واژه‌های کلیدی: کشاورزی، تبخیر-تعرق، هارگریوز، پیش‌بینی، مدل درختی

مقدمه

تبخیر به عنوان تبدیل آب مایع به بخار آب به دلیل اختلاف فشار بین سیستم زمین-اتمسفر تعریف می‌شود (کیم و همکاران ۲۰۱۵). به طور کلی اصطلاح تبخیر-تعرق از دست دادن آب به صورت بخار از سطح خاک و گیاه را بیان می‌کند. تبخیر-تعرق یک جزء ضروری از چرخه هیدرولوژیکی است و به طور گسترده در پروژه‌های آبیاری و منابع آب استفاده می‌شود (مولینا و همکاران ۲۰۱۵، فنگ و همکاران ۲۰۲۲). میزان تبخیر در مناطق خشک و نیمه خشک بسیار زیاد است (البلاسمه و همکاران ۲۰۲۲). بنابراین، برآورد دقیق تبخیر برای برنامه ریزی و مدیریت پایدار منابع آب، به ویژه برای شیوه‌های آبیاری، عملیات دریاچه‌ها و مخازن، بودجه بندی آب و مطالعات مربوط به مدل سازی هیدرولوژیکی حیاتی است (کیسی و هیدم ۲۰۱۹، مالک و همکاران ۲۰۲۰)، علاوه بر این، سیستم‌های کشاورزی مانند برنامه ریزی یا شبیه سازی محصول و برنامه ریزی آبیاری تا حد زیادی به ارزیابی دقیق تبخیر بستگی دارد (آکوا و همکاران ۲۰۱۸). متغیرهای مختلف هواشناسی، مانند تابش خورشیدی، دمای هوا، رطوبت نسبی و سرعت باد، می‌توانند به طور تعاملی بر مولفه تبخیر-تعرق تأثیر بگذارند. بنابراین، تاکنون تلاش‌های زیادی برای استخراج روابط ریاضی یا رگرسیونی بین تبخیر و عوامل محرک آن صورت گرفته است (آلن و همکاران ۱۹۹۸، فنگ و همکاران ۲۰۱۶، وانگ و همکاران ۲۰۱۹، یو و همکاران ۲۰۱۹، چن و همکاران ۲۰۲۰). با توجه به اینکه پیچیدگی سیستم‌های هیدرولوژیکی از نظر تعیین تمام پارامترهای دخیل و ناقص بودن اطلاعات آماری، امکان مدل سازی کامل سیستم‌های هیدرولوژیکی را غیرممکن می‌سازد. در نتیجه استفاده از مدل سازی سیستمی که مبتنی بر روابط ریاضی باشد، مورد توجه قرار می‌گیرد

(کیسی و همکاران ۲۰۱۶). برتی و همکاران (۲۰۱۴) در مطالعه‌ای به ارزیابی تبخیر-تعرق مرجع با روش هارگریوز در شمال شرقی ایتالیا پرداختند. برای انجام مطالعه، مجموعه داده‌های آب و هوای کامل مقادیر روزانه جمع آوری شده در طول دوره ۲۰۰۶-۱۹۹۴ از ۳۵ ایستگاه کشاورزی و هواشناسی واقع در دشت ونتو استفاده شد. ده ایستگاه برای کالیبراسیون معادلات هارگریوز تنظیم شده، به منظور نمایش مناطق مختلف دشت انتخاب شدند، در حالی که ۲۵ ایستگاه دیگر برای اعتبارسنجی استفاده شدند. نتایج این مطالعه نشان داد که با کالیبراسیون مناسب، معادله هارگریوز^۱ توانست مقدار خطا را از ۱۸/۹ درصد به ۲/۶ درصد کاهش داده و سبب افزایش دقت تخمین تبخیر-تعرق مرجع شود. فنگ و همکاران (۲۰۱۷) در مطالعه‌ای به ارزیابی مدل‌های جنگل تصادفی (RF^۲) و رگرسیون شبکه‌های عصبی تعمیم یافته (GRNN^۳) برای مدل سازی تبخیر-تعرق مرجع روزانه در جنوب غربی چین طی سال‌های ۲۰۰۹ تا ۲۰۱۴ پرداختند. نتایج این مطالعه نشان داد با اینکه هر دو مدل RF و GRNN برای تخمین تبخیر-تعرق مرجع روزانه عمل مناسبی را از خود نشان دادند اما مدل RF به طور کلی بهتر از GRNN عمل کرد. البلتقی و همکاران (۲۰۲۲) در تحقیقی به برآورد شاخص تبخیر-تعرق بارندگی استاندارد شده با استفاده از تکنیک‌های داده-محور در بنگلادش طی سال‌های ۲۰۱۳-۱۹۸۰ پرداختند. نتایج این تحقیق نشان داد مدل M5P با داشتن بالاترین ضریب همبستگی (۰/۷۵-۰/۹۸) عملکرد بهتری را نسبت به سایر مدل‌های مورد استفاده داشته است. تبخیر-تعرق به عنوان یکی از مولفه‌های اصلی چرخه هیدرولوژیکی، در معرض تغییرات غیرخطی است و کمبود داده‌های تبخیر اندازه گیری شده در بسیاری از ایستگاه‌های هواشناسی برای دوره‌های زمانی خاص وجود دارد،

^۱ Hargreaves equation

^۲ Random forest

^۳ Generalized regression neural networks

درصد داده‌ها برای آموزش مدل‌ها و ۳۰ درصد داده‌ها برای مرحله آزمون مورد استفاده قرار گرفتند. شهرستان ارومیه به مختصات ۳۷ درجه و ۳۲ دقیقه عرض شمالی و ۴۵ درجه و ۴ دقیقه طول شرقی در ارتفاع ۱۳۳۲ متر از دریا در مرکز استان آذربایجان غربی واقع شده است (روستایی و همکاران ۲۰۲۰). شهرستان یزد نیز با داشتن مختصات ۵۲ درجه و ۵۰ دقیقه طول شرقی و ۲۹ درجه ۵۲ دقیقه عرض شمالی با ارتفاع ۱۲۱۶ متر از سطح دریا در مرکز استان یزد قرار دارد (مالکی و مودت ۲۰۱۳). برخورداری از شرایط منحصر به فرد جغرافیایی و طبیعی باعث شده‌است که شهرستان ارومیه از نظر کشاورزی و صنایع وابسته به آن توسعه فراوانی پیدا کند و به‌عنوان قطب برتر کشاورزی در کشور شناخته شده و در تولید بسیاری از محصولات کشاورزی رتبه اول کشور را دارا می‌باشد. لذا پیش‌بینی هرچه دقیق‌تر پارامتر تبخیر-تعرق سبب مدیریت بهتر آب در این بخش پرمصرف خواهد شد. در مقابل شهرستان یزد با داشتن بارندگی کمتر از (1 mm year^{-1}) ، ۱۰۰، بیش از (mm day^{-1}) ۳۲۰۰ تبخیر سالانه داشته و لذا پیش‌بینی دقیق‌تر پارامتر تبخیر-تعرق سبب بهبود مدیریت منابع آبی خواهد شد. نقشه موقعیت جغرافیایی ایستگاه‌های مورد مطالعه در شکل ۱ آورده شده است.

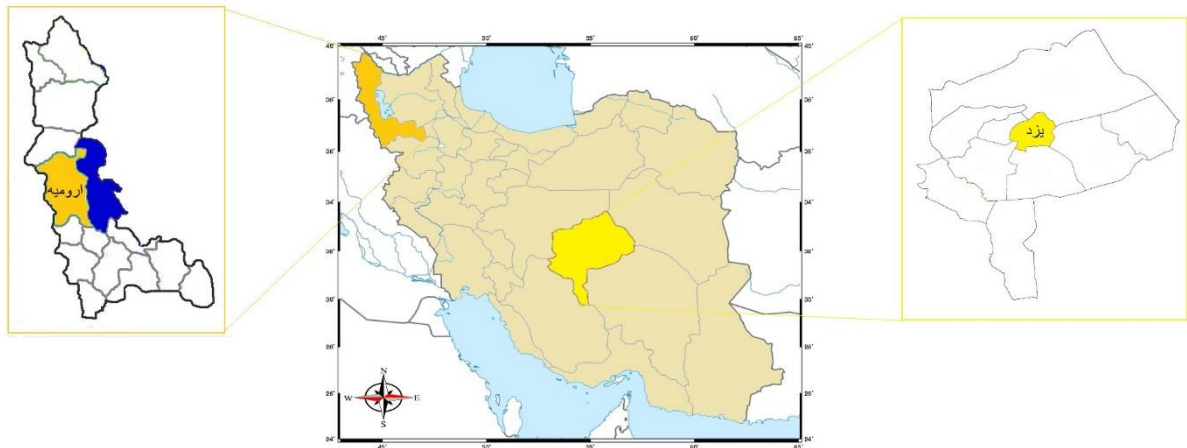
جدول ۱ پارامترهای آماری روزانه مورد استفاده در طی دوره آماری را نشان می‌دهد. پارامترهای به‌کار رفته در هر دو ایستگاه توسط مستخرج از داده‌های هواشناسی بوده و توسط نرم‌افزار WEKA نسخه ۳.۸ محاسبه شده‌اند.

مشکلی که با توزیع نابرابر ایستگاه‌ها تشدید می‌شود. با توجه به دقت مناسب مدل‌های درختی و هارگیوز در مطالعات پیشین استفاده از مدلی ترکیبی با دقت و عملکرد بالا بیش از پیش اهمیت پیدا می‌کند. در این تحقیق ابتدا با استفاده از روش‌های درختی و مدل ترکیبی هارگیوز-درختی مقادیر تبخیر روزانه پیش‌بینی شده و سپس با معیارهای ارزیابی مختلف دقت مدل‌ها بررسی گردید. هدف از بکارگیری مدل درختی با روش هارگیوز، استفاده از مدلی ترکیبی با قابلیت اطمینان بالا بوده تا بتوان با استفاده از مدل ترکیبی مقدار دقت پیش-بینی را نسبت به مدل‌های منفرد درختی و هارگیوز بهبود داده و در نهایت منجر به پیش‌بینی هرچه دقیق‌تر مقدار تبخیر شود. همچنین بررسی مطالعات پیشین نشان داد که تاکنون مطالعه‌ای درباره‌ی ترکیب مدل‌های درختی با مدل هارگیوز انجام نشده است لذا مطالعه حاضر برای اولین بار به پیش‌بینی مقادیر تبخیر-تعرق روزانه با استفاده از ترکیب مدل‌های درختی با روش تجربی هارگیوز می‌پردازد. با در نظر گرفتن موارد مذکور، هدف از مطالعه حاضر، پیش‌بینی مقادیر تبخیر-تعرق روزانه با استفاده از روش‌های هارگیوز و درختی در ایستگاه‌های ارومیه و یزد می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه

در تحقیق حاضر مقادیر تبخیر-تعرق روزانه در طی دوره آماری ۲۱ ساله در سال‌های ۲۰۲۱-۲۰۰۰ در ایستگاه‌های ارومیه و یزد بکار برده شد. همچنین ۷۰



شکل ۱- نقشه موقعیت جغرافیایی ایستگاه‌های مورد مطالعه.

جدول ۱- پارامترهای آماری روزانه داده های مورد استفاده.

داده های مورد استفاده							
ایستگاه	پارامترهای آماری	سرعت باد (m s ⁻¹)	ساعات آفتابی (hr)	رطوبت نسبی (%)	دمای ماکزیمم (°C)	دمای مینیمم (°C)	تبخیر-تعرق (mm day ⁻¹)
ارومیه	میانگین	۲/۷۰۶	۸/۲۳۹	۵۶/۹۴	۱۹/۵۲۷	۵/۶۲۴	۲/۳۶۵
	مینیمم	۰/۱۲۵	۰	۲۰/۳۷۵	-۵/۶	-۱۶/۲	۰/۲۲۰
	ماکزیمم	۷/۶۲۵	۱۳/۸	۹۹/۶۲۵	۲۹/۶	۲۳/۷	۸/۷۸۷
	انحراف استاندارد	۰/۷۸۳	۳/۹۱۹	۱۶/۹۳	۱۰/۶۷۸	۸/۰۰	۱/۸۴۱
یزد	میانگین	۲/۵۰۲	۹/۵۱	۲۵/۹۹۱	۲۷/۸۲	۱۴/۲۷	۲/۰۴۳
	مینیمم	۰/۵	۰	۴/۳۷۵	-۱/۸	-۵/۴	۰/۳۱۷
	ماکزیمم	۸/۱۲۵	۱۳/۵	۹۶/۶۲۵	۴۴/۹	۳۳/۱	۶/۲۸۳
	انحراف استاندارد	۰/۸۹۶	۲/۷۶۰	۱۷/۰۰۲	۹/۶۹	۹/۰۲	۰/۹۳۱

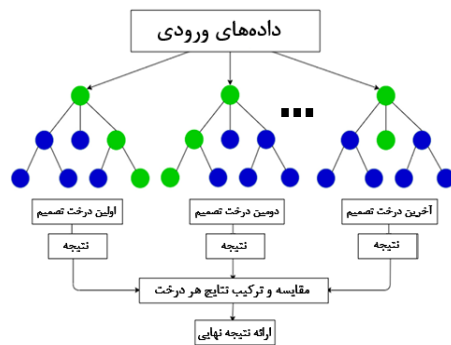
روزانه (ET) به‌عنوان پارامتر هدف می‌باشند. نحوه انتخاب ترکیبات مختلف ورودی، برای تخمین مقدار تبخیر-تعرق روزانه در مدل‌های مورد مطالعه براساس داشتن همبستگی بالاتر نسبت به سایر ترکیبات، صورت گرفته است.

جدول ۲ سناریوهای مختلفی را که به‌عنوان ورودی و خروجی مدل‌ها در نظر گرفته شده‌اند، نشان می‌دهد. پارامترهای حداقل دما (T_{min})، حداکثر دما (T_{max})، رطوبت نسبی (RH)، سرعت باد (U) و ساعات آفتابی (SSH) به‌عنوان پارامترهای ورودی و تبخیر-تعرق

جدول ۲- ترکیب های مختلف ورودی، برای تخمین مقدار تبخیر روزانه در مدل های مورد مطالعه.

سناریو	پارامترهای ورودی				خروجی
۱	T_{max}				ET
۲	T_{max}	T_{min}			ET
۳	T_{max}	T_{min}	SSH		ET
۴	T_{max}	T_{min}	SSH	RH	ET
۵	T_{max}	T_{min}	SSH	RH	U

استفاده شد. همچنین شکل ۲ نیز نمای کلی مدل جنگل تصادفی را نشان می دهد.



شکل ۲- نمای کلی مدل جنگل تصادفی.

الگوریتم M5P

الگوریتم درختی M5P شامل مجموعه ای از نمونه های آموزشی (T) است که هر نمونه آموزشی دارای مجموعه ای از ویژگی هاست که مقادیر ورودی نامیده می شوند و دارای هدف متناظری هستند که در واقع مقدار خروجی مدل است. هر الگوریتم دارای دو مشخصه n و M است که به ترتیب گره و مدل نام دارند. در این تحقیق، ما از الگوریتم M5P مشابه مطالعات موجود استفاده می کنیم (سلامتین و دولال ۲۰۰۳، سلامتین و سبیک ۲۰۰۴). یک درخت تصمیم معمولاً از چهار بخش ریشه، شاخه، گره ها و برگ ها تشکیل شده است که گره ها با دایره نشان داده می شوند و شاخه ها نشان دهنده اتصال بین گره ها می باشند. روش M5P قادر است مجموعه داده های

جنگل تصادفی (RF^۱)

مدل جنگل تصادفی، اولین بار توسط Breimen و همکاران تعیین شد (بریمن ۲۰۰۱). مدل جنگل تصادفی مجموعه ای از فرآیندهای تصادفی برای طبقه بندی داده ها است. که در حقیقت از ترکیب مدل های پایه کمتر پیش بینی کننده برای تولید مدل های پیش بینی با دقت بهتر بکار می رود. مدل های RF به دلیل ماهیت ساده، مفروضات کم و عملکرد بالا، به طور گسترده ای در یادگیری ماشین (ML)^۲ استفاده می شوند. جنگل تصادفی روشی است برای میانگین گیری با هدف کاهش واریانس با استفاده از درخت های تصمیم عمیقی که از قسمت های مختلف داده های آموزشی ایجاد شده باشند. در این روش معمولاً افزایش جزئی سوگیری و از دست رفتن کمی از قابلیت تفسیر اتفاق افتاده اما در کل عملکرد مدل را به طور چشمگیری افزایش خواهد داد. یک جنگل رگرسیون قدرت پیش بینی یک درخت رگرسیون منفرد را ندارد. به طوریکه یک درخت تنها به یک معیار تقسیم شده و به مجموعه داده آموزشی بسیار حساس است. با این حال تغییرات جزئی در مجموعه داده و معیار تقسیم می تواند ساختارهای مختلف درخت را ترسیم کرده و توضیحات متفاوتی را ارائه دهد. بنابراین، مدل های RF متغیرها را بر اساس اهمیت آنها برای دستیابی به بهترین مدل طبقه بندی می کنند (بریمن ۱۹۹۶، بریمن ۲۰۰۱). در این مطالعه برای اجرای مدل RF از نرم افزار WEKA نسخه ۳.۸

¹ Random forest

² Machine learning

محاسبه تابش برون جوی به منبع اصلی مراجعه گردد (هارگریوز و آلن ۲۰۰۳). لازم به ذکر است که در دماهای معمولی هر $2MG/47$ معادل (mm) ۱ تبخیر روزانه در نظر گرفته می‌شود.

معیارهای ارزیابی مدل

در این پژوهش برای ارزیابی عملکرد سناریوهای مختلف تعریف شده برای روش‌های مورد مطالعه از پارامترهای آماری ضریب تعیین (R^2)^۲، جذر میانگین مربعات خطا ($RMSE$)^۴، ضریب نش‌ساتکلیف (NS)^۵ و شاخص توافق ویلموت (WI)^۶ استفاده شده است. مقدار ضریب تعیین بین ۰ تا ۱ متغیر بوده و هرچه به یک نزدیکتر باشد دقت مدل افزایش می‌یابد. معیار جذر میانگین مربعات خطا نشان‌دهنده میزان خطای مدل بوده و هرچه به صفر نزدیکتر باشد مدل دارای خطای کمتری خواهد بود. معیارهای ضریب نش‌ساتکلیف و شاخص توافق ویلموت نیز بین -۱ تا ۱ متغیر بوده و هرچقدر به یک نزدیکتر باشند دقت مدل افزایش خواهد یافت.

$$R^2 = \left(\frac{\sum_{i=1}^N (O_i - \bar{O}) - \frac{1}{N} (P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \right)^2 \quad [۲]$$

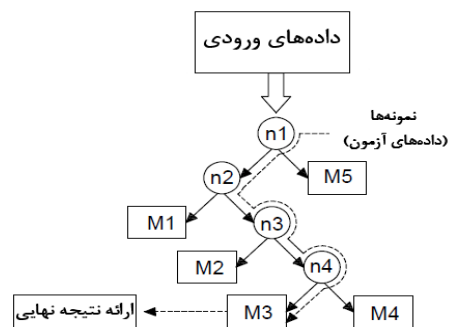
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad [۳]$$

$$NS = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \right] \quad [۴]$$

$$WI = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right] \quad [۵]$$

در روابط ۲-۵، پارامترهای O_i و P_i به ترتیب مقادیر پیش‌بینی و مشاهداتی، پارامترهای \bar{O} و \bar{P} به ترتیب

بزرگ را به همراه داده‌های مفقود شده با تقسیم فضاها و ورودی به زیر فضاها کوچکتر بازیابی کند. به‌طور کلی، حداقل تعداد نمونه، اندازه دسته‌ای، درختان رگرسیون ساخته‌شده، تعداد رقم‌های اعشاری و قابلیت‌های چاپ نشده و کنترل نشده همگی از مزایای مدل‌های M5P هستند. برای مطالعات بیشتر به خسروی و همکاران (۲۰۱۸) مراجعه گردد. در این مطالعه برای اجرای مدل RF از نرم‌افزار WEKA نسخه ۳.۸ استفاده شد. همچنین شکل ۳ نیز نمای کلی مدل درختی M5P را نشان می‌دهد.



شکل ۳- نمای کلی مدل درختی M5P.

مدل هارگریوز^۱

این مدل توسط هارگریوز و آلن پیشنهاد گردید (هارگریوز و آلن ۲۰۰۳). در این پژوهش، از این مدل تحت عنوان مدل هارگریوز (Hs) نام برده می‌شود.

$$ET = 0.0023R_a (T_a + 17.8) (T_{max} - T_{min})^{0.5} \quad [۷]$$

در رابطه ۸، ET تبخیر تعرق گیاه مرجع ($mm \ day^{-1}$)

^۱، R_a معادل تابش برون جوی ($mm \ day^{-1}$)، T_a و T_{min} به ترتیب متوسط دما، متوسط بیشینه و متوسط کمینه دمای هوا ($^{\circ}C$) می‌باشند. برای اطلاع از نحوه

^۱ Hargreaves model

^۲ Megajoule

^۳ Coefficient of determination

^۴ Root mean squared error

^۵ Nash-Sutcliffe coefficient

^۶ Willmott's index of agreement

۰/۴ و ۰/۳۸) عملکرد بهتری نسبت به سایر مدل‌های RF و M5P از خود نشان دادند.

مقایسه ترکیبات مختلف برای مدل ترکیبی M5P-Hs در ایستگاه یزد نشان داد که مدل M5P-Hs5 با داشتن ضریب تعیین و نش ساتکلیف ۰/۹۴، جذر میانگین مربعات خطا (mm day^{-1}) ۰/۲۴ و شاخص ویلموت ۰/۹۸ بالاترین عملکرد را در میان تمام مدل‌های ایستگاه یزد از خود نشان داد. همچنین مدل M5P-Hs5 توانست میزان خطای مدل منفرد M5P1 را تا میزان ۸۵/۷۱ درصد کاهش دهد، که نشان از عملکرد قابل قبول مدل می‌باشد. برای مدل‌های RF-Hs نیز سناریو پنجم توانست دقت بالاتری را با داشتن ضریب تعیین و نش ساتکلیف ۰/۹۲، جذر میانگین مربعات خطا (mm day^{-1}) ۰/۲۶ و شاخص ویلموت ۰/۹۸ در مقایسه با سایر ترکیبات RF و RF-Hs نشان دهد. همچنین مدل RF-Hs5 توانست میزان خطای مدل منفرد RF1 را تا ۸۴/۴۴ درصد کاهش دهد. مقایسه مدل‌های M5P نیز نشان داد که مدل M5P5 مقدار خطای کمتری (mm day^{-1}) ۰/۳۵ را نسبت به سایر ترکیبات M5P داشته و توانست خطای مدل M5P1 را تا ۵۲/۶۳ درصد کاهش دهد. در نهایت در میان مدل‌های RF مدل RF5 خطای کمتری (mm day^{-1}) ۰/۳۶ را نسبت به سایر ترکیبات RF از خود نشان داد. به‌طوریکه سناریو پنجم توانست میزان خطای سناریو اول را به میزان ۵۶ درصد کاهش دهد. مقایسه بین نتایج ایستگاه‌های ارومیه و یزد نشان می‌دهد که در تمام سناریوها مقادیر همه شاخص-نشان‌های عملکرد برای ایستگاه ارومیه بیشتر از ایستگاه یزد بوده است. با توجه به جدول ۱ در پارامترهای حداقل دما، رطوبت نسبی، ساعات آفتابی و تبخیر-تفرق مقدار محدوده بین مینیمم و ماکزیمم هر پارامتر در ایستگاه ارومیه بیشتر از ایستگاه یزد بوده است. از طرفی با افزایش بازه داده‌ها شاخص ضریب تعیین دارای مقدار بیشتر و شاخص جذر میانگین مربعات خطا به‌طور بالعکس دارای مقدار کمتری خواهد بود (منارد ۲۰۰۰). لذا علت کاهش ضریب تعیین و میزان خطا در ایستگاه یزد

میانگین مقادیر پیش‌بینی و مشاهداتی می‌باشند. همچنین i نشانگر آمین داده و N نیز نشانگر تعداد کل داده‌ها می‌باشد.

نتایج و بحث

در این مطالعه، پنج ترکیب مختلف از متغیرهای پیش‌بینی‌کننده پارامترهای حداقل دما (T_{\min})، حداکثر دما (T_{\max})، رطوبت نسبی (RH)، سرعت باد (U) و ساعات آفتابی (SSH)، به‌عنوان ورودی مدل‌ها در جدول ۲ در نظر گرفته شد. در این بخش، مدل‌های RF، RF-Hs، M5P و M5P-Hs بر روی پنج مجموعه داده برای پیش‌بینی مقادیر تبخیر-تفرق روزانه استفاده شد. برای ایجاد مدل‌های ترکیبی، خروجی روش تجربی هارگریوز به-عنوان ورودی مدل‌های درختی در مدل‌های RF-Hs و M5P-Hs در نظر گرفته شدند. تطابق بین مقادیر اندازه-گیری شده و پیش‌بینی شده تبخیر در جدول ۳ از نظر ضریب تعیین، جذر میانگین مربعات خطا، ضریب نش ساتکلیف و شاخص توافق ویلموت، طی مراحل اعتبارسنجی گزارش شده است.

ارزیابی ترکیبات مدل RF-Hs برای ایستگاه ارومیه نشان داد که مدل RF-Hs5 با داشتن ضریب تعیین ۰/۹۹، جذر میانگین مربعات خطا (mm day^{-1}) ۰/۲۷، ضریب نش ساتکلیف ۰/۹۹ و شاخص ویلموت ۰/۹۹ کم‌ترین خطا را بین تمام مدل‌های ایستگاه ارومیه داشته و بالاترین دقت را بدست آورد. همچنین از میان مدل‌های M5P-Hs، مدل M5P-Hs5 با داشتن ضریب تعیین و ضریب نش ساتکلیف ۰/۹۷، جذر میانگین مربعات خطا (mm day^{-1}) ۰/۳۳ و شاخص ویلموت ۰/۹۹ دقت مناسبی را پس از مدل RF-Hs5 نشان داد. در نهایت ارزیابی مدل‌های درختی RF و M5P نیز نشان داد که سناریو پنجم هر دو مدل به‌ترتیب با داشتن جذر میانگین مربعات خطای کمتر (mm day^{-1})

ورودی داشته است. به طوریکه با افزودن پارامتر سرعت باد تمام مدل‌های مورد استفاده به بالاترین دقت خود رسیده‌اند. لذا سرعت باد از جمله کلیدی‌ترین پارامترهای تبخیر-تعرق می‌باشد.

را می‌توان به محدود بودن بازه پارامترهای مورد استفاده نسبت داد. ارزیابی تاثیر پارامترهای ورودی در ایجاد سناریو برای مدل‌ها نشان می‌دهد که پارامتر سرعت باد (U) نقش مهم‌تری نسبت به سایر پارامترهای

جدول ۳- پارامترهای ارزیابی مدل‌های مورد مطالعه در دوره آزمون.

ایستگاه	ارومیه				یزد				
	مدل	R ²	RMSE (mm day ⁻¹)	NS	WI	R ²	RMSE (mm day ⁻¹)	NS	WI
RF1		۰/۸۷	۰/۷۴	۰/۸۴	۰/۹۶	۰/۵۳	۰/۶۴	۰/۵۳	۰/۸۳
RF2		۰/۸۶	۰/۷۵	۰/۸۴	۰/۹۶	۰/۵۱	۰/۶۶	۰/۵۰	۰/۸۳
RF3		۰/۹۰	۰/۶۳	۰/۸۸	۰/۹۷	۰/۶۲	۰/۵۷	۰/۶۲	۰/۸۸
RF4		۰/۹۵	۰/۴۵	۰/۹۴	۰/۹۹	۰/۸۳	۰/۳۹	۰/۸۳	۰/۹۵
RF5		۰/۹۶	۰/۴۰	۰/۹۵	۰/۹۹	۰/۸۶	۰/۳۶	۰/۸۵	۰/۹۶
RF-Hs1		۰/۸۶	۰/۷۶	۰/۸۳	۰/۹۶	۰/۶۳	۰/۵۷	۰/۶۳	۰/۸۸
RF-Hs2		۰/۹۰	۰/۶۵	۰/۸۸	۰/۹۷	۰/۶۹	۰/۵۲	۰/۶۹	۰/۹۰
RF-Hs3		۰/۹۲	۰/۵۸	۰/۹۰	۰/۹۸	۰/۷۲	۰/۴۹	۰/۷۲	۰/۹۱
RF-Hs4		۰/۹۲	۰/۵۷	۰/۹۰	۰/۹۸	۰/۸۹	۰/۳۱	۰/۸۹	۰/۹۷
RF-Hs5		۰/۹۹	۰/۲۷	۰/۹۹	۰/۹۹	۰/۹۲	۰/۲۶	۰/۹۲	۰/۹۸
M5P1		۰/۸۷	۰/۷۳	۰/۸۴	۰/۹۶	۰/۵۹	۰/۶۰	۰/۵۹	۰/۸۶
M5P2		۰/۸۹	۰/۶۶	۰/۸۷	۰/۹۷	۰/۶۱	۰/۵۸	۰/۶۱	۰/۸۶
M5P3		۰/۹۱	۰/۶۰	۰/۸۹	۰/۹۷	۰/۶۵	۰/۵۵	۰/۶۵	۰/۸۸
M5P4		۰/۹۶	۰/۴۲	۰/۹۵	۰/۹۹	۰/۸۴	۰/۳۸	۰/۸۴	۰/۹۵
M5P5		۰/۹۶	۰/۳۸	۰/۹۶	۰/۹۹	۰/۸۷	۰/۳۵	۰/۸۶	۰/۹۶
M5P-Hs1		۰/۸۹	۰/۶۸	۰/۸۶	۰/۹۷	۰/۶۹	۰/۵۲	۰/۶۹	۰/۹۰
M5P-Hs2		۰/۹۰	۰/۶۵	۰/۸۸	۰/۹۷	۰/۷۳	۰/۴۹	۰/۷۳	۰/۹۱
M5P-Hs3		۰/۹۲	۰/۵۷	۰/۹۰	۰/۹۸	۰/۷۳	۰/۴۸	۰/۷۳	۰/۹۲
M5P-Hs4		۰/۹۷	۰/۳۸	۰/۹۶	۰/۹۹	۰/۸۹	۰/۳۱	۰/۸۹	۰/۹۷
M5P-Hs5		۰/۹۷	۰/۳۳	۰/۹۷	۰/۹۹	۰/۹۴	۰/۲۴	۰/۹۴	۰/۹۸

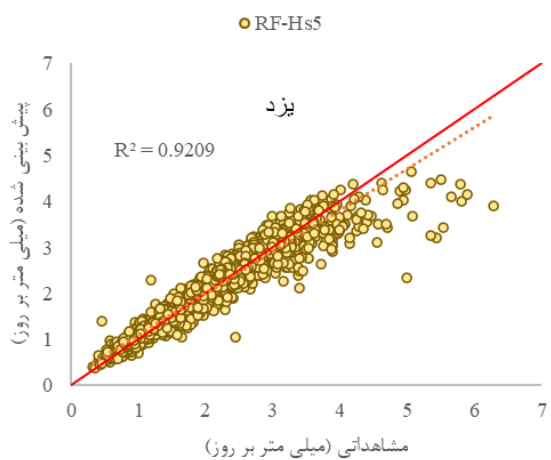
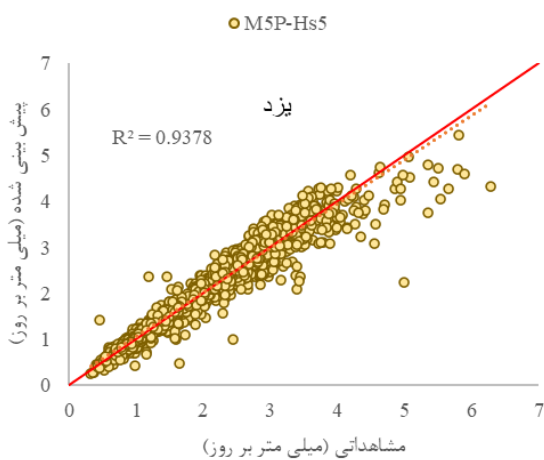
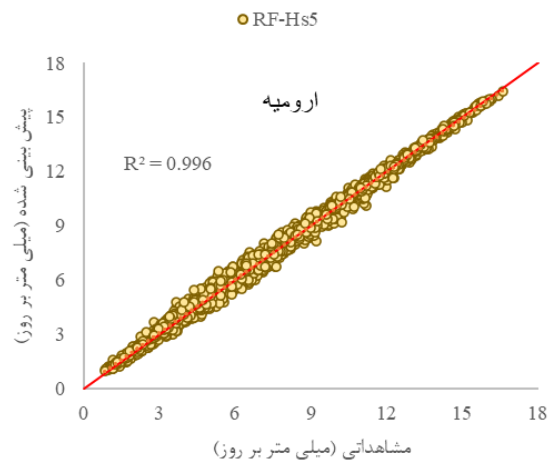
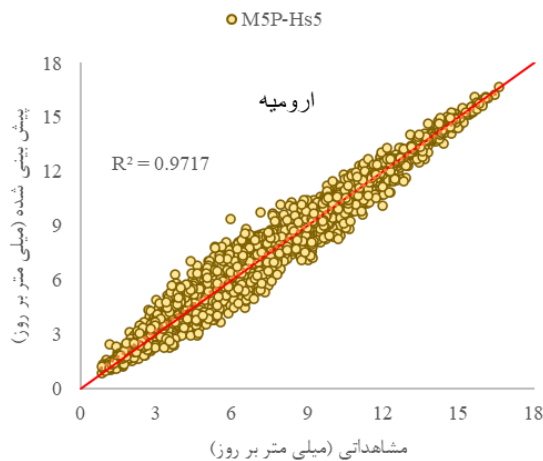
نیمه خشک با استفاده از مدل جنگل تصادفی پرداختند و نشان دادند مدل جنگل تصادفی ابزاری قابل اعتماد می‌باشد، مطابقت دارد.

نمودارهای پراکنش مقادیر تبخیر-تعرق روزانه در شکل ۴ که برای بهترین سناریو (سناریو ۵) رسم شده است، مقایسه بین دو مدل RF-Hs5 و M5P-Hs5 برای

نتایج تحقیق حاضر با نتایج میرهاشمی و پناهی (۲۰۱۵)، که به ارزیابی مدل‌های داده کاوی در پیش‌بینی دمای میانگین و تبخیر-تعرق بالقوه ماه آینده در ایستگاه هواشناسی سینوپتیک یزد پرداختند و نشان دادند که مدل درختی M5P بهترین عملکرد را از خود نشان داد و همچنین با نتایج الوارز و همکاران (۲۰۲۱)، که در تحقیقی به واکنش تبخیر-تعرق به تغییر اقلیم در مناطق

مدل M5P-Hs5 دارای همبستگی بیشتری نسبت به مدل RF-Hs5 بوده است.

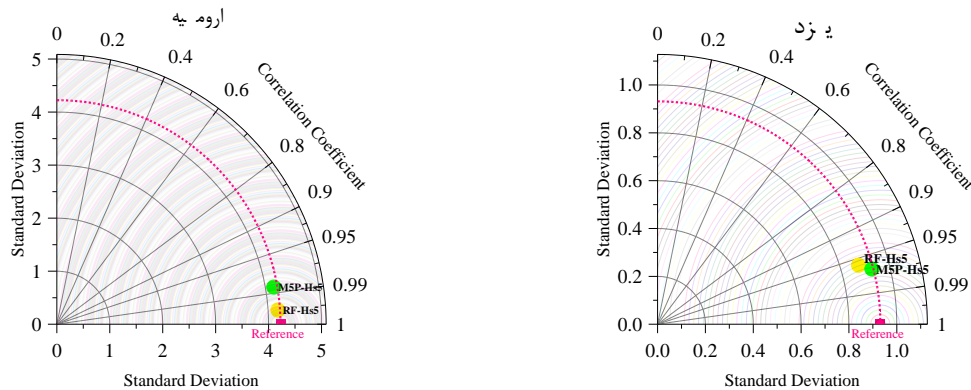
ایستگاه ارومیه نشان می‌دهد که در مدل RF-Hs5 داده‌های مشاهداتی و پیش‌بینی شده همبستگی بیشتری نسبت به یکدیگر داشته و لذا ضریب تعیین مدل RF-Hs5 بیشتر از مدل M5P-Hs5 بوده است. برای ایستگاه یزد نیز با اینکه دو مدل همبستگی تقریباً مشابهی را داشتند اما



شکل ۴- نمودارهای پراکنش مقادیر تبخیر-تعرق روزانه برای بهترین سناریو در میان مدل‌ها.

بیشتری بوده است. برای ایستگاه یزد نیز ضریب همبستگی مدل M5P-Hs5 نسبت به مدل RF-Hs5 بیشتر بوده و لذا دارای عملکرد بهتری بوده است. همچنین مقایسه بین انحراف از معیار دو مدل نیز نشان می‌دهد مدل RF-Hs5 دارای انحراف از معیار کمتری نسبت به مدل M5P-Hs5 بوده در نتیجه داده‌ها به مقدار میانگین نزدیک‌تر بوده و دارای پراکندگی کمتری بوده‌اند.

نمودارهای تیلور مقادیر تبخیر-تعرق روزانه در شکل ۵ برای بهترین سناریوی مدل‌ها رسم شده است. آنچه که از دو نمودار مشخص است اینست که هر دو مدل همبستگی بالایی با مقادیر مشاهداتی از خود نشان داده‌اند. مقایسه مدل‌های RF-Hs5 و M5P-Hs5 برای ایستگاه ارومیه نشان می‌دهد با اینکه دو مدل دارای انحراف از معیار تقریباً یکسانی بوده‌اند اما مدل RF-Hs5 با داشتن همبستگی بیشتر از مدل M5P-Hs5 دارای دقت



شکل ۵- نمودارهای تیلور مقادیر تبخیر-تعرق روزانه برای بهترین سناریو در میان مدل‌ها.

نتیجه‌گیری کلی

استفاده از پارامترهای آماری مورد مقایسه قرار گرفت و مشخص شد که در ایستگاه ارومیه سناریو پنجم مدل‌های RF-Hs و M5P-Hs به ترتیب بهترین عملکرد را با داشتن کمترین خطا نسبت به تمامی سناریوها ارائه داد. همچنین برای ایستگاه یزد نیز نتیجه گرفته شد که سناریو پنجم مدل‌های RF-Hs و M5P-Hs به ترتیب بالاترین دقت را با داشتن کمترین خطا از خود نشان داد. در حالت کلی سناریو پنجم مدل‌های ترکیبی RF-Hs و M5P-Hs در پیش‌بینی مقادیر تبخیر-تعرق روزانه نتایج مناسبی را ارائه داد و برای برنامه‌ریزی‌های آبیاری پیشنهاد می‌گردد.

پیش‌بینی تبخیر برای محققان از اهمیت خاصی برخوردار است، از این‌رو محققان روش‌های مختلفی جهت بهبود عملکرد مدل‌های پیش‌بینی ارائه نموده‌اند. برای پیش‌بینی تبخیر، علی‌رغم اهمیت فراوان این مؤلفه در درک تعادل آب در حوزه‌های آبخیز، متأسفانه مطالعات کمی در مورد استفاده از الگوریتم‌های هارگریوز-درختی انجام شده است. در این پژوهش با استفاده از روش‌های RF، M5P، Hs، در ایستگاه‌های ارومیه و یزد مقادیر تبخیر-تعرق روزانه در طی دوره آماری ۲۰۰۰-۲۰۲۱ برآورد گردید. نتایج به‌دست آمده با

منابع مورد استفاده

- Acquah SJ, Yan H, Zhang C, Wang G, Zhao B, Wu H and Zhang H, 2018. Application and evaluation of Stanghellini model in the determination of crop evapotranspiration in a naturally ventilated greenhouse. *International Journal of Agricultural and Biological Engineering* 11(6): 95-103.
- Albalasmeh AA, Mohawesh O, Gharaibeh MA, Alghamdi AG, Alajlouni MA and Alqudah AM, 2022. Effect of hydrogel on corn growth, water use efficiency, and soil properties in a semi-arid region. *Journal of the Saudi Society of Agricultural Sciences* 20(1): 13-27.
- Allen RG, Pereira LS, Raes D and Smith M, 1998. Crop evapotranspiration-Guidelines for computing crop water requirements. *FAO Irrigation and Drainage Paper* 300(9): 5094-5109.
- Azorin-Molina C, Vicente-Serrano SM, Sanchez-Lorenzo A, McVicar TR, Morán-Tejeda E, Revuelto J, El Kenawy A, Martín-Hernández N and Tomas-Burguera M, 2015. Atmospheric evaporative demand observations, estimates and driving factors in Spain. *Journal of Hydrology* 523(1): 262-277.
- Berti A, Tardivo G, Chiaudani A, Rech F and Borin M, 2014. Assessing reference evapotranspiration by the Hargreaves method in north-eastern Italy. *Agricultural Water Management* 140(1): 20-34.
- Breiman L, 1996. Bagging predictors. *Machine Learning* 24(2): 123-140.
- Breiman L, 2001. Random forests. *Machine Learning* 45(1): 5-32.

- Chen H, Huang JJ and McBean E, 2020. Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random Forest and support vector regression, for a cabbage farmland. *Agricultural Water Management* 228: 105-121.
- Elbeltagi A, AlThobiani F, Kamruzzaman M, Shaid S, Roy DK, Deb L, Islam MM, Kundu PK and Rahman MM, 2022. Estimating the Standardized Precipitation Evapotranspiration Index Using Data-Driven Techniques: A Regional Study of Bangladesh. *Water* 14(11): 1764-1778.
- Fang Q, Wang G, Zhang S, Peng Y, Xue B, Cao Y and Shrestha S, 2022. A novel ecohydrological model by capturing variations in climate change and vegetation coverage in a semi-arid region of China. *Environmental Research* 211: 113-127.
- Feng Y, Cui N, Gong D, Zhang Q and Zhao L, 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management* 193: 163-173.
- Feng Y, Cui N, Zhao L, Hu X and Gong D, 2016. Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. *Journal of Hydrology* 536: 376-383.
- Hargreaves GH and Allen RG, 2003. History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering* 29(1): 153-163.
- Khosravi K, Mao L, Kisi O, Yaseen ZM and Shahid S, 2018. Quantifying hourly suspended sediment load using data mining models: case study of a glacierized Andean catchment in Chile. *Journal of Hydrology* 567(1): 165-179.
- Kim S, Shiri J, Singh VP, Kisi O and Landeras G, 2015. Predicting daily pan evaporation by soft computing models with limited climatic data. *Hydrological Sciences Journal* 60(6): 1120-1136.
- Kisi O, Genc O, Dinc S and Zounemat-Kermani M, 2016. Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree. *Computers and Electronics in Agriculture* 122(1): 112-117.
- Kisi O and Heddad S, 2019. Evaporation modelling by heuristic regression approaches using only temperature data. *Hydrological Sciences Journal* 64(6): 653-672.
- Maleki S and Mavedat E, 2013. Evaluation of seismic vulnerability spectrum of cities based on various intensity scenarios using μd , TOPSIS, and GIS Models (Case study of Yazd). *Journal of Geography and Environmental Hazards* 2(1): 127-142. (In Persian with English abstract).
- Malik A, Kumar A, Kim S, Kashani MH, Karimi V, Sharafati A, Ghorbani MA, Al-Ansari N, Salih SQ and Yaseen ZM, 2020. Modeling monthly pan evaporation process over the Indian central Himalayas: Application of multiple learning artificial intelligence model. *Engineering Applications of Computational Fluid Mechanics* 14(1): 323-338.
- Menard S, 2000. Coefficients of determination for multiple logistic regression analysis. *The American Statistician* 54(1): 17-24.
- Mirhashemi SH and Panahi M, 2015. Evaluation of a data mining model in predicting of " average temperature" and potential evapotranspiration month for the next month in the synoptic weather station Yazd. *Biological Forum* 7(1): 1469-1473.
- Rostaei S, Mokhtari D and Khodaei Gheshlagh F, 2020. Evaluating the risk of desertification using the spectral indices in the surrounding area of Lake Urmia. *Quantitative Geomorphological Research* 9(3): 1-17. (In Persian with English abstract).
- Ruiz-Álvarez M, Gomariz-Castillo F and Alonso-Sarría F, 2021. Evapotranspiration response to climate change in semi-arid areas: Using random forest as multi-model ensemble method. *Water* 13(2): 222-235.
- Solomatine DP and Dulal KN, 2003. Model trees as an alternative to neural networks in rainfall—runoff modelling. *Hydrological Sciences Journal* 48(3): 399-411.
- Solomatine DP and SIEK MBL, 2004. Flexible and optimal M5 model trees with applications to flow predictions. *Hydroinformatics Journal* 33(5): 1719-1726.
- Wang S, Lian J, Peng Y, Hu B and Chen H, 2019. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agricultural Water Management* 221(1): 220-230.
- Yu T, Cui N, Zhang Q and Hu X, 2019. Applicability evaluation of daily reference crop evapotranspiration models in Northwest China. *Journal of Drainage and Irrigation Machinery Engineering* 8(1): 710-717.