

مقاله پژوهشی

کاربرد مدل‌های درختی و مبتنی بر کرنل در تعیین تبخیرتعرق مرجع روزانه در دو منطقه مرطوب و خشک ایران

فاطمه میکائیلی^۱، سعید صمدیان فرد^{۲*}

تاریخ دریافت: ۱۴۰۰/۰۲/۲۰ تاریخ پذیرش: ۱۴۰۰/۰۵/۱۹

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

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

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

چکیده

با توجه به واقع شدن ایران در اقلیم خشک و نیمه‌خشک، تبخیرتعرق یکی از موثرترین مولفه‌ها در بررسی وضعیت بیلان آبی است. برآورد دقیق این پارامتر در محاسبه دقیق نیاز آبی گیاهان و به تبع آن در طراحی و مدیریت سیستم‌های آبیاری و منابع آب از اهمیت ویژه‌ای برخوردار است. ارتباط پیچیده و غیرخطی عوامل موثر در فرآیند تبخیرتعرق باعث شده است که امروزه محققین با استفاده از روش‌های نوین در پی شناخت و پیش‌بینی دقیق این پارامتر باشند. هدف از پژوهش حاضر، بررسی توانایی مدل رگرسیون بردار پشتیبان (SVR)، مدل جنگل تصادفی (RF) و مدل درختی M5P در پیش‌بینی مقادیر روزانه تبخیرتعرق گیاه مرجع در دو ایستگاه آستارا و سیرجان به ترتیب واقع در مناطق مرطوب و خشک ایران با استفاده از داده‌های هواشناسی حداقل، متوسط و حداکثر دما، رطوبت نسبی، تابش خورشیدی و سرعت باد در بازه زمانی سال‌های ۲۰۲۰-۲۰۰۰ است. بدین منظور، چهار سناریو ترکیبی از پارامترهای هواشناسی برای واسنجی و صحت‌سنجی روش‌های مورد مطالعه مد نظر قرار گرفت. در نهایت، دقت روش‌های مذکور و روش‌های تجربی در برآورد تبخیرتعرق روزانه گیاه مرجع با استفاده از معیارهای آماری جذر میانگین مربعات خطا، ضریب همبستگی، شاخص پراکندگی، ضریب نش-ساتکلیف و ضریب ويلموت مورد مقایسه قرار گرفت. نتایج حاصل از داده‌های صحت‌سنجی نشان داد که مدل‌های SVR₃ (سناریو سه با روش رگرسیون بردار پشتیبان) و M5P₃ (سناریو سه با روش مدل درختی M5P) در ایستگاه آستارا با در نظر گرفتن تمامی پارامترهای هواشناسی و با دارا بودن ضریب همبستگی ۰/۹۹۳، جذر میانگین مربعات خطای ۰/۲۰۱ و همچنین مدل SVR₃ در ایستگاه سیرجان نیز با ضریب همبستگی ۰/۹۸۲، جذر میانگین مربعات خطای ۰/۴۱۰ در مقایسه با روش‌های تجربی هارگریوز-سامانی، مک کینک، تورک و دالتون نتایج بهتری در تخمین مقادیر تبخیرتعرق روزانه گیاه داشته‌اند.

واژه‌های کلیدی: تبخیرتعرق مرجع، جنگل تصادفی، درخت M5P، رگرسیون بردار پشتیبان، روش‌های تجربی

Application of Tree and Kernel- Based Models for Estimating Daily Reference Evapotranspiration in Humid and Arid Regions of Iran

F Mikaeili¹, S Samadianfard^{2*}

Received: May 10, 2021

Accepted: August 10, 2021

1-M.Sc. student, Dept. of Water Eng., Faculty of Agric., University of Tabriz, Iran

2- Assoc. Prof., Dept. of Water Eng., Faculty of Agric., University of Tabriz, Iran

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

Abstract

Background and Objectives

The gradual increase in the world's population requires continues increase in agricultural production. Climate change is one of the challenges of our society and frequent droughts affect large areas of the world, which requires more accurate management of water resources, both globally and in local catchments. Accurate estimation of components of the hydrological cycle is essential for proper irrigation scheduling. Most of the precipitation received by the earth is returned to the earth's atmosphere by the process of evapotranspiration (ET). On the other hand, ET is one of the most effective components in determining water requirements of plants and consequently in the designing irrigation systems, water resources management, environmental analysis, drought analysis, etc. The complex and nonlinear relationship between the factors affecting the process of evapotranspiration, has caused researchers today to use new methods to accurately identify and predict this parameter. Reference evapotranspiration is a concept that uses the crop coefficient to obtain the actual water requirement. According to the FAO proposal, the FAO- Penman- Monteith equation was introduced as a benchmark method for calculating reference evapotranspiration values when measurements of this parameter are not available and there is no access to lysimetric data. One of the major advantages of this model is its physical basis and global validity, but this equation needs a large number of meteorological parameters that are often not available, instead empirical equations with low meteorological variables or modern methods such as artificial intelligence and machine learning methods can be used.

Methodology

In this study, meteorological data from two stations of Astara located in the humid region and Sirjan located in the arid region of Iran in the period of 2000-2020 were studied to predict the crop evapotranspiration values. As mentioned, the FAO- Penman- Monteith method has used as a standard method for calibration and evaluation of the other functional equations and machine learning methods. In this study, four types of empirical equations including Hargreaves –Samani, Makkink, Turk and Dalton were evaluated against the FAO- Penman- Monteith model. Also, modelling was performed using Support Vector Regression, Random forest and M5P Tree model. In this study, using different meteorological parameters in accurate prediction of evapotranspiration utilizing four combined scenarios, calibration calculations were performed on 70% of data and validation calculations were performed on 30% of testing data implementing Weka software. Finally, statistical parameters including root mean squared error (RMSE), correlation coefficient (R), scatter index (SI), Nash-Sutcliffe coefficient (NS) and Wilmot index (WI) were used to determine the performance of each mentioned methods in estimating reference evapotranspiration values.

Findings

The obtained results showed that the SVR₃ and M5P₃ models in Astara station with all meteorological parameters and having R= 0.993, RMSE= 0.201 and also, the SVR₃ model in Sirjan station with R= 0.982, RMSE= 0.410 compared to the studied empirical methods provided better results in estimating the reference evapotranspiration and scenario 3 with all meteorological parameters was introduced as the top scenario. Among the empirical methods, Hargreaves- Samani was superior model only in Astara station. At Sirjan station, none of the empirical models performed better than the machine methods.

Conclusion

Accurate estimation of reference evapotranspiration in water resource management is essential. In this study, meteorological data from Astara and Sirjan stations were used to evaluate the ability of machine learning methods including SVR, RF and M5P to estimate the values of reference evapotranspiration and compared the results with empirical methods. The results showed the high accuracy of the SVR₃ model for both stations, but M5P₃ model only for humid area. Empirical methods except for Hargreaves- Samani had poor performance compared to data- driven models. Finally, the use of SVR and M5P methods in irrigation scheduling are recommended.

Keywords: Empirical methods, M5P, Random Forest, Reference Evapotranspiration, Support Vector Machine

مقدمه

می‌گیرد وابسته به آب است و یکی از موارد پرمصرف آب در گیاه تبخیرتعرق است، پس کاهش میزان آب اثرات نامطلوب بر میزان فتوسنتز، تولید محصول، کیفیت محصول و... خواهد داشت (علیزاده و همکاران ۲۰۰۱). از رویکردهای اصلی اندازه‌گیری تبخیرتعرق می‌توان به روش تعادل آب (گیتزنز ۱۹۸۲)، روش انتقال شار بخار آب (هاریک ۱۹۶۲) و یا نصب لایسیمترها اشاره کرد. زمان‌بر بودن، هزینه‌های نسبتاً زیاد، مشکلات احداث، نیاز به دقت بالا کاربرد آن‌ها را محدود می‌کند. رویکرد دیگر برای تعیین مقادیر واقعی تبخیرتعرق گیاه استفاده از تبخیرتعرق مرجع و ضرایب گیاهی است. تبخیرتعرق مرجع^۲ از سطح یک گیاه (چمن فرضی) و با ارتفاع ۱۲ سانتی‌متر که این پوشش کمبود آب و هیچ نوع آفت و بیماری ندارد و دارای ضریب بازتابش ۰/۲۳ و مقاومت روزنه ۷۰ ثانیه بر متر است (آلن و همکاران ۱۹۹۸). با استفاده از این تعریف می‌توان تبخیرتعرق مرجع را با

افزایش تدریجی جمعیت جهان به طور مداوم به افزایش در تولید محصولات کشاورزی نیاز دارد. کشاورزی بخشی است که بیشترین مقدار آب را مصرف می‌کند و از طرف دیگر تغییر اقلیم یکی از چالش‌های کنونی جامعه ماست. خشکسالی‌های مکرر مناطق وسیعی از جهان را تحت تاثیر قرار می‌دهد که مدیریت دقیق‌تر و بیشتر منابع آب را هم در سطح جهانی و هم در هر کدام از حوضه‌های آبریز می‌طلبد. برآورد دقیق از اجزا تشکیل‌دهنده چرخه هیدرولوژیکی به‌ویژه تبخیرتعرق برای برنامه‌ریزی صحیح در زمینه‌های مختلف ضروری است (گراناتا ۲۰۱۹). بیشترین مقدار بارش دریافتی توسط سطح زمین، با فرآیند تبخیرتعرق به جو زمین بازگردانده می‌شود. تبخیرتعرق^۱ دو فرآیند جداگانه است که آب از یک طرف با تبخیر از سطح خاک و از طرف دیگر با تعرق توسط گیاه از دسترس خارج می‌شود، از طرفی چون هر فرآیندی که در گیاه انجام

² Reference Evapotranspiration

¹ Evapotranspiration

نشان داد که بجز ایستگاه اردبیل تفاوت معنی‌داری بین نتایج دو مدل خطی و غیرخطی وجود ندارد. فنگ و همکاران (۲۰۱۷) به ارزیابی کارایی روش‌های جنگل تصادفی (RF)^۷ و شبکه‌های عصبی رگرسیون تعمیم یافته (GRNN)^۸ با استفاده از دو ترکیب ورودی شامل داده‌های کامل هواشناسی و فقط داده‌های دما و تابش فرا زمینی در مدل‌سازی تبخیرتغرق مرجع روزانه پرداختند و نتایج حاکی از این بود که هر دو مدل عملکرد خوبی در پیش‌بینی تبخیرتغرق مرجع روزانه از خود نشان دادند. تائو و همکاران (۲۰۱۸) در منطقه بورکینافاسو تبخیرتغرق مرجع را با استفاده از مدل فازی هیبریدی با الگوریتم کرم شب تاب (FA)^۹ برآورد نمودند و گزارش کردند که تاثیر استفاده از الگوریتم کرم شب تاب الهام گرفته از طبیعت به طور قابل توجهی عملکرد مدل بهینه‌سازی ANFIS^{۱۰} را بهبود بخشید و همچنین استراتژی مدل‌سازی بکار گرفته شده می‌تواند یک سیستم هوشمند خبره برای پیش‌بینی تبخیرتغرق مرجع در صحرای غرب آفریقا به ارمغان بیاورد. قره داغی و همکاران (۲۰۱۹) از پیش‌بینی‌های مدل منطقه‌ای WRF^{۱۱} برای افزایش دقت برآورد تبخیرتغرق مرجع استفاده کردند و گزارش کردند که نه تنها از نظر هماهنگی زمانی، بلکه به لحاظ مقدار نیز همانندی بسیار زیادی بین مقادیر برآورد شده تبخیرتغرق مرجع حاصل از پس پردازش آماری برونداد مدل پیش‌بینی و تحقیقاتی آب و هوا WRF با مقادیر محاسبه شده تبخیرتغرق مرجع توسط داده‌های ایستگاه‌های هواشناسی متناظر در نقاط مورد مطالعه وجود دارد و دامنه تغییرات ضریب تعیین (R^2) برونداد مدل هم بین ۰/۸۱۳ تا ۰/۹۲۱ بدست آمد. گراناتا (۲۰۱۹) با استفاده از چهار الگوریتم ماشین یادگیری شامل بگینگ^{۱۱}، رگرسیون بردار پشتیبان (SVR)^{۱۱}، جنگل

استفاده از متغیرهای هواشناسی بر مبنای فیزیک و معادلات تجربی بدست آورد (کریمی و همکاران ۲۰۲۰). براساس پیشنهاد سازمان فائو، معادله فائو- پنمن- مونتهیث (FPM)^۱ به عنوان روشی مبنای محاسبه مقادیر تبخیرتغرق مرجع هنگامی که اندازه‌های این پارامتر در دسترس نیست، معرفی شد (آلن و همکاران ۱۹۹۸). از عمده‌ترین مزایای این معادله می‌توان به اساس فیزیکی، اعتبار جهانی و همچنین عدم نیاز به کالیبراسیون محلی مقدماتی اشاره کرد. از طرف دیگر، نیاز به تعداد بالایی از پارامترهای هواشناسی که در برخی از مناطق و ایستگاه‌ها در دسترس نمی‌باشد، نقطه ضعف عمده این معادله می‌باشد (کریمی و همکاران ۲۰۲۰). به عنوان جایگزینی برای مدل فائو- پنمن- مونتهیث، معادلات تجربی با تکیه بر متغیرهای هواشناسی محدود، ممکن است استفاده شود. در زمینه برآورد تبخیرتغرق مطالعات زیادی انجام گرفته که در ذیل به چند نمونه از این مطالعات اشاره می‌گردد: ستاری و همکاران (۲۰۱۳) با استفاده از مدل شبکه عصبی مصنوعی (ANN)^۲ و مدل درختی M5 در ایستگاه بناب به پیش‌بینی تبخیرتغرق مرجع روزانه پرداختند. نتایج حاصل از بررسی‌ها نشان داد که روش شبکه عصبی مصنوعی در مقایسه با روش M5 با اختلاف کمی پیش‌بینی دقیق‌تری از تبخیرتغرق مرجع روزانه ارائه می‌دهد، ولی روش M5 روابط ساده خطی و کاربردی‌تری ارائه می‌دهد. زینالزاده و خان محمدی (۲۰۱۶) با مقایسه کارایی مدل‌های سری زمانی آرما^۳ و غیرخطی بیلینیر^۴ در شبیه‌سازی و پیش‌بینی تبخیرتغرق مرجع در سه ایستگاه اردبیل، تبریز و رشت نتیجه گرفتند که هر دو مدل خطی و غیرخطی می‌توانند سری زمانی تبخیرتغرق مرجع را با دقت قابل قبول برآورد نمایند و همچنین نتایج آزمون دایبلد و ماریانو

⁷ Firefly Algorithm

⁸ Adaptive Neuron Fuzzy Inference System

⁹ Weather Research and Forecasting

¹⁰ Bagging

¹¹ Support Vector Regression

¹ FAO- Penman- Montheith

² Artificial Neural Network

³ Arma

⁴ Bilinear

⁵ Random Forest

⁶ Generalized Regression Neural Networks

پیشین انجام شده، تبخیرتعرق اساس برنامه ریزی آبیاری است و برآورد صحیح از میزان این فرآیند می تواند ارزیابی صحیحی از میزان مصرف آب به ویژه در نواحی خشک و نیمه خشک داشته باشد (لیب و همکاران ۲۰۱۲).

در این پژوهش روش دقیق و استاندارد فائو-پنمن- مونتیت برای محاسبه تبخیرتعرق گیاه مرجع به عنوان روش مینا مورد نظر قرار گرفته است. در مطالعه حاضر، برای دو ایستگاه واقع در مناطق مرطوب و خشک ایران چهار الگوی ورودی مختلف از متغیرهای هواشناسی روزانه تعریف شد و مقادیر تبخیرتعرق گیاه مرجع حاصل از روش فائو-پنمن- مونتیت با استفاده از روش های داده محور رگرسیون بردار پشتیبان (SVR)، جنگل تصادفی (RF) و مدل درختی M5P مدل سازی شده و در نهایت با بررسی و ارزیابی دقت هر یک از روش های مذکور با روش های تجربی هارگریوز- سامانی، مک کینک، تورک و دالتون دقیق ترین روش پیشنهاد شد.

مواد و روش ها

در این مطالعه برای بررسی و برآورد مقادیر تبخیرتعرق مرجع در مقیاس روزانه، داده های هواشناسی دو ایستگاه سینوپتیک آستارا و سیرجان به ترتیب واقع در استان های گیلان و کرمان در یک بازه زمانی ۲۱ ساله مربوط به سال های ۲۰۲۰-۲۰۰۰ مورد استفاده قرار گرفت (شکل ۱). اطلاعات کلی مکان های مورد مطالعه و همچنین مقادیر شاخص خشکی (I_A)^۳ (بی نام ۱۹۹۷) در جدول ۱ قابل ملاحظه است. این شاخص نسبت بین بارش سالانه و مقادیر سالانه تبخیرتعرق مرجع در هر ایستگاه است. براساس مقادیر حاصل از این شاخص، در حالت کلی شش طبقه بندی اقلیمی قابل شناسایی است، بدین صورت که $I_A < 0.05$ به عنوان منطقه خیلی خشک، $0.05 < I_A < 0.2$ به عنوان منطقه خشک، $0.2 < I_A < 0.5$ جزو مناطق

تصادفی و مدل درختی M5P به مقایسه عملکرد سه مدل از تبخیرتعرق که تنها در پارامترهای ورودی باهم تفاوت داشتند، پرداخت. نتایج نشان داد که عملکرد مدل M5P با متغیرهای تابش خورشیدی، شار گرما، رطوبت خاک، سرعت باد، میانگین رطوبت نسبی و دما بهترین بود. گونزالس دل سرو و همکاران (۲۰۲۰) با مدل سازی تبخیرتعرق مرجع روزانه در منطقه نیمه خشک جنوب هند با استفاده از ANFIS و مدل های تجربی گزارش کردند در این مورد مدل های تجربی عملکرد بهتری را نسبت به مدل های ANFIS نشان دادند و نتیجه گرفته شد که روش ریت چه^۱ روش تخمین خوبی برای مقادیر تبخیرتعرق مرجع برای یک منطقه خشک در جنوب هند است و علاوه بر این، نتایج مدل های ANFIS امیدوارکننده بود و می تواند به عنوان روش تخمینی استفاده شود. کریمی و همکاران (۲۰۲۰) برای برآورد مقادیر تبخیرتعرق مرجع در نواحی مرطوب ایران با جایگزینی متغیرهای هواشناسی گمشده در مدل های کلاسیک و RF به ارزیابی تاثیر متغیرهای هواشناسی محاسبه شده بر دقت عملکرد هر یک از مدل ها پرداختند و گزارش کردند که براساس نتایج بدست آمده با جایگزینی متغیرهای هواشناسی گمشده با مقادیر محاسباتی مدل های ترکیبی (با تکیه بر ارزش تخمینی سرعت باد) نتایجی با دقت بالا ارائه دادند، در حالی که استفاده از مقادیر تابش خورشیدی تخمین زده شده دقت عملکرد مدل را کاهش داد. فنگ و تیئن (۲۰۲۱) به برآورد تبخیرتعرق مرجع با استفاده از الگوریتم نزدیک ترین همسایگی (K-NN)^۲ و با داده های آب و هوایی محدود در منطقه ای خشک در چین پرداختند که نتیجه پیش بینی با نتایج معادله فائو-پنمن- مونتیت مطابقت داشت. همچنین براساس مقایسه عملکرد کلی، مشخص شد که مدل پیش بینی K-NN که به دمای حداکثر و حداقل و رطوبت نسبی به عنوان پارامترهای ورودی نیاز داشت، بهترین دقت را ارائه داد. با توجه به مطالعات

¹ Ritchie

² K- Nearest Neighbor

³ Aridity Index

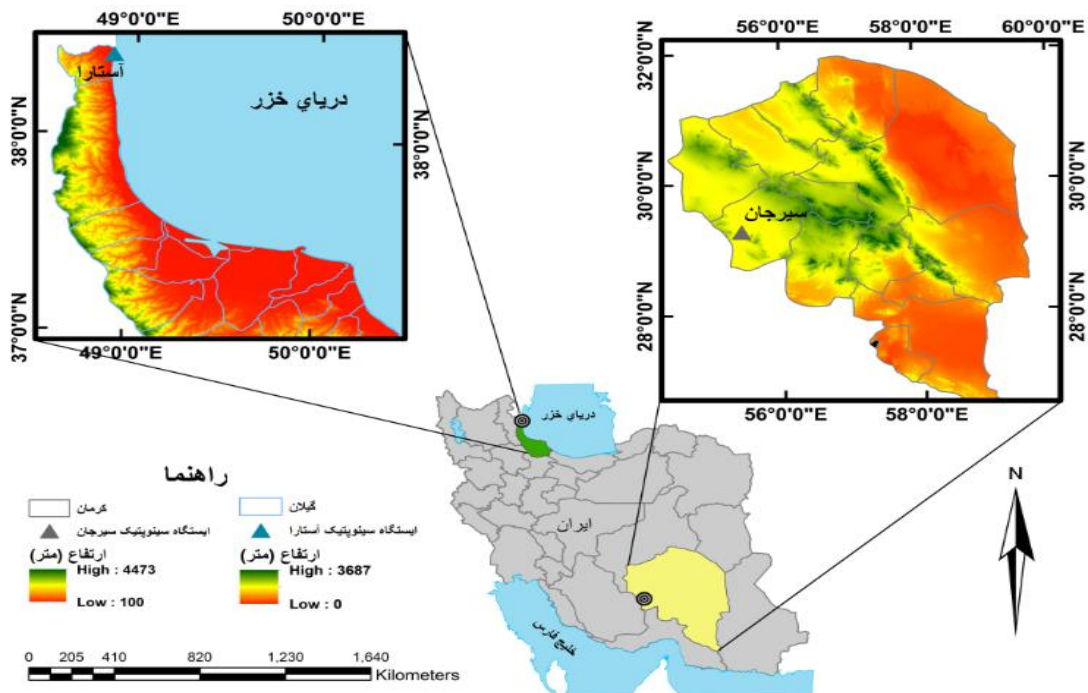
مدل‌سازی با استفاده از روش‌های رگرسیون بردار پشتیبان (SVR)، الگوریتم جنگل تصادفی (RF) و مدل درختی M5P انجام گرفت. برای اجرای مدل‌ها از نرم‌افزار Weka استفاده گردید که در این پژوهش ۷۰٪ داده‌ها برای آموزش و ۳۰٪ برای آزمون در نظر گرفته شدند. همچنین لازم به ذکر است که در تحقیق حاضر از کرنل PUK به دلیل داشتن دقت بالا و کمترین خطا استفاده گردید.

جدول ۳ خلاصه‌ای از معادلات تجربی بکار گرفته شده به همراه متغیرهای هواشناسی مورد نیاز برای محاسبه مقادیر تبخیرتقرق مرجع در مقیاس روزانه را نشان می‌دهد. همانطور که اشاره شد، معادله استاندارد فائو-پنمن-مونتیت به عنوان روش معیار برای کالیبراسیون و ارزیابی سایر معادلات کاربردی و مدل‌های یادگیری ماشین استفاده شد. در این مطالعه سه نوع معادله در برابر مدل استاندارد فائو-پنمن-مونتیت شامل هارگریوز-سامانی مبتنی بر دما، مک کینک و تورک مبتنی بر تابش و معادله دالتون مبتنی بر انتقال جرم ارزیابی شد.

نیمه‌خشک، $0.05 < I_A < 0.165$ منطقه نیمه‌مرطوب خشک، $0.165 < I_A < 0.175$ منطقه نیمه‌مرطوب تر و $I_A > 0.175$ به‌عنوان منطقه مرطوب طبقه‌بندی می‌شوند. داده‌های هواشناسی روزانه شامل حداقل، متوسط و حداکثر دما، تابش خورشیدی، میانگین رطوبت نسبی و سرعت باد می‌باشند. بدین منظور، مشخصات آماری پارامترهای مذکور در ایستگاه‌های مورد مطالعه در جدول ۲ ارائه شده است. لازم به ذکر است که تعداد کل داده‌های مورد استفاده برای ایستگاه‌های آستارا و سیرجان به ترتیب ۶۹۲۸ و ۷۵۹۳ می‌باشند و داده‌ها از سازمان هواشناسی کشور اخذ گردیده‌اند.

جدول ۱- خلاصه ایستگاه‌های هواشناسی مورد مطالعه.

ایستگاه	عرض جغرافیایی (°E)	طول جغرافیایی (°N)	ارتفاع از سطح دریا (m)	I_A
آستارا	۴۸/۵۱	۳۸/۲۱	-۲۱/۱	۱/۵۲
سیرجان	۵۵/۴۱	۲۹/۲۸	۱۷۳۹/۴	۰/۱۴



شکل ۱- موقعیت مکانی ایستگاه‌های مورد مطالعه.

جدول ۲- مشخصات آماری داده‌های مورد استفاده.

ایستگاه	پارامتر	واحد	میانگین	مینیمم	ماکزیمم	انحراف معیار	ضریب تغییرات	چولگی
آستارا	T_{min}	(°C)	۱۲/۶۳	-۷/۸۰	۲۸/۰۰	۷/۳۴۶	۰/۵۸۱	-۰/۱۷۸
	T_{max}	(°C)	۲۰/۱۷	-۱/۴۰	۳۶/۸۰	۸/۳۹۸	۰/۴۱۶	-۰/۰۶۵
	T_{mean}	(°C)	۱۶/۲۵	-۳/۷۰	۳۰/۹۰	۷/۷۶۷	۰/۴۷۸	-۰/۰۸۰
	RH	(%)	۷۹/۷۸	۲۴/۵۰	۱۰۰/۰۰	۱۰/۳۳۲	۰/۱۳۰	-۰/۹۷۷
	R_s	(MJ m ⁻² day ⁻¹)	۵/۸۸	۱/۴۸	۱۲/۱۲	۳/۱۶۱	۰/۵۳۸	۰/۳۹۴
	U_2	(m s ⁻¹)	۱/۳۲	۰/۰۰	۱۲/۷۵	۰/۹۱۶	۰/۶۹۵	۱/۶۵۹
سیرجان	T_{min}	(°C)	۹/۸۱	-۲۷/۵۰	۲۸/۴۰	۸/۴۸۱	۰/۸۶۴	-۰/۱۸۳
	T_{max}	(°C)	۲۵/۷۵	-۱/۰۰	۴۳/۶۰	۸/۸۱۸	۰/۳۴۲	-۰/۲۴۴
	T_{mean}	(°C)	۱۸/۲۴	-۴/۸۰	۳۴/۷۰	۸/۷۷۸	۰/۴۸۱	-۰/۱۷۶
	RH	(%)	۳۲/۸۴	۸/۵۰	۹۸/۸۸	۱۵/۹۰۱	۰/۴۸۴	۱/۰۹۵
	R_s	(MJ m ⁻² day ⁻¹)	۸/۳۳	۲/۰۴	۱۲/۲۳	۲/۴۵۸	۰/۲۹۵	-۰/۳۷۱
	U_2	(m s ⁻¹)	۲/۶۵	۰/۰۰	۱۱/۷۵	۱/۵۴۲	۰/۵۸۳	۰/۴۵۲

خورشیدی نیاز دارد (معادله تورک به داده رطوبت نسبی هم نیاز دارد). براساس رژیم‌های آب و هوایی در مناطق مختلف این مدل‌ها به عنوان رویکردهای قابل قبول برای تخمین مقادیر تبخیرتعرق مرجع اثبات شده‌اند. معادله دالتون (دالتون ۱۸۰۲) مبتنی بر انتقال جرم هم معادله‌ای ساده و بر پایه قانون گاز دالتون است که در آن فقط داده‌های دما و سرعت باد نیاز می‌باشد.

معادله هارگریوز- سامانی یکی از ساده‌ترین معادلات برای برآورد مقادیر تبخیرتعرق مرجع در مقیاس روزانه و ماهانه با شرایط اقلیمی مختلف است، این معادله تنها به داده‌های ثبت شده دما برای تخمین مقادیر تبخیرتعرق مرجع نیاز دارد (ونگ و همکاران ۲۰۰۸). معادلات مک کینک (مک کینک ۱۹۵۷) و تورک (تورک ۱۹۶۱) مبتنی بر تابش به داده‌های دما و تابش

جدول ۳- مشخصات معادلات تجربی مورد مطالعه.

معادلات ET ₀	ورودی‌های هواشناسی	روابط	معیار
$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_{mean} + 273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)}$	T_{mean}, R_s, U_2, RH		فائو- پنمن- مونتیت
$ET_0 = 0.0023R_a(T_{mean} + 17.8)\sqrt{T_{max} - T_{min}}$	$T_{mean}, T_{max}, T_{min}$		* مبتنی بر دما هارگریوز- سامانی
$ET_0 = 0.61 \frac{\Delta}{\Delta + \gamma} \frac{R_s}{\lambda} - 0.12$	$T_{mean}, T_{max}, T_{min}, R_s$		* مبتنی بر تابش مک کینک
$ET_0 = a_T 0.013 \frac{T_{mean}}{T_{mean} + 15} \frac{23.8856R_s + 50}{\lambda}$ $R_H \geq 50 \rightarrow a_T = 1$ $R_H < 50 \rightarrow a_T = 1 + \frac{50 - R_H}{70}$	$T_{mean}, T_{max}, T_{min}, R_s, RH$		تورک
$ET_0 = (0.3648 + 0.07223U_2)(e_s - e_a)$	$T_{max}, T_{min}, T_{dew}, U_2$		* مبتنی بر انتقال جرم دالتون

ET_0 ← تبخیرتعرق گیاه مرجع R_n ، $(mm.day^{-1})$ ← تابش خالص در سطح پوشش گیاهی $(MJ.m^{-2}.day^{-1})$ ← G شار گرما به داخل خاک $(MJ.m^{-2}.day^{-1})$ ← T_{mean} متوسط دمای هوا $(^{\circ}C)$ ← u_2 سرعت باد در ارتفاع دو متری از سطح زمین $(m.s^{-1})$ ← e_a فشار بخار واقعی (kPa) ← e_s فشار بخار اشباع (kPa) ← Δ شیب منحنی فشار بخار $(kPa.^{\circ}C^{-1})$ ← γ ضریب سایکرومتری $(kPa.^{\circ}C^{-1})$ ← R_a تابش فرازمینی $(mm.day^{-1})$ ← T_{max} حداکثر دمای هوا $(^{\circ}C)$ ← T_{min} حداقل دمای هوا $(^{\circ}C)$ ← R_s تابش خورشیدی روزانه $(MJ.m^{-2}.day^{-1})$ ← λ گرمای نهان تبخیر $(MJ.kg^{-1})$ ← R_H رطوبت نسبی (%) و $(e_s - e_a)$ در معادله دالتون برحسب (hPa)

روش رگرسیون بردار پشتیبان

الگوریتم جنگل تصادفی

این الگوریتم یک نوع تکنیک نظارت شده است که توسط وپنیک (۱۹۹۵) معرفی شد که این روش هم برای گروه‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها بکار می‌رود و به دو دسته کلی طبقه‌بندی کننده بردار پشتیبان و رگرسیون بردار پشتیبان تقسیم می‌شوند. با کارایی بیشتر مشکلات تخمین عملکرد چند بعدی را به دلیل توانایی آن در تشخیص همبستگی بین داده‌های ورودی و خروجی حل می‌کند و همگرایی سریع نشان می‌دهد (آئودل و همکاران ۲۰۱۹). برای مجموعه نقاط آموزش $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_i, Y_i)\}$ که $X_i \in R^n$ مقادیر ورودی، $Y_i \in R$ مربوط به هر یک از X_i ها و n هم تعداد کل نقاط (کل داده‌ها) می‌باشد که همبستگی بین ورودی‌ها و خروجی‌ها براساس تابع خطی زیر قابل توصیف است (یائو و همکاران ۲۰۱۸):

$$f(x_i) = w^T \varphi(x_i) + b \quad [1]$$

که $w \in R^n$ و $b \in R$ ضرایب قابل تنظیم هستند که به ترتیب بردارهای وزن و بایاس^۱ را نشان می‌دهند. در بعضی موارد برای جداسازی داده‌ها به ساختارهای پیچیده و غیر خطی نیاز است که ماشین بردار پشتیبان با بکارگیری مجموعه‌ای از توابع ریاضی که کرنل نام دارد، داده‌های اصلی را در فضای جدیدی نگاشت می‌کند (جودی و ستاری ۲۰۱۶). بنابراین موضوع اصلی، پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM^۲ تاکنون تجربه نکرده است، پیش‌بینی کند (سیاسر و هنار ۲۰۱۹).

به عنوان رویکرد گروهی مبتنی بر درخت، جنگل تصادفی از ابعاد رگرسیون بالا استفاده می‌کند که در آن رشد درخت براساس درختان متصل زیادی است (بريمن ۲۰۰۱). به منظور به وجود آوردن درخت رگرسیونی از پارتیشن‌بندی بازگشتی و رگرسیون چندگانه استفاده می‌شود. جنگل تصادفی برای جلوگیری از تطابق درخت‌های رگرسیونی مختلف، تنوع درختان را از طریق ایجاد زیرمجموعه‌های مختلف از داده‌های آموزشی (کیسه‌بندی) کم می‌کند. کیسه‌بندی روشی است که از طریق نمونه‌برداری تصادفی از مجموعه داده‌های اصلی همراه با جایگزینی برای ایجاد داده‌های آموزشی بکار می‌رود (نوروزی و همکاران ۲۰۱۵). پس از تشکیل درختان در مجموعه که نتیجه روش کیسه‌بندی است، برای بدست آوردن خروجی نهایی، متوسط همه درخت‌های پیش‌بینی شده محاسبه می‌شود. (شیری ۲۰۱۸). بایاس درختان کیسه‌ای و منفرد مشابه است اما واریانس به مقادیر همبستگی در بین درختان بستگی دارد (هیستی و همکاران ۲۰۰۹). تشکیل رگرسیون‌های مبتنی بر RF، با رشد درختان براساس بردار تصادفی θ شروع می‌شود، بنابراین درخت برآورد شده $h(x, \theta)$ ممکن است مقادیر عددی بگیرد. میانگین مربعات خطای تعمیم از هر پیش‌بینی‌کننده $h(x)$ به این صورت است که:

$$E_{x,y} = (Y - h(x))^2 \quad [2]$$

تشکیل RF با میانگین‌گیری از زامین درخت منفرد صورت می‌گیرد. با افزایش تعداد درختان جنگل، خطا بصورت

² Support Vector Machine

¹ Bias

زیر برآورد می شود (بریمن ۲۰۰۱، کریمی و همکاران ۲۰۱۸):

$$Sd(T) = \sqrt{\frac{1}{N}(\sum_{i=1}^N y_i^2 - \frac{1}{N}(\sum_{i=1}^N y_i)^2)} \quad [5]$$

که T مجموعه ای از نمونه هاست، T_i نشان دهنده ای از نمونه هاست که i امین نتیجه تست پتانسیلی را دارند، Sd_i بیانگر انحراف معیار، y_i مقدار عددی ویژگی هدف نمونه i و N تعداد داده ها می باشد (ستاری و همکاران ۲۰۱۳). پس از اینکه مجموع مجذور انحراف از میانگین داده ها به صفر نزدیک می شود، درخت بزرگی ایجاد می شود که کار با آن سخت است. برای رسیدن به یک درخت کارآمد باید شاخه های اضافی هرس شوند که برای هرس کردن دو روش وجود دارد: ۱- هرس کردن قبل از شکل گیری درخت ۲- هرس کردن بعد از شکل گیری درخت. روش اول زمانی استفاده می شود که در ابتدا داده های زیادی وجود داشته باشد و در نهایت درخت بهینه ایجاد می شود.

معیارهای ارزیابی دقت روش های مورد استفاده

برای ارزیابی عملکرد هر یک از روش های SVR، RF و M5P و روش های تجربی HS، MK، TU و DN در تخمین مقادیر تبخیر تعرق گیاه مرجع از پارامترهای آماری جذر میانگین مربعات خطا (RMSE)^۱، ضریب همبستگی (R)^۲، شاخص پراکندگی (SI)^۳، ضریب نش-ساتکلیف (NS)^۴ و ضریب ویلموت (WI)^۵ استفاده گردید. حد مطلوب معیارهای RMSE و SI نزدیک صفر و معیارهای R، NS و WI نزدیک یک می باشند.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad [6]$$

$$R = \frac{\text{cov}(O_i, P_i)_i}{\sqrt{\text{var}(O_i) \times \text{var}(P_i)}} \quad [7]$$

$$E_{x,y} = (Y_{avj} h(x, \theta_j))^2 \rightarrow E_{x,y} (y - E_{\theta} h(x, \theta))^2 \quad [3]$$

در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می کنند و هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش بینی کننده ها نسبتاً کم باشد، یک روش پیش بینی کارآمد به حساب می آید.

مدل درختی M5P

این الگوریتم به عنوان زیر مجموعه ای از روش های یادگیری ماشینی و داده کاوی است، که خروجی را بصورت یک سازه درختی با استفاده از داده های ورودی تخمین می زند و یک مدل درختی برای پیش بینی ویژگی های عددی پیوسته است (صمدیان فرد و پناهی ۲۰۱۸). درخت تصمیم جهت نمایش طبقه بندی کننده ها و رگرسیون ها است و متشکل از ریشه، شاخه، گره ها و برگ ها می باشد. در درخت تصمیمی که عمل طبقه بندی را انجام می دهد، برگ ها بیانگر کلاس ها هستند و گره ها با دایره مشخص می شوند و شاخه ها نشان دهنده اتصال بین گره ها هستند که بازه ای از مقادیر مختلف ویژگی های مجموعه را نشان می دهند. تعیین محل انشعاب توسط یکی از متغیرهای پیش بینی کننده انجام می پذیرد و بازه های انشعاب طوری انتخاب می شوند که مجموع مجذور انحراف از میانگین داده های هر گره را به حداقل برسانند (فلاحی و همکاران ۲۰۱۱). برای ایجاد یک درخت تصمیم از معیار انشعاب براساس کاهش عملکرد انحراف استاندارد (SDR) مقادیر هر کلاس (آنتروپی) استفاده می شود.

$$SDR = Sd(T) - \sum_{i=1}^N \frac{|T_i|}{|T|} Sd(T_i) \quad [4]$$

⁴ Nash-Sutcliffe Coefficient

⁵ Wilmot Coefficient

¹ Root Mean Squared Error

² Correlation Coefficient

³ Scatter Index

تا ۲۰۲۰) با استفاده از نرم افزار Weka که در دانشگاه وایکاتو نیوزلند توسعه داده شده و شامل مجموعه‌ای از به‌روزترین الگوریتم‌های یادگیری ماشینی و ابزاری برای پیش‌پردازش داده‌ها می‌باشد، انجام شد.

جدول ۴- پارامترهای دخیل در هر سناریو در روش‌های

RF، SVR، و MSP.

پارامترهای ورودی	شماره سناریو
T _{max} , T _{min} , T _{mean}	۱
T _{max} , T _{min} , T _{mean} , R _s	۲
T _{max} , T _{min} , T _{mean} , R _s , RH	۳
T _{max} , T _{min} , T _{mean} , U ₂	۴

با تعریف چهار سناریو از پارامترهای هواشناسی، محاسبات پیش‌بینی تبخیرتغرق با روش‌های RF، SVR و M5P انجام گرفت و دقت روش‌های مذکور و روابط تجربی HS، MK، TU و DN در مرحله صحت‌سنجی براساس معیارهای آماری (روابط ۶ تا ۱۰) محاسبه شده و در جدول ۵ ارائه گردید.

$$SI = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}}{\bar{O}} \quad [8]$$

$$NS = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \right] \quad [9]$$

$$WI = \left| 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \right] \right| \quad [10]$$

در روابط ۶ تا ۱۰ تبخیرتغرق مرجع تخمین‌زده شده، O_i تبخیرتغرق مرجع محاسبه شده با روش فائو- پنمن- موتیث و N تعداد کل داده‌ها می‌باشد.

نتایج و بحث

به منظور امکان‌سنجی استفاده از ترکیب پارامترهای مختلف هواشناسی در پیش‌بینی دقیق مقادیر تبخیرتغرق گیاه مرجع، چهار سناریو از ترکیب‌های مختلف هواشناسی براساس ورودی‌های معادلات تجربی تعریف گردید (جدول ۴). سپس محاسبات واسنجی بر روی ۷۰٪ از داده‌ها (۲۰۰۰ تا ۲۰۱۴) و محاسبات صحت‌سنجی بر روی ۳۰٪ از داده‌های باقی‌مانده (۲۰۱۵)

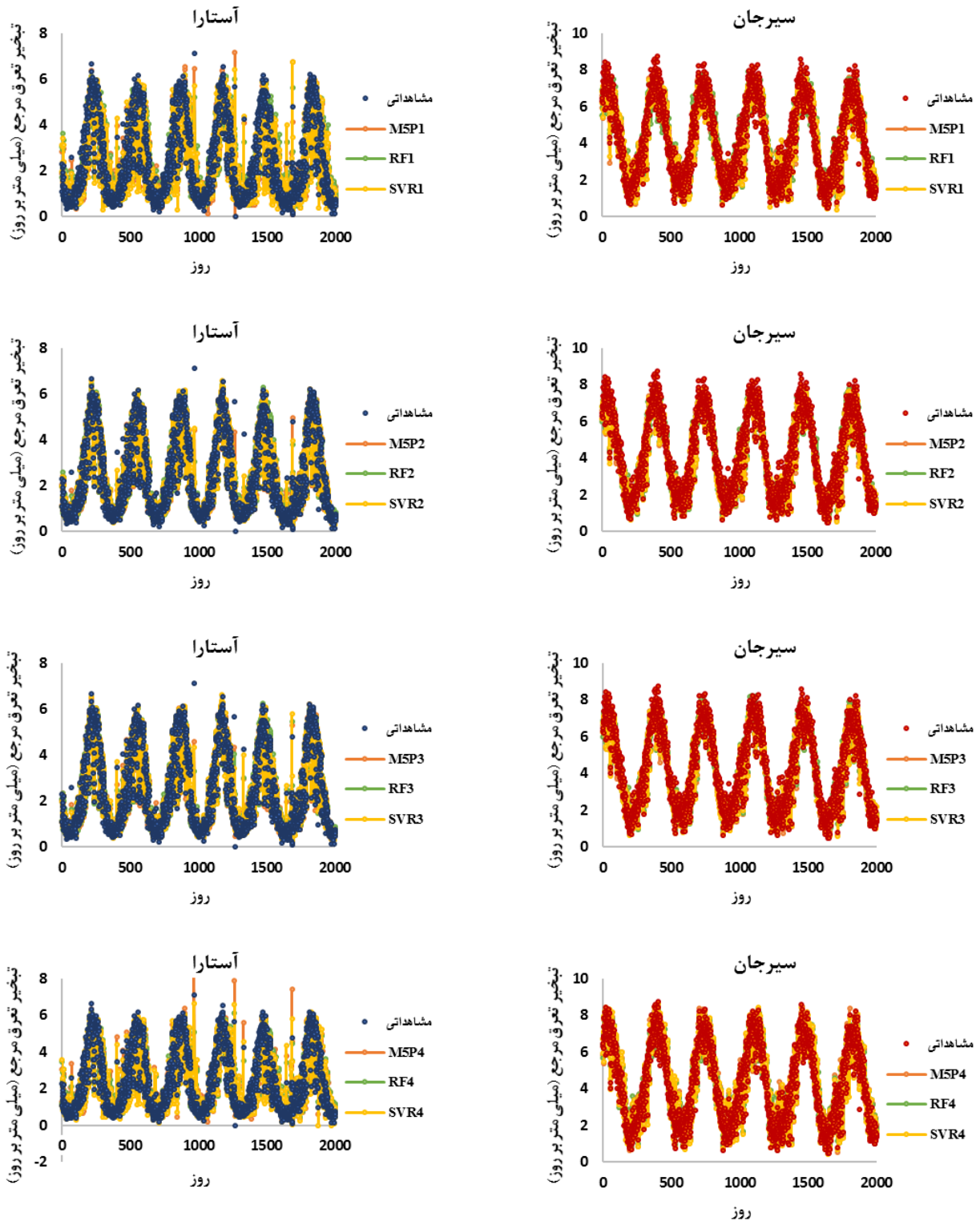
جدول ۵- نتایج سناریوهای تعریف شده برای روش‌های RF، SVR، M5P و روش‌های تجربی HS، MK، TU و DN.

ایستگاه	مدل	RMSE	R	SI	NS	WI	ایستگاه	مدل	RMSE	R	SI	NS	WI
آستارا	SVR ₁	۰/۶۳۴	۰/۹۳۲	۰/۲۵۹	۰/۸۶۸	۰/۹۶۴	سیرجان	SVR ₁	۰/۶۲۱	۰/۹۵۸	۰/۱۴۵	۰/۹۱۶	۰/۹۷۸
	SVR ₂	۰/۲۱۲	۰/۹۹۳	۰/۰۸۷	۰/۹۸۵	۰/۹۹۶		SVR ₂	۰/۴۲۸	۰/۹۸۱	۰/۱۰۰	۰/۹۶۰	۰/۹۹۰
	SVR ₃	۰/۲۰۱	۰/۹۹۳	۰/۰۸۲	۰/۹۸۷	۰/۹۹۷		SVR ₃	۰/۴۱۰	۰/۹۸۲	۰/۰۹۶	۰/۹۶۳	۰/۹۹۱
	SVR ₄	۰/۵۴۷	۰/۹۵۰	۰/۲۲۳	۰/۹۰۲	۰/۹۷۴		SVR ₄	۰/۴۳۰	۰/۹۸۰	۰/۱۰۱	۰/۹۶۰	۰/۹۹۰
	M5P ₁	۰/۶۲۸	۰/۹۳۳	۰/۲۵۷	۰/۸۷۰	۰/۹۶۴		M5P ₁	۰/۶۷۸	۰/۹۵۰	۰/۱۵۹	۰/۹۰۰	۰/۹۷۳
	M5P ₂	۰/۲۱۶	۰/۹۹۲	۰/۰۸۸	۰/۹۸۵	۰/۹۹۶		M5P ₂	۰/۴۴۷	۰/۹۷۹	۰/۱۰۵	۰/۹۵۶	۰/۹۸۹
	M5P ₃	۰/۲۰۱	۰/۹۹۳	۰/۰۸۲	۰/۹۸۷	۰/۹۹۷		M5P ₃	۰/۴۳۷	۰/۹۸۰	۰/۱۰۲	۰/۹۵۸	۰/۹۸۹
	M5P ₄	۰/۵۷۳	۰/۹۴۵	۰/۲۳۴	۰/۸۹۲	۰/۹۷۰		M5P ₄	۰/۴۷۹	۰/۹۷۵	۰/۱۱۲	۰/۹۵۰	۰/۹۸۷
	RF ₁	۰/۶۸۵	۰/۹۲۰	۰/۲۸۰	۰/۸۴۶	۰/۹۵۶		RF ₁	۰/۷۰۴	۰/۹۴۵	۰/۱۶۵	۰/۸۹۲	۰/۹۷۱
	RF ₂	۰/۲۵۵	۰/۹۸۹	۰/۱۰۴	۰/۹۷۹	۰/۹۹۵		RF ₂	۰/۴۷۲	۰/۹۷۶	۰/۱۱۱	۰/۹۵۱	۰/۹۸۷
	RF ₃	۰/۲۳۶	۰/۹۹۱	۰/۰۹۷	۰/۹۸۲	۰/۹۹۵		RF ₃	۰/۴۵۷	۰/۹۷۷	۰/۱۰۷	۰/۹۵۴	۰/۹۸۸
	RF ₄	۰/۶۰۶	۰/۹۳۸	۰/۲۴۷	۰/۸۷۹	۰/۹۶۶		RF ₄	۰/۴۸۰	۰/۹۷۵	۰/۱۱۲	۰/۹۵۰	۰/۹۸۷
	HS	۰/۵۳۸	۰/۹۶۲	۰/۲۱۸	۰/۹۰۰	۰/۹۷۴		HS	۰/۷۱۵	۰/۹۵۷	۰/۱۷۳	۰/۸۸۴	۰/۹۷۱
	TU	۲/۳۹۴	۰/۹۸۳	۰/۹۶۸	۰/۹۷۲	۰/۵۳۶		TU	۳/۵۶۵	۰/۹۵۷	۰/۸۶۵	۰/۸۹۲	۰/۵۲۰
	MK	۱/۹۶۰	۰/۹۷۹	۰/۷۹۳	۰/۳۲۲	۰/۶۲۶		MK	۳/۱۵۵	۰/۹۵۵	۰/۷۶۵	۰/۲۶۴	۰/۵۵۸
	DN	۰/۹۱۶	۰/۸۸۲	۰/۳۷۰	۰/۷۱۲	۰/۹۲۰		DN	۲/۰۰۴	۰/۹۵۸	۰/۴۸۶	۰/۰۸۶	۰/۸۷۰

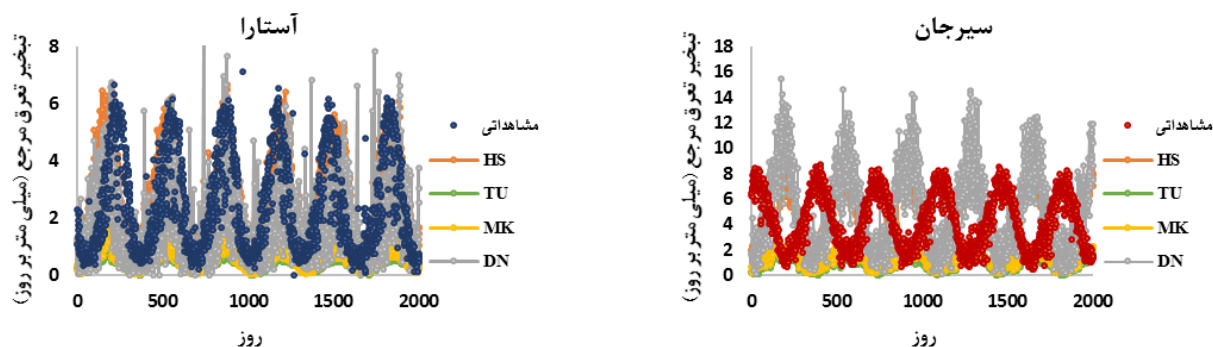
با توجه به نتایج ارائه شده در جدول ۵، در ایستگاه آستارا مدل های SVR_3 و $M5P_3$ با جذر میانگین مربعات خطای $0/201$ ، ضریب همبستگی $0/993$ ، شاخص پراکندگی $0/082$ ، ضریب نش-ساتکلیف $0/987$ و ضریب ویلموت $0/997$ بهترین عملکرد را نسبت به سناریوهای مختلف تعریف شده برای هر یک از مدل های RF ، SVR و $M5P$ و همچنین در مقایسه با روش های تجربی HS ، MK ، TU و DN برای دوره های روزانه مناسب نبوده و احتمالاً برای دوره های ده روزه و یا ماهانه مناسب باشند. سناریوهای ۲ و ۳ به عنوان دو سناریو برتر در پیش بینی مقادیر تبخیر تعرق مرجع برای هر سه روش SVR ، RF و $M5P$ در هر دو ایستگاه آستارا و سیرجان انتخاب شدند. مدل SVR_3 با توجه به اینکه در هر دو ایستگاه عملکرد بهتری داشته است، به عنوان روش برتر معرفی گردید. از طرفی سناریو شماره ۱ در ایستگاه آستارا با توجه به اینکه تنها از پارامتر دما استفاده می نماید، در هر سه مدل نسبت به روش هارگریوز سامانی عملکرد ضعیف داشته است. از روش هارگریوز سامانی در صورت کمبود متغیرهای ورودی در ایستگاه آستارا می توان برای تخمین مقادیر تبخیر تعرق مرجع استفاده کرد. در مقایسه دقت عملکرد روش های تجربی به ترتیب HS ، DN ، MK و TU جایگاه های اول تا چهارم را به خود اختصاص دادند. در ایستگاه آستارا مدل های بکار گرفته شده نسبت به ایستگاه سیرجان عملکرد بهتری داشتند و می توان نتیجه گرفت که مدل های مذکور برای مناطق مرطوب در پیش بینی مقادیر تبخیر تعرق نسبت به مناطق خشک قابلیت بالایی داشته باشند. بررسی نتایج در دو اقلیم نشان می دهد که در اقلیم خشک تغییرات تبخیر تعرق نسبت به زمان بالابوده و این در حالی است که در اقلیم مرطوب تغییرات پایین می باشد. همچنین نمودارهای تغییرات مقادیر تبخیر تعرق مرجع پیش بینی شده برای روش های یادگیری ماشینی و روش های تجربی در مقایسه با نتایج حاصل از روش فائو-پنمن-مونتیث در مرحله صحت سنجی در شکل های ۲ و ۳، نمودارهای پراکنش تنها برای مدل های برتر در شکل ۴ و نمودارهای

با توجه به نتایج ارائه شده در جدول ۵، در ایستگاه آستارا مدل های SVR_3 و $M5P_3$ با جذر میانگین مربعات خطای $0/201$ ، ضریب همبستگی $0/993$ ، شاخص پراکندگی $0/082$ ، ضریب نش-ساتکلیف $0/987$ و ضریب ویلموت $0/997$ بهترین عملکرد را نسبت به سناریوهای مختلف تعریف شده برای هر یک از مدل های RF ، SVR و $M5P$ و همچنین در مقایسه با روش های تجربی HS ، MK ، TU و DN داشته است. در ایستگاه سیرجان مدل SVR_3 با اختلاف کمی از مدل $M5P_3$ با داشتن کمترین جذر میانگین مربعات خطای $0/410$ ، ضریب همبستگی برابر با $0/982$ ، شاخص پراکندگی $0/096$ ، ضریب نش-ساتکلیف $0/963$ و ضریب ویلموت $0/991$ بالاترین دقت را در مقایسه با سایر روش های مورد مطالعه دارد. همچنین در مرتبه دوم نیز در ایستگاه آستارا مدل RF_3 با دارا بودن مقادیر $0/991$ برای ضریب همبستگی و شاخص پراکندگی برابر با $0/097$ ، ضریب نش-ساتکلیف $0/982$ و ضریب ویلموت برابر با $0/995$ و با داشتن خطای کمتر دقت بالاتری را در مقایسه با روش های تجربی مورد مطالعه دارد. در ایستگاه سیرجان نیز همانطور که اشاره شد مدل $M5P_3$ با داشتن ضریب همبستگی $0/980$ ، شاخص پراکندگی برابر با $0/102$ ، ضریب نش-ساتکلیف مساوی با $0/958$ و ضریب ویلموت $0/989$ در جایگاه بعدی برای پیش بینی تبخیر تعرق مرجع قرار می گیرد. با بررسی های به عمل آمده در مقایسه عملکرد با الگوهای مختلف روش های یادگیری ماشینی با عملکرد روابط تجربی مشاهده می گردد که در ایستگاه آستارا دو سناریو از تعداد چهار سناریو تعریف شده برای هر یک از مدل های SVR (SVR_2 و SVR_3)، RF (RF_2 و RF_3) و $M5P$ ($M5P_2$ و $M5P_3$) دقت بیشتر و خطای کمتر نسبت به روش تجربی HS در تخمین مقادیر تبخیر تعرق مرجع دارا می باشند. در ایستگاه سیرجان تمامی مدل های مورد استفاده با ترکیب های مختلف از پارامترهای هواشناسی نسبت به روش های تجربی دارای دقت بالا می باشند. در ایستگاه سیرجان علاوه بر سه روش تجربی MK ، TU و

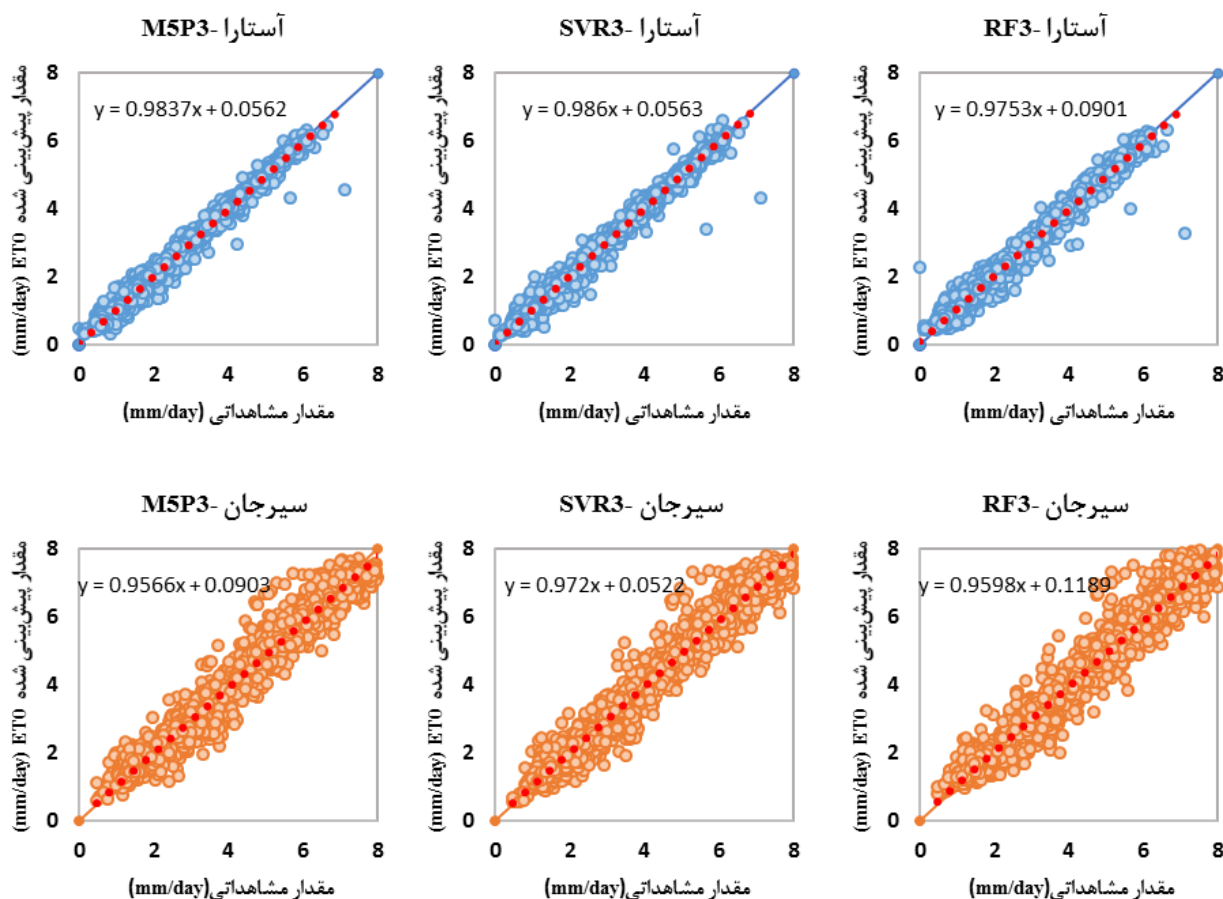
چارک جعبه‌ای برای مدل‌های برتر در شکل ۵ قابل مشاهده می‌باشند.



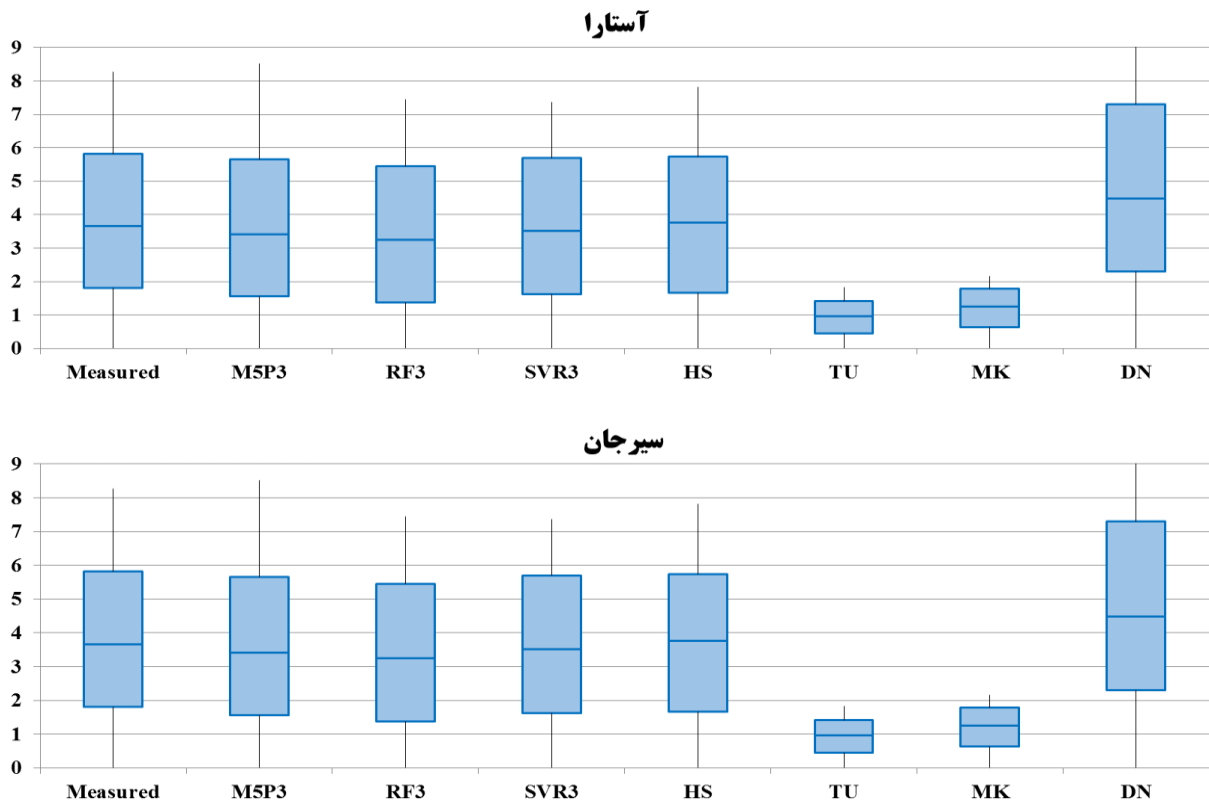
شکل ۲- نمودار تغییرات زمانی مقادیر تبخیرتعرق مرجع با استفاده از روش فائو-پنمن-مانتیت و روش‌های SVR, RF, MSP به ترتیب در آستارا و سیرجان.



شکل ۳- نمودار تغییرات زمانی مقادیر تبخیر تعرق مرجع با استفاده از روش های تجربی هارگریوز- سامانی، مک کینک، تورک و دالتون به ترتیب در آستارا و سیرجان.



شکل ۴- نمودارهای پراکنش تبخیر تعرق مرجع (میلی متر بر روز) با استفاده از مدل های SVR، RF، M5P3 برای مدل های برتر به ترتیب در آستارا و سیرجان.



شکل ۵- نمودارهای چارک جعبه‌ای تبخیرتغرق مرجع (میلی متر بر روز) با استفاده از مدل‌های $M5P_3$, RF_3 , SVR_3 برای مدل‌های برتر به ترتیب در آستارا و سیرجان.

تبخیرتغرق احتمالا عامل دقت بالای مدل‌های مذکور است. همچنین مقایسه نتایج بدست آمده با نتایج حاصل از پژوهش صمدیان فرد و پناهی (۲۰۱۸)، برآورد تبخیرتغرق گیاه مرجع در ایستگاه‌های مناطق نیمه‌خشک با استفاده از روش‌های $M5P_3$ و SVR_3 نشان داد که مدل‌های $M5P_3$ و SVR_3 به عنوان دقیق‌ترین مدل‌های نتیجه شده در این مطالعه با داشتن ضریب و یلموت 0.997 و خطای برابر با 0.201 برآورد نزدیکی از مقادیر تبخیرتغرق گیاه مرجع با مدل SVR_3 با ضریب و یلموت 0.993 و خطای 0.398 گزارش شده در پژوهش آن‌ها ارائه نمود. علاوه بر این، مقایسه نتایج بدست آمده با نتایج حاصل از پژوهش کریمی و همکاران (۲۰۲۰) تطابق مناسبی داشته، به صورتی که در پیش‌بینی تبخیرتغرق مرجع در نواحی مرطوب ایران مدل RF_6 کریمی و همکاران (۲۰۲۰) دارای ضریب نش-ساتکلیف 0.966 و

همانطور که اشاره شد دقت بالای مدل SVR_3 و سناریو ۳ برای هر دو ایستگاه آستارا و سیرجان از شکل‌های ۲ و ۳ نیز قابل مشاهده است، چرا که انطباق نقاط برآورد شده از روش‌های مذکور بر نقاط حاصل از روش فائو-پنمن-مونتیث بیشتر است. از شکل ۴ نیز ملاحظه می‌شود که پراکنش نقاط رسم حول خط نیم‌ساز برای SVR_3 کمتر بوده و نشان از دقت بالای این مدل است. علاوه بر این، نمودارهای جعبه‌ای ارائه شده در شکل ۴ نیز دقت بالا و تطابق بهتر مدل‌های $M5P_3$ و SVR_3 در ایستگاه آستارا و مدل SVR_3 در ایستگاه سیرجان را نشان می‌دهند. در نهایت، بررسی تحلیل‌های انجام شده نشان داد که مدل $M5P_3$ و SVR_3 در ایستگاه آستارا و مدل SVR_3 در ایستگاه سیرجان پیش‌بینی‌های دقیق‌تری از مقادیر تبخیرتغرق گیاه مرجع در مقیاس روزانه داشته‌اند. استفاده از تمامی پارامترهای موثر در پدیده

هارگریوز-سامانی، مک کینک، تورک و دالتون نشان داد که سناریویی که در آن حداقل، متوسط و حداکثر دما، تابش خورشیدی، رطوبت نسبی و سرعت باد به عنوان پارامترهای ورودی مورد استفاده قرار گرفتند، از دقت بالایی برخوردار بود و در ایستگاه آستارا مدل های SVR₃ و M5P₃ عملکرد یکسان و بالایی نسبت به سایر مدل ها داشتند. همچنین، مدل تجربی هارگریوز از برخی از مدل های ماشینی عملکرد بهتری داشت. در ایستگاه سیرجان نیز به عنوان منطقه خشک، مدل SVR₃ مطلوب ترین عملکرد را ارائه داد و روش های تجربی ضعیف ترین عملکرد را ارائه دادند. در نهایت، به عنوان یک نتیجه کلی می توان چنین اظهار داشت که مدل های SVR₃ و M5P₃ توانایی بالایی در پیش بینی مقادیر تبخیر تعرق گیاه داشته و در زمینه محاسبه نیاز آبی گیاهان و برنامه ریزی های آبیاری پیشنهاد می گردند.

مدل RF₃ مطالعه حاضر دارای ضریب نش-ساتکلیف ۰/۹۸۲ می باشد. بنابراین با استناد بر یافته های تحقیق حاضر، استفاده از مدل های SVR₃ و M5P₃ در مناطق مرطوب و مدل SVR₃ در مناطق خشک برای برآورد دقیق مقادیر تبخیر تعرق گیاه مرجع توصیه می گردد.

نتیجه گیری کلی

برآورد دقیق مقدار تبخیر تعرق مرجع در مدیریت منابع آب و بهینه سازی مصارف آب در بخش کشاورزی امری ضروری است. لذا در این مطالعه از داده های هواشناسی دو ایستگاه آستارا و سیرجان برای بررسی توانایی روش های یادگیری ماشینی SVR، RF و M5P برای برآورد مقادیر تبخیر تعرق گیاه مرجع استفاده گردید و مقایسه نتایج حاصل شده با روش های تجربی

منابع مورد استفاده

- Alizade A, Mirshahi B, Hashemi Niya SM and Sanayi Nezhad H, 2001. Evaluation of accuracy and performance of potential evapotranspiration calculated by Hargreaves-Samani and evapotranspiration pan methods in synoptic stations of Khorasan province. *Newar Scientific and Technical Journal* 42: 51-70. (In Persian with English abstract).
- Allen RG, Pereira LS, Raes D and Smith M, 1998. Crop evapotranspiration- guidelines for computing crop water requirements. Irrigation and Drainage Paper no. 56, FAO, Rome, Italy.
- Anonymous, 1997. World Atlas of Desertification. United Nations Environmental Program (UNEP). Editorial commentary by N. Middleton and D.S.G. Edward Arnold, London.
- Ayodele T, Ogunjuyigbe A, Amedu A and Munda J, 2019. Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. *Renewable Energy Focus* 29: 78-93.
- Breiman L, 2001. Random forests. *Machine Learning* 45 (1): 5-32.
- Chen D, Gong L, Xu CY and Halldin S, 2007. A high-resolution, gridded data set for monthly temperature normals (1971-2000) in Sweden. *Geografiska Annaler: Series A, Physical Geography* 89 (4): 249-261.
- Chena H, Huang JJ and McBeana E, 2020. Partitioning of daily evapotranspiration using a modified shuttle worth wallace model, random forest and support vector regression, for a cabbage farmland. *Agricultural Water Management* 228: 105923.
- Dalton J, 1802. Experimental essays on the constitution of mixed gases; on the force of steam of vapour from waters and other liquids in different temperatures, both in a torricellian vacuum and in air on evaporation and on the expansion of gases by heat. *Memoirs and Proceedings of the Manchester Literary & Philosophical Society* 5: 535-602.
- Fallahi MR, Varvani H and Goliyan S, 2012. Precipitation forecasting using regression tree model to flood control. 5th International Watershed and Water and Soil Resources Management. 29 February & 1 March, Kerman, Iran.
- Feng K and Tian J, 2021. Forecasting reference evapotranspiration using data mining and limited climatic data. *European Journal of Remote Sensing* 54(S2): 363-371.
- 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. *Agricultural Water Management* 193: 163-173.

- Gharahdaghi MH, Homae M, Mirlatifi SM and Noroozi AA, 2019. Using forecasts of WRF regional model to improve the accuracy of reference evapotranspiration estimation. *Iranian Soil and Water Research* 51(1): 165-177. (In Persian with English abstract).
- Gonzalez del Cerro RT, Subathra MSP, Manoj Kumarc N, Verrastro S and George ST, 2021. Modelling the daily reference evapotranspiration in semi-arid region of south India: A case study comparing ANFIS and empirical models. *Information Processing in Agriculture* 8:173– 184.
- Granata F, 2019. Evapotranspiration evaluation models based on machine learning algorithms- A comparative study. *Agricultural Water Management* 217: 303-315.
- Guitjens JC, 1982. Models of alfalfa yield and evapotranspiration. *Journal of the Irrigation and Drainage Division* 108 (IR3): 212–222.
- Harbeck GE, 1962. A practical field technique for measuring reservoir evaporation utilizing mass-transfer theory. *U.S. Geological Survey*, 272-E:101–105.
- Hastie T, Tibshirani R and Friedman J, 2009. *The Elements of Statistical Learning*. Springer, New York.
- Judi A and Sattari MT, 2015. Estimation of bridge base scour depth in a aqueous structures by gaussian process regression method. *Journal of Applied Research in Irrigation and Drainage Structures Engineering*. 16(65):19-36. (In Persian with English abstract).
- Karimi S, Shiri J and Martic P, 2020. Supplanting missing climatic inputs in classical and random forest models for estimating reference evapotranspiration in humid coastal areas of Iran. *Computers and Electronics in Agriculture* 176:1-13.
- Leib B, Sassenrath G and Schmidt AM, 2012. Irrigation scheduling tools. Pp.32-37 In: Perry C and Barnes E, (eds.). *Cotton Irrigation Management for Humid Regions*. Cotton, Incorporated, Cary, NC.
- Makkink GF, 1957. Testing the penman formula by means of lysimeters. *Journal Institute of Water Engineering* 11: 277-288.
- Noruzi H, Asghari Moghaddam A and Nadiri AA, 2015. Determination of vulnerable areas of Malekan plain aquifer to nitrate using random forest method. *Journal of Environmental Science* 41(4): 923-942.
- Samadianfard S and Panahi S, 2018. Estimating daily reference evapotranspiration using data mining methods of support vector regression and M5 model tree. *Journal of Watershed Management Research* 10(18):157-167. (In Persian with English abstract).
- Sattari MT, Nahrein F and Azimi V, 2013. M5 model trees and neural networks based prediction of daily ETO (case study: Bonab Station). *Iranian Journal of Irrigation and Drainage* 1(7):104-113. (In Persian with English abstract).
- 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.
- 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. *Iranian Journal of Irrigation and Drainage* 2: 378-388. (In Persian with English abstract).
- Tao H, Diop L, Bodian A, Djaman K, Ndiaye PM and Mundher Yaseen Z, 2018. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso. *Agricultural Water Management* 208: 140–151.
- Turc L, 1961. Estimation of irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date. *Annals of Agronomy* 12: 13-49.
- Vapnik VN, 1995. *The Nature of Statistical Learning Theory*. Springer, New York. 314P.
- Wang YM, Traore S and Kerh T, 2008. Neural network approach for estimating reference evapotranspiration from limited climatic data in Burkina Faso. *WSEAS Transactions on Computers* 7: 704-713.
- Yao W, Zhang C, Hao H, Wang X and Li X, 2018. A support vector machine approach to estimate global solar radiation with the influence of fog and haze. *Renewable Energy* 128:155–162.
- Zeinyal Zade K and Khan Mohammadi N, 2018. Comparison of the efficiency of linear and nonlinear time series models in simulating and predicting reference evapotranspiration. *Journal of Geography and Planning* 63: 139-160.