Water and Soil Science, Vol.33, NO.4, Winter 2023, pp. 167-184.

DOI: 10.22034/ws.2021.49033.2450 ISSN:2717-3755 https://water-soil.tabrizu.ac.ir/

Research Article

Estimation of Daily Evaporation from Class A Pan Using Five Data Mining Methods (Case Study: Tabriz Synoptic Station)

KH Seifzadeh[1], D Zare Haghi[2*], S Samadianfard[3], MR Neyshabouri[4], F Mikaeili[5]

Received: October 28, 20212 Accepted: May 22, 2022
Revised: April 17, 2022 Published online: December 22, 2023

1-Former M.Sc. student, Dept. of Soil Sci. and Eng., Faculty of Agric., University of Tabriz, Iran.
2- Assoc. Prof, Dept. of Soil Sci. and Eng., Faculty of Agric., University of Tabriz, Iran
3-Assoc. Prof, Dept. of Water Sci. and Eng., Faculty of Agric., University of Tabriz, Iran
4-Prof, Dept. of Soil Sci. and Eng., Faculty of Agric., University of Tabriz, Iran
5-M.Sc. student, Dept. of Water Sci. and Eng., Faculty of Agric., University of Tabriz, Iran
*Corresponding Author, Email: davoudhaghi@tabrizu.ac.ir

Abstract

Background and Objectives

Evaporation is one of the main components of hydrological cycle and one of the effective climatic variables in arid areas such as Iran. Accurate estimation of evaporation rate plays an important role in sustainable development and optimal management of water resources. Evaporation is one of the essential processes, because it depends on meteorological variables such as solar radiation, air temperature, wind speed, relative humidity and atmospheric pressure, which are related to the topography and the climate of the region. Class A pan-evaporation is one of the standard and direct tools for measuring evaporation, which is used all over the world due to its ease of application for determining evaporation. However, in most stations accurate evaporation recording is not practical due to instrument limitations and maintenance problems. On the other hand, the temporal and spatial distribution of evaporation stations compared to meteorological stations is limited, therefore according to the problems mentioned, the use of meteorological variables for estimating the rate of evaporation from the pan will be useful. In different regions, the impact of different climatic factors on changes evaporation from the pan has not be fully understood, so the relatively accurate estimation and prediction of this phenomenon is an effective step in the relevant fields. In recent years, for estimating the amount of evaporation from the pan, a variety of intelligent systems and software calculations such as data mining methods have been developed.

Methodology

In this study, meteorological data of Tabriz Station in the period of 2003 to 2018 have been used to estimate the evaporation values from the class A pan. For this purpose, a simple correlation between meteorological variables and evaporation from class A pan was created and based on the result of this correlation, in the studied station the minimum temperature and relative humidity were inversely and the maximum and average temperature were directly affected by evaporation. Thus, ten combined scenarios were defined and modeling was performed using support vector regression (SVR), Gaussian process regression (GPR), M5tree, random forest (RF) and linear regression (LR) methods. It should be noted that in this study, 70% of the data were selected for training and 30% for testing. Finally, the performance of each method in estimating evaporation values was

©: The Author(S)
Publisher: University of Tabriz

evaluated using root mean squared error (RMSE), mean absolute error (MAE), Nash- Sutcliffe coefficient (NS) and Akaike information criterion (AIC).

Findings

The results showed that GPR10 method with RMSE = 1.90 mm day-1, MAE = 1.48, NS = 0.81 and SVR10 method with RMSE = 1.92 mm day-1, MAE = 1.51, NS = 0.8 had reasonable performance in estimating the values of daily evaporation from class A pan. The GPR method showed its higher capability to estimate daily evaporation values in all definition scenarios with the least error and the most accuracy. The SVR model with appropriate results was in the second place. The results of statistical parameters for random forest model were even weaker than the results of linear regression. In general, scenario number 10 with all meteorological variables and scenario number 1 with only the input of minimum temperature variable had the best and weakest results among all defined scenarios, respectively. Scenarios 6 to 10 have more accuracy and less error and modeling structures with the least number of variables has the least accuracy. Also, wind speed and solar radiation variables were introduced as the most effective factors for estimating the evaporation rate from class A pan.

Conclusion

Evaporation is one of the important processes that cause the losses of half of precipitation in arid and semi- arid regions. Accordingly, the rate of evaporation and its modeling as one of the hydrological variables in agricultural research and factors related to soil and water are of importance. Therefore, accurate estimation of this phenomenon is essential. In this study, meteorological data from Tabriz Station were utilized to assessment the capability of machine learning methods. Evaporation values were estimated using five data mining methods including SVR, GPR, M5, RF and LR. Conclusively, the results of evaluation criteria indicated that GPR and SVR models using all variable of meteorological data performed more accurate than others. Finally, both of them are recommended to estimate the amount of evaporation from class A pan.

Keywords: Evaporation, Gaussian process regression, Linear regression, Random Forest, Support vector regression.

نشریه دانش آب و خاک / جلد ۳۳ شماره ۴ صفحه‌های ۱۶۷ تا ۱۸۴/ سال ۱۴۰۲

مقاله پژوهشی

برآورد تبخیر روزانه از تشت کلاس A با استفاده از پنج روش داده‌کاوی (مطالعه موردی: ایستگاه هواشناسی تبریز)

خدیجه سیف‌زاده[۱]، داوود زارع حقی [۲*]، سعید صمدیان فرد[۳] ، محمدرضا نیشابوری[۴]، فاطمه میکائیلی[۵]

تاریخ دریافت: ۱۴۰۰/۰۸/۰۶ تاریخ پذیرش: ۱۴۰۱/۰۳/۰۱

تاریخ ویرایش:۱۴۰۱/۰۱/۲۸ تاریخ انتشار انلاین :۱۴۰۲/۱۰/۰۱

۱- دانش آموخته کارشناسی ارشد فیزیک و حفاظت خاک، دانشکده کشاورزی، دانشگاه تبریز

۲- دانشیار، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز

۳- دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

۴- استاد، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز

۵- دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

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

چکیده

تبخیر یکی از عوامل اثرگذار در چرخه هیدرولوژیکی است که تخمین صحیح آن نقش مهمی در توسعه پایدار و مدیریت بهینه منابع آب در کشورهای مواجه با بحران آب ایفا می‌کند. هدف از این پژوهش، ارزیابی عملکرد روش‌های داده‌کاوی جهت برآورد تبخیر روزانه از تشت کلاس A در ایستگاه تبریز می‌باشد. در این پژوهش از داده‌های هواشناسی روزانه ایستگاه تبریز در طی دوره ١٦ ساله (۲۰۱۸- ۲۰۰۳) استفاده گردید. برآورد میزان تبخیر از تشت کلاس A با استفاده از روش‌های رگرسیون بردار پشتیبان (SVR)، رگرسیون فرآیند گاوسی (GPR)، مدل درختی M5، جنگل تصادفی (RF) و رگرسیون خطی (LR) انجام گرفت. ۱۰ سناریو ترکیبی بر اساس همبستگی بین متغیرهای هواشناسی و تبخیر برای واسنجی و صحت‌سنجی روش‌های مورد مطالعه مدنظر قرار گرفت. نتایج بررسی‌های آماری نشان داد که در ایستگاه تبریز، مقادیر تخمینی تبخیر روش GPR با جذر میانگین مربعات خطای برابر با ۱/۹ میلی‌متر بر روز و ضریب نش- ساتکلیف برابر با ۰/۸۱ و در روش SVR با جذر میانگین مربعات خطای برابر با ۱/۹۲ میلی‌متر بر روز و ضریب نش- ساتکلیف ۰/۸۰، از عملکرد مناسبی در شبیه‌سازی مقدار تبخیر روزانه از تشت کلاس A برخوردار بوده‌اند. در نهایت برای ایستگاه هواشناسی تبریز، مدل‌های GPR و SVR برای سناریو شماره ۱۰ با همه متغیرها و دارا بودن بهترین عملکرد، به‌عنوان مدل‌هایی با دقت مناسب پیشنهاد گردید. همچنین متغیرهای سرعت باد و تابش خورشیدی به‌عنوان موثرترین متغیرها در برآورد میزان تبخیر از تشت کلاس A معرفی شدند.

واژه‌های کلیدی: تبخیر، جنگل تصادفی، رگرسیون بردار پشتیبان، رگرسیون خطی، رگرسیون فرآیند گاوسی.

مقدمه

تبخیر به‌عنوان یکی از اجزای اصلی در چرخه آب و از متغیرهای اقلیمی موثر در مناطق خشک و کم آبی مثل ایران بوده و تخمین این پدیده در زمینه‌های مختلفی از قبیل مطالعات کشاورزی، هیدرولوژیکی، هواشناسی، بهره‌برداری از مخازن سدها، طراحی سیستم‌های آبیاری و زهکشی، زمان‌بندی آبیاری و مدیریت منابع آب اهمیت فراوانی دارد. بر همین اساس آگاهی از مقدار تبخیر و مدل‌سازی آن به عنوان یکی از مهم‌ترین متغیرهای هیدرولوژیکی در تحقیقات کشاورزی و عوامل مربوط به آب و خاک از اهمیت زیادی برخوردار است (افخمی و همکاران ۲۰۱۸). تبخیر فرآیندی غیرخطی و پیچیده است (ژیانگ هو و همکاران ۲۰۱۸)، چرا که به متغیرهای هواشناسی نظیر تابش خورشیدی، دمای هوا، سرعت باد، رطوبت نسبی و فشار اتمسفر بستگی دارد که خود این متغیرها در ارتباط با توپوگرافی منطقه و اقلیم منطقه می‌باشند (ویسنت و همکاران ۲۰۱۸). تشت تبخیر کلاس A یکی از ابزارهای استاندارد و مستقیم اندازه‌گیری تبخیر است که در سرتاسر دنیا به علت سهولت در کاربرد به عنوان شاخصی در تعیین تبخیر از دریاچه‌ها و مخازن مورد استفاده قرار می‌گیرد (ایرمیک و همکاران ۲۰۰۲). در بیشتر ایستگاه‌ها، ثبت دقیق تبخیر به دلیل محدودیت‌های ابزاری و مشکلات نگهداری عملی نیست (شادمانی و معروفی ۲۰۱۱). از سوی دیگر توزیع زمانی و مکانی ایستگاه‌های تبخیر سنجی نسبت به ایستگاه‌های هواشناسی محدود است، بنابراین با توجه به مشکلات ذکر شده کاربرد متغیرهای هواشناسی در تخمین میزان تبخیر از تشت مفید خواهد بود. (کیشی ۲۰۰۶، دهقانی و همکاران ۲۰۱۰). در مناطق مختلف هنوز به میزان تاثیر متغیرهای مختلف آب و هوایی در تغییرات تبخیر از تشت به طور کامل پی برده نشده است، از این رو برآورد و پیش‌بینی نسبتا دقیق این پدیده برای برنامه‌ریزی، مدیریت منابع آب و مدل‌سازی دارای اهمیت ویژه‌ای است. در سال‌های اخیر انواع سیستم‌های هوشمند و محاسبات نرم‌افزاری مانند روش داده‌کاوی، برای تخمین تبخیر از تشت با استفاده از داده‌های هواشناسی توسعه یافته‌اند (گویال ۲۰۱٤)، که ماهیت این روش و مدل‌های آن محاسباتی بوده و دقت و کارآیی این مدل‌ها به نوع و تعداد ورودی آن‌ها بستگی دارد (ترزی ۲۰۰۷). در زمینه محاسبه تبخیر مطالعاتی در سطح جهان و ایران انجام گرفته است که در ادامه به چند نمونه از این مطالعات اشاره می‌گردد. طبری و همکاران (۲۰۱۰) با استفاده از شبکه عصبی مصنوعی و رگرسیون غیرخطی در استان همدان تبخیر روزانه از تشت را تخمین زدند و نتیجه گرفتند که تبخیر به‌دست آمده به روش شبکه عصبی مصنوعی دارای خطای کم بوده و دما و سرعت باد موثرترین عوامل در برآورد تبخیر از تشت بودند. ستاری و همکاران (۲۰۱۳) در تحقیق خود جهت برآورد دقیق مقدار تبخیر و تعرق مرجع روزانه در منطقه بناب استان آذربایجان شرقی، ابتدا بر اساس یک روش استاندارد و متداول فائو- پنمن- مونتیث و داده‌های هواشناسی ایستگاه بناب، مقدار تبخیر و تعرق مرجع را محاسبه کردند. سپس با ارائه سناریوهای مختلفی از ترکیب پنج متغیر به عنوان ورودی مدل شبکه‌های عصبی مصنوعی و مدل درختی M5 در مقیاس زمانی روزانه سعی در برآورد دقیق‌تری از مقادیر تبخیر و تعرق مرجع به عنوان خروجی مدل داشتند. نتایج حاصل نشان داد که اگرچه روش شبکه عصبی مصنوعی با اختلاف خیلی کمی در مقایسه با روش M5 پیش‌بینی نسبتا دقیق‌تری از میزان تبخیر و تعرق مرجع ارائه می‌کند، اما روش M5 روابط ساده خطی، قابل فهم‌تر و کاربردی‌تری برای پیش‌بینی تبخیر و تعرق ارائه می‌کند. گویال و همکاران (۲۰۱٤) در مطالعه‌ای توانایی شبکه عصبی مصنوعی، حداقل مربعات رگرسیون بردار پشتیبان، سیستم استنتاج عصبی فازی- تطبیقی را در برآورد تبخیر

روزانه از تشت در آب و هوای نیمه‌گرمسیری در کارسو هندوستان بررسی کردند. براساس مقایسه‌های انجام گرفته، سیستم استنتاج عصبی فازی- تطبیقی و حداقل مربعات رگرسیون بردار پشتیبان در برآورد تبخیر از تشت با توجه به داده‌های در دسترس، نتایج بهتری داشته است. ژیانگ هو و همکاران (۲۰۱۸) در منطقه آبخیز دریاچه پویانگ در کشور چین به بررسی تبخیر روزانه از تشت با استفاده از چهار مدل تجربی و سه مدل یادگیری ماشینی شامل مدل درختی M5، جنگل‌های تصادفی (RF)[1] و درخت تصمیم‌گیری پرداختند، نتایج حاکی از آن بود که روش پریستلی تیلور بالاترین دقت را در تخمین دارد و مدل درختی M5 و مدل‌های تجربی با ترکیب ورودی مشابه از نظر دقت پیش‌بینی به طور متوسط عملکرد بهتری دارند. محمدی و همکاران (۲۰۱۸) در پژوهش خود با استفاده از داده‌های تبخیر روزانه ایستگاه سینوپتیک رودسر کارایی الگوریتم هیبریدی کرم شب تاب (FA)[2] را در برآورد تبخیر روزانه در مقایسه با شبکه عصبی مصنوعی مورد ارزیابی قرار دادند. دستاوردهای حاصل از این پژوهش نشان داد که FA در برآورد تبخیر، در همه سناریوهای تعریف شده موفق‌تر از شبکه عصبی مصنوعی عمل کرده است. صمدیان فرد و همکاران (۲۰۱۸) در ایستگاه‌های تبریز و جلفا برای برآورد و شبیه‌سازی مقادیر تبخیر از تشت، از روش‌های برنامه‌ریزی ژنتیک (GP)[3]، رگرسیون بردار پشتیبان (SVR)[4] و شبکه عصبی مصنوعی استفاده کردند. نتایج حاصل نشان دادند که در ایستگاه‌های تبریز و جلفا به ترتیب GP و SVR عملکرد مناسبی در شبیه‌سازی مقدار تبخیر داشته‌اند و برای ایستگاه تبریز، مدل با متغیرهای ورودی دما و سرعت باد و برای ایستگاه جلفا مدل با ورودی همه متغیرها به عنوان مدل‌هایی با دقت مناسب برای شبیه‌سازی تبخیر پیشنهاد شدند. قاسم و همکاران (۲۰۱۹) در مطالعه خود کاربرد رگرسیون بردار پشتیبان موجک و شبکه‌های عصبی مصنوعی موجک را برای تخمین تبخیر از تشت در ایستگاه‌های تبریز و آنتالیا مورد بررسی قرار دادند. نتایج به‌دست آمده نشان دادند که شبکه عصبی مصنوعی عملکرد بهتری در برآورد تبخیر دارد. آلسومایی (۲۰۲۰) با بررسی کاربرد شبکه‌های عصبی مصنوعی در مدل‌سازی تبخیر از تشت در اقلیم‌های فراخشک، گزارش کرد که مدل‌های شبکه عصبی مصنوعی جهت‌گیری قابل توجهی در پیش‌بینی نرخ بالای تبخیر از تشت (<۲۵ میلی‌متر بر روز) نشان داد. مالیک و همکاران (۲۰۲۱) با استفاده از روش جدید رگرسیون بردار پشتیبان بهینه‌سازی با الگوریتم ازدحام سلپ در ارتباط با آزمون گاما، به تخمین تبخیر از تشت در مناطق مختلف زراعی- اقلیمی پرداختند. نتایج حاصل نشان دادند که روش جدید SVR-SSA-5[5] بهترین عملکرد را در بین سایر روش‌های مورد استفاده ارائه کرده و در نتیجه این مدل به‌عنوان مدلی قوی و قابل اعتماد در پیش‌بینی تبخیر از تشت انتخاب شد.

با توجه به مطالعات انجام شده در زمینه تخمین تبخیر، می‌توان به اهمیت برآورد تبخیر پی برد. در پژوهش حاضر، به برآورد تبخیر با استفاده از داده‌های هواشناسی با کاربرد پنج مدل روش داده‌کاوی، شامل رگرسیون بردار پشتیبان (SVR)، رگرسیون فرآیند گاوسی (GPR)[6]، مدل درختی M5، الگوریتم جنگل تصادفی (RF) و رگرسیون خطی (LR)[7] پرداخته شده است. در سالیان اخیر، توانایی روش‌های داده محور مذکور در زمینه‌های مختلف علوم و مهندسی مورد توجه و بررسی محققین قرار گرفته است، لذا این تحقیق بر پایه ارزیابی همزمان روش‌های اشاره شده در

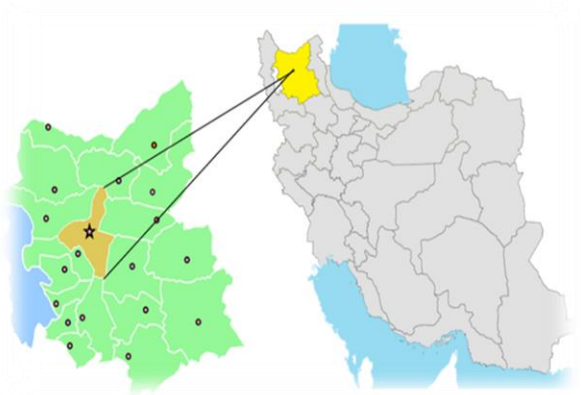
[1] Random Forest
[2] Firefly Algorithm
[3] Genetic Programming
[4] Support Vector Regression
[5] Support Vector Regression- Slap Swarm Algorithm
[6] Gaussian Process Regression
[7] Linear Regression

تخمین مقادیر تبخیر روزانه از تشت در تبریز پایه‌ریزی گردید. در نهایت، با بررسی و تحلیل آماری، مناسب‌ترین روش پیشنهاد شد.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

استان آذربایجان‌شرقی از بزرگ‌ترین و پرجمعیت‌ترین نواحی شمال غربی ایران محسوب می‌شود و از نظر موقعیت جغرافیایی در محدوده ٤٥ درجه و ٧ دقیقه الی ٤٨ درجه و ٢٠ دقیقه طول شرقی و ٣٦ درجه و ٤٥ دقیقه الی ٣٩ درجه و ٢٦ دقیقه عرض شمالی قرار گرفته است. مرکز استان آذربایجان‌شرقی کلان‌شهر تبریز است که در حدود ١٣٤٨ تا ١٥٦١ متر از سطح دریا ارتفاع دارد که دارای آب و هوای سرد کوهستانی بوده و کل محدوده آن را کوه‌ها و ارتفاعات تشکیل داده‌اند (صمدیان فرد و پناهی ٢٠١٨). میانگین بارش سالانه ایستگاه تبریز ٣١٠ میلی‌متر بوده و تعداد روزهای یخبندان در سال ١٠٤ روز می‌باشد (شکل ١).



شکل ١- موقعیت مکانی منطقه مورد مطالعه.

روش پژوهش

در این پژوهش، با استفاده از داده‌های روزانه هواشناسی ١٦ سال آماری مربوط به سال‌های ٢٠١٨-٢٠٠٣ و داده‌های اندازه‌گیری شده تبخیر از تشت (E_{pan}) برای بازه زمانی ذکر شده، به تخمین میزان تبخیر پرداخته شد. در مطالعه حاضر، امکان‌سنجی استفاده از متغیرهای هواشناسی روزانه ایستگاه هواشناسی تبریز شامل: حداکثر دما (T_{max})، حداقل دما (T_{min})، متوسط دما (T_{mean})، رطوبت نسبی(RH)، سرعت باد در ارتفاع دو متری (U_2) و تابش خورشیدی (S) و تبخیر از تشت در ترکیب‌های مختلف به‌منظور تخمین تبخیر روزانه مدنظر قرار گرفت. بدین منظور، مشخصات آماری متغیرهای مذکور در ایستگاه‌های مورد مطالعه در جدول ١ ارائه شده است. اطلاعات درج شده در جدول ١ بیانگر اهمیت تخمین مولفه تبخیر در وضعیت بیلان آبی منطقه مورد مطالعه می‌باشد، بطوریکه میزان تبخیر در حدود ٣٠ میلی‌متر در روز نیز میزان اهمیت برآورد صحیح آن را بیش از پیش نمایان می‌سازد.

جدول۱- مشخصات آماری متغیرهای اقلیمی ایستگاه تبریز در طول دوره ۱۶ ساله (۲۰۱۸- ۲۰۰۳).

مشخصات آماری					
متغیر	واحد	میانگین	کمینه	بیشینه	انحراف معیار
T_{min}	(°C)	۱۳/٤۸	-۷/٦	۲۸/۸	٦/٤۹
T_{max}	(°C)	۲٦/۲۸	۰/٤	٤۱	۷/۵۲
T_{mean}	(°C)	۱۹/٦٦	-۳/۵	۳۵/۱	٦/۹۳
U_2	(m/s)	۳/۹۳	۰/۲۵	۱۰/۷۵	۱/۵۷
RH	(%)	٤۲/٤۲	۱۳/۵	۹٤	۱٤/۹۳
S	($MJ\ m^{-2}\ day^{-1}$)	۹/۲۳	۰	۱۳/۹	۳/۱۹
E_{pan}	($mm\ day^{-1}$)	۷/۵٦	۰	۳۰/٤	٤/۳۰

مدل‌سازی با استفاده از روش‌های رگرسیون بردار پشتیبان (SVR)، رگرسیون فرآیند گاوسی (GPR)، مدل درختی M5، الگوریتم جنگل تصادفی (RF) و رگرسیون خطی (LR) انجام گرفت و برای اجرای مدل‌ها از نرم‌افزار Weka استفاده گردید. در این پژوهش ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای آزمون در نظرگرفته شدند. علاوه‌براین، همبستگی بین شش پارامتر هواشناسی و تبخیر در ایستگاه تبریز در جدول ۲ ارائه شده است. آن‌چنان که در جدول ۲ نشان داده شده است، دمای حداقل روزانه و رطوبت نسبی به صورت معکوس و دمای حداکثر و متوسط دمای روزانه، به صورت مستقیم بر روی تبخیر تاثیرگذار می‌باشند. با توجه به وضعیت اقلیمی منطقه مورد مطالعه، بین متغیرهای میانگین دما، رطوبت نسبی هوا و مولفه تبخیر همبستگی بالایی مشاهده می‌شود.

جدول۲- همبستگی بین شش متغیر اقلیمی و تبخیر در ایستگاه هواشناسی تبریز

متغیر	T_{min} (°C)	T_{max} (°C)	T_{mean} (°C)	S ($MJ\ m^{-2}\ day^{-1}$)	RH (%)	U_2 ($m\ s^{-1}$)	E_{pan} ($mm\ day^{-1}$)
T_{min}	۱	۰/۹۲٤	۰/۹۵۸	۰/٤٤٤	-۰/٦۱۰	۰/٤۳٤	۰/۸۱۵
T_{max}		۱	۰/۹۸۸	۰/٦۲۵	-۰/۷٦۵	۰/۳۰۸	-۰/۸۰۳
T_{mean}			۱	۰/۵۸۲	-۰/۷۳۷	۰/۳۳۵	۰/۸۱۹
S				۱	-۰/٦٤۵	۰/۱٤۸	۰/۵۱۳
RH					۱	-۰/۲۰۹	-۰/٦۳۷
U_2						۱	۰/٤٦۷
E_{pan}							۱

رگرسیون بردار پشتیبان (SVR)

رگرسیون بردار پشتیبان یک نوع سیستم یادگیری نظارت شده است که در سال ۱۹۹۵ توسط واپنیک معرفی شد. ماشین بردار پشتیبان، هم برای گروه‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها در مسائل رگرسیونی به‌کار می‌رود. ماشین‌های بردار پشتیبان شامل دو دسته‌اند، طبقه‌بندی کننده بردار پشتیبان و رگرسیون بردار پشتیبان. در برخی مواقع، به ساختارهای پیچیده و غیرخطی برای جداسازی داده‌ها نیاز است. در این صورت، ماشین بردار پشتیبان

داده‌های اصلی را با به‌کارگیری مجموعه ای از توابع ریاضی که کرنل نام دارند، در فضای جدیدی نگاشت و بازآرایی می‌کند که به این کار تبدیل نگاشت گفته می‌شود (جودی و ستاری ۲۰۱۶). بساک و همکاران (۲۰۰۷) برای کاربرد ماشین‌های بردار پشتیبان در مسائل رگرسیون از تابع خطایی استفاده کردند که خطاهایی که در یک ε-Insensitive به نام فاصله معین از مقادیر واقعی می‌باشند، را نادیده می‌گیرد. این تابع به‌صورت زیر تعریف می‌شود:

$$L\big(y,f(x,a)\big)=|y-f(x,a)|_{\varepsilon}\rightarrow \begin{cases}0 & for\ |y-f(x,a)|\le\varepsilon\\ |y-f(x,a)|-\varepsilon & if\ |y-f(x,a)|>\varepsilon\end{cases} \quad [۱]$$

این تابع خطا مقادیر خطای کمتر از ε را در نظر نمی‌گیرد.

رگرسیون فرآیند گاوسی (GPR)

رگرسیون فرآیند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد محدودی از آنها با توزیع‌های گاوسی سازگار شده‌اند. این فرآیند به‌طور کامل توسط تابع میانگین و تابع کواریانس معین می‌شود. این فرآیند تعمیمی طبیعی از توزیع گاوسی است که میانگین و کواریانس آن به‌ترتیب یک بردار و یک ماتریس است. مدل‌های رگرسیون فرآیند گاوسی برمبنای این فرض هستند که داده‌ها باید حاوی اطلاعاتی درباره همدیگر باشند. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی بوده، درحالی‌که فرآیند گاوسی بیانگر توزیع بین توابع می‌باشد. در نتیجه مدل‌های فرایند گاوسی به دلیل دانش قبلی درباره وابستگی‌های تابعی و داده‌ها برای تعمیم به هیچ فرآیند اعتبارسنجی نیاز ندارند (جودی و ستاری ۲۰۱۵).

فرآیند گاوسی $f(x)$ توسط توابع میانگین $m(x)$ و کواریانس به شکل زیر تعریف می‌شود:

$$m(x)=E(f(x)) \quad [۲]$$

$$k(x,\acute{x})=E(f(x)-m(x))(f(\acute{x})m(\acute{x})) \quad [۳]$$

که در روابط فوق $k(x,\acute{x})$ تابع کواریانس بوده که در نقاط x و $\acute{x}$ محاسبه می‌شود. فرآیند گاوسی $f(x)$ می‌تواند به‌صورت زیر بیان گردد:

$$f(x)\sim GP(m(x),k(x,\acute{x})) \quad [٤]$$

که معمولاً جهت ساده‌سازی، مقدار تابع میانگین برابر با صفر در نظر گرفته می‌شود (صاقبیان و روشنگر ۲۰۱۹).

مدل درختی M5

این الگوریتم، زیرمجموعه‌ای از روش‌های یادگیری ماشینی و داده‌کاوی است. روش‌های مبتنی بر درخت تصمیم به‌عنوان یکی از شناخته شده‌ترین تکنیک‌های داده‌کاوی، ویژگی هدف را به‌عنوان خروجی به‌صورت یک مدل با سازه درختی با استفاده از داده‌های ورودی تخمین یا طبقه‌بندی می‌نمایند. این مدل با ارائه روابط خطی ساده و قابل فهم، توانایی پیش‌بینی مقادیر تبخیر را با استفاده از داده‌های هواشناسی دارد. داده‌کاوی به‌عنوان روش‌های مبتنی بر داده و الگوریتم یادگیری ماشینی با بهره‌گیری از داده‌های ثبت شده به کشف روابط و الگوهای پنهان در داخل داده‌ها می‌پردازد (ستاری و همکاران ۲۰۱۳). برای ایجاد یک درخت تصمیم از معیار انشعاب براساس کاهش عملکرد انحراف استاندارد (SDR) مقادیر هر کلاس (آنتروپی) استفاده می‌شود.

$$SDR=Sd(T)-\sum_{i=1}^{N}\frac{|T_i|}{|T|}Sd(T_i) \quad [۵]$$

$$Sd(T) = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right)} \qquad [۶]$$

که T مجموعه‌ای از نمونه‌هاست، T_i نشان‌دهندهی مجموعه‌ای از نمونه‌هاست که i امین نتیجه تست پتانسیلی را دارند، S_d بیانگر انحراف معیار، y_i مقدار عددی ویژگی هدف نمونه و N تعداد داده‌ها می‌باشد.

جنگل تصادفی (RF)

جنگل تصادفی هم توانایی رگرسیون و هم توانایی طبقه‌بندی را با هم دارد. الگوریتم جنگل تصادفی در صورتی قادر به کشف درست اطلاعات پنهان است که ورودی‌های مناسبی انتخاب گردد. مدل پیش‌بینی کننده جنگل تصادفی بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوط استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی انجام می‌دهد. روش کار در یک جنگل تصادفی چنین است که داده‌های نمونه‌ای با عمل جایگذاری به n مجموعه نمونه تقسیم می‌شود و از هر دسته، نمونه‌ای برای آموزش یک درخت استفاده می‌شود. تمام درخت‌ها عمق مشخصی دارند و در هر گره تقسیم، یک ویژگی به تصادف از بین مجموعه ویژگی‌ها انتخاب می‌شود و شاخه‌بندی بر اساس آن صورت می‌گیرد. این الگوریتم به‌علت استفاده از چندین دسته نمونه، مشکل داده‌های پرت و گم‌شده را ندارد (ذبیحی و همکاران ۲۰۱۵). 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:

$$E_{x,y} = (Y - h(x))^2 \qquad [۷]$$

تشکیل RF با میانگین‌گیری از j امین درخت منفرد صورت می‌گیرد. با افزایش تعداد درختان جنگل، خطا بصورت زیر برآورد می‌شود (بریمن ۲۰۰۱، کریمی و همکاران ۲۰۲۰):

$$E_{x,y} = \left(Y_{avj} h(x,\theta_j)\right)^2$$

$$\rightarrow E_{x,y}(y - E_\theta h(x,\theta))^2 \qquad [۸]$$

رگرسیون خطی (LR)

تحلیل رگرسیون خطی این امکان را فراهم می‌آورد تا تغییرات متغیر وابسته را از طریق متغیرهای مستقل پیش‌بینی و سهم هریک از متغیرهای مستقل را در تبیین متغیر وابسته تعیین کند. رگرسیون رابطه نزدیکی با ضریب همبستگی دارد، بدین معنا که برای انجام رگرسیون، باید ضریب همبستگی را محاسبه کرد. هرچه ضریب همبستگی بین متغیرها قوی‌تر باشد، پیش‌بینی نیز دقیق‌تر خواهد بود (کلانتری ۲۰۰۶).

معیارهای ارزیابی دقت روش‌های مورد استفاده

برای ارزیابی دقت روش‌های موردمطالعه در برآورد تبخیر، از معیارهای آماری ضریب نش-ساتکلیف (NS)[1]، جذر میانگین مربعات خطا (RMSE)[2]، متوسط خطای مطلق (MAE)[3] و معیار ارزیابی اطلاعات (AIC)[4] استفاده گردید.

$$NS = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \qquad [۹]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \qquad [۱۰]$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}[O_i - P_i] \qquad [۱۱]$$

[1] Nash-Sutcliffe coefficient
[2] Root mean square error
[3] Mean absolute error
[4] Akaike information criterion

$$AIC = n \times ln(RMSE) + 2k \quad [۱۲]$$

معیار ارزیابی اطلاعات (AIC) برای تشخیص مدل مناسب از بین مدل‌های موجود معرفی شد و به‌عنوان یک شاخص سنجش و انتخاب مدل مناسب استفاده می‌گردد. در این روابط، n تعداد داده‌ها، O_i مقادیر مشاهداتی تبخیر از تشت، P_i مقادیر تخمینی تبخیر و Kتعداد متغیر در الگوی ورودی مدل می‌باشد. مدلی بهترین نتیجه را خواهد داشت که برای آن، سه معیار NS، RMSE و MAE به ترتیب مقادیری نزدیک به یک، صفر و صفر و برای شاخص AIC کمترین مقدار را داشته باشد (قربانی و همکاران ۲۰۱۶).

نتایج و بحث

به منظور بررسی امکان استفاده از ترکیب‌های متفاوتی از داده‌های هواشناسی برای پیش‌بینی دقیق مقادیر تبخیر، ۱۰ سناریو مختلف بر اساس همبستگی بین متغیرهای هواشناسی و تبخیر از تشت، تعریف گردید (جدول ۳). سپس با استفاده از نرم‌افزار Weka، محاسبات واسنجی بر روی دو سوم از کل داده‌ها و محاسبات صحت‌سنجی هم بر روی یک سوم از داده‌های باقیمانده انجام شد. با تعریف ترکیب‌های متفاوت از متغیرهای هواشناسی موثر بر تبخیر، محاسبات مربوط به پیش‌بینی مقدار تبخیر با روش‌های SVR، GPR، M5، RF و LR با استفاده از نرم‌افزار Weka انجام پذیرفته و دقت روش‌های مذکور در جدول ٤ ارائه شده‌اند.

جدول ۳- متغیرهای دخیل در هر سناریو در روش های مورد مطالعه.

شماره سناریو	متغیرهای ورودی (مقیاس روزانه)	متغیر خروجی
۱	T_{min}	E_{pan}
۲	T_{max}	E_{pan}
۳	T_{mean}	E_{pan}
٤	T_{min}، T_{max}، T_{mean}، ssh	E_{pan}
٥	T_{min}، T_{max}، T_{mean}، RH	E_{pan}
٦	T_{min}، T_{max}، T_{mean}، U_2	E_{pan}
۷	T_{min}، T_{max}، T_{mean}، ssh، RH	E_{pan}
۸	T_{min}، T_{max}، T_{mean}، ssh، U_2	E_{pan}
۹	T_{min}، T_{max}، T_{mean}، RH، U_2	E_{pan}
۱۰	T_{min}، T_{max}، T_{mean}، ssh، RH، U_2	E_{pan}

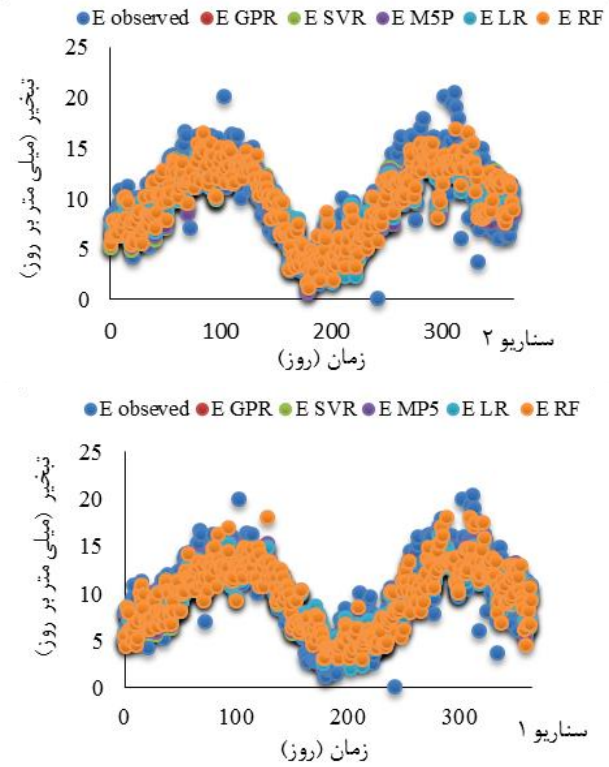
جدول ۴- نتایج سناریوهای تعریف شده برای روش‌های مورد مطالعه.

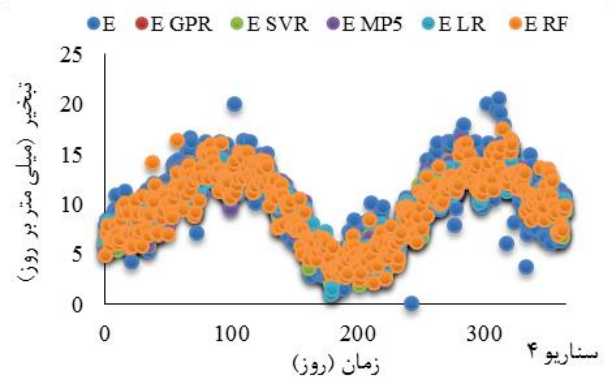
شماره سناریو	شاخص‌های آماری	مدل‌ها				
		SVR	GPR	M5	RF	LR
۱	RMSE*	۲/۲۹	۲/۲۵	۲/۲۶	۲/۳۵	۲/۳۰
	MAE	۱/۷۷	۱/۷۹	۱/۷۹	۱/۸۸	۱/۷۹
	NS	۰/۷۳	۰/۷۳	۰/۷۳	۰/۷۱	۰/۷۲
	AIC	۱۸۲۲/۶۳	۱۸۳۰/۵۳	۱۸۳۷/۷۰	۱۹۲۹/۱۸	۱۸۶۶/۱۸
۲	RMSE	۲/۲٤	۲/۲۳	۲/۲۵	۲/۰۸	۲/۳۱
	MAE	۱/۷۵	۱/۷۷	۱/۷۷	۱/٦٤	۱/۸۲
	NS	۰/۷٤	۰/۷٤	۰/۷۳	۰/۷۷	۰/۷۲
	AIC	۱۸۱۵/۷٦	۱۸۱۱/۸۲	۱۸۲٦/۸۳	۱٦٤۱/۵۸	۱۸۸۸/٦٤
۳	RMSE	۲/۱۵	۲/۱۷	۲/۱٦	۲/۳۳	۲/۲۲
	MAE	۱/٦۸	۱/۷۲	۱/۷۱	۱/۸۳	۱/۷۵
	NS	۰/۷٦	۰/۷۵	۰/۷۵	۰/۷۱	۰/۷٤
	AIC	۱۷۲۰/۱۹	۱۷٤۵/٦۲	۱۷۳۱/٤۲	۱۸۹۹/۸۷	۱۸۰٤/۸۰
٤	RMSE	۲/۰٤	۲/۰۵	۲/۰۹	۲/۲۳	۲/۱۳
	MAE	۱/۵۹	۱/٦۱	۱/٦٤	۱/۷۷	۱/٦۵
	NS	۰/۷۸	۰/۷۸	۰/۷۷	۰/۷٤	۰/۷٦
	AIC	۱٦۲۲/۹۸	٦۲۲۲/۸۷	۱٦۵۸/۸۱	۱٦۵۸/۸۱	۱٦۹٤/۵۱
۵	RMSE	۲/۰٦	۲/۰۵	۲/۰۰	۲/۲۳	۲/۱۱
	MAE	۱/٦۰	۱/٦۱	۱/٦۱	۱/۷۳	۱/٦۵
	NS	۰/۷۸	۰/۷۸	۰/۷۷	۰/۷٤	۰/۷٦
	AIC	۱٦۲۵/۱۲	۱٦۱۳/۸۹	۱٦٤٦/۰۵	۱۸۹۹/۸۷	۱٦۱۲/۲٤
٦	RMSE	۱/۹۵	۱/۹٦	۱/۹۸	۲/۰٦	۲/۰٤
	MAE	۱/٤۹	۱/۵۲	۱/۵	۱/٦۱	۱/۵۸
	NS	۰/۸۰	۰/۸۰	۰/۷۹	۰/۷۷	۰/۷۸
	AIC	۱۵۱۲/۸۲	۱۵۲۱/۹۸	۱۵٤۷/۱۰	۱٦٤۱/۱۸	۱٦۵۷/۲۸
۷	RMSE	۲/۰۲	۲/۰۱	۲/۰۳	۲/۱٤	۲/۰۸
	MAE	۱/۵۷	۱/۵۷	۱/۵۹	۱/٦۸	۱/٦۱
	NS	۰/۷۸	۰/۷۹	۰/۷۸	۰/۷٦	۰/۷۷
	AIC	۱۵۹٤/۷۷	۱۵۸۲/٦٤	۱٦۰۹/۸۹	۱۷۳۰/٦۱	۱۵۸۷/۱۷
۸	RMSE	۱/۹۳	۱/۹۳	۱/۹۵	۲/۰۱	۲/۰۱
	MAE	۱/٤۹	۱/۵۰	۱/۵۳	۱/۵٦	۱/۵۵
	NS	۰/۸۰	۰/۸۰	۰/۸۰	۰/۷۹	۰/۷۹
	AIC	۱٤۹۷/۷۹	۱٤۹۵/۷۱	۱۵۲۰/۹۳	۱۵۸٦/۳۲	۱۵۷۳/٦٦
۹	RMSE	۱/۹۵	۱/۹۳	۱/۹۵	۲/۰۰	۲/۰۰
	MAE	۱/۵۱	۱/۵۰	۱/۵۱	۱/۵٦	۱/۵۵
	NS	۰/۸۰	۰/۸۰	۰/۸۰	۰/۷۹	۰/۷۹
	AIC	۱۵۱٦/٦۹	۱٤۹٤/۸۵	۱۵۱۵/۵٦	۱۵۷۳/۲۲	۱۵٤۹/۵۷
۱۰	RMSE	۱/۹۲	۱/۹۰	۱/۹۳	۱/۹٦	۱/۹۸
	MAE	۱/۵۱	۱/٤۸	۱/۵۰	۱/۵۳	۱/۵۳
	NS	۰/۸۰	۰/۸۱	۰/۸۰	۰/۸۰	۰/۷۹
	AIC	۱٤۸۷/۷٦	۱٤٦۵/۱۲	۱٤۹٦/۰۹	۱۵۲۷/۰۳	۱۵٤۹/۵۷

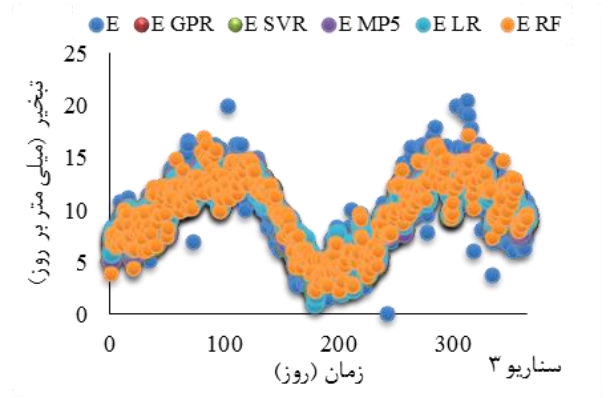
*واحد RMSE: میلی متر بر روز

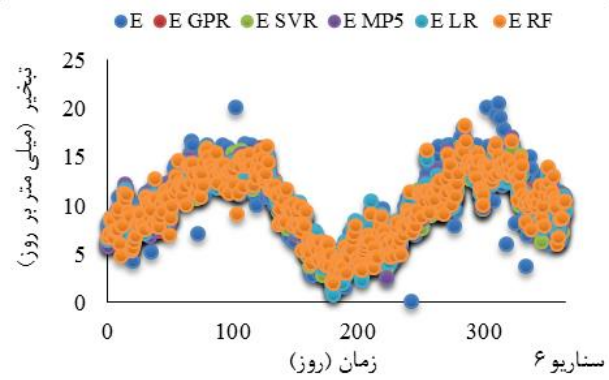
چنانچه از جدول ٤ مشخص است، مدل GPR-10 با داشتن کمترین خطا و دارا بودن پارامترهای آماری جذر میانگین مربعات خطای ۱/۹۰ میلی‌متر بر روز، میانگین خطای مطلق ۱/٤۸، ضریب نش- ساتکلیف ۰/۸۱

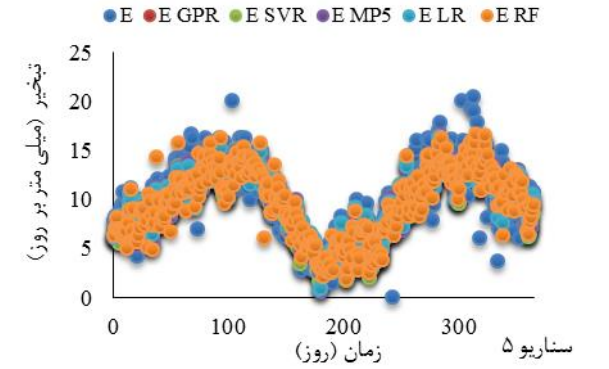
و معیار ارزیابی اطلاعات ۱٤٦٥/۱۲ بهترین عملکرد را در مقایسه با سناریوهای مختلف تعریف شده برای مدل‌های SVR، GPR، M5، RF و LR داشته است. در مرتبه دوم نیز مدل SVR-10 با جذر میانگین مربعات خطای برابر با ۱/۹۲ میلی‌متر بر روز، میانگین خطای مطلق ۱/۵۱، ضریب نش- ساتکلیف ۰/۸ و معیار ارزیابی اطلاعات مساوی با ۱٤۸۷/۷۹ قرار گرفته است. ضعیف‌ترین عملکرد مربوط به مدل RF-1 با جذر میانگین مربعات خطای ۲/۳۵ میلی‌متر بر روز، میانگین خطای مطلق ۱/۸۸، ضریب نش ساتکلیف ۰/۷۱ و معیار ارزیابی اطلاعات ۱۹۲۹/۱۸ می‌باشد. باتوجه به نتایج حاصل شده از جدول ٤، در همه سناریوهای تعریف شده الگوی ۱۰ بهترین و الگوی ۱ ضعیف‌ترین نتایج را دارا می‌باشند. همچنین، سناریوهای ٦،۸،۹،۱۰ برای روش‌های GPR و SVR و سناریوهای ۸،۹،۱۰ برای روش M5 و همچنین ۱۰ و ۹ برای روش‌های RF و LR به‌عنوان سناریوهای برتر در پیش‌بینی تبخیر از تشت کلاس A انتخاب شدند. در حالت کلی، با توجه به نتایج حاصل شده در برآورد تبخیر روزانه از تشت با استفاده از روش‌های مختلف، می‌توان چنین اظهار داشت که به‌ترتیب مدل‌های GPR، SVR، M5، LR و RF توانایی بیشتری در پیش‌بینی مقادیر تبخیر را دارا می‌باشند. همچنین نمودارهای تغییرات زمانی مقادیر تبخیر پیش‌بینی شده از تشت برای سناریوهای تعریف شده در روش‌های GPR، SVR، M5، RF و LR در شکل ۲ و نمودار های پراکنش برای سناریو برتر برای هر روش مذکور در شکل ۳ نشان داده شده‌اند.

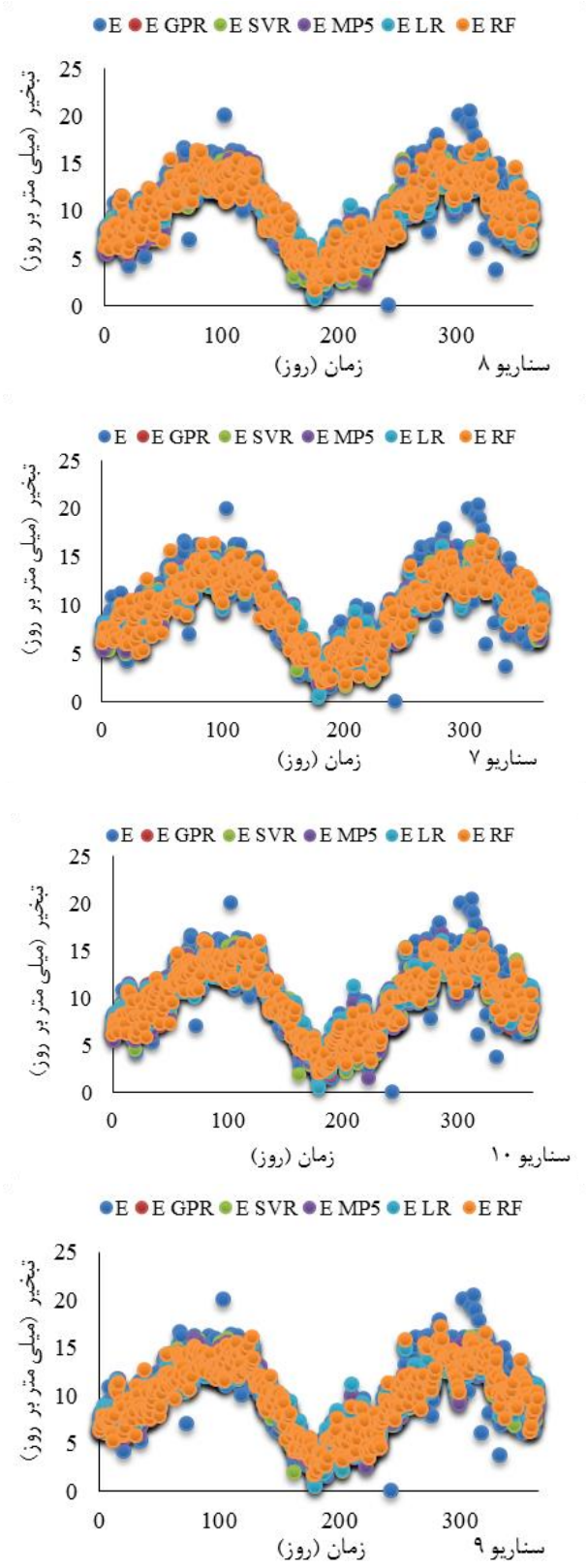



E
E GPR
E SVR
E MP5
E LR
E RF
تبخیر (میلی متر بر روز)
زمان (روز)
سناریو ۴

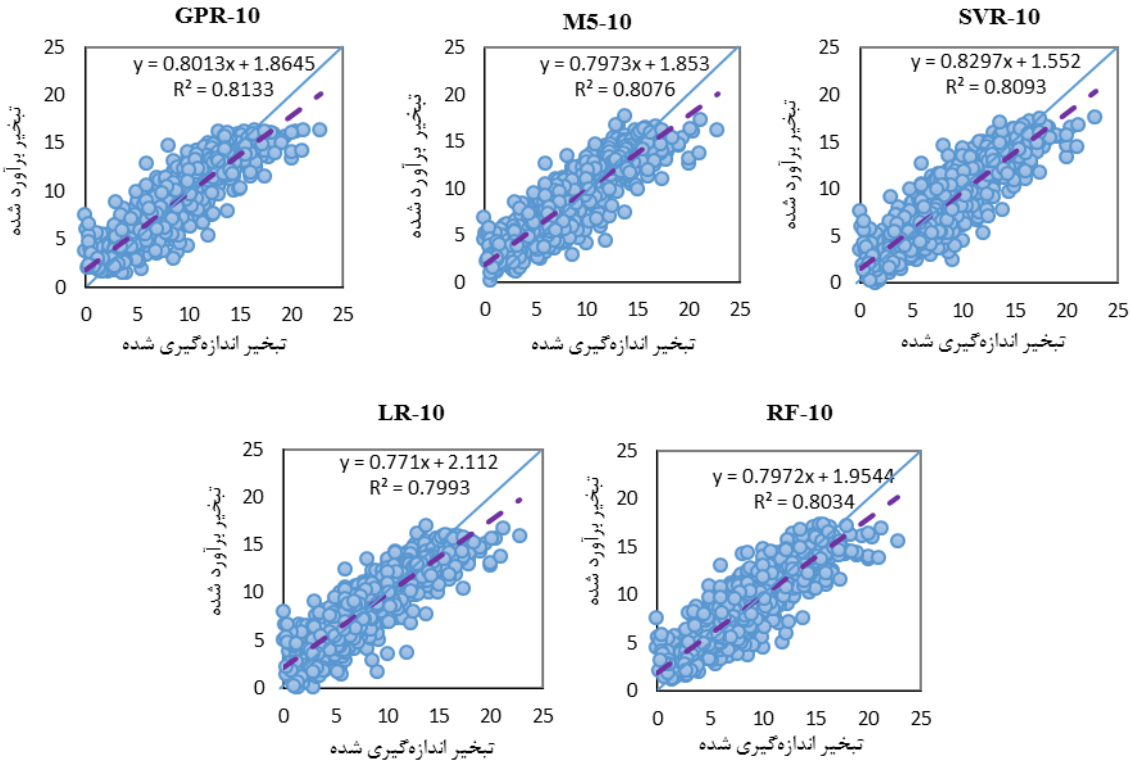

E
E GPR
E SVR
E MP5
E LR
E RF
تبخیر (میلی متر بر روز)
زمان (روز)
سناریو ۳


E
E GPR
E SVR
E MP5
E LR
E RF
تبخیر (میلی متر بر روز)
زمان (روز)
سناریو ۶


E
E GPR
E SVR
E MP5
E LR
E RF
تبخیر (میلی متر بر روز)
زمان (روز)
سناریو ۵



شکل ۲- نمودارهای تغییرات زمانی مقادیر تبخیر از تشت کلاس A با استفاده از روش‌های SVR، GPR، M5، RF و LR.



شکل ۳- نمودارهای پراکنش تبخیر از تشت کلاس A (میلی متر بر روز) با استفاده از روش‌های SVR، GPR، M5، RF و LR برای الگوی شماره ۱۰.

از شکل ۲ می‌توان چنین استنباط نمود که پیش‌بینی‌های مدل‌های GPR و SVR تطابق بیشتری با مقادیر حاصل از تشت تبخیر دارد، چرا که انطباق نقاط برآورد شده از روش‌های مذکور نسبت به روش‌های دیگر بر نقاط مشاهداتی بیشتر است. روند اشاره شده قبلی در مورد بالا بودن دقت سناریو ۱۰ برای همه روش‌ها از شکل ۳ قابل نتیجه‌گیری است و همچنین از شکل ۳ به‌وضوح مشاهده می‌گردد که پراکنش نقاط رسم شده حول خط یک به یک (خط نیم‌ساز) برای سناریو ۱۰ کمتر می‌باشد (لازم به‌ذکر می‌باشد که خط ممتد مربوط به نیمساز و خط منقطع مربوط به خط برازش می‌باشد). در جدول ۵ نیز نتایج حاصل از آنالیز هر یک از روش‌های مورد استفاده در مطالعه حاضر نشان داده شده است.

جدول ۵- آنالیز حساسیت مدل‌های SVR، GPR، M5، RF و LR.

مدل	متغیرها						شاخص‌های آماری	
	T_{min}	T_{max}	T_{mean}	S	RH	W	RMSE	MAE
SVR	■	■	■	■	■	■	۱/۹۲۷	۱/۵۱۲
	■	■	■		■	■	۱/۵۹٤	۱/۵۱٦
	■	■	■	■		■	۱/۹۳۸	۱/٤۹۹
	■	■	■	■	■		۲/۰۲۳	۱/۵۷۵
GPR	■	■	■	■	■	■	۱/۹۰۸	۱/٤۸۵
	■	■	■		■	■	۱/۹۳۵	۱/۵۰۹
	■	■	■	■		■	۱/۹۳۵	۱/۵۰۹
	■	■	■	■	■		۲/۰۱۱	۱/۵۷۲

M5	۱/۹۲٤	۱/٥۱۲	■	■	■		■	■		
	۱/۹٥۰	۱/٥۱۳	■	■	■	■	■			
	۱/۹۲۷	۱/٥۱۳	■	■	■	■		■		
	۱/۹۳٤	۱/٥۰۸	■	■	■	■	■	■		
	۱/۹٥۳	۱/٥۱۰	■	■		■	■	■		
	۱/۹٥۷	۱/٥۳٦	■		■	■	■	■		
	۱/۰۲۳	۱/٥۷٥		■	■	■	■	■		
	۱/۹۳۹	۱/٥۱۳	■	■	■		■	■		
	۱/۹٤٦	۱/٥۱٦	■	■	■	■	■			
	۱/۸۳٤	۱/٥۱۲	■	■	■	■		■		
	۱/۹۸۰	۱/٥۳۲	■	■	■	■	■	■		
	۲/۰۰٤	۱/٥٥۸	■	■		■	■	■		
	۲/۰۱٥	۱/٥٥۷	■		■	■	■	■		
LR	۲/۰۸۰	۱/٦۱۷		■	■	■	■	■		
	۱/۹۹۱	۱/٥٤٤	■	■	■		■	■		
	۲/۰۰۸	۱/٥٥٥	■	■	■	■	■			
	۱/۹۸٤	۱/٥۳٦	■	■	■	■		■		
RF	۱/۹۱۳	۱/٤۹۰	■	■	■		■	■		
	۱/۹۲۹	۱/٤۹٥	■	■	■	■	■			
	۱/۹۰٦	۱/٤۸٤	■	■	■	■		■		
	۱/۹٦۱	۱/٥۳٦	■	■	■	■	■	■		
	۲/۰۰۳	۱/٥٦۰	■	■		■	■	■		
	۲/۰۱٥	۱/٥٦٦	■		■	■	■	■		
	۲/۱٤۹	۱/٦۸۷		■	■	■	■	■		
	۱/۹۸۲	۱/٥٥٥	■	■	■		■	■		
	۲/۰۱۰	۱/٥٤٤	■	■	■	■	■			
	۱/۹۸۰	۱/٥۲۸	■	■	■	■		■		

■متغیر مورد استفاده در تحلیل حساسیت

براساس نتایج جدول ۵ متغیرهای دما (متوسط و حداکثر) و تابش خورشیدی بیشترین و سرعت باد و رطوبت نسبی کمترین ضریب تاثیر را با تبخیر از تشت دارند. آنچه مسلم است متغیرهای هواشناسی دما و تابش خورشیدی در فرآیند تبخیر، از تأثیرگذاری زیادی در منطقه مورد مطالعه برخوردار می‌باشند. با توجه به نتایج تحلیل حساسیت، مدل M5 و SVR کمترین حساسیت را نسبت به تغییرات دمای متوسط و بیشترین حساسیت را به تغییرات سرعت باد نسبت به سایر متغیرها دارا می‌باشند. در این راستا مدل GPR کمترین حساسیت را نسبت به تغییرات دما از خود نشان داد. اثرگذاری متغیرهای ورودی بر میزان تبخیر، از بیشترین به کمترین آنها به‌ترتیب شامل: سرعت باد، تابش خورشیدی، دما و رطوت نسبی می‌باشند. نتایج این مطالعه با نتایج حاصل از تحقیق طبری و همکاران (۲۰۱۰) و شادمانی و معروفی (۲۰۱۱) که درآن بیشترین تاثیر مربوط به متغیر سرعت باد و با نتایج تحقیق سیاح فرد و همکاران (۲۰۱۰) که در آن بیشترین حساسیت مربوط به متغیر ورودی تابش خورشیدی است، مطابقت دارد.

نتیجه‌گیری کلی

تبخیر یکی از فرآیندهای مهم و موثری است که سبب از بین رفتن بیش از نیمی از نزولات جوی در مناطق خشک و نیمه خشک می‌شود. تبخیر از مخازن آب با توجه به اقلیم خشک و نیمه خشک ایران و محدودیت منابع آب، تلفات آب را درپی خواهد داشت. یکی از مسائل مهم در بحث تبخیر، ثبت مقدار تبخیر است و این کار با استفاده از تشت تبخیر صورت می‌گیرد. و با توجه به میزان بالای تبخیر در ایران و نبود ابزار اندازه‌گیری مناسب می‌توان مقادیر تبخیر از تشت را با متغیرهای هواشناسی و مدل مناسب با دقت خوبی برآورد نمود. در این پژوهش پنج مدل از روش‌های داده‌کاوی برای برآورد تبخیر از تشت کلاس

A ارزیابی گردید و بر اساس نتایج به دست آمده، متغیرهای تابش خورشیدی و سرعت باد به‌عنوان متغیرهای تاثیرگذار بر میزان تخمین تبخیر معرفی گردیدند. همچنین استفاده از ترکیبی با شش متغیر ورودی، حالت ایده‌آل مدل‌های کاربردی پژوهش حاضر بود. براساس نتایج حاصل از شاخص‌های AIC و NS، به‌ترتیب مدل‌های SVR، GPR و M5 کاربردی‌ترین مدل‌ها در تخمین تبخیر برای منطقه مورد مطالعه معرفی شدند.

برای مطالعات بعدی در زمینه تخمین مقدار تبخیر پیشنهاد می‌گردد نتایج حاصل از معادلات تجربی و روش های مورد استفاده در مطالعه حاضر برای اقلیم هایی با تنوع آب و هوایی مختلف بررسی گردد.

منابع مورد استفاده

Afkhami H, Ekhtesasi MR and Habibipour A, 2018. Evaluate the efficiency of several data mining methods to predict evaporation (Case study: Yazd Synoptic Station). Journal of Range and Watershed Management 71(3): 579- 594 (In Persian with English abstract).

Alsumaiei AA, 2020. Utility of artificial neural networks in modeling pan evaporation in hyper-arid climates. Water 2020, 12 (5), 1508 https://doi.org/10.3390/w12051508

Basak D, Pal S and Patranabis DC, 2007. Support vector regression. Neural Information Processing 11: 203-225.

Breiman L, 2001. Random forests. Machine Learning 45 (1): 5–32.

Dehghani AA, Piri M, Hesam M and Dehghani N, 2010. Estimation of daily evaporation from evaporation pan using three multilayer perceptron neural networks, radial and Germanic base function. Water and Soil Conservation Research (Agricultural Sciences and Natural Resources) 17(2): 49- 67 (In Persian with English abstract).

Ghorbani MA, Ahmadzadeh H, Isazadeh M and Terzi O, 2016. Comparative study of artificial neural network and support vector machine models for river flow prediction. Environmental Earth Science 75: 476-490.

Goyal MKB, Bharti J, Quilty J, Adamowski and Pandey A, 2014. Modeling of daily pan evaporation in sub-tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert System with Application 41: 5267–5276.

Irmak S, Haman DZ and Jones JW, 2002. Evaluation of class A pan coefficients for estimating reference evapotranspiration in humid location. Journal of Irrigation and Drainage Engineering, 128(3): 153-159 (In Persian with English abstract).

Joudy AR and Sattari MT, 2015. Estimation of scour depth of piers in hydraulic structures using Gaussian process regression. Irrigation and Drainage Structures Engineering Research 16(65): 19-36 (In Persian with English abstract).

Joudy AR and Sattari MT, 2016. Evaluation of the performance of kernel-based methods in estimating the suspended sediment load of the river (Case study: Sufi Chai Maragheh river). Natural Geography Research 48(3): 413- 429 (In Persian with English abstract).

Kalantari KH, 2006. Data Processing and Analysis in Socio-Economic Research, Sharif Publishing (In Persian).

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.

Kisi O, 2006. Daily pan evaporation modeling using a neuro-fuzzy computing technique. Journal of Hydrology 329: 636– 646.

Malik A, Tikhamarine Y, Al-Ansarid N, Shahid Sh, Singh Sekhon H, Kumar Pal R, Rai P, Pandey K, Singh P, Elbeltagi A and Shauket Sammen S, 2020. Daily pan-evaporation estimation in different agro-

climatic zones using novel hybrid support vector regression optimized by Salp swarm algorithm in conjunction with gamma test. Engineering Applications of Computational Fluid Mechanics 15(1): 1075-1094.

Mohammadi B, 2018. Estimation of daily evaporation based on artificial neural network hybrid model and firefly optimization algorithm, Iran. MSc Thesis, University of Tabriz (In Persian with English abstract).

Qasem SN, Samadianfard S, Kheshtgar S, Jarhan S, Kisi O, Shamshirband S and Chau KW, 2019. Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. Engineering Applications of Computational Fluid Mechanics 13(1): 177-187.

Saghebian M and Roushangar K 2019. Prediction of total and bedform roughness coefficient in alluvial channels based on experimental data via Gaussian process regression method. Iranian Journal of Irrigation and Drainage 2: 437-499 (In Persian with English abstract).

Samadianfard S, Hashemi S and Ezadyar M, 2018. Estimation of daily evaporation from evaporation pan using machine learning methods. Iranian Journal of Irrigation and Drainage 4(12): 1015- 1042 (In Persian with English abstract).

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 18: 157-167(In Persian with English abstract).

Sattari MT, Nahrin F and Azimi V, 2013. Prediction of daily reference evapotranspiration using artificial neural network model and M5 tree model (Case study: Bonab Station). Iranian Journal of Irrigation and Drainage 1(7): 104-113 (In Persian with English abstract).

Sayah M, 2010. Sensitivity analysis of different evaporation estimation methods using Artificial Neural Networks, MSc Thesis, University of Tabriz, Iran (In Persian with English abstract).

Shadmani M and Maroufi S, 2011. Comparison of several methods for estimating daily evaporation from the pan - a case study of Kerman region. Agricultural Technologies and Natural Resources, Water and Soil Sciences 15(55): 69- 84 (In Persian with English abstract).

Tabari H, Maroofi S and Sabziparvar AA, 2010. Estimation of daily pan evaporation using artificial neural network and multivariate non- liner regression. Irrigation Science 28: 399-406.

Terzi O, 2007. Data mining approach for estimation evaporation from free water surface. Journal of Applied Sciences, 7 (4): 593-596.

Vapnik VN, 1995. The Nature of Statistical Learning Theory. Springer, New York. 314P.

Vicente Serrano SM, Bidegain M, Tomas Burguera M, Dominguez-Castro F, El Kenawy A, McVicar T, Azorin-Molina C, López-Moreno J, Nieto R, Gimeno L and Giménez A, 2018. A comparison of temporal variability of observed and model-based pan evaporation over Uruguay (1973–2014). International Journal of Climatology 38 (1): 337-350.

Xianghui L, Yan J, Lifeng W, Junliang F, Fucang Zh and Zhijun L, 2018. Daily pan evaporation modeling from local data using three tree-based machine learning models. Journal of Hydrology 566: 668–684.

Zabihi M, Pourghasemi H and Behzadfar M, 2015. Groundwater potential mapping using Shannon entropy and stochastic forest models in Bojnourd plain. Echo Hydrology 2(2): 221-232 (In Persian with English abstract).