

## Research Article

# Estimating Reference Evapotranspiration in Three Arid, Semi-Arid and Humid Climates Using Gradient Boosted Tree, Generalized Linear Model and Random Forest

M Izadyar<sup>1</sup>, S Hashemi<sup>2</sup>, S Samadianfard<sup>\*3</sup>

Received: August 17, 2019      Received in Revised: February 8, 2022

Accepted: April 6, 2021      Published online: September 23, 2023

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

2-PhD. Student, Dept. of Water Sci. Eng., Faculty of Agric., University of Tabriz, Iran

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

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

## Abstract

### Background and Objectives

Evapotranspiration is one of the main components of water balance in agriculture and is one of the effective and efficient factors for accurate irrigation planning and management. Direct measurement of evapotranspiration values is time consuming and costly. On the other hand, modeling such a complex process in which many parameters interact with each other is so difficult that it is not possible to simplify the issue without multiple assumptions. Therefore, accurate estimation of this parameter has always been considered by the researchers. In the other point of view, the FAO-56 method was used as the accurate and accepted method for calculating reference evapotranspiration. One of the weaknesses of this model is its dependence on various meteorological variables. Therefore, it is necessary to use methods which need low number of meteorological variables and estimate the reference evapotranspiration with high accuracy. Additionally, due to the use of many meteorological variables and the complexity of the calculations, it is difficult to use FAO-56 method in all regions. Therefore, in the recent years, many researchers implemented machine learning methods to estimate reference evapotranspiration. Most studies in the field of reference evapotranspiration estimation use experimental models that require all the effective reference evapotranspiration parameters to provide an acceptable estimate. Hence, the aim of the current study was to present a superior model from three machine learning models, including random forest (RF), gradient boosted tree (GBT) and generalized linear model (GLM) for estimating reference evapotranspiration in three synoptic stations located at arid, semi-arid and wet climates of Iran. To the best of our knowledge, the proposed GBT and GLM methods have not been used for estimating reference evapotranspiration in the mentioned stations.

### Methodology

In this research, the FAO-56 method was used to estimate the reference evapotranspiration. Also, three machine learning methods including GBT, GLM and RF were implemented to estimate the amount of reference evapotranspiration. Daily parameters of some fundamental and effective meteorological variables on evapotranspiration during 21-years statistical period (2000-2020) were collected in three stations located at different climates including Yazd station (arid), Birjand station (semi-arid) and Sari station (wet). In order to investigate the possibility of using different



combinations of meteorological parameters to estimate the reference evapotranspiration as accurately as possible, seven different combinations of meteorological parameters were defined. The accuracy of the utilized methods was evaluated using three criteria such as correlation coefficient, scattering index and Nash-Sutcliffe coefficient. Additionally, Taylor diagrams were implemented for evaluating the accuracy of the used methods. It should be noted that the Taylor diagram shows the three parameters of root mean square error, correlation coefficient and standard deviation simultaneously in one figure. Also, the most suitable combination of meteorological parameters that had good accuracy for estimating reference evapotranspiration, was suggested.

### **Findings**

The results showed that in the best model at Birjand Station, and Yazd stations scenario number three by two meteorological variables of temperature and wind speed and in Sari station the scenario number two with temperature and relative humidity, the gradient boosted tree model was reinforced with Nash-Sutcliffe coefficient of 0.804, 0.826 and 0.733, with correlation coefficient of 0.997, 0.997 and 0.919 and scatter index of 0.249, 0.218 and 0.361 and the generalized linear model with Nash-Sutcliffe coefficient of 0.892, 0.931 and 0.869 correlation coefficient of 0.952, 0.966 and 0.933 and scatter index of 0.185, 0.137 and 0.252, respectively. Finally, the RF method with Nash-Sutcliffe coefficient of 0.954, 0.956 and 0.929, correlation coefficient of 0.978, 0.978 and 0.965 and scatter index of 0.121, 0.110 and 0.186 had good performance for estimating the reference evapotranspiration. On the other hand, in all methods, the scenario number seven using the meteorological parameters of temperature, relative humidity of sunny hours and wind speed in all three stations, presented the most accurate performance. Therefore, all three methods may be proposed as models with high degree of accuracy for estimating reference evapotranspiration.

### **Conclusion**

Reference evapotranspiration is one of the main components of water balance in agriculture and is one of the effective and influential factors for accurate irrigation planning. Therefore, accurate estimation of this parameter has a significant role on reducing excessive water consumption. In this study, three data-driven models of RF, GBT and GLM were used in three stations of Yazd, Birjand and Sari stations. The obtained results indicated that the seventh scenario using all four meteorological parameters in all stations with the highest correlation coefficient, the lowest scatter index and the highest Nash-Sutcliffe coefficient provided most accurate estimates of the reference evapotranspiration and may be recommended for proper estimation of reference evapotranspiration.

**Keywords:** Arid and semi-arid climates, Irrigation, Random Forest, Reference evapotranspiration.

## مقاله پژوهشی

# برآورد تبخیر-تعرق مرجع در سه اقلیم خشک، نیمه خشک و مرطوب با استفاده از روش‌های درخت گرادیان تقویت شده، مدل خطی تعمیم یافته و جنگل تصادفی

مجتبی ایزدیار<sup>۱</sup>، سجاد هاشمی<sup>۲</sup>، سعید صمدیان فرد<sup>۳\*</sup>

تاریخ دریافت: ۱۴۰۰/۰۸/۲۷

تاریخ پذیرش: ۱۴۰۰/۱۲/۰۳

تاریخ دریافت ویرایش: ۱۴۰۰/۱۱/۱۹

تاریخ انتشار آنلاین: ۱۴۰۲/۰۷/۰۱

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

۲- دانشجوی دکتری، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

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

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

## چکیده

تبخیر-تعرق یکی از مولفه‌های اصلی بیلان آب در کشاورزی و از جمله عوامل موثر و تاثیرگذار جهت برنامه‌ریزی دقیق آبیاری است. لذا برآورد دقیق این پارامتر همواره مورد توجه پژوهشگران بوده است. در این راستا و در پژوهش حاضر، توانایی سه روش درخت گرادیان تقویت شده، مدل خطی تعمیم یافته و جنگل تصادفی در برآورد مقدار تبخیر-تعرق گیاه مرجع در سه اقلیم خشک (ایستگاه یزد)، نیمه خشک (ایستگاه بیرجند) و مرطوب (ایستگاه ساری) در بازه زمانی بیست و یک ساله (سال ۲۰۰۰ تا ۲۰۲۰) مورد بررسی قرار گرفت. دقت روش‌های مذکور با استفاده از سه معیار ارزیابی ضریب همبستگی، شاخص پراکندگی داده‌ها و ضریب نش-ساتکلیف مورد بررسی قرار گرفت. نتایج حاصل نشان دادند که در بهینه‌ترین حالت به ترتیب در ایستگاه‌های بیرجند، یزد و ساری مدل گرادیان تقویت شده با مقدار ضریب نش-ساتکلیف ۰/۸۰۴، ۰/۸۲۶ و ۰/۷۳۳، مدل خطی تعمیم یافته با ضرایب ۰/۸۹۲، ۰/۹۳۱ و ۰/۸۶۹ در نهایت روش جنگل تصادفی با ضرایب ۰/۹۵۴، ۰/۹۵۶ و ۰/۹۲۹ عملکرد مناسبی را در برآورد میزان تبخیر-تعرق مرجع داشتند. از طرفی در تمامی روش‌ها ترکیب داده هفتم با استفاده از پارامترهای هواشناسی دما، رطوبت نسبی، ساعات آفتابی و سرعت باد در هر سه ایستگاه مورد پژوهش بهترین عملکرد را ارائه نمود؛ اما در ایستگاه بیرجند و ساری روش درخت گرادیان تقویت شده و در ایستگاه یزد مدل خطی تعمیم یافته نتایج بهتری را نسبت به دیگر مدل‌ها ارائه کردند و می‌توانند در ایستگاه‌های مورد پژوهش به عنوان روشی با دقت بالا در برآورد تبخیر-تعرق مرجع پیشنهاد گردند.

واژه‌های کلیدی: آبیاری، اقلیم خشک، اقلیم نیمه خشک، تبخیر-تعرق مرجع، جنگل تصادفی

## مقدمه

استفاده بهینه از منابع آب در بخش کشاورزی به عنوان بزرگترین بخش مصرف کننده آب، امری حیاتی می باشد. تبخیر-تعرق یکی از مولفه های اصلی بیلان آب در زمینه کشاورزی و از جمله عوامل موثر و تاثیرگذار جهت برنامه ریزی دقیق آبیاری است. لذا برآورد دقیق نیاز آبی انواع گیاهان تاثیر شایانی بر کاهش مصرف بی رویه آب به ویژه در مناطق خشک و نیمه خشک دارد. برای محاسبه و تخمین تبخیر-تعرق مرجع روش های گوناگونی همچون روش پنمن-مانتیث-فائو که مورد پذیرش سازمان خوار و بار جهانی (FAO) می باشد، استفاده نمود. از طرفی تغییرات زمانی-مکانی عناصر آب و هوایی موثر در تبخیر-تعرق به همراه تغییرات آب در گیاه، شرایط پیچیده را برای تخمین این پارامتر بوسیله روش پنمن-مانتیث-فائو و روش های تجربی بوجود آورده است (رحیمی خوب و محمودی ۲۰۱۱). از این رو پژوهشگران سعی بر آن داشته اند که با توجه به پیچیدگی روابط حاکم بر تغییرات پارامترهای موثر بر مقدار تبخیر-تعرق، از روش هایی همچون هوش مصنوعی و یا روش های مبتنی بر یادگیری ماشینی، به عنوان یک راه حل مکمل، جهت شناسایی ارتباط بین عوامل مذکور استفاده نمایند. رحیمی خوب و همکاران (۲۰۱۱) تبخیر-تعرق مرجع را به وسیله چهار سناریو از مدل شبکه عصبی مصنوعی با استفاده از ترکیب متغیرهای هواشناسی محاسبه نمودند. نتایج حاصل نشان داد که متغیرهای دمای بیشینه و کمینه هوا، رطوبت نسبی و سرعت باد؛ حداقل داده های هواشناسی مورد نیاز برای برآورد تبخیر-تعرق هستند. میانگین جذر مربعات خطاها (RMSE) و ضریب تعیین ( $R^2$ ) بین مقادیر بدست آمده از مدل شبکه عصبی با ورودی های فوق به ترتیب ۰/۱۷ میلیمتر در روز و ۰/۹۵ تعیین شدند. نتایج تحقیقات نورانی و سیاح (۲۰۱۳) در دو شهر تبریز و ارومیه که از

مدل های شبکه عصبی مصنوعی، رگرسیون خطی چندمتغیره و سه روش تجربی تعادل انرژی، آیرودینامیک و ترکیبی پنمن برای برآورد تبخیر روزانه استفاده کرده بودند، بیانگر برتری مدل شبکه عصبی نسبت به سایر روش های مذکور بوده است. سپهری و همکاران (۲۰۲۱) در پژوهشی تبخیر-تعرق مرجع را با استفاده از داده های هواشناسی ۱۰ ساله استان تهران و با نرم افزار ET-Calculator محاسبه کردند. ایشان نتایج بدست آمده را با خروجی مدل شبکه عصبی مصنوعی مقایسه نمودند و بیان داشتند که مدل شبکه عصبی مصنوعی با ضریب همبستگی بالا ( $R^2=98\%$ ) تکنیک بسیار مناسبی برای محاسبه تبخیر-تعرق مرجع می باشد. کریمی و همکاران (۲۰۱۳) تبخیر-تعرق مرجع را در دو ایستگاه هواشناسی کشور اسپانیا (سالواتیرا و زامبران) با استفاده از مدل های هوش مصنوعی و معادلات تجربی، با متغیرهای ورودی دمای هوا، تشعشع خورشیدی، سرعت باد و رطوبت نسبی محاسبه کردند. ایشان نتایج حاصل را با خروجی رابطه پنمن-مانتیث-فائو مقایسه نموده و بیان داشتند که مدل های هوش مصنوعی از دقت بالایی نسبت به معادلات تجربی برخوردار هستند. محمدرضاپور (۲۰۱۶) در مطالعه ای تبخیر-تعرق پتانسیل ماهانه را در استان سیستان و بلوچستان محاسبه کردند. ایشان بیان داشتند که میزان تبخیر-تعرق در این منطقه نسبت به متغیرهای دمای متوسط، رطوبت نسبی و سرعت باد بیشترین حساسیت را دارا می باشد. صمدیان فرد و پناهی (۲۰۱۷) روش های رگرسیون بردار پشتیبان و درخت M5 را با معادلات تجربی برآورد تبخیر-تعرق مرجع را در ایستگاه سینوپتیک تبریز مورد بررسی قرار دادند و نتایج حاصله را با مقادیر معادله استاندارد پنمن-مانتیث-فائو مقایسه کردند. نتایج نشان داد که مدل های هوش مصنوعی با در نظر گرفتن تمامی متغیرهای هواشناسی با دارا بودن مقادیر خطای کمتر (جذر میانگین مربعات خطای ۰/۴)

## مواد و روش‌ها

### منطقه مورد مطالعه

ایستگاه‌های مورد پژوهش در مطالعه حاضر از بین ایستگاه‌های هواشناسی سه استان یزد، خراسان شمالی و مازندران انتخاب گردید. در استان‌های مذکور ایستگاه‌های یزد، بیرجند و ساری بررسی شدند. همچنین براساس بررسی انجام شده به وسیله ضریب اقلیم نمای دومارتن ایستگاه‌های یزد، بیرجند و ساری به ترتیب دارای سه اقلیم خشک، نیمه خشک و مرطوب هستند. شهرستان یزد (مرکز استان یزد) واقع در فلات مرکزی ایران با اقلیم خشک اولین ایستگاه انتخابی پژوهش حاضر می‌باشد. شهرستان بیرجند (مرکز استان خراسان شمالی) با اقلیمی نسبتاً مرطوب از لحاظ میزان بارش رتبه دوم را بین ایستگاه‌های مورد مطالعه دارد و در نهایت شهرستان ساری (مرکز استان مازندران) سومین ایستگاه انتخابی در پژوهش حاضر با اقلیم مرطوب می‌باشد. موقعیت جغرافیایی ایستگاه‌ها در جدول ۱ بیان شده است. همچنین شکل ۱ موقعیت سه ایستگاه را بر روی نقشه نشان می‌دهد. در پژوهش حاضر، از متغیرهای هواشناسی دمای هوا (درجه سانتی‌گراد)، رطوبت نسبی (درصد)، سرعت باد (متر بر ثانیه) و ساعات آفتابی ایستگاه‌های مذکور در بازه زمانی بیست و یک ساله (۲۰۲۰-۲۰۰۰) استفاده شد.

### معادله پنمن مانتیث فائو

سازمان خوار و بار جهانی در سال ۱۹۹۸ در نشریه ۵۶ روشی تکامل یافته از رابطه پنمن مانتیث را برای محاسبه تبخیر-تعرق گیاه از سطح مرجع با استفاده از داده‌های هواشناسی معرفی کرد. بر اساس این نشریه در هنگام عدم دسترسی به لایسیمتر، با استفاده از داده‌های هواشناسی مقدار تبخیر-تعرق مرجع را محاسبه و

نسبت به معادلات تجربی، عملکرد بهتری در تخمین تبخیر-تعرق گیاه دارا هستند. سیلوا جونیور و همکاران (۲۰۱۹) در پژوهشی با استفاده از مدل‌های جنگل تصادفی (RF) و جنگل تصادفی مکانی (RF<sub>SP</sub>) میزان تبخیر-تعرق را در منطقه شمال شرقی کشور برزیل مورد مطالعه قرار داده و بیان داشتند که مدل جنگل تصادفی از دقت بالایی نسبت به مدل جنگل تصادفی مکانی برخوردار است و برای برآورد تبخیر-تعرق مرجع در این منطقه مناسب‌تر می‌باشد. کریمی و همکاران (۲۰۲۰) تاثیر داده‌های ورودی از دست رفته در عملکرد معادله پنمن-مانتیث-فائو را در شش منطقه مرطوب ساحلی ایران طی ده سال را مورد بررسی قرار دادند. ایشان همچنین از روش‌های یادگیری ماشینی بوت استرپ مبتنی بر جنگل تصادفی (RF<sup>1</sup>) و چند روش تجربی (هارگریوز-سامانی، پرستی تیلو، مک کینگ و تورک) برای محاسبه تبخیر-تعرق مرجع استفاده کردند و بیان کردند که در صورت جایگزینی ورودی‌های گمشده با ورودی‌های محاسبه‌شده، مدل‌های مبتنی بر ترکیب دقیق‌ترین نتایج را ارائه کردند.

اغلب مطالعات انجام گرفته در زمینه برآورد تبخیر-تعرق مرجع از معادلات تجربی استفاده شده و یا مدل‌های پیشنهادی توسط پژوهشگران نیازمند تمامی متغیرهای موثر تبخیر-تعرق مرجع می‌باشند تا تخمین قابل قبولی ارائه دهند. هدف از پژوهش حاضر، بررسی کارایی سه مدل یادگیری ماشینی، جنگل تصادفی (RF)، درخت گرادیان تقویت شده (GBT<sup>2</sup>) و مدل خطی تعمیم یافته (GLM<sup>3</sup>) در برآورد دقیق تبخیر-تعرق مرجع در سه ایستگاه سینوپتیک در اقلیم های خشک، نیمه خشک و مرطوب ایران می‌باشد، بطوری‌که مدل پیشنهادی نیازمند متغیرهای ورودی کمتری باشد تا در هنگام نبود داده کافی نیز استفاده از این روش امکان‌پذیر باشد.

<sup>3</sup> Generalized Linear Model

<sup>1</sup> Random Forest

<sup>2</sup> Gradient Boosted Tree

به عنوان روشی استاندارد برای ارزیابی روابط تجربی استفاده کرد. رابطه ارائه شده به شرح ذیل می باشد (آلن و همکاران ۱۹۹۸):

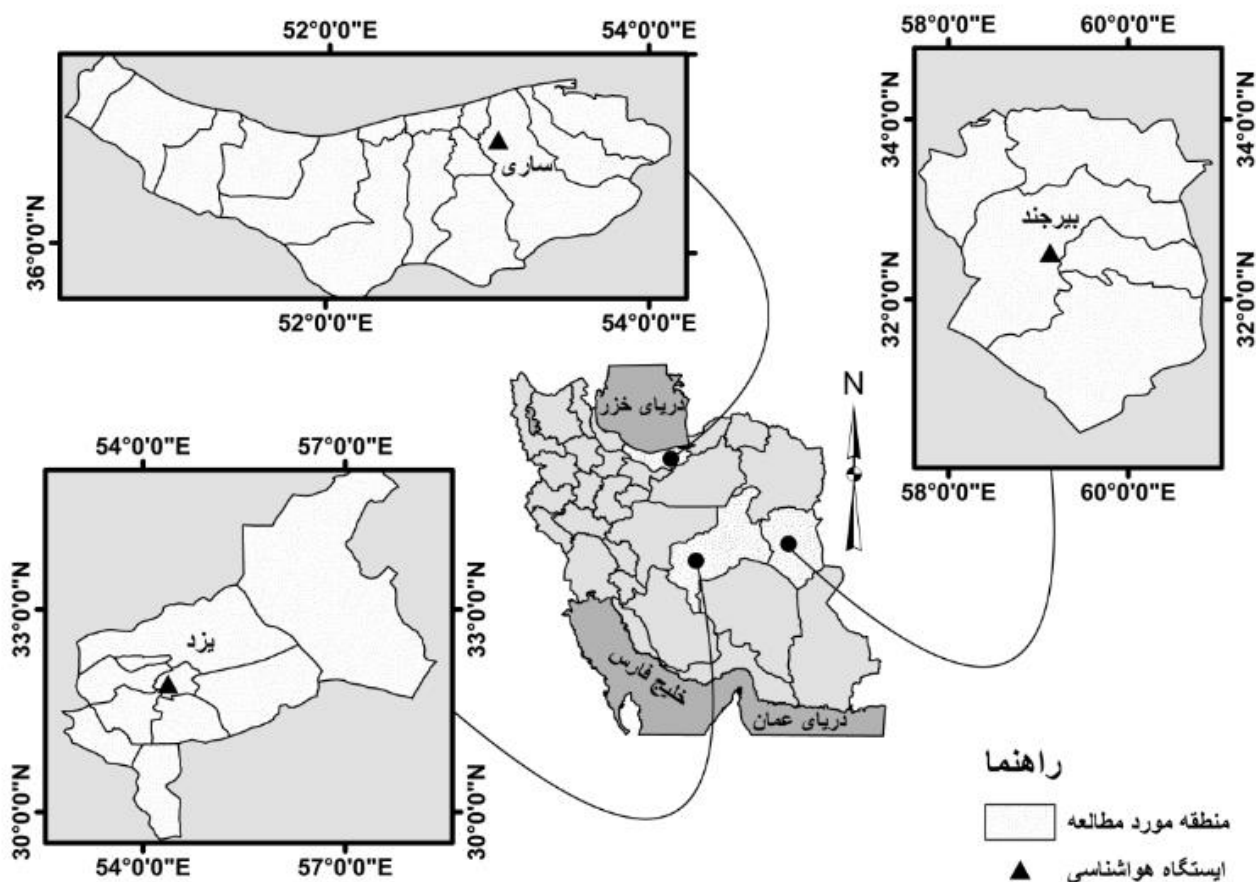
$$ET_0 = \frac{0.408\Delta(Rn - G) + \gamma \left[ \frac{900}{T + 273} \right] U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad [۱]$$

در رابطه فوق،  $ET_0$  تبخیر-تعرق گیاه مرجع (میلیمتر بر روز)،  $Rn$  تابش خالص در سطح پوشش گیاهی (مگاژول بر مترمربع بر روز)،  $G$  شار گرمایی به داخل

خاک (مگاژول بر مترمربع بر روز)،  $T$  متوسط دمای هوا در ارتفاع دو متری از سطح زمین (درجه سانتیگراد)،  $U_2$  متوسط سرعت باد در ارتفاع دو متری از سطح زمین (متر بر ثانیه)،  $e_s$  فشار بخار اشباع (کیلو پاسکال)،  $e_a$  فشار بخار واقعی (کیلو پاسکال)،  $e_s - e_a$  کمبود فشار بخار اشباع (کیلو پاسکال)،  $\Delta$  شیب منحنی فشار بخار اشباع (کیلو پاسکال بر درجه سانتیگراد) و  $\gamma$  ضریب رطوبتی (کیلو پاسکال بر درجه سانتیگراد) می باشند.

جدول ۱- مشخصات ایستگاه‌های مورد مطالعه.

نام ایستگاه	طول ایستگاه	عرض	ارتفاع (متر از سطح دریا)
یزد	۵۴ درجه و ۲۱ دقیقه	۳۱ درجه و ۵۳ دقیقه	۱۲۳۰
بیرجند	۵۹ درجه و ۱۲ دقیقه	۳۲ درجه و ۵۲ دقیقه	۱۴۷۰
ساری	۵۳ درجه و ۵ دقیقه	۳۶ درجه و ۴ دقیقه	۵۴



شکل ۱- موقعیت جغرافیایی ایستگاه‌های مورد پژوهش.

و ایستگاه‌های هواشناسی امکان برآورد و یا ثبت این داده‌ها وجود ندارد؛ بنابراین محققان در

هنگام پیش‌بینی موارد پیچیده، نتایج حاصله از ادغام تصمیم چندین درخت، برتر از نتیجه حاصله از یک درخت خاص خواهد بود. در این روش درخت‌های تصمیم می‌توانند با استفاده از ویژگی داده‌های ارائه شده مقادیر هدف را پیش‌بینی کنند، به طوری که در هر مرحله هر درخت زیرمجموعه برای بار اول با استفاده از داده‌هایی که به اشتباه توسط درخت پیشین پیش‌بینی شده‌اند، آموزش داده می‌شوند. همین امر موجب تمرکز بیشتر مدل بر روی موارد پیچیده شده و بنابراین دقت روش افزایش پیدا می‌کند. درختان گرادیان تصمیم گیرنده، بر خلاف مدل‌های خطی مانند رگرسیون لجستیک، تغییرات غیرخطی و متقاطع را در ویژگی‌های ورودی بیشتر می‌توانند اعمال کنند. به ویژه به دلیل پیکربندی سلسله

روش پنمن مانیتیت فائو (PFM) علی رغم دقیق و استاندارد بودن، به داده ورودی زیادی نیاز دارد که در اکثر مناطق سال‌های اخیر در تلاش هستند تا با استفاده از روش‌های هوش مصنوعی روشی با حداقل میزان خطا را جایگزین کرده و در هنگام نبود داده هواشناسی کافی از این روش‌ها استفاده شود. از جمله این روش‌ها مدل‌های جنگل تصادفی، مدل خطی تصمیم یافته و مدل درخت گرادیان تقویت شده می‌باشد که در این پژوهش مورد بررسی قرار گرفته است.

#### مدل درخت گرادیان تقویت شده (GBT)

روش درخت گرادیان تقویت شده یکی از بهترین الگوریتم‌های یادگیری ماشینی است و در بسیاری از مجموعه داده‌ها کاربرد داشته و دقت قابل قبولی را ارائه کرده است. الگوریتم افزایش بر این اساس می‌باشد که

روزهای خشک و مرطوب را با استفاده از رگرسیون لجستیک مدل سازی کردند و بیان داشتند که اگر متغیر  $P_i$  را به عنوان احتمال بارش روز  $\lambda$ م در مجموعه داده‌ها با متغیر پیش بینی  $X_i$  در نظر گرفته شود، مدل از رابطه زیر حاصل می‌گردد:

$$\ln\left(\frac{P_i}{1-P_i}\right) = X_i' B \quad [3]$$

در مرحله دوم ایشان مقادیر توزیع الگوریتم گاما را با مقدار بارش در روزهای مرطوب مطابقت دادند. مقدار بارندگی در روز مرطوب  $\lambda$ م و به طبع آن مقدار پیش بینی آن  $G_i$  از مجموعه داده‌های موجود برای توزیع گاما با میانگین  $M_i$  در نظر گرفته شد، بطوریکه:

$$\ln_i = G_i' Y \quad [4]$$

پارامتر شکلی توزیع‌های گاما برای همه داده‌های موجود در مجموعه یکسان فرض و با  $\nu$  نشان داده شد. این مورد بر این فرض استوار است که مقادیر بارندگی روزانه دارای یک ضریب تغییر ثابت باشند.

#### مدل جنگل تصادفی (RF)

مدل جنگل تصادفی یک روش قدرتمند در بین روش‌های یادگیری ماشینی است که بر اساس ساخت مجموعه‌ای از درختان تصمیم یا رگرسیون طراحی شده است. برای ساخت هر درخت، آن‌را به عنوان زیرمجموعه تصادفی از بین متغیرها انتخاب کرده و برای تعیین نتیجه پیش‌بینی از آن استفاده می‌شود (کلودالوز و همکاران ۲۰۱۹). دو پارامتر مهم در فرآیند مدل جنگل تصادفی؛ تعداد درختان ( $m_{tree}$ ) و تعداد متغیرهای انتخابی ( $m_{try}$ ) در هر انشعاب می‌باشند. بخش آموزش پارامترهای روش جنگل تصادفی (هم برای طبقه بندی و هم رگرسیون) بصورت زیر می‌باشد (هووبورگ و مک کابه ۲۰۱۸):

مراتبی، آنها مقیاس پذیرتر، قوی‌تر در پیدا کردن نقاط پرت دور افتاده و دارای توانایی زیاد در شبیه‌سازی مرزهای تصمیم‌گیری غیر خطی به طور طبیعی هستند. این نوع مدل برای مدیریت ویژگی‌های آماری با ده‌ها دسته مناسب هستند (شادکانی و همکاران ۲۰۲۰).

#### مدل خطی تعمیم یافته (GLM)

مدل‌های خطی تعمیم یافته به‌طور کلی ابزارهایی تحلیلی برای انواع داده‌های مختلف هستند. این مدل برای اولین بار توسط چندلر و ویاتر (۲۰۰۲) برای موضوعات هیدرولوژی بکار گرفته شد. ایشان بیان داشتند که بطور معمول در محاسبه مدل رگرسیون تعمیم یافته یک پارامتر تصادفی از مجموعه  $Y=(Y_1, \dots, Y_n)$  به پارامتر  $P$  پیش‌بینی بستگی دارد (مقادیر آن را می‌توان در یک ماتریس  $X$  به تعداد پارامتر  $n \times p$  جمع کرد که عنصر  $(i,j)$ ام مقدار پیش‌بینی کننده  $Y_i$  برای  $Y_i$  است، بطوریکه از یک توزیع احتمال برای  $Y$ ، با میانگین بردار  $M=(M_1, \dots, M_n)$  تشکیل شده است به گونه‌ای که

$$g(M) = Xb \quad [2]$$

در معادله فوق  $g(\cdot)$  یک تابع یکنواخت و پارامتر  $b$  یک بردار از ضرایب  $1 \times p$  می‌باشد (منظور از  $g(M)$  یک بردار  $1 \times n$  می‌باشد که عنصر  $\lambda$ م آن با  $g(m_i)$  مشخص شده است). رابطه ۲ حالت توسعه یافته از مدل رگرسیون خطی ساده می‌باشد. عبارت ثابت مدل را می‌توان با جایگذاری در ستون اول ماتریس  $X$  مشخص کرد. هنگامی که  $Y_s$  به عنوان یک یا چند سری زمانی در مدل تعریف می‌شوند، مقادیر سری قبلی را به عنوان متغیرهای پیش‌بینی وارد مدل می‌شوند، هر مقدار  $Y$  در سری‌های قبلی را به طور توزیع ضمنی شرطی بررسی کرده و روش معمول GLM به طور مستقیم ادامه می‌یابد. کوو و استرن (۱۹۸۲) و استرن و کوو (۱۹۸۴) به عنوان توسعه دهندگان این مدل، رویکرد دو مرحله‌ای را اتخاذ کردند. ایشان در مرحله اول در منطقه الگوی کشت



$$SI = \frac{RMSE}{\bar{x}_i} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad [V]$$

در روابط ۵ تا ۷،  $x_i$  مقادیر تبخیر-تعرق PMF،  $y_i$  مقادیر برآورد شده تبخیر-تعرق از طریق مدل‌های مورد پژوهش،  $n$  تعداد داده‌ها و  $\bar{x}_i$  میانگین مقادیر مشاهداتی تبخیر-تعرق می‌باشند.

### نتایج و بحث

به منظور بررسی امکان استفاده از ترکیب‌های مختلفی از داده‌های هواشناسی برای برآورد هرچه دقیق‌تر میزان تبخیر-تعرق و استفاده از پارامترهای هواشناسی کمتر، هفت ترکیب متفاوت از داده‌های هواشناسی تعریف گردید (جدول ۲). سپس هر یک از این سناریوها در سه مدل جنگل تصادفی، درخت گرادیان تقویت شده و مدل خطی تعمیم یافته به عنوان داده ورودی مدل در نظر گرفته شد و تبخیر-تعرق مرجع به وسیله هر یک از مدل‌ها محاسبه گردید.

جدول ۲- ترکیب‌های مختلف پارامترهای ورودی برای مدل‌های استفاده شده.

متغیر	متغیرهای ورودی	شماره
تبخیر-تعرق	دما	۱
تبخیر-تعرق	دما و رطوبت نسبی	۲
تبخیر-تعرق	دما و سرعت باد	۳
تبخیر-تعرق	دما و ساعات آفتابی	۴
تبخیر-تعرق	دما، رطوبت نسبی و ساعات آفتابی	۵
تبخیر-تعرق	دما، رطوبت نسبی و سرعت باد	۶
تبخیر-تعرق	دما، رطوبت نسبی، سرعت باد و ساعات آفتابی	۷

$$R = \frac{\left( \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right)}{\left( \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 \right) \left( \sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 \right)} \quad [5]$$

$$NS = 1 - \frac{\sum_{i=1}^N (y_i - x_i)^2}{\sum_{i=1}^N (y_i - \bar{x}_i)^2} \quad [6]$$

نتایج بدست آمده در ایستگاه بیرجند به وسیله روش داده محور جنگل تصادفی نشان می‌دهد، سناریو

۱- به تعداد  $n$  عدد نرون نمونه از داده‌های اصلی تشکیل می‌شود.

۲- برای هر یک از نرون‌های نمونه، یک درخت طبقه بندی یا رگرسیون بدون تغییر ایجاد می‌شود؛ پس از آن اصلاحات ذیل صورت می‌پذیرد: در هر گره از درخت‌های موجود به جای انتخاب بهترین انشعاب بین همه داده‌های پیش بینی، چند نمونه به طور تصادفی ( $m_{try}$ ) انتخاب و از بین آنها بهترین تقسیم بندی اتخاذ می‌گردد. پیش‌بینی داده‌های جدید با تجمیع نتایج حاصله از پیش‌بینی‌های  $n_{tree}$  درخت انجام می‌شود. سرانجام پیش‌بینی نهایی رایج‌ترین مقدار بازگشتی (برای مسائل طبقه بندی) و یا مقدار متوسط بازگشتی (برای موارد رگرسیونی) توسط درختان تصمیم گیری که جنگل را تشکیل می‌دهند، انجام می‌شود (بریمن ۲۰۰۱).

### معیارهای ارزیابی

به منظور صحت سنجی و ارزیابی روش‌های استفاده شده در پژوهش برای برآورد تبخیر-تعرق مرجع، سه فاکتور ضریب همبستگی ( $R^4$ )، شاخص پراکندگی ( $SI^5$ ) و ضریب نش-ساتکلیف ( $NS^6$ ) با روابط زیر استفاده گردید.

<sup>6</sup> Nash-Sutcliffe coefficient

<sup>4</sup> Correlation coefficient

<sup>5</sup> Scatter index

مناسبی را حاصل کرد. سناریو ششم با دارا بودن چهار پارامتر دما، رطوبت نسبی و سرعت باد بعد از سناریو هفتم در رتبه دوم از نظر معیارهای ارزیابی برای برآورد پارامتر هدف قرار می‌گیرد. ضریب همبستگی، شاخص پراکندگی داده‌ها و در نهایت ضریب نش-ساتکلیف این سناریو در شهرستان بیرجند به ترتیب برابر ۰/۹۹۴، ۰/۰۶۵ و ۰/۹۸۷ می‌باشد؛ در ایستگاه یزد به ترتیب ۰/۹۸۵، ۰/۰۹۱ و ۰/۹۷۰ می‌باشد. همچنین سناریو سوم با استفاده از دو متغیر دما و سرعت باد با قبول اندکی خطا در مواقع فقدان داده می‌تواند به عنوان روش بهینه مورد استفاده قرار گیرد. خلاصه نتایج حاصل از همه سناریوها در جدول ۳ بیان شده است.

هفتم با دارا بودن بالاترین ضریب همبستگی ۰/۹۹۴، کمترین میزان شاخص پراکندگی داده‌ها ۰/۰۶۵ و بیشترین ضریب نش-ساتکلیف ۰/۹۸۷ همچنین با استفاده از تمامی داده‌های هواشناسی موثر در تبخیر-تعرق مرجع بهترین عملکرد را در بین تمامی سناریوهای ایستگاه‌های مورد مطالعه حاصل نمود. از طرفی همین سناریو در ایستگاه یزد با اندکی تفاوت در معیارهای ارزیابی با ضریب همبستگی ۰/۹۹۲، میزان شاخص پراکندگی داده‌های ۰/۰۶۶ و ضریب نش-ساتکلیف ۰/۹۸۴ نتایج خوبی را ارائه کرد. سناریو هفتم در ایستگاه ساری با ضریب همبستگی ۰/۹۷۴، شاخص پراکندگی داده ۰/۱۵۸ و ضریب نش-ساتکلیف ۰/۹۴۹ نیز، برآورد

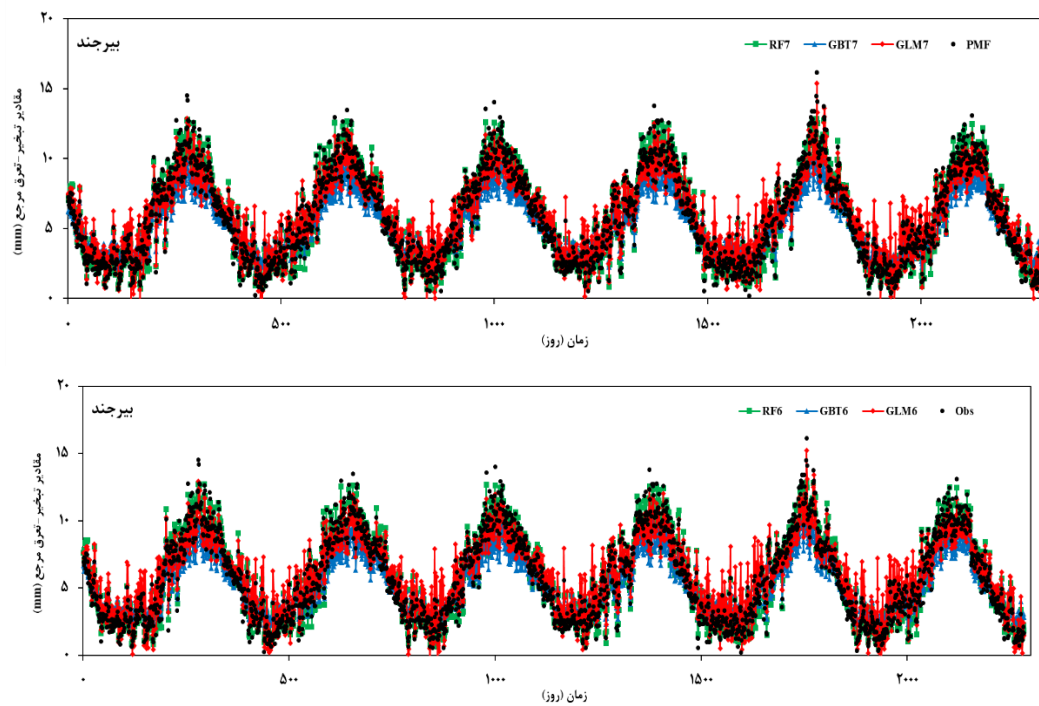
جدول ۳- نتایج حاصل از برآورد تبخیر-تعرق مرجع با روش‌های مختلف.

مدل	یزد			ساری			بیرجند		
	NS	SI	R	NS	SI	R	NS	SI	R
RF1	۰/۸۷۴	۰/۱۸۶	۰/۹۳۶	۰/۷۶۲	۰/۳۴۱	۰/۸۷۶	۰/۷۸۶	۰/۲۶۱	۰/۹۵۶
RF2	۰/۸۷۴	۰/۱۸۵	۰/۹۳۷	۰/۸۴۸	۰/۲۷۲	۰/۹۲۲	۰/۸۳۹	۰/۲۲۶	۰/۹۴۸
RF3	۰/۹۵۶	۰/۱۱۰	۰/۹۷۸	۰/۷۷۲	۰/۳۳۳	۰/۸۷۹	۰/۹۵۴	۰/۱۲۱	۰/۹۷۸
RF4	۰/۸۹۱	۰/۱۷۳	۰/۹۴۵	۰/۹۱۸	۰/۲۰۰	۰/۹۵۹	۰/۸۱۱	۰/۲۴۵	۰/۹۵۹
RF5	۰/۹۰۱	۰/۱۶۴	۰/۹۵۱	۰/۹۲۹	۰/۱۸۶	۰/۹۶۵	۰/۸۵۳	۰/۲۱۶	۰/۹۵۹
RF6	۰/۹۷۰	۰/۰۹۱	۰/۹۸۵	۰/۸۷۰	۰/۲۵۲	۰/۹۳۳	۰/۹۷۶	۰/۰۸۷	۰/۹۸۸
RF7	۰/۹۸۴	۰/۰۶۶	۰/۹۹۲	۰/۹۴۹	۰/۱۵۸	۰/۹۷۴	۰/۹۸۷	۰/۰۶۵	۰/۹۹۴
GBT1	۰/۷۶۱	۰/۲۵۵	۰/۹۳۴	۰/۶۴۲	۰/۴۱۸	۰/۸۷۳	۰/۶۲۱	۰/۳۴۷	۰/۹۵۳
GBT2	۰/۷۶۷	۰/۲۵۲	۰/۹۳۵	۰/۷۳۳	۰/۳۶۱	۰/۹۱۹	۰/۶۸۴	۰/۳۱۷	۰/۹۴۷
GBT3	۰/۸۲۶	۰/۲۱۸	۰/۹۷۷	۰/۶۵۹	۰/۴۰۸	۰/۸۷۴	۰/۸۰۴	۰/۲۴۹	۰/۹۷۷
GBT4	۰/۷۸۰	۰/۲۴۵	۰/۹۴۴	۰/۷۸۲	۰/۳۲۶	۰/۹۵۷	۰/۶۴۷	۰/۳۳۵	۰/۹۵۶
GBT5	۰/۷۸۷	۰/۲۴۱	۰/۹۴۸	۰/۷۹۷	۰/۳۱۴	۰/۹۶۲	۰/۶۹۳	۰/۳۱۲	۰/۹۵۳
GBT6	۰/۸۳۵	۰/۲۱۲	۰/۹۸۴	۰/۷۴۹	۰/۳۵۰	۰/۹۲۹	۰/۸۲۹	۰/۲۳۳	۰/۹۸۷
GBT7	۰/۸۴۹	۰/۲۰۳	۰/۹۹۱	۰/۸۱۵	۰/۳۰۰	۰/۹۷۲	۰/۸۳۵	۰/۲۲۹	۰/۹۹۳
GLM1	۰/۸۶۸	۰/۱۹۰	۰/۹۳۴	۰/۷۳۶	۰/۳۵۹	۰/۸۶۳	۰/۷۶۰	۰/۲۷۹	۰/۹۴۲
GLM2	۰/۸۷۱	۰/۱۸۸	۰/۹۳۵	۰/۸۲۰	۰/۲۹۶	۰/۹۰۶	۰/۷۹۵	۰/۲۵۵	۰/۹۴۹
GLM3	۰/۹۳۱	۰/۱۳۷	۰/۹۶۶	۰/۷۴۹	۰/۳۵۰	۰/۸۶۸	۰/۸۹۲	۰/۱۸۵	۰/۹۵۲
GLM4	۰/۸۷۷	۰/۱۸۳	۰/۹۳۹	۰/۸۶۹	۰/۲۵۲	۰/۹۳۳	۰/۷۸۹	۰/۲۵۹	۰/۹۵۴
GLM5	۰/۸۷۷	۰/۱۸۳	۰/۹۳۹	۰/۸۷۵	۰/۲۴۷	۰/۹۳۵	۰/۸۰۰	۰/۲۵۲	۰/۹۵۵
GLM6	۰/۹۳۹	۰/۱۲۹	۰/۹۷۰	۰/۸۳۰	۰/۲۸۸	۰/۹۱۱	۰/۹۰۴	۰/۱۷۵	۰/۹۶۲
GLM7	۰/۹۵۳	۰/۱۱۳	۰/۹۷۸	۰/۸۹۴	۰/۲۲۸	۰/۹۴۶	۰/۹۲۹	۰/۱۵۰	۰/۹۷۴

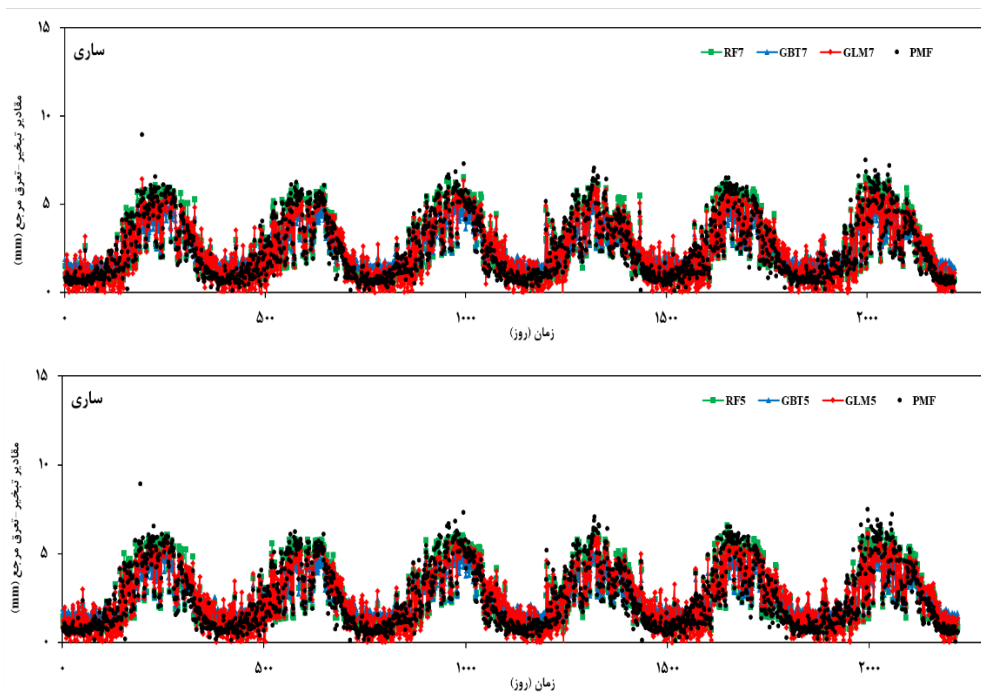
در مدل‌های خطی تعمیم یافته نیز همانند دو مدل قبلی، ترکیب داده هفتم با توجه به اینکه از همه پارامترهای موثر استفاده می‌کند در تمامی ایستگاه‌ها برآورد مطلوب‌تری را حاصل کرده است، بطوریکه بهترین نتیجه مربوط به ایستگاه یزد می‌باشد که معیارهای ارزیابی آن (ضریب همبستگی، میزان شاخص پراکندگی داده و ضریب نش-ساتکلیف) به ترتیب ۰/۹۷۸، ۰/۱۱۳ و ۰/۹۵۳ است. از طرفی سناریو سوم در ایستگاه‌های یزد و بیرجند با اندکی تفاوت در ضریب همبستگی و ضریب نش-ساتکلیف با توجه به اینکه از داده کمی نسبت به سناریو هفتم استفاده می‌کند می‌تواند به عنوان روش بهینه مورد استفاده قرار گیرد. ضعیف‌ترین نتایج در این مدل نیز مربوط به سناریو اول می‌باشد؛ ضریب همبستگی این ترکیب در ایستگاه‌های بیرجند، ساری و یزد به ترتیب ۰/۹۴۲، ۰/۸۶۳ و ۰/۹۳۴ می‌باشد که دلیل استفاده از یک پارامتر ورودی (دما) قابل پیش بینی بود (جدول ۳).

همچنین دیاگرام تیلور برای ارزیابی توانایی روش‌های مورد استفاده در برآورد تبخیر-تعریق مرجع مورد استفاده قرار گرفت. لازم به ذکر است که در دیاگرام تیلور، فاصله طولی از مبدا مختصات نشان‌دهنده انحراف معیار، خط چین‌های شعاعی سبز رنگ نشان‌دهنده ضریب همبستگی و خط چین‌های قطاعی نارنجی رنگ نشان‌دهنده مقادیر جذر میانگین مربعات خطا می‌باشند که با بزرگ شدن قطاع دایره مقدار پارامتر مذکور بیشتر می‌شود. به عبارت دیگر، هر نقطه بر روی نمودار تیلور نشان‌دهنده هم‌زمان سه پارامتر انحراف معیار، ضریب همبستگی و جذر میانگین مربعات خطا می‌باشد.

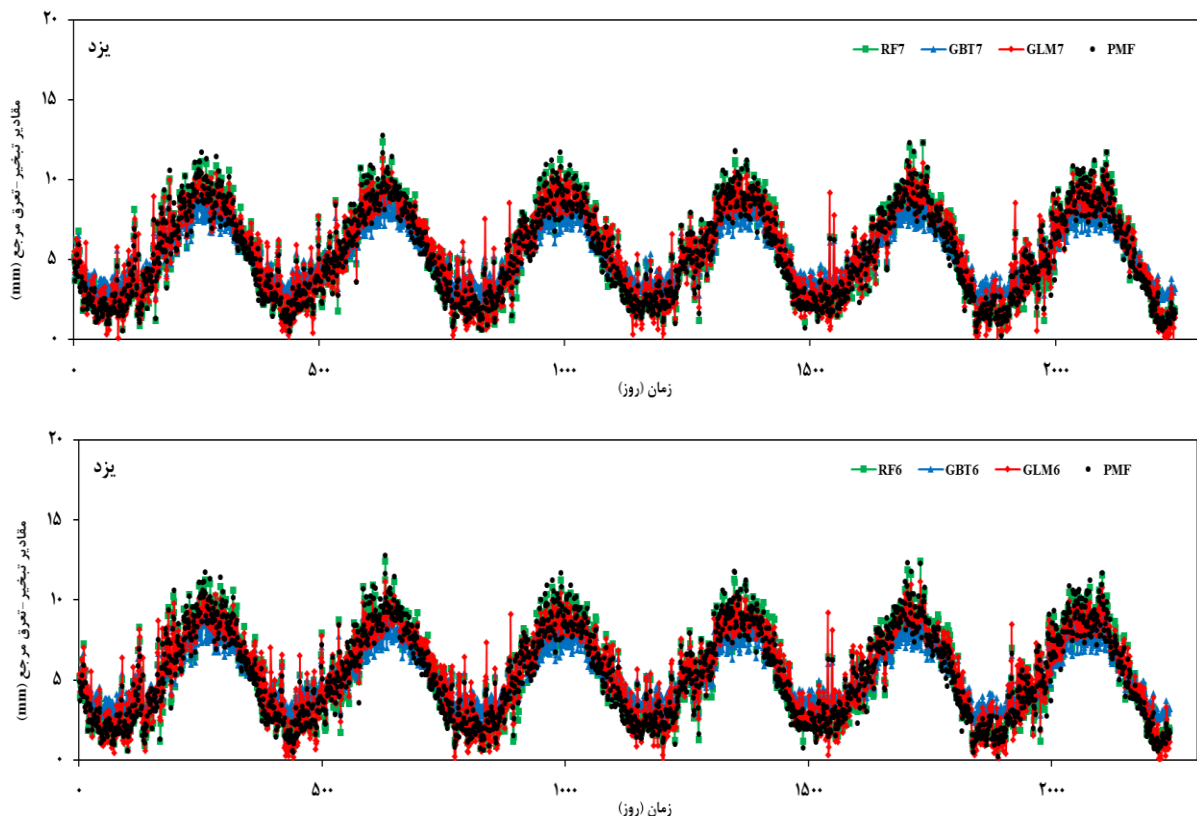
لازم به ذکر است که منظور از عبارت RF، GBT و GLM در جدول فوق همان روش‌های مورد پژوهش می‌باشد که به ترتیب عبارتند از مدل جنگل تصادفی، مدل درخت گردیان تقویت شده و مدل خطی تعمیم یافته. با بررسی توابع موجود در مدل داده محور درخت گردیان تقویت شده، تابع گاوسی بهترین نتیجه را در برآورد تبخیر-تعریق مرجع دارد. بنابراین در این پژوهش نیز تابع گاوسی مورد استفاده قرار گرفت. در این مدل نیز همانند روش جنگل تصادفی از هفت ترکیب مختلف از داده‌ها برای ورودی مدل استفاده شد تا تاثیر جز به جز داده‌ها روی پارامتر هدف مشخص شود و یک روش برای زمان فقدان داده هواشناسی در ایستگاه‌های مورد پژوهش گزارش گردد. نتایج مذکور در جدول ۳ نشان می‌دهد که در مدل درخت گردیان تقویت شده نیز، سناریو هفتم در ایستگاه بیرجند با ضریب همبستگی ۰/۹۹۳، میزان شاخص پراکندگی داده‌ها ۰/۰۶۵ و ضریب نش-ساتکلیف ۰/۹۸۷ بهترین برآورد را از تبخیر-تعریق مرجع داشته است. از طرفی همین سناریو در ایستگاه یزد و ساری با ضریب همبستگی به ترتیب ۰/۹۹۱ و ۰/۹۷۲، میزان شاخص پراکندگی داده‌ها ۰/۲۰۳ و ۰/۳۰۰ در نهایت ضریب نش-ساتکلیف ۰/۸۴۹ و ۰/۸۱۵ سناریو برتر می‌باشد. از طرفی در ایستگاه‌های بیرجند و یزد سناریو سوم با اندکی اختلاف از سناریو هفتم نتایج قابل قبولی ارائه کردند و از آنجایی که داده مورد استفاده در این سناریو به مراتب قابل استحصال‌تر از داده‌های سناریو هفتم می‌باشد می‌تواند به عنوان ترکیب بهینه انتخاب شود. در ایستگاه ساری سناریو پنجم با ضریب همبستگی ۰/۹۶۲، میزان شاخص پراکندگی داده‌ها ۰/۳۱۴ و ضریب نش-ساتکلیف ۰/۷۹۷ پس از سناریو هفتم بهترین نتیجه را داشته است (جدول ۳).



شکل ۲- تغییرات زمانی تبخیر و تعرق مرجع برآورد شده توسط برترین مدل های مورد مطالعه در ایستگاه بیرجند.



شکل ۳- تغییرات زمانی تبخیر و تعرق مرجع برآورد شده توسط برترین مدل های مورد مطالعه در ایستگاه ساری.



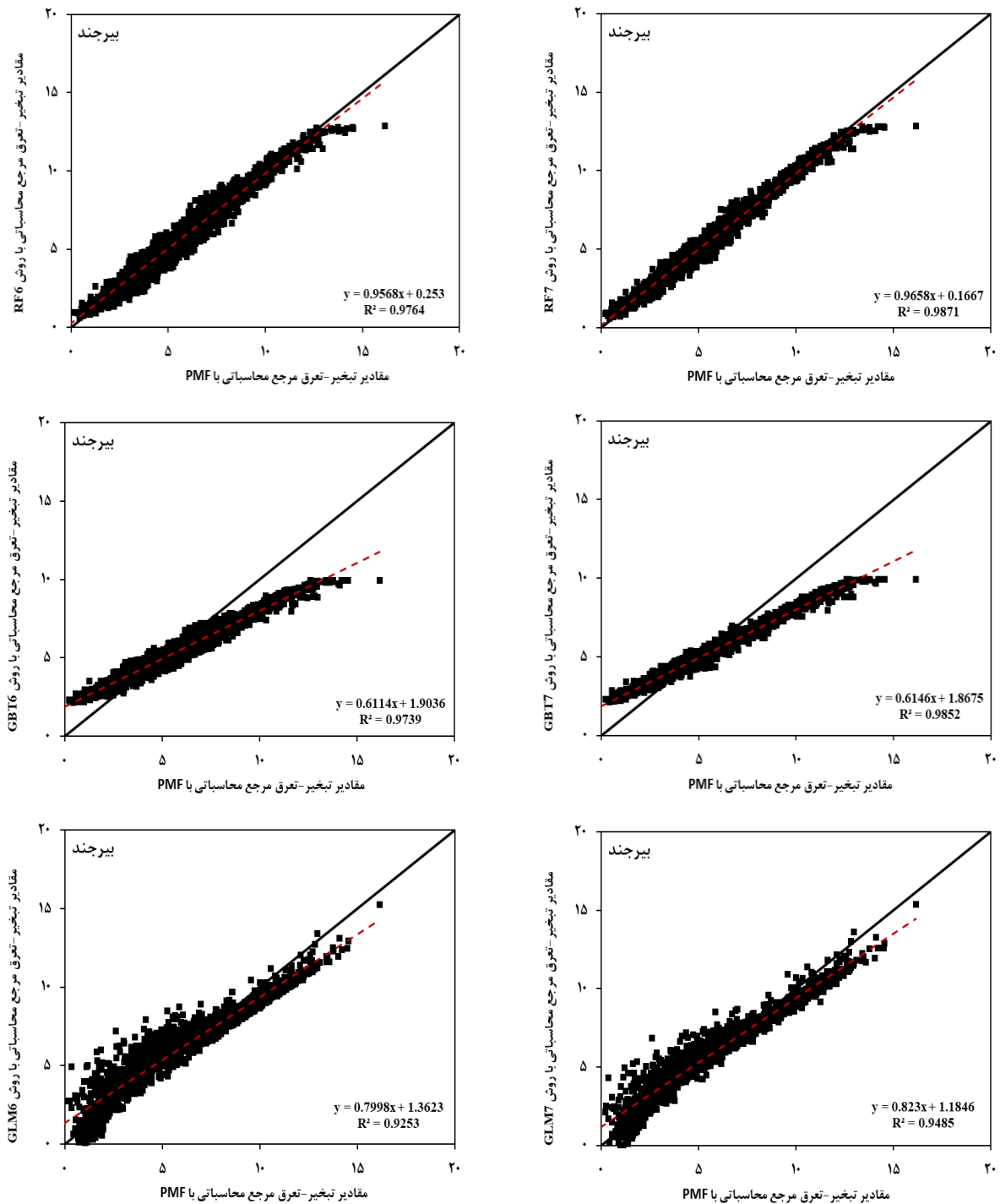
شکل ۴- تغییرات زمانی تبخیر و تعرق مرجع برآورد شده توسط برترین مدل‌های مورد مطالعه در ایستگاه یزد.

تصادفی در شهر بیرجند بهترین نتایج را ارائه نمود و به عنوان روش برتر معرفی می‌گردد. همچنین در پژوهش مشابهی در رابطه با تخمین تبخیر-تعرق گیاه مرجع در دشت سیستان با استفاده از سه روش مدل خطی تعمیم یافته، جنگل تصادفی و درخت گرادیان تقویت شده، سیاسر و همکاران (۲۰۱۹) بیان داشتند که هر سه مدل عملکرد نسبتاً مناسبی را ارائه کردند؛ اما استفاده از مدل درخت گرادیان تقویت شده بیش از هفتاد درصد زمان اجرای مدل را در مقایسه با سایر مدل‌ها کاهش می‌دهد، بنابراین این مدل توسط پژوهشگران مذکور معرفی گردید. علاوه بر این، فنگ و همکاران (۲۰۱۷) در مطالعه‌ای میزان تبخیر-تعرق مرجع را در منطقه جنوب غربی چین مدل‌سازی کردند. ایشان از دو مدل جنگل تصادفی و شبکه عصبی رگرسیون تعمیم یافته (GRNN)<sup>۷</sup> استفاده کردند و بیان داشتند که در حالت کلی هر دو روش برآورد مناسبی را ارائه می‌دهند، اما نتایج حاصله از مدل

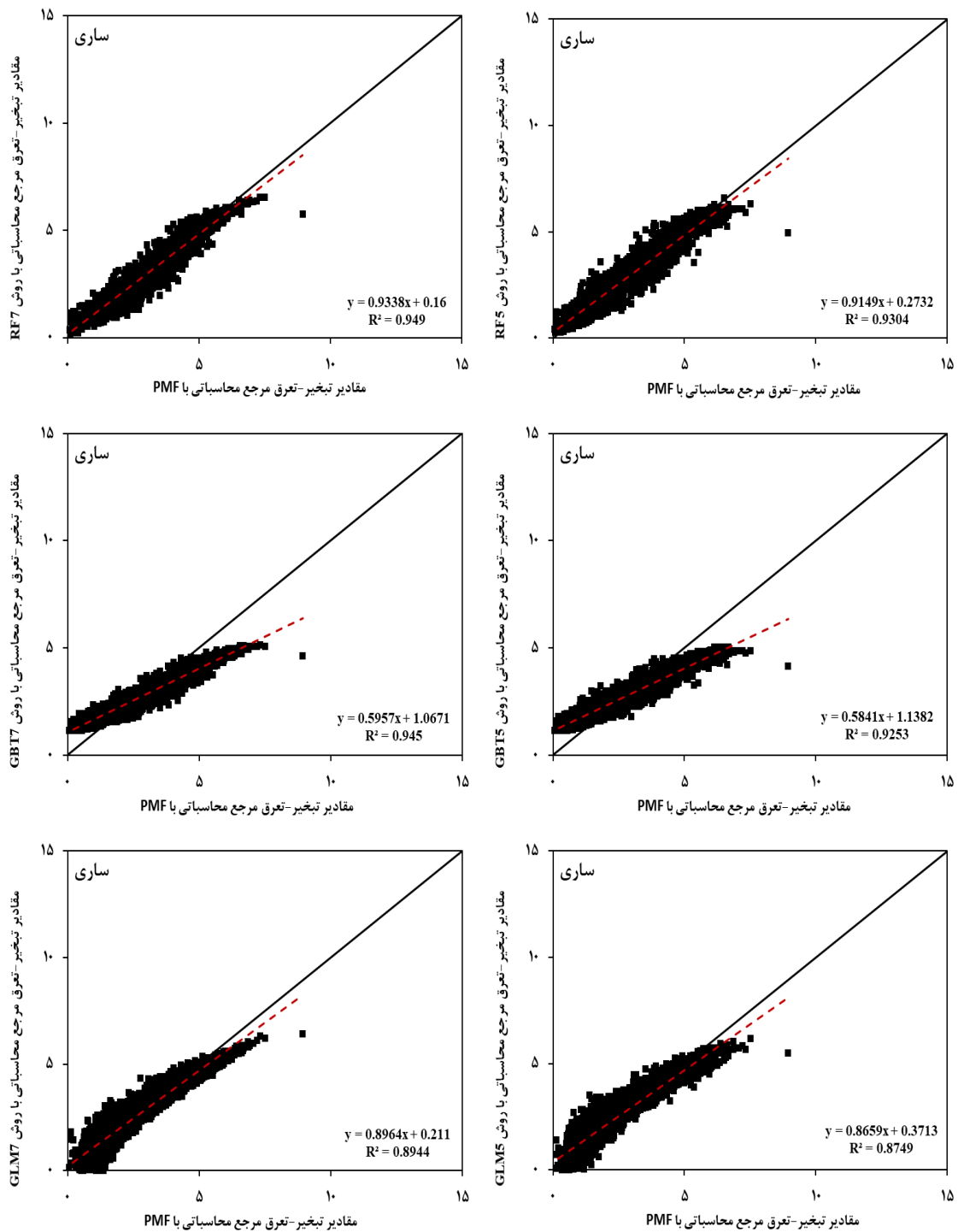
روند اشاره شده در مورد بالا بودن دقت سناریوهای برتر برای هر سه مدل مورد مطالعه از شکل‌های ۲ تا ۴ نیز قابل نتیجه‌گیری است. از شکل‌های ۵ تا ۷ نیز میتوان چنین استنباط نمود که شبیه‌سازی‌های سناریو هفتم برای هر سه ایستگاه بیرجند، یزد و ساری در هر سه مدل تطابق بیشتری با مقادیر مشاهداتی تبخیر-تعرق مرجع (محاسبه شده توسط روش پنمن-مانتیث-فائو) دارد. علاوه بر این، تمایل به کم‌برآوردی در تمامی مدل‌های GBT در ایستگاه‌های مورد مطالعه در شکل‌های ۵ تا ۷ مشاهده می‌گردد. همچنین از نمودار تیلور مشخص است که در همه ایستگاه‌ها، مدل‌های GBT-7 و RF-7 نتایج بهتری نسبت به مدل GLM-7 حاصل کرده‌اند (شکل ۸). به عبارت دیگر، می‌توان مقادیر تبخیر-تعرق مرجع را در تمامی ایستگاه‌ها با داده‌های هواشناسی دما، رطوبت نسبی، ساعات آفتابی و سرعت باد با دقت قابل قبولی شبیه‌سازی کرد، اما مدل جنگل

<sup>7</sup> Generalized Regression Neural Networks

جنگل تصادفی اندکی بهتر از مدل شبکه عصبی رگرسیون تعمیم یافته می باشد.



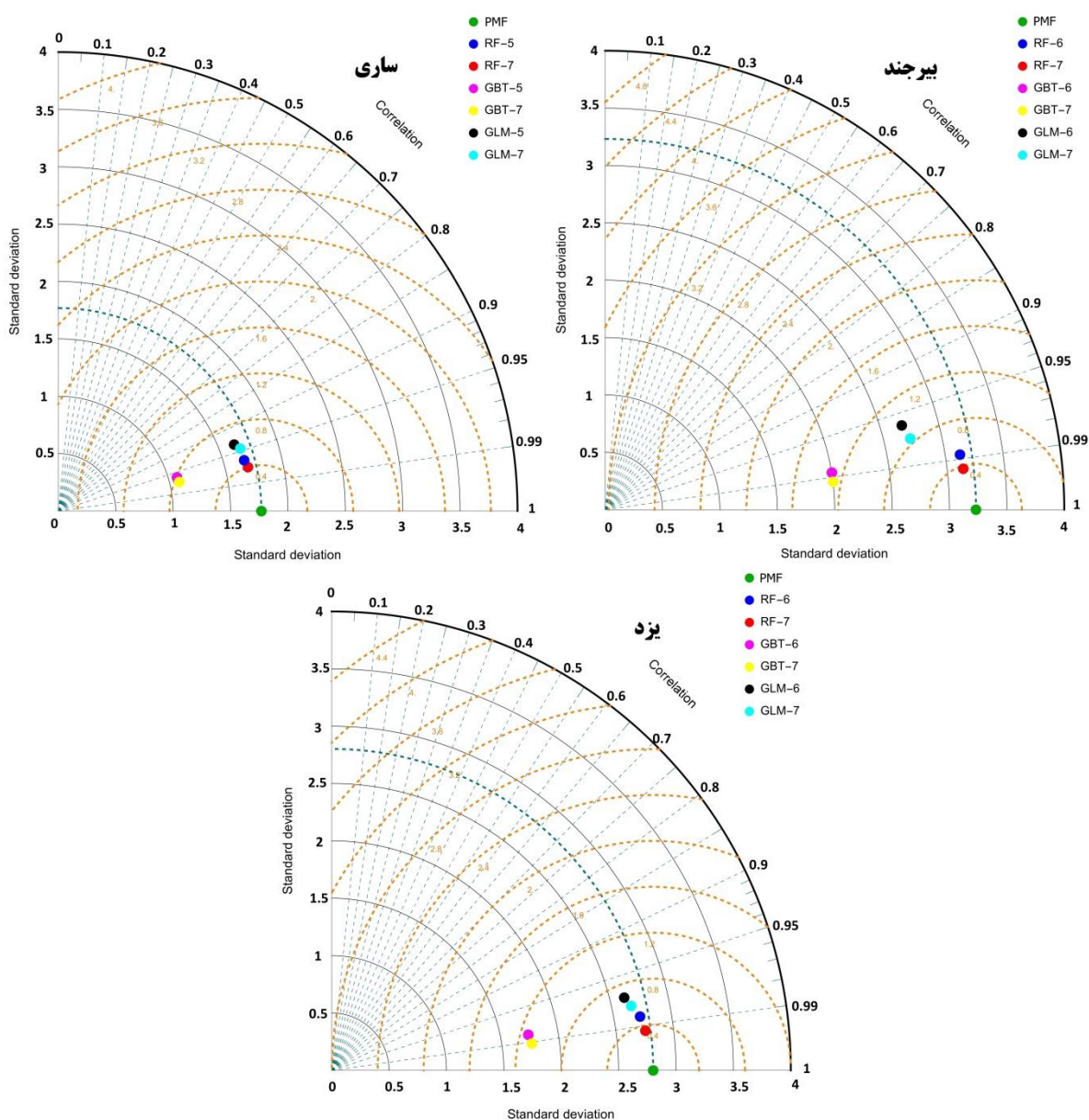
شکل ۵- پراکنش مقادیر تبخیر-تعرق برآورد شده با سناریوهای ششم و هفتم هر روش در ایستگاه بیرجند.



شکل ۶- پراکنش مقادیر تبخیر-تعریق برآورد شده با سناریوهای پنجم و هفتم هر روش در ایستگاه ساری.







شکل ۸- نمودار تیلور روش‌های مورد مطالعه در برآورد تبخیر-تعرق مرجع در سه ایستگاه یزد، ساری و بیرجند.

## نتیجه گیری کلی

تبخیر-تعرق مرجع یکی از مولفه های اصلی بیان آب در زمینه کشاورزی و از جمله عوامل موثر و تاثیرگذار جهت تنظیم دقیق تقویم آبیاری است. لذا برآورد دقیق این پارامتر تاثیر شایانی بر کاهش مصرف بی رویه آب دارد. روش‌هایی همچون پنمن مانیتث فائو که مقبولیت بالایی در بین پژوهشگران دارند بدلیل استفاده از داده‌های هواشناسی زیاد و عدم دسترسی به این داده‌ها در برخی مناطق؛ معرفی یک روش کارآمد

همواره مورد توجه پژوهشگران می‌باشد. در این مطالعه، سه مدل داده محور جنگل تصادفی، درخت گرادیان تقویت شده و مدل خطی تعمیم یافته در سه ایستگاه یزد، بیرجند و ساری استفاده شد. نتایج حاصله نشان داد که سناریو هفتم روش جنگل تصادفی با استفاده از هر چهار متغیر هواشناسی در همه ایستگاه‌ها با ضریب همبستگی ۰/۹۹۲، ۰/۹۷۴ و ۰/۹۹۴؛ شاخص پراکندگی داده ۰/۰۶۶، ۰/۱۵۸ و ۰/۰۶۵ و در نهایت ضریب نش-ساتکیف ۰/۹۸۴، ۰/۹۴۹ و ۰/۹۸۷ به ترتیب در ایستگاه‌های بیرجند،

یزد و ساری برآورد خوبی از تبخیر-تعرق مرجع ارائه داد؛ بنابراین می‌تواند توسط پژوهشگران، هنگام فقدان داده کافی مورد استفاده قرار گیرد.

#### منابع مورد استفاده

- Allen R, Pereira L, Raes D and Smith M, 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56, FAO, Rome.
- Breiman L, 2001. Random forests. *Machine Learning* 45:5-32.
- Chandler RE and Wheater HS, 2002. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research* 38(10): 1-11
- Clodoalves da Silva Júnior J, Medeiros V, Garrozia C, Montenegro A and Gonçalves GE, 2019. Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. *Computers and Electronics in Agriculture* 166:105-116.
- 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.
- Houborg R and McCabe MF, 2018. A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 135:173-188.
- Karimi S, Shiri J And Nazemi AH, 2013. Estimating daily reference crop evapotranspiration using artificial intelligences-based ANFIS and ANN techniques and empirical models. *Water and Soil Science*. 7:139-158 (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* 176:168-171.
- Mohammadrezapour A, 2017. . Monthly forecast of potential evapotranspiration models using support vector machine (SVM), genetic programming and neural - fuzzy inference system. *Journal of Irrigation and Water Engineering* 7:135-150 (In Persian with English abstract).
- Nourani V And SayahFard M, 2013. Sensitivity analysis of ANN inputs in estimating daily evaporation. *Journal of Water and Wastewater* 3:88-100 (In Persian with English abstract).
- RahimiKhob A and Mahmoudi A, 2011. Estimating actual evapotranspiration in a catchment using artificial neural networks with minimum climatic data. Case study: Emame representative catchment. *Iran- Water Resources Research* 4:51-61 (In Persian with English abstract).
- Saggi MK and Jain S, 2019. Application of fuzzy-genetic and regularization random forest (FG-RRF) estimation of crop evapotranspiration (ET<sub>c</sub>) for maize and wheat crops. *Agricultural Water Management* 229:178-192.
- Samadianfard S, Hashemi S and Izdyar M, 2019. Estimation of daily evaporation from panevaporation using machine learning methods. *Iranian Journal of Irrigation and Drainage* 4(12):1004-1015 (In Persian with English abstract).
- SamadianFard S And Panahi S, 2019. Estimating daily reference evapotranspiration using data mining methods of support vector regression and M5 model tree. *Journal of Watershed Management* 8:157-167 (In Persian with English abstract).
- Sepehri S, Abbasi F, Zarei G and Nakhjavani Moghaddam MM, 2021. Investigation of artificial neural network based models and sensitivity analysis for reference evapotranspiration estimating. *Iranian Journal of Irrigation and Drainage* 6:2089-2099 (In Persian with English abstract).
- Shadkani S, Abbaspour A SamadianFard S, Hashemi S, Mosavi A and Shamshir Band S, 2020. Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted

trees for predicting daily suspended sediment load: The case study of the Mississippi River, U.S. International Journal of Sediment Research 36:512-523

ZeionlabediniRezaabad M, GHazanfari S and Salajegheh M, 2020. ANFIS modeling with ICA, BBO, TLBO, and IWO optimization algorithms and sensitivity analysis for predicting daily reference evapotranspiration. Journal of Hydrologic Engineering 25(8):20-33.