Shahid Chamran University of Ahvaz

Irrigation Sciences and Engineering (JISE)
Vol. 48, No. 1, Spring 2025, p. 77-92

Research Paper

Journal homepage: https://jise.scu.ac.ir

EXTENDED ABSTRACT

Comparing the performance of machine learning methods in modeling daily reference ET and its spatial distribution (case study: Zanjan province)

S. Ganjei[1]* and A. H. Nazemi [2]

1- Corresponding Author, Ph.D Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. (siminganjei@tabrizu.ac.ir)*
2- Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.

ARTICLE INFO

Artlcle history:
Received: 25 October 2023
Revised: 13 May 2024
Accepted: 15 May 2024

Keywords:
GEP, GIS, MLR, RF, Reference Evapotranspiration.

TO CITE THIS ARTICLE:

Ganjei, S., Nazemi, A. H. (2025). 'Comparing the performance of machine learning methods in modeling daily reference ET and its spatial distribution (case study: Zanjan province)', Irrigation Sciences and Engineering, 48(1), pp. 77-92. doi: 10.22055/jise.2024.45305.2103.

Introduction

Iran is one of the world's semi-arid and arid regions, it has serious water resource limitations. Two important independent processes of the hydrological cycle are the combined processes of subsoil evaporation and transpiration from plant. It is critical to estimate evapotranspiration accurately since a significant amount of water is lost through this process. The FAO Penman-Monteith method is among the mostly-used indirect methods to calculate evapotranspiration. In this method, the water demands of the plant are calculated using the plant coefficients after estimating the reference evapotranspiration (Allen et al., 1998). Researchers typically use several kinds of models for predicting evapotranspiration in an effort to select a model with the highest accuracy, lowest error, and fewest input variables. This is because the FAO Penman-Monteith method requires a larger number of data points despite having acceptable accuracy. These models can be utilized for prediction since they do not have the problems and expenses of direct methods. This category of models involves machine learning models, that have excellent modeling potential that is not only in the area of evapotranspiration predicts but also in other areas of hydrology (Malekahmadi et al., 2022). The purpose of this research is to compare the performance of machine learning methods including multiple linear regression (MLR), gene expression programming (GEP) and random forest (RF) in modeling daily reference evapotranspiration and its spatial distribution in Zanjan province.

Materials and Methods

In the current research, the meteorological data of ten years (2009–2019) from the synoptic stations of Zanjan, Mahneshan, Khodabande and Khoramdare were used. The input of multiple linear regression (MLR), gene expression programming (GEP) and random forest (RF) models includes four meteorological parameters (average temperature, average relative humidity, wind speed and net radiation). In order to evaluate the accuracy of the models, the daily reference evapotranspiration of FAO Penman Monteith (FPM) was chosen as the standard method. To evaluate the performance of the models, one-third and two-thirds of the data were used for training and validation, respectively. Also,

to compare the performance of the machine learning methods with each other, the evaluation criteria including the Root Mean Square Error (RMSE), Coefficient of determination (R^2), Scatter Index (SI) and Wilmot Index (WI) were calculated and time series, dispersion and violin charts were drawn. On the other hand, the geostatistical technique was used for estimating the reference evapotranspiration in unmeasured points for zoning because of the limited number of available synoptic stations in Zanjan province. In order to analyze the spatial changes of reference evapotranspiration, its zoning was carried out after the highest-performing machine learning models were chosen, considering that it has not been evaluated in this manner before in the study area.

Findings

The random forest (RF) model in Zanjan station has good performance with WI = 0.991, SI = 0.109, and RMSE = 0.373(mm/day), R^2= 0.97; the gene expression programming (GEP) model in Khodabande station has the best performance with WI = 0.966, SI = 0.124, RMSE = 0.478(mm/day), and R^2 = 0.967 and the Multiple linear regression (MLR) model has the best performance in Zanjan station with WI = 0.955, SI = 0.08, RMSE = 0.276(mm/day), and R^2 = 0.983. After the random forest model, the gene expression programming model in Khodabandeh and Mahenshan stations and the multiple linear regression model in Khoramdare stations ranked second. In all stations, except for Zanjan station, where the MLR model performed well, the RF model has high accuracy and less error in estimating daily reference evapotranspiration. Given that the RF model has high accuracy in predicting daily reference evapotranspiration, the values of this model were used for evapotranspiration zoning using the inverse distance weighting (IDW) method. The created map showed that Reference evapotranspiration in Zanjan province is in the range of (3.466-4.273) mm/day. In the northwestern region of this province, its amount increases, and its highest amount is in the range of (4.020-4.273) mm/day at the Mahneshan station, which shows that the amount of precipitation and climatic conditions affect the amount of evaporation-transpiration.

Conclusion

Taking into consideration the important role of accurate evapotranspiration prediction in the design, planning, and management of irrigation networks and systems, in the present research, the synoptic data of 4 stations in Zanjan province for 10 years, including average temperature, average relative humidity, wind speed and net radiation were input to the model. Multiple linear regression (MLR), gene expression programming (GEP) and random forest (RF) were used to estimate daily reference evapotranspiration. According to the values of statistical indicators, the results showed that all three models have high accuracy and less error in estimating daily reference evapotranspiration, but the random forest (RF) model was chosen as the best model with a small difference. Based on the distribution of the values predicted by the machine learning models and comparing it with ET0-FPM values, the random forest (RF) model has less difference. Also, the closest violin chart to the ET0-FPM data is the random forest model chart, which shows the superiority of the random forest model over the other models. The results of zoning using the inverse distance weighting (IDW) method of this model showed that the rate of evapotranspiration is higher in the northwest of Zanjan province, especially in Mahenshan Station with (4.020-4.273) mm/day.

Acknowledgement

We acknowledge that the meteorological dataset utilized in this investigation was provided by the Islamic Republic of Iran Meteorological Organization (IRIMO).

References

1-Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO*, Rome, 300(9), D05109.

2-Malekahmadi, H., Mirzania, E., Khosravi, S., & Ebrahim Zadeh, A. (2022). Daily evapotranspiration modeling using regression and intelligent models (Case Study). *Iranian Water Research Journal*, 16(1), 49-62. Doi.org/ 10.22034/IWRJ.2022.10026.2338 (In Persian).

© 2025 The Author(s). Published by Shahid Chamran University of Ahvaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

علوم و مهندسی آبیاری جلد ۴۸، شماره‌ی ۱، بهار ۱۴۰۴، مقاله پژوهشی، ص. ۹۲-۷۷

مقایسه عملکرد روش‌های یادگیری ماشین در مدل‌سازی تبخیر-تعرق مرجع روزانه وتوزیع مکانی آن(مطالعه موردی: استان زنجان)

سیمین گنجه‌ای[1*] و امیر حسین ناظمی[2]

۱*- نویسنده مسئول، دانشجو دکتری، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران siminganjei@tabrizu.ac.ir

۲- استاد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

دریافت: ۱۴۰۲/۰۸/۰۳ بازنگری: ۱۴۰۳/۰۲/۲۴ پذیرش: ۱۴۰۳/۰۲/۲۶

چکیده

کشور ایران جزء مناطق خشک و نیمه‌خشک جهان است و با محدودیت شدید منابع آب مواجه است. از آنجا که بخش بسیار بزرگی از آب در دسترس به واسطه‌ی تبخیر-تعرق تلف می شود، برآورد دقیق آن حائز اهمیت می‌باشد. هدف از این تحقیق، مقایسه عملکرد روش‌های یادگیری ماشین شامل رگرسیون خطی چندگانه (MLR)، برنامه‌ریزی بیان ژن(GEP) و جنگل تصادفی (RF) در مدل‌سازی تبخیر-تعرق مرجع روزانه وتوزیع مکانی آن در استان زنجان است. برای این منظور از داده‌های هواشناسی ۱۰ساله (۱۳۹۸-۱۳۸۸) ایستگاه‌های سینوپتیک زنجان، ماه‌نشان، خدابنده و خرمدره استفاده شد. چهار پارامتر هواشناسی (میانگین دما، میانگین رطوبت نسبی، سرعت باد و تابش خالص) به‌عنوان ورودی مدل‌ها و تبخیر-تعرق فائو پنمن مانتیث به‌عنوان خروجی مدل‌ها درنظر گرفته شد. برای مقایسه عملکرد روش‌های یادگیری ماشین با یکدیگر، معیارهای ارزیابی محاسبه، نمودارهای سری زمانی، پراکنش و ویولونی رسم گردید. طبق نتایج به‌دست آمده در تمامی ایستگاه‌ها به جز ایستگاه زنجان که مدل رگرسیون خطی چندگانه عملکرد خوبی داشت، مدل جنگل تصادفی دقت بالا و خطای کمتری در برآورد تبخیر-تعرق مرجع روزانه را دارد. سایر مدل‌ها نیز با داشتن دقت نزدیک به مدلRF، عملکرد قابل قبولی دارند. بنابراین از خروجی مدل RF جهت پهنه‌بندی تبخیر-تعرق مرجع روزانه استان زنجان به روش IDW در GIS استفاده گردید. نتایج نشان داد که در شمال غرب استان زنجان بر مقدار تبخیر-تعرق مرجع افزوده می شود که بیشترین مقدار آن در ایستگاه ماه‌نشان(٤/٣-٤) میلی‌متر بر روز است.

کلید واژه‌ها: برنامه‌ریزی بیان ژن، تبخیر-تعرق مرجع، جنگل تصادفی، رگرسیون خطی چندگانه، سامانه اطلاعات جغرافیایی.

مقدمه

فرآیند ترکیبی تبخیر آب از سطح زمین و تعرق از پوشش گیاهی دو فرآیند مهم جدایی ناپذیر چرخه هیدرولوژیکی هستند و در تلفات بارندگی سهم زیادی دارند(Sayyahi et al., 2021). یکی از رایج‌ترین روش‌های غیرمستقیم دربرآورد تبخیر-تعرق مرجع(Reference evapotranspiration - ET_0)، معادله‌ی فائو پنمن مانتیث می‌باشد. در این روش، پس از تعیین نیاز آبی گیاه مرجع، نیاز آبی گیاه موردنیاز بر اساس ضرایب گیاهی به دست می‌آید (Allen et al., 1998). روش مرجع فائو پنمن مانتیث به پارامترهای حداقل و حداکثر دمای هوا، تابش خورشیدی، رطوبت نسبی، ساعات آفتابی و... نیاز دارد. با این وجود، این متغیرهای هواشناسی در دسترس نبوده یا معمولاً غیرقابل اعتماد هستند (Almorox et al., 2015). تعداد زیاد داده‌ها در این روش با وجود دقت قابل قبول آن، همواره پژوهشگران را به سوی استفاده از روش‌های ساده‌تر و درعین حال دقیق‌تر سوق داده است بنابراین از مدل‌های گوناگونی برای برآورد تبخیر-تعرق استفاده می‌کنند و سعی در انتخاب مدلی با دقت بیشتر، خطا کمتر و درعین حال کمترین تعداد متغیر ورودی صورت می‌پذیرد. این مدل‌ها مشکلات و هزینه‌های روش‌های مستقیم را ندارند و می‌توان به‌منظور پیش‌بینی نیز از آن‌ها استفاده کرد، ازجمله این مدل‌ها می‌توان به مدل‌های یادگیری ماشینی(Machine learning) اشاره کرد که علاوه بر در حیطه تخمین تبخیر-تعرق، بلکه در حیطه‌های دیگر علم هیدرولوژی پتانسیل بالایی را برای مدل‌سازی دارند(Malekahmadi et al., 2022).

Faloye et al. (2023) به مدل‌سازی تبخیر-تعرق محصول و کارایی مصرف آب ذرت با استفاده از شبکه عصبی مصنوعی (Artificial Neural Network) و مدل‌های رگرسیون خطی (Multiple Linear Regression) پرداختند. آن‌ها نتیجه گرفتند که با استفاده از یک مدل ساده مانند MLR، بهره‌وری آب محصول را می‌توان با دقت در شرایط مختلف مدیریت خاک و آب پیش‌بینی کرد. Malekahmadi et al. (2022) کارایی برخی از روش های داده محور شامل شبکه‌های عصبی مصنوعی و ترکیب

آن با تبدیل موجک (Wavelet- Artificial Neural Network)، ماشین بردار پشتیبان (Support Vector Regression)، درخت تصمیم (Decision Tree) و رگرسیون خطی چند متغیره برای تخمین نرخ تبخیر–تعرق در ایستگاه اسکاتسبلوف در ایالت نبراسکا بررسی کردند. نتایج آن‌ها نشان داد که در این ایستگاه مدل WANN، عملکرد بهتری در مقایسه با سایر مدل‌ها دارد و مدل‌های MLR و DT با داشتن دقت متوسط از سایر مدل‌ها ضعیف‌تر عمل کردند.

Ghourdoyee Milan et al. (2020) به شبیه‌سازی تراز سطح آب زیرزمینی با استفاده از مدل ماشین بردار پشتیبان، شبکه عصبی مصنوعی و رگرسیون خطی چند متغیره پرداختند. نتایج آن‌ها نشان داد که هر سه مدل به‌کار رفته با دقت قابل قبولی توانایی به شبیه‌سازی تراز سطح آب زیرزمینی هستند. از بین مدل‌ها مدل حداقل مربعات ماشین بردار پشتیبان به‌عنوان بهترین مدل انتخاب گردید. Swain et al. (2017) مدل رگرسیون خطی چندگانه را برای تخمین بارش سالانه در منطقه کاتک هند استفاده کردند. نتایج این پژوهش نشان داد که خروجی داده‌های این مدل با ضریب تبیین بیشتر، دقت بالایی در تخمین بارش دارد. Allahverdipour و Sattari (2023) مقایسهٔ عملکرد روش‌های داده کاوی نوین و روش کلاسیک رگرسیون خطی چندگانه در پیشبینی بارش سالانه در شهر اهواز پرداختند. نتایج آن‌ها نشان داد که روش‌های تحلیل مؤلفه های اصلی و برنامه نویسی بیان ژن (Gene Expression Programming) نسبت به مدل‌های ماشین بردار پشتیبان و رگرسیون خطی چندگانه عملکرد بهتر و دقت بیشتر در پیش‌بینی بارش سالانهٔ اهواز دارند. Mattar و Alazba (2019) به پیش‌بینی تبخیر–تعرق مرجع با دو روش MLR و GEP در مصر پرداختند. طبق نتایج پژوهش آن‌ها مدل GEP عملکرد بهتری از مدل MLR در مدل‌سازی تبخیر–تعرق مرجع دارد. Heramb et al. (2023) تبخیر–تعرق مرجع را در منطقه پانتاناگار به مدت ۲۰۱۰ تا ۲۰۱۹ توسط الگوریتم‌های یادگیری ماشین مدل‌سازی کردند. آن‌ها نتیجه گرفتند که مدل ANN و GEP دقت بالایی نسبت مدل MLR دارند. Mohammadrezapour (2017) برای پیش‌بینی تبخیر–تعرق پتانسیل ماهانه استان سیستان و بلوچستان با استفاده از مدل‌های ماشین بردار پشتیبان، برنامه ریزی ژنتیک و استنتاج فازی عصبی تطبیقی (Adaptive Neuro Fuzzy Inference System) پرداخت. ایشان نتیجه گرفتند که در مدل برنامه‌ریزی ژن الگوی متوسط دمای هوا، رطوبت نسبی، سرعت باد و ساعات آفتابی در همان ماه دارای دقت بالاتری است. Maroufpour et al. (2017) به‌منظور براورد تبخیر–تعرق غرب و شمال غرب ایران، روش ترکیبی روش فائو پنمن– مانتیث را به‌عنوان مرجع در نظر گرفتند و تکنیک‌های یادگیری ماشین را مقایسه کردند. نتایج تحقیق آن‌ها بیانگر دقت بالای برنامه‌ریزی بیان ژن در پیش‌بینی تبخیر–تعرق متوسط مرجع ماهانه می‌باشند. Kisi و Guven (2010) برای داده‌های کالیفرنیا ترکیب‌های مختلفی از پارامترهای هواشناسی را انتخاب کردند و از برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی به مدل‌سازی تبخیر–تعرق مرجع اقدام نموده‌اند. آن‌ها دریافتند که برنامه‌ریزی ژنتیک خطی نسبت به شبکه عصبی مصنوعی دقت بالاتری دارد.

Chen et al. (2020) به بررسی الگوریتم‌های یادگیری ماشین در مزرعه کلم پرداختند. آن‌ها نتیجه گرفتند که در صورت کمبود داده‌های مشاهداتی می‌توان از مدل جنگل تصادفی (Random Forest) استفاده کرد. Poursalehi et al. (2022) عملکرد الگوریتم جنگل تصادفی، مدل درخت تصمیم و شبکهٔ عصبی مصنوعی در پیش‌بینی نوسانات سطح ایستابی آبخوان در دشت بیرجند را بررسی کردند. نتایج پژوهش آن‌ها نشان داد که الگوریتم جنگل تصادفی دقت بالایی در شبیه سازی تراز سطح ایستابی آبخوان دارد.

Siasar و Honar (2019) با استفاده از مدل‌های جنگل تصادفی، الگوریتم چاید (Chid) و ماشین بردار پشتیبان به برآورد تبخیر تعرق مرجع روزانه درشمال استان سیستان وبلوچستان اقدام کردند. طبق پژوهش آن‌ها مدل جنگل تصادفی با سناریو ورودی دمای حداکثر و حداقل، دمای میانگین، سرعت باد، رطوبت میانگین و تبخیر از تشت بیشترین دقت را دارد. Feng et al. (2017) ارزیابی جنگل‌های تصادفی و شبکه‌های عصبی رگرسیون تعمیم‌یافته (Generalized Regression Neural Networks) برای تخمین تبخیر تعرق مرجع روزانه برای جنوب غربی چین استفاده کردند و نتایج نشان داد که مدل جنگل تصادفی عملکرد کمی بهتر از مدل شبکه عصبی رگرسیونی تعمیم یافته دارد. Fernández-Delgado et al. (2014) به طبقه‌بندی مدل‌های هوش مصنوعی از جمله شبکه عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی پرداختند. آنها نتیجه گرفتند که در بین طبقه بندی‌ها، مدل جنگل تصادفی بهترین است.

Rezaei et al. (2021) به برآورد و پهنه‌بندی تبخیر و تعرق پتانسیل شمال شرق ایران با استفاده از روش فائو پنمن مونتیث پرداختند. نتایج تحقیق آن‌ها نشان داد که میزان تبخیر و تعرق روزهای سال به‌طور میانگین از ۰ تا ۶/۹ میلی‌متر در روز متغیر می‌باشد و بیشترین تبخیر وتعرق در ماه‌های جولای (۱۳۴/۴) میلی‌متر در ماه و کمترین تبخیر و تعرق در ماه ژانویه (۱۸/۷) میلی‌متر در ماه اتفاق افتاده است. پهنه‌بندی تبخیر تعرق پتانسیل نشان داد بیشترین تبخیر و تعرق در جنوب غربی محدوده مورد مطالعه (طبس) و کمترین آن مربوط به شمال شرق منطقه مورد مطالعه(قوچان و بجنورد) می‌باشد. Da Silva et al. (2019) از مدل جنگل تصادفی برای پیش‌بینی تبخیر–تعرق مرجع و مقایسه دو روش زمین آماری کریجینگ معمولی (Ordinary Kriging) و وزن‌دهی معکوس فاصله (Inverse Distance Weighting) در شمال شرقی برزیل پرداختند. طبق نتایج پژوهش آن‌ها جنگل تصادفی دقت بیشتر و روش IDW زمان کمتری را در ایجاد درون یابی صرف می کند. Karami و Asadi (2016) به پهنه‌بندی

DOI: 10.22055/jise.2024.45305.2103

تبخیر–تعرق مرجع پرداختند. نتایج آن‌ها نشان داد که روش درون یابی IDW دقت بالایی دارد و میزان تبخیر–تعرق در قسمت‌های شمالی و شرقی دریاچه ارومیه بیشتر است. Sharghi et al. (2010) در ۲۹ ایستگاه کلیماتولوژی و سینوپتیک انتخاب شده در استان یزد به پهنه‌بندی تبخیر–تعرق مرجع به روش فائو پنمن مانتیث پرداختند. نتایج بررسی آن‌ها نشان داد که مناطق مرکزی، جنوبی و برخی نواحی غربی استان از شدت تبخیر–تعرق بیشتری نسبت به مناطق شمالی و شرقی برخوردار هستند.

در همین راستا، این پژوهش با هدف مدل‌سازی دقیق تبخیر–تعرق مرجع روزانه در استان زنجان با اقلیم خشک و نیمه‌خشک انجام گرفته است. با توجه به اهمیت برآورد دقیق تبخیر– تعرق در طراحی، برنامه ریزی و مدیریت سامانه‌های آبیاری، در این مطالعه از داده‌های ده ساله چهار ایستگاه هواشناسی و سه الگوریتم یادگیری ماشین شامل جنگل تصادفی برنامه‌ریزی بیان ژن و رگرسیون خطی چندگانه استفاده شده‌است تا عملکرد این روش‌ها در شرایط اقلیمی منطقه مورد ارزیابی قرار گیرد. کلیه مدل‌ها با ورودی‌های یکسان شامل پارامترهای هواشناسی میانگین دما، میانگین رطوبت نسبی، سرعت باد و تابش خالص آموزش داده شده و دقت آن‌ها در برآورد تبخیر– تعرق مرجع روزانه با استفاده از تحلیل‌های آماری و گرافیکی مقایسه شده است. از سوی دیگر، با توجه به محدودیت تعداد ایستگاه‌های هواشناسی از سامانه اطلاعات جغرافیایی(Geographic Information System) جهت تخمین تبخیر–تعرق در نقاط فاقد داده و انجام پهنه بندی مکانی استفاده شد. بدین منظور، پس از شناسایی مدل برتر، نتایج آن برای برآورد تغییرات مکانی تبخیر– تعرق در سطح استان به کار گرفته شد؛ فرآیندی که تاکنون در این منطقه با چنین دقت و جامعیتی صورت نگرفته است. بهره گیری از مدل‌سازی پیشرفته و ارائه نقشه‌ی کاربردی، این پژوهش را به گامی نوآورانه در جهت بهینه‌سازی مدیریت منابع آب کشاورزی و تصمیم گیری‌های منطقه‌ای در اقلیم‌های خشک و نیمه‌خشک تبدیل کرده است.

مواد و روش‌ها

منطقهٔ مورد مطالعه

استان زنجان با بیش از ۲۲ هزار کیلومتر مربع مساحت، در منطقه شمال غرب ایران قرار گرفته است. محدوده جغرافیایی این استان ما بین ۳۵ درجه و ۳۳ دقیقه تا ۳۷ درجه و ۱۵ دقیقه عرض شمالی از خط استوا و ۴۷ درجه و ۱۰دقیقه تا ۴۹ درجه و ۲۶ دقیقه طول شرقی از نصف النهار گرینویچ قرار دارد همچنین اقلیم غالب منطقه بر اساس طبقه بندی اقلیمی به روش دومارتن، نیمه خشک است. مرجع روزانه فائوپنمن‌مانتیث (FAO Penman Monteith-FPM) به‌عنوان پارامتر خروجی استفاده شد. مشخصات اقلیمی و مشخصات مکانی ایستگاه های مذکور در جدول (۱) و نقشه موقعیت ایستگاه‌های مورد مطالعه در شکل (۱) نشان داده شده است.

جدول ۱- مشخصات جغرافیایی ایستگاه های سینوپتیک استان زنجان

Table 1- Geographical characteristics of synoptic stations in Zanjan province

Station	Average Annual precipitation (mm)	Average ET_0 (mm/day)	Height (m)	Latitude (°N)	Longitude (°E)
Zanjan	0.807	3.463	1659.4	36.6603	48.5217
Mahneshan	0.702	4.269	1284.5	36.74	47.6838
Khodabande	1.134	3.934	1887	36.1438	48.5891
Khoramdare	0.844	3.789	1575	36.1958	49.2108

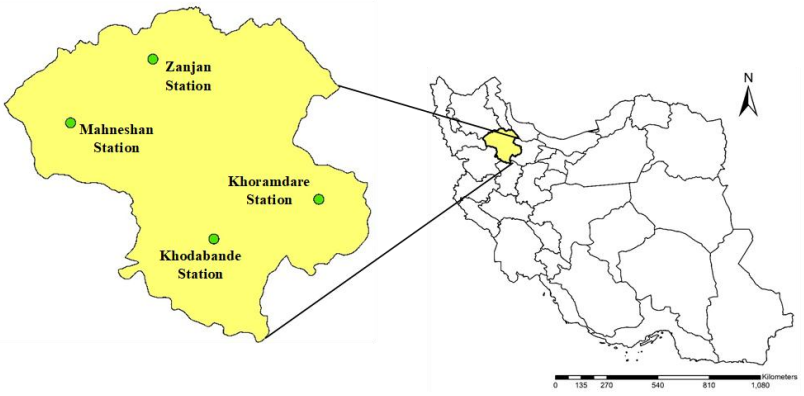


Fig.1- Distribution of stations and the studied area

شکل ۱- پراکنش ایستگاه ها و منطقه مورد مطالعه

داده‌های مورد استفاده و روش تحقیق

در تحقیق حاضر از ایستگاه‌های سینوپتیک زنجان، ماه‌نشان، خدابنده و خرمدره با پارامترهای هواشناسی میانگین دما T_{mean} (^0C)، میانگین رطوبت نسبی RH_{mean} (%)، سرعت باد U_2 (m.s-1)، تابش خالص R_n (MJ.m-2.day-1) به‌صورت روزانه در بازه زمانی ۱۰ سال(۱۳۸۸-۱۳۹۸) به‌عنوان ورودی مدل‌های یادگیری ماشین شامل رگرسیون خطی چندگانه(MLR)، برنامه‌ریزی بیان ژن(GEP) و جنگل تصادفی(RF) و تبخیر-تعرق مرجع روزانه فائو پنمن مانتیث(ET_0-FPM) به عنوان روش استاندارد برای ارزیابی دقت مدل‌ها استفاده شد. برای بررسی میزان عملکرد مدل‌ها ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای صحت سنجی استفاده شد. نتایج به‌دست آمده از مدل‌های یادگیری ماشین با استفاده از شاخص‌های آماری جذر میانگین مربعات خطا ، ضریب تعیین ، نمایه پراکندگی و شاخص ویلوت مقایسه و بهترین مدل انتخاب شد. علاوه بر شاخص‌های آماری از نمودارهای ارزیابی پراکنش و ویولونی برای مقایسه عملکرد مدل‌ها رسم گردید. با انتخاب بهترین مدل به پهنه بندی ET_0-Model پرداخته شد.

مدل استاندارد فائو پنمن مونتیث (FAO56-PM)

معادله مدل تجربی فائو پنمن مانتیث برای محاسبه تبخیر-تعرق مرجع روزانه و روش استاندارد ارزیابی دقت سایر مدل‌ها بر اساس رابطه (۱) است (Allen et al., 1998).

$$ET_0=\frac{0.408\Delta(R_n-G)+\gamma\frac{900}{T_{mean}+273}U_2(e_S-e_a)}{\Delta+\gamma(1+0.34U_2)} \qquad (۱)$$

در رابطه اخیر، ET_0 تبخیر-تعرق گیاه مرجع (میلی متر بر روز)، Δ شیب تابع فشار بخار اشباع (کیلو پاسکال بر درجه سلسیوس)، R_n تابش خالص خورشیدی (مگا ژول بر متر مربع بر روز)، G چگالی شار حرارتی خاک (مگاژول بر متر مربع بر روز)، γ ثابت سایکرومتری (کیلو پاسکال بر درجه سلسیوس)، T_{mean} متوسط دمای هوای (درجه سلسیوس)، U_2 متوسط سرعت باد در یک شبانه روز در ارتفاع دو متری از سطح زمین (متر بر ثانیه)، e_s فشار بخار اشباع (کیلو پاسکال) و e_a نیز فشار بخار واقعی (کیلو پاسکال) می باشد. مقدار چگالی شار حرارتی خاک برای دوره های روزانه برابر با صفر در نظر گرفته می‌شود.

رگرسیون خطی چندگانه(MLR)

رگرسیون خطی چند متغیره یک روش آماری است که ارتباط دو یا چند متغیر مستقل با یک متغیره وابسته را بیان می‌کند (Yu et al., 2018). معادلهٔ (۲) رگرسیون خطی چند متغیره استفاده شده در این مطالعه است.

$$ET_{model}=a_0+a_1X_1+\cdots+a_kX_k+e \qquad (۲)$$

ET_{model} متغیر وابسته، X_1 و X_k متغیرهای مستقل، a_0 و a_k ضرایب رگرسیون و e مقدار باقی‌مانده خطای رگرسیونی است که می‌توان صرف‌نظر کرد. رگرسیون خطی چند متغیره دارای روش‌های رگرسیون همزمان(Enter)، گام به گام(Stepwise)، پیشرونده (Forward) ، پسرونده (Backword) و حذفی (Remove) می‌باشد. در این پژوهش برای بکار بردن رگرسیون چند متغیره از SPSS (Statistical Package for the Social Sciences) به روش هم‌زمان استفاده شده است.

برنامه‌ریزی بیان ژن(GEP)

برنامه‌ریزی بیان ژن توسط فریرا در سال ۱۹۹۹ مطرح شد که ترکیبی از روش‌های الگوریتم ژنتیک (Genetic Algorithm) و برنامه‌ریزی ژنتیک(Genetic Programming) است که در آن کروموزوم‌های خطی و ساده با طول ثابت، شبیه به آن چه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اشکال و اندازه‌های مختلف، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک، ادغام می‌شوند. در این روش پدیده‌های مختلف با استفاده از مجموعه ای از ترمینال‌ها و توابع مدل‌سازی می شود (Ferreira., 2001). در اجرای این مدل ، ابتدا بلوک‌های موجود نظیر پارامترهای ورودی، هدف و نیز تابع ارتباط دهنده آن‌ها می‌باشد، تعریف شده و سپس ساختار مناسب مدل و ضرایب آن مشخص می‌شوند. این روش دارای یک معادله ارتباط دهنده بین پارامترهای ورودی و خروجی بوده بنابراین توانایی انتخاب خودکار پارامترهای مناسب مدل و حذف پارامترهای غیر مرتبط است که این کار باعث کاهش ابعاد پارامتر های ورودی خواهد شد. از مدل برنامه‌ریزی بیان ژن به‌طور گسترده در مدل‌سازی مسائل مربوط به تعیین ساختار پدیده‌ها به‌کار برده می‌شود (Borrelli et al., 2006). در تحقیق حاضر ۷۰ درصد داده‌ها (سال۱۳۸۸-۱۳۹۵) برای آموزش و ۳۰ درصد داده‌ها (سال۱۳۹۶-۱۳۹۸) برای آزمون مدل درنظر گرفته شد و از نرم‌افزار GeneXproTools 4.0 برای مدل‌سازی ET_0 روزانه استفاده شد. در این تحقیق عملگرهای {-,+, $X^{1/3}$,X^3,X^2,Ln,exp,sqrt,/,*,} استفاده شدند. در جدول (۲) مشخصات پارامترهای برنامه‌ریزی ژنتیک به‌کار برده شده، ارائه گردیده است.

DOI: 10.22055/jise.2024.45305.2103

جدول ۲- مشخصات پارامترهای برنامه ریزی ژنتیک

Table 2- Genetic operators employed in implementing the GEP model

Number of chromosomes	30	One point recombination rate	0.3
Head size	8	Two points recombination rate	0.3
Number of genes	3	Gene recombination rate	0.1
Linking function	+	Gene transposition rate	0.1
Fitness function	RMSE	Insertion sequence transposition rate	0.1
Mutation rate	0.044	Root insertion sequence	0.1
Inversion rate	0.1		

جنگل تصادفی(RF)

روش جنگل تصادفی را اولین بار Breiman (2001) با توسعه درخت‌های تصمیم، به‌عنوان یک روش جدید ارائه داده است که تخمین چندین الگوریتم منفرد را با هم با استفاده از قوانین مبتنی ترکیب می کند. این روش در بین روش‌های درختی، روش نسبتاً پیچیده ای است که به‌دلیل افزایش دقت مدل در آن چندین درخت تصمیم آموزش داده می‌شود. نتیجه حاصل تخمین گروهی از درختان تصمیم است. اصول کلی روش‌های آموزش گروهی بر اساس این فرض است که دقت آن‌ها از دیگر الگوریتم‌های آموزشی بیشتر است، چون ترکیبی از چند مدل تخمینی، دقیق‌تر از یک مدل می‌باشد و گروه‌ها توان مجموعه‌های منفرد و منحصر به فرد از طبقه‌ها را بیشتر می‌کنند، در حالی که هم‌زمان نقاط ضعف طبقه‌ها را کاهش می‌دهند (Kotsiantis and Pintelas, 2004). در روش جنگل تصادفی دو ویژگی n گره و m مشخصه به‌ترتیب برای تعداد درختان و تعداد متغیرهای کمکی مورد استفاده در هر زیر مجموعه است. تعداد متغیرها می‌تواند از یک تا تعداد کل متغیرهای کمکی را شامل شود. تعداد درختان توسط کاربر و معمولاً از ۵۰۰ تا ۱۰۰۰ انتخاب می گردد .CART (Classification And RegressionTree) یک الگوریتم درخت باینری در الگوریتم درخت تصمیم است و جنگل تصادفی مجموعه‌ای از درختانCART است و شامل چهار مرحله می‌باشد (Wang et al 2018): ۱-نحوه انتخاب نمونه‌های مختلف بوت استرپ (Bootstropt) X_j (j= تکرار بوت استرپ) به‌طور تصادفی از مجموعه داده اصلی X می باشد. دو سوم از نمونه‌ها در یک نمونه بوت استرپ و یک سوم از نمونه‌ها خارج از مجموعه جاگذاری می‌شود. هر درخت برای تطابق با یک زیر مجموعه خاصی از بوت استرپ ساخته می‌شود. ۲-در یک گره در هر درخت تقسیم جدید به طور تصادفی از همه شاخص‌ها و متغیر ورودی با کمترین میانگین مربع خطا به‌عنوان معیار تقسیم درخت رگرسیون انتخاب می‌شود. ۳-مراحل تقسیم داده‌ها در هر گره داخلی مطابق با مراحل بالا تا وقتی که تمام درختان تصادفی رشد کنند و شرایط توقف حاصل شود، تکرار می‌شود. ۴-در نهایت خروجی طبقه‌بندی براساس یک نتیجه میانگین، از پیش‌بینی‌های تمام درخت‌های منفرد آموزش دیده به‌دست می‌آید. برای مدل جنگل تصادفی از نرم‌افزار Statistica استفاده شد.

پهنه بندی تبخیرتعرق-مرجع

روش‌های مختلفی برای پهنه‌بندی بر اساس نقاط دارای پارامتر معین گزارش شده است که اغلب بر اساس درون‌یابی هستند که هدف از این روش‌ها تعمیم اطلاعات حاصل از نقاط به سطح می‌باشد. برای این منظور نقاط تعیین شده با فواصل مناسب روی نقشه استان مشخص شده و مقادیر پارامترهای اقلیمی در این نقاط اندازه‌گیری و به سایر قسمت‌ها، تعمیم داده شود (Herrera et al., 2019). در روش وزن‌دهی معکوس فاصله (IDW) برای هرکدام از نقاط اندازه‌گیری شده، وزنی برطبق فاصله بین آن نقطه تا موقعیت نقطه‌ای که اندازه‌گیری نشده است، در نظر می‌گیرند. سپس توان وزندهی این وزن‌ها را کنترل می‌کند، به صورتی که توان‌های بزرگتر تأثیر نقاط دورتر از نقطه مورد برآورد را کاهش داده و توان‌های کوچکتر وزن‌ها را به طور یکنواخت‌تری بین نقاط همسایه توزیع می‌کنند. البته باید درنظر گرفت که این روش بدون توجه به موقعیت و آرایش نقاط، فقط فاصله آن‌ها را به حساب می‌آورد، یعنی نقاطی که دارای فاصله یکسانی از نقطه برآورد هستند دارای وزن یکسانی می‌باشند. مقدار عامل وزنی طبق رابطه (۳) محاسبه می‌گردد (Sanjari, 2023):

$$X^{*}=\frac{\sum_{i=1}^{n}\frac{X_i}{D_i^{p}}}{\sum_{i=1}^{n}\frac{1}{D_i^{p}}} \qquad (۳)$$

که در آن:X^{*} مقدار برآورد شده، X_i مقدار متغیر مشاهداتی در ایستگاه i ام، D_i فاصله بین نقطه مقدار مشاهداتی و مقدار برآورد شده، p توان وزن‌دهی و n تعداد نقاط مشاهداتی می‌باشد. برای پهنه بندی تبخیر-تعرق از روش وزن دهی معکوس فاصله (IDW) در ARC GIS 10.6 استفاده شد.

شاخص های ارزیابی

از چهار معیار ارزیابی شامل جذر میانگین مربعات خطا (RMSE)، ضریب تعیین (R^2)، نمایه پراکندگی(SI) و شاخص ویلوت (WI) جهت مقایسه مدل‌ها با یکدیگر و انتخاب بهترین مدل استفاده شد.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(ET_m-ET_o)^2} \quad (۴)$$

$$R^2=\left[\frac{\sum_{i=1}^{n}(ET_m-\overline{ET_m})(ET_o-\overline{ET_o})}{\sqrt{\sum_{i=1}^{n}(ET_m-\overline{ET_m})^2\sum_{i=1}^{n}(ET_o-\overline{ET_o})^2}}\right]^2 \quad (۵)$$

$$SI=\frac{RMSE}{\overline{ET_o}} \quad (۶)$$

$$WI=1-\left(\frac{\sum_{i=1}^{n}(ET_o-ET_m)^2}{\sum_{i=1}^{n}(|ET_m-\overline{ET_o}|+|ET_o-\overline{ET_m}|)^2}\right) \quad (۷)$$

در روابط اخیر ET_0 و ET_m به‌ترتیب مقادیر مدل استاندارد فائو پنمن مونتیث و تبخیر-تعرق محاسبه شده توسط مدل ، $\overline{ET_0}$ و $\overline{ET_m}$ به‌ترتیب متوسط مقادیر مشاهداتی و مدل، n نمایانگر تعداد داده‌ها و i نیز گام زمانی مورد نظر می‌باشد. هر چه مقدار ضریب تعیین و شاخص ویلوت بیشتر و مقدار جذر میانگین مربعات خطا و نمایه پراکندگی کمتر باشد مدل دقت بیشتری در مدل‌سازی دارد.

علاوه بر ارزیابی‌های آماری از نمودارهای ارزیابی پراکنش و ویولونی برای مقایسه عملکرد مدل‌ها رسم گردید.

نتایج و بحث

نتایج شاخص‌های آماری به‌دست آمده از مدل‌های یادگیری ماشین در جدول (۳) و روابط ریاضی به‌دست آمده از مدل GEP در جدول (۴) گزارش شده است. طبق جدول (۳)، مدل جنگل تصادفی در ایستگاه زنجان عملکرد خوبی با داشتن WI=0.99, SI=0.109, RMSE=0.373(mm/day), R^2=0.97 و عملکرد ضعیف در ایستگاه خرمدره با داشتن WI=0.988, SI=0.133, RMSE=0.496(mm/day), R^2=0.957 همچنین برای مدل برنامه‌ریزی بیان ژن در ایستگاه خدابنده با WI=0.966, SI=0.124, RMSE=0.478(mm/day), R^2=0.967 بهترین عملکرد و در ایستگاه خرمدره با WI=0.987, SI=0.14, RMSE=0.525(mm/day), R^2=0.953 ضعیف‌ترین عملکرد و برای مدل رگرسیون خطی چندگانه در ایستگاه زنجان با WI=0.955, SI=0.08, RMSE=0.276(mm/day), R^2=0.983 بهترین عملکرد و در ایستگاه خدابنده WI=0.987, SI=0.148, RMSE=0.572(mm/day), R^2=0.952 ضعیف‌ترین عملکرد را داراست. در تمامی ایستگاه ها به جز ایستگاه زنجان که مدل رگرسیون خطی چندگانه عملکرد خوبی داشت، مدل جنگل تصادفی دارای دقت بالا و خطای کمتری در برآورد تبخیر-تعرق مرجع روزانه می‌باشد. بعد از مدل جنگل تصادفی، مدل برنامه ریزی بیان ژن در ایستگاه‌های خدابنده و ماه‌نشان و مدل رگرسیون خطی چندگانه در ایستگاه‌های خرمدره در رتبه دوم قرار گرفتند.

جدول ۳- نتایج آماری مدل های یادگیری ماشین در مدل‌سازی تبخیر-تعرق مرجع

Table 3- Statistical results of machine learning models in reference evapotranspiration modeling

Station Name		Model					
		MLR		GEP		RF	
		Test	Train	Test	Train	Test	Train
Khoramdare	WI	0.988	0.988	0.987	0.987	0.988	0.991
	SI	0.137	0.136	0.14	0.136	0.133	0.113
	RMSE (mm/day)	0.513	0.501	0.525	0.523	0.496	0.431
	R^2	0.955	0.956	0.953	0.951	0.957	0.968
Kodabande	WI	0.987	0.982	0.966	0.989	0.99	0.991
	SI	0.148	0.171	0.124	0.129	0.126	0.115
	RMSE (mm/day)	0.572	0.68	0.478	0.511	0.488	0.454
	R^2	0.952	0.931	0.967	0.961	0.968	0.971
Mahneshan	WI	0.988	0.989	0.989	0.991	0.99	0.991
	SI	0.139	0.124	0.127	0.106	0.117	0.11
	RMSE (mm/day)	0.548	0.545	0.502	0.468	0.488	0.473
	R^2	0.962	0.958	0.97	0.969	0.968	0.969
Zanjan	WI	0.955	0.992	0.997	0.987	0.991	0.993
	SI	0.08	0.103	0.144	0.134	0.109	0.096
	RMSE (mm/day)	0.276	0.357	0.499	0.463	0.373	0.335
	R^2	0.983	0.97	0.955	0.951	0.97	0.9748

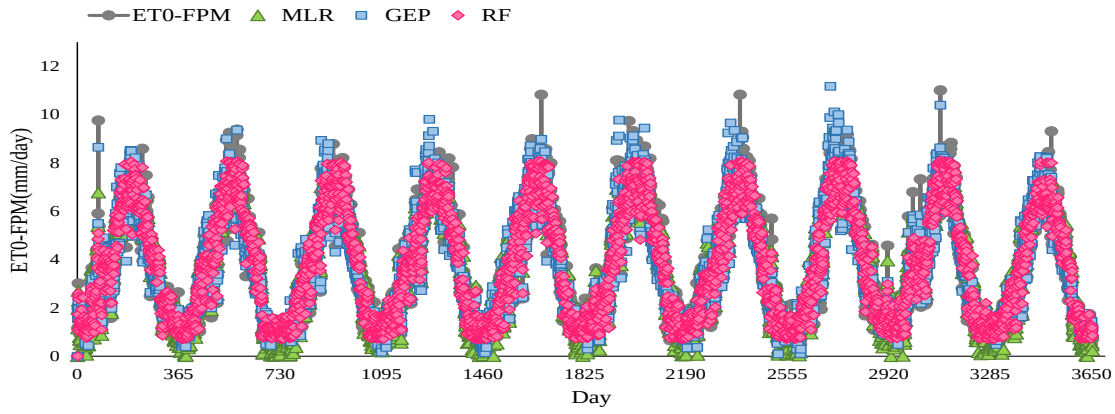
DOI: 10.22055/jise.2024.45305.2103

روند تغییرات تبخیر–تعرق مرجع روزانه مدل‌های یادگیری ماشین با داده‌های ET_0-FPM در شکل (۲) نشان داده شده است. همان‌طور که مشاهده می‌شود داده‌های ET_0 مدل جنگل تصادفی در مقایسه با مدل‌های GEP و MLR به داده‌های ET_0-FPM نزدیک تر است پس می‌توان دریافت که مدل جنگل تصادفی در برآورد تبخیر–تعرق مرجع روزانه از دقت بالایی برخوردار است. همچنین نحوه پراکنش مقادیر پیش‌بینی شده توسط مدل‌ها و مقادیر مشاهداتی ET_0-FPM در شکل (۳) ارائه شده است. هر چقدر میزان پراکنش از خط نیمساز کمتر باشد، اختلاف مقادیر پیش‌بینی شده با مقادیر ET_0-FPM کمتر بوده و دقت مدل بیشتر است. براساس شکل (۳) سایر مدل ها نیز با وجود داشتن نتایج قابل قبول، دقتی پایین‌تری نسبت به مدل جنگل تصادفی دارند. نمودار ویولونی که نمودار اصلاح شده و کامل تری از نمودار جعبه ای است در شکل (۴) رسم شده است. وضعیت پراکندگی داده ها نسبت به مقادیر بیشینه، کمینه و میانه برای مقادیر مشاهداتی و مدل‌ها ترسیم شده است. نزدیکترین نمودار ویولونی به داده‌های ET_0-FPM نمودار مدل جنگل تصادفی است. نتایج پژوهش حاضر با یافته‌های Wu . et al (2020) مبنی بر مدل جنگل تصادفی در مناطق خشک که داده‌های هواشناسی محدود است در شبیه‌سازی تبخیر–تعرق عملکرد خوبی دارد، در یک راستا است. همچنین نتایج این مطالعه با نتایج Aiyelokun و Agbede (2021) که با استفاده از مدل جنگل تصادفی و با ترکیب چهار پارامتر حداکثر دما، رطوبت نسبی، سرعت باد و ساعات آفتابی به پیش‌بینی تبخیر–تعرق مرجع پرداختند و نشان دادند که این مدل با دقت بالا تبخیر–تعرق مرجع را مدل‌سازی می‌کند، مطابقت دارد.

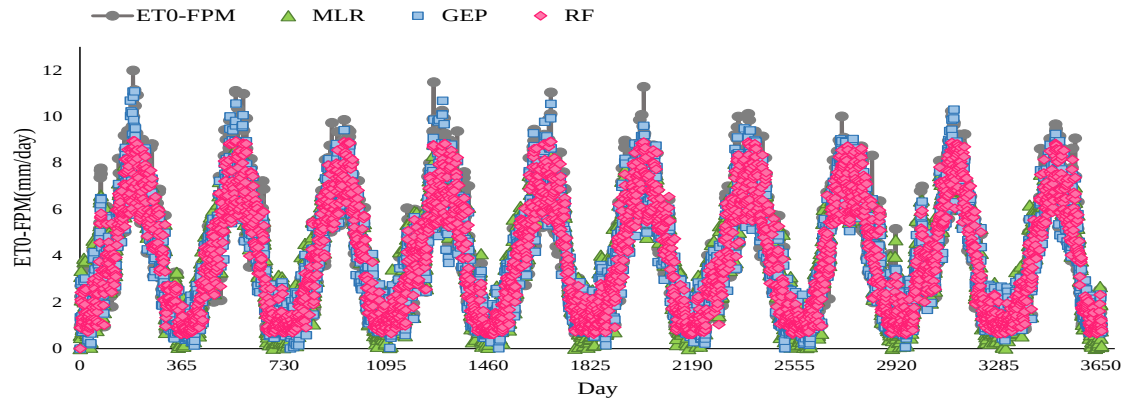
جدول ٤- روابط ریاضی مدل GEP

Table 4- Mathematical expressions of GEP model

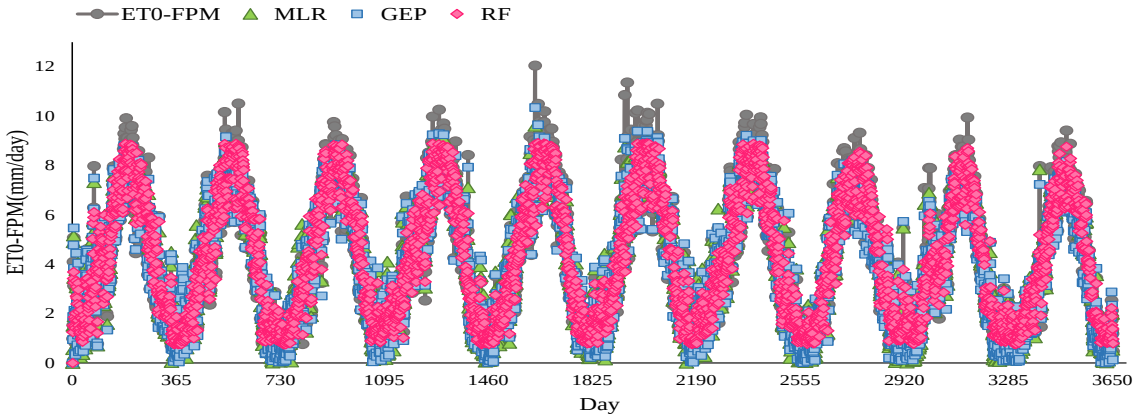
Station Name	Mathematical expressions
Khoramdare	$\sqrt[3]{(R_n - T_{mean} - 7.9)^2} + \frac{(U_2+R_n)^2 - RH_{mean}}{RH_{mean}+R_n+4} + (0.19)^3$
Khodabande	$\frac{\sqrt{R_n}^3}{RH_{mean}} - \frac{exp(U_2 - 4.75) + T_{mean}}{(R_n * U_2)3.42} + \frac{2U_2 + T_{mean} + 2.52}{23.14}$
Mahneshan	$1\text{-}\log\left(\frac{RH_{mean}-U_2}{\sqrt[3]{U_2}}\right) + \left[\sqrt{R_n} - \log(U_2)\right]^3 + \sqrt{\sqrt{({T_{mean}}^2 * U_2)} + 5.44} + R_n$
Zanjan	$\frac{((1.52 + U_2) * T_{mean} * \log(R_n))}{T_{mean} + RH_{mean} + 6.19} + \frac{U_2 - R_n - 6.86 - (RH_{mean} * R_n)}{U_2 * R_n * T_{mean} * RH_{mean}} + \sqrt{R_n}$



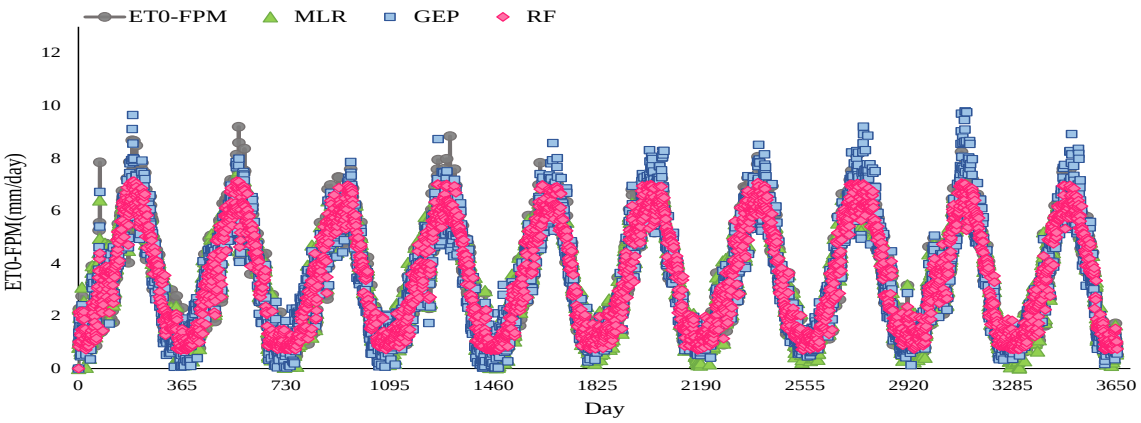
A)



B)



C)



D)

Fig. 2- Trends in daily reference evapotranspiration changes resulting from machine learning models with ET_0-FPM data

A) Khoramdare B) Khodabande C) Mahneshan D) Zanjan

شکل ۲-روند تغییرات تبخیر-تعرق مرجع روزانه مدل‌های یادگیری ماشین با داده‌های ET_0-FPM

A) خرمدره B)خدابنده C)ماه‌نشان D)زنجان

DOI: 10.22055/jise.2024.45305.2103

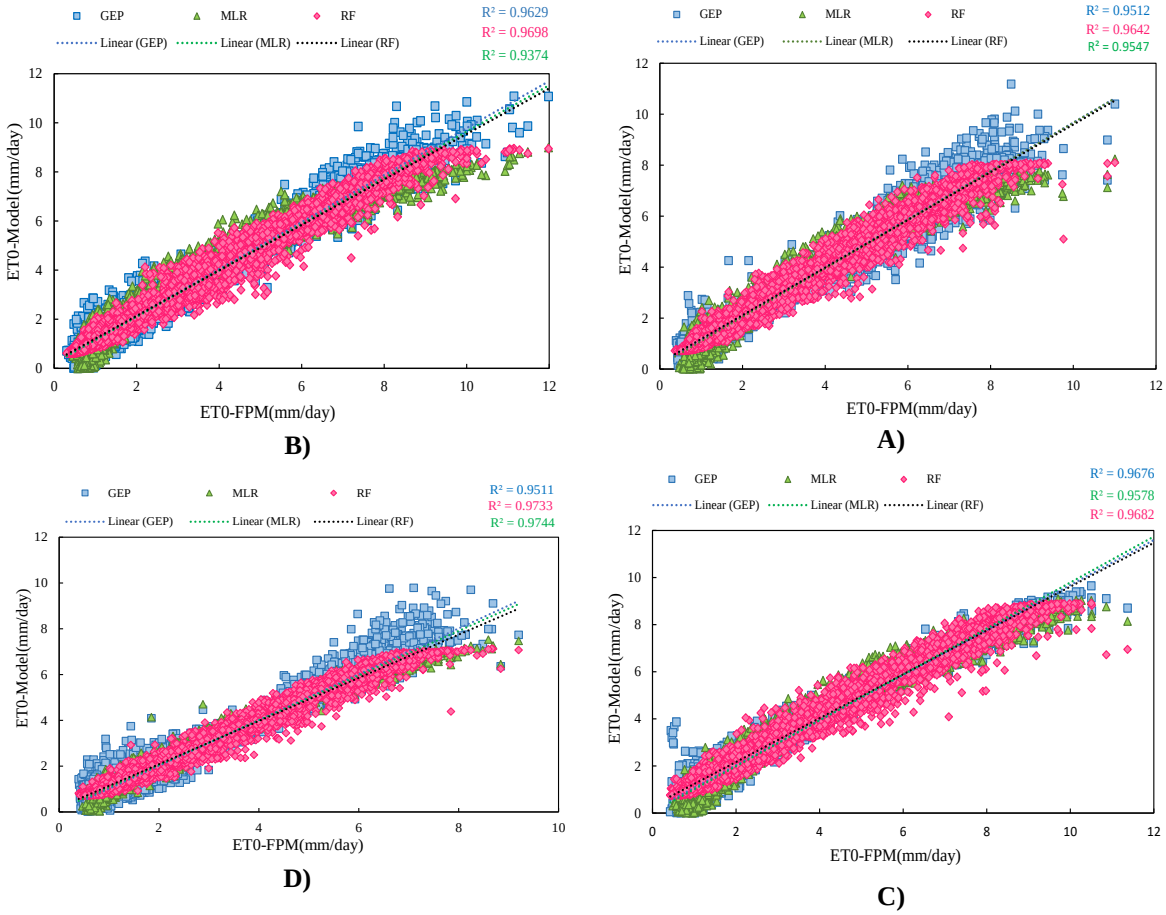
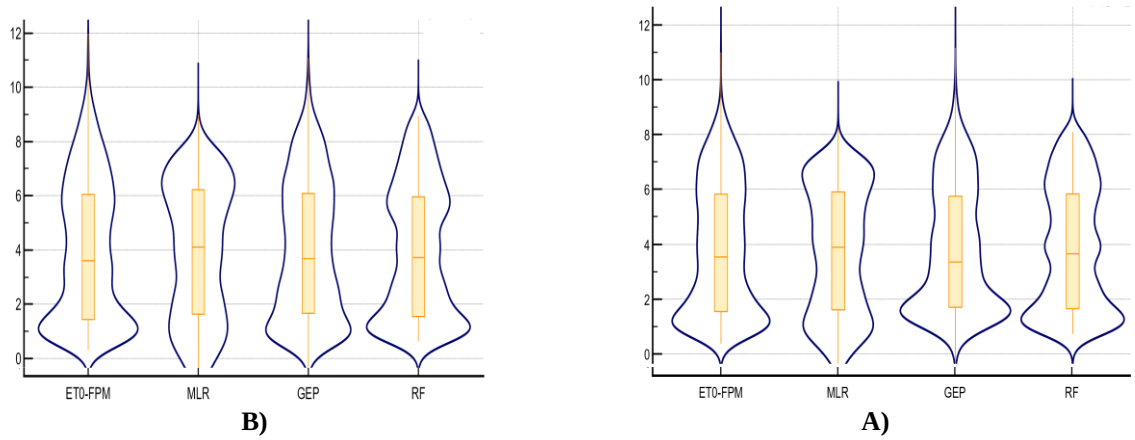


Fig. 3- Distribition of predicted values by machine learning models and obseverd ET0-FPM values
A) Khoramdare B) Khodabande C) Mahneshan D) Zanjan

شکل ۳- پراکنش مقادیر پیش‌بینی شده توسط مدل‌های یادگیری ماشین و مقادیر مشاهداتی ET_0-FPM
A) خرمدره B)خدابنده C)ماه‌نشان D)زنجان



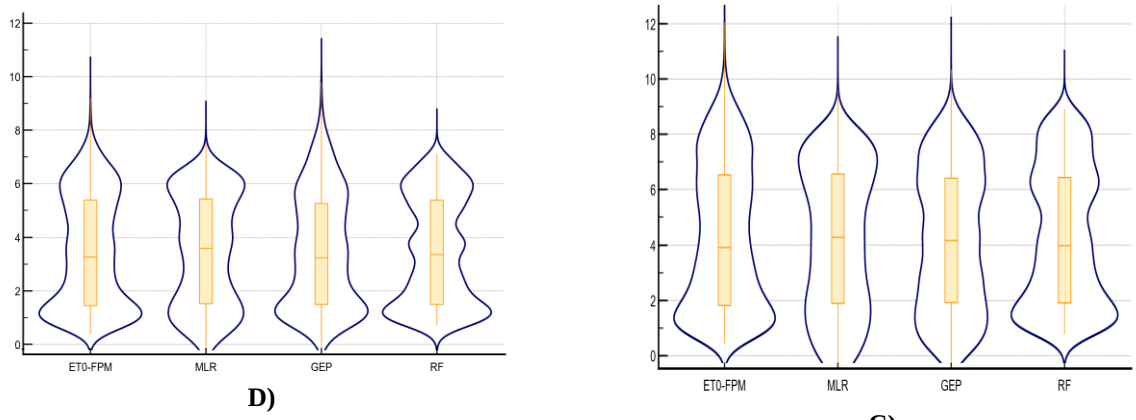


Fig. 4- Violin chart of machine learning models and ET0-FPM
A) Khoramdare B) Khodabande C) Mahneshan D) Zanjan
شکل ٤- نمودار ویولونی مدل‌های یادگیری ماشین و ET0-FPM A)خرمدره B)خدابنده C)ماه‌نشان D)زنجان

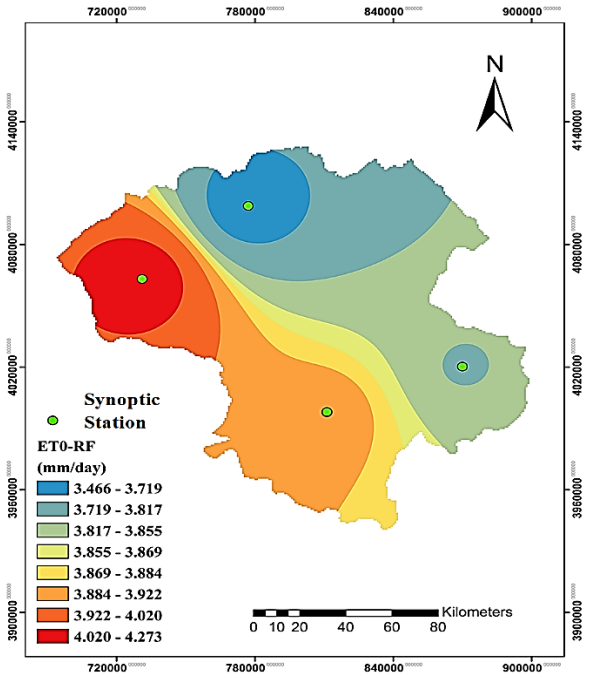


Fig. 5- Zoning of ET_0 resulting from Random Forest model in Zanjan province
شکل ۵- پهنه‌بندی تبخیر-تعرق مرجع مدل جنگل تصادفی در استان زنجان

با توجه به اینکه مدل جنگل تصادفی دارای دقت بالایی در پیش‌بینی تبخیر-تعرق مرجع روزانه است، از مقادیر این مدل برای پهنه‌بندی تبخیر-تعرق به روش وزن‌دهی‌معکوس‌فاصله (IDW) استفاده شد. طبق شکل (۵)، تبخیر-تعرق مرجع در استان زنجان در بازه‌ی(۳/۴-۴/۲) میلی‌متر بر روز است. در مناطق شمال‌غرب این استان بر مقدار آن افزوده می‌شود و بیشترین مقدار آن در بازه‌ی(۴-۴/۳) میلی‌متر بر روز در ایستگاه ماه‌نشان است که مقدار بارش و شرایط اقلیمی بر میزان تبخیر-تعرق اثر گذار است. در نهایت نتایج این تحقیق با نتایج Da Silva et al. (2019) که روش IDW برای ایجاد نقشه توزیع مکانی تبخیر-تعرق مرجع عملکرد خوبی دارد، در یک راستا است. به طور کلی از نتایج شکل‌ها می‌توان چنین استنباط کرد که مدل جنگل تصادفی عملکرد مناسبتری داشته است و می‌تواند مقدار تبخیر-تعرق روزانه منطقه مورد مطالعه را با استفاده از ورودی‌های منتخب با دقت نسبتا بالایی پیش‌بینی کند. همچنین نقشه ایجاد شده می‌تواند در حیطه مکانیابی بسیاری از پروژه‌های کشاورزی و منابع طبیعی و همچنین عمرانی در استان زنجان به عنوان یکی از نقشه‌های مهم مورد توجه قرار گیرد.

نتیجه‌گیری

با توجه به اهمیت پدیده تبخیر-تعرق مرجع در مطالعه‌های هیدرولوژیک و مدیریت بهینه منابع آبی، در تحقیق حاضر از

DOI: 10.22055/jise.2024.45305.2103

داده‌های سینوپتیک چهار ایستگاه استان زنجان به مدت ۱۰ سال شامل میانگین دما، میانگین رطوبت نسبی، سرعت باد و تابش خالص به عنوان ورودی مدل‌های رگرسیون خطی چندگانه(MLR)، برنامه‌ریزی بیان ژن(GEP) و جنگل تصادفی(RF)، برای تخمین تبخیر–تعرق مرجع روزانه استفاده گردید سپس مقادیر تخمینی به‌دست آمده از مدل‌ها با مقادیر روش فائو پنمن مانتیث مقایسه شدند. با توجه به مقادیر شاخص‌های آماری، نتایج نشان داد که هر سه مدل در تخمین تبخیر–تعرق مرجع روزانه از دقت بالا و خطای کمتری برخوردار هستند ولی مدل جنگل تصادفی(RF) با اختلاف کمی به‌عنوان مدل بهتر انتخاب شد نتایج حاصل از پهنه‌بندی به روش IDW این مدل نشان داد که در شمال‌غربی استان زنجان مخصوصا در ایستگاه ماه‌نشان مقدار تبخیر–تعرق بیشتر است.

تشکر و قدردانی

از اداره هواشناسی کشور برای در اختیار قرار دادن داده‌های لازم برای انجام این تحقیق تشکر و قدردانی می‌شود.

References

1-Aiyelokun, O. O., and Agbede, O. A., 2021. Development of random forest model as decision support tool in water resources management of Ogun headwater catchments. *Applied Water Science*, 11, pp.1-9.

2- Allahverdipour, P. and Sattari, M.T., 2023. Comparing the performance of the multiple linear regression classic method and modern data mining methods in annual rainfall modeling (Case study: Ahvaz city). *Water and Soil Management and Modelling*, *3*(2), pp.125-142. [In Persian]

3- Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9), p.D05109.

4-Almorox, J., Quej, V.H. and Martí, P., 2015. Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, *528*, pp.514-522.doi.org/10.1016/j.jhydrol.2015.06.057

5- Borrelli, A., De Falco, I., Della Cioppa, A., Nicodemi, M. and Trautteur, G., 2006. Performance of genetic programming to extract the trend in noisy data series. *Physica A: Statistical Mechanics and its Applications*, *370*(1), pp.104-108. Doi: 10.1016/j.physa.2006.04.025.

6- Breiman, L., 2001. Random forests. *Machine Learning*, *45*, pp.5-32.

7- Chen, H., Huang, J.J. and McBean, E., 2020. Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random Forest and support vector regression, for a cabbage farmland. *Agricultural Water Management*, *228*, p.105923.doi.org/10.1016/j.agwat.2019.105923

8- Da Silva Júnior, J. C., Medeiros, V., Garrozi, C., Montenegro, A., & Gonçalves, G. E. 2019. Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. *Computers and Electronics in Agriculture*, 166, 105017. doi.org/10.1016/j.compag.2019.105017

9- Faloye, O.T., Ajayi, A.E., Babalola, T., Omotehinse, A.O., Adeyeri, O.E., Adabembe, B.A., Ogunrinde, A.T., Okunola, A. and Fashina, A., 2023. Modelling Crop Evapotranspiration and Water Use Efficiency of Maize Using Artificial Neural Network and Linear Regression Models in Biochar and Inorganic Fertilizer-Amended Soil under Varying Water Applications. *Water*, 15(12), p.2294. https://doi.org/10.3390/w15122294

10- 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*, pp.163-173. doi.org/10.1016/j.agwat.2017.08.003

11- Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems?. *The Journal of Machine Learning Research*, *15*(1), pp.3133-3181.

12- Ferreira, C., 2001. Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.

13- Ghourdoyee Milan, S., Aryaazar, N., Javadi, S. and Razdar, B., 2020. Simulation of groundwater head using LS-SVM and comparison with ANN & MLR. *Hydrogeology*, *5*(1), pp.118-133. Doi.org/10.22034/HYDRO.2020.10455 (In Persian)

14- Heramb, P., Singh, P.K., Rao, K.R. and Subeesh, A., 2023. Modelling reference evapotranspiration using gene expression programming and artificial neural network at Pantnagar, India. *Information Processing in Agriculture*. doi.org/10.1016/j.inpa.2022.05.007

15- Herrera, S., Kotlarski, S., Soares, P.M., Cardoso, R.M., Jaczewski, A., Gutiérrez, J.M. and Maraun, D., 2019. Uncertainty in gridded precipitation products: Influence of station density, interpolation method and grid resolution. *International Journal of Climatology*, *39*(9), pp.3717-3729. doi.org/10.1002/joc.5878

16- Karami, M. and Asadi, M., 2016. Estimates and Zoning of reference evapotranspiration by FAO-Penman-Monteith (Case Study: North West of Iran). *International Journal of Scientific Research in Science, Engineering and Technology*, *2*(1), pp.210-216.

17- Kisi, O. and Guven, A., 2010. Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering*, *136*(10), pp.715-723.

18- Kotsiantis, S. and Pintelas, P., 2004. Combining bagging and boosting. *International Journal of Computational Intelligence*, *1*(4), pp.324-333.

19-Malekahmadi, H., Mirzania, E., Khosravi, S., and Ebrahim Zadeh, A., 2022. Daily evapotranspiration modeling using regression and intelligent models (Case Study). *Iranian Water Research Journal*, 16(1), pp.49-62. Doi.org/ 10.22034/IWRJ.2022.10026.2338 [In Persian]

20- Maroufpour, S., Maroufpoor, E. and Sanikhani, H., 2017. Monthly Evapotranspiration Modeling in the Absence of Climatic Data Using the Soft Computing Methods in West and Northwest of Iran. *Water Resources Engineering*, *10*(34), pp.37-50. Dor: 20.1001.1.20086377.1396.10.34.4.4 [In Persian]

21- Mattar, M.A. and Alazba, A.A., 2019. GEP and MLR approaches for the prediction of reference evapotranspiration. *Neural Computing and Applications*, *31*, pp.5843-5855.doi.org/10.1007/s00521-018-3410-8.

22- Mohammadrezapour O., 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(3), pp.135-150. (In Persian).

23- Poursalehi, F., KhasheiSiuki, A. and Hashemi, S.R., 2022. Investigating the performance of random forest algorithm in predicting water table fluctuations Compared with two models of decision tree and artificial neural network (Case study: unconfined aquifer of Birjand plain). *Iranian Journal of Ecohydrology*, *8*(4), pp.961-974.Dio.org/ 10.22059/IJE.2022.327263.1526 (In Persian).

24- Rezaei, H., Habibi Nokhandan, M., Ghalhari, F. and Abbas, G., 2021. Estimation and mapping evapotranspiration using climatic parameters in GIS: A Case Study in Khorasan Province. *Journal of Meteorology and Atmospheric Science*, *4*(2), pp.115-127.Doi.org/ 10.22034/JMAS.2022.336879.1170 (In Persian).

25-Sanjari, S., 2023. ARC GIS application guide, *Mehrgan Qalam Publications*, 263. (In Persian).

26- Sayyahi, F., Farzin, S. and Karami, H., 2021. Forecasting daily and monthly reference evapotranspiration in the Aidoghmoush basin using multilayer perceptron coupled with water wave optimization. *Complexity*, *2021*, pp.1-12.doi.org/10.1155/2021/6683759

27- Sharghi, T., Bari Abarghuei, H., Asadi, M.A. and Kousari, M.R., 2010. Estimation of reference evapotranspiration using FAO-Penman-Monteith method and its zonation in Yazd province. *Journal of Arid Biome*, *1*(1), pp.25-33. Dor: 20.1001.1.2008790.1389.1.1.3.7 (In Persian).

28- 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 & Drainage*, *13*(2), pp.378-388.Dor.org/ 20.1001.1.20087942.1398.13.2.10.8 (In Persian).

29- Swain, S., Patel, P. and Nandi, S., 2017, April. A multiple linear regression model for precipitation forecasting over Cuttack district, Odisha, India. In *2017 2nd International Conference for Convergence in Technology (I2CT)* (pp. 355-357). IEEE.doi.org/ 10.1109/I2CT.2017.8226150

30- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J. and Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science*, *8*, pp.1-12. doi.org/10.1007/s13201-018-0742-6

31- Wu, M., Feng, Q., Wen, X., Deo, R.C., Yin, Z., Yang, L. and Sheng, D., 2020. Random forest predictive model development with uncertainty analysis capability for the estimation of evapotranspiration in an arid oasis region. *Hydrology Research*, *51*(4), pp.648-665.doi.org/10.2166/nh.2020.012

32- Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J. and Wu, M., 2018. Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. *Water Resources Management*, *32*, pp.301-323. doi.org/10.1007/s11269-017-1811-6