

ارزیابی یادگیری ماشین و سنجش از دور در تخمین تبخیر و تعرق مرجع

جوانشیر عزیزی مبصر^۱، علی رسولزاده^۲، امین اکبری مجد^۳

۱-نویسنده مسئول، گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران،

رایانامه: ja_mobaser@uma.ac.ir

۲- گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران، رایانامه:

rasoulzadeh@uma.ac.ir

۳- گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران، رایانامه:

Amin.a.m.ce@gmail.com

چکیده

تبخیر و تعرق مرجع (ET_o) یکی از اساسی‌ترین متغیرها در تعیین نیاز آبی محصول و برنامه‌ریزی و طراحی سامانه‌های آبیاری است. مدل‌های یادگیری ماشین برای رفع محدودیت‌های مدل‌های تجربی و برآورد تبخیر تعرق (ET) توسعه داده شده‌اند. در سال‌های اخیر، محاسبه ET با استفاده از فناوری‌های سنجش از دور توسعه و بهبود یافته است. در این مطالعه برای تخمین تبخیر و تعرق مرجع در دشت اردبیل کارایی سه مدل جنگل تصادفی (RF)، رگرسیون خطی چندگانه (MLR) و ماشین بردار پشتیبان (SVM) بررسی شد. اطلاعات هواشناسی برای داده‌های مدل از سال ۱۳۸۵ تا ۱۴۰۱ از طریق سنجش از دور و ایستگاه‌های سینوپتیک دریافت شد. ET_o به عنوان پارامتر هدف، با استفاده از روش فائو پنمن مانیت در محدوده پنج ایستگاه سینوپتیک محاسبه شد. در مرحله ساخت و آموزش مدل‌ها، سری زمانی پارامترهای ورودی و هدف در محل چهار ایستگاه سینوپتیک و با تشکیل یک سری زمانی تصادفی و ترکیبی از کل داده‌ها انجام و در مرحله ارزیابی نهایی مدل، تنها از داده‌های ایستگاه پنجم استفاده شد. آماره‌های ارزیابی مورد استفاده، R^2 ، NSE و RMSE بودند. نتایج حاصل برای شاخص‌های آماری، برای مدل RF به ترتیب برابر با ۰/۷، ۰/۵۵۸ و ۱۰/۷۶، برای SVM برابر ۰/۷۱، -۱ و ۱۳/۶ و برای MLR ۰/۷۱، ۰/۶۸۸- و ۲۱ بود که با مقایسه نتایج، مدل RF نسبت به سایر مدل‌ها از دقت بالاتری برخوردار بود. مطالعه حاضر نشان داد که مدل جنگل تصادفی می‌تواند یک مدل مطمئن و نسبتاً دقیق برای پیش‌بینی ET_o با استفاده از داده‌های RS برای مناطق فاقد آمار باشد.

کلمات کلیدی: MLR، RF، SVM، تبخیر و تعرق مرجع، حوضه فاقد آمار، یادگیری ماشین.

۱. مقدمه

تبخیر و تعرق^۱ (ET) که یکی از اجزای اصلی چرخه هیدرولوژیکی است (Callañaupa Gutierrez et al., 2021; Duhan, Singh, & Arya, 2020; Eslamian, Gohari, Zareian, & Firoozfar, 2012; Mahesh Chand, Shivam, Sanjay, Vishnu, & Sompal, 2022; Singh, Satpute, Prasad, & Sharma, 2022) عوامل متعددی شامل متغیرهای هواشناسی، ویژگی‌های خاک، الگوی کشت و شیوه‌های مدیریت مزرعه (Allan, Pereira, & Smith, 1998; Córdova, Carrillo-Rojas, Crespo, Wilcox, & Céleri, 2015) است. تبخیر و تعرق مرجع (ET_o) پرکاربردترین متغیر برای تعیین نیاز آبی محصول، برنامه‌ریزی آبیاری، بیان آب هیدرولوژیکی، شبیه‌سازی عملکرد محصول و طراحی سیستم‌های آبیاری است (Sanjay, Mahesh Chand, & Sunil, 2021; Zhang et al., 2015). بنابراین، برآورد دقیق آن برای مدیریت آب آبیاری موجود، به ویژه در مناطق کم‌آب بسیار مهم خواهد بود.

ET_o را می‌توان با استفاده از روش‌های برگرفته از داده‌های مشاهداتی^۲ بر اساس رویکردهای موازنه انرژی و انتقال جرم و یا به‌طور غیرمستقیم با استفاده از تکنیک‌های تجربی محاسبه کرد. با این حال، روش‌های مستقیم در مقایسه با روش‌های غیرمستقیم، که فقط به داده‌های مشاهداتی خاص در منطقه مطالعاتی نیاز دارند، از نظر زمان و هزینه مقرون به صرفه هستند (Allen, Pereira, Raes, & Smith, 1998; Blaney, 1952; Monteith, 1965; Thornthwaite, 1948). روش فائونمن-ماتیت (PM) (Penman & Keen, 1948) در برابر داده‌های لایسمتر تحت شرایط آب و هوایی مختلف جهان اعتبارسنجی شده و نسبت به سایر مدل‌های تجربی برای تخمین ET_o مورد پذیرش‌تر است. (Suleiman & Hoogenboom, 2007) در مطالعه‌ای نشان دادند روش PM می‌تواند بازده آبیاری را در آب و هوای مرطوب گرجستان، در مناطق ساحلی و کوهستانی افزایش دهد. (Duhan et al., 2020) دریافتند که PM-ET_o ارتباط نزدیکی با تبخیر تحت (ET_{pan}) دارد و شاخص خوبی برای ET_{pan} در یک منطقه نیمه خشک در مرکز هند است. (Vishwakarma et al., 2021) گزارش دادند که روش PM در برابر ET_{pan} در بین ۳۰ مدل در مناطق مرطوب و نیمه گرمسیری هند عملکرد بهتری داشت. اشکال روش PM تعداد زیاد متغیرهای ورودی برای تخمین ET است (Fan et al., 2018; Feng, Peng, Cui, Gong, & Zhang, 2017).

مدل‌های یادگیری ماشین^۳ (ML) که اخیراً توسعه یافته‌اند در مقایسه با مدل‌های تجربی تحت شرایط محیطی مختلف دقیق‌تر هستند (Del Cerro, Subathra, Kumar, Verrastro, & George, 2021; Fakhar & Kaviani, 2024; Heravi & Zolfaghari, 2024; Mendes Reis et al., 2019). این مدل‌ها مقدار ET_o را به‌عنوان یک پدیده غیرخطی و پیچیده به راحتی شبیه‌سازی می‌کنند. مدل‌های ML زیرمجموعه‌ای از هوش مصنوعی (AI) هستند که از تجربه قبلی (آموزش داده‌ها) برای یادگیری و استفاده از آن برای اجرای وظایف جدید استفاده می‌کند. همان‌طور که هوش مصنوعی روابط غیرخطی بین متغیرهای ورودی و خروجی را ترسیم می‌کند، دقت پیش‌بینی را نیز بهبود می‌بخشد (Chia, Huang, & Koo, 2022; Fan et al., 2018). کاربرد مدل‌های ML برای پیش‌بینی ET در سال‌های اخیر به شدت افزایش یافته است (Chia, Huang, Koo, et al., 2022; Duhan et al., 2020; Feng et al., 2017). الگوریتم‌های مختلف ML که برای پیش‌بینی ET_o در سطح جهانی استفاده شده شامل شبکه عصبی فازی-عصبی تطبیقی^۴ (Keshtegar, Kisi, Ghohani Arab, & Zounemat-Kermani, 2018)، ماشین بردار حداقل مربعات پشتیبانی^۵ (Mendes Reis et al., 2019)، منطق فازی^۶ (Malik et al., 2022)، شبکه عصبی ادراک چند لایه^۷ (Seifi & Riahi, 2020)، ماشین بردار ارتباط^۸ (Bachour, Maslova, Ticlavilca, Walker, & McKee, 2016)، اسپلاین رگرسیون چند متغیره^۹ (Mehdizadeh, Behmanesh, & Khalili, 2017) و بسیاری موارد دیگر است.

طبق تحقیق (Mendes Reis et al., 2019)، مدل‌های شبکه عصبی مصنوعی تخمین‌های بهتری از ET_o در مقایسه با روش‌های تجربی مانند روش هارگریوز-سامانی در منطقه نیمه خشک برزیل تنها با استفاده از داده‌های دما ارائه می‌دهند. (Seifi & Riahi, 2020) در مطالعه‌ای نشان دادند مدل ماشین بردار پشتیبان - حداقل مربعات^{۱۰} در مقایسه با مدل‌های ANN و سیستم

استنتاج عصبی فازی تطبیقی (ANFIS) برای پیش‌بینی ETo روزانه در شرایط خشک (ایستگاه زاهدان) ایران کارآمدتر است. با توجه به تحقیق (Dou & Yang, 2018)، ANN بهتر از ANFIS و مدل‌های ماشین یادگیری افراطی برای پیش‌بینی ETo در اکوسیستم‌های زراعی و علفزار عمل می‌کند. (Nema, Khare, & Chandniha, 2017) در مطالعه‌ای توانایی شبکه عصبی مصنوعی برای پیش‌بینی ET ماهانه را در منطقه آب و هوایی نیمه مرطوب دهرادون، هند را بررسی کردند. (Granata, 2019) نشان دادند درخت رگرسیون M5P به عنوان بهترین مدل بعد از SVR و Random Forest در آب و هوای مرطوب نیمه گرمسیری فلوریدا، ایالات متحده است.

در سال‌های اخیر، تخمین ET با استفاده از فناوری سنجش از دور، عمدتاً در زمینه‌های کشاورزی و هیدرولوژیکی، بهبود یافته است (Fakhar & Kaviani, 2024; Negin & Gholmohamadi, 2023; Niasati, Ebadi, & Kiani, 2021; Siasar & Dindarlou, 2020; Rodrigues & Braga, 2021; Rajput et al., 2022; Elbeltagi et al., 2023). جداگانه نشان دادند خطای تخمین، پس از اعمال تصحیح بایاس در داده‌های NASA POWER برای تخمین ETo کاهش می‌یابد و محصولات تحلیل مجدد NASA POWER انتخاب مناسبی برای تخمین ETo منطقه‌ای هستند. همچنین در مناطقی که پارامترهای هواشناسی در دسترس نیستند، کاربرد سنجش از دور در حوضه‌های فاقد آمار بسیار کمک‌کننده خواهد بود. در این مطالعه با بررسی صورت گرفته مشاهده شد که در محدوده مطالعاتی مورد نظر، تنها پنج ایستگاه هواشناسی وجود دارد که سه ایستگاه با نام‌های اردبیل، فرودگاه اردبیل و نمین در داخل مرز دشت اردبیل و دو ایستگاه دیگر با نام‌های نیر و سرعین در ناحیه غربی دشت و خارج از مرز آن، ولی در همسایگی و مجاورت آن واقع شده‌اند. در بین ایستگاه‌ها به جز ایستگاه اردبیل، سایر ایستگاه‌ها به طور متوسط از سال ۱۳۸۵ به بعد راه‌اندازی شده‌اند که نشان از عدم وجود داده‌های هواشناسی کافی در محدوده مطالعاتی داشت. همچنین موقعیت جغرافیایی ایستگاه‌ها در این منطقه به‌صورتی است که هیچ اطلاعاتی هواشناسی را برای نواحی شمالی و جنوبی دشت ارائه نمی‌دهند که این موضوع با توجه به وسعت دشت اردبیل، نشان از عدم وجود اطلاعات کافی در سطح آن بود. پس در این مطالعه تلاش کردیم تا برای تخمین ETo در منطقه دشت اردبیل ضمن بررسی کارایی سه مدل جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون خطی چندگانه با استفاده از داده‌های سنجش از دور، امکان تخمین ETo در مناطق فاقد آمار نیز بررسی گردد.

۲. مواد و روش‌ها

۲.۱. منطقه مورد مطالعه

این تحقیق در محدوده‌ی دشت اردبیل انجام شده است. دشت اردبیل در منطقه‌ی مرکزی استان اردبیل و در شمال غرب ایران قرار دارد. این دشت با مساحتی در حدود ۱۲۹۱ کیلومتر مربع، در محدوده‌ی مختصات مابین ۳۱ درجه و ۳۸ دقیقه تا ۳۸ درجه شمالی و ۴۸ درجه و ۹ دقیقه تا ۴۸ درجه و ۳۹ دقیقه شرقی واقع شده است. این منطقه از غرب به کوه سبلان و شرق به کوه‌های بزغوش و طالش محدود شده است. متوسط بارش سالانه‌ی آن در ایستگاه هواشناسی اردبیل حدود ۳۲۷ میلی‌متر ثبت شده است (Nouri-Khajehbolagh, Khaledian, & Kavooosi-Kalashami, 2020). در این دشت، اردیبهشت، پربارش‌ترین ماه و میانگین دمای سالانه ۷ درجه سلسیوس است (شکل ۱).

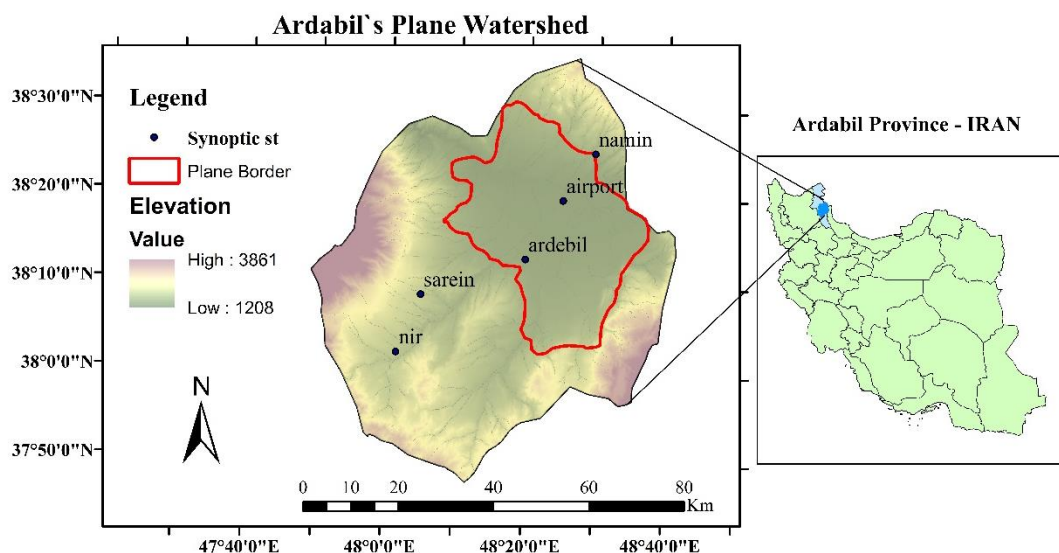


Figure 1. Scope and location of the study area

۲.۲. داده‌های مورد استفاده

در این مطالعه و در بازه زمانی ۱۳۸۵ تا ۱۴۰۱، برای پارامترهای ورودی از داده‌های سنجنش از دور و برای پارامتر خروجی (پارامتر هدف) از داده‌های مربوط به تبخیر و تعرق مرجع با گام‌های ده روزه (دهه اول، دوم و سوم هر ماه) به روش ارائه شده در نشریه FAO56 (پنمن مانیت) استفاده شد (Abdzaad Gohari, Nik Akhtar, Ebrahimipak, & Tafteh, 2023). داده‌های سنجنش از دور از سامانه GEE¹¹ شامل: متوسط حجم آب موجود در خاک (vsw)، متوسط دمای اتمسفر (T2)، حجم تبخیر آب (pe)، شاخص سطح برگ کم تراکم (LAIl)، شاخص سطح برگ پرتراکم (LAIh) و میانگین دمای خاک (st1) و داده‌های مشاهداتی مورد نیاز برای محاسبه تبخیر و تعرق مرجع شامل: سرعت باد در ارتفاع دو متری، دمای حداکثر، حداقل و متوسط، رطوبت نسبی، در محل پنج ایستگاه هواشناسی، به نام‌های ایستگاه سینوپتیک نیر، سرعین، اردبیل، فرودگاه اردبیل و نمین که در محدوده و یا نزدیک به محدوده دشت بودند، محاسبه شد (شکل ۱).

شناسایی دقیق و انتخاب داده‌های ورودی کلید توسعه یک مدل ML است. بنابراین، همیشه باید داده‌های ورودی را از نظر در دسترس بودن و ارتباط در نظر گرفت. متغیرهای موثر بر ET عبارتند از عوامل آب و هوایی (حداکثر و حداقل دما، سرعت باد، رطوبت نسبی، مدت و شدت تابش خورشید)، خواص خاک (رطوبت خاک، نوع محصول، شیوه‌های کشاورزی، بافت خاک)، و ویژگی‌های گیاه (نوع گیاه، مرحله رشد، شکل برگ، کل پتانسیل آب درون برگ) (Allen et al., 1998). داده‌های اولیه را می‌توان عمدتاً از دو منبع به دست آورد، یعنی داده‌های مشاهداتی و داده‌های RS (Maeda, Wiberg, & Pellikka, 2011). داده‌های RS دریافتی از سنجنده ERA5 (جدول ۱) که در این مطالعه از آن‌ها استفاده شد، با در نظر گرفتن عوامل مرتبط با تبخیر و تعرق شامل متوسط حجم آب موجود در خاک تا عمق ۷ سانتی‌متر از زمین (volumetric soil water)، متوسط دمای اتمسفر تا ارتفاع ۲ متر از سطح زمین (temperature)، میزان حجم تبخیر آب از سطح زمین (reference evaporation sum)، شاخص سطح برگ پرتراکم (leaf_area_index_low_vegetation)، شاخص سطح برگ کم تراکم (leaf_area_index_high_vegetation)، میانگین دمای خاک تا عمق ۷ سانتی‌متر از سطح زمین (soil temperature) بود به کمک پلت فرم GEE دریافت شد (Muñoz Sabater et al., 2021).

Table 1. Meteorological data received from the Google Earth Engine platform

Variable	index	Unit*	Dataset Provider	Resolution (m)	Time Step
volumetric_soil_water_layer_1	vsw	(Volume fraction)	ERA5-Land Daily	11132	Daily
temperature_2m	T2	(°C)			
Reference_evaporation_sum	pe	(m)			
leaf_area_index_low_vegetation	LAIl	(Area fraction)			
leaf_area_index_high_vegetation	LAIh	(Area fraction)			
soil_temperature_level_1	st1	(°C)			

۲.۳. روش مطالعه

در دهه‌های اخیر مدل‌های متعددی مبتنی بر RS برای تخمین ET در مقیاس منطقه‌ای توسعه داده شده‌اند. مدل‌ها براساس فرضیه‌های نظری مختلف مانند تعادل انرژی و نظریه شباهت مونین-اوپوخوف^{۱۲} است. این مدل‌ها را می‌توان به چهار گروه تقسیم کرد.

گروه اول شامل مدل‌های تعادل انرژی سطحی مانند SEB (Ball, Woodrow, & Berry, 1987) و SEBAL (Bastiaanssen, Menenti, Feddes, & Holtslag, 1998) است که براساس تعادل انرژی سطحی توسعه یافته‌اند و برای برآورد ET قابل قبول هستند. با این حال، وابستگی آن‌ها به مقاومت‌های آیرودینامیکی و سطحی، به برخی متغیرهای مهم مانند سرعت باد یا کمبود فشار بخار (VPD) نیاز دارد که معمولاً از طریق داده‌های RS قابل بازیابی نیستند (Amani & Shafizadeh-Moghadam, 2023).

گروه دوم شامل مدل‌های مبتنی بر پنمن مانیت (PM) و پرستلی تیلور (Cleugh, Leuning, Mu, & Running, 2007; Leuning, Zhang, Rajaud, Cleugh, & Tu, 2008; Mu, Heinsch, Zhao, & Running, 2007; Mu, Zhao, & Running, 2011; Peng et al., 2019) هستند. معادله PM همه عوامل اصلی را که بر ET تاثیر می‌گذارند را در نظر می‌گیرد، اما برآورد مقاومت سطحی (Gs)، که یک پارامتر کلیدی در روش PM برای تخمین ET است، همچنان به عنوان یک چالش باقی مانده است (Jiang, Kang, Tong, & Li, 2016). به همین دلیل، مدل‌های پرستلی-تیلور مقاومت آیرودینامیکی و سطحی را نادیده می‌گیرند (Amani & Shafizadeh-Moghadam, 2023).

گروه سوم، شامل مدل‌هایی که دمای سطح زمین (LST) را برای محاسبه ET استفاده می‌کنند. از جمله این مدل‌ها TEFM (Moran, Carlson, Capehart, & Gillies, 1995; Goward, Cruickshanks, & Hope, 1985; Price, 1984)، TIM (Nishida, Nemani, Running, & Glassy, 2003)، S-SEBI (Roerink, Su, & Menenti, 2000)، TTEFM (et al., 1996) هستند. این مدل‌ها به اندازه‌گیری دمای سطح زمین برای تخمین ET تکیه می‌کنند، که در تخمین ET با استفاده از داده‌های ماهواره‌ای که اغلب تحت تاثیر پوشش ابرها و مکان‌هایی که شار گرما به‌طور قابل توجهی کوچک است، مانند مناطقی با پوشش گیاهی متراکم، بی‌اثر هستند (Chen & Liu, 2020).

گروه چهارم مدل‌های مبتنی بر RS برای تخمین ET، مدل‌های مبتنی بر داده را به دو دسته تجربی یا نیمه تجربی تقسیم می‌کند (M. Jung, Reichstein, & Bondeau, 2009; Martin Jung et al., 2010; Tramontana et al., 2016). به‌طور کلی مطالعات قبلی نشان داده‌اند که مدل‌های فیزیکی و روش‌های تحلیلی مانند سنجش از دور، به دلیل ناهمگونی محیط‌های زیرسطحی، شرایط اقلیمی پیچیده در نزدیکی سطح، تنوع دینامیکی فرآیند انتقال آب و حرارت، از رویکردهای تجربی و آماری متعارف، دقیق‌تر نیستند (Kalma, McVicar, & McCabe, 2008).

ET به‌طور قابل توجهی با تغییرات مکان و زمان تغییر می‌کند. با این حال، روش‌های فیزیکی به تنهایی نمی‌توانند همه این عوامل را در نظر بگیرند. استفاده از مدل‌های داده‌مبنای^{۱۳} برای تخمین ET در سال‌های اخیر به دلیل توانایی آن‌ها در شناسایی روابط پیچیده به سرعت افزایش یافته است. روش‌های یادگیری ماشینی (ML) به دانش یا فرضیه‌های قبلی نیاز ندارند و روابط پیچیده را مدیریت می‌کنند. بسیاری از محققان از مدل‌های یادگیری ماشین برای ادغام مشاهدات زمینی، RS و داده‌های آب و

هوایی برای تخمین ET محلی و جهانی استفاده کرده‌اند (Amani & Shafizadeh-Moghadam, 2023). با توجه به بررسی منابع، در این مطالعه و در محدوده مورد مطالعه به منظور ارزیابی کارایی مدل‌های یادگیری ماشین در ترکیب با داده‌های سنجش از دور برای برآورد میزان تبخیر و تعرق مرجع، از سه مدل رگرسیون خطی چندگانه و جنگل تصادفی و ماشین بردار پشتیبان استفاده شد. در مرحله اول برای همگام‌سازی داده‌های روزانه ماهواره‌ای با داده‌های محاسبه شده برای تبخیر و تعرق، مقدار پارامترهای هواشناسی در بازه‌های ده روزه (دهه اول، دوم و سوم هر ماه) محاسبه شد. دلیل اصلی این تقسیم‌بندی ضمن کاهش حجم محاسبات، کاهش خطاهای ناشی از نقص داده‌های روزانه در ایستگاه‌های سینوپتیک بود.

۲.۴. مدل‌های یادگیری ماشین

برای تعریف و ساخت مدل‌ها، تنها از داده‌های چهار ایستگاه هواشناسی شامل ایستگاه‌های نیر، سرعین، اردبیل و فرودگاه اردبیل با ساخت یک سری‌زمانی ترکیبی - تصادفی که از اطلاعات آن مطابق مرحله قبل به دست آمده بود، استفاده شد و از ایستگاه نمین که از داده‌های آن در تعریف و ساخت مدل‌ها استفاده نشده بود، برای ارزیابی نهایی مدل‌ها استفاده شد. در مرحله ساخت و آموزش مدل از ۷۰ درصد داده‌ها سری‌زمانی ترکیبی - تصادفی برای آموزش و از ۳۰ درصد باقی‌مانده برای اعتبار سنجی و تست استفاده گردید، در (شکل ۲) فرآیند کلی مطالعه نمایش داده شده است.

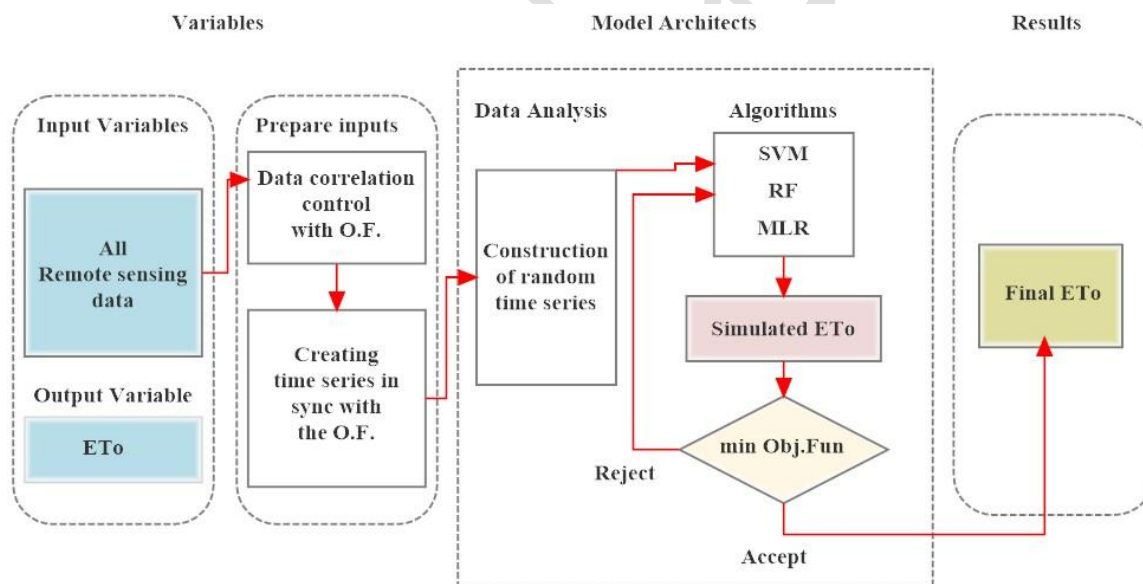


Figure 2. The flowchart of the modeling process

۲.۵. مدل جنگل تصادفی (RF)

مدل RF یک طبقه‌بندی کننده تخصصی (رگرسیون) است که از مجموعه‌ای از درخت‌های تصمیم تشکیل شده و همان‌طور که توسط (Breiman, 2001) پیشنهاد شده است از بردارهای تصادفی مستقل و با توزیع یکسان تشکیل می‌یابد. بریمن در سال ۲۰۰۱ مدل‌های Bagging را به جنگل تصادفی ارتقا داد. این تکنیک، جزئی از یادگیری گروهی است که به طور قابل توجهی دقت یادگیری را برای وظایف طبقه‌بندی و رگرسیون بهبود بخشیده است. هر درخت تصمیم در RF با استفاده از یک زیر مجموعه تصادفی از داده‌های آموزشی ساخته و آموزش داده می‌شود. در نتیجه، درختان از کل مجموعه داده استفاده نمی‌کنند و در هر گره، ویژگی بهینه از مجموعه‌ای از ویژگی‌های موجود انتخاب می‌شود. مدل جنگل تصادفی از دو فرآیند استفاده می‌کند: تعداد درختان رگرسیون برای رشد (ntree) و تعداد ویژگی‌هایی که به‌طور تصادفی در هر تقسیم (mtry) نمونه برداری شده‌اند (Breiman, 2001; Dong, Yu, Cao, Shi, & Ma, 2020). در نتیجه این روش، هنگامی که همه درختان آموزش داده

شدند - بدون اعمال هیچ گونه اصلاحی (هرس) - هر مورد جدید توسط همه درختان ارزیابی می‌شود و پیش‌بینی نهایی با میانگین‌گیری نتایج بدست می‌آید. مدل جنگل تصادفی پیش‌بینی‌ها را با دقت بالا ارائه می‌دهد و خطر بیش از حد برازش را به حداقل می‌رساند و آن را برای مدل‌سازی ویژگی‌های هیدرولوژیکی مناسب می‌کند (Breiman, 2001; Ruiz-Álvarez, 2021; Gomariz-Castillo, & Alonso-Sarria, 2021). درختان تصمیم به عنوان یک مثال ایده‌آل برای مدل‌سازی عمل می‌کنند زیرا می‌توانند به طور موثر ساختارهای تعامل پیچیده موجود در داده‌ها را به تصویر بکشند. در این مدل‌ها، در حالی که پیش‌بینی‌های تک‌تک درختان به نوبت در مجموعه آموزشی بسیار حساس هستند، استفاده از چندین درخت به طور متوسط این حساسیت را کاهش می‌دهد، مشروط بر اینکه درختان همبستگی نداشته باشند (Wang et al., 2017). هنگامی که مدل آموزش داده شد، پیش‌بینی را می‌توان به صورت زیر به دست آورد (Ruiz-Álvarez et al., 2021):

$$\hat{Y}_i = \frac{1}{M} \sum_{m=1}^M T_m(f_i) \quad (1)$$

که در آن M تعداد درخت‌ها، T_m درخت تصمیم منفرد و f_i بردار پیش‌بینی کننده است. در مطالعه حاضر از مدل RF در تخمین ETo با استفاده از متغیرهای آب و هوایی استفاده شده است.

۲.۶. ماشین بردار پشتیبان (SVM)

SVM یک روش جدید یادگیری ماشین مبتنی بر تئوری یادگیری آماری است و به عنوان یکی از رویکردهای محاسباتی توسعه یافته، توسط ولادیمیر وپنیک^{۱۴} (Vapnik, 2013) طبقه بندی می‌شود. براساس اصل کمینه‌سازی ریسک ساختاری^{۱۵}، SVM می‌تواند قوانین تصمیم‌گیری را دریافت کند و به خطاهای کوچکی برای مجموعه تست‌های مستقل دست یابد و از این رو می‌تواند مسائل یادگیری را به طور موثر حل کند (Deng, Wu, & Shao, 2008). اخیراً از SVM برای حل مسائلی مانند توابع غیرخطی، حداقل محلی^{۱۶} و بعد بالا^{۱۷} استفاده شده است. در بسیاری از کاربردهای عملی، SVM می‌تواند از دقت بالاتری برای پیش‌بینی بلندمدت در مقایسه با سایر روش‌های محاسباتی اطمینان حاصل کند (Deris, Zain, & Sallehuddin, 2011). SVM مبتنی بر مفهوم سطوح تصمیم^{۱۸} است که مرزهای تصمیم را مشخص می‌کند. SVM با استفاده از یک مدل خطی، برای پیاده سازی مرزهای کلاس غیرخطی، از طریق برخی از بردارهای ورودی، یک ابرصفحه^{۱۹} ایجاد می‌کند (Samanta, Al-Balushi, & Al-Araimi, 2003). در SVM، مقداری وابستگی مجهول و غیرخطی برای مثال در تابع هدف $\gamma = f(x)$ بین برخی از بردارهای ورودی با بعد بالا (x) و خروجی اسکالر γ (یا بردار خروجی (y) مانند SVM چند کلاسه) وجود دارد و هیچ اطلاعاتی در مورد توابع احتمال مشترک اساسی وجود ندارد. مجموعه داده‌های آموزشی $D = \{(x_i, y_i) \in X \times Y, I = \{1, 1\}\}$ که در آن ۱ مخفف جفت داده‌های آموزشی است و به اندازه مجموعه داده‌های آموزشی D است. اغلب y_i به صورت d_i بیان می‌شود، جایی که d مخفف مقدار مورد نظر است. بنابراین، SVM بخشی از تکنیک‌های یادگیری با نظارت است (Deris et al., 2011).

به صورت معمول، وقتی که SVM با داده‌های خطی روبرو می‌گردد، از فناوری به نام توابع هسته کرنل^{۲۰} استفاده می‌کند. تابع هسته می‌تواند داده‌ها را از محیطی با ابعاد کمتر به محیطی با بعد بالاتر منتقل کند. از بین توابع هسته مختلفی که پیشنهاد شده است، ساختار چهار نوع هسته شامل توابع چندجمله‌ای، پایه شعاعی (RBF)، خطی و سیگموئیدی جز پرکاربردترین توابع هسته هستند (Bui, Pradhan, Löfman, Revhaug, & Dick, 2012).

۲.۷. رگرسیون خطی چندگانه

یک روش آماری شناخته شده به نام تحلیل رگرسیون خطی چندگانه (MLR) مجموعه‌ای از تخمین‌های پارامتر را برای پیش‌بینی نتیجه یک متغیر وابسته ترکیب می‌کند. برای نشان دادن رابطه خطی بین متغیرهای مستقل و وابسته از رگرسیون خطی

چندگانه استفاده می‌شود. با توجه به وجود چندین متغیر مستقل، رگرسیون چندگانه به عنوان بسط رگرسیون حداقل مربعات ساده عمل می‌کند. روش‌های مختلفی برای تحلیل رگرسیون وجود دارد که رگرسیون خطی رایج‌ترین آنها است. این روش فرض می‌کند که متغیرها به صورت خطی با یکدیگر مرتبط هستند. به طور کلی، دو نوع رگرسیون خطی وجود دارد: رگرسیون خطی ساده و رگرسیون خطی چندگانه، که هر کدام از تکنیک‌های مختلفی برای محاسبه و به حداقل رساندن خطا استفاده می‌کنند. معیار ارزیابی که در مدل رگرسیون خطی چندگانه استفاده می‌شود، حداقل کردن مجموع مربعات خطا است. چون که میانگین مقدارهای خطا صفر در نظر گرفته شده، پس زمانی مجموع مربعات خطا، حداقل ممکن را خواهند داشت که توزیع داده‌ها نرمال باشند. در نتیجه، نرمال بودن داده‌های مربوط به متغیر وابسته یا باقیمانده‌ها یکی از فرضیه‌های مهم برای مدل رگرسیونی خطی است (Fattahi dolatabadi, Babazadeh, Najafi, & Sedghi, 2018; Gokcekus, Kassem, & Woyea, 2015; Piña-Monarez & Ortiz-Yañez, 2015; Krzywinski & Altman, 2015; Nolan, Fienen, & Lorenz, 2015). این رابطه معمولاً به صورت زیر تعریف می‌شود:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon \quad (2)$$

که برای n داده مشاهداتی Y_i متغیر وابسته، x_i متغیر مستقل، β_0 ضریب ثابت، β_p ضریب متغیر مستقل، ε خطای مدل است.

۲.۸. معیارهای ارزیابی

آماره‌های ارزیابی مورد استفاده در این مطالعه RMSE (Hyndman & Koehler, 2006)، R^2 (N. Draper, 1998; M. Nash & Sutcliffe, 1970) NSE، (H. Glantz, 1990; Steel & Torrie, 1960) است.

$$RMSE = \sqrt{MSE}; MSE = \frac{1}{n} \sum_{i=1}^n (Y_o - Y_m)^2 \quad (3)$$

$$R^2 = 1 - \frac{RSS}{TSS}; (RSS: \sum_{i=1}^n (Y_o - Y_m)^2, TSS: \sum_{i=1}^n (Y_o - \bar{Y}_o)^2) \quad (4)$$

$$NSE = 1 - \frac{\sum_{t=1}^T (Y_o - Y_m)^2}{\sum_{t=1}^T (Y_o - \bar{Y}_o)^2} \quad (5)$$

در تمامی روابط: $\bar{Y}_o$ میانگین داده‌ای مشاهداتی، Y_m : خروجی مدل، Y_o : داده مشاهداتی، n : تعداد داده‌ها، cov : کوواریانس، σ_{Y_o} & σ_{Y_m} : انحراف معیار پارامترهای است. در RMSE این مقدار هرچه کمتر باشد، مدل بهتر است. یک مدل کامل (یک مدل فرضی که همیشه مقدار دقیق مورد انتظار را پیش‌بینی می‌کند) دارای مقدار میانگین مربعات خطای ریشه صفر خواهد بود (Hyndman & Koehler, 2006). پارامتر R^2 مقداری بین صفر و یک (۰٪ تا ۱۰۰٪) را است، عدد صفر بیان‌کننده‌ی عدم ارتباط مابین متغیرهای وابسته و مستقل در اطراف میانگین است و عدد یک نشان می‌دهد که مدل، همه‌ی تغییرپذیری داده‌های پاسخ در اطراف میانگین را پوشش می‌دهد (J. V. Draper, Kaber, & Usher, 1998; P. O. Glantz et al., 1993). برای شرایط ایجاد یک مدل کامل با واریانس خطای تخمینی صفر، ضریب کارایی نش-ساتکلیف برابر با ۱ است، (NSE = 1). در مدلی که واریانس خطای تخمینی، برابر با واریانس سری‌های زمانی مشاهده‌شده باشد، ضریب کارایی نش-ساتکلیف برابر با صفر خواهد بود، (NSE = 0). در واقع، NSE = 0 نشان می‌دهد که مدل، دارای توان پیش‌بینی مشابه، با میانگین سری‌های زمانی، از نظر مجموع مجذور خطا است بود (Forner, Abate, Mengoli, Palù, & Gussetti, 2015; McCuen, Knight, & Cutter, 2006).

۳. نتایج و بحث

۳.۱. جمع آوری داده‌ها

در گام اول داده‌های ورودی از طریق سنجش از دور شامل پارامترهای متوسط حجم آب موجود در خاک (vsw)، متوسط دمای اتمسفر (T2)، حجم تبخیر آب (pe)، شاخص سطح برگ کم‌تراکم (LAIl)، شاخص سطح برگ پرتراکم (LAIh) و میانگین دمای خاک (st1)، همچنین داده خروجی مربوط به پارامتر تبخیر و تعرق مرجع (ETo) در محل پنج ایستگاه هواشناسی شامل ایستگاه‌های نیر، سرعین، اردبیل، فرودگاه اردبیل و نمین در حد فاصل سال‌های ۱۳۸۵ تا ۱۴۰۱ و با گام‌های زمانی ده روزه (شامل ۶۱۲ گام)، دریافت و یا محاسبه شد که آمار توصیفی آن‌ها برای کل طول دوره آماری به شرح جدول‌های (۲)، (۳)، (۴)، (۵) و (۶) است.

Table 2. Descriptive Statistics of Ardabil Synoptic Station

variable	N	Mean	Minimum	Median	Maximum
vsw	612	0.36795	0.208	0.3929	0.4713
T2	612	9.714	-10.111	10.721	24.124
pe	612	-3.5291	-8.4	-3.5833	0
LAIl	612	1.0805	0.897	1.024	1.442
LAIh	612	2.1648	2.105	2.1434	2.2823
St1	612	11.674	-0.246	11.733	26.565
ETo	612	29.048	5.55	26.205	65.89

Table 3. Descriptive Statistics of Nir Synoptic Station

variable	N	Mean	Minimum	Median	Maximum
vsw	506	0.3177	0.1232	0.34005	0.417
T2	506	8.327	-11.832	8.966	23.529
pe	506	-3.405	-8.167	-3.5	0
LAIl	506	1.0657	0.8982	0.9868	1.4572
LAIh	506	2.3852	2.359	2.3789	2.4366
St1	506	10.032	-1.189	9.984	25.05
ETo	506	28.281	5.3	27.14	71.35

Table 4. Descriptive Statistics of Sarein Synoptic Station

variable	N	Mean	Minimum	Median	Maximum
vsw	609	0.41393	0.243	0.44067	0.4952
T2	609	8.12	-11.801	9.134	22.949
pe	609	-3.9162	-8.8	-4	0
LAIl	609	1.2273	1.081	1.177	1.5338
LAIh	609	1.9761	1.88	1.944	2.1795
St1	609	10.074	-0.803	9.97	24.217
ETo	609	26.526	5.61	25.69	62.88

Table 5. Descriptive Statistics of Airport Synoptic Station

variable	N	Mean	Minimum	Median	Maximum
vsw	609	0.3101	0.1307	0.3416	0.4164
T2	609	9.192	-10.551	10.087	23.78
pe	609	-2.8318	-7.4	-2.7	0
LAIl	609	0.9829	0.692	0.9422	1.4914
LAIh	609	2.457	2.444	2.455	2.4825
St1	609	11.909	-0.059	11.839	26.897
ETo	609	29.679	5.27	26.9	77.58

Table 6. Descriptive Statistics of Namin Synoptic Station

variable	N	Mean	Minimum	Median	Maximum
vsw	528	0.3129	0.1208	0.34605	0.4153
T2	528	9.393	-10.77	10.097	23.661
pe	528	-3.387	-8.636	-3.286	0
LAIl	528	1.5215	1.174	1.4808	2.0445
LAIh	528	2.3343	2.222	2.3208	2.5083
St1	528	11.748	0.067	11.568	26.183
ETo	528	28.172	5.44	25.435	81.71

با هدف ساخت یک مدل که توانایی درک و ایجاد ارتباط، مابین پارامترهای ورودی و تبخیر و تعرق مرجع را داشته باشد، از ترکیب تصادفی داده‌های چهار ایستگاه هواشناسی شامل ایستگاه‌های نیر، سرعین، اردبیل و فرودگاه استفاده شد. از داده ایستگاه نمین که نظر مکانی شرقی‌ترین ایستگاه منطقه مطالعه بود، به جهت اینکه مابین سایر ایستگاه‌ها نباشد، به عنوان ایستگاهی که با آن ارزیابی نهایی مدل‌ها انجام شود، استفاده گردید.

۳.۲. انتخاب داده‌های نهایی

در این مرحله و به منظور نشان دادن میزان همبستگی بین پارامترهای ورودی و تبخیر و تعرق مرجع، برای اختصار تنها نتیجه همبستگی مابین داده‌های ایستگاه هواشناسی اردبیل در جدول (۷) ارائه شده است.

Table 7. Correlation between ETo and input parameters (Ardabil Station)

	Station	vsw	T2	pe	LAI1	LAIh	St1
ETO	Ardabil	-0.754	0.892	-0.887	0.769	0.761	0.916
	Nir	-0.522	0.843	-0.871	0.733	0.772	0.834
	Sarein	-0.561	0.862	-0.891	0.787	0.785	0.858
	AirPort	-0.698	0.864	-0.909	0.830	0.826	0.858
	Namin	-0.662	0.780	-0.843	0.858	0.860	0.781

به دلیل تغییرات دمایی موجود در سطح دشت که در جدول (۲) تا (۶) نیز قابل مشاهده است و برای تشخیص الگو و طبقه‌بندی بین ورودی‌ها و تابع هدف و در نتیجه در مناطق مختلف محدوده مطالعاتی، از تمامی داده‌های ورودی استفاده شد، که به عنوان نمونه سری زمانی داده‌ها برای ایستگاه اردبیل مطابق شکل (۳) است.

Soil Water Time Series

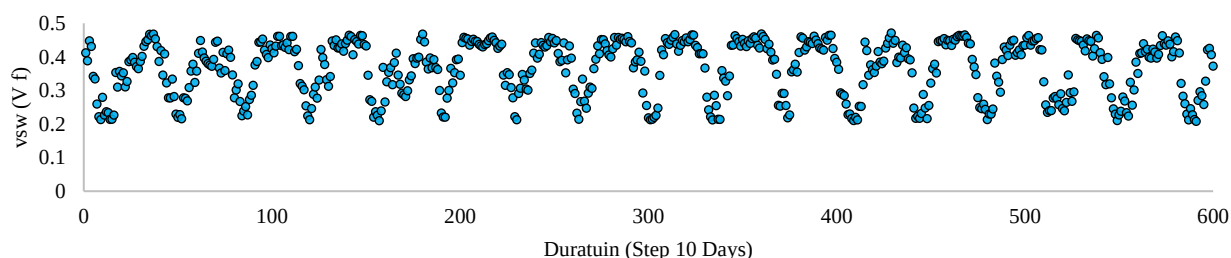


Figure 3.a. Soil water Time series

Temperature Time Series

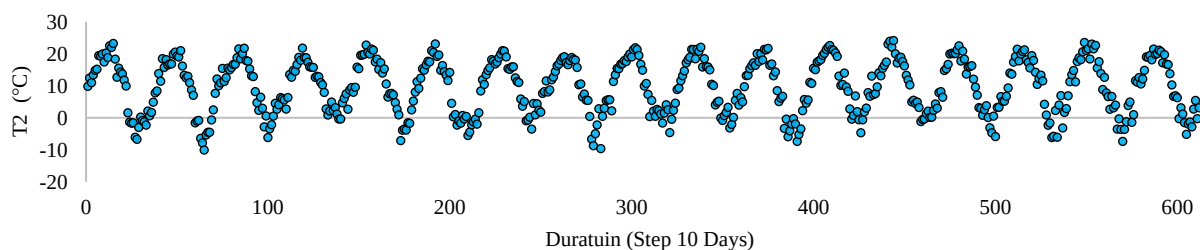


Figure 3.b. Temperature Time series

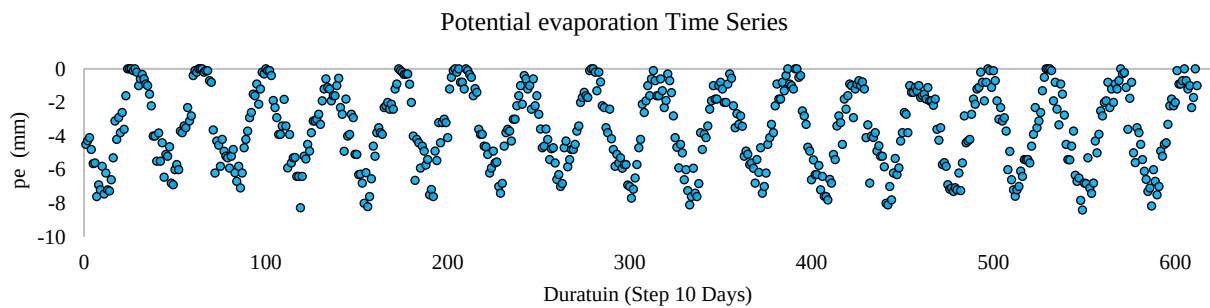


Figure 3.c. Reference evaporation Time series

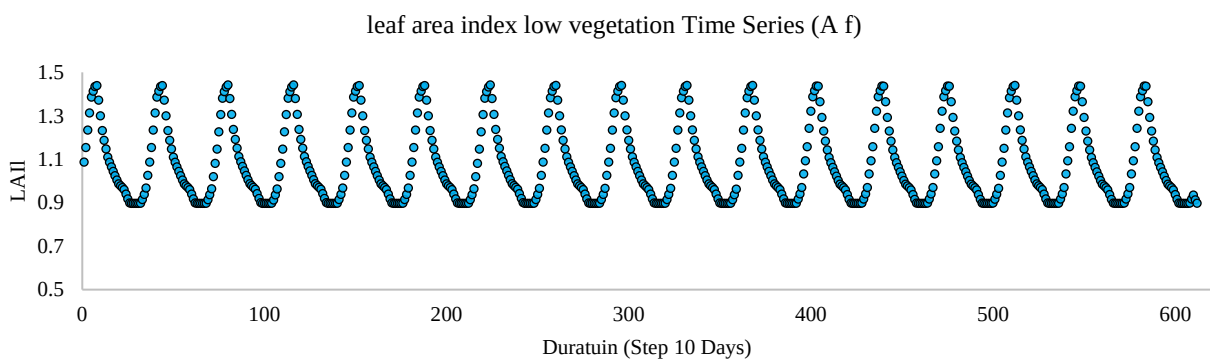


Figure 3.d. LAI low Time series

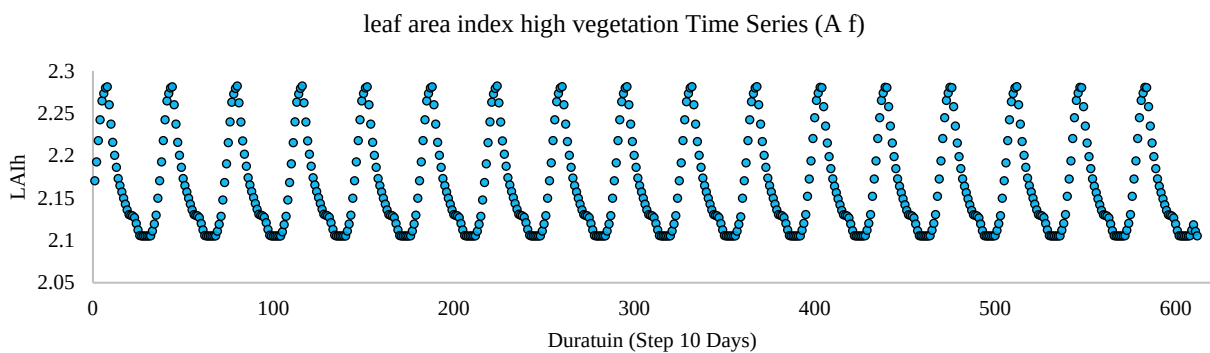


Figure 3.e. LAI high Time series

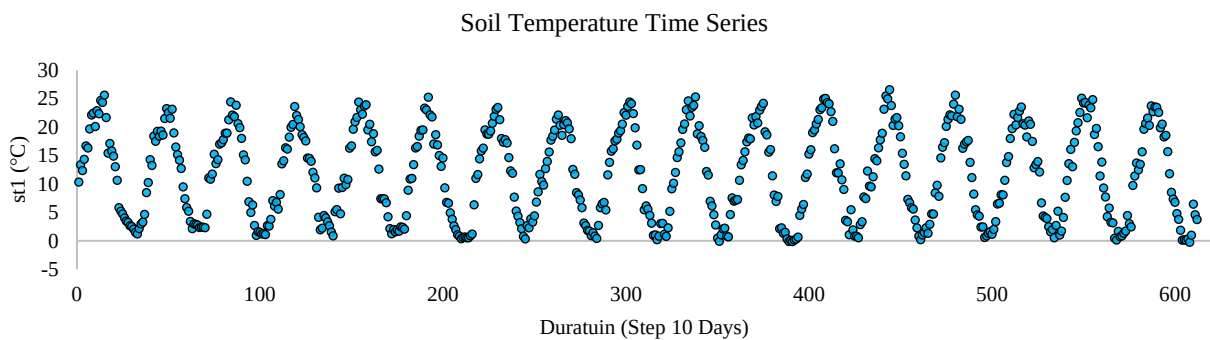


Figure 3.f. Soil Temperature Time series

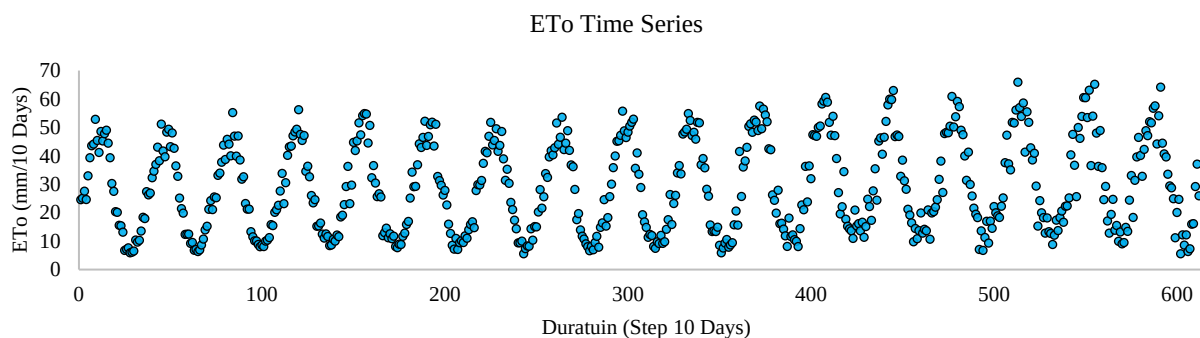


Figure 3.g. Evapotranspiration Time series
Figure 3. Model's Parameters Time series in Ardabil station

۳.۳. مدل سازی

در این بخش با تکمیل مراحل جمع آوری داده ها و ساخت سری های زمانی مربوط به پارامترهای ورودی و هدف، نسبت به ساخت مدل ها و بررسی نتایج در دو بخش برای هر مدل پرداخته شد. در بخش اول مدل ها با استفاده از ترکیب تصادفی داده های چهار ایستگاه که یک سری زمانی با طول دوره ۱۶۰۰ واحدی را تشکیل می داد، آموزش و یا ساخته شدند. سپس ارزیابی نهایی مدل ها با استفاده از داده های ایستگاه نمین انجام گرفت.

در انتهای بررسی دقت مدل ها در هر بخش، ترکیب داده های ورودی (جدول ۸) در دقت تخمین تابع هدف براساس آماره R^2 بررسی شد تا تاثیر وجود پارامترهای مختلف بر دقت مدل ها نیز مشخص گردد؛ ترتیب حذف داده ها براساس همبستگی کمتر به بیشتر در داده های ورودی با تابع هدف بود.

Table 8. Input Variants Combination

COMB num	Combination
Comb 1	st1, T2, pe, LAI1, LAIh, vsw
Comb 2	st1, T2, pe, LAI1, LAIh,
Comb 3	st1, T2, pe, LAI1
Comb 4	st1, T2, pe
Comb 5	st1, T2
Comb 6	St1

برای محاسبه ضرایب و در نتیجه تشکیل معادله رگرسیون خطی چندگانه از نرم افزارهای Excel و Minitab استفاده شد که ضرایب پارامترهای ورودی به شرح جدول (۹) و نتیجه آن مطابق شکل (۴) به دست آمد؛ با این توضیح که برای امکان مقایسه بهتر، تنها ۳۰۰ داده اول در شکل (۴) ترسیم شده است.

Table 9. Constant coefficients of multiple linear regression equation

Parameters	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.000					
vsw	-40.972	2.813	-14.566	0.000	-46.489	-35.454
T2	0.598	0.077	7.749	0.000	0.447	0.749
pe	-0.698	0.092	-7.607	0.000	-0.878	-0.518
LAI1	23.188	1.148	20.202	0.000	20.937	25.440
LAIh	4.127	0.407	10.149	0.000	3.330	4.925
St1	0.521	0.097	5.378	0.000	0.331	0.711

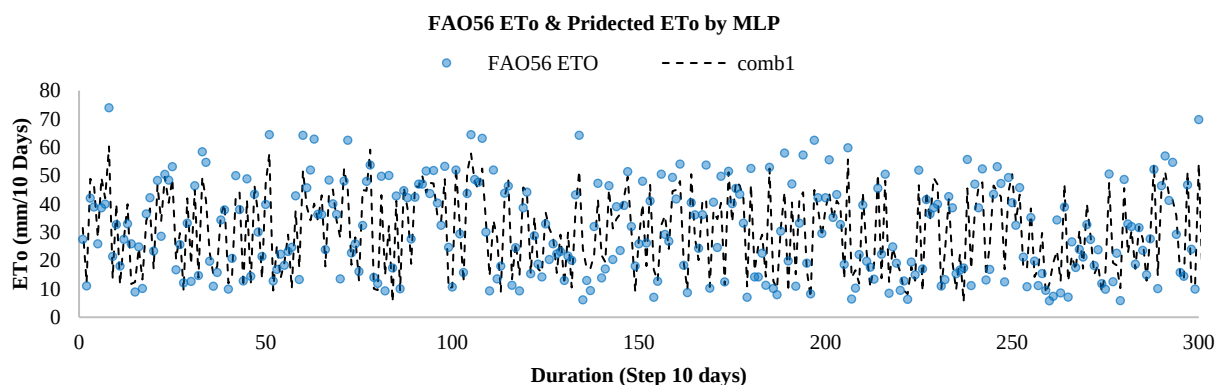


Figure 4. Reference Evapotranspiration for random time series in MLR method

از مدل ساخته شده برای برآورد تبخیر و تعرق مرجع ایستگاه نمین استفاده شد که نتایج آن در شکل (۵) و جدول (۱۰) ارائه شده است.

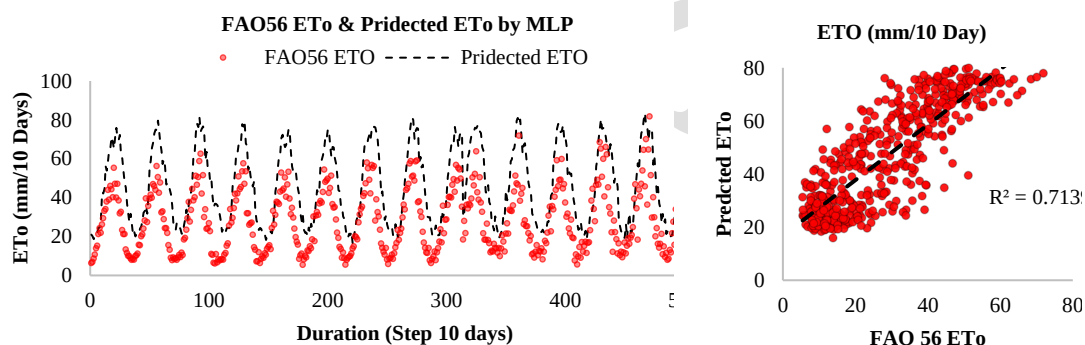


Figure 5. Estimated reference evapotranspiration for Namin station in MLR method

برای ساخت مدل جنگل تصادفی از نرم افزار Matlab استفاده شد که ساختار مدل از ۱۰ درخت تصمیم گیری و ۵۰ واحد شاخه (برگ) ساخته و نتیجه آن به مطابق شکل (۶) به دست آمد؛ با این توضیح که برای امکان مقایسه بهتر، تنها ۳۰۰ داده اول در شکل (۶) ترسیم شده است.

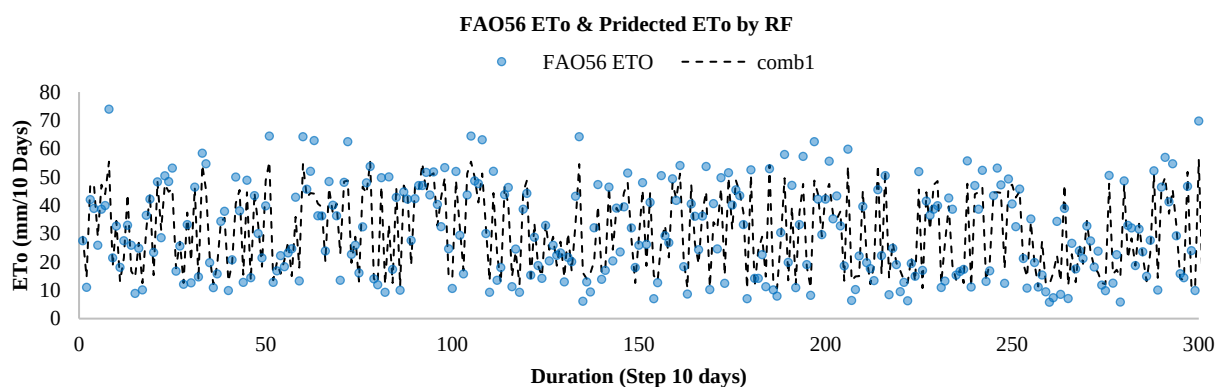


Figure 6. Reference Evapotranspiration for random time series in RF method

از مدل ساخته شده برای برآورد تبخیر و تعرق مرجع ایستگاه نمین استفاده شد که نتایج آن در شکل (۷) و جدول (۱۰) ارائه شده است.

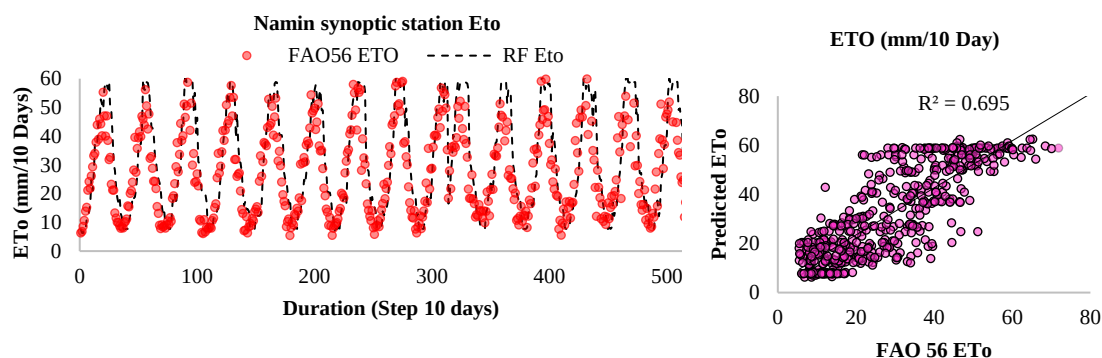


Figure 7. Estimated reference evapotranspiration for Namin station in RF method

برای ساخت مدل ماشین بردار پشتیبان از نرم افزار Matlab استفاده شد که ساختار مدل به استفاده از تابع خطی کرنل تعریف و نتیجه آن به مطابق شکل (۸) به دست آمد؛ با این توضیح که برای امکان مقایسه بهتر، تنها ۳۰۰ داده اول در شکل (۸) ترسیم شده است.

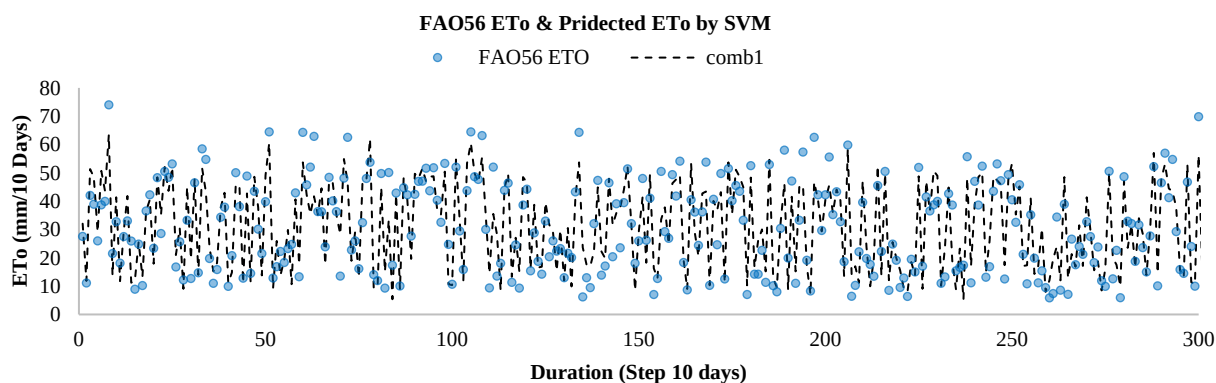


Figure 8. Reference Evapotranspiration for random time series in SVM method

از مدل ساخته شده برای برآورد تبخیر و تعرق مرجع ایستگاه نمین استفاده شد که نتایج آن در شکل (۹) و جدول (۱۰) ارائه شده است.

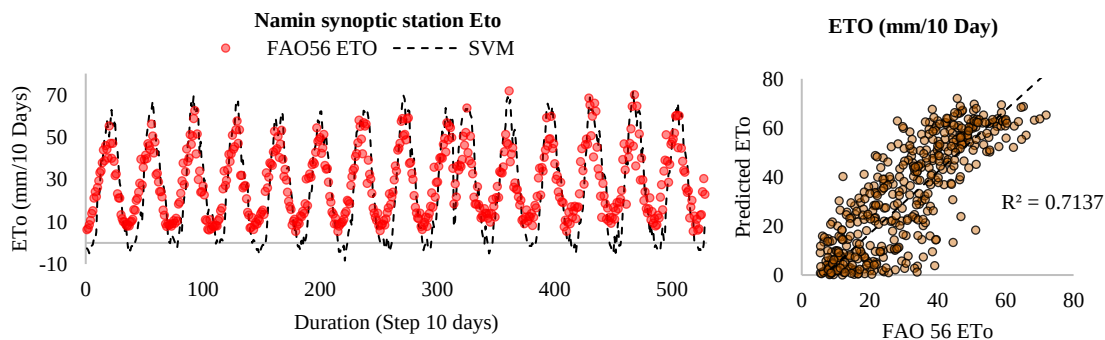


Figure 9. Estimated reference evapotranspiration for Namin station in SVM method

مطابق نتایج ارایه شده در جدول (۱۰) مقادیر حاصل از آماره‌های R^2 ، NSE و RMSE نشان داد که مدل RF از دقت بالاتری در برآورد تبخیر و تعرق مرجع با داده‌های یکسان و مشترک دارد. مقادیر حاصل از آماره R^2 در وهله اول برای هر سه مدل قابل قبول است اما مقدار NSE علیرغم قابل قبول بودن آن در مرحله تعریف مدل، در محل ایستگاه نمین به سبب منفی بودن قابل قبول نیست، که این موضوع در دو مدل MLR و SVM نشان از عدم کارایی مدل‌ها در صورت تغییر داده است. که این درخصوص جنگل تصادفی مشاهده نشده و نشان از کارایی این مدل در استفاده از داده‌های سنجش از دور برای تخمین ETo است. با بررسی نتایج سایر تحقیقات، برآورد تبخیر و تعرق پتانسیل، با استفاده از مدل‌های یادگیری ماشین، می‌تواند دقیق‌تر از مدل‌های تجربی باشد، اما نکته مهم در این رابطه این است که مدل‌های تجربی به‌طور معمول، در مقابل تغییرات داده‌های ورودی و یا اضافه شدن داده‌ها پایدار هستند، اما مدل‌های یادگیری ماشین، می‌توانند کم‌ترین دقت پیش‌بینی‌کننده را زمانی که یک مجموعه داده جدید به مدل اضافه می‌شود، داشته باشند و ناپایدار شود (Dong et al., 2020; Hassan, Khalil, Kaseb, & Kassem, 2017). چراکه بسیاری از مدل‌های داده‌مبنا، اغلب یک مجموعه داده مشخص را برای آموزش و آزمایش در نظر می‌گیرند، که ممکن است این موضوع باعث کاهش خطا در یادگیری مدل شده یا نتیجه نسبتاً خوب و دقیقی ارایه کند، زیرا دقت عملکرد فقط برای یک الگوی خاص از مجموعه داده‌های اصلی تعریف شده و شاید آن مجموع داده‌های منتخب، الگوی مناسبی را برای کل محدوده مطالعاتی در نظر نگیرد (Martí, González-Altozano, López-Urrea, Mancha, & Shiri, 2015; Shiri, 2018; Shiri et al., 2014). پس در نظر گرفتن یک روش اعتبارسنجی برای امکان ارزیابی عملکرد دقیق مدل، کافی است. چراکه عملکرد هر مدل بر اساس الگوهای داده‌ها است. به‌طوری‌که یک مدل همسان، عملکردهای متفاوتی را در مطالعات مختلف تحت شرایط آب و هوایی مختلف نشان داده‌است (Fan et al., 2018; Wu & Fan, 2019)، که در این صورت استفاده از ترکیب داده‌های ورودی از نقاط مختلف منطقه مطالعاتی می‌تواند در ثبات و پایداری مدل کمک کننده باشد (Akbari Majd, 2024). به همین منظور در این تحقیق یک سری‌زمانی ترکیبی و تصادفی بدون رعایت ترتیب زمانی و مکانی با استفاده از کل داده‌های چهار ایستگاه منتخب برای ساخت و آموزش مدل تشکیل شد که این سری‌زمانی باعث پایداری مدل در مقابل داده‌های جدید شد.

همچنین در هر سه مدل RMSE دارای مقدار بالایی است که توصیه شده برای افزایش بیشتر همپوشانی بین تابع هدف و ورودی‌های مدل از روش‌های داده‌کاوی در صورت وجود یا عدم وجود داده‌های پرت (Chatfield & Xing, 2019; Coucke & Soumali, 2017; Heckert et al., 2002; Hund, Massart, & Smeyers-Verbeke, 2000; Kalteh & Hjorth, 2009; Little et al., 2013; Organization, 2008) استفاده کرد.

Table 10. Models' accuracy

Model	Input Dataset	RMSE	NSE	R^2
MLR	Models Data (4 stations)	6.44	0.838	0.84
	Namin Station	21.02	-0.688	0.71
RF	Models Data (4 stations)	5.65	0.875	0.88
	Namin Station	10.76	0.558	0.7
SVM	Models Data (4 stations)	6.59	0.83	0.84
	Namin Station	13.6	-1	0.71

مشابه این تحقیق توسط (Fang et al., 2022) برای توسعه روش‌های تخمین ET انجام شد که برای ورودی‌های مدل از داده ماهواره‌های NOAA GOES استفاده کردند. نتایج نشان داد که روش یادگیری ماشینی درخت رگرسیون برای تولید ET روزانه تحت هر شرایط آب و هوایی با دقت رضایت‌بخش با حجم زیادی از داده‌های ماهواره‌ای امکان‌پذیر و مؤثر است. علاوه بر این، (Alavi, Albaji, Golabi, Ali Naseri, & Homayouni, 2024) برای تخمین ETC بر اساس چهار الگوریتم ML با استفاده از داده‌های چندمنبعی RS و اندازه‌گیری‌های هواشناسی و زمینی (Meteo-GM) را طی سه فصل رشد نیشکر (۲۰۱۸-

(۲۰۲۱) در خوزستان انجام دادند. نتایج نشان داد که الگوریتم RFR دقیق‌ترین تخمین ETc را در تمام سناریوها ارائه می‌کند. همچنین (Amani & Shafizadeh-Moghadam, 2023) بیان نمودند که ترکیب تصاویر سنجش از دور و مدل‌های یادگیری ماشین (ML) ظرفیت قابل توجهی برای تخمین ET فراهم کرده است. بنابراین با در نظر گرفتن آماره‌های مورد استفاده در این تحقیق و مقایسه با نتایج سایر محققین استفاده از مدل‌های هوشمند MLR، RF و SVM نتایج قابل قبولی را برای محاسبه ارائه می‌نماید.

مشابه این تحقیق توسط (Dos Santos, Mantovani, Bufon, & Fernandes-Filho, 2024) انجام شد که در آن، تبخیر و تعرق واقعی برای محصول نیشکر با استفاده از تکنیک‌های یادگیری ماشین و داده‌های حاصل از سنسورهای OLI و TIRS و کاربرد ماهواره‌های Landsat-8,9 برای تولید متغیر مشاهده‌شده، مدل‌سازی شد، که نتایج نشان داد الگوریتم‌های یادگیری ماشین قادر به پیش‌بینی تبخیر و تعرق واقعی بدون نیاز به استفاده از ایستگاه‌های هواشناسی و داده‌های حرارتی هستند. در مطالعه‌ای دیگری که (Noghankar, Raeini, Gholami Sefidkouhi, & mobini, 2023) انجام دادند، روشی برای پیش‌بینی کوتاه‌مدت نقشه‌های ET روزانه برنج با استفاده از تصاویر ماهواره‌ای و الگوریتم‌های یادگیری ماشین ارائه شد. نتایج نشان داد که مقادیر RMSE برای مدل RVM در دو سناریوی مطالعه به ترتیب ۰/۵۶ و ۰/۸۲ بدست آمد که نشان دهنده خطای کمتر نسبت به مطالعه حاضر است و دلیل آن می‌تواند به رابطه تبخیر و تعرق واقعی روزانه با داده‌های سنجش از دور در منطقه مطالعه فوق بستگی داشته باشد که در آنجا عامل گیاهی نیز وارد محاسبات شده است، در حالی که در دشت اردبیل تبخیر و تعرق پتانسیل محاسبه شده است. علاوه بر این، بر اساس داده‌های Landsat 8 OLI_TIRS و داده‌های تصویر سنجش از دور تحقیقی توسط (Song et al., 2023) از یک مدل رگرسیون خطی چندگانه (MLR) و یک مدل (ELM) برای تحلیل تبخیر و تعرق، به عنوان پارامتر هدف، استفاده کردند. بهترین مدل برای ET در سال‌های ۲۰۱۶، ۲۰۱۷ و ۲۰۲۱ MLR با R^2 بیشتر از ۰/۸ بود. بهترین مدل برای ET در سال‌های ۲۰۱۸-۲۰۱۹، ELM با R^2 به ترتیب ۰/۸۳ و ۰/۶۲ بود که نتایج آن‌ها مشابه نتایج به‌دست آمده در این مطالعه است.

Table 11. The accuracy of models by Random-Combination Inputs

Models	R^2		
	SVM	MLR	RF
FAO56 ETO	1.0000	1.0000	1.0000
ETO_Comb1	0.8758	0.8375	0.8376
ETO_Comb2	0.8804	0.8346	0.8280
ETO_Comb3	0.8718	0.7765	0.8143
ETO_Comb4	0.8346	0.7729	0.7943
ETO_Comb5	0.8204	0.3325	0.7748
ETO_Comb6	0.8081	0.4350	0.7683

باتوجه به نتایج جدول (۱۱) مشاهده شد به طور نسبی با حذف داده‌ها به ترتیب از ترکیب شماره یک تا شش، مقدار آماره R^2 در هر سه مدل کاهش می‌یابد، که باتوجه به این موضوع، بهترین ترکیب در این مدل، ترکیب شماره یک یا دو است. پس باتوجه به قابلیت استخراج پارامترهای مورد نیاز از طریق سنجش از دور برای هر نقطه دلخواه در محدوده مطالعاتی حداقل، استفاده از سه ترکیب اول شامل پارامترهای متوسط حجم آب موجود در خاک (vsw)، متوسط دمای اتمسفر (T_2)، حجم تبخیر آب (pe)، شاخص سطح برگ کم‌تراکم (LAIl)، شاخص سطح برگ پرتراکم (LAIh) توصیه می‌شود. برای بررسی حساسیت مدل‌ها می‌توان از مقایسه بین خطای استاندارد و مقادیر پارامترهای برآورد شده در جدول (۹) استفاده کرد (Rasoulzadeh & Ghoorabjiri, 2014; Rasoulzadeh & Yaghoubi, 2014)؛ در این صورت مدل به بیشترین حساسیت را به شاخص سطح برگ کم‌تراکم و کمترین حساسیت را به دمای خاک کمترین را داشت.

مشابه این تحقیق توسط (Saremi, 2015)، درخصوص تأثیر ترکیب پارامترهای ورودی بر میزان حساسیت ETo نسبت به پارامترهای اقلیمی با استفاده از شبکه عصبی انجام شد. نتایج این تحقیق نشان داد که لزوماً افزایش تعداد داده‌ها در لایه ورودی منجر به بهبود نتایج مدل‌های هوشمند نمی‌شود، اما تعداد داده‌های بیش‌تر در خیلی از مواقع نتایج قابل قبول‌تری ارائه می‌دهد. مشابه تحقیقات مدل‌های هوشمند، برای مدل‌های تجربی نیز تحقیقی که توسط (Martí, López-Urrea, Mancha, González-) و (Altozano, & Román, 2024) روی معادلات هارگریوز-سامانی و فائو پنمن-مانتیت برای برآورد تبخیر و تعرق مرجع با استفاده از سه ترکیب ورودی و واسنجی با روش لایسیمتری انجام شد، که براساس نتایج ارایه شده مدل‌های وابسته به میانگین دمای هوا و تابش خورشیدی، دقت بیش‌تری نسبت به مدل‌های متکی بر میانگین دمای هوا و رطوبت نسبی داشتند.

۳.۴. نتیجه‌گیری

در بین مدل‌های مورد مطالعه در این تحقیق، مدل RF به عنوان مدل برتر نسبت به MLR و SVM بود. نکته قابل بحث در مورد نتایج، این است که هر سه مدل در زمان استفاده از داده‌های چهار ایستگاه منتخب (اردبیل، فرودگاه، نمین و نیر)، علیرغم RMSE بالا، دارای دقت قابل قبولی هستند. اما این موضوع در زمان استفاده از اطلاعات ایستگاه نمین که از داده‌های آن برای ساخت و آموزش مدل استفاده نشده بود، صادق نبود. چراکه با توجه به نتایج آماره‌های ارزیابی دقت دو مدل MLR و SVM به سبب مقادیر منفی NSE برای ایستگاه نمین قابل قبول نبودند.

مطالعه حاضر نشان داد که مدل‌های یادگیری ترکیبی مانند جنگل تصادفی و درخت تصادفی می‌تواند یک مدل یادگیری ماشین مطمئن با در نظر گرفتن دقت و ثبات، برای پیش‌بینی ETo با استفاده از مجموعه داده‌های آب و هوایی استخراج شده از طریق سنجش از دور، در مناطق مختلف آب و هوایی دشت اردبیل باشد و ممکن است برای اقلیم‌های مشابه در سراسر جهان نیز قابلیت استفاده داشته باشد. با این حال، تحقیقات آینده برای ارزیابی عملکرد مدل‌های RF برای دوره‌های مختلف، به عنوان مثال، مقیاس‌های ساعتی یا ماهانه یا سایر مناطق مشابه جهان با آب و هوای متفاوت مورد نیاز است.

۴. جمع‌بندی

در این مطالعه کارایی مدل‌های RF، MLR و SVM را برای تخمین دقیق ETo محاسبه شده با روش فائو پنمن - مانتیت با گام‌های ده روزه، با ترکیبی از پارامترهای ورودی آب و هوایی دریافت شده از طریق سنجش از دور و در محل پنج ایستگاه هواشناسی در منطقه دشت اردبیل، طی سال‌های ۱۴۰۱-۱۳۸۵ بررسی شد. نتایج نشان داد که از بین متغیرهای آب و هوایی دریافت شده، سه ترکیب از پارامترها دقت رضایت‌بخشی را برای تخمین ETo ارایه می‌دهد. اما به سبب در نظر گرفتن الگوهای متفاوت آب و هوایی و جلوگیری از عدم ثبات در صورت تغییر داده‌ها بهترین ترکیب برای دشت اردبیل، در هر سه مدل ترکیب $st1$, $T2$, pe , $LAIh$, $LAIh$, vsw بود. از بین مدل‌ها، مدل RF دقت پیش‌بینی برتری را نسبت به MLR و SVM نشان داد (مقدار بالاتر R^2 ، NSE و مقدار کمتر RMSE) اما با در نظر گرفتن مقدار RMSE در هر سه مدل برای افزایش تطابق میزان برآوردی و واقعی ETo استفاده از داده‌کاوی اعم از بررسی و حذف روند یا تناوب پیشنهاد شد.

علیرغم نتیجه بهتر در استفاده هم‌زمان از هر شش پارامتر در ورودی مدل‌ها با توجه به ضرایب ثابت به دست آمده در مدل MLR و دقت مدل‌های RF و SVM می‌توان نتیجه گرفت پارامترهای دمایی و شاخص سطح برگ در ترکیب با سایر پارامترها امکان بهتری در برآورد تبخیر و تعرق مرجع با روش فائو پنمن-مانتیت در استفاده از داده‌های سنجش از دور دارند. بنابراین استفاده از ترکیب یادگیری ماشین و داده‌های سنجش از دور در تخمین تبخیر و تعرق مرجع در مناطق فاقد آمار نتایج قابل قبولی خواهد داشت. از این رو نتایج این مطالعه برای محققانی که عملکرد مدل‌های مختلف یادگیری ماشین را، شبیه‌سازی یا ارزیابی می‌کنند مفید خواهد بود. همچنین، برای سیاست‌گذاران و دست‌اندرکاران برای مدیریت منابع آب کشاورزی در دشت اردبیل

به سبب عدم پراکنش مناسب ایستگاه‌های هواشناسی و در نتیجه عدم وجود داده‌های هواشناسی در مناطق مختلف دشت، قابل استفاده خواهد بود.

۵. تشکر و قدردانی

نویسندگان از گروه مهندسی آب دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی که موجبات تسهیل انجام این تحقیق را فراهم نموده‌اند، تشکر و قدردانی می‌نمایند.

۶. تعارض منافع

در این مقاله تعارض منافع وجود ندارد و این مسئله مورد تأیید نویسندگان مقاله است.

۷. پی‌نوشت‌ها

-
- ¹ Evapotranspiration
 - ² Micro weather methods
 - ³ Machine Learning
 - ⁴ neuro-fuzzy neural network
 - ⁵ least square-support vector machine
 - ⁶ fuzzy logic
 - ⁷ multiple-layer perception neural network
 - ⁸ relevance vector machine
 - ⁹ multivariate regression spline
 - ¹⁰ LS-SVM
 - ¹¹ Google Earth Engine
 - ¹² Monin-Obukhov similarity theory
 - ¹³ Data-Driven
 - ¹⁴ Vladimir Vapnik
 - ¹⁵ Structural Risk Minimization
 - ¹⁶ local minimum
 - ¹⁷ high dimension
 - ¹⁸ decision planes
 - ¹⁹ hyperplane
 - ²⁰ kernel

- Abdzaad Gohari, A., Nik Akhtar, A., Ebrahimipak, N., & Tafteh, A. (2023). Using NIAZAB System to Determine Soybean Water Use Based on the Inverse Solution of the Production Functions under Different Irrigation Conditions. *Journal of Water Research in Agriculture*, 37(2), 159-169. doi:http://10.22092/jwra.2023.361550.978
- Akbari Majd, A., Azizi Mobaser, J., Rasoulzadeh, A., Hasanpour Kashani, M., & Kisi, O. (2024). Enhancing the accuracy of metaheuristic neural networks in predicting underground water levels using meteorological data and remote sensing: A case study of Ardabil Plain, Iran. *Ain Shams Engineering Journal*, 103061. doi:https://doi.org/10.1016/j.asej.2024.103061
- Alavi, M., Albaji, M., Golabi, M., Ali Naseri, A., & Homayouni, S. (2024). Estimation of sugarcane evapotranspiration from remote sensing and limited meteorological variables using machine learning models. *Journal of Hydrology*, 629, 130605. doi:https://doi.org/10.1016/j.jhydrol.2023.130605
- Allan, R., Pereira, L., & Smith, M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56* (Vol. 56).
- 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* (Vol. 300).
- Amani, S., & Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. *Agricultural water management*, 284, 108324. doi:https://doi.org/10.1016/j.agwat.2023.108324
- Bachour, R., Maslova, I., Ticlavilca, A. M., Walker, W. R., & McKee, M. (2016). Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. *Stochastic Environmental Research and Risk Assessment*, 30, 103-117. doi:https://
- Ball, J. T., Woodrow, I. E., & Berry, J. A. (1987). A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions. In J. Biggins (Ed.), *Progress in Photosynthesis Research: Volume 4 Proceedings of the VIIth International Congress on Photosynthesis Providence, Rhode Island, USA, August 10–15, 1986* (pp. 221-224). Dordrecht: Springer Netherlands.
- Bastiaansen, W. G. M., Menenti, M., Feddes, R. A., & Holtslag, A. A. M. (1998). A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of Hydrology*, 212-213, 198-212. doi:https://doi.org/10.1016/S0022-1694(98)00253-4
- Blaney, H. F. (1952). Determining water requirements in irrigated areas from climatological and irrigation data. doi:https://
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5-32. doi:https://10.1023/A:1010933404324
- Bui, D., Pradhan, B., Löfman, O., Revhaug, I., & Dick, Ø. (2012). Landslide susceptibility assessment in the Hoa Binh Province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. *Geomorphology*, 171-172. doi:https://10.1016/j.geomorph.2012.04.023
- Callañaupa Gutierrez, S., Segura Cajachagua, H., Saavedra Huanca, M., Flores Rojas, J., Silva Vidal, Y., & Cuxart, J. (2021). Seasonal variability of daily evapotranspiration and energy fluxes in the Central Andes of Peru using eddy covariance techniques and empirical methods. *Atmospheric Research*, 261, 105760. doi:https://doi.org/10.1016/j.atmosres.2021.105760
- Carlson, T. N., Capehart, W. J., & Gillies, R. R. (1995). A new look at the simplified method for remote sensing of daily evapotranspiration. *Remote Sensing of Environment*, 54(2), 161-167. doi:https://doi.org/10.1016/0034-4257(95)00139-R
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R*: Chapman and hall/CRC.
- Chen, J. M., & Liu, J. (2020). Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sensing of Environment*, 237, 111594. doi:https://doi.org/10.1016/j.rse.2019.111594

- Chia, M. Y., Huang, Y. F., & Koo, C. H. (2022). Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agricultural water management*, 261, 107343. doi:https://doi.org/10.1016/j.agwat.2021.107343
- Chia, M. Y., Huang, Y. F., Koo, C. H., Ng, J. L., Ahmed, A. N., & El-Shafie, A. (2022). Long-term forecasting of monthly mean reference evapotranspiration using deep neural network: A comparison of training strategies and approaches. *Applied Soft Computing*, 126, 109221. doi:https://doi.org/10.1016/j.asoc.2022.109221
- Cleugh, H. A., Leuning, R., Mu, Q., & Running, S. W. (2007). Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment*, 106(3), 285-304. doi:https://doi.org/10.1016/j.rse.2006.07.007
- Córdova, M., Carrillo-Rojas, G., Crespo, P., Wilcox, B., & Céleri, R. (2015). Evaluation of the Penman-Monteith (FAO 56 PM) Method for Calculating Reference Evapotranspiration Using Limited Data. Application to the Wet Páramo of Southern Ecuador. *Mountain Research and Development*, 35, 230-239. doi:https://10.1659/MRD-JOURNAL-D-14-0024.1
- Coucke, W., & Soumali, M. (2017). Demystifying EQA statistics and reports. *Biochemia Medica*, 27, 37-48. doi:https://10.11613/BM.2017.006
- Del Cerro, R. T. G., Subathra, M., Kumar, N. M., Verrastro, S., & George, S. T. (2021). Modelling the daily reference evapotranspiration in semi-arid region of South India: A case study comparing ANFIS and empirical models. *Information Processing in Agriculture*, 8(1), 173-184. doi:https://
- Deng, C., Wu, J., & Shao, X. (2008, 2008/). *Reliability Assessment of Machining Accuracy on Support Vector Machine*. Paper presented at the Intelligent Robotics and Applications, Berlin, Heidelberg.
- Deris, A. M., Zain, A. M., & Sallehuddin, R. (2011). Overview of Support Vector Machine in Modeling Machining Performances. *Procedia Engineering*, 24, 308-312. doi:https://doi.org/10.1016/j.proeng.2011.11.2647
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241-258. doi:https://10.1007/s11704-019-8208-z
- Dos Santos, R. A., Mantovani, E. C., Bufon, V. B., & Fernandes-Filho, E. I. (2024). Improving actual evapotranspiration estimates through an integrated remote sensing and cutting-edge machine learning approach. *Computers and Electronics in Agriculture*, 225, 109258. doi:https://doi.org/10.1016/j.compag.2024.109258
- Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture*, 148, 95-106. doi:https://
- Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors*, 40(3), 354-375. doi:10.1518/001872098779591386
- Draper, N. (1998). *Applied regression analysis*: McGraw-Hill. Inc.
- Duhan, D., Singh, D., & Arya, S. (2020). Effect of projected climate change on reference evapotranspiration in the semiarid region of central India. *Journal of Water and Climate Change*, 12(5), 1854-1870. doi:https://10.2166/wcc.2020.168
- Elbeltagi, A., Srivastava, A., Li, P., Jiang, J., Jinsong, D., Rajput, J., . . . Awad, A. (2023). Forecasting actual evapotranspiration without climate data based on stacked integration of DNN and meta-heuristic models across China from 1958 to 2021. *Journal of Environmental Management*, 345, 118697. doi:https://doi.org/10.1016/j.jenvman.2023.118697
- Eslamian, S., Gohari, A., Zareian, M. J., & Firoozfar, A. R. (2012). Estimating Penman-Monteith Reference Evapotranspiration Using Artificial Neural Networks and Genetic Algorithm: A Case Study. *Arabian Journal for Science and Engineering*, 37. doi:https://10.1007/s13369-012-0214-5
- Fakhar, M. S., & Kaviani, A. (2024). Accurate estimation of actual evapotranspiration using remote sensing data for improved water management in the Moghan plain. *Water and Irrigation Management*, -. doi:https://10.22059/jwim.2024.369438.1123
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Xiukang, W., . . . Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration

- using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225-241. doi:<https://10.1016/j.agrformet.2018.08.019>
- Fang, L., Zhan, X., Kalluri, S., Yu, P., Hain, C., Anderson, M., & Laszlo, I. (2022). Application of a Machine Learning Algorithm in Generating an Evapotranspiration Data Product From Coupled Thermal Infrared and Microwave Satellite Observations. *Front Big Data*, 5, 768676. doi:<https://10.3389/fdata.2022.768676>
- Fattahi dolatabadi, K., Babazadeh, H., Najafi, P., & Sedghi, H. (2018). A Model for Irrigation Scheduling Using the Difference between Air and Leaf Temperature of Corn. *Journal of Water Research in Agriculture*, 32(2), 305-320. doi:<https://10.22092/jwra.2018.116972>
- Feng, Y., Peng, Y., Cui, N., Gong, D., & Zhang, K. (2017). Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Computers and Electronics in Agriculture*, 136, 71-78. doi:<https://doi.org/10.1016/j.compag.2017.01.027>
- Forner, G., Abate, D., Mengoli, C., Palù, G., & Gussetti, N. (2015). High Cytomegalovirus (CMV) DNAemia Predicts CMV Sequelae in Asymptomatic Congenitally Infected Newborns Born to Women With Primary Infection During Pregnancy. *J Infect Dis*, 212(1), 67-71. doi:10.1093/infdis/jiu627
- Glantz, M. H. (1990). On the interactions between climate and society. *Population and Development Review*, 16, 179-200. doi:<https://10.1007/s11707-008-0045-6>
- Glantz, P. O., Rangert, B., Svensson, A., Stafford, G. D., Arnvidarson, B., Randow, K., . . . Hultén, J. (1993). On clinical loading of osseointegrated implants. A methodological and clinical study. *Clin Oral Implants Res*, 4(2), 99-105. doi:10.1034/j.1600-0501.1993.040206.x
- Gokcekus, H., Kassem, Y., & Woyea, L. T. (2023). A Prediction of Rainfall of Haifa Using MLR and ARIMA Models. *International Journal of Engineering and Applied Physics*, 3(1), 612-624. doi:<https://ijeap.org/ijeap/article/view/109>
- Goward, S. N., Cruickshanks, G. D., & Hope, A. S. (1985). Observed relation between thermal emission and reflected spectral radiance of a complex vegetated landscape. *Remote Sensing of Environment*, 18(2), 137-146. doi:[https://doi.org/10.1016/0034-4257\(85\)90044-6](https://doi.org/10.1016/0034-4257(85)90044-6)
- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural water management*, 217, 303-315. doi:<https://doi.org/10.1016/j.agwat.2019.03.015>
- Hassan, M. A., Khalil, A., Kaseb, S., & Kassem, M. A. (2017). Exploring the reference of tree-based ensemble methods in solar radiation modeling. *Applied Energy*, 203, 897-916. doi:<https://doi.org/10.1016/j.apenergy.2017.06.104>
- Heckert, N., Filliben, J., Croarkin, C., Hembree, B., Guthrie, W., Tobias, P., & Prinz, J. (2002). *Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods*: NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD.
- Heravi, H., & Zolfaghari, A.-A. (2024). Using Machine Learning Method to Estimate Evapotranspiration (Case Study: Semnan Province). *Iranian Journal of Soil and Water Research*, 55(5), 781-797. doi:<https://10.22059/ijswr.2024.371452.669652>
- Hund, E., Massart, D. L., & Smeyers-Verbeke, J. (2000). Inter-laboratory studies in analytical chemistry. *Analytica Chimica Acta*, 423(2), 145-165. doi:[https://doi.org/10.1016/S0003-2670\(00\)01115-6](https://doi.org/10.1016/S0003-2670(00)01115-6)
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- Jiang, X., Kang, S., Tong, L., & Li, F. (2016). Modification of evapotranspiration model based on effective resistance to estimate evapotranspiration of maize for seed production in an arid region of northwest China. *Journal of Hydrology*, 538, 194-207. doi:<https://doi.org/10.1016/j.jhydrol.2016.04.002>
- Jung, M., Reichstein, M., & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, 6(10), 2001-2013. doi:<https://10.5194/bg-6-2001-2009>
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., . . . Zhang, K. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467(7318), 951-954. doi:10.1038/nature09396

- Kalma, J. D., McVicar, T. R., & McCabe, M. F. (2008). Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surveys in Geophysics*, 29(4), 421-469. doi:<https://10.1007/s10712-008-9037-z>
- Kalteh, A. M., & Hjorth, P. (2009). Imputation of missing values in precipitation-runoff process database. *Hydrology Research*, 40. doi:<https://10.2166/nh.2009.001>
- Keshtegar, B., Kisi, O., Ghohani Arab, H., & Zounemat-Kermani, M. (2018). Subset modeling basis ANFIS for prediction of the reference evapotranspiration. *Water resources management*, 32, 1101-1116. doi:<https://>
- Krzywinski, M., & Altman, N. (2015). Multiple linear regression: when multiple variables are associated with a response, the interpretation of a prediction equation is seldom simple. *Nature Methods*, 12, 1103+. doi:<https://doi.org/10.1038/nmeth.3665>
- Leuning, R., Zhang, Y. Q., Rajaud, A., Cleugh, H., & Tu, K. (2008). A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research*, 44(10). doi:<https://doi.org/10.1029/2007WR006562>
- Little, T., Thompson, B., Cox, S., Woods, C., von Eye, A., Buskirk, T., . . . Wang, L. (2013). The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis. doi:<https://10.1093/oxfordhb/9780199934898.001.0001>
- Maeda, E. E., Wiberg, D. A., & Pellikka, P. K. E. (2011). Estimating reference evapotranspiration using remote sensing and empirical models in a region with limited ground data availability in Kenya. *Applied Geography*, 31(1), 251-258. doi:<https://doi.org/10.1016/j.apgeog.2010.05.011>
- Mahesh Chand, S., Shivam, P., Sanjay, S., Vishnu, P., & Sompal, S. (2022). Estimating seasonal reference evapotranspiration using limited weather data. *Journal of Agrometeorology*, 24(1), 99-102. doi:<https://10.54386/jam.v24i1.786>
- Malik, A., Saggi, M. K., Rehman, S., Sajjad, H., Inyurt, S., Bhatia, A. S., . . . Yaseen, Z. M. (2022). Deep learning versus gradient boosting machine for pan evaporation prediction. *Engineering Applications of Computational Fluid Mechanics*, 16(1), 570-587. doi:<https://>
- Martí, P., González-Altozano, P., López-Urrea, R., Mancha, L. A., & Shiri, J. (2015). Modeling reference evapotranspiration with calculated targets. Assessment and implications. *Agricultural water management*, 149, 81-90. doi:<https://doi.org/10.1016/j.agwat.2014.10.028>
- Martí, P., López-Urrea, R., Mancha, L. A., González-Altozano, P., & Román, A. (2024). Seasonal assessment of the grass reference evapotranspiration estimation from limited inputs using different calibrating time windows and lysimeter benchmarks. *Agricultural water management*, 300, 108903. doi:<https://doi.org/10.1016/j.agwat.2024.108903>
- McCuen, R., Knight, Z., & Cutter, A. (2006). Evaluation of the Nash–Sutcliffe Efficiency Index. *Journal of Hydrologic Engineering - J HYDROL ENG*, 11. doi:10.1061/(ASCE)1084-0699(2006)11:6(597)
- Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2017). Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture*, 139, 103-114. doi:<https://>
- Mendes Reis, M., Silva, A., Junior, J., Tuffi, S., Azevedo, A., & Lopes, É. (2019). Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. *Computers and Electronics in Agriculture*, 165, 104937. doi:<https://10.1016/j.compag.2019.104937>
- Monteith, J. L. (1965). *Evaporation and environment*. Paper presented at the Symposia of the society for experimental biology.
- Moran, M. S., Rahman, A. F., Washburne, J. C., Goodrich, D. C., Weltz, M. A., & Kustas, W. P. (1996). Combining the Penman-Monteith equation with measurements of surface temperature and reflectance to estimate evaporation rates of semiarid grassland. *Agricultural and Forest Meteorology*, 80(2), 87-109. doi:[https://doi.org/10.1016/0168-1923\(95\)02292-9](https://doi.org/10.1016/0168-1923(95)02292-9)
- Mu, Q., Heinsch, F. A., Zhao, M., & Running, S. W. (2007). Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, 111(4), 519-536. doi:<https://doi.org/10.1016/j.rse.2007.04.015>
- Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8), 1781-1800. doi:<https://doi.org/10.1016/j.rse.2011.02.019>

- Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., . . . Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13, 4349-4383. doi:<https://doi.org/10.5194/essd-13-4349-2021>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282-290. doi:[https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Negin, S., & Gholmohamadi, M. H. (2023). *Comparison of machine learning models for estimating reference evapotranspiration using satellite and ground data*. Paper presented at the 13th International Congress on Civil Engineering. <https://civilica.com/doc/1853201>
- Nema, M. K., Khare, D., & Chandniha, S. K. (2017). Application of artificial intelligence to estimate the reference evapotranspiration in sub-humid Doon valley. *Applied Water Science*, 7, 3903-3910. doi:<https://doi.org/10.1007/s13201-017-0693-1>
- Niasati, Z., Ebadi, H., & Kiani, A. (2021). Estimation of reference evapotranspiration using remote sensing data in Hamedan-Bahar Plain. *Iranian Water Researches Journal*, 15(4), 45-58. doi:<https://doi.org/10.22034/iwrj.2021.11165>
- Nishida, K., Nemani, R. R., Running, S. W., & Glassy, J. M. (2003). An operational remote sensing algorithm of land surface evaporation. *Journal of Geophysical Research: Atmospheres*, 108(D9). doi:<https://doi.org/10.1029/2002JD002062>
- Noghankar, H., Raeini, M., Gholami Sefidkouhi, M. A., & Mobini, M. (2023). Prediction of daily evapotranspiration images of rice using machine learning. *Iranian Journal of Soil and Water Research*, 53(12), 2793-2807. doi:<https://doi.org/10.22059/ijswr.2022.350978.669391>
- Nolan, B. T., Fienen, M. N., & Lorenz, D. L. (2015). A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology*, 531, 902-911. doi:<https://doi.org/10.1016/j.jhydrol.2015.10.025>
- Nouri-Khajebozlagh, R., Khaledian, M., & Kavooosi-Kalashami, M. (2020). Comparison of Water Productivity Indicators for Major Crops in Ardabil Plain. *Iranian Journal of Irrigation & Drainage*, 14(3), 894-904. doi:<https://doi.org/10.22034/ijid.2020.14.3.894>
- Organization, W. M. (2008). *Manual on Low-flow Estimation and Prediction*: World Meteorological Organization.
- Peng, L., Zeng, Z., Wei, Z., Chen, A., Wood, E. F., & Sheffield, J. (2019). Determinants of the ratio of actual to reference evapotranspiration. *Global Change Biology*, 25(4), 1326-1343. doi:<https://doi.org/10.1111/gcb.14577>
- Penman, H. L., & Keen, B. A. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 193(1032), 120-145. doi:<https://doi.org/10.1098/rspa.1948.0037>
- Piña-Monarez, M. R., & Ortiz-Yañez, J. F. (2015). Weibull and lognormal Taguchi analysis using multiple linear regression. *Reliability Engineering & System Safety*, 144, 244-253. doi:<https://doi.org/10.1016/j.ress.2015.08.004>
- Price, J. C. (1984). Land surface temperature measurements from the split window channels of the NOAA 7 Advanced Very High Resolution Radiometer. *Journal of Geophysical Research: Atmospheres*, 89(D5), 7231-7237. doi:<https://doi.org/10.1029/JD089iD05p07231>
- Rajput, J., Singh, M., Lal, K., Khanna, M., Sarangi, A., Mukherjee, J., & Singh, S. (2022). Performance evaluation of soft computing techniques for forecasting daily reference evapotranspiration. *Journal of Water and Climate Change*, 14(1), 350-368. doi:<https://doi.org/10.2166/wcc.2022.385>
- Rasoulzadeh, A., & Ghoorabjiri, M. (2014). Comparing hydraulic properties of different forest floors. *Hydrological Processes*, 28. doi:<https://doi.org/10.1002/hyp.10006>
- Rasoulzadeh, A., & Yaghoubi, A. (2014). Inverse modeling approach for determining soil hydraulic properties as affected by application of cattle manure. *International Journal of Agricultural and Biological Engineering*, 7, 27-35. doi:<https://doi.org/10.3965/j.ijabe.20140702.004>
- Rodrigues, G. C., & Braga, R. P. (2021). Estimation of Daily Reference Evapotranspiration from NASA POWER Reanalysis Products in a Hot Summer Mediterranean Climate. *Agronomy*, 11(10), 2077. doi:<https://doi.org/10.3390/agr11102077>
- Roerink, G. J., Su, Z., & Menenti, M. (2000). S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 25(2), 147-157. doi:[https://doi.org/10.1016/S1464-1909\(99\)00128-8](https://doi.org/10.1016/S1464-1909(99)00128-8)

- Ruiz-Álvarez, M., Gomariz-Castillo, F., & Alonso-Sarría, F. (2021). Evapotranspiration Response to Climate Change in Semi-Arid Areas: Using Random Forest as Multi-Model Ensemble Method. *Water*, 13(2), 222. doi:<https://doi.org/10.3390/w13020222>
- Samanta, B., Al-Balushi, K. R., & Al-ArAIMI, S. A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 16(7), 657-665. doi:<https://doi.org/10.1016/j.engappai.2003.09.006>
- Sanjay, S., Mahesh Chand, S., & Sunil, G. (2021). Assessment of irrigation water requirements for different crops in central Punjab, India. *Journal of Agrometeorology*, 23(4), 481-484. doi:<https://10.54386/jam.v23i4.183>
- Saremi, M. (2015). Determination of Effective Parameters in Estimating Reference Crop Evapotranspiration Using Artificial Neural Networks (Case study: Lorestan province). *Iranian Journal of Irrigation & Drainage*, 9(4), 614-623. doi:https://idj.iaid.ir/article_55085.html?lang=en
- Seifi, A., & Riahi, H. (2020). Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. *Journal of Water and Climate Change*, 11(1), 217-240. doi:<https://doi.org/10.1016/j.jhydrol.2018.04.042>
- Shiri, J. (2018). Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. *Journal of Hydrology*, 561, 737-750. doi:<https://doi.org/10.1016/j.jhydrol.2018.04.042>
- Shiri, J., Sadraddini, A., Nazemi, A., Kisi, O., Landaras, G., fakheri fard, A., & Marti, P. (2014). Generalizability of Gene Expression Programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *Journal of Hydrology*, 508, 1-11. doi:<https://10.1016/j.jhydrol.2013.10.034>
- Siasar, H., & Dindarlou, A. (2020). Estimation of daily reference evapotranspiration using models of deep learning, random forest and decision tree (Case study: Sistan Plain). *Iranian Water Research Journal*, 14(36), 99-108. doi:<https://doi.org/10.1016/j.jhydrol.2013.10.034>
- Singh, M., Satpute, S., Prasad, V., & Sharma, K. K. (2022). Trend analysis of temperature, rainfall, and reference evapotranspiration for Ludhiana district of Indian Punjab using non-parametric statistical methods. *Arabian Journal of Geosciences*, 15. doi:<https://10.1007/s12517-022-09517-1>
- Song, E., Zhu, X., Shao, G., Tian, L., Zhou, Y., Jiang, A., & Lu, J. (2023). Multi-Temporal Remote Sensing Inversion of Evapotranspiration in the Lower Yangtze River Based on Landsat 8 Remote Sensing Data and Analysis of Driving Factors. *Remote Sensing*, 15(11), 2887. doi:<https://doi.org/10.3390/rs15112887>
- Steel, R. G. D., & Torrie, J. H. (1960). *Principles and procedures of statistics*.
- Suleiman, A. A., & Hoogenboom, G. (2007). Comparison of Priestley-Taylor and FAO-56 Penman-Monteith for Daily Reference Evapotranspiration Estimation in Georgia. *Journal of Irrigation and Drainage Engineering*, 133(2), 175-182. doi:[https://10.1061/\(ASCE\)0733-9437\(2007\)133:2\(175\)](https://10.1061/(ASCE)0733-9437(2007)133:2(175))
- Thornthwaite, C. W. (1948). An Approach Toward a Rational Classification of Climate. *Soil Science*, 66, 55-94. doi:<https://doi.org/10.1016/j.jhydrol.2013.10.034>
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., . . . Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, 13(14), 4291-4313. doi:<https://10.5194/bg-13-4291-2016>
- Vapnik, V. (2013). *The nature of statistical learning theory*: Springer science & business media.
- Vishwakarma, D., Pandey, K., Kaur, a., Kushwaha, N. L., Kumar, R., Ali, R., . . . Kuriqi, A. (2021). Methods to estimate evapotranspiration in humid and subtropical climate conditions. *Agricultural water management*, 261. doi:<https://10.1016/j.agwat.2021.107378>
- Wang, L., Kisi, O., Hu, B., Bilal, M., Zounemat-Kermani, M., & Li, H. (2017). Evaporation modelling using different machine learning techniques. *International Journal of Climatology*, 37(S1), 1076-1092. doi:<https://doi.org/10.1002/joc.5064>
- Wu, L., & Fan, J. (2019). Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLOS ONE*, 14(5), e0217520. doi:<https://10.1371/journal.pone.0217520>

Zhang, K., Pan, S.-m., Zhang, W., Xu, Y., Cao, L., Hao, Y.-p., & Wang, Y. (2015). Influence of climate change on reference evapotranspiration and aridity index and their temporal-spatial variations in the Yellow River Basin, China, from 1961 to 2012. *Quaternary International*, 380. doi:<https://10.1016/j.quaint.2014.12.037>

نسخه آماده انتشار

Assessment of machine learning and remote sensing in quantifying Reference Evapotranspiration

Javanshir Azizi Mobaser  | Ali Rasoulzadeh  | Amin Akbari Majd 

1. Corresponding Author, Department of Water Engineering, Faculty of Agriculture and Natural Resources University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: ja_mobaser@uma.ac.ir
2. Department of Water Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: rasoulzadeh@uma.ac.ir
3. Department of Water Science and Engineering, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: amin.a.m.ce@gmail.com

Abstract

Estimating crop water requirements and designing irrigation systems effectively depends heavily on determining evapotranspiration. Machine learning models have been developed to estimate evapotranspiration (ET) and circumvent the limitations of empirical models. In recent years, ET estimate has been improved and refined through the use of remote sensing technology. Assessment of remote sensing and machine learning for determining reference evapotranspiration This study examined the effectiveness of three models for estimating reference evapotranspiration in the Ardabil plain: random forest (RF), multiple linear regression (MLR), and support vector machine (SVM). From 2006 to 2023, synoptic stations and remote sensing provided meteorological data for the model. The FAO Penman-Monteith method was used to calculate ETo, the target parameter, within a five-synoptic station range. The time series of input and target parameters were recorded at the four synoptic stations during the model's construction and training phases. A random time series and a combination of all the data were then used in the model's final evaluation phase, which only used the data from the fifth station. R^2 , NSE, and RMSE were the evaluation statistics that were employed. The RF model's statistical index results were 0.7, 0.558, and 10.76, the SVM's were 0.71, -1, and 13.6, and the MLR's were 0.71, -0.688, and 21. Comparing the outcomes, it was found that the RF model was more accurate than the others. The current study demonstrated that, for areas lacking statistics, the random forest model can be a dependable and reasonably accurate model for predicting ETo using RS data.

Keywords: Machine Learning, MLR, Reference Evapotranspiration, RF, SVM, Ungagged catchment