

مقایسه عملکرد مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادیان بوستینگ در برآورد تبخیر- تعرق گیاه مرجع (مطالعه موردی: دشت سیستان)

هادی سیاسر^{۱*}، تورج هنر^۲ و محمد عبداللهی‌پور^۳

(تاریخ دریافت: ۱۳۹۷/۹/۲۷؛ تاریخ پذیرش: ۱۳۹۸/۴/۲۳)

چکیده

تخمین تبخیر- تعرق گیاه مرجع یکی از مهم‌ترین مؤلفه‌ها در مطالعات هیدرولوژی، برنامه‌ریزی آبیاری و مدیریت منابع آب است. در این پژوهش به ارزیابی مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادیان بوستینگ در شرایط داده‌های اقلیمی منطقه سیستان پرداخته شد. داده‌های مورد استفاده شامل دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد و تبخیر از تشت بین سال‌های ۱۳۸۷ تا ۱۳۹۶ بود. با ارائه الگوهای مختلف شامل ترکیبی از پارامترهای هواشناسی به‌عنوان ورودی‌های مدل در مقیاس زمانی روزانه، مقدار تبخیر- تعرق توسط مدل‌های پیشنهادی به‌عنوان خروجی مدل برآورد شد. همچنین قابلیت پیش‌بینی این مدل‌ها، در مقایسه نتایج آنها با نتایج روش فائو- پنمن- مانتیث به‌عنوان روش مبنای ارزیابی شد. در بین الگوهای مورد بررسی، الگوی MI که دارای بیشترین تعداد پارامترهای هواشناسی ورودی (۱۰ پارامتر) بود، با کمترین خطا $RMSE = 0/633$ و $MAE = 0/451$ و بیشترین ضریب همبستگی $R = 0/993$ برای مدل درخت گرادیان بوستینگ، بهترین عملکرد را در بین مدل‌ها و الگوهای بالا نشان داد. با انجام فرایند حساسیت‌سنجی پارامترها، دمای هوا به‌عنوان مهم‌ترین متغیر مؤثر برای الگوهای انتخابی در هر سه مدل مطالعه‌شده، به‌دست آمد. بنابراین افزایش دقت و صحت داده‌های دما، می‌تواند به افزایش دقت مدل‌های پیشنهادی کمک کند. استفاده از مدل درخت گرادیان بوستینگ توانست بیش از ۷۰ درصد، زمان اجرای مدل را در مقایسه با سایر مدل‌های مورد مطالعه کاهش دهد. در نهایت این پژوهش، مدل درخت گرادیان بوستینگ را برای برآورد تبخیر- تعرق در منطقه دشت سیستان توصیه می‌کند.

واژه‌های کلیدی: روش فائو- پنمن- مانتیث، پارامترهای هواشناسی، حساسیت‌سنجی، دمای هوا

۱. بخش کشاورزی، دانشگاه پیام نور، تهران، ایران

۲. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

۳. بخش تحقیقات آبیاری و فیزیک خاک، مؤسسه تحقیقات خاک و آب، کرج، ایران

*: مسئول مکاتبات: پست الکترونیکی: hadisiasar@pnu.ac.ir

مقدمه

ایران کشوری با اقلیم خشک و نیمه‌خشک از توزیع نامناسب زمانی و مکانی بارش‌های جوی برخوردار است. در چنین شرایط اقلیمی بارش‌های جوی، جوابگوی نیازهای آبی مصرف‌کنندگان بخش‌های مختلف نبوده و از این‌روست که مدیریت منابع آب محدود کشور به‌ویژه در بخش کشاورزی که بیشترین سهم در بین مصرف‌کنندگان مختلف را داراست، از دغدغه‌های اصلی مدیران بخش آب و کشاورزی کشور است. مدیریت منابع آب از مهم‌ترین عوامل افزایش بهره‌وری اقتصادی در بخش کشاورزی محسوب می‌شود. در شرایط کمبود آب استفاده کارآمد از آب موجود برای تولید پایدار محصولات مورد نیاز در نواحی خشک و نیمه‌خشک کشور از جمله منطقه سیستان ضروری است. تقاضای آب و مصرف آن در اکوسیستم‌های کشاورزی به‌شدت وابسته به پارامترهای اقلیمی است. در برنامه‌ریزی و مدیریت منابع آب برای آبیاری نیاز به تجزیه و تحلیل متغیرهای هیدرولوژیکی نظیر تبخیر-تعرق وجود دارد.

مطالعات زیادی برای تخمین مقدار تبخیر-تعرق با استفاده از روش‌های مختلف در شرایط متفاوت اقلیمی انجام شده است (۲۸، ۳۵ و ۳۷). اشنایدر و همکاران (۳۱) گزارش دادند که برآورد تبخیر-تعرق برای برنامه‌ریزی آبیاری و به‌ویژه برای آبیاری با تناوب زیاد و گیاهان با ریشه سطحی مفید است.

پژوهشگران بسیاری، روش فائو-پنمن-مانتیت را در مقایسه با لایسیمتر به‌عنوان دقیق‌ترین روش برای برآورد تبخیر-تعرق گیاه مرجع معرفی کرده‌اند (۹). این روش می‌تواند در مناطق مختلف با اقلیم و آب‌وهوای متفاوت، مقادیر تبخیر-تعرق قابل اعتمادی ارائه دهد (۳۳). دهقانی سانچ و همکاران (۵) با بررسی شش مدل مختلف برای تخمین تبخیر-تعرق نشان دادند در مناطق نیمه‌خشک ایران روش فائو-پنمن-مانتیت و در مناطق مرطوب روش پنمن بهترین نتایج نزدیک به نتایج لایسیمتر را ارائه می‌کند.

بنابراین در این پژوهش، برای منطقه خشک دشت سیستان

از روش فائو-پنمن-مانتیت به‌عنوان روش مبنا برای ارزیابی‌ها استفاده شد. مهم‌ترین چالش در روش فائو-پنمن-مانتیت، نیاز به تنوع زیادی از پارامترهای هواشناسی است که در مناطق با پارامترهای اندازه‌گیری شده کم، محدودیت ایجاد می‌کند. تجهیز ایستگاه‌های هواشناسی برای اندازه‌گیری این داده‌ها، به‌ویژه در کشورهای در حال توسعه، هزینه زیادی دربر دارد. بنابراین استفاده از روش‌های مناسب محاسبه تبخیر-تعرق که بتواند با بهره‌گیری از پارامترهای قابل اندازه‌گیری اندکی که در بیشتر ایستگاه‌های هواشناسی موجود است، با سرعت مناسب، نتایج قابل قبولی داشته شده باشد، توصیه شده است (۳۲).

تبخیر-تعرق، یک فرایند پیچیده فیزیکی پویا و غیرخطی است و بنابراین تعیین رابطه‌ای که توانایی نمایش همه فرآیندهای دخیل در آن را داشته باشد، دشوار است (۳۸). با پیشرفت تکنولوژی در دهه‌های اخیر ابزارهای یادگیری و روش‌های نوین هوشمند، برای حل برخی از مشکلات، ارائه شده و به‌طور گسترده‌ای استفاده می‌شوند. بنابراین از آنجا که روش شبکه‌های عصبی مصنوعی به‌عنوان یک ابزار هوشمند غیرخطی با داشتن ویژگی‌های تخمین احتمالاتی و تعمیم‌دهی که می‌تواند روابط محاسباتی را برای حل مسائل غیرخطی تطبیق دهد، توانایی زیادی در برآورد تبخیر-تعرق دارد (۲۷ و ۳۶). پرا و همکاران (۲۴) نشان دادند که عملکرد مدل‌های پیش‌بینی بر اساس متغیر دما در میان متغیرهای مختلف هواشناسی، قابل اطمینان‌تر است. به‌طور کلی روش‌های مختلف هوش مصنوعی وجود دارد و انتخاب مناسب‌ترین روش با بهترین عملکرد ضروری است.

در زمینه کاربرد روش‌های هوش مصنوعی در برآورد تبخیر-تعرق مرجع پژوهش‌های متعددی صورت گرفته است (۱۵)، ولی کاربرد روش‌های هوش مصنوعی در پیش‌بینی مقادیر تبخیر-تعرق کمتر مورد توجه قرار گرفته و تحقیقات محدودی در این زمینه موجود است. دیامانتوپولو و همکاران (۶)، عملکرد شبکه‌های عصبی مصنوعی را در تخمین تبخیر-تعرق پتانسیل با حداقل متغیرهای هواشناسی بررسی

را در برآورد تبخیر- تعرق در شرایط محدودیت داده‌های هواشناسی و تنها با استفاده از داده‌های دما و تشعشع نشان داد. در پژوهشی در ساری به‌منظور دستیابی به تخمین تبخیر- تعرق گیاه مرجع، از روش رگرسیون چند متغیره (MLR) و روش‌های نوین هوش مصنوعی شامل، شبکه عصبی پرسپترون چندلایه (MLP) و سیستم استنتاجی فازی- عصبی تطبیقی (Adaptive Neuro-Fuzzy Inference System, ANFIS) استفاده شد. نتایج پژوهش نشان داد زمانی که از چهار پارامتر برای ورودی مدل استفاده شود، مدل‌های شبکه عصبی (MLP) و رگرسیونی (MLR) از دقت و کارایی بالاتری در شبیه‌سازی تبخیر- تعرق گیاه مرجع برخوردارند درحالی که با کاهش پارامتر ورودی به سه و یا کمتر، مدل شبکه عصبی (ANFIS) در مقایسه با دو مدل دیگر دارای دقت بیشتری در برآورد این پارامتر مهم اقلیمی بود (۱۶).

ساگی و جین (۲۹)، با بررسی روش‌های آموزش عمیق، مدل خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادبان بوسستینگ برای محاسبه تبخیر- تعرق، مناسب بودن این روش‌ها و همچنین یک ترکیب الگویی از متغیرهای دمای حداکثر، دمای حداقل، رطوبت نسبی، سرعت باد و تشعشع خورشیدی و ساعات آفتابی را تأیید کردند.

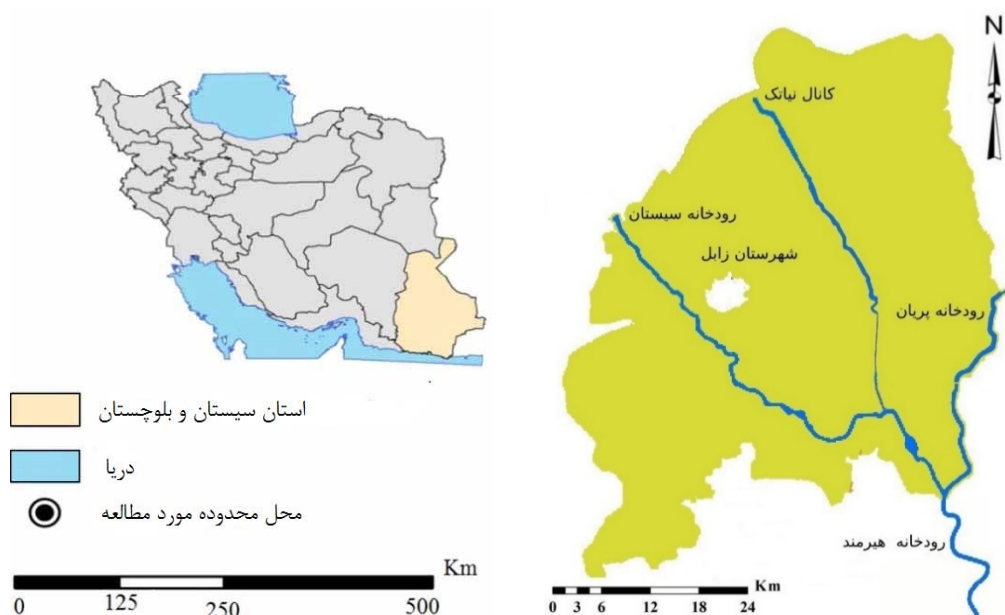
هدف از انجام این پژوهش، شناسایی پارامترهای تأثیرگذار در محاسبه تبخیر- تعرق مرجع روزانه در دشت سیستان و بررسی الگوهای مختلف از ترکیب پارامترهای هواشناسی به‌عنوان ورودی مدل‌های خطی تعمیم‌یافته (Generalized Linear Model, GLM)، جنگل تصادفی (Random Forest, RF) و درخت گرادبان بوسستینگ (Gradient Boosting Trees, GBT) به‌عنوان تکنیک‌های نوین محاسباتی و انتخاب مدل برتر بر اساس الگوی منتخب است.

مواد و روش‌ها

دشت سیستان در شمال استان سیستان و بلوچستان و در مختصات جغرافیایی ۳۰ درجه و ۱۸ دقیقه تا ۳۱ درجه و

کردند. آنها در بررسی خود از داده‌های یک ایستگاه خودکار هواشناسی در شمال یونان استفاده کردند و نتیجه گرفتند که با در نظر گرفتن متوسط، حداکثر و حداقل دما، مدل‌های انتخابی شبکه عصبی مصنوعی، تخمین‌های روزانه تبخیر- تعرق پتانسیل را در مقایسه با معادله هارگریوز اصلاح شده، به‌صورت قابل توجهی بهبود می‌دهند. لو و همکاران (۱۷) با بررسی چهار شبکه عصبی مصنوعی نشان دادند که این روش می‌تواند ابزاری مناسب برای پیش‌بینی تبخیر- تعرق باشد. مطابق تحقیقات آنها، دقت پیش‌بینی تبخیر- تعرق ۷۵/۵۳ تا ۷۸/۱۴ درصد، میانگین خطای مطلق برابر ۰/۹۹ تا ۱/۰۹ میلی‌متر در روز، خطای جذر میانگین مربعات برابر ۰/۸۷ تا ۱/۳۶ میلی‌متر در روز و مقدار ضریب همبستگی بین ۰/۷ تا ۰/۷۵ محاسبه شد.

صیادی و همکاران (۳۰)، عملکرد شبکه‌های تابع پایه شعاعی (Radial Basis Function, RBF) و شبکه‌های پرسپترون چندلایه (Multi Layer Perceptron, MLP) را در برآورد مقدار تبخیر- تعرق مرجع با استفاده از داده‌های هواشناسی سال‌های ۱۳۸۳-۱۳۳۰ در ایستگاه تبریز بررسی کردند و نتیجه گرفتند که تنها با استفاده از دو پارامتر دمای میانگین و سرعت باد به‌عنوان ورودی می‌توان میزان تبخیر- تعرق گیاه مرجع را با استفاده از این دو نوع شبکه با دقت قابل قبولی تخمین زد. کیشی (۱۵)، با مدل‌سازی تبخیر ماهانه در سه ناحیه از ایالت کالیفرنیا، با استفاده از پارامترهای هواشناسی (دمای هوا، سرعت باد، رطوبت، فشار و تابش خورشیدی) و روش شبکه عصبی مصنوعی، تبخیر را برآورد کردند. نتایج آنها نشان داد که شبکه‌های عصبی مصنوعی تابع پایه شعاعی و شبکه‌های پرسپترون چندلایه در مقایسه با روش‌های رگرسیون خطی چندمتغیره (Multiple Linear Regression, MLR) و Stewart-Stephens با معیارهای سنجش آماری مختلف، از کارایی و دقت بالاتری برخوردار است. جین و همکاران (۱۳)، مقدار تبخیر- تعرق را با استفاده از روش شبکه‌های عصبی محاسبه کردند. نتایج تحقیقات آنها ضمن تأیید کارایی شبکه‌های عصبی در تخمین این پدیده، دقت مناسب این روش



شکل ۱. موقعیت جغرافیایی استان سیستان و بلوچستان و دشت سیستان

جدول ۱. میانگین آماری پارامترهای استفاده شده در مدل‌سازی تبخیر- تعرق پتانسیل روزانه (ایستگاه سینوپتیک زابل)

پارامتر آماری	دما (درجه سانتی‌گراد)	رطوبت نسبی (درصد)	سرعت باد (متر بر ثانیه)	بارش (میلی‌متر)	ساعات آفتابی (ساعت)	ارتفاع از سطح دریا (متر)
میانگین	۲۳/۴۵	۲۹/۸۵	۱۳/۳۱	۰/۰۳	۹/۱۱	
میانه	۲۳/۸۰	۲۷/۵۰	۱۳	۰	۹/۷۰	
واریانس	۱۰۲/۴۵	۲۲۸/۵۵	۳۸/۹۲	۰/۰۳۸	۷/۸۲	۴۸۹/۲۰
چولگی	-۲/۳۴	۰/۶۴	۰/۱۹	۷/۱۴	-۱/۵۴	
کشیدگی	-۱/۲۲	-۰/۴۲	-۰/۹۸	۵۳/۸۸	۲/۲۸	

سینوپتیک زابل جمع‌آوری شده است. در این ایستگاه، آمار بلندمدت مؤلفه‌های اقلیمی شامل دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت نسبی حداکثر، رطوبت نسبی حداقل، رطوبت نسبی میانگین، ساعات آفتابی، سرعت باد، بارش و تبخیر از تشت، به‌صورت روزانه در طول دوره آماری ۱۳۸۷ تا ۱۳۹۶ جمع‌آوری شد. جدول (۱)، میانگین متغیرهای استفاده شده برای مدل‌سازی تبخیر- تعرق پتانسیل روزانه در طول دوره آماری برای دشت سیستان را نشان می‌دهد.

داده‌های خام هواشناسی روزانه بین سال‌های مذکور جمع‌آوری و با استفاده از نرم‌افزارهای SPSS، Minitab و

۲۰ دقیقه عرض شمالی و ۶۱ درجه و ۱۰ دقیقه تا ۶۱ درجه و ۵۰ دقیقه طول جغرافیایی شرقی واقع شده است (۲۵) (شکل ۱). این دشت با بارندگی میانگین سالانه ۵۰ میلی‌متر (حدود هفت درصد متوسط بارندگی در جهان) و میزان تبخیر سالانه ۴۰۰۰-۵۰۰۰ میلی‌متر از شرایط نامساعد محیطی برخوردار بود و بر اساس شاخص خشکی دومارتن جزو نواحی فراخشک محسوب می‌شود. از علائم مشخصه دشت سیستان بادهای موسومی ۱۲۰ روزه است که از اواخر ماه اردیبهشت به‌صورت پیوسته و گاهی گسسته می‌وزد. داده‌های استفاده شده در این پژوهش از ایستگاه هواشناسی

در ارتفاع دو متری ($m \ s^{-1}$)، e_s : فشار بخار اشباع (Kpa)، e_a : فشار بخار واقعی (Kpa) است.

متغیرهای ورودی به مدل‌ها

در این تحقیق با توجه به داده‌های ورودی روزانه که شامل شاخص روزانه دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد و تبخیر از تشت هستند، الگوهای مختلفی (۲۵ الگو) که بتواند ترکیبات مختلف از پارامترهای هواشناسی را دربر گیرد، ارائه شد (جدول ۲).

مدل خطی تعمیم‌یافته (GLM)

مدل‌های خطی تعمیم‌یافته، به‌طور کلی ابزارهایی تحلیلی برای انواع داده‌های مختلف هستند و روابط ریاضی آن توسط نلد و باکر (۲۲)، تهیه شد. تئوری این مدل توسط مک کالاک (۲۱)، توسعه داده شد و به‌وسیله چندلر و ویاتر (۴) برای موضوعات هیدرولوژی و هواشناسی کاربردی شد. مدل‌های خطی تعمیم‌یافته شامل طیف گسترده‌ای از مدل‌های آماری نظیر رگرسیون خطی برای پاسخ توزیع شده نرمال، مدل‌های لجستیک برای داده‌های باینری، مدل‌های خطی برای داده‌های شمارشی و همچنین بسیاری از مدل‌های آماری مفید از طریق فرم‌سازی مدل کلی آن می‌شود. مدل‌های خطی تعمیم‌یافته می‌تواند برای مواقعی که مشاهدات توزیع نرمالی ندارند و زمانی که سایر روش‌های مدل رگرسیون مناسب نیستند، به‌کار رود. عملکرد این مدل، در بین روش‌های معمول مدل‌سازی، قابل قبول است (۲۶). بسیاری از پژوهشگران از این روش در مدل‌سازی و پیش‌بینی‌های بارش استفاده کرده‌اند (۴ و ۲۳).

مدل جنگل تصادفی (RF)

مدل جنگل تصادفی در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است و برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با سرعت بالایی انجام می‌دهد. این روش که بیشتر

MATLAB (R2017b) پیش پردازش‌های لازم به‌منظور حذف یا اصلاح داده‌های پرت و کامل کردن مجموعه داده‌ها انجام شد. در این تحقیق، روش مبنا برای محاسبه مقدار تبخیر-تعرق روزانه روش فائو-پنمن-مانتیت است. از بین کل داده‌های موجود، حدود ۸۰ درصد برای آموزش و ۲۰ درصد بقیه برای آزمون مدل‌های منتخب اختصاص یافت. ترکیب‌های متنوعی از داده‌های هواشناسی به‌عنوان ورودی مدل‌ها و مقادیر محاسبه شده متناظر، به‌عنوان خروجی مدل‌ها در نظر گرفته شد (جدول ۲). نتایج حاصل از الگوهای ترکیبات مختلف، بررسی شده و بهترین سناریوی انتخابی، برای پیش‌بینی مقادیر تبخیر-تعرق، وارد مدل‌ها شد.

معادله پنمن-مانتیت

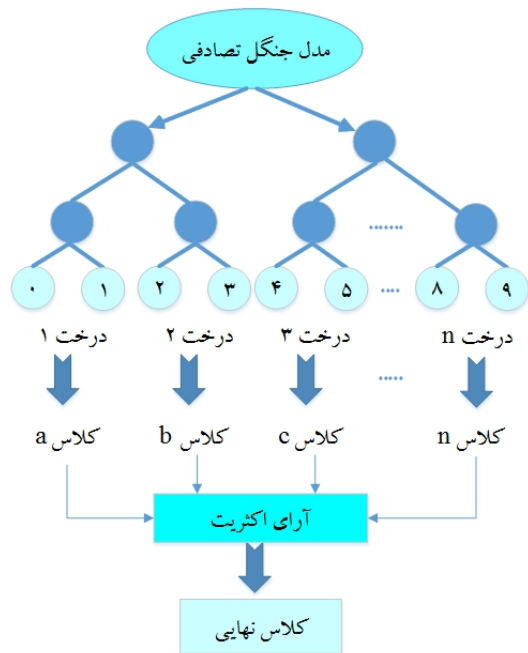
روش فائو-پنمن-مانتیت، تکامل یافته روش ترکیبی پنمن-مانتیت ۱۹۶۵ است که توسط آلن و همکاران (۱) برای محاسبه تبخیر-تعرق در نشریه ۵۶ سازمان فائو ارائه شد. این روش با دقت بالا در محدوده وسیعی از مناطق و اقلیم‌ها برآورد صحیحی از تبخیر-تعرق گیاه مرجع ارائه می‌کند و از سوی سازمان خوار و بار جهانی (FAO) به‌عنوان روش استاندارد برای محاسبه تبخیر-تعرق گیاه مرجع از روی داده‌های هواشناسی و اقلیمی و همچنین برای ارزیابی سایر روش‌ها پیشنهاد شده است. در این تحقیق، به‌منظور برآورد مقدار تبخیر-تعرق مرجع روزانه (میلی‌متر بر روز) از روش فائو-پنمن-مانتیت مطابق معادله زیر استفاده شد (۱):

$$E_{PM56} = \frac{0.408 \Delta (R_n - G)}{\Delta + \gamma (1 + 0.34 u_2)} + \frac{90.0 \gamma}{[\Delta + \gamma (1 + 0.34 u_2)]} \frac{u_2 (e_s - e_a)}{(T + 273)} \quad (1)$$

در رابطه فوق E_{PM56} : تبخیر-تعرق گیاه مرجع ($mm \ day^{-1}$)، Δ : شیب فشار منحنی فشار بخار اشباع ($Kpa \ C^{-1}$)، R_n : تابش خالص در سطح پوشش گیاهی ($MJ \ m^{-2} \ d^{-1}$)، G : چگالی شار گرمای خاک ($MJ \ m^{-2} \ d^{-1}$)، γ : ضریب سایکرومتری رطوبتی ($Kpa \ C^{-1}$)، T_{mean} : میانگین دمای روزانه ($^{\circ}C$)، u_2 : سرعت باد

جدول ۲. الگوهای به کار برده شده در مدل‌های پیشنهادی برای برآورد تبخیر-تعرق پتانسیل

تعداد متغیرها	پارامترهای ورودی شبکه	الگو
۱۰	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد، تبخیر از تشت	M1
۹	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد، تبخیر از تشت	M2
۹	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، ساعات آفتابی، سرعت باد، تبخیر از تشت	M3
۸	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، ساعات آفتابی، تبخیر از تشت	M4
۸	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، سرعت باد، تبخیر از تشت	M5
۷	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداقل، رطوبت میانگین، سرعت باد، تبخیر از تشت	M6
۶	دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت میانگین، سرعت باد، تبخیر از تشت	M7
۵	دمای حداکثر، دمای میانگین، رطوبت میانگین، سرعت باد، تبخیر از تشت	M8
۴	دمای میانگین، رطوبت میانگین، سرعت باد، تبخیر از تشت	M9
۴	دمای حداکثر، رطوبت میانگین، سرعت باد، تبخیر از تشت	M10
۳	دمای حداکثر، رطوبت میانگین، تبخیر از تشت	M11
۳	دمای حداکثر، سرعت باد، تبخیر از تشت	M12
۲	دمای حداکثر، سرعت باد	M13
۱	دمای حداکثر	M14
۱	سرعت باد	M15
۱	تبخیر از تشت	M16
۲	دمای حداکثر، دمای حداقل	M17
۱	رطوبت میانگین	M18
۲	دمای حداکثر، تبخیر از تشت	M19
۲	دمای حداکثر، رطوبت میانگین	M20
۲	دمای حداکثر، رطوبت حداکثر	M21
۲	دمای میانگین، سرعت باد	M22
۲	دمای حداکثر، سرعت باد	M23
۱	ساعات آفتابی	M24
۱	دمای حداقل	M25



شکل ۲. ساختار و اجزای مدل جنگل تصادفی (۲۹)

که برای ارائه پیش‌بینی‌های خروجی نسبت به دیگر الگوریتم‌ها کندتر عمل می‌کند.

در این پژوهش، به‌منظور انتخاب مناسب‌ترین خط برازش‌یافته توسط این روش در احتمالات متفاوت، متغیر اندازه گره (Node size) که نشان‌دهنده تعداد برگ‌ها در هر شاخه است، با آزمون و خطا تعیین شد.

مدل درخت گرادبان بوس‌تینگ (GBT)

روش درخت گرادبان بوس‌تینگ جزء بهترین الگوریتم‌های یادگیری است و برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با دقت بالایی انجام می‌دهد. در این روش، درخت‌ها یکی پس از دیگری آموزش داده می‌شوند. هر درخت زیرمجموعه، در درجه اول با داده‌هایی که به اشتباه توسط درخت پیشین پیش‌بینی شده‌اند، آموزش داده می‌شوند (۱۸). این امر موجب می‌شود مدل بیشتر روی موارد پیچیده و کمتر بر مسائلی که پیش‌بینی در آنها آسان است، متمرکز شود. بنابراین این روش نسبت به بسیاری از روش‌ها مانند روش رگرسیون خطی و روش دسته‌بندی (bagging) نتایج بهتری دارد (۳). ایده اولیه

برای پیش‌بینی تشعشع خورشیدی ساعتی شناخته شده است (۲۹)، اولین بار توسط بریمن (۳)، ارائه شد. این روش، بر خلاف مدل‌های کلاسیک مانند رگرسیون که تنها بر پایه یک مدل تکیه دارند با استفاده از تعداد زیادی درخت تصمیم، از اطلاعات بیشتری در داده‌ها استفاده می‌کند. سپس تمام درختان با هم، برای پیش‌بینی ترکیب می‌شوند، تا بتوان استنباط بهتری از متغیرها داشت. این روش، یک مدل ناپارامتری است و الگوریتم آن روش دسته‌بندی را به کار می‌گیرد (۳). در شکل (۲)، ساختار و اجزای این مدل نشان داده شده است.

جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن n داده مشاهده‌ای مستقل ساخته می‌شود:

$$(Y_i, X_i), i=1, 2, \dots, n \quad (2)$$

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چند نمونه بوت استرپ (bootstrap) از داده‌ها شرکت دارند و در ساخت هر درخت به‌طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. در طی فرآیند نمونه‌گیری حدود یک سوم از داده‌ها نمونه‌گیری نمی‌شوند و به‌عنوان نمونه خارج از کیسه در نظر گرفته می‌شوند و از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد ناریب خطا استفاده می‌شود. سپس روی هر نمونه بوت‌استرپ، یک درخت گسترش داده می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به‌صورت تصادفی، m متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت m/M برابر با یک‌سوم است و برای کلاسه‌بندی برابر با $m = \sqrt{M}$ پیشنهاد شده است. پس از ساخت تمام درخت داده‌های مرحله آزمون، به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به‌دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها، مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش درخت رگرسیون جنگل تصادفی به‌ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد، یک روش پیش‌بینی کارآمد است. مشکل این الگوریتم آن است

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{n}} \quad (5)$$

در این روابط، O_i : مقادیر واقعی، \bar{O}_i : میانگین مقادیر واقعی، P_i : مقادیر تبخیر- تعرق برآوردشده، \bar{P}_i : میانگین مقادیر برآوردشده، N : تعداد داده‌ها است.

تحلیل حساسیت

هدف از تحلیل حساسیت در این پژوهش شناسایی و تعیین درجه اهمیت هر یک از پارامترها در سری داده‌های ورودی بر عملکرد دینامیکی سیستم است. به منظور آنالیز حساسیت در هر سری داده ورودی، یک پارامتر حذف و مدل با همان داده‌های سری آموزش و آزمون، اجرا و مقادیر ضریب همبستگی و خطا تعیین شد. با حذف هر پارامتر هر چه خطای مدل افزایش بیشتری نشان دهد، بیانگر تأثیر بیشتر آن پارامتر بر فرایند شبیه‌سازی است.

برای انتخاب مناسب‌ترین مدل بر اساس معیارهای خطای ارائه شده در قسمت قبل، از روش رتبه‌بندی برای داده‌های آزمون استفاده شد. در این روش به هر یک از مدل‌ها بر اساس معیارهای ذکر شده رتبه‌ای داده می‌شود و در پایان پس از مقایسه مجموع رتبه‌های به‌دست آمده، مناسب‌ترین مدل برای تخمین بهترین معادله انتخاب می‌شود. شیوه رتبه‌بندی این مدل‌ها بدین صورت است که در هر روش، مدلی که کمترین میزان خطای RMSE را دارد، پایین‌ترین رتبه (رتبه ۱) و به مدل‌ها یا ترکیبات ورودی دیگر در هر یک از دسته‌های ورودی بر حسب میزان معیار خطاهای مذکور، رتبه‌های دو الی آخر تعلق می‌گیرد. اما در مورد معیارهای ضریب همبستگی (R)، مدل یا ترکیبی که بیشترین مقدار این معیارها را دارا است، رتبه یک اختصاص می‌یابد. همچنین به مدلهایی که میزان معیار خطای آنها یکسان باشند، رتبه یکسان تعلق می‌گیرد. در نهایت پس از انتخاب مدل‌ها یا ترکیبات مناسب هر روش و مقایسه بین آنها، روش برتر در برآورد تبخیر- تعرق مرجع در منطقه مورد مطالعه انتخاب می‌شود.

این روش توسط بریمن (۲) این روش را به‌عنوان روشی که می‌تواند الگوریتم بهینه‌سازی را روی یک تابع هزینه مناسب تفسیر کند، ارائه شد.

ارزیابی عملکرد مدل‌ها

بررسی نحوه عملکرد و مقایسه قدرت پیش‌بینی به‌منظور مقایسه عملکرد مدل‌ها لازم و ضروری است. در این مطالعه، علاوه بر ترسیم نمودارهای مقادیر واقعی در برابر مقادیر پیش‌بینی شده توسط مدل‌ها، شاخص‌های کمی ضریب همبستگی (R ، Correlation coefficient)، میانگین خطای مطلق (MAE ، Mean Absolute Error) و خطای جذر میانگین مربعات ($RMSE$ ، Root Mean Square Error) به‌عنوان معیارهای ارزیابی به‌کار گرفته شد (روابط ۳ تا ۵).

الف) ضریب همبستگی: بیان‌کننده همبستگی بین مقادیر برآورد شده مدل و داده‌های واقعی (روش فائو- پنمن - مانیتث) است (رابطه ۳). بر این اساس هر چه مقدار آن به یک نزدیک‌تر باشد، نشان‌دهنده نزدیکی بیشتر مقادیر برآورد شده به مقادیر واقعی است (۷). در واقع ضریب همبستگی ارتباط بین دو متغیر را اندازه‌گیری می‌کند و یک ابزار ریاضی است که در پایه‌ریزی تحلیل‌های اقلیمی کاربرد زیادی دارد (۳۴). اما با توجه به اینکه این ضریب به داده‌های خارج از محدوده حساسیت زیادی دارد، از این روش نمی‌تواند به‌تنهایی برای ارزیابی یک مدل به‌کار رود (۷).

ب) میانگین خطای مطلق: برای نشان دادن میزان همخوانی میان داده‌های برآورد شده مدل و داده‌های واقعی است. هر چه این شاخص که بدون علامت است، به صفر نزدیک‌تر باشد، نشان‌دهنده انحراف کمتر و دقت بالاتر مدل است (۱۰).

ج) خطای جذر میانگین مربعات: این پارامتر نشان‌دهنده میزان خطای مدل است که بهترین مقدار آن صفر است.

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^N (|O_i - P_i|)}{N} \quad (4)$$

نتایج و بحث

با محاسبه مقادیر تبخیر-تعرق مرجع با روش فائو-پنمن-مانتیت و در نظر گرفتن این مقادیر به‌عنوان مقادیر هدف در آموزش، شبکه مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادبان بوستینگ ایجاد و عملکرد آنها بررسی شد.

نتایج پارامترهای آماری مدل‌های پیشنهادی برای انتخاب بهترین الگو در جدول (۳) ارائه شده است. ارزیابی کلی نتایج حاصل از مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادبان بوستینگ و همچنین شاخص‌های آماری محاسبه‌شده، نشان‌دهنده توانایی این مدل‌های هوشمند در شبیه‌سازی میزان تبخیر-تعرق روزانه منطقه مورد مطالعه است.

بررسی مقایسه مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادبان بوستینگ نشان می‌دهد که می‌توان یک الگوی مشخصی برای دشت سیستان را با دقت مناسب برای برآورد تبخیر-تعرق پتانسیل روزانه معرفی کرد و باید برای هر مدل، الگوی برتر را معرفی کرد. مهم‌ترین چالش‌های یک مدل یادگیری ماشین، انتخاب بهترین الگوریتم ممکن، انتخاب متغیرهای نشانگر مناسب و در دسترس بودن مجموعه داده‌های ورودی است (۱۰).

پس از طراحی اجزای مختلف مدل‌های بالا، خروجی این مدل‌ها دریافت شد و با شاخص‌های آماری ذکر شده مورد ارزیابی قرار گرفت. نتایج هر یک از مدل‌ها در جدول (۳) دیده می‌شود. با تعیین بهترین مدل مناسب، ترکیب متغیرهای متفاوت در مدل خطی تعمیم‌یافته بررسی شد و الگوهای M_1 و M_3 با بالاترین ضریب همبستگی $R=0/95$ و کمترین خطا $RMSE=1/12$ به‌عنوان بهترین الگوها انتخاب شدند (جدول ۳).

در مدل جنگل تصادفی با الگوهای مختلف، بهترین ترکیب، الگوی M_{10} با متغیرهای دمای حداکثر، رطوبت میانگین، سرعت باد و تبخیر از تشت بود که با بیشترین ضریب همبستگی $R=0/981$ و کمترین خطای $RMSE=0/073$ به‌عنوان بهترین الگو در این مدل برگزیده شد. همچنین در رتبه‌های بعدی به ترتیب الگوهای M_8 ، M_3 و M_1 در حد قابل

قبول بودند. مهم‌ترین پارامترها به ترتیب شامل دمای میانگین، دمای حداکثر، سرعت باد، دمای میانگین و تبخیر از تشت بودند. گراناتا (۱۰) با بررسی الگوریتم‌های مختلف یادگیری ماشین شامل الگوریتم‌های درخت تصمیم M_5P ، روش دسته‌بندی (bagging)، روش جنگل تصادفی و روش رگرسیون بردار پشتیبانی (Support Vector Regression, SVR) و همچنین الگوهای مختلف متغیرهای هواشناسی و مقایسه آنها با روش فائو-پنمن-مانتیت در منطقه‌ای با اقلیم مرطوب، استفاده از روش جنگل تصادفی را برای برآورد تبخیر-تعرق رضایت‌بخش دانست. این مدل برای الگوهای ساده‌تر با متغیرهای هواشناسی اندک مانند الگوی شامل تشعشع خورشیدی خالص، سرعت باد، دمای میانگین و رطوبت نسبی میانگین، و همچنین الگوی تشعشع خورشیدی خالص، رطوبت نسبی میانگین و دمای میانگین نتایج بهتر از سایر روش‌ها ارائه داد. درحالی که برای سایر الگوهای پیچیده‌تر مانند الگوی تشعشع خورشیدی خالص، فلاکس گرمای محسوس، رطوبت خاک، سرعت باد، رطوبت نسبی میانگین و دمای میانگین، نتایج سایر الگوریتم‌ها مناسب‌تر بود.

فنگ و همکاران (۸) برای منطقه‌ای خشک در چین نشان داد که روش جنگل تصادفی و روش شبکه‌های عصبی رگرسیونی تعمیم‌یافته (Generalized Regression Neural Networks, GRNN)، نتایج مناسبی در تخمین تبخیر-تعرق روزانه داشتند و نتایج روش جنگل تصادفی اندکی مناسب‌تر بود.

در رابطه با مدل درخت گرادبان بوستینگ، الگوی M_1 با بیشترین ضریب همبستگی $R=0/993$ و کمترین خطا $RMSE=0/633$ به‌عنوان بهترین الگو در این مدل برگزیده شد و الگوهای M_3 و M_5 در رده‌های بعدی قرار گرفتند. نمودار پراکندگی مقادیر برآورده شده در برابر مقادیر مشاهداتی در مرحله آموزش و آزمون بر اساس نوع مدل در شکل (۳) ارائه شده است.

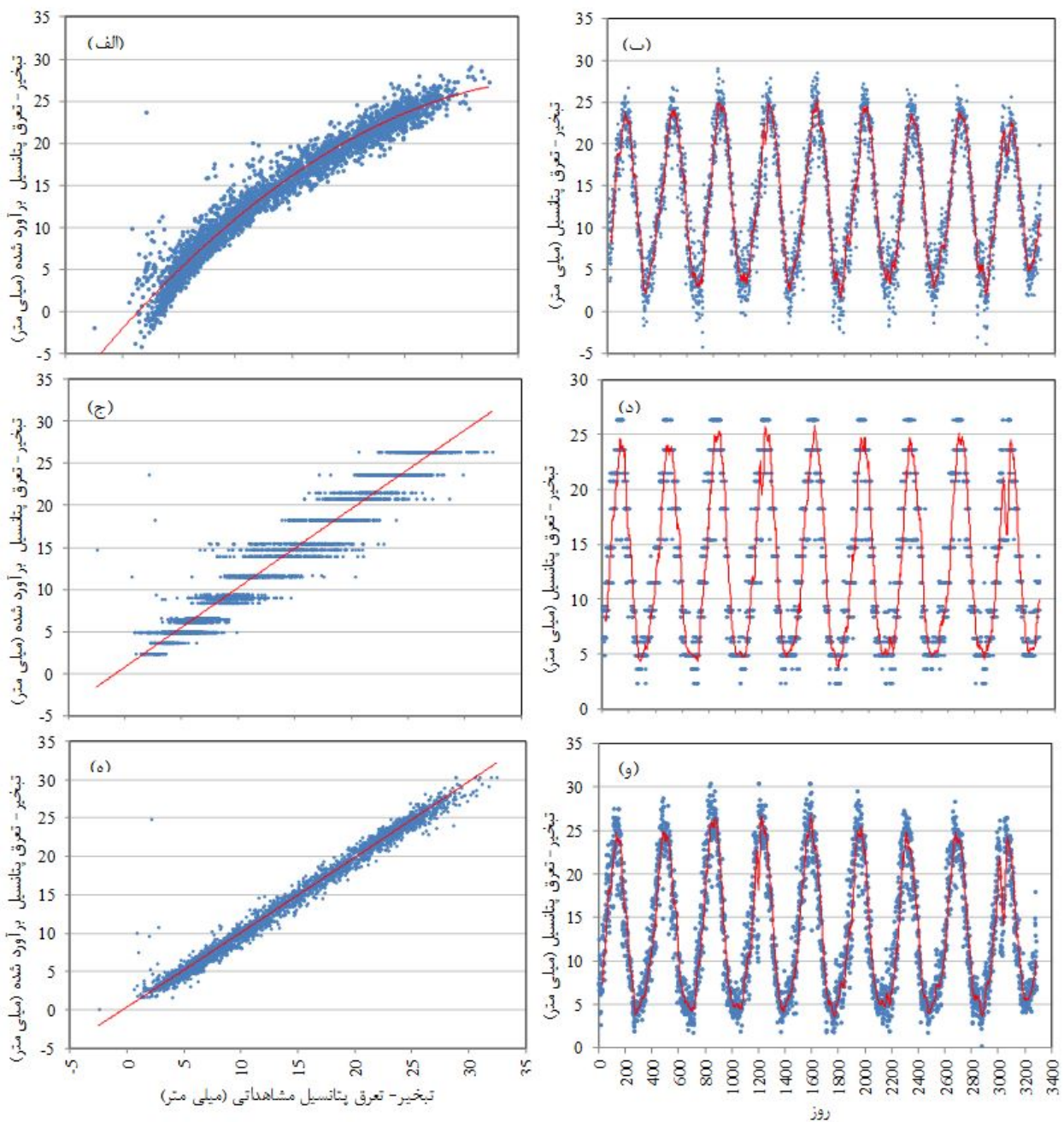
نتایج همبستگی بالای بین تبخیر-تعرق برآورده‌شده از

جدول ۳. ضریب همبستگی و خطای الگوهای مختلف در مدل‌های پیشنهادی در ایستگاه مورد مطالعه

مدل پیشنهادی								
درخت گردیان بوستینگ			جنگل تصادفی			مدل خطی تعمیم‌یافته		
شاخص‌های آماری								
الگو	ضریب همبستگی	خطای مطلق	خطای جذر میانگین مربعات	ضریب همبستگی	خطای مطلق	خطای جذر میانگین مربعات	ضریب همبستگی	خطای مطلق
M1	۰/۹۵۰	۱/۳۴۲	۱/۷۱۲	۰/۹۷۱	۰/۹۳۹	۱/۱۸۷	۰/۹۹۳	۰/۴۵۱
M2	۰/۹۴۸	۱/۳۶۰	۱/۷۵۱	۰/۹۷۲	۰/۹۲۳	۱/۲۹۲	۰/۹۹۰	۰/۴۹۵
M3	۰/۹۵۰	۱/۳۴۲	۱/۷۱۲	۰/۹۷۷	۰/۸۳۹	۱/۱۸۷	۰/۹۹۳	۰/۴۵۱
M4	۰/۹۱۲	۱/۶۴۴	۲/۲۷۴	۰/۹۲۷	۱/۳۹۲	۲/۰۷۵	۰/۹۳۲	۱/۳۴۰
M5	۰/۹۵۰	۱/۳۴۵	۱/۷۱۵	۰/۹۷۶	۰/۸۴۵	۱/۱۹۵	۰/۹۹۳	۰/۴۶۲
M6	۰/۹۴۵	۱/۴۰۲	۱/۷۹۴	۰/۹۷۴	۰/۸۸۴	۱/۲۵۹	۰/۹۹۳	۰/۴۶۸
M7	۰/۹۴۵	۱/۴۰۱	۱/۷۹۴	۰/۹۷۴	۰/۸۷۴	۱/۲۴۹	۰/۹۷۴	۰/۴۷۴
M8	۰/۹۴۴	۱/۴۱۴	۱/۸۱۵	۰/۹۸۱	۰/۸۰۰	۱/۰۸۲	۰/۹۹۲	۰/۴۹۰
M9	۰/۹۳۳	۱/۵۰۸	۱/۹۷۷	۰/۹۷۹	۰/۸۲۷	۱/۱۲۴	۰/۹۸۷	۰/۵۹۳
M10	۰/۹۴۵	۱/۴۰۵	۱/۸۰۳	۰/۹۸۱	۰/۸۰۰	۱/۰۷۳	۰/۹۹۳	۰/۴۷۴
M11	۰/۸۹۸	۱/۸۱۱	۲/۴۴۶	۰/۹۱۲	۱/۵۹۴	۲/۲۶۷	۰/۹۱۴	۱/۵۵۸
M12	۰/۹۳۳	۱/۵۸۹	۱/۹۸۷	۰/۹۶۵	۱/۱۰۶	۱/۴۴۹	۰/۹۷۱	۰/۹۹۸
M13	۰/۹۱۷	۱/۸۳۱	۲/۲۱۲	۰/۹۶۴	۱/۱۱۳	۱/۴۴۶	۰/۹۶۷	۱/۰۶۵
M14	۰/۷۶۳	۲/۹۹۳	۳/۷۳۵	۰/۷۸۸	۲/۶۶۶	۳/۵۲۲	۰/۸۳۷	۲/۶۵۶
M15	۰/۵۰۱	۴/۳۰۲	۵/۴۱۸	۰/۵۰۳	۴/۲۸۵	۵/۴۰۸	۰/۵۰۱	۴/۲۸۹
M16	۰/۸۲۵	۲/۴۱۴	۳/۲۰۰	۰/۸۳۳	۲/۳۳۴	۳/۱۲۵	۰/۸۳۷	۲/۲۹۲
M17	۰/۸۲۵	۲/۵۳۲	۳/۱۸۳	۰/۸۷۹	۱/۹۳۸	۲/۶۷۱	۰/۸۸۱	۱/۹۶۲
M18	۰/۵۵۹	۳/۹۴۵	۵/۰۹۰	۰/۷۳۷	۲/۹۶۸	۳/۹۲۷	۰/۷۳۹	۲/۹۷۳
M19	۰/۸۸۸	۱/۹۱۲	۲/۵۷۱	۰/۸۹۳	۱/۸۲۰	۲/۵۰۵	۰/۸۹۵	۱/۷۷۷
M20	۰/۸۰۶	۲/۶۷۷	۳/۳۷۶	۰/۸۷۲	۱/۹۵۹	۲/۷۳۷	۰/۸۷۲	۱/۹۹۳
M21	۰/۸۳۹	۲/۴۱۸	۳/۰۷۴	۰/۸۸۴	۱/۸۵۴	۲/۶۰۳	۰/۸۸۴	۱/۸۹۶
M22	۰/۹۰۲	۱/۹۵۶	۲/۴۰۰	۰/۹۵۴	۱/۲۷۱	۱/۶۴۷	۰/۹۵۶	۱/۲۳۱
M23	۰/۹۱۷	۱/۸۳۱	۲/۲۱۲	۰/۹۶۴	۱/۱۱۳	۱/۴۴۶	۰/۹۶۷	۱/۰۶۵
M24	۰/۲۸۹	۵/۵۶۴	۶/۴۶۵	۰/۴۵۶	۴/۴۵۳	۵/۶۴۴	۰/۴۶۰	۴/۴۵۰
M25	۰/۸۱۲	۲/۶۱۲	۳/۲۲۲	۰/۸۵۶	۲/۱۱۲	۲/۹۰۱	۰/۸۶۰	۲/۰۸۶

زمانی که داده‌های هواشناسی محدودی در دسترس باشند، استفاده کرد. جابلون و سالی (۱۲) مقدار تبخیر- تعرق

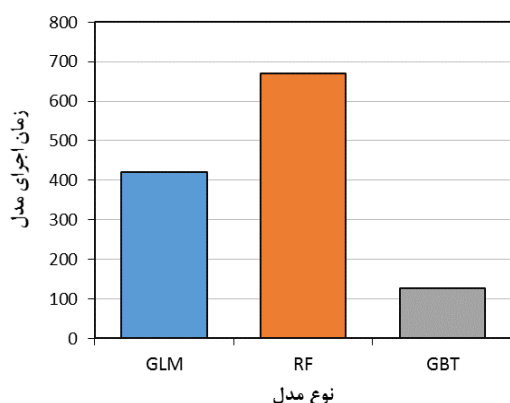
مدل‌های پیشنهادی با روش فائو- پنمن- مانیتث نشان داد که می‌توان از این مدل‌ها برای برآورد تبخیر- تعرق روزانه در



شکل ۳. همبستگی و پراکنش بین مدل فانو- پنمن- مانیت با (الف و ب) مدل درخت گرادبان بوستینگ (GBT)، (ج و د) مدل خطی تعمیم‌یافته (GLM) و (ه و و) مدل جنگل تصادفی (RF) (الگوی یک)

استفاده از داده‌های روزهای قبل (previous days strategy) و نیز روش کلاسترینگ (clustering) بهره‌گیری کرد. نکته دیگر در بحث مدل‌سازی با مدل‌های مذکور، زمان مورد نیاز برای محاسبات شبیه‌سازی است. با توجه به اینکه در برآورد تبخیر- تعرق روزانه، تغییرات کوتاه‌مدت پارامترهای

محاسباتی را با استفاده از دو مجموعه محدود و کامل از داده‌های هواشناسی در تونس مقایسه کردند و تفاوت بین نتایج آنها برای مناطق مورد مطالعه اندک بود. فریرا و همکاران (۹) نشان دادند که در زمان محدودیت داده‌های ورودی برای مدل‌های شبکه عصبی مصنوعی و SVM می‌توان از روش



شکل ۴. زمان اجرای مدل‌های خطی تعمیم‌یافته (GLM)، جنگل تصادفی (RF) و درخت گرادیان بوس‌تینگ (GBT) (زمان بر حسب ثانیه)

بر اساس نتایج، در مجموع پارامتر دما به‌عنوان مهم‌ترین پارامتر مؤثر برای هر سه مدل بررسی شده به‌دست آمد. بنابراین افزایش دقت و صحت داده‌های دما، می‌تواند به افزایش دقت مدل‌های استفاده‌شده کمک کند. لو و همکاران (۱۷) با مطالعه پارامترهای مختلف هواشناسی در تخمین تبخیر-تعرق توسط مدل‌های شبکه عصبی مصنوعی نتایج مشابهی مبنی بر اهمیت صحت داده‌های دما بر دقت نتایج پیش‌بینی شده را گزارش کردند. از طرف دیگر، مدل‌هایی که بر اساس دما هستند، به‌علت اندازه‌گیری این پارامتر در بیشتر ایستگاه‌های هواشناسی، مدل‌هایی با اهمیت و مقبولیت بیشتر به‌شمار می‌آیند (۱۹ و ۲۰). تراژوره و همکاران (۳۶)، گزارش دادند که مدل شبکه عصبی مصنوعی بر پایه دما می‌تواند عملکرد بهتری نسبت به روش تجربی هارگریوز داشته باشد. همچنین نتایج پژوهش حقیقت‌جو و آخوند علی (۱۱) نشان داد که روش‌های مبتنی بر دمای هوا، تبخیر-تعرق دشت سیستان را بهتر برآورد می‌کند. بنابراین در مدل درخت گرادیان بوس‌تینگ، بهترین الگوی الگوی $M1$ به دلیل ضریب همبستگی بالا $R = 0.993$ و خطای کم $RMSE = 0.633$ در منطقه مورد مطالعه انتخاب شد. همچنین بهترین الگوهای این مدل در رتبه‌های بعدی به ترتیب $M5$ و $M3$ بودند.

هواشناسی مؤثر بر تبخیر-تعرق در نظر گرفته می‌شود، انتخاب پارامترهای مؤثر و در صورت امکان حذف برخی از آنها به‌علت کاهش متغیرهای ورودی، می‌تواند کمک زیادی در برآورد سریع و کم‌هزینه تبخیر-تعرق روزانه و فرایند مدل‌سازی کند. کاهش زمان محاسبات به لحاظ صرفه‌جویی در وقت برای کاربر بسیار مهم است (۱۴) و در میان مدل‌های منتخب همان‌طور که در شکل (۴) دیده می‌شود، کمترین زمان اجرا مربوط به مدل درخت گرادیان بوس‌تینگ بود. در مقایسه با مدل‌های خطی تعمیم‌یافته و جنگل تصادفی، استفاده از مدل درخت گرادیان بوس‌تینگ توانست به ترتیب ۷۰ و ۸۱ درصد، زمان اجرای مدل را کاهش دهد.

مدل درخت گرادیان بوس‌تینگ بلافاصله مقادیر خروجی را شبیه‌سازی می‌کند و نیازی به شکل محاسباتی ندارد که یک برتری نسبت به سایر مدل‌های مشابه است. با این حال باید توجه کرد که هر چند مدل‌های داده محور، نیازی به معادلات با پایه‌های فیزیکی که بیشتر به صورت غیرخطی و پیچیده‌اند، ندارند، اما درک فرایندها و تأثیر عوامل مورد نظر در داده‌های ورودی، روی خروجی‌های آن بسیار مهم بود و در صورت عدم این درک، نمی‌توان انتظار نتایج مطلوبی را داشت.

تحلیل حساسیت

نتایج نشان داد در مدل خطی تعمیم‌یافته مهم‌ترین پارامترهای هواشناسی به ترتیب دمای حداکثر، سرعت باد، دمای میانگین، تبخیر-تعرق و تابش و در مدل جنگل تصادفی به ترتیب پارامترهای دمای میانگین، دمای حداکثر، سرعت باد، دمای میانگین و تبخیر از تشت هستند. جدول (۳) مقادیر همبستگی و خطا را با حذف هر یک از پارامترهای ورودی نشان می‌دهد. با حذف هر پارامتر، مقدار بیشتر خطا و ضریب همبستگی کمتر در مرحله آزمون نشان می‌دهد که تأثیر آن پارامتر از دیگر پارامترها بیشتر بوده است. همچنین در مدل درخت گرادیان بوس‌تینگ، مهم‌ترین پارامترها در الگوسازی تبخیر-تعرق مرجع، به ترتیب شامل دمای میانگین، سرعت باد و دمای حداکثر بودند.

نتیجه گیری

بوستینگ توانست به ترتیب ۷۰ و ۸۱ درصد، زمان اجرای مدل را در مقایسه با مدل‌های خطی تعمیم‌یافته و جنگل تصادفی کاهش دهد که از امتیازات دیگر این روش محسوب می‌شود. همچنین نتایج به دست آمده نشان داد که می‌توان مقدار تبخیر-تعرق گیاه مرجع در منطقه مورد مطالعه را با استفاده از تعداد پارامترهای کمتری نسبت به آنچه در روش فائو-پنمن-مانتیت به کار می‌رود، با دقت قابل قبولی برآورد کرد. از بین تمامی پارامترها در مدل درخت گرادیان بوستینگ به ترتیب دمای میانگین، سرعت باد، دمای حداکثر بیشترین تأثیر و اهمیت را در میزان تبخیر-تعرق روزانه داشتند. از این‌رو استفاده از این پارامترها برای تعیین تبخیر-تعرق روزانه توصیه می‌شود. همچنین با توجه به اهمیت پارامتر دما در الگوهای انتخاب شده، ضرورت استفاده از داده‌های با دقت و صحت مناسب برای این پارامتر در مدل‌سازی با روش‌های پیشنهادی ضروری است. از نتایج پژوهش حاضر می‌توان در بهینه‌سازی شبکه‌های آبیاری و زهکشی و منابع آبی در مناطق خشک استفاده کرد. پیشنهاد می‌شود نتایج مدل‌های پیشنهادی در این پژوهش در سایر اقلیم‌های کشور ارزیابی شود.

در این پژوهش دقت مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادیان بوستینگ در تخمین تبخیر-تعرق پتانسیل روزانه دشت سیستان بررسی شد. در پیش‌بینی‌های تبخیر-تعرق روزانه در مدل‌های مورد مطالعه، از الگوهای متفاوت شامل ۲۵ الگوی مختلف از داده‌های هواشناسی استفاده شد. برای این منظور داده‌های دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد و تبخیر از تشت به عنوان پارامترهای ورودی و تبخیر-تعرق پتانسیل روزانه به عنوان پارامتر خروجی مدل انتخاب شد. مقادیر برآورد شده از هر یک از مدل‌ها با روش فائو-پنمن-مانتیت مقایسه شدند. مدل‌سازی به صورت روزانه برای سال‌های ۱۳۸۷ تا ۱۳۹۶ و ارزیابی مدل‌ها با استفاده از ضریب همبستگی (R) و مقادیر خطای جذر میانگین مربعات (RMSE) انجام شد. بررسی نتایج مدل‌ها در ایستگاه نشان داد که الگوی M1 و مدل درخت گرادیان بوستینگ از دقت بالاتری ($R = 0.993$ و $RMSE = 0.633$)، برای مدل‌سازی تبخیر-تعرق پتانسیل مرجع برخوردارند. استفاده از مدل درخت گرادیان

منابع مورد استفاده

- Allen, R. G., L. S. Pereira, D. Raes and M. Smith. 1998. Crop evapotranspiration Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome 300(9): D05109.
- Breiman, L. 1999. Using adaptive bagging to debias regressions. Technical Report 547, Statistics Dept. UCB.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32.
- Chandler, R. E. and H. S. Wheater. 2002. Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resources Research* 38(10): 10-1.
- DehghaniSanij, H., T. Yamamoto and V. Rasiyah. 2004. Assessment of evapotranspiration estimation models for use in semi-arid environments. *Agricultural Water Management* 64(2): 91-106.
- Diamantopoulou, M., P. Georgiou and D. Papamichail. 2011. Performance evaluation of artificial neural networks in estimating reference evapotranspiration with minimal meteorological data. *Global NEST Journal* 13(1): 18-27.
- Dou, X. and Y. Yang. 2018. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture* 148: 95-106.
- Feng, Y., N. Cui, D. Gong, Q. Zhang and L. Zhao. 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management* 193: 163-173.
- Ferreira, L. B., F. F. Da Cunha, R. A. De Oliveira and E. I. Fernandes Filho. 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—a new approach. *Journal of Hydrology* 572: 556-570.
- Granata, F. 2019. Evapotranspiration evaluation models based on machine learning algorithms-A comparative study. *Agricultural Water Management* 217: 303-315.
- Haghighatjou, P. and A. M. AkhondAli. 2008. Computation of evapotranspiration of Sistan plain based on solar data. In: Proceeding of the Second National Conference on Management of Irrigation and Drainage Networks.

- Ahvaz, Iran. (In Farsi).
12. Jabloun, M. and A. Sahli. 2008. Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data: application to Tunisia. *Agricultural Water Management* 95: 707-715.
 13. Jain, S., P. Nayak and K. Sudheer. 2008. Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrological Processes: An International Journal* 22(13): 2225-2234.
 14. Keskin, M. E., O. Terzi, E. D. Taylan and D. Küçükyaman. 2009. Meteorological drought analysis using data-driven models for the Lakes District, Turkey. *Hydrological Sciences Journal* 54(6): 1114-1124.
 15. Kişi, Ö. 2009. Modeling monthly evaporation using two different neural computing techniques. *Irrigation Science* 27(5): 417-430.
 16. Koulaian, A., M. A. Gholami Sefidkouhi and M. Ziatabar Ahmadi. 2013. Evaluation of new computational methods on estimation of daily evapotranspiration (Case study: Dasht-e-Naz, Sari). In: Proceeding of the 12th National Conference on Irrigation and Evaporation Reduction. Kerman, Iran. (In Farsi).
 17. Luo, Y., S. Traore, X. Lyu, W. Wang, Y. Wang, Y. Xie, X. Jiao and G. Fipps. 2015. Medium range daily reference evapotranspiration forecasting by using ANN and public weather forecasts. *Water Resources Management* 29(10): 3863-3876.
 18. Malohlava, M. and A. Candel. 2018. Gradient boosting machine with H₂O. Available online at: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/gbm.html> .Accessed 24 FEBRUARY 2020.
 19. Mattar, M. A., A. A. Alazba, B. Ablewi, B. Gharabaghi and M. A. Yassin. 2016. Evaluating and calibrating reference evapotranspiration models using water balance under hyper-arid environment. *Water Resources Management* 30: 3745-3767.
 20. Mendicino, G. and A. Senatore. 2013. Regionalization of the Hargreaves coefficient for the assessment of distributed reference evapotranspiration in Southern Italy. *Journal of Irrigation and Drainage Engineering* 139: 349-362.
 21. McCullagh, P. 1984. Generalized linear models. *European Journal of Operational Research* 16(3): 285-292.
 22. Nelder, J. A. and R. J. Baker. 1972. Generalized Linear Models. Wiley Online Library, New Jersey.
 23. Nykodym, T., T. Kraljevic, N. Hussami, A. Rao and A. Wang. 2018. Generalized linear modeling with H₂O. Available online at: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glm.html> .Accessed 24 FEBRUARY 2020.
 24. Perera, K. C., A. W. Western, B. Nawarathna and B. George. 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. *Agricultural and Forest Meteorology* 194: 50-63.
 25. Piri, H. 2012. Assessment of computational methods of estimation of potential evapotranspiration using lysimeter daya (Case study: Sistan plain). *Journal of Irrigation and Water Engineering* 3(9): 50-62. (In Farsi).
 26. Pregibon, D. and T. J. Hastie. 2017. Generalized linear models. In: Statistical Models in S, 195-247. Routledge.
 27. Rahimikhoob, A. 2014. Comparison between M5 model tree and neural networks for estimating reference evapotranspiration in an arid environment. *Water Resources Management* 28(3): 657-669.
 28. Rongfan, C., S. Shanlei, C. Haishan and S. Zhou. 2018. Changes in reference evapotranspiration over China during 1960–2012: attributions and relationships with atmospheric circulation. *Hydrological Processes* 32(19): 3032-3048.
 29. Saggi, M. K. and S. Jain. 2019. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. *Computers and Electronics in Agriculture* 156: 387-398.
 30. Sayyadi, H., A. Oladghaffari, A. Faalian and A. A. Sadraddini. 2009. Comparison of RBF and MLP neural networks performance for estimation of reference crop evapotranspiration. *Water and Soil Science* 19(1): 1-12. (In Farsi).
 31. Snyder, R. L., C. Palmer, M. Orang and M. Anderson. 2009. National weather service reference evapotranspiration forecast. *Crop Water Use* 4: 1-6.
 32. Tabari, H. 2010. Evaluation of reference crop evapotranspiration equations in various climates. *Water Resources Management* 24(10): 2311-2337.
 33. Tabari, H., M. E. Grismer and S. Trajkovic. 2013. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science* 31(2): 107-117.
 34. Tabari, H., O. Kisi, A. Ezani and P. H. Talaei. 2012. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology* 444: 78-89.
 35. Tang, D., Y. Feng, D. Gong, W. Hao and N. Cui. 2018. Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. *Computers and Electronics in Agriculture* 152: 375-384.
 36. Traore, S., Y. M. Wang and T. Kerh. 2010. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Water Resources Management* 24(5): 707-714.
 37. Valipour, M., M. A. Gholami Sefidkouhi and M. Raeini-Sarjaz. 2017. Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events. *Agricultural Water*

Management 180: 50-60.

38. Wen, X., J. Si, Z. He, J. Wu, H. Shao and H. Yu. 2015. Support-Vector-Machine-Based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resources Management* 29: 3195-3209.

Comparing of Generalized Linear Models, Random Forest and Gradient Boosting Trees in Estimation of Reference Crop Evapotranspiration (Case Study: The Sistan Plain)

H. Siasar^{1*}, T. Honar² and M. Abdolahipour³

(Received: December 18-2018 ; Accepted: July 14-2019)

Abstract

The estimation of reference crop evapotranspiration (ET_o) is one the important factors in hydrological studies, irrigation planning, and water resources management. This study attempts to explore the possibility of predicting this key component using three different methods in the Sistan plain: Generalized Linear Models (GLM), Random Forest (RF) and Gradient Boosting Trees (GBT). The maximum and minimum temperature, mean temperature, maximum and minimum humidity, mean humidity, rainfall, sunshine hours, wind speed, and pan evaporation data were applied for years between 2009 to 2018. Using various networks, the ET_o as output parameter was estimated for different scenarios including the combination of daily scale meteorological parameters. In order to evaluate the capabilities of different models, results were compared with the ET_o calculated by FAO Penman-Monteith as the standard method. Among studied scenarios, M1 covering the maximum number of input parameters (10 parameters) showed the highest accuracy for GBT model, with the lowest RMSE (0.633) and MAE (0.451) and the maximum coefficient of regression (R = 0.993). Air temperature was found as the most sensitive parameters during sensitivity analysis of studied models. It indicated that accuracy and precision of temperature data can improve the results. Application of the GBT model could decrease the time consumed to run the model by 70%. Therefore, the GBT model is recommended for estimation of ET_o in the Sistan plain.

Keywords: FAO Penman-Monteith method, Meteorological parameters, Sensitivity analysis, Air temperature

1. Department of Agriculture, Payame Noor University (PNU), Tehran, Iran.
2. Department of Water Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran.
3. Department of Irrigation and Soil Physics, Soil and Water Research Institute, Karaj, Iran.
*: Corresponding Author: Email: hadisiasar@pnu.ac.ir