نشریه آبیاری و زهکشی ایران
شماره۲، جلد ۱۳، خرداد - تیر ۱۳۹۸، ص. ۳۸۸-۳۷۸

Iranian Journal of Irrigation and Drainage
No. 2, Vol. 13, Jan.-Jul. 2019, p. 378-388

کاربرد مدل های ماشین بردار پشتیبان، چاید و جنگل تصادفی در برآورد تبخیر تعرق مرجع روزانه درشمال استان سیستان وبلوچستان

هادی سیاسر[1]،تورج هنر

تاریخ دریافت: ۱۳۹۷/۰۸/۲۷ تاریخ پذیرش: ۱۳۹۸/۰۲/۲۸

چکیده

تبخیر- تعرق به عنوان یکی از اجزا مهم چرخه هیدرولوژیک نقش بسیار بزرگی در برنامه ریزی ومدیریت منابع آبی در مناطق خشک ونیمه خشک ایفا می کند. اندازه گیری دقیق تبخیر -تعرق نیازمند ابزاری گران قیمت می باشد که امکان استفاده از آن ها در همه نقاط وجود ندارد. از این رو محققان همواره به دنبال روابط و روش های کاربردی، کم هزینه و با دقت مناسب برای برآورد صحیح مقدار این پارامتر بوده اند. روش های متعددی برای برآورد صحیح تبخیر تعرق در سراسر دنیا توسعه داده شده است .از جمله این روش ها می توان به معادلات تجربی و روش های داده محور از جمله شبکه های عصبی- مصنوعی، ماشین بردار پشتیبان و مدل های درختی اشاره نمود. لذا هدف از این پژوهش بررسی دقت و قابلیت مدل های ماشین بردار پشتیبان ، درخت تصمیم نوع چاید[2] وجنگل تصادفی در تخمین تبخیر-تعرق مرجع می باشد. داده های مورد استفاده در این تحقیق شامل دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد وتبخیراز تشت از ایستگاه هواشناسی دشت سیستان بین سالهای ۲۰۱۸-۲۰۰۹ می باشند. در این پژوهش، ابتدا با استفاده از داده های هواشناسی و مدل فائو پنمن مانتیث مقادیر تبخیر تعرق محاسبه و سپس با ارائه سناریوهای ترکیبی مختلف از پارامترهای هواشناسی به عنوان ورودی مدل های مورد مطالعه (ماشین بردار پشتیبان ، درخت تصمیم نوع چاید وجنگل تصادفی) در مقیاس زمانی روزانه سعی در برآوردی دقیق تر از تبخیر-تعرق مرجع به عنوان خروجی مدل ها شده است. در این تحقیق به منظور مقایسه مدل های مختلف ازشاخص های آماری ضریب همبستگی (R) و قدرمطلق خطا(MAE) استفاده شده است. بررسی نتایج نشان دادند که از میان مدل های ماشین بردار پشتیبان، چاید وجنگل تصادفی، مدل جنگل تصادفی با الگوی M7 بیشترین دقت را با ضریب همبستگی(R=0.983) و کمترین میانگین قدرمطلق خطا(MAE=0.798) دارد. بنابراین این پژوهش مدل جنگل تصادفی را برای برآورد تبخیر وتعرق در منطقه دشت سیستان توصیه می کند.

واژه های کلیدی: مدل چاید ، مدل ماشین بردار پشتیبان خطی، فائو پنمن-مانتیث، مدل جنگل تصادفی

مقدمه

تبخیر و تعرق یکی از مولفه های اصلی بیلان آب در یک منطقه و از جمله عوامل اثرگذار جهت برنامه ریزی دقیق آبیاری می باشد. لذا، برآورد دقیق نیاز آبی گیاه، تاثیر شایانی بر کاهش معضل بحران آب، به ویژه در مناطق خشک و نیمه خشک خواهد داشت. روشهای متعددی برای محاسبه تبخیر-تعرق وجود دارد که در یک تقسیم بندی کلی، میتوان آنها را به دو دسته روشهای مستقیم و غیرمستقیم تقسیم نمود. از یکی از رایجترین روشهای غیر مستقیم، استفاده مقادیر برآورد شده تبخیر-تعرق گیاه مرجع با استفاده از معادله فائوپنمن-مانتیث اصلاح شده می باشد. در این روش، پس از تعیین نیاز آبی گیاه مرجع، نیاز آبی گیاه مورد نیاز بر اساس ضرایب گیاهی به دست می آید(Allen et al.,1998) .در سالهای اخیر روشهای مبتنی بر داده کاوی وهوش مصنوعی در حل مسائلی از مهندسی آب که داده های کافی جهت برآورد مدل وجود داشته باشد به طور چشمگیری گسترش یافته اند. روش استفاده ازمدل های ماشین بردار پشتیبان،درخت تصمیم وجنگل تصادفی به عنوان روشهایی معتبر در مدل سازی فرآیندهای پیچیده غیرخطی مطرح است. اساسا "تبخیر و تعرق یک فرایند پیچیده و غیرخطی است و از این رو روش های ذکرشده بعنوان یک روش جدید که از قابلیت زیادی در مدل سازی فرایندهای پیچیده و غیرخطی برخوردار است، می تواند جهت مدل سازی تبخیر- تعرق بکار گرفته شود. شبکه های مبتنی بر هوش مصنوعی جزو روش های جعبه سیاه بوده و بدون در نظر گرفتن

۱- دانش‌آموخته دکتری مهندسی آب دانشگاه شیراز و هیأت علمی دانشگاه پیام نور
۲- دانشیار بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شیراز
(*-نویسنده مسئول: (Email : hadisiasar@pnu.ac.ir

فرایند فیزیکی حاکم بر سیستم و تنها با کمک داده های ورودی و خروجی قادر به ایجاد رابطه بین آن دو می باشد.تعداد داده های موردنیاز در روش پنمن–مانتیث به رغم دقت قابل قبول آن، همواره ذهن محققان را به سمت استفاده از روش های ساده تر و در عین حال دقیق معطوف ساخته است. نتایج پژوهش های پیشین حاکی از این واقعیت است که مدلهای خطی، به دلیل تغییرپذیری زیاد مولفه تبخیر–تعرق عملکرد مناسبی در برآورد آن نداشته و باید از مدلهای غیرخطی استفاده نمود که خود نیازمند برآوردپارامترهای زیادی خواهد بود (wang et al,2009) .در این راستا، مدل های هوشمند با توجه به توانایی درحل پدیده های غیرخطی و پیچیده می تواند مفید واقع شود.

Chen et al (2010) از SVM برای ریزمقیاس کردن بارش روزانه استفاده کردند و با روش آنالیز چندمتغیره مقایسه نمودند و نشان دادند نتایج پیش بینی های حاصل از SVM دقیقتر است. Tripathi et al (2006) از روش SVM برای برآورد بارندگی ماهانه در هند استفاده کردند. آنها پیشنهاد کردند که SVM به عنوان گزینه برتر نسبت به سایر روشهای معمول، برای پیش بینی بارش مورد استفاده قرار گیرد. کیسی وکیلیک مدل درختی و شبکه عصبی مصنوعی را برای برآورد تبخیر و تعرق مرجع تعمیم دادند. نتایج نشان داد که عملکرد مدل درختی دقیق تر است. هدف از این تحقیق ارزیابی مدل های شبکه عصبی مصنوعی ومدل درختی تحت شرایط مختلف حداقل داده های اقلیمی در منطقه وانتخاب سناریوهای کاربردی برای تخمین تبخیر و تعرق در این منطقه می باشد Kisi) (and Kilic, 2015. گو و همکاران در تحقیقی برآورد تبخیر و تعرق گیاه مرجع بصورت روزانه با استفاده از مدل ماشین بردار پشتیبان صورت پذیرفت که نتایج تحقیق نشان داد که مدل ماشین بردار پشتیبان می تواند داده های تبخیر و تعرق گیاه مرجع را تا۹۰ ٪با مقادیر محاسبه شده با روش پنمن– مانتیث تقریب بزند (Gou ,et al,2011) .

آیرماک و همکاران با استفاده از ایستگاههای هواشناسی در فلوریدا با اقلیم مرطوب، 21 روش تخمین مقادیر روزانه ET0 با مقادیر به دست آمده از روش پنمن مانتیث - -فائو مقایسه کردند. مدل های تابشی عملکرد بهتری از مدل های دمایی نشان دادند و از بین روشهایی که فقط از داده های دما استفاده می کردند، روش هارگریوز سامانی - دارای کمترین مقدار خطا بود (Irmak et al, 2005). Dehghanisanij et al (2004) عملکرد مدل های تخمین تبخیر و تعرق در دو منطقه کرج(خشک و نیمه خشک) و تاتوری ژاپن (مرطوب) با استفاده از دو سال داده- های لایسیمتری -1995) (1994ارزیابی - نمود و نتایج نشان دهنده این است که مدل پنمن مانتیث در منطقه - کرج و مدل پنمن در تاتوری مناسب ترین عملکرد را داشتند.

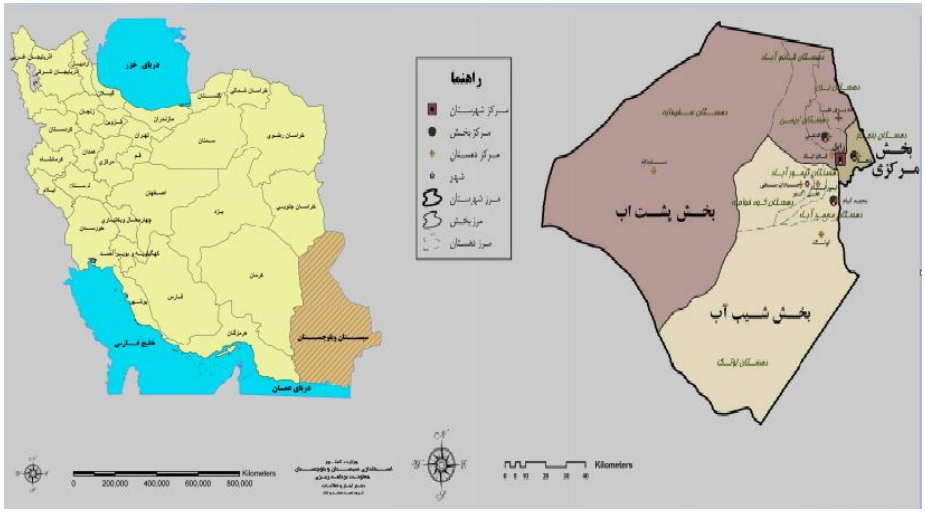
Rahimi Khoob (2008) با استفاده از داده های هواشناسی ایستگاه های منطقه خوزستان در ایران مقدار تبخیر– تعرق را با روش شبکه های عصبی مصنوعی و با دقت بالایی تخمین زدند. Mogadam nia et al (2009) برای مدل سازی تبخیر روزانه منطقه چاه نیمه، واقع در استان سیستان و بلوچستان از شبکه های عصبی مصنوعی و سیستم استنتاج عصبی فازی استفاده نمودند و درپایان بیان داشتند که نتایج حاصل از شبکه عصبی برای پیش بینی تبخیر نسبت به مدلهای تجربی و سیستم استنتاج فازی– عصبی از دقت بالاتری برخوردارند. Piri et al (2009) برای مناطق گرم و خشک مقدار تبخیر– تعرق را با شبکه های عصبی مصنوعی تخمین زدند. بولستیکس و همکاران قابلیتهای روش جنگل تصادفی را در زمینه آنالیز داده های بیوانفورماتیک بررسی کردند و بسیاری از جنبه‌های جنگل تصادفی را مورد بحث قرار دادند. همچنین در این تحقیق پیاده سازی های مختلف جنگل تصادفی معرفی و مقایسه شده است (Boulesteix et al, 2012). گیسالسون الگوریتم جنگل تصادفی را به منظور طبقه بندی پوشش اراضی با استفاده از داده های سنجش از دور بررسی کرد. این تحقیق نشان داد که روش جنگل تصادفی تا حد زیادی دقت طبقه بندی پوشش اراضی را بهبود می‌بخشد و احتمال برازش اضافی در آن وجود ندارد (Gislason,2004). همچنین روش جنگل تصادفی پتانسیل لازم به منزله ابزار مدل مکانی برای ارزیابی آسیب پذیری در مباحث زیست محیطی و منابع آب را دارا است (Booker and Snelder,2012).

با توجه به اهمیت تعیین دقیق و به موقع تبخیر –تعرق پتانسیل در محاسبات بیلان آبی، شبیه سازی تولیدات گیاهی و برنامه ریزی های آبیاری از یک سو و نبود داده های مناسب هواشناسی از سوی دیگر، ارائه یک مدل ساده، کم هزینه و دقیق را در ارائه این پارامتر ضروری می نماید. لذا هدف از انجام این تحقیق بررسی دقت مدلهای ماشین بردار پشتیبان خطی،درخت تصمیم نوع چاید وجنگل تصادفی در تخمین تبخیر – تعرق دشت سیستان بصورت روزانه می باشد.

مواد و روش ها

دشت سیستان در شمال استان سیستان و بلوچستان در $30^{\circ}.18'$ تا$31^{\circ}.20'$ عرض شمالی وعرض های جغرافیایی وطول جغرافیایی $61^{\circ}.10'$ تا $61^{\circ}.50'$ شرقی واقع است. دشت سسیتان با بارندگی میانگین سالیانه ۵۰ میلی متر (در حدود ۷ درصد متوسط بارندگی در جهان) و میزان تبخیر سالیانه ۴۰۰۰–۵۰۰۰ میلی متر از شرایط نامساعد محیطی برخوردار بوده و بر اساس شاخص خشکی دومارتن جزو نواحی فراخشک محسوب می شود. از علائم مشخصه دشت سیستان بادهای موسومی ۱۲۰ روزه می باشد که از اواخر اردیبهشت ماه به صورت پیوسته و گاهی گسسته با سرعت متوسط ۱۶ گره در ساعت می وزد.

داده های مورد استفاده در این تحقیق ازایستگاه زابل جمع آوری شده است. در این ایستگاه، آمار بلندمدت مولفه های اقلیمی شامل دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت نسبی حداکثر، رطوبت نسبی حداقل، رطوبت نسبی میانگین، ساعات آفتابی، سرعت باد، بارش وتبخیر ازتشت، به صورت روزانه در طول دوره آماری ۲۰۱۸-۲۰۰۹ جمع آوری شد. جدول (۱) میانگین متغیرهای مورد استفاده برای مدل سازی تبخیر– تعرق پتانسیل روزانه در طول دوره آماری برای دشت سیستان را نشان می دهد:



شکل ۱- موقعیت جغرافیایی دشت سیستان

پارامترهای ورودی مدل

در این مطالعه از داده های هواشناسی متداول مانند داده های شاخص روزانه، دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد وتبخیر از تشت در ایستگاه سینوپتیک دشت سیستان که قابل اندازه گیری هستند استفاده شده است.

جدول ۱ - میانگین آماری پارامترهای مورد استفاده در مدلسازی تبخیر وتعرق پتانسیل روزانه

ایستگاه/زابل	دمای میانگین (درجه سانتیگراد)	رطوبت نسبی(٪)	سرعت باد	بارش (میلیمتر)	ساعات آفتابی (ساعت)	ارتفاع از سطح دریا (متر)
میانگین	۲۳،۴۵	۲۹،۸۵	۱۳،۳۱	۰،۰۳۱	۹،۱۰۶	۴۸۹،۲
میانه	۲۳،۸۰	۲۷،۵۰	۱۳،۰۰	۰،۰۰	۹،۷۰۰	
واریانس	۱۰۲،۴۵	۲۲۸،۵۵	۳۸،۹۲	۰،۰۳۸	۷،۸۱۷	
Skewness	-۲،۳۴	۰،۶۳۸	۰،۱۹۳	۷،۱۴۲	-۱،۵۴۴	
Kutosis	-۱،۲۱۹	-۰،۴۲	-۰،۹۸۵	۵۳،۸۷۶	۲،۲۸۴	

داده های خام هواشناسی روزانه بین سال های ۲۰۱۸-۲۰۰۹جمع آوری و با استفاده از نرم افزار CropWat و مدل فائو پنمن مانتیث مقادیر تبخیر و تعرق روزانه محاسبه گردید. سپس از داده های مذکور در محاسبه مقدار روزانه تبخیر–تعرق مرجع مورد استفاده قرار گرفت. از بین کل داده های موجود، حدود ۸۰ درصد جهت آموزش و ۲۰ درصد بقیه جهت تست مدل ماشین بردار پشتیبان خطی، درخت تصمیم نوع چاید وجنگل تصادفی اختصاص یافت. ترکیب های متنوعی از داده های هواشناسی به عنوان ورودی مدل ها و مقادیر محاسبه شده متناظر، به عنوان خروجی مدلها در نظر گرفته شده و شبکه آموزش داده شد. نتایج حاصل از سناریوها و ترکیب های های مختلف مورد ارزیابی قرار گرفت و برترین سناریوی حاصل از این ارزیابی جهت پیش بینی مقادیر تبخیر–تعرق ، وارد مدل ها گردید.

معادله پنمن –مانتیث

روش موسوم به فائو پنمن–مانتیث تکامل یافته ی روش ترکیبی پنمن مانتیث ۱۹۶۵ است که توسط آلن وهمکاران (۱۲) برای محاسبه تبخیر وتعرق در نشریه ۵۶ سازمان فائو با نام E_{PM56} ارائه شده است.

این روش با دقت بالا در محدوده وسیعی ازمناطق و اقلیم ها برآورد صحیحی از تبخیر وتعرق گیاه مرجع ارائه می کند و از سوی سازمان خوار وبار جهانی(FAO) به عنوان روش استاندارد برای محاسبه تبخیروتعرق گیاه مرجع از روی داده های هواشناسی واقلیمی وهمچنین برای ارزیابی سایر روشها پیشنهاد شده است. بنابراین به منظور برآورد مقدار ET0 روزانه (میلیمتر بر روز) از روش فائو پنمن- مانتیث مطابق معادله زیر استفاده گردید (آلن و همکاران،۱۹۹۸):

$$E_{PM56}=\frac{0.408\Delta(R_n-G)}{\Delta+\gamma(1+0.34u_2)}+\frac{900\gamma}{[\Delta+\gamma(1+0.34u)]}\frac{u_2(e_s-e_a)}{(T+273)} \quad (۱)$$

در رابطه فوق ET_{PM56} تبخیر وتعرق گیاه مرجع $(mm\cdot day^{-1})$، شیب فشار منحنی فشار بخار اشباع$(kpaC^{-1})$، 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 (℃)، u_2 سرعت باد در ارتفاع دو متری $(m\ s^{-1})$،e_s فشار بخار اشباع (Kpa)،e_a فشار بخار واقعی (Kpa) است.

در تحقیق حاضر با استفاده از داد ه های هواشناسی روزانه ایستگاه مورد مطالعه، تبخیر –تعرق مرجع روزانه تعیین گردید.

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه سازی مقید است که از اصل استقرای کمینه سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینهٔ کلی می شود.

در یک مدل رگرسیونی SVM لازم است وابستگی تابع متغیر وابسته(y) به مجموعه ای از متغیرهای مستقل(x) تخمین زده شود .فرض براین است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز[1] مشخص می شود:

$$y=f(x)+noise \quad (۲)$$

بنابراین موضوع اصلی پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است پیش بینی کند. در روش Ɛ-SVM از توابع هسته ای (کرنل) به منظور جداسازی صفحات و بهینه سازی فواصل موجود بین داده های مشاهداتی و تخمینی استفاده می شود اما در مدل-V SVMاز روش برنامه ریزی غیرخطی برای کاهش خطای پیش بینی استفاده میشود (ریاحی و همکاران ۲۰۰۹).

در این مطالعه مدلε-SVM به دلیل کاربرد گسترده آن در مسائل رگرسیونی استفاده گردید. برای این مدل، تابع خطا به صورت زیر تعریف می شود:

$$C\sum_{i=1}^{N}\xi_i^{*}+C\sum_{i=1}^{N}\xi_i+\frac{1}{1}W^TW \quad (۳)$$

تابع خطای مذکور لازم است که با توجه به محدودیتهای زیر حداقل گردد:

$$W^T\phi(x_i)+b-y_i\le\varepsilon+\xi_i \quad (۴)$$

$$y_i-W^T\phi(x_i)+b\le\varepsilon+\xi_i \quad (۵)$$

$$\xi_i.\xi_i^{*}\ge 0 \quad (۶)$$

که در این روابط C ثابت گنجایش، W بردار ضرایب، W^T ترانهاده بردار ضرایب، ξ_i^{*}و ξ_i ضرایب کمبود،b ضریب ثابت، N الگوی آموزش مدل و ϕ تابع کرنل است.

ماشینهای بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسئله را از طریق توابع کرنل تغییر می دهند. انتخاب کرنل برای SVMبه حجم داده های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی های مسئله را داشته باشد. در عمل کرنل خطی به کار برده می شد. در جدول (۲) معادلات برخی از کرنل های رایج ارائه شد:

جدول ۲ -توابع کرنل رایج در ماشینهای بردار پشتیبان (همل،۲۰۰۹)

نوع تابع	تابع کرنل
خطی	$k(x_i.x_j)=x^T{}_i\cdot x_j$

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت معادله ۴ خواهد بود که کنترل کننده میزان نوسان تابع گوسی و همچنین کنترل کننده نتایج پیش بینی و تعمیم دهنده مدل SVM است (یو وچن و همکاران۲۰۰۶).

$$f(x_i)=\sum_{i=1}^{l}(-\partial_i-\partial_i^{*})K(x_i.x_j)+b \quad (۷)$$

الگوریتم CHAID[2]

چاید مخفف مجذور مربع شناسایی کننده تعاملات خودکار است. این روش آماری بسیار موثر برای بخش بندی و رشد درخت، به وسیله کاس توسعه یافته است. با استفاده از مفهوم تست آماری به عنوان یک معیار، چاید تمام مقادیر صفت پیش بینی کننده بالقوه را ارزیابی می‌کند. این الگوریتم مقادیری را که به صورت آماری همگن در نظر گرفته شده اند(شباهت) با توجه به متغیر هدف و حفظ تمام مقادیری که ناهمگن (غیر مشابه) هستند، ادغام می کند.(Ramaswami, M., and R. Bhaskaran,2010). سپس این الگوریتم بهترین پیش بینی کننده را از شاخه در درخت تصمیم انتخاب می کند به طوریکه هر فرزندگره از یک گروه از ارزش‌های

1- Noise

2- Chi-squared Automatic Interaction Detector

همگون نسبت به صفت انتخاب شده می باشند. این فرآیند به صورت بازگشتی تا آنجا که درخت به طور کامل رشد می‌کند، ادامه می‌یابد. آزمون آماری مورد استفاده بستگی به سطح اندازه گیری میدان هدف دارد. اگر صفت هدف پیوسته باشد، یک آزمونF صورت می گیرد.اگرصفت هدف دسته ای باشد،یک آزمون مجذور مربع به کار برده می‌شود. چاید یک روش باینری نیست و می‌تواند بیش از دو دسته در هر سطح خاص ازدرخت ایجاد کند. بنابراین گرایش به ساخت درخت نسبت به روش رشد باینری گسترده تر است. این الگوریتم برای تمام انواع متغیرها پاسخگو است و دو متغیروزنی و فراوانی را می پذیرد .مدل مذکور همچنین مقادیر مفقودی را با یک طبقه بندی معتبر ایجاد می کند(Kisi, Ozgur, et al.,2016).

مدل جنگل تصادفی

مدل جنگل تصادفی برای بسیاری از مجموعه داده ها، دسته بندی را با سرعت بالایی انجام می دهد و بر خلاف مدل های کلاسیک چون رگرسیون[1] که تنها بر پایه یک مدل تکیه دارند با استفاده از صدها و هزاران درخت از اطلاعات بیشتری در داده ها استفاده می کند تا بتوان استنباط بهتری از متغیرها داشت. این روش یک تکنیک مدل ناپارامتری[2] و متعلق به خانواده روش های دسته جمعی است. این الگوریتم از جمله دسته بندی هایی است که متد Bagging را به کار می گیرد و حاوی چندین درخت تصمیم است که خروجی آن از خروجی های درختان افرادی به دست می آید (بریمن ،۲۰۰۱).

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می کنند. با استفاده از روش بوت استرپ به تعداد زیاد (به عنوان مثال ۲۰۰۰ مرتبه) نمونه های n تایی از مجموعه دادههای مشاهداتی اولیه، نمونه برداری همراه با جایگذاری می شوند. در طی فرآیند نمونه گیری حدود یک سوم از داده ها نمونه گیری نمیشوند و به عنوان نمونه خارج از کیسه (از این داده ها برای تعیین متغیرهای مهم و همچنین برآورد ناریب خطا استفاده می شود) در نظر گرفته می شوند. سپس بر روی هر نمونه بوت استرپ یک درخت گسترش داده می شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به صورت تصادفی m متغیر برای تقسیم شدن انتخاب میشود. برای حالت رگرسیونی نسبت $\frac{m}{M}$برابر با یک سوم است و برای کلاسه بندی برابر با $m = \sqrt{M}$پیشنهاد شده است پس از ساخت تمام درخت داده های تست به درخت معرفی شده و به تعداد درختها برای بردار ورودی یک خروجی به دست میآید. با میانگین گیری این خروجی ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی ها مقادیر صدکها و دامنه عدم قطعیت محاسبه میشود .روش درخت رگرسیون جنگل تصادفی به ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیشبینی کننده ها نسبتاً کم باشد یک روش پیش بینی کارآمد است (Svetnik et al. 2003) .

در این پژوهش محاسبات مدل جنگل تصادفی در نرم افزار Matlab 2017 انجام شد. به منظور انتخاب مناسبترین خط برازش یافته توسط این روش در احتمالات متفاوت، متغیر اندازه گره[3] که نشان دهنده تعداد برگها در هر شاخه است با آزمون و خطا تعیین شد.

متغیرهای ورودی به مدلها

در این تحقیق با توجه به داده های ورودی روزانه که شامل شاخص روزانه ،دمای حداکثر،دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل ، رطوبت میانگین ،بارش ،ساعات آفتابی، سرعت باد وتبخیر ازتشت می باشند الگوهای مختلفی ارائه شده است. الگوهای ورودی در جدول(۳) زیر آورده شده است. برای انتخاب مناسب ترین مدل نفوذ ،بر اساس معیارهای خطای ذکرشده، از روش رتبه بندی برای داده های تست استفاده می گردد . بدین معنا که به هر یک از مدل ها بر اساس معیارهای ذکرشده رتبه ای داده می‌شود و در پایان پس از مقایسه مجموع رتبه های به دست آمده، مناسب ترین مدل برای تخمین بهترین معادله نفوذ انتخاب می شود. شیوه رتبه بندی این مدل‌ها بدین صورت است که در هر روش، مدلی که کمترین میزان خطایMAE را دارد، پایین ترین رتبه ،یعنی رتبه ۱ ؛و به مدل ها یا ترکیبات ورودی دیگر در هر یک از دسته های ورودی بر حسب میزان معیار خطاهای مذکور، رتبه های ۲ الی آخر تعلق می گیرد. اما در مورد معیارهای ضریب همبستگی (R) ، مدل یا ترکیبی که بیشترین مقدار این معیارها را دارا است، رتبه ۱ اختصاص می یابد. همچنین به مدل هایی که میزان معیار خطای آنها یکسان باشند، رتبه یکسان تعلق می گیرد. در نهایت پس از انتخاب مدل ها یا ترکیبات مناسب هر روش و مقایسه بین آنها، روش برتر در برآورد تبخیر وتعرق مرجع در منطقه مورد مطالعه انتخاب می گردد.

معیارهای آماری برای مقایسه مدل ها

در این تحقیق به منظور مقایسه مدل های مختلف ازشاخص های آماری ضریب همبستگی (R) برای تعیین همبستگی بین مقادیر پیش بینی شده و اندازه گیری شده وهمچنین میانگین قدرمطلق خطا(MAE) برای نشان دادن میزان همخوانی میان مجموعه های داده های مقادیر مشاهداتی ومدل سازی شده به عنوان معیارهای ارزیابی مورد استفاده قرار گرفتند (روابط۸تا۱۰).

1- Regression
2- Nonlinear
3- Nodesize

جدول۳- الگوهای استفاده شده توسط مدل های مورد استفاده در این تحقیق جهت تخمین تبخیر وتعرق پتانسیل

سناریو	پارامترهای ورودی شبکه	تعداد متغیر
۱	دمای حداکثر،دمای حداقل،دمای میانگین،رطوبت حداکثر،رطوبت حداقل،رطوبت میانگین،بارش،ساعات آفتابی،سرعت باد،تبخیر از تشت	۱۰
۲	دمای حداقل،دمای میانگین،رطوبت حداکثر،رطوبت حداقل،رطوبت میانگین،بارش،ساعات آفتابی،سرعت باد،تبخیر از تشت	۹
۳	دمای حداکثر،دمای حداقل،دمای میانگین،رطوبت حداکثر،رطوبت حداقل،رطوبت میانگین، ساعات آفتابی،سرعت باد،تبخیر از تشت	۹
۴	دمای حداکثر،دمای حداقل،دمای میانگین،رطوبت حداکثر،رطوبت حداقل،رطوبت میانگین ،ساعات آفتابی، تبخیر از تشت	۸
۵	دمای حداکثر،دمای حداقل،دمای میانگین،رطوبت حداکثر،رطوبت حداقل،رطوبت میانگین ، سرعت باد،تبخیر از تشت	۸
۶	دمای حداکثر،دمای حداقل،دمای میانگین، رطوبت حداقل،رطوبت میانگین ، سرعت باد،تبخیر از تشت	۷
۷	دمای حداکثر،دمای حداقل،دمای میانگین، رطوبت میانگین ، سرعت باد،تبخیر از تشت	۶
۸	دمای حداکثر، ،دمای میانگین، رطوبت میانگین ، سرعت باد،تبخیر از تشت	۵
۹	دمای میانگین، رطوبت میانگین ، سرعت باد،تبخیر از تشت	۴
۱۰	دمای حداکثر، رطوبت میانگین ، سرعت باد،تبخیر از تشت	۴
۱۱	دمای حداکثر، رطوبت میانگین ، تبخیر از تشت	۳
۱۲	دمای حداکثر، سرعت باد،تبخیر از تشت	۳
۱۳	دمای حداکثر، سرعت باد	۲
۱۴	دمای حداکثر	۱
۱۵	سرعت باد	۱
۱۶	تبخیر ازتشت	۱
۱۷	دمای حداکثر-دمای حداقل	۲
۱۸	رطوبت میانگین	۱
۱۹	دمای حداکثر-تبخیر از تشت	۲
۲۰	دمای حداکثر -رطوبت میانگین	۲
۲۱	دمای حداکثر-رطوبت حداکثر	۲
۲۲	دمای میانگین-سرعت باد	۲
۲۳	دمای حداکثر-سرعت باد	۲
۲۴	ساعات آفتابی	۱
۲۵	دمای حداقل	۱

$$R = \frac{\sum_{i=1}^{N} O_i P_i}{\sqrt{\sum_{i=1}^{N} O^2{}_i \sum_{i=1}^{N} p^2{}_i}} \quad (۸)$$

$$MAE = \frac{\sum_{i=1}^{N} (|O_i - P_i|)}{N} \quad (۹)$$

در این روابط: O_i تبخیر و تعرق به روش فائو- پنمن –مانتیث؛ P_i تبخیر و تعرق محاسبه شده به هر روش؛ R ضریب همبستگی که هر چه این شاخص به یک نزدیکتر باشد نشان دهنده همبستگی بالای مدل می باشد و MAEمیانگین قدرمطلق خطا که هرچه این شاخص به صفر نزدیک تر باشد نشان دهنده انحراف کمتر و دقت بالاتر مدل می‌باشد.

نتایج وبحث

پس از طراحی اجزای مختلف مدل های ماشین بردار پشتیبان خطی ،درخت تصمیم نوع چاید و جنگل تصادفی، خروجی این مدل‌ها دریافت شد. برای انتخاب مناسب ترین مدل ،بر اساس معیارهای خطای ذکرشده، از روش رتبه بندی برای داده های تست استفاده می گردد. بدین معنا که به هر یک از مدل ها بر اساس معیارهای ذکرشده رتبه ای داده می‌شود و در پایان پس از مقایسه مجموع رتبه های به دست آمده، مناسب ترین مدل برای تخمین بهترین معادله تبخیر- تعرق انتخاب می شود.

شیوه رتبه بندی این مدل‌ها بدین صورت است که در هر روش، مدلی که کمترین میزان خطایMAE را دارد، پایین ترین رتبه،یعنی رتبه ۱ و به مدل ها یا ترکیبات ورودی دیگر در هر یک از دسته های ورودی بر حسب میزان معیار خطاهای مذکور، رتبه های ۲ الی آخر تعلق می گیرد. اما در مورد معیارهای ضریب همبستگی (R) ، مدل یا ترکیبی که بیشترین مقدار این معیارها را دارا است، رتبه ۱ اختصاص می یابد. همچنین به مدل هایی که میزان معیار خطای آنها یکسان باشند، رتبه یکسان تعلق می گیرد.

بررسی مقایسه سه مدل ماشین بردارپشتیبان خطی، درخت تصمیم نوع چاید وجنگل تصادفی نشان می دهد که می‌توان یک الگوی مشخصی برای دشت سیستان را با دقت مناسب برای برآورد تبخیر و تعرق پتانسیل روزانه معرفی نمود و باید برای هر مدل، الگوی برتر را معرفی کرد. در نهایت پس از انتخاب مدل ها یا ترکیبات مناسب هر روش و مقایسه بین آنها، روش برتر در برآورد تبخیر وتعرق مرجع در منطقه مورد مطالعه انتخاب می گردد. جدول ۴ضریب همبستگی و خطای سنایوهای مختلف در سه مدل ماشین بردار پشتیبان خطی، درخت تصمیم نوع چاید وجنگل تصادفی را نشان میدهد. همچنین در این تحقیق تحلیل حساسیت به منظور شناسایی وتعیین درجه اهمیت هر یک از پارامترها در سری داده های ورودی بر عملکرد دینامیکی سیستم صورت گرفته است. به منظور آنالیز حساسیت در هر سری داده ورودی یک پارامتر حذف و مدل با همان داده های تست، اجرا ومقادیر ضریب همبستگی وخطا تعیین گردید. با حذف هر پارامتر هر چه خطای مدل افزایش بیشتری نشان دهد، بیانگر تاثیر بیشتر آن پارامتر بر فرایند شبیه سازی می باشد. جهت انجام تحلیل حساسیت برای سری داده های ورودی از مدلهای SVM-linear، CHAID وRF به دلیل دقیق تر بودن نتایج آن استفاده شد.

جدول ٤-ضریب همبستگی و خطای سنایوهای مختلف در سه مدل ماشین بردار پشتیبان خطی،درخت تصمیم نوع چاید وجنگل تصادفی در ایستگاه مورد مطالعه

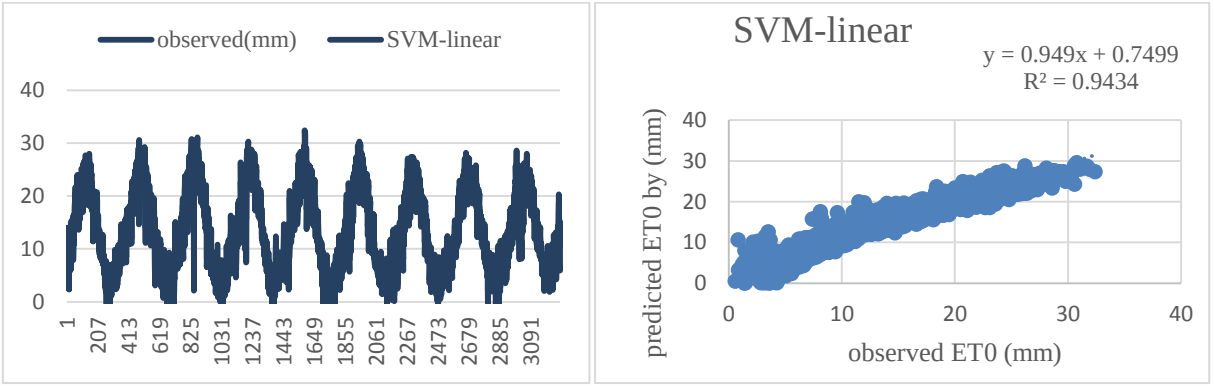
سناریو	SVM-Linear		CHaid		Random Forest	
	R	MAE	R	MAE	R	MAE
M1	۰,۹۷۱	۱,۳۹۸	۰,۹۵۴	۱,۴۸	۰,۹۸۴	۰,۸۱۹
M2	۰,۹۵۸	۲,۱۴۹	۰,۹۵۲	۱,۵۲	۰,۹۷۴	۱,۱۰۳
M3	۰,۹۷۱	۱,۴۰۲	۰,۹۵۴	۱,۴۸	۰,۹۸۳	۰,۸۰۷
M4	۰,۹۴۳	۱,۷۳۲	۰,۹۴۵	۱,۵۸۷	۰,۹۵۵	۱,۴۱۳
M5	۰,۹۷	۱,۴۱۵	۰,۹۵۴	۱,۴۸۱	۰,۹۸۳	۰,۷۹۸
M6	۰,۹۷	۱,۴۱۵	۰,۹۵۴	۱,۴۸۱	۰,۹۸۳	۰,۷۹۸
M7	۰,۹۷	۱,۴۱۵	۰,۹۵۴	۱,۴۸۱	۰,۹۸۳	۰,۷۹۸
M8	۰,۹۶۹	۱,۴۱۹	۰,۹۵	۱,۶۱۳	۰,۹۸۳	۰,۸۰۴
M9	۰,۹۳۱	۱,۹۹۴	۰,۹۲۴	۲,۰۴۲	۰,۹۳۹	۱,۷۵۴
M10	۰,۹۶۹	۱,۴۱۹	۰,۹۵	۱,۶۱۳	۰,۹۸۳	۰,۸۰۴
M11	۰,۹۴	۱,۸۰۲	۰,۹۳۷	۱,۸۲۶	۰,۹۴۳	۱,۶۶۱
M12	۰,۹۶۵	۱,۵۴۶	۰,۹۴۹	۱,۶۲۹	۰,۹۷۴	۱,۱۶۸
M13	۰,۹۵۴	۱,۸۳۱	۰,۹۷۱	۱,۲۹۳	۰,۹۷۱	۱,۸۱۱
M14	۰,۸۶۴	۳,۰۵۶	۰,۸۷۹	۲,۶۶۱	۰,۸۷۱	۲,۸۱۹
M15	۰,۶۶۴	۴,۳۹۶	۰,۶۶۸	۴,۳۶۳	۰,۶۶۳	۴,۳۸۷
M16	۰,۸۹۶	۲,۳۹۴	۰,۹	۲,۳۱۴	۰,۹۰۲	۲,۳۱۶
M17	۰,۹۰۵	۲,۴۸۶	۰,۹۲۸	۱,۹۳۵	۰,۹۳۲	۱,۹۱۵
M18	۰,۷۵۵	۳,۷۹۵	۰,۸۰۳	۳,۳۴۹	۰,۸۰۶	۳,۳۲۴
M19	۰,۹۳۵	۱,۸۹۷	۰,۹۳۴	۱,۸۸۲	۰,۹۳۵	۱,۸۴۷
M20	۰,۸۹۲	۲,۶۶۸	۰,۹۱۹	۲,۰۶۶	۰,۹۱۷	۲,۰۸۱
M21	۰,۸۶۴	۳,۰۵۶	۰,۸۷۹	۲,۶۶۱	۰,۸۷۱	۲,۸۱۹
M22	۰,۹۲۶	۲,۲۴۹	۰,۹۴۲	۱,۷۶۶	۰,۹۴۴	۱,۷۷
M23	۰,۹۵۴	۱,۸۳۱	۰,۹۷۱	۱,۲۹۳	۰,۹۷۱	۱,۲۸۲
M24	۰,۵۳۱	۵,۳۵۹	۰,۶۵۹	۴,۴۳۷	۰,۶۶۲	۴,۴۳۴
M25	۰,۹	۲,۵۳۲	۰,۹۲۲	۲,۰۳۵	۰,۹۲۲	۲,۰۳

همانطور که در بخشهای قبلی بیان گردید، در این پژوهش ۲۵ سناریو با ترکیب پارامترهای مختلف اقلیمی به عنوان ورودی مدل ایجاد شدند. نتایج این تحقیق نشان می دهد که در مدل ماشین بردار پشتیبان خطی هنگامی که پارامترهای هواشناسی به عنوان ورودی مدل معرفی شدند، بهترین سناریو M1 با بالاترین ضریب همبستگی R=0.971 و وکمترین مقدار خطا MAE=1.398 در مدل، بوده است.

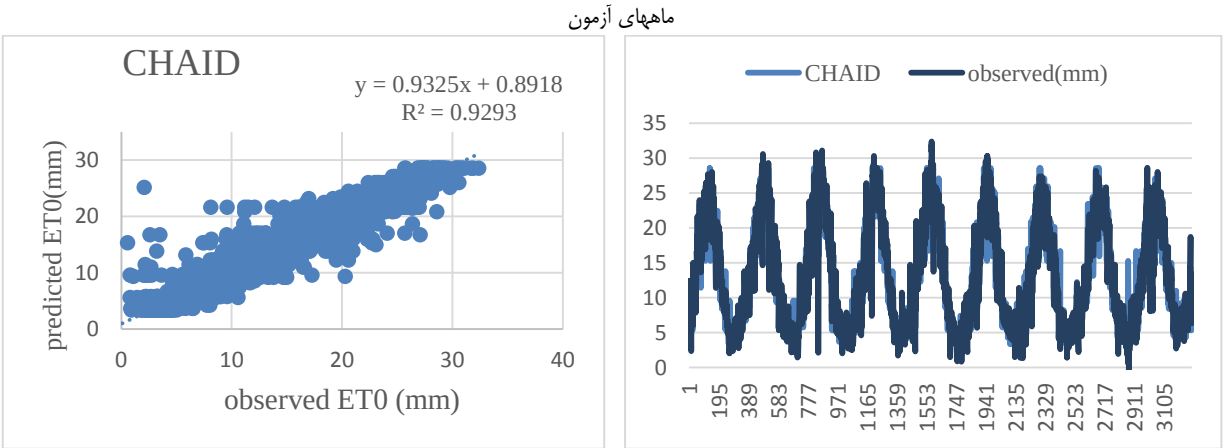
همچنین در مدل درخت تصمیم نوع چاید بهترین سناریوM13 با R=0.971 وکمترین خطاMAE=1.293 به عنوان بهترین سناریو انتخاب گردید. و درنهایت در مدل جنگل تصادفی الگوهای M5 ، M6 و M7 به عنوان بهترین الگوها با بیشترین ضریب همبستگی R=0.983 وکمترین خطا (MAE=0.798)، به عنوان الگوی برتر شناخته شدند. نتایج آنالیز حساسیتها نشان می دهد در ماشین بردار پشتیبان خطی به ترتیب پارامترهای دمای حداکثر، تبخیر از تشت، سرعت باد و رطوبت میانگین به عنوان پارامترهای تاثیر گذار در مدل شناخته شدند. در مدل درخت تصمیم نوع چاید بترتیب دمای حداقل، رطوبت میانگین ،دمای حداکثرو تبخیر از تشت بیشترین تاثیررا در مدل و در مدل جنگل تصادفی سرعت باد ،دمای حداکثر ورطوبت میانگین به عنوان مهمترین پارامترها، اهمیت نسبی بالایی در برآورد تبخیر و تعرق از منطقه مورد مطالعه دارد. شکلهای ۲ تا ۴ پراکنش داده های مشاهداتی و برآورد شده توسط بهترین الگو از هر مدل نشان داده شده است.

نتیجه گیری

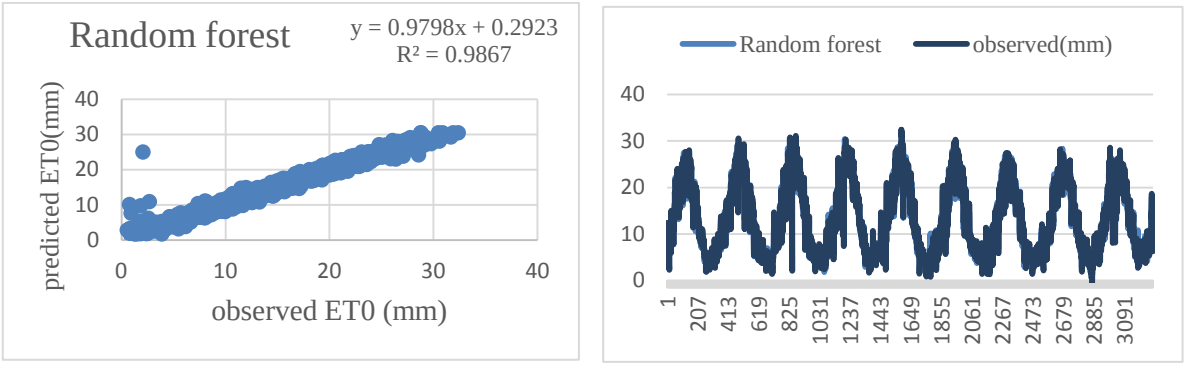
در این پژوهش توانایی سه مدل روش ماشین بردار پشتیبان خطی، درخت تصمیم نوع چاید و جنگل تصادفی در الگوسازی تبخیر وتعرق مرجع روزانه دردشت سیستان مورد ارزیابی قرار گرفت. مدل سازی برای دشت سیستان از سال ۲۰۰۹ تا ۲۰۱۸ به صورت روزانه انجام شده است. برای این منظور از داده های دمای حداکثر، دمای حداقل، دمای میانگین، رطوبت حداکثر، رطوبت حداقل، رطوبت میانگین، بارش، ساعات آفتابی، سرعت باد وتبخیر از تشت به عنوان پارامترهای ورودی مدل وتبخیر و تعرق گیاه مرجع روزانه به عنوان خروجی از مدل انتخاب شدند.



شکل ۲- سمت راست، پراکنش مدل ماشین بردار پشتیبان با مدل فائو-پنمن مانتیث ب)سمت چپ:نتایج مدل بدست آمده تبخیر- تعرق مرجع برای



شکل ۳- سمت راست، پراکنش مدل چاید با مدل فائو-پنمن مانتیث ب)سمت چپ:نتایج بدست آمده تبخیر-تعرق مرجع برای ماههای آزمون



شکل ٤- الف)سمت راست، پراکنش مدل جنگل تصادفی با مدل فائو-پنمن مانتیث ب)سمت چپ:نتایج مدل بدست آمده تبخیر-تعرق مرجع برای ماههای آزمون

مقادیر برآورد شده از هر یک از مدل ها با روش فائو-پنمن-مانتیث مقایسه شدند. ارزیابی مدل ها با استفاده از ضریب همبستگی(R) و مقادیر میانگین قدرمطلق خطا(MAE) صورت گرفته است. بررسی نتایج مدل ها در ایستگاه مورد مطالعه نشان داد که سناریوی M7در مدل جنگل تصادفی با ضریب همبستگی R=0.983 و MAE=0.798 برای مدل سازی تبخیر و تعرق گیاه مرجع نسبت به بقیه مدل ها به دلیل اینکه اساس آن برپایه قوانین منطقی بیان شده است از دقت بالاتری نسبت به مدل های مذکور برخوردار است. از این رو استفاده از این پارامترها برای تعیین تبخیر و تعرق روزانه توصیه می شود.

تقدیر و تشکر

بدینوسیله نویسندگان این مقاله، از دانشگاه زابل که این پژوهش با حمایت مالی این دانشگاه با کد پژوهانه UOZ-GR-9618-18 صورت گرفته است کمال تشکر و سپاسگزاری را دارند.

منابع

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 No. 56. Food and Agricultural Organization of the United Nations, Rome, 300p.

Booker, D.J. and Snelder, T. H. 2012. Comparing methods for estimating flow duration curves at ungauged sites. Journal of Hydrology 434–435, 78–94.

Boulesteix, A.L, Janitza, S. Kruppa, J, and König IR. 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.24 (2), pp: 493-507.

Breiman, L., 2001. Application and analysis of random forests and machine learning. Journal of Water Management, 15(1): 5-32.

Chen S.T., Yu P.Sh., and Tang H.Y. 2010. Statistical downscaling of daily Precipition using support vector machines and multivariate analysis. Journal of Hydrology, 385:13-23.

Dehghanisanij, H., T. Yamamoto and V. Rasiah. 2004. Assessment of evapotranspiration estimation models for use in semiarid environments, Agricultural Water Management, 64: 91-106.

Gislason, PO. Benediktsson, JA. and Sveinsson, JR. 2004. Random forest classification of multisource remote sensing andgeographic data. Journal of Geoscience and Remote Sensing Symposium, Vol. 2, pp: 1049-52.

Guo J., Zhou J., Qin H., Zou Q. and Li Q. (2011). Monthly stream flow forecasting based on improved support vector machine model, Expert Sys. Appl., 38 (10), 13073-13081.

Irmak, S., Haman, D., Jones, J.W., 2002. Evaluation of class 'A' pan coefficients for estimating reference evapotranspiration in a humid location. J. Irrig. Drain. Eng. ASCE 128 (3), 153–159.

Kisi O. (2008). The potential of different ANN techniques in evapotranspiration modeling. Hydrol.Proc., 22(14), 2449-2460.

Kisi O. 2009. Modeling monthly evaporation using two different neural computing, techniques Irrigation Science, 27(5): 417-430

Kisi O. Kilic Y. 2015. An investigation on generalization ability of artificial neural networks and M5 model tree in modeling reference evapotranspiration. TheorAppl Climatol. 1-13p

Kisi, Ozgur, Onur Genc, Semih Dinc, and Mohammad Zounemat-Kermani. "Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree." Computers and Electronics in

Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *arXiv preprint arXiv:1002.1144*.

Tripathi Sh., Srinivas V.V., and Nanjundiah R.S. 2006. Downscaling of precipitation for climate change scenarios: A support vector machine approach.Journal of Hydrology, 330:62- 640.

Wang, K., & Liang, S. (2008). An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture. Journal of Hydrometeorology, 9(4), 712-727.

Yu P.S, Chen S.T., Chang I.F. 2006. Support vector regression for real-time flood stage forecasting. Hydrology,328: 704-716.

Agriculture 122 (2016): 112-117.

Moghaddamnia A., Ghafari Gousheh M., Piri J., Amin S., and Han D. 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Advances in Water Resources. 32 : 88–97.

Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C., Komaki, C.B, and Bogaert, P. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. Journal of Geoderma, Vol. 232, pp: 97–106

Piri, J., S. Amin, A. Moghaddamnia, A. Keshavarz, D. Han and R. Remesan. (2009). "Daily Pan Evaporation Modeling in a Hot and Dry Climate." J. Hydrol. Eng., 14(8): 803–811.

Iranian Journal of Irrigation and Drainage
No. 2, Vol. 13, Jan.-Jul. 2019, p. 378-388

نشریه آبیاری و زهکشی ایران
شماره۲، جلد ۱۳، خرداد - تیر ۱۳۹۸، ص. ۳۸۸-۳۷۸

Application of Support Vector Machine, CHAID and Random forest Models, in Estimated Daily Reference Evapotranspiration in Northern Sistan and Baluchestan Province

Hadi Siasar[1*], Tooraj Honar [2]
Recived: Nov.18, 2018 Accepted: May.18, 2019

Abstract

Evapotranspiration as an important component of the hydrological cycle plays a very important role in the planning and management of water resources in dry and hyper dry areas.

Accurate estimation of evapotranspiration requires a costly tool which can not be used anywhere. Hence, researchers are always looking for applied relationships and practices that are low-cost and accurate for the correct estimation of the value of this parameter. Several methods have been developed for the accurate estimation of evapotranspiration throughout the world. Among of these methods can be pointed out empirical equations, Artificial Neural Network, support vector machin, and tree models. Therefore, the purpose of this study was to investigate the accuracy and the capability of linear supporting vector machine models, the decision tree of the type of chiad and the random forest model in the estimation of reference evapotranspiration.The data used in this study include maximum temperature, minimum temperature, average temperature, maximum moisture content, minimum humidity, average humidity, precipitation, sunny hours, wind speed, and a shift from the meteorological station of Sistan Plain between2009-2018. In this study, using meteorological data and the FAO Penman-Monteith model, the values of evapotranspiration were calculated and then by providing different combination scenarios of the meteorological parameters as inputs of the studied models on a daily basis, an attempt was made to find a more accurate estimate of the refrence evapotranspiration as the output of the models. In this research, correlation coefficient (R) and Mean Absolute amount of Error (MAE) were used to compare different model. The results showed among the support model carriers, the random knife, the random forest model with M7 patterns has the highest accuracy with the correlation coefficient (R = 0.983) and the lowest mean error magnitude (MAE = 0.798). Therefore, this research recommends a random forest model for estimating evapotranspiration in the area of Sistan plain.

Keywords: Chid model, linear support vector model, FAO Penman-Monteith, random forest model

1- Assistant Professor., Department of Agriculture, Payame Noor University
2- Faculty Member, Department of Water Science and Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran
(*-Corresponding Authour Email:hadisiasar@pnu.ac.ir)