

پیش‌بینی تبخیر و تعرق ماهانه گیاه مرجع در شمال غرب کشور با استفاده از برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان

فرشاد احمدی^۱، فریدون رادمش^۲، رسول میرعباسی نجف آبادی^۳، سجاد آیشم^{۴*}

تاریخ دریافت: ۱۳۹۳/۶/۸ تاریخ پذیرش: ۱۳۹۳/۱۲/۱۳

چکیده

در مطالعه حاضر به منظور تخمین تبخیر-تعرق گیاه مرجع با استفاده از دو مدل ماشین بردار پشتیبان (SVM) و برنامه‌ریزی ژنتیک (GP) در مقیاس زمانی ماهانه، ۶ ایستگاه سینوپتیک در منطقه شمال غرب کشور در دوره آماری ۲۸ ساله (۲۰۱۰-۱۹۷۳) انتخاب گردید. در ابتدا مقادیر تبخیر-تعرق مرجع ماهانه برای ایستگاه‌های منتخب توسط روش فائو-پنمن-مانتیت محاسبه و به عنوان خروجی مدل‌های SVM و GP در نظر گرفته شد. سپس یک رابطه رگرسیونی بین متغیرهای اقلیمی مختلف مؤثر در پدیده تبخیر-تعرق به دست آمده و الگوهای مختلف ورودی برای مدل‌های مورد استفاده مشخص شد که بر این اساس رطوبت نسبی با داشتن کم‌ترین اثر از ورودی‌ها حذف گردید. همچنین در مطالعه حاضر، به منظور بررسی اثر حافظه در پیش‌بینی تبخیر-تعرق از گام‌های زمانی (تاخیر) یک، دو، سه، چهار و پنج ماهه نیز به عنوان ورودی برای مدل‌ها استفاده شد. به طور کلی، برای هر مدل ۱۰ الگوی ورودی ایجاد گردید. هرچند نتایج حکایت از دقت بالا و خطای کم هر دو مدل در پیش‌بینی تبخیر-تعرق مرجع ماهانه در شمال غرب ایران داشت، ولی کارایی مدل SVM به مراتب بهتر از مدل GP بود. همچنین نتایج مطالعه حاضر نشان داد که در صورت نبود اطلاعات هواشناسی کافی، می‌توان از حافظه نیز در پیش‌بینی تبخیر-تعرق گیاه مرجع استفاده کرد.

واژه‌های کلیدی: برنامه‌ریزی ژنتیک، تبخیر-تعرق گیاه مرجع، حافظه، ماشین بردار پشتیبان

مقدمه

است. این روش توسط (Vapnik, 1998) بر پایه تئوری یادگیری آماری^۷ بنا نهاده شده است. ماشین بردار پشتیبان روشی برای طبقه‌بندی دوتائی با ویژگی‌های دلخواه است و از این روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود (Pai and Hong, 2007). ماشین بردار پشتیبان در اصل یک دسته‌بندی کننده دو کلاسه است که کلاس‌ها را توسط یک مرز خطی از هم جدا می‌کند. در این روش نزدیک‌ترین نمونه‌ها به مرز تصمیم‌گیری را بردارهای پشتیبان می‌نامند. این بردارها معادله مرز تصمیم‌گیری را مشخص می‌کنند. الگوریتم‌های شبیه‌سازی هوشمند کلاسیک مانند شبکه‌های عصبی مصنوعی، معمولاً قدر مطلق خطا یا مجموع مربعات خطای داده‌های آموزشی را حداقل می‌کنند، ولی مدل‌های SVM، اصل حداقل‌سازی خطای ساختاری را به کار می‌گیرند (Hamel, 2009). اخیراً این مدل‌ها در گستره وسیعی از مسائل هیدرولوژیکی به کار رفته‌اند. به عنوان مثال، می‌توان به تحقیقات محققینی نظیر (Adamowski and Prasher, 2012) در پیش‌بینی جریان روزانه رودخانه و (Yu et al, 2005)، (Karahhan et al, 2014) در پیش‌بینی سیلاب اشاره

تبخیر-تعرق از مهم‌ترین اجزای چرخه هیدرولوژی بوده و برآورد دقیق آن در مطالعات زیادی از قبیل بیان آب، برنامه‌ریزی آبیاری، مدیریت منابع آب و طراحی سیستم‌های آبیاری کاربرد دارد. با توجه به عوامل مختلف تأثیرگذار در پدیده تبخیر-تعرق و ماهیت غیرخطی آن، تهیه یک مدل ریاضی برای این پدیده با در نظر گرفتن تمامی عوامل، کاری دشوار می‌باشد. بنابراین، استفاده از مدل‌های هوشمند مانند ماشین بردار پشتیبان و برنامه‌ریزی ژنتیک می‌تواند ابزارهای مناسبی برای تخمین متغیرهای غیرخطی همچون تبخیر-تعرق باشد. ماشین بردار پشتیبان^۵ (SVM) یکی از روش‌های یادگیری تحت نظارت^۶ است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده

۱- دانشجوی دکترای مهندسی منابع آب دانشگاه شهید چمران اهواز

۲- استادیار گروه مهندسی منابع آب دانشگاه شهید چمران اهواز

۳- استادیار گروه مهندسی آب دانشگاه شهرکرد

۴- دانشجوی کارشناسی ارشد مهندسی منابع آب دانشگاه ارومیه

(Email: S.ayashm@gmail.com)

*- نویسنده مسئول

5- Support Vector Machine (SVM)

6- Supervised learning

7- Statistical Learning Theory

داد که بهینه‌سازی شبکه عصبی مصنوعی موفق‌تر از منطق فازی بود، به‌طوریکه الگوریتم ژنتیک در ترکیب با شبکه عصبی توانست خطای آزمون شبکه را نسبت به عصبی-فازی به میزان ۲۴ درصد کاهش دهد. اما ترکیب منطق فازی نه تنها سبب بهبود نتایج شبکه عصبی مصنوعی نشد، بلکه دقت پیش‌بینی را به میزان ۱۲ درصد کاهش داد. ستاری و همکاران (۱۳۹۲) تبخیر-تعرق مرجع روزانه ایستگاه سینوپتیک بناب را با استفاده از شبکه عصبی مصنوعی و مدل درختی M5 پیش‌بینی کردند. نتایج نشان داد که گرچه روش شبکه عصبی مصنوعی با اختلاف خیلی کمی در مقایسه با روش M5 پیش‌بینی نسبتاً دقیق‌تری از میزان تبخیر-تعرق مرجع دارد، اما روش M5 روابط ساده خطی، قابل فهم‌تر و کاربردی‌تری را برای پیش‌بینی تبخیر-تعرق ارائه می‌کند.

جیا بینگ مقدار تبخیر-تعرق گیاه مرجع را با استفاده از سه روش منطق فازی، شبکه عصبی مصنوعی و سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) برآورد کردند. نتایج به دست آمده حاکی از آن بود که مدل ANFIS با ورودی‌های تعداد ساعات آفتابی و حداکثر دما توانسته با دقت بیش‌تری تبخیر-تعرق گیاه مرجع را پیش‌بینی کند (Jia Bing, 2004). دوگان با استفاده از مدل ANFIS تبخیر-تعرق گیاه مرجع را در ایستگاه مورگان سانفرانسیسکو محاسبه کرده و دقت بالای این مدل را در برآورد تبخیر-تعرق مورد تایید قرار دادند (Dogan, 2009). طبری و همکاران دقت مدل‌های ANFIS و SVM را در پیش‌بینی تبخیر-تعرق گیاه سیب زمینی مورد ارزیابی قرار دادند. ورودی‌های مختلفی برای هر دو مدل معرفی شد. نتایج نشان داد که گرچه دقت هر دو مدل در تخمین تبخیر-تعرق گیاه سیب‌زمینی بسیار نزدیک به هم بود، اما مدل SVM با ورودی‌های درجه حرارت متوسط، رطوبت نسبی، تابش خورشیدی، تعداد ساعات آفتابی و سرعت باد عملکرد بهتری در مقایسه با سیستم استنتاج عصبی-فازی تطبیقی داشت (Tabari et al, 2013).

با عنایت به موارد فوق می‌توان دریافت که تخمین پارامترهای هیدرولوژیکی همچون تبخیر-تعرق مرجع در تعیین نیاز آبی گیاهان و طراحی پروژه‌های عمرانی آبیاری بسیار مهم بوده و از دیرباز مورد توجه محققین امر قرار گرفته است. بدین منظور، روش‌های متعددی از جمله مدل‌های تجربی-نیمه تجربی، سری‌های زمانی و مدل‌های هوشمند توسعه یافته‌اند که در این میان مدل‌های هوشمند با الهام گرفتن از طبیعت قادر به تخمین پارامترهای مربوط به پدیده‌های طبیعی با دقت قابل قبول می‌باشند و نیز دقت قابل توجهی نسبت به سایر روش‌ها دارند. هم‌چنین دو مدل GP و SVM به صورت بسیار محدودی در زمینه پیش‌بینی تبخیر-تعرق به کار گرفته شده است. لذا هدف از تحقیق حاضر، توسعه مدل‌های مناسب GP و SVM جهت پیش‌بینی تبخیر-تعرق ماهانه گیاه مرجع در منطقه شمال غرب ایران و مقایسه عملکرد این دو روش می‌باشد.

برنامه‌ریزی ژنتیک^۱ (GP) شاخه‌ای از الگوریتم‌های تکاملی است که توانایی مدل‌سازی فرایندهای کاملاً غیرخطی و پویا را دارند. روش برنامه‌ریزی ژنتیک اولین بار توسط (Koza, 1992) توسعه داده شد. این روش جزو روش‌های الگوریتم تکاملی محسوب می‌شود که مبنای آن‌ها بر اساس نظریه تکامل داروین استوار است. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، جواب مناسب را ارائه می‌نمایند. روش برنامه‌ریزی ژنتیک، جدیدترین شیوه از بین روش‌های الگوریتم تکاملی می‌باشد که به دلیل دارا بودن دقت کافی، از کاربرد بیش‌تری برخوردار است (Alvisi et al, 2005). تفاوت اساسی موجود بین الگوریتم ژنتیک^۲ (GA) و برنامه‌ریزی ژنتیک به طبیعت هر یک از افراد بر می‌گردد، به نحوی که افراد در الگوریتم ژنتیک، ردیف‌های خطی با طول ثابت می‌باشند (کروموزم‌ها) ولی در برنامه‌ریزی ژنتیک، همانند شاخه‌های مجزا هستند. هم‌چنین در برنامه‌ریزی ژنتیک بر ساختار درختی مجموعه‌ها تاکید می‌شود، ولی الگوریتم ژنتیک بر اساس سیستم ارقام دو دویی عمل می‌نماید. روش GP در زمینه‌های مختلف مهندسی آب مورد استفاده محققین قرار گرفته است، از جمله تحقیقات انجام یافته با GP می‌توان به مطالعات (Guvem, 2009) برای پیش‌بینی جریان روزانه رودخانه‌ها، (Khu et al, 2001) در پیش‌بینی رواناب ساعتی، (Aytek et al, 2008) برای تعیین رابطه بارش-رواناب، (Ghorbani et al, 2010) در روندیابی سیلاب و (Zahiri and Azamatullah, 2014) در پیش‌بینی دبی در مقاطع مرکب رودخانه‌ها اشاره کرد.

تاکنون مطالعات فراوانی برای پیش‌بینی تبخیر-تعرق گیاه مرجع^۳ (ET₀) با روش‌های مختلف صورت گرفته است که در ادامه به برخی از این مطالعات اشاره می‌شود. زارع ابیانه و همکاران (۱۳۸۸) از شبکه عصبی مصنوعی^۴ و سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) برای برآورد تبخیر-تعرق گیاه مرجع در منطقه همدان استفاده کردند. نتایج حاصله نشان دهنده برتری مدل ANFIS با ورودی‌های درجه حرارت بیشینه و کمینه و ساعات آفتابی بود. زارع ابیانه و همکاران (۱۳۹۰) تبخیر-تعرق واقعی گیاه سیر را با استفاده از روش‌های شبکه عصبی مصنوعی، سیستم استنتاج عصبی-فازی تطبیقی و روش تلفیقی عصبی-ژنتیک^۵ برآورد نمودند. نتایج نشان

- 1- Genetic Programming
- 2- Genetic Algorithm
- 3- Grass reference crop evapotranspiration
- 4- Artificial Neural Networks (ANN)
- 5- Adaptive Neural Network Based Genetic Algorithm (ANNGA)

مواد و روش‌ها

منطقه مورد مطالعه

برای محاسبه تبخیر- تعرق ماهانه گیاه مرجع در منطقه شمال غرب کشور شامل استان‌های آذربایجان غربی و شرقی، زنجان و کردستان، ایستگاه‌های سینوپتیک مورد بررسی قرار گرفت و در نهایت ۶ ایستگاه به دلیل دارا بودن اطلاعات طولانی مدت در دوره آماری ۳۸ ساله (۲۰۱۰-۱۹۷۳) و پراکنش مناسب در سطح منطقه مورد مطالعه انتخاب شدند. ایستگاه‌های منتخب متوسط دمای منطقه و بارش سالانه را به ترتیب ۱۲/۸ درجه سانتی‌گراد و ۳۶۰ میلی‌متر نشان می‌دهند که براساس طبقه‌بندی اقلیمی دومارتن، اقلیم این منطقه با شاخص ۱۵/۷۹ از نوع آب و هوای نیمه خشک می‌باشد. در جدول ۱ مشخصات ایستگاه‌های منتخب ارائه و شکل ۱ نیز موقعیت مکانی این ایستگاه‌ها را نشان می‌دهد.

برنامه‌ریزی ژنتیک (GP)

برنامه‌ریزی ژنتیک تعمیم یافته الگوریتم ژنتیک می‌باشد که برای اولین بار بر اساس تئوری داروین ارائه شد. به این ترتیب که جمعیتی در جهت تکامل به صورت انتخابی، جمعیت نامناسب را رها کرده و فرزندی اصلاح شده ایجاد می‌کنند. در برنامه‌ریزی ژنتیک، ابتدا

ساختارهای موجود که شامل متغیرهای ورودی و هدف و نیز تابع ارتباط دهنده آن‌ها می‌باشد، تعریف گردیده و سپس ساختار مناسب مدل و ضرایب آن تعیین می‌شود. این روش شامل یک معادله ارتباط- دهنده بین متغیرهای ورودی و خروجی بوده و لذا قادر به انتخاب خودکار متغیرهای مناسب مدل و حذف متغیرهای غیرمرتبط است که این امر سبب کاهش ابعاد متغیرهای ورودی خواهد شد. فرآیند گام به گام حل یک مسئله با استفاده از برنامه‌ریزی ژنتیک متشکل از ۵ مرحله به شرح زیر می‌باشد (Ghorbani et al, 2010):

- ۱) انتخاب مجموعه ترمینال: که همان متغیرهای مستقل مسئله و متغیرهای حالت سامانه می‌باشند. ۲) انتخاب مجموعه توابع: که شامل عملگرهای حسابی، توابع آزمون و توابع بولی می‌باشد. ۳) شاخص اندازه‌گیری دقت مدل که بر مبنای آن می‌توان مشخص نمود که توانایی مدل در حل یک مسئله خاص تا چه اندازه می‌باشد. ۴) مؤلفه‌های کنترل: مقادیر مؤلفه‌های عددی و متغیرهای کیفی که برای کنترل اجرای برنامه‌ها استفاده می‌شوند. ۵) شرط توقف اجرای برنامه: که معیاری برای حصول نتیجه و توقف اجرای برنامه می‌باشد.
- در مطالعه حاضر از برنامه GeneXpro Tools (Ferreira, 2001) برای توسعه و اجرای مدل‌های مبتنی بر برنامه‌ریزی ژنتیک استفاده به عمل آمد.



شکل ۱- موقعیت مکانی ایستگاه‌های منتخب در شمال غرب ایران

جدول ۱- مشخصات ایستگاه‌های منتخب در دوره آماری ۲۰۱۰-۱۹۷۳

ردیف	ایستگاه	طول جغرافیایی		عرض جغرافیایی		ارتفاع (متر)	متوسط دما (°C)	متوسط بارش سالانه (mm)	نوع اقلیم در طبقه بندی دومارتن
		درجه	دقیقه	درجه	دقیقه				
۱	ارومیه	۴۵	۰۵	۳۷	۳۲	۱۳۲۸/۰	۱۱/۹	۳۳۸	نیمه خشک
۲	تبریز	۴۶	۱۷	۳۸	۰۵	۱۳۶۱/۰	۱۲/۸	۲۸۳	نیمه خشک
۳	خوی	۴۴	۵۸	۳۸	۳۳	۱۱۰۳/۰	۱۲/۱	۲۸۹	نیمه خشک
۴	زنجان	۲۹	۴۸	۳۶	۴۱	۱۶۶۳/۰	۱۱/۰	۳۱۱	نیمه خشک
۵	سقز	۴۶	۱۶	۳۶	۱۵	۱۵۲۲/۸	۱۳/۶	۴۸۷	مدیترانه‌ای
۶	سنندج	۴۷	۰۰	۳۵	۲۰	۱۳۷۳/۴	۱۳/۶	۴۴۹	نیمه خشک

آن در مسائل رگرسیونی مورد استفاده قرار گرفت. برای این مدل، تابع خطا به صورت زیر تعریف می‌شود:

$$\frac{1}{2}W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (2)$$

تابع خطای فوق لازم است که با توجه به محدودیت‌های زیر کمینه گردد (Hamel, 2009):

$$\begin{aligned} W^T \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - W^T \phi(x_i) - b &\leq \varepsilon + \xi_i^* \end{aligned} \quad (3)$$

$\xi_i, \xi_i^* \geq 0$
 که در این روابط، C ثابت گنجایش، W بردار ضرایب، W^T ترانهاده بردار ضرایب، ξ_i, ξ_i^* ضرایب کمبود، b ضریب ثابت، N الگوهای آموزش مدل و ϕ تابع کرنل است. اطلاعات کمی در مورد انتخاب تابع غیرخطی مناسب ϕ در دسترس می‌باشد. ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مساله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVM به حجم داده‌های 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 ورودی‌های مساله را داشته باشد. در عمل چهار نوع کرنل خطی^۱، کرنل چند جمله‌ای^۲، کرنل تانژانت هیپربولیک^۳ و کرنل گوسی^۴ (RBF) به کار گرفته می‌شوند. در جدول ۲ معادلات برخی از کرنل‌های رایج ارائه شده اند.

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت معادله ۴ خواهد بود که کنترل‌کننده میزان نوسان تابع گوسی و همچنین کنترل‌کننده نتایج پیش‌بینی و تعمیم‌دهنده مدل SVM است (Yu et al, 2005).

$$f(x_i) = \sum_{i=1}^l (-\partial_i - \partial_i^*) K(x_i, x_j) + b \quad (4)$$

انتخاب داده‌های ورودی

توانمندی پیش‌بینی مدل‌های هوشمند به انتخاب صحیح نوع اطلاعات ورودی در لایه اول بستگی دارد. بنابراین، در مطالعه حاضر دو حالت برای داده‌های ورودی در نظر گرفته شد. در حالت اول به منظور بررسی نقش و توانایی حافظه در پیش‌بینی تبخیر-تعرق، از تاخیرهای زمانی یک، دو، سه و چهار ماهه استفاده گردید و هدف پاسخ به این سوال است که آیا می‌توان از حافظه تبخیر-تعرق برای

برنامه‌ریزی ژنتیک به استنتاج برنامه‌های رایانه‌ای با اندازه‌ها و شکل‌های مختلف می‌پردازد. یکی از نقاط قوت GP آن است که معیار تنوع ژنتیکی در آن بسیار ساده بوده و لذا عملگرهای ژنتیک در سطح کروموزوم عمل می‌نمایند. همچنین یکی دیگر از نقاط قوت این روش، طبیعت منحصر به فرد چند ژنی بودن آن است که زمینه ارزیابی مدل‌های پیچیده‌ای را که شامل چندین زیرمدل می‌باشند، فراهم می‌آورد (Ferreira, 2001). در تحقیق حاضر، مراحل مدل-سازی تبخیر-تعرق ماهانه گیاه مرجع در منطقه شمال غرب کشور به ترتیب زیر انجام شد.

گام اول، انتخاب تابع برازش مناسب می‌باشد که در این مطالعه تابع جذر میانگین مربعات خطا به عنوان تابع برازش انتخاب گردید. گام دوم، انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به منظور تولید کروموزوم‌ها می‌باشد. در مسئله حاضر، مجموعه ترمینال‌ها متشکل از مقادیر تبخیر-تعرق ماهانه با تاخیرهای زمانی و اطلاعات اقلیمی می‌باشد. در این مطالعه از چهار عملگر اصلی که شامل $\{+, -, \times, \div, X^2, X^3, \sqrt{X}, \sqrt[3]{X}, \text{Log}(X)\}$ استفاده شد. گام سوم، شامل انتخاب ساختار و معماری کروموزوم‌ها می‌باشد. اندازه طول هر راس و تعداد ژن‌ها بر اساس تعداد متغیرهای ورودی انتخاب شد. گام چهارم انتخاب تابع پیوندی است. در نهایت، در گام پنجم عملگرهای ژنتیک و نرخ هریک از آن‌ها تعیین گردید.

ماشین بردار پشتیبان

در یک مدل رگرسیونی SVM لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به علاوه یک مقدار اضافی نویز^۱ مشخص می‌شود.

$$y = f(x) + \text{Noise} \quad (1)$$

بنابراین موضوع اصلی، پیدا کردن فرم تابع f است که بتواند به صورت صحیح، موارد جدیدی را که SVM تاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به وسیله آموزش مدل SVM بر روی یک مجموعه داده به عنوان مجموعه آموزش که شامل فرآیندی به منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVM شناخته شده است که عبارتند از: الف) مدل‌های رگرسیونی SVM نوع اول که مدل‌های $SVM - \nu$ نیز نامیده می‌شوند و ب) مدل‌های رگرسیونی SVM نوع دوم که با نام $SVM - \varepsilon$ شناخته شده هستند. در این مطالعه، مدل $SVM - \varepsilon$ به دلیل کاربرد گسترده

2- Linear kernel

3- Polynomial kernel

4- Hyperbolic tangent kernel

5- Radial Base Function kernel

1- Noise

$$RMSE = \left(\frac{\sum_{i=1}^n ((ET_{0i}(FPM) - ET_{0i}(cal))^2)}{n} \right)^{0.5} \quad (۴)$$

که در روابط فوق $ET_{0i}(FPM)$ مقادیر مشاهداتی تبخیر- تعرق گیاه مرجع (به دست آمده از معادله فائو- پمن- مانتیث) در گام زمانی n ام، $ET_{0i}(cal)$ مقدار محاسبه شده در همان زمان، n تعداد داده‌ها و $ET_0(FPM)$ میانگین مقادیر مشاهداتی می‌باشد. هر چه مقدار ضریب همبستگی بالاتر و RMSE کمتر باشد، نتایج دقیق‌تر و قابل اعتمادتر خواهند بود.

نتایج

داده‌های ورودی به مدل‌ها

در این تحقیق ۱۰ الگو با ترکیبات مختلف از داده‌های موجود به مدل‌های SVM و GP معرفی گردید. در حالت اول فقط از حافظه تبخیر- تعرق استفاده شد و با تاخیرهای یک، دو، سه و چهار ماهه الگوهای موردنظر ساخته شد. در حالت دوم، با ارائه مدل رگرسیونی، ارتباط بین متغیرهای مستقل هواشناسی و تبخیر- تعرق گیاه مرجع برای هر یک از ایستگاه‌ها مشخص گردید. جدول ۳ روابط رگرسیونی را برای ایستگاه‌های مورد مطالعه نشان می‌دهد. با توجه به این جدول چنین استنباط می‌شود که تابش خورشیدی و رطوبت نسبی به ترتیب بیش‌ترین و کم‌ترین اثرگذاری را در تبخیر- تعرق گیاه مرجع دارا هستند (به دلیل داشتن بزرگ‌ترین و کوچک‌ترین ضرایب همبستگی).

جدول ۳- معادله کلی رگرسیون برای تبخیر- تعرق گیاه مرجع ایستگاه‌های مورد مطالعه

ردیف	ایستگاه	معادله رگرسیون	ضریب همبستگی
۱	ارومیه	$ET_0 = 0.141T_{max} - 0.045T_{min} - 0.005RH_{ave} + 0.009U_2 + 0.053n + 0.141R$	۰/۸۶
۲	تبریز	$ET_0 = 0.0316T_{max} + 0.078T_{min} - 0.024RH_{ave} + 0.019U_2 + 0.093n + 0.119R$	۰/۸۱
۳	خوی	$ET_0 = 0.0365T_{max} + 0.064T_{min} - 0.018RH_{ave} + 0.027U_2 - 0.023n + 0.127R$	۰/۸۳
۴	زنجان	$ET_0 = 0.042T_{max} + 0.046T_{min} - 0.006RH_{ave} + 0.013U_2 - 0.145n + 0.120R$	۰/۸۴
۵	سقز	$ET_0 = 0.036T_{max} + 0.061T_{min} - 0.002RH_{ave} + 0.003U_2 - 0.038n + 0.158R$	۰/۸۸
۶	سندج	$ET_0 = 0.078T_{max} + 0.018T_{min} - 0.004RH_{ave} + 0.005U_2 + 0.014n + 0.142R$	۰/۸۷

*: در روابط فوق ET_0 تبخیر- تعرق گیاه مرجع (میلی‌متر در روز)، T_{min} و T_{max} به ترتیب دمای هوای کمینه و بیشینه (درجه سانتی‌گراد)، RH_{ave} متوسط رطوبت نسبی (درصد)، U_2 سرعت باد در ارتفاع دو متری (کیلومتر بر روز)، n تعداد ساعات آفتابی (ساعت) و R تابش خورشیدی ($MJ/m^2/day$) می‌باشند.

پیش‌بینی و مدل‌سازی خود پدیده نیز استفاده کرد یا خیر؟ در حالت دوم پارامترهای اقلیمی مؤثر در تبخیر- تعرق توسط روش رگرسیون شناسایی شد. سپس ترکیبات مختلف ورودی با داده‌های اقلیمی ساخته و به مدل‌ها معرفی گردید. در این حالت متغیری که کم‌ترین ضریب رگرسیون را به خود اختصاص داده بود، حذف شد تا از خطا و پیچیدگی مدل کاسته شود.

جدول ۲- توابع کرنل رایج در ماشین‌های بردار پشتیبان (Hamel., 2009)

نوع تابع	تابع کرنل
خطی	$K(x_i, x_j) = x_i^T \cdot x_j$
چند جمله‌ای	$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + C)^d$
تانژانت هیپربولیک	$K(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + C)$
RBF	$K(x_i, x_j) = \exp\left(-\gamma x_i - x_j ^2\right)$

معیارهای ارزیابی مدل

در این تحقیق، برای ارزیابی مدل‌های مورد نظر از معیارهای ضریب همبستگی و جذر میانگین مربعات خطا استفاده می‌شود:

$$R = \left(1 - \frac{\sum_{i=1}^n (ET_{0i}(FPM) - ET_{0i}(cal))^2}{\sum_{i=1}^n (ET_{0i}(FPM) - \overline{ET_0(FPM)})^2} \right)^{0.5} \quad (۵)$$

جدول ۴- الگوهای مختلف معرفی شده به مدل‌های GP و SVM

حالت‌های ورودی	شماره الگو	الگو	الگوهای ورودی به مدل‌های GP و SVM	خروجی مدل‌های GP و SVM
حالت اول: استفاده از حافظه سری تبخیر- تعرق	۱	M1	$ET_0(t-1)$	$ET_0(t)$
	۲	M2	$ET_0(t-1), ET_0(t-2)$	
	۳	M3	$ET_0(t-1), ET_0(t-2), ET_0(t-3)$	
	۴	M4	$ET_0(t-1), ET_0(t-2), ET_0(t-3), ET_0(t-4)$	
	۵	M5	$ET_0(t-1), ET_0(t-2), ET_0(t-3), ET_0(t-4), ET_0(t-5)$	
	۶	M6	Tmin, Tmax	
	۷	M7	Tmin, Tmax, Tave	
	۸	M8	Tmin, Tmax, Tave, R	
	۹	M9	Tmin, Tmax, Tave, R, U_2	
	۱۰	M10	Tmin, Tmax, Tave, R, U_2, n	
حالت دوم: استفاده از داده‌های اقلیمی				

جدول ۵- دقت الگوهای مختلف ورودی مدل GP در پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع در شمال غرب ایران

ایستگاه	مرحله	معیارهای ارزیابی	الگوهای ورودی									
			M10	M9	M8	M7	M6	M5	M4	M3	M2	M1
ارومیه	آموزش	R	۰/۸۴۹	۰/۹۵۶	۰/۹۶۴	۰/۹۷۱	۰/۹۷۳	۰/۹۵۳	۰/۹۵۸	۰/۹۶۹	۰/۹۷۹	۰/۹۸۰
		RMSE	۱/۰۹	۰/۶۱	۰/۵۴	۰/۴۸	۰/۴۲	۰/۶۲	۰/۵۹	۰/۵۲	۰/۴۲	۰/۴۳
	صحت سنجی	R	۰/۸۴۴	۰/۹۴۱	۰/۹۵۰	۰/۹۶۵	۰/۹۶۱	۰/۹۶۶	۰/۹۷۳	۰/۹۷۰	۰/۹۸۹	۰/۹۸۲
		RMSE	۱/۱۷	۰/۷۶	۰/۷۰	۰/۶۰	۰/۶۵	۰/۶۶	۰/۶۵	۰/۷۸	۰/۵۰	۰/۵۵
تبریز	آموزش	R	۰/۸۳۹	۰/۹۵۴	۰/۹۵۳	۰/۹۶۷	۰/۹۶۸	۰/۹۶۲	۰/۹۷۶	۰/۹۷۲	۰/۹۵۱	
		RMSE	۱/۵۳	۰/۸۴	۰/۸۷	۰/۷۰	۰/۶۸	۰/۸۱	۰/۶۰	۰/۶۵	۰/۸۶	
	صحت سنجی	R	۰/۸۳۹	۰/۹۴۶	۰/۹۴۱	۰/۹۶۵	۰/۹۶۰	۰/۹۶۹	۰/۹۷۹	۰/۹۷۷	۰/۹۷۲	
		RMSE	۱/۶۰	۰/۸۹	۱/۰۴	۰/۷۴	۰/۸۱	۰/۸۶	۰/۶۲	۰/۸۳	۰/۷۰	
خوی	آموزش	R	۰/۸۴۲	۰/۹۴۲	۰/۹۶۲	۰/۹۶۶	۰/۹۵۳	۰/۹۴۳	۰/۹۴۱۴	۰/۹۸۰	۰/۹۳۷	
		RMSE	۰/۹۷	۰/۶۱	۰/۵۰	۰/۴۷	۰/۷۲	۰/۵۹	۰/۶۱	۰/۳۵	۰/۶۲	
	صحت سنجی	R	۰/۸۳۴	۰/۹۲۳	۰/۹۵۶	۰/۹۶۲	۰/۹۴۲	۰/۹۵۲	۰/۹۵۰	۰/۹۷۸	۰/۹۳۴	
		RMSE	۱/۱۳	۰/۸۰	۰/۶۷	۰/۶۱	۰/۸۹	۰/۷۱	۰/۶۸	۰/۶۰	۰/۸۴	
زنجان	آموزش	R	۰/۸۴۶	۰/۹۵۸	۰/۹۶۱	۰/۹۴۷	۰/۹۵۱	۰/۹۳۹	۰/۹۵۳	۰/۹۷۱	۰/۹۷۱	
		RMSE	۱/۲۷	۰/۷۰	۰/۶۵	۰/۷۹	۰/۷۰	۰/۸۳	۰/۷۲	۰/۵۷	۰/۵۷	
	صحت سنجی	R	۰/۸۴۳	۰/۹۵۷	۰/۹۶۰	۰/۹۶۶	۰/۹۶۱	۰/۹۶۸	۰/۹۷۰	۰/۹۷۸	۰/۹۸۸	
		RMSE	۱/۲۷	۰/۷۲	۰/۶۸	۰/۶۱	۰/۶۹	۰/۶۴	۰/۶۰	۰/۳۸	۰/۴۵	
سقز	آموزش	R	۰/۸۵۲	۰/۹۵۷	۰/۹۷۱	۰/۹۶۵	۰/۹۶۷	۰/۹۷۱	۰/۹۲۰	۰/۹۶۵	۰/۹۶۱	
		RMSE	۱/۱۳	۰/۶۳	۰/۵۷	۰/۵۲	۰/۴۹	۰/۸۵	۰/۷۷	۰/۵۶	۰/۶۰	
	صحت سنجی	R	۰/۸۳۶	۰/۹۴۹	۰/۹۶۲	۰/۹۶۵	۰/۹۶۰	۰/۹۶۲	۰/۹۶۷	۰/۹۸۶	۰/۹۷۱	
		RMSE	۱/۳۰	۰/۷۵	۰/۶۸	۰/۶۰	۰/۷۳	۰/۹۷	۰/۸۶	۰/۶۶	۰/۶۸	
سندج	آموزش	R	۰/۸۴۶	۰/۹۵۸	۰/۹۶۱	۰/۹۴۷	۰/۹۴۹	۰/۹۳۹	۰/۹۵۳	۰/۹۷۱	۰/۹۷۱	
		RMSE	۱/۲۷	۰/۷۰	۰/۶۵	۰/۷۹	۰/۷۶	۰/۸۳	۰/۷۲	۰/۵۷	۰/۵۷	
	صحت سنجی	R	۰/۸۴۳	۰/۹۴۷	۰/۹۶۶	۰/۹۶۹	۰/۹۶۴	۰/۹۶۸	۰/۹۷۰	۰/۹۷۸	۰/۹۸۸	
		RMSE	۱/۲۸	۰/۸۲	۰/۶۱	۰/۵۸	۰/۶۱	۰/۶۳	۰/۶۰	۰/۳۸	۰/۴۵	

* واحد RMSE بر حسب میلی‌متر در روز می‌باشد.

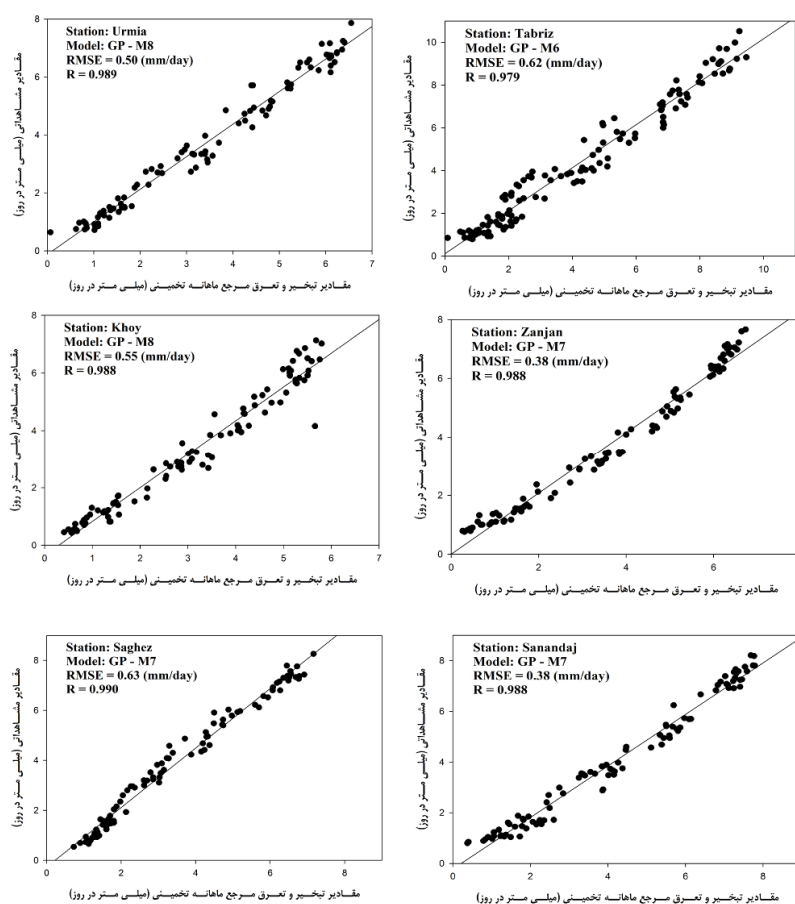
بایستی سعی نمود مؤثرترین داده‌های مشاهداتی را به عنوان داده‌های آموزشی انتخاب کرد. در این مطالعه، برای انتخاب هرچه بهتر و دقیق‌تر اطلاعات ورودی، از روش رگرسیون و حافظه‌سازی برای سری تبخیر- تعرق کمک گرفته شد و الگوهای ورودی به صورت M1 تا M10 (جدول ۲) به مدل GP ارائه شد. جدول ۵ مقایسه شاخص‌های آماری مربوط به نتایج حاصل از کاربرد روش برنامه‌ریزی ژنتیک در پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع در شمال غرب ایران را برای الگوهای ورودی مختلف نشان می‌دهد.

با توجه به جدول ۵، می‌توان نتیجه گرفت که در کل عملکرد GP در پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع در این منطقه مناسب می‌باشد. با انجام مقایسه بین حالتی که از حافظه سری به عنوان ورودی‌های مدل استفاده می‌شود (حالت اول) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر تبخیر- تعرق بهره گرفته می‌شود (حالت دوم)، حالت دوم از دقت بیش‌تری برخوردار است.

جدول ۴ ترکیب‌های مختلف ورودی و متغیرهای انتخاب شده برای مدل‌های SVM و GP را نشان می‌دهد. همان‌گونه که مشاهده می‌گردد رطوبت نسبی به دلیل تأثیرگذاری پایین از ترکیبات ورودی کنار گذاشته شده است.

نتایج مدل‌سازی تبخیر- تعرق ماهانه گیاه مرجع با استفاده از برنامه‌ریزی ژنتیک

در مطالعه حاضر، برای مدل‌سازی تبخیر- تعرق ماهانه گیاه مرجع با استفاده از روش GP، داده‌های ۳۰ سال (از ژانویه ۱۹۷۳ تا دسامبر ۲۰۰۲) برای آموزش و ۸ سال (از ژانویه ۲۰۰۳ تا دسامبر ۲۰۱۰) برای صحت‌سنجی انتخاب شدند. انتخاب ورودی‌های مختلف و تأثیرگذار در پدیده (که در برنامه‌ریزی ژنتیک به عنوان داده‌های آموزشی از آن‌ها یاد می‌شود) به منظور آموزش ماهیت سازوکار حاکم بر پدیده، نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت الگو نیز می‌شود. لذا در الگوسازی تبخیر- تعرق نیز



شکل ۲- نمودارهای پراکنده مقادیر مشاهداتی و تخمینی تبخیر- تعرق ماهانه گیاه مرجع در بهترین حالت مدل GP برای ایستگاه‌های منتخب شمال غرب ایران

تبخیر- تعرق ماهانه گیاه مرجع استفاده شد. برای هر الگوی ورودی با تغییر ϵ ، C و γ شبکه‌های مختلفی ساخته و آموزش داده شد و در نهایت ساختاری که دارای کم‌ترین خطا و بیش‌ترین دقت بود، به عنوان مناسب‌ترین الگو انتخاب شد. جدول ۷ شاخص‌های آماری مربوط به نتایج حاصل از کاربرد مدل SVM را برای هر الگوی ورودی نشان می‌دهد. با توجه به این جدول می‌توان نتیجه گرفت که در کل عملکرد SVM در پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع در ایستگاه‌های مورد بررسی بسیار مناسب می‌باشد. حال اگر از حافظه خود سری تبخیر- تعرق ماهانه برای مدل‌سازی استفاده گردد، بر اساس جدول ۷ مشاهده می‌شود که دقت مدل‌سازی رفته رفته تا چهار تاخیر زمانی رو به بهبود بوده و در تمامی ایستگاه‌ها در گام زمانی چهارم بیشینه دقت حاصل شده و بعد از آن عملکرد مدل کاهش می‌یابد. با انجام مقایسه بین حالتی که از حافظه سری به عنوان ورودی‌های مدل استفاده می‌شود (حالت اول) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر تبخیر- تعرق بهره گرفته می‌شود (حالت دوم) می‌توان نتیجه گرفت حالت دوم از دقت بیش‌تری برخوردار است، هرچند اختلاف دقت بین دو حالت آنچنان زیاد نبوده و می‌توان با بیش‌تر کردن میزان خطای قابل قبول در شرایطی که داده‌های هواشناسی به طور کامل در دسترس نباشد از حافظه تبخیر- تعرق به عنوان ورودی‌های مدل استفاده کرد.

هم‌چنین جدول ۷ نشان می‌دهد که مدل SVM برای ایستگاه‌های ارومیه و سنندج با ورودی‌های دمای بیشینه، متوسط و کمینه، تابش خورشیدی و سرعت باد (الگوی M9) بهترین عملکرد و برای ایستگاه‌های تبریز، خوی، زنجان و سقز الگوی M10 با ورودی‌های دمای بیشینه، متوسط و کمینه، تابش خورشیدی، سرعت باد و ساعات آفتابی بالاترین دقت و کم‌ترین خطا را داشته است. الگوی M6 نیز با ورودی‌های دمای بیشینه و کمینه پایین‌ترین دقت و بیش‌ترین خطا را در تخمین تبخیر- تعرق ماهانه گیاه مرجع برای تمامی ایستگاه‌ها ارائه کرده است. شکل ۴ نمودارهای پراکندگی مقادیر مشاهداتی و تخمینی را در بهترین حالت مدل SVM برای هر ایستگاه نشان می‌دهد.

نتیجه‌گیری و بحث

در مطالعه حاضر از داده‌های ایستگاه‌های سینوپتیک ارومیه، تبریز، خوی، زنجان، سقز و سنندج در دوره آماری ۳۸ ساله (۲۰۱۰-۱۹۷۳) برای پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع با استفاده از دو روش برنامه‌ریزی ژنتیک (GP) و ماشین بردار پشتیبان (SVM) استفاده شد. به طور کلی می‌توان نتایج حاصله را به شرح زیر خلاصه نمود:

هم‌چنین جدول ۵ نشان می‌دهد که مدل GP برای ایستگاه‌های ارومیه و خوی با ورودی‌های دمای بیشینه، متوسط و کمینه، تابش خورشیدی و سرعت باد (الگوی M9)، برای ایستگاه‌های زنجان، سقز و سنندج الگوی M8 (با ورودی‌های دمای بیشینه، متوسط، کمینه و تابش خورشیدی) و برای ایستگاه تبریز الگوی M7 (با ورودی‌های دمای بیشینه، متوسط و کمینه) بهترین عملکرد را داشته است. شکل ۲ نمودارهای پراکندگی مقادیر مشاهداتی و تخمینی را در بهترین حالت مدل GP برای هر ایستگاه نشان می‌دهد.

نتایج مدل‌سازی تبخیر- تعرق ماهانه گیاه مرجع با استفاده از ماشین بردار پشتیبان

در مطالعه حاضر، برای مدل‌سازی تبخیر- تعرق ماهانه گیاه مرجع در ابتدا بایستی مقادیر بهینه مشخصه‌های مدل SVM شامل ϵ و C تعیین گردد. هم‌چنین در این مطالعه تابع کرنل مورد استفاده، تابع RBF انتخاب شد، چرا که از دقت بهتری در برآورد تبخیر- تعرق ماهانه گیاه مرجع نسبت به سایر توابع کرنل برخوردار بود. در این تابع نیز مشخصه γ بایستی تعیین شود. بنابراین، در حالت کلی برای پیش‌بینی تبخیر- تعرق توسط مدل SVM رگرسیونی، لازم است که مقادیر بهینه سه مشخصه مذکور به دست آید که بدین منظور دو مشخصه ϵ و C توسط الگوریتم بهینه‌سازی جستجوی شبکه^۱ و متغیر γ نیز به صورت آزمون و خطا محاسبه شد. البته قابل ذکر است که الگوریتم بهینه‌سازی جستجوی شبکه بسیار کند عمل می‌کند و زمان محاسباتی زیادی را به خود اختصاص می‌دهد. برای حل این مشکل، طبق توصیه (Chen and Yu, 2006) از برنامه اصلاح‌شده الگوریتم جستجوی شبکه که به نام الگوریتم جستجوی شبکه دو مرحله‌ای^۲ معروف است به همراه اعتبارسنجی متقاطع^۳ استفاده شد. برای این منظور ابتدا با انتخاب شبکه‌هایی با ابعاد بزرگ، محدوده مشخصه‌های ϵ و C به ازای مقدار ثابت مشخصه γ تعیین شد. سپس با مشخص شدن محدوده مذکور و تقسیم آن به شبکه‌هایی با ابعاد ریزتر مقادیر دقیق دو مشخصه ϵ و C مشخص گردیدند. روند مذکور برای دیگر مقادیر γ نیز تکرار شد و بدین طریق مدل‌های متفاوتی با تغییر در مقدار γ حاصل شد. حال می‌توان از بین مدل‌های توسعه داده شده، مدل با کم‌ترین خطا را تعیین کرده و مشخصه‌های آن را به عنوان مقادیر بهینه ϵ ، C و γ انتخاب نمود. در جدول ۶ مقادیر بهینه ϵ ، C و γ برای تمامی ورودی‌های مدل SVM ارائه شده است. در مطالعه حاضر از مدل SVM نیز با همان الگوهای ورودی GP برای تخمین

- 1- Grid Search
- 2- Two-Steps Grid Search
- 3- Cross-Validation

جدول ۶- مقادیر بهینه ضرایب C، ε و γ برای الگوهای مختلف ورودی‌های مدل SVM

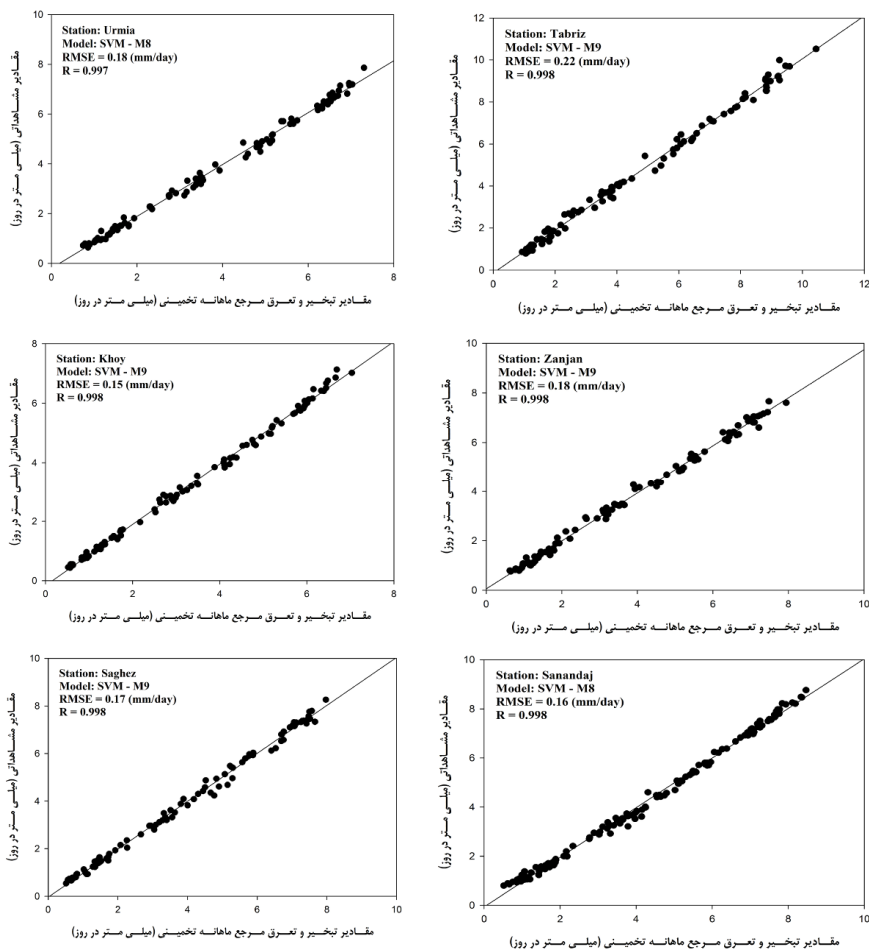
ایستگاه	ضرایب مدل SVM	الگوهای ورودی									
		M10	M9	M8	M7	M6	M5	M4	M3	M2	M1
ارومیه	C	۵	۵	۸	۶	۷	۸	۱۰	۶	۱۰	۴
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۵۰
	γ	۰/۱۷	۰/۲۰	۰/۲۵	۰/۳۳	۰/۵۰	۰/۴۵	۰/۲۵	۰/۳۳	۰/۵	۱/۰۰
تبریز	C	۳	۳	۱۰	۱۰	۸	۸	۹	۹	۷	۲
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۳۰
	γ	۰/۱۶	۰/۲۰	۰/۲۵	۰/۳۰	۰/۵۵	۰/۳۵	۰/۳۵	۰/۴۳	۰/۶۵	۱/۰۰
خوی	C	۱۰	۱۰	۹	۵	۸	۸	۱۰	۵	۶	۷
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۲
	γ	۰/۱۷	۰/۲۹	۰/۳۵	۰/۴۱	۰/۶۰	۰/۵۰	۰/۴۵	۰/۳۱	۰/۵۹	۱/۰۰
زنجان	C	۵	۴	۱۰	۹	۸	۸	۹	۹	۸	۱۰
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۲۰
	γ	۰/۱۷	۰/۲۰	۰/۱۵	۰/۳۰	۰/۵۰	۰/۳۹	۰/۳۲	۰/۲۱	۰/۴۵	۱/۰۰
سقز	C	۸	۱۰	۲	۷	۶	۶	۷	۸	۸	۱۰
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۵۰
	γ	۰/۱۶	۰/۲۰	۰/۱۵	۰/۱۳	۰/۷۲	۰/۵۰	۰/۳۲	۰/۴۲	۰/۶۱	۱/۰۰
سنندج	C	۸	۷	۱۰	۸	۴	۸	۱۰	۶	۶	۱۰
	ε	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۱۰	۰/۵۰
	γ	۰/۱۸	۰/۲۰	۰/۲۵	۰/۴۳	۰/۲۸	۰/۳۸	۰/۲۵	۰/۱۲	۰/۵۰	۱/۰۰

جدول ۷- دقت الگوهای مختلف ورودی مدل SVM در پیش‌بینی تبخیر- تعرق ماهانه گیاه مرجع در شمال غرب ایران

ایستگاه	مرحله	معیارهای ارزیابی	الگوهای ورودی									
			M10	M9	M8	M7	M6	M5	M4	M3	M2	M1
ارومیه	آموزش	R	۰/۹۹۷	۰/۹۹۸	۰/۹۹۱	۰/۹۶۶	۰/۹۷۸	۰/۹۸۳	۰/۹۸۶	۰/۹۷۴	۰/۹۶۴	۰/۸۳۸
		RMSE	۰/۱۵	۰/۱۴	۰/۲۷	۰/۵۳	۰/۵۴	۰/۵۰	۰/۴۳	۰/۴۷	۰/۵۵	۱/۲۳
	صحت سنجی	R	۰/۹۹۸	۰/۹۹۷	۰/۹۹۴	۰/۹۷۷	۰/۹۷۸	۰/۹۶۹	۰/۹۷۴	۰/۹۶۴	۰/۹۵۰	۰/۸۳۸
		RMSE	۰/۲۰	۰/۱۸	۰/۴۵	۰/۵۸	۰/۶۲	۰/۶۱	۰/۵۷	۰/۶۴	۰/۷۲	۱/۲۵
تبریز	آموزش	R	۰/۹۹۷	۰/۹۹۷	۰/۹۹۲	۰/۹۸۲	۰/۹۸۲	۰/۹۷۸	۰/۹۷۷	۰/۹۷۳	۰/۹۶۴	۰/۸۳۹
		RMSE	۰/۲۲	۰/۲۳	۰/۳۸	۰/۵۰	۰/۵۳	۰/۵۹	۰/۶۱	۰/۶۵	۰/۷۵	۱/۵۳
	صحت سنجی	R	۰/۹۹۸	۰/۹۹۷	۰/۹۹۲	۰/۹۸۴	۰/۹۸۳	۰/۹۷۰	۰/۹۷۳	۰/۹۶۵	۰/۹۵۷	۰/۸۳۹
		RMSE	۰/۲۲	۰/۲۵	۰/۵۰	۰/۵۸	۰/۶۰	۰/۷۹	۰/۷۵	۰/۸۲	۰/۹۰	۱/۶۰
خوی	آموزش	R	۰/۹۹۷	۰/۹۹۷	۰/۹۸۲	۰/۹۵۳	۰/۹۵۲	۰/۹۷۳	۰/۹۷۰	۰/۹۶۱	۰/۹۴۶	۰/۸۴۲
		RMSE	۰/۱۶	۰/۱۶	۰/۳۴	۰/۵۴	۰/۵۵	۰/۴۰	۰/۴۳	۰/۵۰	۰/۵۸	۰/۹۷
	صحت سنجی	R	۰/۹۹۸	۰/۹۹۷	۰/۹۸۶	۰/۹۶۲	۰/۹۶۳	۰/۹۶۵	۰/۹۶۹	۰/۹۵۰	۰/۹۳۱	۰/۸۳۶
		RMSE	۰/۱۵	۰/۱۷	۰/۴۷	۰/۶۶	۰/۶۵	۰/۶۷	۰/۵۸	۰/۷۱	۰/۸۰	۱/۱۴
زنجان	آموزش	R	۰/۹۹۷	۰/۹۹۷	۰/۹۷۶	۰/۹۴۷	۰/۹۴۷	۰/۹۵۷	۰/۹۵۰	۰/۹۶۸	۰/۹۵۷	۰/۸۴۸
		RMSE	۰/۱۹	۰/۲۱	۰/۵۱	۰/۶۹	۰/۷۰	۰/۴۲	۰/۵۰	۰/۵۴	۰/۶۳	۱/۱۶
	صحت سنجی	R	۰/۹۹۸	۰/۹۹۷	۰/۹۹۰	۰/۹۷۸	۰/۹۷۸	۰/۹۶۸	۰/۹۷۰	۰/۹۶۴	۰/۹۴۷	۰/۸۴۱
		RMSE	۰/۱۸	۰/۲۰	۰/۵۸	۰/۴۸	۰/۴۸	۰/۵۱	۰/۵۶	۰/۵۹	۰/۶۹	۱/۱۸
سقز	R	۰/۹۹۷	۰/۹۹۷	۰/۹۸۳	۰/۹۵۰	۰/۹۵۰	۰/۹۷۵	۰/۹۷۶	۰/۹۷۱	۰/۹۵۷	۰/۸۲۲	
	RMSE	۰/۱۸	۰/۱۹	۰/۴۴	۰/۶۸	۰/۶۸	۰/۴۹	۰/۴۷	۰/۵۲	۰/۶۳	۱/۳۶	

۰/۹۹۶	۰/۹۹۵	۰/۹۸۹	۰/۹۸۲	۰/۹۸۲	۰/۹۶۳	۰/۹۶۹	۰/۹۶۳	۰/۹۵۶	۰/۸۱۶	R	صحت سنجی
۰/۱۷	۰/۲۰	۰/۷۹	۰/۷۰	۰/۷۱	۰/۷۵	۰/۶۲	۰/۶۷	۰/۷۱	۱/۴۳	RMSE	
۰/۹۹۶	۰/۹۹۷	۰/۹۷۸	۰/۹۵۶	۰/۹۵۳	۰/۹۷۵	۰/۹۸۲	۰/۹۷۵	۰/۹۵۶	۰/۸۰۱	R	آموزش
۰/۲۲	۰/۲۲	۰/۵۲	۰/۶۸	۰/۷۰	۰/۶۵	۰/۴۴	۰/۵۲	۰/۶۹	۱/۲۷	RMSE	
۰/۹۹۸	۰/۹۹۸	۰/۹۹۱	۰/۹۷۴	۰/۹۷۲	۰/۹۶۹	۰/۹۷۷	۰/۹۷۱	۰/۹۶۷	۰/۷۹۲	R	سنندج
۰/۱۸	۰/۱۶	۰/۴۱	۰/۵۴	۰/۵۵	۰/۶۲	۰/۵۰	۰/۵۶	۰/۵۹	۱/۲۷	RMSE	

※: واحد RMSE بر حسب میلی متر در روز می‌باشد.



شکل ۴- نمودارهای پراکندگی مقادیر مشاهده‌ای و تخمینی تبخیر-تعرق ماهانه گیاه مرجع در بهترین حالت مدل SVM برای ایستگاه‌های منتخب شمال غرب ایران

دست می‌یابند. به طوریکه مدل GP در ایستگاه‌های ارومیه و خوی با الگوی M9، در ایستگاه‌های زنجان، سقز و سنندج با الگوی M8 و در ایستگاه تبریز با الگوی M7 به بیشینه دقت رسیده است. همچنین مدل SVM در ایستگاه‌های ارومیه و سنندج با الگوی M9 و در سایر ایستگاه‌ها با الگوی M10 بهترین عملکرد را داشته است. در همه ایستگاه‌ها مدل SVM از دقت بالاتری نسبت به مدل GP برخوردار بود. نتایج حاصله در این قسمت با مطالعه (Tabari et al, 2013) که

با توجه به نتایج تحلیل‌های آماری و شاخص‌های RMSE و R، دو مدل SVM و GP از دقت بسیار خوب و قابل قبولی در پیش‌بینی تبخیر-تعرق ماهانه گیاه مرجع در ایستگاه‌های شمال غرب کشور برخوردار هستند. علی‌رغم اینکه عملکرد مدل SVM بهتر از مدل GP بود، اما در کل نتایج حاصله اختلاف زیادی را بین دو مدل نشان داد. هر دو مدل با ورودی‌های مختلفی به بالاترین دقت در مدل‌سازی

- 340: 63-77.
- Dogan, E. 2009. Reference Evapotranspiration Estimation using adaptive neuro-fuzzy inference system, *Journal of Irrigation and Drainage*. 58.5: 617-628.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*. 13.6: 87-129.
- Ghorbani, M.A., Kisi, O and Aalinezhad, M. 2010. A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods. *Applied Mathematical Modelling*. 34.12: 4050-4057.
- Guven, A. 2009. Linear genetic programming for time-series modeling of daily flow rate. *Journal of Earth System Sciences*. 118.2: 157-173.
- Hamel, L. 2009. Knowledge Discovery with Support Vector Machines. Hoboken, N.J. John Wiley.
- Jia Bing, C. 2004. Prediction of daily reference evapotranspiration using adaptive neurofuzzy inference system. *Trans of the Chinese society of Agricultural Engineering*. 20.4: 13-16.
- Karahan, H., Iplikci, S., Yasar, M and Gurarslan, G. 2014. River flow estimation from upstream flow records using support vector machines. *Journal of Applied Mathematics*. 1:1-7.
- Khu, S.T., Liang, S.Y., Babovic, V., Madsen, H and Muttil, N. 2001. Genetic programming and its application in real-time runoff forming. *Journal of American Water Resources Association*. 37 .2: 439-451.
- Koza, J.R. 1992. Genetic Programming: on the programming of computers by means of natural selection. Cambridge, MA: MIT Press.
- Pai, P.F and Hong, W.C. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process* 21:819-827.
- Tabari, H., Martinez, C., Ezani, A and Hosseinzadeh Talae, P. 2013. Applicability of support vector machines and adaptive neuro-fuzzy inference system for modeling potato crop evapotranspiration. *Irrigation Sciences*. 31.4: 575-588.
- Vapnik, V.N. 1998. *Statistical Learning Theory*. Wiley, New York.
- Yu, P.S., Chen, S.T., Chang I.F. 2005. Flood stage forecasting using support vector machines. *Geophysical Research Abstracts*. 7 :41-76.
- Zahiri, A and Azamathulla, H.M.D. 2014. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Computing and Applications*. 24.2: 413-420.
- نشان دهنده عملکرد بهتر SVM در پیش‌بینی تبخیر-تعرق می‌باشد، مطابقت دارد.
- اینکه کدامیک از مدل‌ها بهتر بوده و کدام الگو می‌بایست انتخاب شود، می‌تواند بسته به نظر کارشناس و دسترسی به داده‌های اقلیمی تعیین گردد. به عنوان مثال، برای ایستگاه سقز اگر بیشینه دقت مدنظر بوده و تمامی داده‌های اقلیمی در دسترس باشند، می‌توان مدل SVM با الگوی M10 را انتخاب کرد، ولی اگر داده ساعات آفتابی و سرعت باد به هر دلیلی کنار گذاشته شود، مدل GP با الگوی M8 بهترین انتخاب خواهد بود.
- در مواردی مانند پیش‌بینی جریان رودخانه، از حافظه سری جریان برای پیش‌بینی استفاده می‌گردد که دلیل آن نبود داده‌های بیش‌تر و یا بررسی اثر حافظه می‌باشد. در این تحقیق نیز سعی گردید اثر حافظه در پیش‌بینی تبخیر-تعرق ماهانه گیاه مرجع بررسی گردد. نتایج حاکی از آن بود که با افزایش میزان خطای قابل قبول می‌توان از حافظه نیز برای تخمین تبخیر-تعرق گیاه مرجع استفاده کرد که در این حالت مدل SVM عملکرد بهتری نسبت مدل GP دارد.

منابع

- زارع ابیانه، ح.، قاسمی، ع.، بیات ورکشی، م.، محمدی، ک.، سبزی - پرور، ع.ا. ۱۳۸۸. ارزیابی کارایی دو نرم افزار شبکه عصبی مصنوعی در پیش‌بینی تبخیر-تعرق گیاه مرجع. *مجله دانش آب و خاک*. ۲۰۱۹: ۲۱۲-۲۰۱.
- زارع ابیانه، ح.، بیات ورکشی، م.، معروفی، ص. ۱۳۹۰. محاسبه تبخیر و تعرق واقعی گیاه سیر به روش مدل سازی چندگانه تحت شرایط کاشت لایسیمتر. *مجله پژوهش‌های حفاظت آب و خاک*. ۲۰۱۸: ۱۵۸-۱۴۱.
- ستاری، م.ت.، نهرین، ف.، عظیمی، و. ۱۳۹۲. پیش‌بینی تبخیر - تعرق مرجع روزانه با استفاده از مدل شبکه عصبی مصنوعی و مدل درختی M5 (مطالعه موردی: ایستگاه بناب). *نشریه آبیاری و زهکشی ایران*. ۱۰.۷: ۱۱۳-۱۰۴.
- Adamowski, J and Prasher, S.O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *Journal of Water and Land Development*. 17.8: 89-97.
- Alvisi, S., Mascellani, G., Franchini, M and Bardossy, A. 2005. Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences*. 10.1: 1-17.
- Aytek, A and Kisi, O. 2008. A genetic programming approach to suspended sediment modeling. *Journal of Hydrology*. 351.3: 288-298.
- Chen, S.T., Yu, P.S. 2007. Real-time probabilistic forecasting of flood stages. *Journal of Hydrology*

Prediction of Monthly Grass Reference Crop Evapotranspiration in Northwest of Iran Using Genetic Programming and Support Vector Machine

F. Ahmadi¹, F. Radmaneh², R.I Mirabbasi Najaf Abadi³, S. Ayashm^{4*}

Received: Agu.30, 2014

Accepted: Mar.4, 2015

Abstract

In this study, for estimating grass reference crop evapotranspiration using support vector machine (SVM) and genetic programming (GP) techniques, monthly meteorological data of 6 synoptic stations in northwest of Iran during a 38 year period (1973-2010) were collected. At first, the monthly grass reference crop evapotranspiration values for selected station were calculated using the FAO-56 Penman-Monteith equation and considered as the output of SVM and GP models. In the next step, a regression equation was made based on effective climatic variables on evapotranspiration and different input patterns were defined for developing SVM and GP models. Since the relative humidity had the lowest effect on evapotranspiration, it was removed from the model's input variables. In order to consider the effect of memory of time series in the accuracy of evapotranspiration prediction, the evapotranspiration values with one, two, three, four and five months lag time were used in model's input patterns. Ten input patterns were made for every model. However, the obtained results showed the high accuracy of both SVM and GP models in predicting monthly grass reference crop evapotranspiration in northwest of Iran, but the SVM method had better performance than GP method. Also, the results showed that if there is no enough meteorological data, the memory of time series can be used for predicting monthly grass reference crop evapotranspiration.

Keywords: Genetic programming, Memory of time series, Grass reference crop evapotranspiration, Support vector machine

1 - Ph.D Candidate of Water Resources Engineering, Shahid Chamran University, Ahvaz

2 - Assistant Professor, Water Engineering Department, Shahid Chamran University, Ahvaz

3 - Assistant Professor, Water Engineering Department, Shahre Kord University, Shahre Kord

4 - Msc student, Water Engineering Department, Agriculture Faculty Urmia University, Urmia

(*: Corresponding Author, Email: S.ayashm@gmail.com)