

مقاله علمی-پژوهشی

ارزیابی مدل‌های هوشمند در تدقیق پیش‌بینی دبی اوج سیلاب

انیس حسنی^۱، فرشته مدرسی^۲، کیومرث ابراهیمی^{۳*}

تاریخ دریافت: ۱۳۹۹/۱۲/۰۶ تاریخ پذیرش: ۱۴۰۰/۰۲/۰۷

چکیده

با توجه به کمبود ایستگاه‌های اندازه‌گیری پارامترهای هیدرولوژیکی و هواشناسی در کشور، استفاده از مدل‌های داده‌مبناء در برآورد دبی حداکثر سیلاب ضروری است. در این تحقیق از مدل‌های برنامه‌ریزی بیان‌ژن و ماشین بردار پشتیبان برای پیش‌بینی دبی اوج در حوضه ماهنشان-انگوران استفاده شد و کارایی این دو مدل مورد ارزیابی قرار گرفته است. بدین منظور از داده‌های مشاهداتی ۳۶ سال (۱۳۹۰-۱۳۵۴) دبی حداکثر روزانه، بارش متناظر آن روز و دمای میانگین ماهانه در سه ایستگاه مهرآباد، ینگکند و قره‌گونی استفاده شد. دبی‌های اوج مشاهداتی و پیش‌بینی‌شده در هر دو مدل بر اساس معیارهای ارزیابی جذر میانگین مربعات خطا (RMSE)، ضریب تبیین (R^2) و معیار نش-ساتکلیف (NSE) مقایسه شد. میانگین مقادیر RMSE در مرحله‌ی صحت‌سنجی برای مدل GEP به ترتیب در ایستگاه‌های ینگکند، قره‌گونی و مهرآباد برابر با ۰/۰۴۹، ۰/۰۸۰ و ۰/۰۵۰ و در مرحله‌ی آموزش برابر با ۰/۰۴۲، ۰/۰۶۰ و ۰/۰۴۷ محاسبه شد. میانگین مقادیر R^2 در مرحله‌ی صحت‌سنجی در این ایستگاه‌ها به ترتیب برابر با ۰/۸۸، ۰/۸۶ و ۰/۸۷ و برای مرحله‌ی آموزش برابر با ۰/۷۷، ۰/۷۶ و ۰/۸۰ به دست آمد. همچنین مقادیر RMSE در مدل SVM برای مرحله‌ی صحت‌سنجی به ترتیب در این ایستگاه‌ها برابر با ۰/۰۴۲، ۰/۰۴۰ و ۰/۰۵۴ و در مرحله‌ی آموزش برابر با ۰/۰۵۳، ۰/۰۶۴ و ۰/۰۴۴ محاسبه شد. مقادیر R^2 در مرحله‌ی صحت‌سنجی برابر با ۰/۶۶، ۰/۸۵، ۰/۷۳ و برای مرحله‌ی آموزش برابر با ۰/۸۶، ۰/۸۸ و ۰/۹۱ و مقادیر NSE برای صحت‌سنجی برابر با ۰/۵۶، ۰/۷۵، ۰/۶۱ و برای مرحله آموزش برابر با ۰/۷۱، ۰/۷۷ و ۰/۸۰ حاصل شد. با بررسی معیارهای ارزیابی، مدل GEP عملکرد نسبتاً بهتری داشته است و این مدل برای پیش‌بینی سیل منطقه ماهنشان-انگوران مناسب‌تر است.

واژه‌های کلیدی: پیش‌بینی سیل، داده‌کاوی، مدیریت

مقدمه

همکاران، ۱۳۸۷). به‌طور کلی کنترل و کاهش خسارات سیل به دو شیوه‌ی سازه‌ای و غیرسازه‌ای قابل انجام است. امروزه مشخص شده است که روش‌های سازه‌ای به‌تنهایی برای کاهش خسارات سیل کافی نیستند. از جمله روش‌های غیرسازه‌ای که به کمک روش‌های سازه‌ای می‌تواند در کنترل و کاهش خسارات سیل مفید باشد سامانه‌های هشدار سیل است (Chang et al., 2018). بخش اصلی سیستم هشدار سیل، مدل پیش‌بینی سیل است. در پیش‌بینی و هشدار سیل هدف از شبیه‌سازی، شناسایی مبنای فیزیکی فرایند بارش-رواناب نبوده و فقط دقت و سرعت شبیه‌سازی اهمیت پیدا می‌کند. از طرفی با توجه به غیرخطی بودن فرایند بارش-رواناب و همچنین وابسته بودن آن به بسیاری از فرایندهای فیزیوگرافیک و هیدرولوژیک دارای همبستگی‌های داخلی حوضه، باید از مدل‌هایی استفاده شود که از قابلیت برقراری نگاشت غیرخطی بین بارش و رواناب برخوردار باشند؛ بنابراین استفاده از روش‌های هوش مصنوعی راهکاری مفید

سیل یک حادثه طبیعی است که سبب سرریز شدن آب به زمین-هایی که در شرایط عادی خشک هستند، می‌شود (Huang et al., 2019). تا به امروز هیچ کشوری صرف‌نظر از ثروت ملی و پیشرفت فناوری نتوانسته است نواحی سیل‌خیز خود را کاملاً از خطر محفوظ دارد (قاسمی و همکاران، ۱۳۹۲). سطح مناطق سیل‌خیز کشور حدود ۹۱ میلیون هکتار برآورد شده است؛ به عبارتی ۵۵ درصد از سطح کشور ما در تولید رواناب مستقیم نقش داشته که حدود ۴۲ درصد آن دارای شدت سیل‌خیزی متوسط تا خیلی زیاد است (شعبانلو و

۱- دانشجوی کارشناسی ارشد مهندسی منابع آب، دانشگاه تهران، کرج، ایران
۲- ستادیار گروه علوم و مهندسی آب، دانشگاه فردوسی مشهد، مشهد، ایران
۳- استاد گروه مهندسی آبیاری و آبادانی، دانشگاه تهران، کرج، ایران
* نویسنده مسئول: EbrahimiK@ut.ac.ir

تلقی می‌شود.

برنامه‌ریزی بیان ژن متعلق به خانواده‌ی الگوریتم‌های تکاملی است، مانند تمام الگوریتم‌های تکاملی دیگر در برنامه‌ریزی بیان ژن هم از جمعیت افراد استفاده می‌شود و با تولید یک جمعیت تصادفی اولیه شروع به کار می‌کند و سپس با توجه به عملکردهای انتخاب نسل-های جدیدی را تولید می‌کند. با تکرار این فرآیند برای تعداد مشخصی از نسل‌ها تکامل اتفاق می‌افتد که این به معنای یافتن بهترین پاسخ است (Ferreira., 2001).

ماشین بردار پشتیبان یکی از روش‌های یادگیری تحت نظارت^۱ است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. این روش که توسط وپنیک (Vapnik., 1998) بر پایه یادگیری آماری بنا شده است، روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این‌رو روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود (Pai et al., 2007).

به علت کارایی بالا، این دو روش در مطالعات پیشین برای مدل‌سازی‌های هیدرولوژیکی مورداستفاده قرار گرفته‌اند؛ که از آن جمله می‌توان به تحقیقات زیر اشاره نمود:

در مطالعه‌ای برای مدل‌سازی بارش-رواناب در حوضه رودخانه جونباتا در پنسیلوانیای آمریکا از دو روش شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن استفاده شد. مقایسه نتایج نشان داد که روش برنامه‌ریزی بیان ژن در مقایسه با شبکه عصبی مصنوعی عملکرد بهتری داشته است (Aytek and Kisi., 2008). همچنین، در مطالعه‌ای عملکرد سه روش برنامه‌ریزی ژنتیک، شبکه عصبی و نرو فازی در روند یابی سیلاب رودخانه قزل ایرماق ترکیه مورد ارزیابی قرار گرفته. نتایج مطالعه نشان داد مدل برنامه‌ریزی ژنتیک با دقت بیشتری هیدروگراف خروجی را شبیه‌سازی می‌کند (Khatibi et al., 2010). در تحقیقی دیگر مدل ماشین بردار پشتیبان را جهت پیش-بینی دبی روزانه رودخانه قره‌سو مورد ارزیابی قرار دادند نتایج کار حاکمی از دقت بالا و کارایی مناسب این مدل بود (محرم‌پور و همکاران، ۱۳۹۰). همچنین برای پیش‌بینی دبی اوج در حوضه آبخیز طالقان از دو مدل شبکه عصبی و رگرسیون چند متغیره غیرخطی استفاده شده است. نتایج این تحقیق نشان داد که عملکرد شبکه عصبی بهتر از رگرسیون چند متغیره غیرخطی است (خسروی و همکاران، ۱۳۹۱). در مقاله‌ای به مقایسه عملکرد دو مدل برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان برای پیش‌بینی جریان روزانه رودخانه باراندوز چای پرداخته شد و فقط از داده‌های دبی جریان برای پیش‌بینی جریان رودخانه استفاده شد. نتایج حاکمی از بهتر بودن عملکرد مدل برنامه‌ریزی بیان ژن داشت (احمدی و همکاران، ۱۳۹۳). در مطالعه‌ای کارایی چهار مدل شبکه عصبی مصنوعی (ANN)، شبکه عصبی رگرسیون

تعمیم‌یافته (GRNN)، رگرسیون بردار پشتیبان (SVR)، و K-نزدیک‌ترین همسایگی را برای پیش‌بینی جریان ماهانه بر اساس سیگنال‌های اقلیمی در شرایط خطی و غیرخطی مورد مقایسه و ارزیابی قرار دادند. نتایج این تحقیق نشان داد که در شرایط غیرخطی کارایی مدل SVR بر سایر مدل‌ها برتری دارد در حالی که در شرایط خطی، مدل ANN دارای بهترین کارایی است (Modaresi et al., 2018).

همان‌طور که در تحقیقات صورت گرفته‌ی بالا مشاهده می‌شود، مطالعات در قالب‌های کلی ۱: بررسی سیل و سیل‌خیزی ۲: بررسی مدل‌های هوشمند به‌خصوص دو مدل برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان است. با توجه به سیل‌خیز بودن نواحی بسیاری در کشور (در حدود ۵۵ درصد مناطق)، و با توجه به سیل‌خیز بودن منطقه مورد مطالعاتی در این تحقیق هدف از انجام مقاله موردنظر، استفاده از مدل برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان که دو مدل دقیق‌تری برای پیش‌بینی‌ها هستند، جهت پیش‌بینی دبی اوج حوضه آبریز موردنظر است.

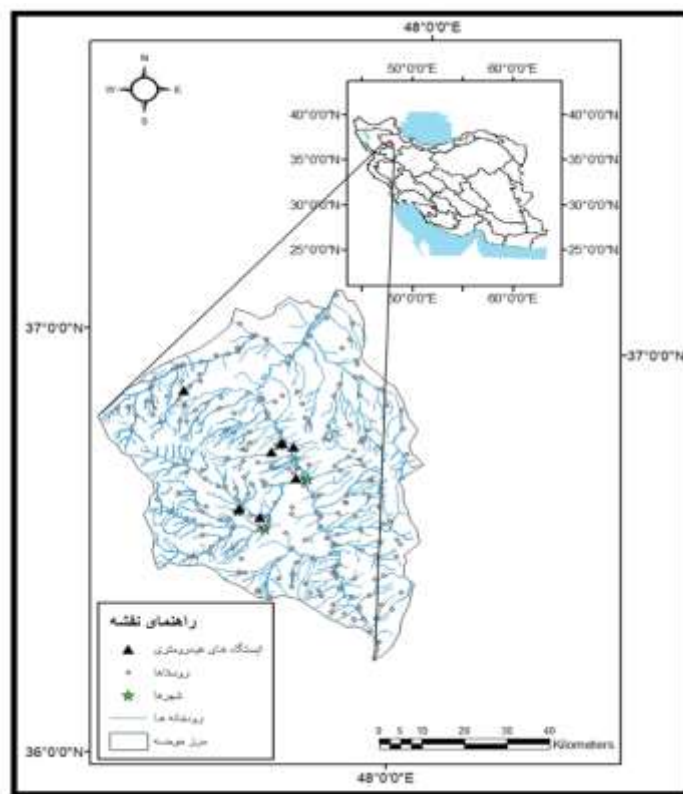
مواد و روش‌ها

موقعیت منطقه مورد مطالعه

جهت انجام تحقیق حاضر حوضه آبریز ماهنشان-انگوران به‌عنوان حوضه مطالعاتی انتخاب شد. این حوضه در غرب استان زنجان قرار دارد و یکی از زیر حوضه‌های سفیدرود و در محدوده‌ی حوضه دریای خزر واقع است. در این حوضه به دلیل وقوع بارش‌های مناسب مشکلی از نظر کمبود منابع آب‌های سطحی و زیرزمینی وجود ندارد. متوسط بارندگی بلندمدت در این منطقه معادل ۲۵۱ میلی‌متر و مهم‌ترین رودخانه‌ی این حوضه قزل‌اوزن است. این حوضه دارای مساحتی در حدود ۴۸۳۸/۷۴ کیلومترمربع و محیط آن ۳۳۱/۸۷ کیلومتر است. حداکثر و حداقل ارتفاع حوضه به ترتیب برابر با ۳۲۸۶ و ۱۱۰۸ متر از سطح دریا است. شکل ۱ موقعیت این حوضه در نقشه‌ی کشوری را نشان می‌دهد.

روش تحقیق

ابتدا بر پایه وجود اطلاعات سه ایستگاه هیدرومتری در منطقه موردنظر، محدوده مطالعه به سه واحد هیدرولوژیک تقسیم شد. این واحدها ایستگاه‌های هواشناسی متناظر نیز دارند. آمار حداکثر دبی روزانه و بارش متناظر، همچنین دمای میانگین ماهانه در واحدهای هیدرولوژیک ینگی‌کند، مهرآباد و قره‌گونی در طول سال‌های ۱۳۵۴ تا ۱۳۹۰ استخراج و ۷۰ درصد داده‌ها در مرحله آموزش و ۳۰ درصد آن‌ها در مرحله صحت‌سنجی مدل‌ها در نظر گرفته شد.



شکل ۱- موقعیت حوضه آبریز ماهنشان-انگوران در نقشه‌ی کشوری

شد (Ferreira, 2001). GEP روشی مناسب برای مدل‌سازی فرایندهای غیرخطی و پیچیده است.

اولین مرحله در روند یابی سیلاب به کمک این نرم‌افزار، تولید جمعیت اولیه از راه‌حل‌هاست که به کمک یک سری توابع و پایانه‌ها صورت می‌گیرد. سپس کروموزوم‌ها به صورت ساختار درختی^۲ نشان داده می‌شوند؛ در مرحله بعد باید کارایی یا سازگاری هر عضو جمعیت کروموزوم‌ها را به کمک تابع برازش ارزیابی کرد. به طور کلی فرآیند گام‌به‌گام مدل برنامه‌ریزی بیان ژن به صورت الگوریتم در شکل ۲ ارائه شده است:

ماشین بردار پشتیبان (SVM)

مدل ماشین بردار پشتیبان، در سال ۱۹۹۲ توسط یک ریاضیدان روس به نام وپنیک بر پایه نظریه یادگیری آماری بنا گردیده است (Vapnik, 1998). ماشین بردار پشتیبان یک روش داده‌کاوی است که بر پایه نظریه یادگیری آماری بنا شده است. این روش بر پایه بهینه‌سازی مقید است، ماشین بردار پشتیبان دسته‌بندی کننده‌ای است که جزء شاخه‌ی روش‌های کرنلی در یادگیری ماشین محسوب می‌شود.

روش‌های داده‌کاوی

برنامه‌ریزی بیان ژن (GEP)

برنامه‌ریزی ژنتیک تعمیم‌یافته الگوریتم ژنتیک است که برای اولین بار بر اساس تئوری داروین ارائه شد. برنامه‌ریزی ژنتیک یک فن برنامه‌ریزی خودکار است که راه حل مسئله را با استفاده از برنامه کامپیوتری ارائه می‌کند. تفاوت اساسی موجود بین GP و GA به طبیعت هر یک از اجزای درون برنامه برمی‌گردد، به نحوی که اجزا در GA، ردیف‌های خطی با طول ثابت می‌باشند (کروموزوم‌ها). ولی در GP، شاخه‌های مجزا هستند. همچنین در GP بر ساختار درختی مجموعه‌ها تأکید می‌شود ولی GA، بر اساس سیستم ارقام دودویی عمل می‌نماید.

در این مطالعه از برنامه Genexprotools نسخه ۵/۰ برای توسعه و اجرای مدل‌های مبتنی بر GP استفاده شد.^۱ GEP ویرایش جدیدی از برنامه‌ریزی ژنتیک است که به استنتاج برنامه‌های رایانه‌ای با اندازه‌ها و شکل‌های مختلف می‌پردازد. این روش توسط فریرا ارائه

2 Expression tree structure

1 Gene Expression Programming

دسته‌بندی باشند و سپس با یافتن خطوط پشتیبان، سعی دریافتن معادله خطی دارد که بیشترین فاصله را بین دودسته ایجاد می‌کند. در این مطالعه از نرم‌افزار تاناگرا نسخه ۱/۴ جهت پیش‌بینی توسط مدل ماشین بردار پشتیبان استفاده شد.

این نوع سیستم یادگیری هم برای دسته‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها به کار می‌رود، به طوری که کمترین خطا در دسته‌بندی داده‌ها و یا تابع برازش رخ دهد. ماشین بردار پشتیبان داده‌ها را با توجه به دسته‌های از پیش تعیین شده آن‌ها، به یک فضای جدید می‌برد به گونه‌ای که داده‌ها به صورت خطی قابل تفکیک و



شکل ۲- فرآیند گام به گام مدل برنامه‌ریزی بیان ژن (Dey and Das., 2016)

در هر دو روش، در تمامی الگوها پس از نرمال‌سازی داده‌ها، ۷۰ درصد داده‌ها جهت آموزش و ۳۰ درصد برای تست کردن و صحت-سنجی به طور تصادفی در نظر گرفته و وارد مدل‌ها شدند.

معیارهای ارزیابی

در این پژوهش معیار سنجش عملکرد مدل‌ها، جذر میانگین مربعات خطا (RMSE)، ضریب تبیین (r^2) و معیار نش-ساتکلیف (NSE) در نظر گرفته شد؛ که هر مدل با داشتن RMSE کمتر، r^2 و NSE بیشتر به عنوان مدل مناسب‌تر انتخاب شده است (Lu et al., 2018):

برای روند یابی سیل حوضه آبریز ماهنشان-انگوران متغیرهای میانگین حداکثر دبی‌های روزانه در هر ماه (Q_t)، حداکثر دبی روزانه ماه قبل (Q_{t-1})، حداکثر دبی روزانه دو ماه قبل (Q_{t-2})، بارش متناظر (P_t)، بارش یک ماه قبل (P_{t-1}) و دمای میانگین ماهانه (T_t)، در پنج ترکیب متفاوت به مدل‌ها معرفی شد تا بهترین حالت ممکن جهت پیش‌بینی هیدروگراف خروجی انتخاب شود (جدول ۱).

جدول ۱- ترکیب ورودی‌های مختلف برای منطقه مورد مطالعه

نام مدل	متغیرهای ورودی
مدل ۱	Q_{t-1}
مدل ۲	Q_{t-1}, Q_{t-2}
مدل ۳	Q_{t-1}, P_t
مدل ۴	Q_{t-1}, P_t, P_{t-1}
مدل ۵	$Q_{t-1}, P_t, P_{t-1}, T_t$

دبی اوج مربوط به آن رسم شد (جدول ۲). با توجه به شکل‌های ۳ و ۴ بهترین حالت برای ایستگاه‌های ینگگی کند و قره‌گونی در مدل برنامه‌ریزی بیان ژن، الگوی چهارم بود؛ از این رو ساختار رابطه‌ای که متغیرهای ورودی و خروجی را به هم مرتبط می‌سازد به صورت رابطه ۴ تعریف شده است:

$$Q_t = f(Q_{t-1}, P_t, P_{t-1}) \quad (۴)$$

یعنی زمانی که دبی اوج با یک ماه تأخیر و همچنین بارش متناظر و بارش با یک ماه تأخیر در نظر گرفته شد که در این حالت RMSE و r^2 و NSE مربوط به مرحله آموزش برای ایستگاه ینگگی کند به ترتیب برابر با ۰/۰۴۴، ۰/۹۳، و ۰/۸۲ و برای ایستگاه قره‌گونی مهرآباد از بین الگوهای ورودی پنجم بهترین برازش را داشت که ترکیبی از دبی اوج با یک ماه تأخیر، بارش متناظر و بارش با یک ماه تأخیر و دمای میانگین ماهانه بود که ساختار رابطه آن در رابطه ۵ نمایش داده شده است و میزان RMSE و r^2 و NSE آن برای آموزش به ترتیب برابر با ۰/۰۲۷، ۰/۹۸ و ۰/۸۸ برآورد شد.

$$Q_t = f(Q_{t-1}, P_t, P_{t-1}, T) \quad (۵)$$

بهترین ترکیب ورودی برای هر سه ایستگاه در مدل ماشین بردار پشتیبان نیز نظیر مدل برنامه‌ریزی بیان ژن بود به طوری که در دو ایستگاه ینگگی کند و قره‌گونی ترکیب چهارم دارای بهترین عملکرد است؛ و در ایستگاه مهرآباد الگوی پنجم نتیجه بهتری را ارائه داده است (جدول ۳).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{obs,i} - Q_{pre,i})^2}{n}} \quad (۱)$$

$$R^2 = \frac{\left[\sum_{i=1}^n (Q_{pre,i} - Q_{mean\ pre,i}) (Q_{obs,i} - Q_{mean\ obs,i}) \right]^2}{\sum_{i=1}^n (Q_{pre,i} - Q_{mean\ pre,i})^2 \sum_{i=1}^n (Q_{obs,i} - Q_{mean\ obs,i})^2} \quad (۲)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{pre,i})^2}{\sum_{i=1}^n (Q_{obs,i} - Q_{mean\ obs,i})^2} \quad (۳)$$

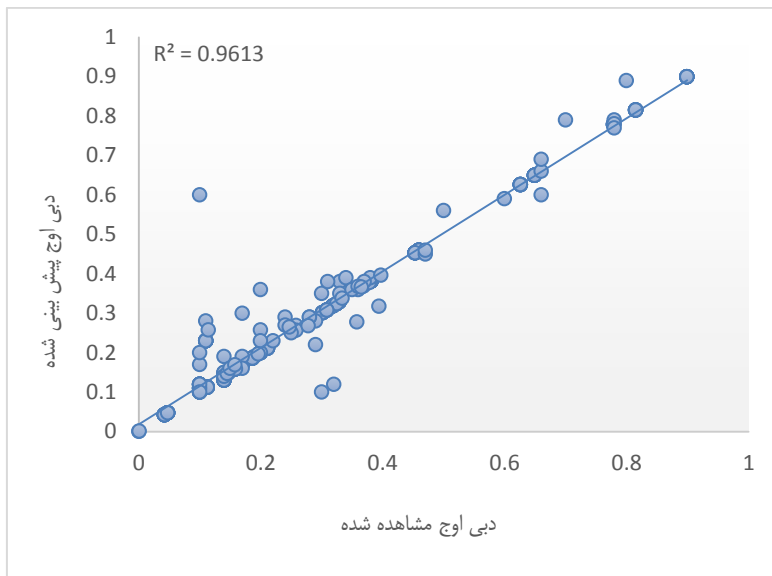
در روابط بالا، $Q_{obs,i}$ = دبی مشاهده شده؛ $Q_{pre,i}$ = دبی پیش‌بینی شده؛ $Q_{mean\ pre,i}$ = میانگین دبی پیش‌بینی شده؛ $Q_{mean\ obs,i}$ = میانگین دبی مشاهده شده؛ i = گام‌های زمانی و n = تعداد کل گام‌های زمانی است.

نتایج

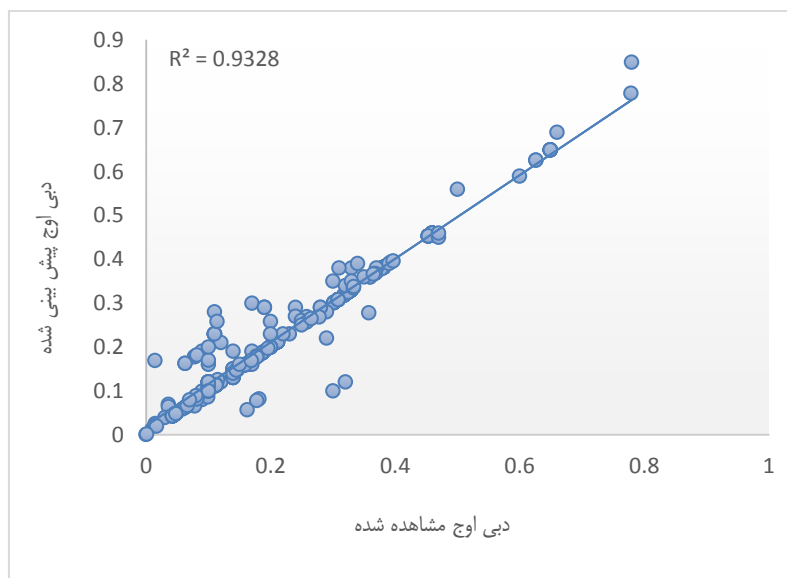
در هر سه ایستگاه الگوهای مختلف ورودی به مدل برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان اضافه شدند، پس از نرمال‌سازی داده‌های استخراج شده و آموزش و صحت‌سنجی با استفاده از معیارهای ارزیابی بهترین مدل با توجه به معیارهای ارزیابی در هر دو مرحله‌ی آموزش و صحت‌سنجی، در هر ایستگاه تعیین و روند تغییرات

جدول ۲- مقایسه دقت الگوهای مختلف در مدل برنامه‌ریزی ژنتیک

ایستگاه	الگوی ورودی	آموزش			صحت‌سنجی		
		NSE	RMSE	R^2	NSE	RMSE	R^2
ینگگی کند	مدل ۱	-۰/۹۱	-۰/۳۹	-۰/۸۱	-۰/۸۶	-۰/۴۷	-۰/۷۲
	مدل ۲	-۰/۸۱	-۰/۴۱	-۰/۷۲	-۰/۸۹	-۰/۰۷	-۰/۷۵
	مدل ۳	-۰/۹۰	-۰/۴۴	-۰/۶۹	-۰/۸۵	-۰/۴۵	-۰/۷۰
	مدل ۴	۰/۹۳	-۰/۴۴	۰/۸۲	۰/۹۶	۰/۳۶	۰/۸۵
	مدل ۵	-۰/۹۱	-۰/۴۲	-۰/۸۰	-۰/۸۴	-۰/۵۱	-۰/۷۳
قره‌گونی	مدل ۱	-۰/۸۸	-۰/۴۶	-۰/۷۶	-۰/۸۲	-۰/۷۷	-۰/۷۳
	مدل ۲	-۰/۸۹	-۰/۶۵	-۰/۷۶	-۰/۸۰	-۰/۸۴	-۰/۶۹
	مدل ۳	-۰/۸۷	-۰/۶۲	-۰/۷۵	-۰/۸۴	-۰/۶۵	-۰/۷۵
	مدل ۴	۰/۹۴	۰/۴۹	۰/۸۳	۰/۹۳	۰/۸۵	۰/۸۲
	مدل ۵	-۰/۸۵	-۰/۷۶	-۰/۷۰	-۰/۹۰	-۰/۰۹	-۰/۷۸
مهرآباد	مدل ۱	-۰/۹۵	-۰/۵۷	-۰/۸۲	-۰/۹۱	-۰/۷۰	-۰/۷۹
	مدل ۲	-۰/۸۱	-۰/۶۵	-۰/۶۹	-۰/۶۲	-۰/۷۹	-۰/۵۱
	مدل ۳	-۰/۹۵	-۰/۳۸	-۰/۸۲	-۰/۹۲	-۰/۵۲	-۰/۸۰
	مدل ۴	-۰/۹۱	-۰/۴۸	-۰/۸۰	-۰/۹۳	-۰/۳۹	-۰/۸۱
	مدل ۵	۰/۹۸	۰/۲۷	۰/۸۸	۰/۹۵	۰/۴۷	۰/۸۴



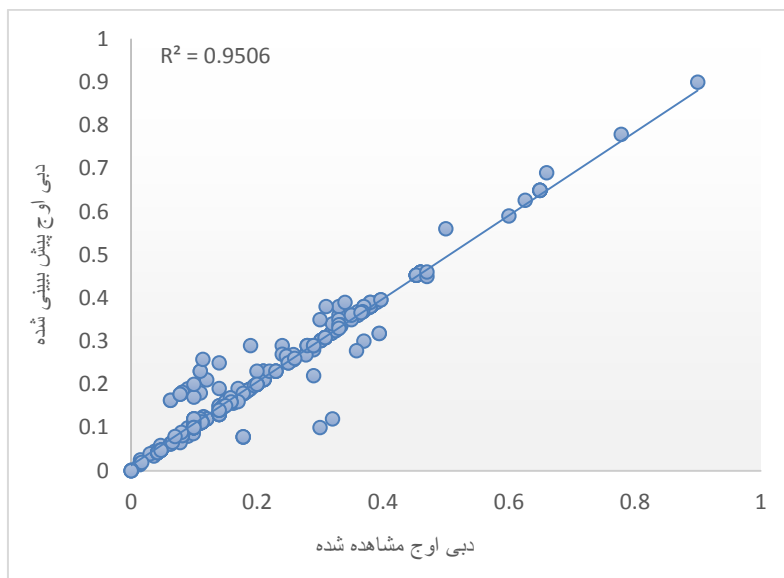
شکل ۳- همبستگی داده‌های مشاهداتی و محاسباتی در مرحله صحت‌سنجی ایستگاه ینگ‌کند مدل چهارم



شکل ۴- همبستگی داده‌های مشاهداتی و محاسباتی در مرحله صحت‌سنجی ایستگاه قره‌گونی مدل چهارم

باشد آن مدل نتایج بهتری ارائه می‌دهد. با توجه به شکل‌های ۶ الی ۱۱ عملکرد مدل برنامه‌ریزی بیان ژن در این مطالعه نسبت به ماشین بردار پشتیبان کمی بهتر است. پس می‌توان نتیجه گرفت مدل برنامه‌ریزی بیان ژن مدل دقیق‌تری برای انجام محاسبات پیش‌بینی سیل در این منطقه است.

در ادامه بهترین ساختار از هر مدل در هر یک از ایستگاه‌ها مورد مقایسه قرار گرفت که نتایج آن در جدول (۴) ارائه شده است. نتایج نشان داد که دقت و عملکرد هر دو مدل برای پیش‌بینی سیلاب قابل قبول است؛ اما از آنجایی که هرچه ضریب تبیین و معیار نش ساتکلیف بیشتر و نزدیک به یک و جذر میانگین مربعات خطا کمتر



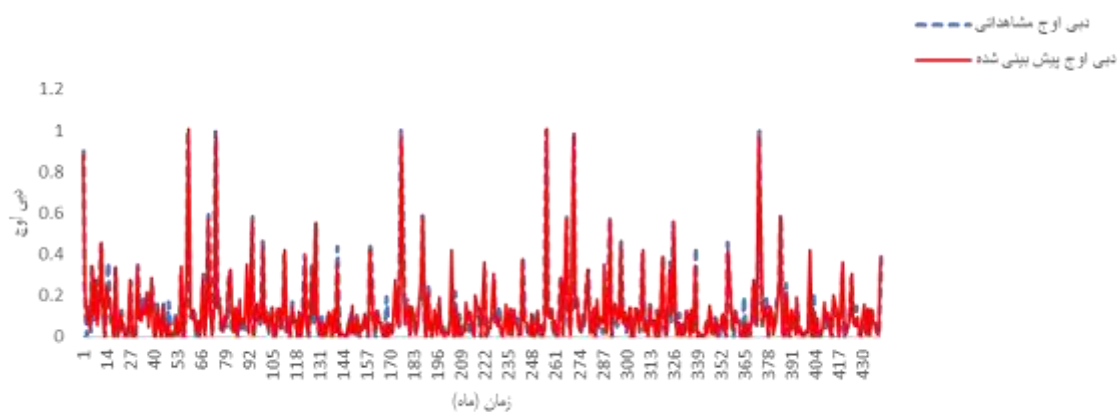
شکل ۵ - همبستگی داده‌های مشاهداتی و محاسباتی در مرحله صحت‌سنجی ایستگاه مهرآباد مدل پنجم

جدول ۳- مقایسه دقت الگوهای مختلف در مدل مائین بردار پشتیبان

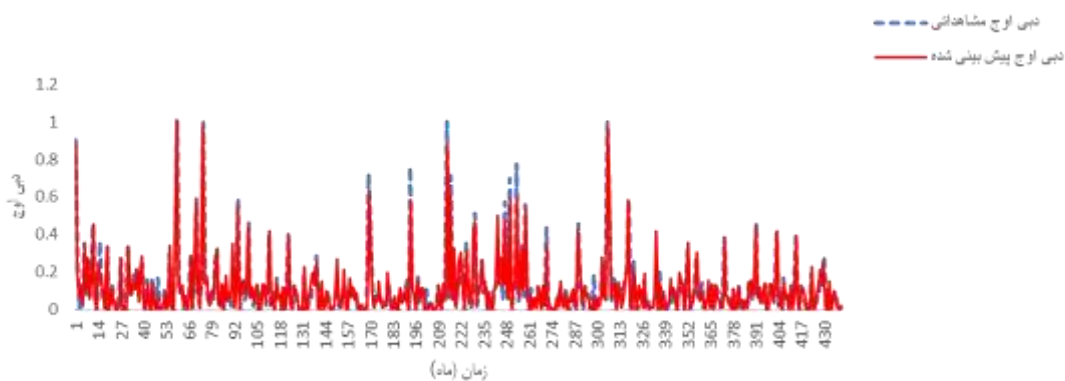
ایستگاه	الگوی ورودی	آموزش			صحت‌سنجی		
		NSE	RMSE	R ²	NSE	RMSE	R ²
ینگى کند	مدل ۱	-/۰۸۷	-/۰۵۰	-/۰۷۵	-/۰۶۳	-/۰۷۶	-/۰۵۱
	مدل ۲	-/۰۸۴	-/۰۵۹	-/۰۷۲	-/۰۶۶	-/۰۲۲	-/۰۵۴
	مدل ۳	-/۰۸۶	-/۰۵۳	-/۰۶۳	-/۰۵۸	-/۰۳۱	-/۰۴۷
	مدل ۴	+/۰۸۸	-/۰۵۰	+/۰۷۳	+/۰۷۸	+/۰۴۱	+/۰۶۵
	مدل ۵	-/۰۸۵	-/۰۵۳	-/۰۷۱	-/۰۶۴	-/۰۴۱	-/۰۶۲
قره‌گونی	مدل ۱	-/۰۸۶	-/۰۶۸	-/۰۷۵	-/۰۷۶	-/۰۷۲	-/۰۶۳
	مدل ۲	-/۰۸۹	-/۰۶۰	-/۰۷۸	-/۰۷۵	-/۰۳۰	-/۰۶۴
	مدل ۳	-/۰۸۷	-/۰۶۶	-/۰۷۶	-/۰۹۲	-/۰۲۹	-/۰۸۱
	مدل ۴	+/۰۹۰	+/۰۶۲	+/۰۷۹	+/۰۹۲	+/۰۳۰	+/۰۸۰
	مدل ۵	-/۰۸۸	-/۰۶۳	-/۰۷۶	-/۰۹۱	-/۰۴۲	-/۰۸۷
مهرآباد	مدل ۱	-/۰۹۳	-/۰۴۶	-/۰۸۰	-/۰۷۲	-/۰۶۱	-/۰۶۱
	مدل ۲	-/۰۸۷	-/۰۴۶	-/۰۷۵	-/۰۶۶	-/۰۸۰	-/۰۵۳
	مدل ۳	-/۰۹۰	-/۰۴۷	-/۰۷۸	-/۰۶۵	-/۰۴۷	-/۰۵۲
	مدل ۴	-/۰۹۳	-/۰۴۲	-/۰۸۱	-/۰۷۴	-/۰۵۳	-/۰۶۱
	مدل ۵	+/۰۹۴	+/۰۴۱	+/۰۸۲	+/۰۸۹	+/۰۳۳	+/۰۷۸

جدول ۴- مقایسه نتایج حاصل از بهترین ترکیب داده‌ها برای هر ایستگاه

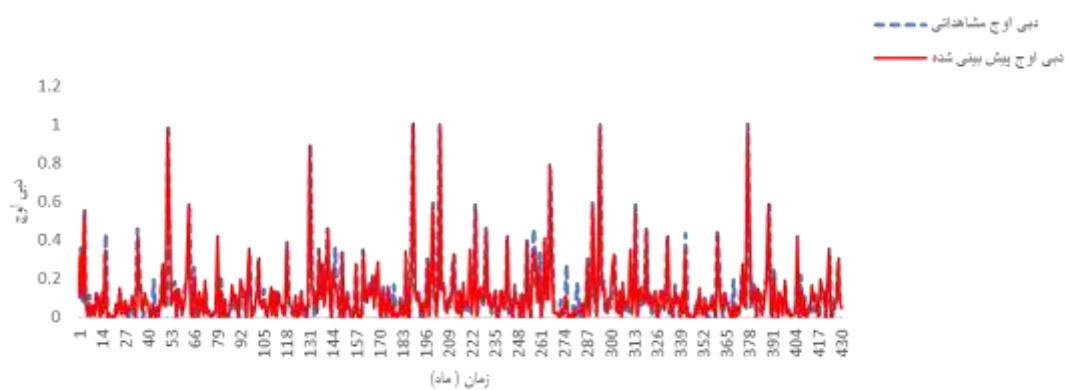
مدل	ایستگاه	جذر میانگین مربعات خطا (RMSE)		ضریب تبیین (R ²)		نش-ساتکیف (NSE)	
		آموزش	تست	آموزش	تست	آموزش	تست
GEP	قره‌گونی	-/۰۹۳	-/۰۹۶	-/۰۴۴	-/۰۳۶	-/۰۸۲	-/۰۸۵
	مهرآباد	-/۰۹۸	-/۰۹۵	-/۰۲۷	-/۰۴۷	-/۰۸۸	-/۰۸۴
	ینگى کند	-/۰۸۸	-/۰۷۸	-/۰۵۰	-/۰۴۱	-/۰۷۳	-/۰۶۵
SVM	قره‌گونی	-/۰۹۰	-/۰۹۲	-/۰۶۲	-/۰۳۰	-/۰۷۹	-/۰۸۰
	مهرآباد	-/۰۹۴	-/۰۸۹	-/۰۴۱	-/۰۳۳	-/۰۸۲	-/۰۷۸



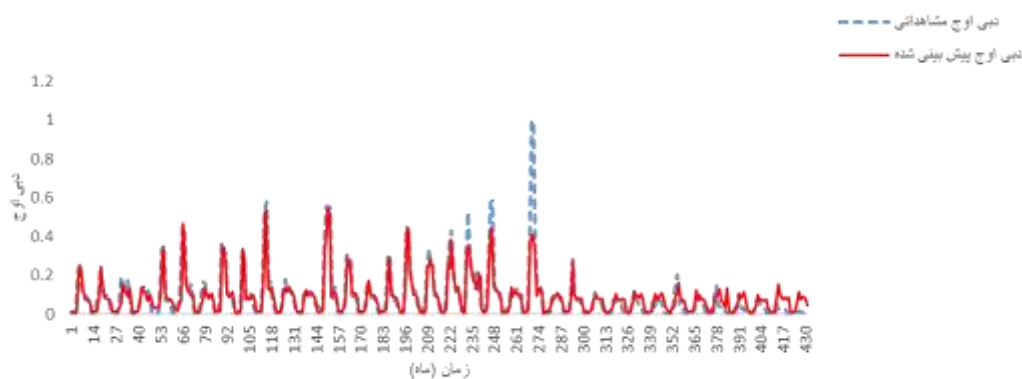
شکل ۶- روند تغییرات دبی اوج مشاهده‌ای و پیش‌بینی شده با مدل GEP در ایستگاه ینگ‌کند



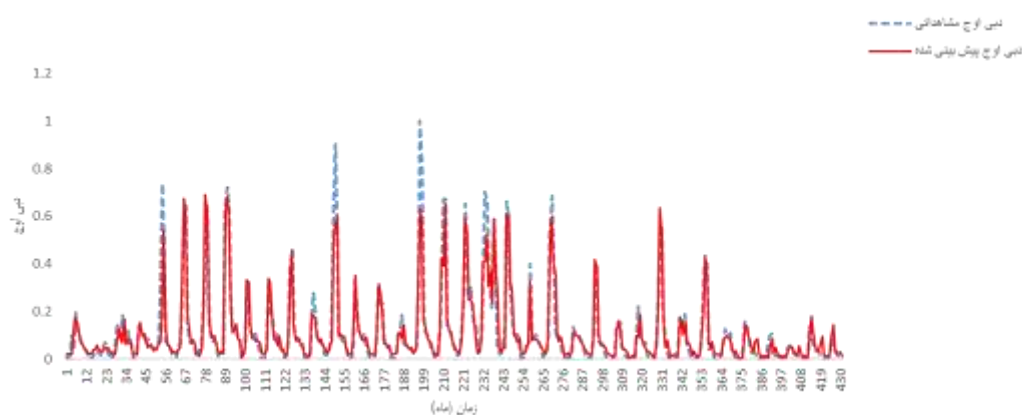
شکل ۷- روند تغییرات دبی اوج مشاهده‌ای و پیش‌بینی شده با مدل GEP در ایستگاه قره‌گونی



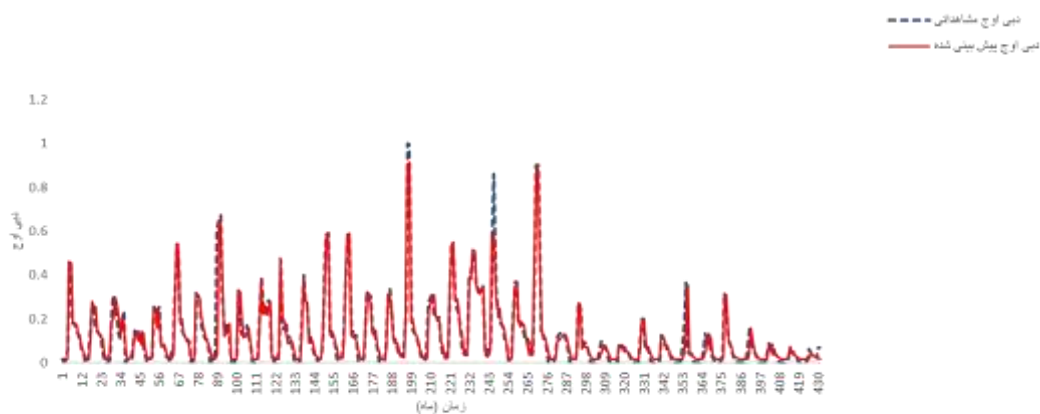
شکل ۸- روند تغییرات دبی اوج مشاهده‌ای و پیش‌بینی شده با مدل GEP در ایستگاه مهرآباد



شکل ۹- روند تغییرات دبی اوج مشاهداتی و پیش‌بینی شده با مدل SVM در ایستگاه ینگکند



شکل ۱۰- روند تغییرات دبی اوج مشاهداتی و پیش‌بینی شده با مدل SVM در ایستگاه قره‌گونی



شکل ۱۱- روند تغییرات دبی اوج مشاهداتی و پیش‌بینی شده با مدل SVM در ایستگاه مهرآباد

بحث و نتیجه‌گیری

در این تحقیق دقت روند یابی سیلاب حوضه‌ی ماهنشان- انگوران، دبی‌های اوج با استفاده از مدل‌های برنامه‌ریزی بیان ژن و

ماشین بردار پشتیبان در سه ایستگاه ینگکند، قره‌قونی و مهرآباد ارزیابی شد. ورودی‌های مختلف شامل پارامترهای دبی، بارندگی و دما با ۵ ترکیب مختلف در نظر گرفته شد. نتایج نشان داد که در هر دو مدل و در هر سه ایستگاه، نوع و اهمیت پارامترهای ورودی برای

قاسمی، ع.، حاجی بابایی، ا. و شمسایی، ا. ۱۳۹۲. مدیریت سیلاب، سامانه پیش‌بینی و هشدار سیل. کنفرانس ملی مدیریت سیلاب.

محرم پور، م.، محرابی، ع. و کاتوزی، م. ۱۳۹۰. به‌کارگیری ماشین بردار پشتیبان SVM برای پیش‌بینی دبی روزانه. چهارمین کنفرانس مدیریت منابع آب ایران.

Aytek, A. and Kisi, O. 2008. A genetic programming approach to suspended sediment modeling. *Journal of Hydrology*. 351: 288-298

rmatics Integration Platform for Regional Flood Inundation Warning Systems. *Journal of Water*. 11(1).

Dey P. and Das A. 2016. A utilization of GEP (gene expression programming) metamodel and PSO (particle swarm optimization) tool to predict and optimize the forced convection around a cylinder. *Journal of energy*. 19: 447-458.

Ferreira, C. 2001. Gene expression program-ming a new adaptive algorithm for solving problems. *Complex Systems* 13(2): 87-129.

Huang, W., Cao, Z., Huang, M., Duan, W., Ni, Y. and Yang, w. 2019. A New Flash Flood Warning Scheme Based on Hydrodynamic Modelling. *Journal of Water*. 11(6).

Khatibi, R., Ghorbani, M.A., Hasanpourkashani, M., and Kisi, O. 2010. Comparison of three artificial intelligence techniques for discharge routing". *Journal of hydrology*. 403(3-4): 201-212.

Lu X., Ju Y., Wu L., Fan J., Zhang F., Li Z. 2018. Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models. *Journal of Hydrology*. 556: 668-684.

Modaresi, F., Araghinejad, S. and Ebrahimi, K. 2018. A comparative assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for monthly streamflow forecasting in linear and nonlinear conditions. *Water Resources Management*. 32(1): 243-258.

Pai, P.F. and Hong W.C. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process*. 21:819-827.

Vapnik V.N. 1998. *Statistical Learning Theory*. Wiley, New York.

پیش‌بینی سیل در مقایسه با یکدیگر متفاوت است. در مدل GEP و SVM در دو ایستگاه ینگگی کند و قره‌گونی ترکیب چهارم که در آن دبی اوج با یک ماه تأخیر و همچنین بارش متناظر و بارش با یک ماه تأخیر در نظر گرفته شد و در ایستگاه مهرآباد زمانی که دبی اوج با یک ماه تأخیر، بارش متناظر و بارش با یک ماه تأخیر و دمای میانگین ماهانه بهترین برازش را نتیجه داد. نتایج بررسی فوق حاکی از دقت بالای مدل برنامه‌ریزی بیان ژن نسبت به مدل ماشین بردار پشتیبان در مدل‌سازی بارش-رواناب و پیش‌بینی سیل است که با تحقیقات مختلف از جمله (احمدی و همکاران، ۱۹۹۳)، (Aytek & Kisi., 2008) و (Khatibi et al., 2010) هم‌خوانی دارد، بنابراین می‌توان با در نظر گرفتن پیش‌بینی‌ها، تلاش برای پیش‌هشدار سیل و جلوگیری از خسارت‌های احتمالی این پدیده به عمل آورد. به همین جهت پیشنهاد می‌شود که با استفاده از داده‌های دبی‌های متوسط حداکثر روزانه ماه قبل، دو ماه قبل، بارش متناظر و بارش با یک ماه تأخیر و همچنین دمای میانگین ماهانه در دیگر حوضه‌های آبریز کشور به بررسی روند دبی اوج پرداخته شود تا از این طریق از خسارت‌ها و صدمات احتمالی جلوگیری کرد.

سیاسگزارى

بدین‌وسیله از دانشگاه تهران، شرکت مدیریت منابع آب، مؤسسه تحقیقات آب وزارت نیرو و سازمان هواشناسی کشور به دلیل تأمین امکانات و داده‌های لازم جهت انجام این تحقیق و تهیه مقالات مربوطه تشکر و قدردانی می‌شود.

منابع

احمدی، ف.، رادمش، ف. و میر عباسی نجف‌آبادی، ر. ۱۳۹۳. مقایسه روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوز چای). *مجله آب‌و‌خاک*. ۶ (۲۸): ۱۱۶۲-۱۱۷۱.

خسروی، م.، سلاجقه، ع.، مهدوی، م. و محسنی ساروی، م. ۱۳۹۱. پیش‌بینی سیل با استفاده از شبکه عصبی مصنوعی و رگرسیون چند متغیره غیرخطی. *نشریه مرتع و آبخیزداری*، مجله منابع طبیعی ایران، ۶۵ (۳): ۳۴۱-۳۳۹.

شعبانلو، س.، صدقی، خ.، ثقفیان، ب. و موسوی جهرمی، ح. ۱۳۸۷. پهنه‌بندی سیلاب در شبکه رودخانه‌های استان گلستان با استفاده از GIS. *مجله پژوهش آب ایران*. ۳ (۲): ۱۱-۲۲.

Evaluation of Intelligent Prediction Models Towards Precision of Flood Peak Flows

A. Hasani¹, F. Modaresi², K. Ebrahimi^{3*}

Received: Feb. 24, 2021

Accepted: Apr. 27, 2021

Abstract

Due to the lack of hydrological and meteorological stations in our country, the use of data-based models in estimating the maximum flood discharge is essential. In this research the efficiency of Gene Expression Programming Models and Support Vector Machine are evaluated involving peak flood discharge prediction of Mahneshan-Angoran basin as a case study, central Iran. For this purpose, observational data of 36 annual maximum daily flow (1975-2011), corresponding rainfall and average monthly temperature of three stations including Mehrabad, Yangikand and Qarahgoni were used. The observed and predicted peak discharge flows in both models were compared based on the Root Mean Square Error (RMSE), explanatory coefficient (R^2) and Nash-Sutcliffe (NSE) criteria. The mean values of RMSE in the validation stage for the GEP model in Yingikand, Qarahgooni and Mehrabad stations are equal to 0.049, 0.080 and 0.050, respectively, and in the training stage are equal to 0.042, 0.060 and 0.047, the mean values of R^2 in the validation stage in above mentioned stations are equal to 0.88, 0.86 and 0.87, respectively, and for the training stage estimated equal to 0.89, 0.89 and 0.92. NSE values in the validation stage is equal to 0.75 for all three stations. However that is equal to 0.77, 0.76 and 0.80 for the mentioned stations in the training stage. Also, the RMSE values in the SVM model for the validation stage are equal to 0.042, 0.040, 0.054, respectively, and in the training stage are equal to 0.053, 0.064 and 0.044. R^2 values in the validation stage are equal to 0.66, 0.85 and 0.73. Also for the training stage are equal to 0.86, 0.88 and 0.91. NSE values for validation are equal to 0.56, 0.75 and 0.61 and for the training stage are equal to 0.71, 0.77 and 0.80. According to the evaluation criteria, the GEP model performed relatively better and this model is more suitable for predicting floods in Mahneshan-Angoran basin.

Keywords: Data mining, Flood forecasting, Management

1- MSc student of Water Resources Engineering, Department of Irrigation & Reclamation Engineering, University of Tehran, Karaj, Iran

2- Assistant Professor, Department of Water Science and Engineering, Ferdowsi University of Mashhad, Iran

3- Professor, Department of Irrigation & Reclamation Engineering, University of Tehran, Karaj, Iran

(*Corresponding Author Email: EbrahimiK@ut.ac.ir)