



DOI:10.52547/GeoSpa.23.3.97

\*علی ابراهیم‌زاده<sup>۱</sup>  
بیتا باقری صیقلانی<sup>۲</sup>  
وحید نورانی<sup>۳</sup>

## شبیه‌سازی فرآیند بارش-رواناب در حوضه آبریز اهرچای با استفاده از روش‌های داده‌کاوی

تاریخ پذیرش: ۱۴۰۰/۰۴/۰۳

تاریخ دریافت: ۱۳۹۹/۰۸/۰۸

### چکیده

شبیه‌سازی فرآیند بارش-رواناب گامی موثر در مدیریت منابع آب، آبخیزداری، بحران کمبود آب و کنترل سیلاب می‌باشد. پیچیدگی ذاتی فرآیند بارش-رواناب، تغییرات زمانی-مکانی و عوامل موثر بر آن، شبیه‌سازی با مدل‌های فیزیکی و یا هیدرولوژیکی را مشکل می‌نماید. لذا در سال‌های اخیر روش‌های فراکاوشی همچون ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی کاربرد گسترده‌ای در مطالعات هیدرولوژی و به طور کلی پدیده‌هایی که رابطه مشخصی برای آن‌ها وجود ندارد، پیدا کرده است. حوضه آبریز اهرچای واقع در شمال غرب کشور، بدلیل تامین آب شرب، کشاورزی و صنعت نقش بسزائی در توسعه منطقه دارد. در این مقاله به ارزیابی مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی در شبیه‌سازی فرآیند بارش-رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری تازه‌کند، رواسجان، اشدلق، برمیس، اورنگ و کاسین پرداخته شده است. به منظور تعیین ترکیب ورودی مدل‌ها، پس از تهیه لیستی از متغیرهای مستقل مرتبط با رواناب هر ایستگاه، به انتخاب ورودی‌های مناسب با استفاده از دو معیار همبستگی خطی پیرسون و اطلاعات متقابل جزئی پرداخته شده و ترکیب‌های ورودی بدست آمده با استفاده از هر معیار، در محل ایستگاه‌های اشاره شده، مورد بررسی قرار گرفته است. نتایج حاکی از دقت مناسب مدل‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن، و برتری نسبی مدل شبکه عصبی مصنوعی می‌باشد. در تعیین متغیرهای ورودی نیز معیار همبستگی خطی پیرسون، باعث افزایش دقت مدل شبیه‌سازی و بهبود نتایج گردیده است.

**کلید واژه‌ها:** ایستگاه هیدرومتری، برنامه‌ریزی بیان ژن، شبکه عصبی مصنوعی، ماشین بردار پشتیبان.

\* ۱- گروه مهندسی عمران، دانشگاه آزاد اسلامی واحد مرند، مرند، ایران. (نویسنده مسئول). E-mail: ebrahimzadeh@marand-iau.ac.ir

۲- گروه معماری، دانشگاه آزاد اسلامی واحد میانه، میانه، ایران

۳- دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران.

## مقدمه

در سال‌های اخیر به دلیل تغییرات اقلیمی و فعالیت‌های بشری، تغییرات محسوسی در بارش‌ها و جریان رودخانه‌ها به وجود آمده و این تغییرات سبب وقوع سیلاب‌ها و خشکسالی‌های شدیدی شده است. بنابراین نیاز به عملیات جامع برای کاهش تلفات آب، مصرف بیش از حد آن و خطرات رو به رشد خشکسالی احساس می‌گردد. بارش به عنوان یکی از موثرترین مولفه‌های هیدرولوژیک حوضه آبخیز دارای تغییرات مکانی و زمانی است. همچنین خصوصیات متفاوت توپوگرافی، نوع خاک و کاربری اراضی در مناطق مختلف باعث ایجاد تغییر در پاسخ هیدرولوژیکی حوضه‌های آبریز می‌شود (Mostafazadeh et al., 2015). علاوه بر این، کمیت رواناب ناشی از وقوع بارش بستگی به عوامل متعددی از جمله شرایط رطوبتی اولیه و خصوصیات بارش نظیر مقدار بارش، شدت بارش و تداوم آن دارد (Sadeghi and Mostafazadeh, 2016). ارتباط بارش-رواناب و نیز مقایسه درصد ضریب رواناب در آبخیزهای مختلف، امکان مقایسه پاسخ هیدرولوژیک حوضه‌های آبخیز را فراهم می‌آورد (Tetzlaff et al., 2010). پاسخ آبخیز به وسیله میزان فصلی بودن جریان، حجم، شدت، فراوانی و مدت بارندگی و رواناب مشخص می‌شود (Mostafazadeh et al., 2015). برنامه‌ریزی بهره‌برداری از منابع آب‌های سطحی و زیرزمینی، ساماندهی رودخانه‌ها و هشدار سیل نیاز به برآورد آبدهی رودخانه‌ها و رواناب حوضه آبخیز دارد و در این میان پیچیدگی موضوع و عدم قطعیت پارامترهای موثر، همواره باعث ایجاد خطا و دشواری در شبیه‌سازی این فرآیند می‌گردد (Lee and Resdi, 2016). لذا انتخاب مدلی که با استفاده از عوامل تاثیرگذار، رواناب حاصل از بارندگی را با دقت قابل قبولی پیش‌بینی کند، امری ضروری به نظر می‌رسد. بر اساس اهمیت ورودی‌های مدل برای بدست آوردن الگوی کلی از فرآیند بارش-رواناب، مشارکت داده‌های موثر حوضه آبریز از اهمیت زیادی برخوردار است (Nourani and Komasi, 2013). در سال‌های اخیر روش‌های فراکاشی همچون ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی کاربرد گسترده‌ای در مطالعات هیدرولوژی و کلا پدیده‌هایی که الگوریتم مشخصی برای آنها وجود ندارد، پیدا کرده‌اند و این مدل‌ها در یک گستره وسیع از مسائل هیدرولوژیکی مورد استفاده قرار گرفته‌اند. مقاله حاضر به ارزیابی مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی در شبیه‌سازی فرآیند بارش-رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری تازه‌کند، رواسجان، اشلق، برمیس، اورنگ و کاسین پرداخته است. حوضه آبریز اهرچای، واقع در شمال غرب کشور، بخش قابل توجهی از حوضه آبریز ارس را پوشش داده و به دلیل تامین آب شرب، کشاورزی و صنعت نقش بسزایی در توسعه منطقه دارد.

(Nikpour et al (2019)، به بررسی بارش-رواناب رودخانه دره‌رود پرداختند و در این راستا از مدل‌های شبکه‌های عصبی مصنوعی، برنامه‌ریزی بیان ژن، موجک-عصبی و حداقل مربعات ماشین بردار پشتیبان، استفاده کردند. نتایج حاکی از آن بود که عملکرد مدل‌ها مشابه یکدیگر بوده و در عین حال مدل موجک-عصبی در تخمین مقادیر پیک جریان، بصورت جزئی عملکرد بهتری دارد. (Baba Ali and Dehghani (2019). به بررسی عملکرد مدل شبکه عصبی موجک در تخمین دبی روزانه پرداختند. نتایج حاصل نشان داد که استفاده از مدل شبکه عصبی موجک می-

تواند در زمینه تخمین دبی روزانه موثر باشد و در نوبه خود برای تسهیل توسعه و پیاده سازی استراتژی های مدیریت منابع آب مفید باشد. (Shafeizadeh et al (2019) شبیه سازی پیوسته بارش- رواناب با استفاده از شبکه های عصبی مصنوعی بر مبنای انتخاب متغیرهای موثر ورودی با الگوریتم اطلاعات متقابل جزئی (PMI<sup>۴</sup>) را انجام دادند. نتایج نشان داد که بکارگیری الگوریتم مذکور، جهت تعیین متغیرهای ورودی موثر در شبکه های عصبی مصنوعی موجب افزایش دقت شبکه ها در شبیه سازی پیوسته بارش- رواناب می شود. (Nourani et al (2017) ترکیب دو تکنیک داده- کاوی درخت تصمیم و قوانین انجمنی را در استخراج وابستگی های مهم بین بیشینه بارش ماهانه ایستگاه های سینوپتیکی ارومیه و تبریز و دمای سطح دریا های سیاه، مدیترانه و سرخ بررسی کردند. نتایج نشان داد که بین هر یک از دمای سطوح آب دریا های مورد نظر و بیشینه بارش ماهانه ایستگاه های سینوپتیکی همبستگی نسبی وجود دارد. (Mohammadi et al (2018) از تلفیق روش های برنامه ریزی خطی، تبدیل موجک متقاطع و فیلترکالمن به عنوان یک مدل کنترل کننده خطا جهت تجزیه و تحلیل نه رخداد مرکب و منفرد بارش- رواناب حوضه آبریز صوفی چای استفاده کردند. نتایج نشان داد روش زمان-مساحت اصلاح شده عملکرد نسبتاً ضعیف تری نسبت به دو روش دیگر دارد. (Ghorbani et al (2019) در تحقیقی از فیلتر کالمن همادی توام برای برآورد و کاهش عدم قطعیت در یک مدل مفهومی هیدرولوژیکی بارش- رواناب جهت مدل سازی حوضه بهشت آباد در استان چهارمحال و بخیاری استفاده نمودند و برای مدل خود زیرمدل ذوب برف را نیز اضافه نمودند. نتایج تحقیق آنها نشان داد که به روزرسانی متغیرهای حالت، خطای پایین تری را در هر یک از دسته های مختلف جریان و به خصوص جریان های پایه ارائه می دهد. (Mohammadi et al (2019) در تحقیقی دیگر به شبیه سازی وقایع بارش- رواناب با اعمال دیاگرام های اختلاف فاز پرداختند. آنها در مطالعه خود و بدون اتکاء به ساختارهای از پیش تعیین شده، رابطه بین ورودی و خروجی سیستم را صرفاً بر اساس طبیعت اطلاعات ثبت شده استخراج کردند. ایشان همچنین اختلاف فاز اتفاق افتاده بین سیگنال های بارش- رواناب را با استفاده از آنالیز موجک متقاطع محاسبه نمودند و دیاگرام های اختلاف فاز را برای وقایع منفرد و مرکب حوضه مورد مطالعه خود ترسیم نمودند. Nikpour & Abdollahi (2020) از روش های شبکه عصبی مصنوعی و عصبی-موجک برای شبیه سازی بارش ماهانه در ۴ ایستگاه سینوپتیکی واقع در استان اردبیل استفاده کردند. نتایج بیانگر عملکرد قابل قبول و برتری مدل عصبی-موجک، با بیشترین ضریب تعیین و کمترین ریشه میانگین مربعات خطا برای همه ایستگاه ها و به ازای بهترین الگو در مرحله صحت سنجی بود. (Abdollahzadeh et al (2016) از مدل های ANN<sup>۵</sup> و GEP<sup>۶</sup> و همچنین ترکیب آنها با موجک (WANN<sup>۷</sup> و WGEP<sup>۸</sup>) برای تخمین جریان روزانه رودخانه های دائمی و فصلی در منطقه ای نیمه خشک واقع در کوه های زاگرس استفاده کردند. نتایج نشان داد که از بین چهار مدل مذکور، مدل WANN عملکرد بهتری را

4-Pointwise Mutual Information

5-Artificial Neural Networks

6-Gene Expression Programming

7-Wavelet Artificial Neural Network

8-Wavelet Gene Expression Programming

در پیش‌بینی دبی رودخانه‌های دائمی و فصلی دارد. (Sofian et al (2018)، با بکارگیری مدل‌های پس انتشار و تابع پایه شعاعی، بارش ماهانه منطقه‌ای واقع در اندونزی را شبیه‌سازی کردند. برای این منظور از داده‌های بارش در بازه سالهای ۱۹۹۴ تا ۲۰۱۳ استفاده کردند. بر اساس شاخص‌های آماری خطا، مدل  $RBF^9$  عملکرد بهتری داشت. (Yasean et al (2019)، از تلفیق سه الگوریتم تکاملی مختلف به نام‌های بهینه‌سازی مجموعه ذرات<sup>۱۰</sup>، الگوریتم ژنتیک<sup>۱۱</sup> و تکامل دیفرانسیلی<sup>۱۲</sup>، با مدل  $ANFIS^{13}$  برای پیش‌بینی سری زمانی بارش ماهانه حوضه آبریز پاهانگ واقع در مالزی استفاده کردند. شبیه‌سازی بارش در سه مقیاس زمانی ماهانه، فصلی و سالانه انجام شد. نتایج حاکی از آن بود که مدل‌های هیبرید از عملکرد بهتری نسبت به مدل  $ANFIS$  برخوردارند. (Liu and Shi (2019)، از مدل برنامه‌ریزی ژنتیک به همراه یک الگوریتم بازگشتی برای پیش‌بینی طولانی مدت بارش ماهانه برای حوضه آبریزی واقع در کشور چین استفاده کردند. آنها با بکارگیری داده‌های بارش طی سالهای ۱۹۶۱ تا ۲۰۱۴ توانستند میزان بارش ماهانه تا سال ۲۰۵۰ را پیش‌بینی نمایند. (Yadava et al(2020)، به منظور بررسی بارش-رواناب در حوضه رودخانه از شبکه عصبی مصنوعی استفاده کردند. مقادیر خطای نسبی در مرحله آموزش برابر  $30/52$  و در مرحله تست برابر  $53/08$  بدست آمده است. نتایج حاصل نشان داد که مدل  $ANN$  دقت قابل قبولی دارد. (Samantary et al(2020)، با استفاده از داده‌های ۴۸ سال گذشته، از روش شبکه عصبی مصنوعی برای پیش‌بینی بارش استفاده کردند. عملکرد مدل مبتنی بر  $ANN$  و  $SVM^{14}$  با شبیه‌سازی سیلاب‌های با دوره بازگشت طولانی با یکدیگر مورد مقایسه قرار گرفت که نتایج قابل قبولی بدست آمد.

انتخاب متغیرهای ورودی مدل‌ها با استفاده از دو معیار همبستگی خطی پیرسون و اطلاعات متقابل جزئی انجام شده و با توجه به معیارهای ارزیابی ضریب تبیین، ریشه میانگین مربعات خطا و نش‌ساتکلیف، مدل مناسب انتخاب گردیده است. اهداف مقاله حاضر شامل شبیه‌سازی رواناب رودخانه اهرچای با استفاده از مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و شبکه‌های عصبی مصنوعی، همچنین بررسی و استفاده از دو معیار همبستگی خطی پیرسون و اطلاعات متقابل جزئی در تعیین ترکیب ورودی مدل‌ها از میان متغیرهای دما، بارش و تاخیرهای دبی و نیز مقایسه مدل‌های مذکور و انتخاب بهترین مدل با توجه به معیارهای ضریب تبیین، ریشه میانگین مربعات خطا و نش‌ساتکلیف جهت شبیه‌سازی فرآیند بارش-رواناب می‌باشد.

## مواد و روش‌ها

در این بخش ضمن معرفی منطقه مورد مطالعه، مشخصات مدل‌های مورد نظر شامل ماشین بردار پشتیبان، برنامه-ریزی بیان ژن و شبکه عصبی مصنوعی بیان شده است. همچنین به بررسی معیارهای ارزیابی پرداخته شده و در

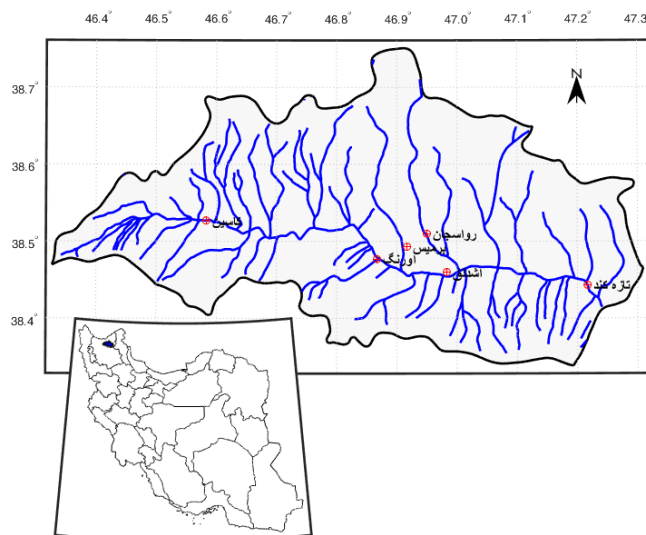
- 9-Radial Basis Function
- 10-Particle Swarm Optimization
- 11-Genetic Algorithm
- 12-Differential Evolution
- 13-Adaptive Neuro-Fuzzy Inference System
- 14-Support Vector Machines

نهایت به منظور شبیه سازی فرآیند بارش- رواناب، به تعیین ترکیب ورودی مدل ها با استفاده از دو معیار همبستگی جزئی پیرسون ( $PC^{10}$ ) و اطلاعات متقابل جزئی (PMI) پرداخته شده است.

الف) منطقه مورد مطالعه

حوضه آبریز اهرچای، از زیر حوضه های دره رود که خود از زیر حوضه های رودخانه ارس واقع در شمال غرب ایران است، می باشد. این حوضه بین مختصات جغرافیایی  $38^{\circ} 20'$  تا  $38^{\circ} 45'$  عرض شمالی و  $46^{\circ} 30'$  تا  $47^{\circ} 40'$  طول شرقی واقع شده است. وسعت این حوضه که از غرب به شرق گسترده شده ۲۲۳۲ کیلومتر مربع می باشد که ۱۷۳۱ کیلومتر مربع از آن را مناطق کوهستانی و ۵۰۲ کیلومتر مربع را دشت ها فرا گرفته اند. این حوضه در محدوده سیاسی استان آذربایجان شرقی قرار داشته و نقاط شهری مهم واقع در این حوضه شهرهای اهر و ورزقان می باشند. رودخانه اصلی حوضه آبریز اهرچای، رودخانه هم نام با حوضه یعنی اهرچای است که دارای شاخه های فرعی زیادی می باشد. رودخانه اهرچای در خروجی حوضه به رود قره سو پیوسته و رود دره رود را تشکیل می دهند. موقعیت جغرافیایی حوضه آبریز اهرچای و ایستگاه های هیدرومتری در شکل (۱) نشان داده شده است.

مهمترین ایستگاه اندازه گیری دبی حوضه، به لحاظ قدمت دوره آماری، ایستگاه تازه کند اهر می باشد که در سال ۱۳۵۱ تأسیس شده است. سایر ایستگاه های هیدرومتری منطقه شامل رواسجان، اشدلق، برمیس، اورنگ و کاسین می باشد که در این مقاله به شبیه سازی دبی رواناب رودخانه در این ایستگاه ها پرداخته شده است. ایستگاه های تازه کند، اشدلق، اورنگ و کاسین بر روی رودخانه اهرچای و دو ایستگاه رواسجان و برمیس به ترتیب بر روی شاخه های فرعی برمیس چای و کشان چای واقع شده اند. مشخصات جغرافیایی محل ایستگاه ها و خصوصیات آماری داده های هیدرومتری در جدول (۱) ارائه شده است.



شکل ۱: موقعیت جغرافیایی حوضه آبریز اهرچای و ایستگاه های هیدرومتری

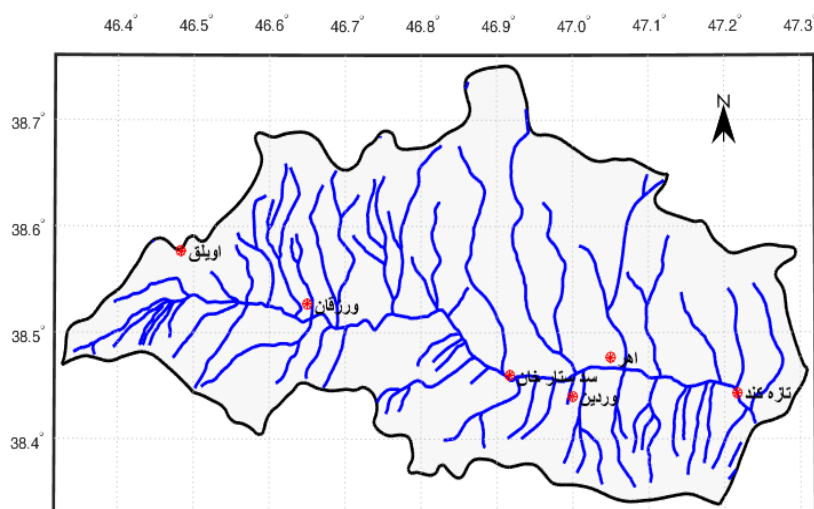
Figure 1: The geographical location of Ahr chay watershed and hydrometric stations

جدول ۱- مختصات جغرافیایی ایستگاه‌های هیدرومتری و خصوصیات آماری داده‌های رواناب

Table 1- Geographical coordinates of hydrometric stations and statistical properties of runoff data

ایستگاه	رودخانه	مختصات جغرافیایی			خصوصیات آماری داده‌های رواناب					
		طول	عرض	ارتفاع	کمینه	میانگین	بیشینه	انحراف معیار	چولگی	کشیدگی
تازه‌کند	اهرچای	۴۷-۱۳	۳۸-۲۶	۱۲۲۵	۰/۰۰	۱/۴۱	۷۸/۰۰	۲/۰۴	۱۳/۷۷	۴۶۰/۹۶
رواسجان	کشان‌چای	۴۶-۵۷	۳۸-۳۰	۱۴۶۰	۰/۰۰	۰/۴۱	۹/۱۰	۰/۸۲	۵/۴۹	۳۷/۵۱
اشدلق	اهرچای	۴۶-۵۹	۳۸-۲۷	۱۳۲۰	۰/۰۰	۰/۷۵	۱۷۸/۰۰	۴/۱۶	۲۸/۹۵	۱۰۳۵/۲۸
برمیس	برمیس‌چای	۴۶-۵۵	۳۸-۳۲	۱۳۵۰	۰/۰۰	۰/۰۳	۲/۳۴	۰/۰۹	۱۰/۷۹	۱۷۹/۴۹
اورنگ	اهرچای	۴۶-۵۲	۳۸-۲۸	۱۴۳۰	۰/۰۰	۱/۴۷	۳۱/۷۰	۲/۰۶	۴/۵۹	۴۱/۸۴
کاسین	اهرچای	۴۶-۳۵	۳۸-۳۱	۱۶۸۰	۰/۰۰	۱/۰۷	۵۹/۰۰	۲/۳۵	۱۰/۴۸	۱۷۰/۴۱

در این بررسی استفاده از داده‌های بارش روزانه ایستگاه‌های هواشناسی تازه‌کند، اهر، وردین، سد ستارخان، ورزقان و اوئلیق و داده‌های دمای متوسط روزانه ایستگاه سد ستارخان در شبیه‌سازی رواناب رودخانه مورد بررسی قرار گرفته است. موقعیت ایستگاه‌های هواشناسی مذکور در شکل (۲) و مختصات جغرافیایی ایستگاه‌ها و خصوصیات آماری داده‌های بارش و دمای متوسط روزانه در جدول (۲) ارائه شده است. داده‌های مورد استفاده در این بررسی مربوط به بازه زمانی مهر ماه سال ۱۳۸۱ تا پایان شهریور ماه سال ۱۳۹۳ می‌باشند.



شکل ۲: موقعیت جغرافیایی ایستگاه‌های هواشناسی

Figure 2: The geographical location of meteorological stations

جدول ۲- مختصات جغرافیایی ایستگاه های هواشناسی و خصوصیات آماری داده های بارندگی و دمای میانگین روزانه

Table 2- Geographical coordinates of meteorological stations and statistical characteristics of rainfall data and average daily temperature

ایستگاه	نوع داده	مختصات جغرافیایی		خصوصیات آماری داده ها					
		عرض	طول	کمینه	میانگین	بیشینه	انحراف معیار	چولگی	کشیدگی
تازه کند	بارش	۳۸-۲۶	۴۷-۱۳	-۱۵/۴۰	۰/۷۰	۳۷/۰۰	۲/۸۶	۵/۹۱	۴۶/۶۶
اهر	بارش	۳۸-۲۸	۴۷-۰۳	-۱۵/۵۴	۰/۷۹	۴۱/۷۰	۲/۷۵	۵/۷۸	۴۸/۹۱
وردین	بارش	۳۸-۲۶	۴۶-۵۸	-۱۶/۰۲	۰/۹۳	۳۹/۶۰	۲/۸۹	۵/۳۱	۴۱/۱۰
ورزقان	بارش	۳۸-۳۱	۴۶-۳۹	-۱۶/۸۷	۰/۹۸	۴۲/۰۰	۳/۱۱	۶/۰۱	۵۳/۲۶
اویلق	بارش	۳۸-۳۴	۴۶-۲۹	-۱۵/۷۴	۱/۲۶	۴۳/۳۴	۳/۹۶	۴/۶۵	۳۳/۵۷
سد ستار خان	بارش	۳۸-۲۷	۴۶-۵۵	-۱۶/۳۸	۰/۹۷	۴۴/۱۱	۳/۳۲	۵/۷۴	۴۷/۲۵
سد ستار خان	دما	۳۸-۲۷	۴۶-۵۵	-۱۴/۳۰	۲/۳۳	۳۸/۳۹	۴/۴۵	۳/۳۹	۳۸/۱۸

(ب) مشخصات مدل ها

مشخصات ماشین بردار پشتیبان

به منظور تخمین رواناب با استفاده از مدل ماشین بردار پشتیبان می توان انواع تابع کرنل را مورد بررسی قرار داد، در واقع انتخاب تابع کرنل مناسب در استفاده از این مدل بسیار حائز اهمیت می باشد و نتایج مختلفی را به همراه خواهد داشت. در این پژوهش علاوه بر تابع پایه شعاعی به عنوان رایج ترین تابع کرنل، توابع کرنل خطی و چندجمله ای نیز مورد بررسی قرار گرفته اند. مقادیر بهینه پارامترهای ماشین بردار پشتیبان و پارامترهای  $\sigma$  مربوط به تابع کرنل پایه شعاعی و  $t$  و  $d$  (درجه تابع) مربوط به تابع کرنل چندجمله ای از طریق سعی و خطا محاسبه شده است. با برنامه نویسی مرتبط با هر سه تابع کرنل در محیط متلب، فرآیند محاسبات این مدل انجام شده است.

مشخصات برنامه ریزی بیان ژن

روند مدل سازی با استفاده از روش برنامه ریزی بیان ژن را به طور خلاصه می توان به پنج مرحله تقسیم نمود:

- مرحله نخست شامل تعیین تابع برازش می باشد که جذر میانگین مربعات خطا به عنوان معیار خطای تابع برازش انتخاب شده است.
- مرحله دوم شامل انتخاب مجموعه ترمینال ها و توابع برای ایجاد کروموزوم ها می باشد.
- مرحله سوم شامل انتخاب کروموزوم ها است که مربوط به انتخاب طول راس و تعداد ژن ها می باشد. طول راس برابر با هشت انتخاب شده و تعداد سه ژن نیز در هر کروموزوم در نظر گرفته شده است، همچنین در هر اجرا تعداد ۳۰ کروموزوم انتخاب گردیده است.
- مرحله چهارم شامل انتخاب تابع پیوند است که تعیین کننده ارتباط بین زیر درخت ها می باشد. در این بررسی سه ژن انتخابی زیر درخت ها (ETS) از طریق تابع مجموع به هم پیوند داده می شوند.

- مرحله پنجم شامل انتخاب عملگرهای ژنتیکی و نرخ آنها است. در این مورد ترکیبی از کلیه عملگرهای بهسازی از قبیل جهش، وارون‌سازی، سه نوع ترانهش و سه نوع ترکیب استفاده شده است. در این پژوهش جهت پیش بینی میزان دبی به روش برنامه‌ریزی بیان ژن از نرم افزار GeneXpro Tools 4.0 استفاده شده است. این نرم افزار ابزار تحلیل داده به شمار می‌آید که به کمک داده‌های موجود امکان ایجاد مدل‌های شبیه سازی داده می‌شود.

#### مشخصات شبکه عصبی مصنوعی

بطور کلی فرآیند مدل‌سازی با روش شبکه عصبی مصنوعی شامل مراحل زیر می‌باشد:

- تهیه زوج‌های آموزش: داده‌های مربوط به نرون‌های ورودی و خروجی موجود دسته‌بندی شده و به صورت زوج‌های آموزش ارائه می‌شوند.
- نرمالیزه کردن زوج‌های آموزش: بدلیل استفاده از توابع محرک، بایستی زوج‌های آموزش و صحت‌سنجی در بازه مورد قبول تابع (یعنی منفی یک تا مثبت یک و یا صفر تا مثبت یک) قرار گیرند. در این پژوهش ابتدا داده‌های ورودی و خروجی، با استفاده از رابطه (۱) در محدوده منفی یک تا مثبت یک قرار گرفتند.

$$X_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{رابطه (۱)}$$

- یادگیری شبکه: در این قسمت با روش‌های گوناگون بهترین وزن نرون‌ها بدست می‌آید. آموزش شبکه‌های پرسپترون چندلایه با استفاده از الگوریتم آموزش پس انتشار خطای لونیگ-مارکواریت (به دلیل همگرایی سریع و دقیق به خطای حداقل) انجام و حداکثر تعداد تکرارها در فرآیند یادگیری شبکه ۱۰۰۰ در نظر گرفته شده است. یکی از نکات مهم در آموزش شبکه‌های عصبی، تصمیم‌گیری در مورد توقف فرآیند آموزش است، زیرا اگر آموزش شبکه به درستی متوقف نشود، شبکه مستعد مشکل فوق‌برازشی یا ضعف‌برازشی می‌شود. در این مواقع برای جلوگیری از اینگونه مشکلات از تکنیک توقف آموزش استفاده می‌شود. بدین منظور، کل داده‌ها به سه دسته (آموزش، اعتبارسنجی و تست) تقسیم می‌شوند، هر زمان که خطای اعتبارسنجی شبکه زیاد شود، فرآیند آموزش متوقف خواهد شد.
- تست شبکه: در این قسمت شبکه برای زوج داده‌های تست، مورد آزمون قرار گرفته و میزان مؤثر بودن شبکه مورد بررسی قرار گرفته است.
- یافتن بهترین ساختار برای شبکه: تعداد نرون‌های موجود در لایه پنهان هم با سعی و خطا (تکرار دو مرحله قبلی) و به منظور یافتن مدلی با کمترین مقدار خطای داده‌های تست تعیین شد. روند کار با تعداد نرون لایه میانی برابر با یک آغاز و افزودن نرون‌های اضافی تا حداکثر ۲۰ نرون ادامه پیدا کرد. به منظور مدل‌سازی شبکه عصبی مصنوعی از جعبه ابزار شبکه عصبی موجود در نرم‌افزار MATLAB استفاده شده است.



## ج) معیارهای ارزیابی

پس از تحلیل داده‌ها با انواع روش‌ها و مدل‌ها لازم است عملکرد آنها مورد ارزیابی قرار گیرد. بدین منظور روش‌های متنوعی وجود دارد که عمده‌ترین این روش‌ها مقایسه مقادیر هدف و محاسباتی مدل‌ها با استفاده از معیارهای ارزیابی می‌باشد. در این مقاله از جمله معیارهای پرکاربرد در مهندسی آب شامل ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطا ( $RMSE^{16}$ ) و معیار نش ساتکلیف (E) استفاده شده، که روابط آنها به ترتیب شامل (۲)، (۳) و (۴) می‌باشد. در این روابط  $Q_{xi}$  و  $Q_{yi}$ ، به ترتیب مقادیر هدف و محاسباتی دبی رودخانه در گام زمانی  $i$  ام،  $N$  تعداد داده‌ها،  $\bar{Q}_x$  و  $\bar{Q}_y$  نیز میانگین مقادیر مشاهداتی و محاسباتی دبی رودخانه می‌باشد.

$$R^2 = \frac{\left(\sum_{i=1}^N (Q_{xi} - \bar{Q}_x)(Q_{yi} - \bar{Q}_y)\right)^2}{\sum_{i=1}^N (Q_{xi} - \bar{Q}_x)^2 * \sum_{i=1}^N (Q_{yi} - \bar{Q}_y)^2} \quad \text{رابطه (۲)}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{xi} - Q_{yi})^2} \quad \text{رابطه (۳)}$$

$$E = 1 - \frac{\sum_{i=1}^N (Q_{xi} - Q_{yi})^2}{\sum_{i=1}^N (Q_{xi} - \bar{Q}_x)^2} \quad \text{رابطه (۴)}$$

مدلی بهترین نتیجه را خواهد داشت که برای سه معیار فوق به ترتیب مقادیری نزدیک به یک، صفر و یک را به همراه داشته باشد.

به منظور تعیین ترکیب ورودی مناسب، در ابتدا لیستی از متغیرهای مستقل مرتبط با رواناب هر ایستگاه تهیه شده است. سپس به انتخاب ورودی‌های مناسب از میان لیست متغیرها پرداخته شده است. دو معیار همبستگی خطی پیرسون (PC) و اطلاعات متقابل جزئی (PMI) در انتخاب ورودی‌های مناسب بکار گرفته شده و ترکیب‌های ورودی بدست آمده از هر روش در شبیه‌سازی بارش-رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری مورد بررسی قرار گرفته است. علاوه بر این دو ترکیب، مدل‌سازی به کمک کلیه متغیرهای مستقل مرتبط نیز انجام شده است (Total). بدین ترتیب برای هر ایستگاه از سه ترکیب ورودی استفاده شده است. در ایستگاه‌های تازه‌کند، اشدلق، برمیس و اورنگ تعداد متغیرهای ورودی انتخاب شده با معیار PMI کمتر از تعداد متغیرهای ورودی انتخاب شده با معیار PC بوده و در ایستگاه‌های رواسجان و کاسین تعداد متغیرهای ورودی انتخاب شده با معیار PC کمتر از تعداد متغیرهای ورودی انتخاب شده با معیار PMI می‌باشد. در روش انتخاب غیرخطی می‌باشد، به داده‌های بارندگی و در روش خطی PC به تأخیرهای دبی همان ایستگاه اهمیت بیشتری داده شده است. در همه ترکیب‌های ورودی، دبی همان ایستگاه با یک روز تأخیر وجود دارد.

داده‌های مورد استفاده در این تحقیق شامل داده‌های روزانه ۱۲ سال در بازه زمانی مهر ماه سال ۱۳۸۱ تا پایان شهریور ماه سال ۱۳۹۳ می‌باشند. برای هر ۶ ایستگاه، داده‌های ۱۰ سال اول (مهر ماه سال ۱۳۸۱ تا پایان شهریور ماه سال ۱۳۹۱) جهت آموزش و داده‌های دو سال آخر (مهر ماه سال ۱۳۹۱ تا پایان شهریور ماه سال ۱۳۹۳) جهت

صحت‌سنجی مدل‌های SVM، GEP و ANN انتخاب شده است. خصوصیات آماری داده‌های آموزش و صحت‌سنجی مدل رواناب حوضه آبریز اهرچای در محل هر یک از شش ایستگاه در جدول (۳) ارائه شده است.

جدول ۳- خصوصیات آماری داده‌های آموزش و صحت‌سنجی مدل رواناب حوضه آبریز اهرچای

Table 3- Statistical characteristics of the training and validation data of the Ahrchay watershed runoff model

ایستگاه	مرحله	تعداد	کمینه	میانگین	بیشینه	انحراف معیار	چولگی	کشیدگی
کاسین	آموزش	۳۶۵۲	۰/۰۰	۱/۱۲	۵۹/۰۰	۲/۵۳	۹/۹۲	۱۵۰/۰۰
	تست	۷۳۱	۰/۰۰	۰/۸۴	۱۲/۲۰	۰/۹۶	۴/۶۶	۴۴/۶۵
اورنگ	آموزش	۳۶۵۲	۰/۰۰	۱/۵۳	۳۱/۷۰	۲/۱۹	۴/۴۶	۳۸/۵۶
	تست	۷۳۱	۰/۰۰	۱/۱۹	۱۰/۹۰	۱/۱۷	۲/۲۷	۱۴/۱۱
برمیس	آموزش	۳۶۵۲	۰/۰۰	۰/۰۴	۲/۳۴	۰/۱۰	۱۰/۳۴	۱۶۰/۶۷
	تست	۷۳۱	۰/۰۰	۰/۰۳	۰/۳۰	۰/۰۴	۲/۹۸	۱۴/۳۲
اشدلق	آموزش	۳۶۵۲	۰/۰۰	۰/۸۴	۱۷۸/۰۰	۴/۵۴	۲۶/۵۴	۸۶۸/۰۸
	تست	۷۳۱	۰/۰۰	۰/۳۱	۵/۸۹	۰/۴۶	۶/۱۶	۵۶/۰۸
رواسجان	آموزش	۳۶۵۲	۰/۰۰	۰/۳۲	۹/۱۰	۰/۸۹	۵/۰۰	۳۱/۴۲
	تست	۷۳۱	۰/۰۰	۰/۰۸	۲/۲۱	۰/۱۴	۹/۵۵	۱۳۶/۳۲
تازه‌کند	آموزش	۳۶۵۲	۰/۰۰	۱/۴۹	۷۸/۰۰	۲/۲۰	۱۳/۰۶	۴۰۶/۷۳
	تست	۷۳۱	۰/۰۰	۱/۰۳	۶/۱۳	۰/۷۷	۱/۹۳	۱۱/۹۸

## نتایج و بحث

در این بخش، استفاده از مدل‌های SVM، GEP و ANN در شبیه‌سازی رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری تازه‌کند، رواسجان، اشدلق، برمیس، اورنگ و کاسین و همچنین کارایی معیارهای PC و PMI در انتخاب ترکیب مناسب ورودی‌های مدل مورد بررسی قرار گرفته و بهترین مدل شبیه‌سازی انتخاب شده است. جذر میانگین مربعات خطا (RMSE)، ضریب تبیین ( $R^2$ ) و معیار نش‌سانکلیف (E) معیارهای ارزیابی بکار رفته در این مقاله می‌باشند.

الف) شبیه‌سازی با مدل ماشین بردار پشتیبان (SVM)

به منظور شبیه‌سازی با استفاده از مدل SVM، توابع کرنل پایه شعاعی، چندجمله‌ای و خطی مورد بررسی قرار گرفته و از طریق سعی و خطا مقادیر بهینه پارامترهای مورد نیاز مدل ماشین بردار پشتیبان بدست آمده است.

نتایج مرحله تست نشان داد که در ایستگاه‌های تازه‌کند، اورنگ و کاسین انتخاب ترکیب ورودی با معیار PC و در ایستگاه‌های رواسجان و اشدلق انتخاب ترکیب ورودی با معیار PMI نتایج بهتری به همراه داشته است. در ایستگاه برمیس هم که استفاده از کل ورودی‌های پیشنهادی (Total) بهترین نتایج را به همراه داشته، نتایج ترکیب ورودی PC به ترکیب ورودی Total بسیار نزدیک می‌باشد. با مقایسه نتایج مشاهده شد که در ایستگاه‌های تازه‌کند، برمیس و

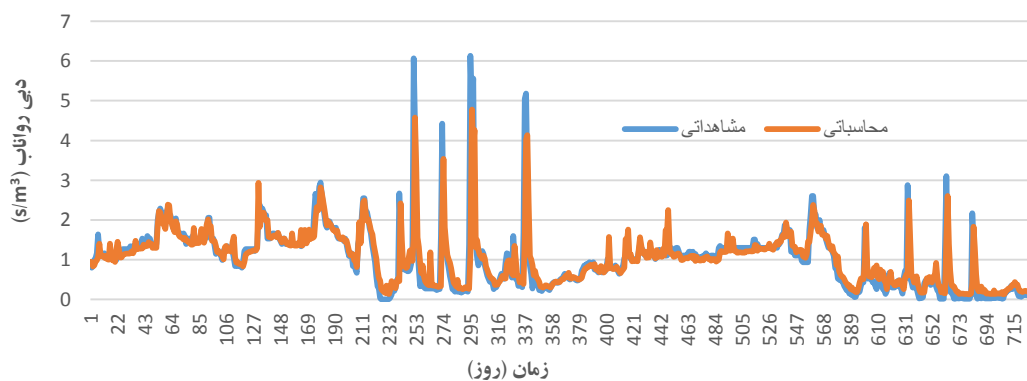
اورنگ که تعداد متغیرهای ورودی ترکیب PC بیشتر از PMI است، ترکیب ورودی PC نتایج بهتری داشته و در ایستگاه رواسجان که تعداد متغیرهای ورودی ترکیب PMI بیشتر از PC است، ترکیب ورودی PMI نتایج بهتری به همراه داشته است. در ایستگاههای اشدلق و کاسین ترکیبهای ورودی با تعداد کمتر، نتایج بهتری داشته است. این موضوع نشان دهنده کارایی مدل SVM در شبیه سازی مدل هایی است که دارای تعداد زیادی متغیر ورودی می باشند، هر چند که در مدل SVM استفاده از تمام متغیرهای ورودی مرتبط نتایج ضعیف تری به همراه داشته و باید متغیرهای کم اهمیت از ورودی مدل ها حذف گردند. همچنین نتایج حاکی از آن بود که در ایستگاههای تازه کند، رواسجان، اشدلق استفاده از تابع هسته پایه شعاعی، در ایستگاه برمیس تابع هسته چند جمله ای و در ایستگاههای اورنگ و کاسین تابع هسته خطی نتایج مطلوب تری داشته و نمی توان برای همه ایستگاهها یک نوع تابع هسته خاص را پیشنهاد نمود. نتایج حاصل از مدل سازی SVM با ترکیب ورودی های PC، PMI و Total برای شش ایستگاه تازه کند، رواسجان، اشدلق، برمیس، اورنگ و کاسین در جدول (۴) ارائه شده است.

پس از بررسی نتایج حالت های مختلف ورودی و توابع مختلف هسته و انتخاب بهترین مدل SVM برای هر ایستگاه، به ترسیم نمودارهای مربوطه پرداخته شده است. نمودارهای سری زمانی دبی رواناب مشاهداتی و محاسباتی دوره تست شش ایستگاه مورد مطالعه به ترتیب در شکل های (۳) تا (۸) نشان داده شده است.

جدول ۴- نتایج حاصل از مدل SVM با ترکیب ورودی های PC، PMI و Total

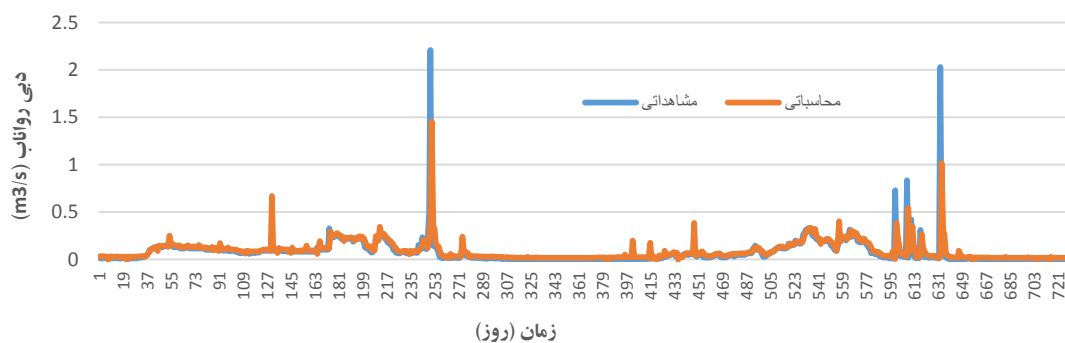
Table 4- The results of SVM model with the combination of PC, PMI and Total inputs

تست			آموزش			هسته	ورودی	ایستگاه
E	RMSE	R <sup>2</sup>	E	RMSE	R <sup>2</sup>			
۰/۶۸	۰/۴۳	۰/۶۸	۰/۶۰	۱/۴۰	۰/۶۰	پایه شعاعی	PC	تازه کند
۰/۰۹	۰/۱۳	۰/۲۰	۰/۸۶	۰/۳۴	۰/۸۶	پایه شعاعی	PMI	رواسجان
۰/۳۶	۰/۳۷	۰/۴۴	۰/۶۳	۲/۷۶	۰/۶۳	پایه شعاعی	PMI	اشدلق
۰/۸۸	۰/۰۲	۰/۸۸	۰/۶۵	۰/۰۶	۰/۶۵	چند جمله ای	Total	برمیس
۰/۷۰	۰/۶۴	۰/۷۰	۰/۸۵	۰/۸۵	۰/۸۵	خطی	PC	اورنگ
۰/۶۹	۰/۵۴	۰/۶۹	۰/۵۷	۱/۶۶	۰/۵۷	خطی	PC	کاسین



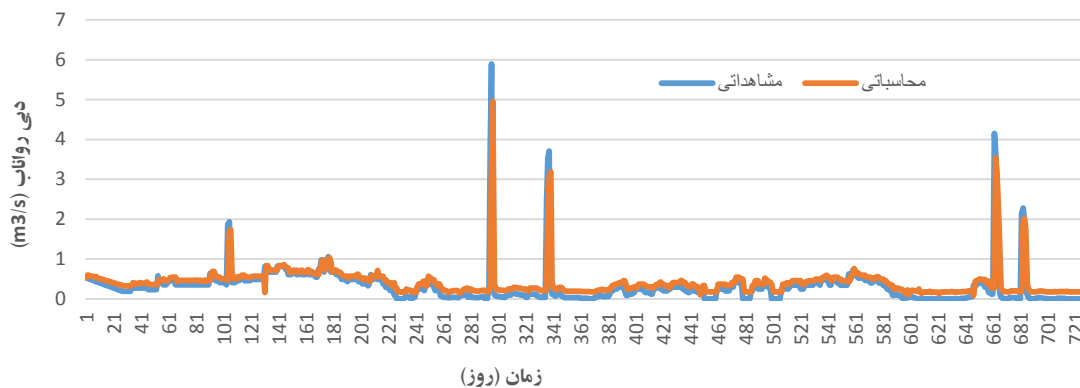
شکل ۳: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه تازه کند برای دوره تست

Figure 3: Time series graph of observed and calculated runoff discharge with SVM model at Tazekand station for the test period



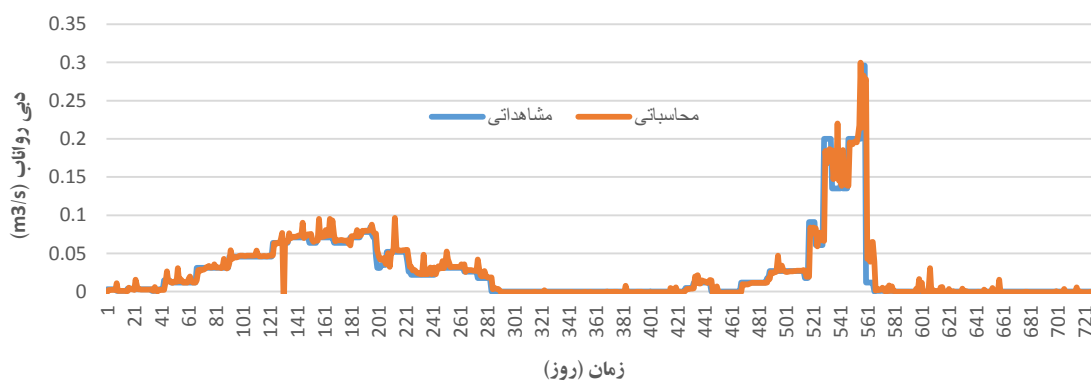
شکل ۴: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه روسجان برای دوره تست

Figure 4: Time series graph of observed and calculated runoff discharge with SVM model at Ravasjan station for the test period



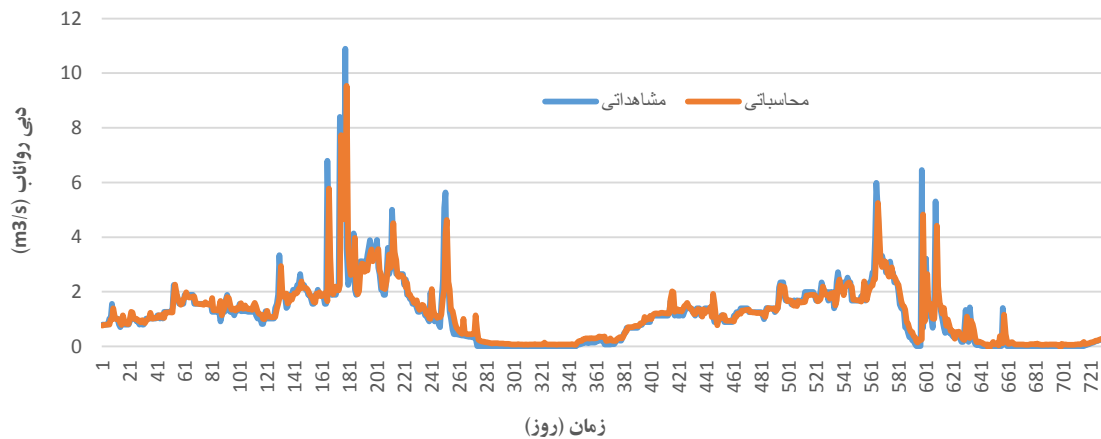
شکل ۵: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه اشدلق برای دوره تست

Figure 5: Time series graph of observed and calculated runoff discharge with SVM model at Oushdilaq station for the test period



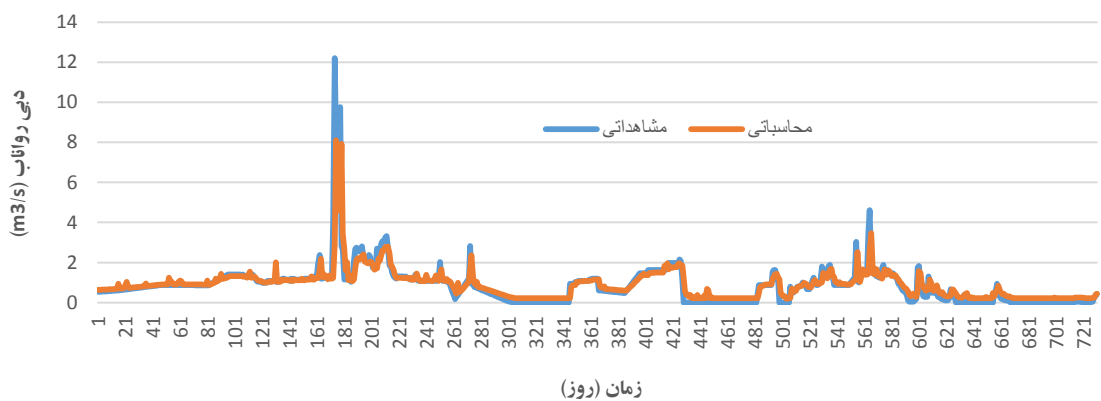
شکل ۶: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه برمیس برای دوره تست

Figure 6: Time series graph of observed and calculated runoff discharge with SVM model at Bermis station for the test period



شکل ۷: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه اورنگ برای دوره تست

Figure 7: Time series graph of observed and calculated runoff discharge with SVM model at Owrang station for the test period



شکل ۸: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل SVM در ایستگاه کاسین برای دوره تست

Figure 8: Time series graph of observed and calculated runoff discharge with SVM model at Kasin station for the test period

ب) شبیه سازی با مدل برنامه ریزی بیان ژن (GEP)

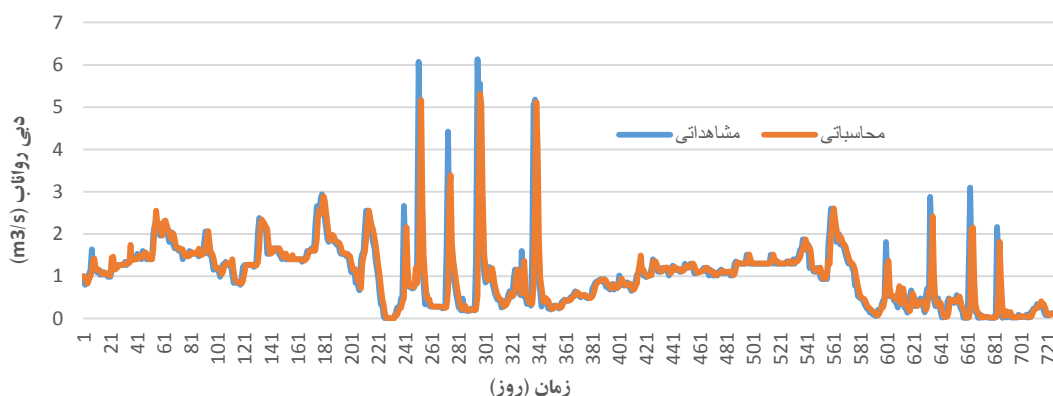
نتایج مرحله تست به روش GEP حاکی از آن است که در ایستگاه‌های رواسجان و کاسین انتخاب ترکیب ورودی با معیار PC و در ایستگاه‌های تازه‌کند، اشدلق، برمیس و اورنگ انتخاب ترکیب ورودی با معیار PMI نتایج بهتری به همراه داشته است. قابل ملاحظه است که در هیچ یک از ایستگاه‌ها استفاده از کل ورودی‌های پیشنهادی (Total) بهترین نتایج را به همراه نداشته و با کاهش تعداد متغیرهای ورودی بر دقت مدل افزوده شده است. با بررسی تعداد متغیرهای ترکیب‌های مختلف ورودی مشخص شد که در تمام ایستگاه‌ها ترکیب ورودی با تعداد متغیر کمتر نتیجه بهتری به همراه داشته است و GEP برای مدل‌های با تعداد ورودی کم مناسب می‌باشد. نتایج حاصل از مدل GEP در جدول (۵) ارائه شده است.

جدول ۵- نتایج حاصل از مدل GEP با ترکیب ورودی‌های PC، PMI و Total

Table 5- The results of GEP model with the combination of PC, PMI and Total inputs

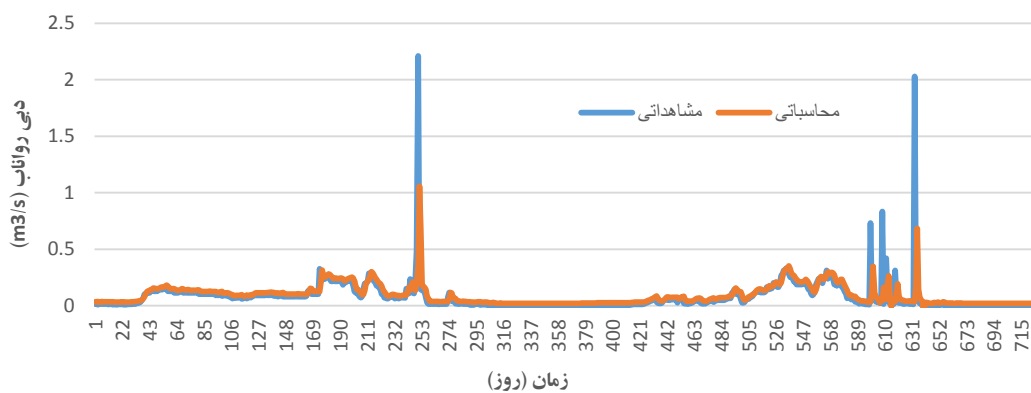
ایستگاه	ورودی	آموزش			تست		
		E	RMSE	R <sup>2</sup>	E	RMSE	R <sup>2</sup>
تازه کند	PMI	۰/۵۴	۱/۴۹	۰/۵۵	۰/۵۱	۰/۵۴	۰/۵۵
رواسجان	PC	۰/۸۳	۰/۳۷	۰/۸۳	۰/۲۱	۰/۱۲	۰/۲۴
اشدلق	PMI	۲/۶۴	۰/۶۶	۰/۶۶	۰/۲۶	۰/۳۹	۰/۴۲
برمیس	PMI	۰/۰۶	۰/۶۵	۰/۶۵	۰/۹۱	۰/۰۱	۰/۹۱
اورنگ	PMI	۰/۸۱	۰/۸۶	۰/۸۶	۰/۵۳	۰/۸۰	۰/۶۲
کاسین	PC	۱/۶۲	۰/۵۹	۰/۶۸	۰/۶۸	۰/۵۵	۰/۶۸

پس از بررسی نتایج حالت‌های مختلف ورودی و انتخاب بهترین مدل GEP هر ایستگاه، نمودارهای سری زمانی دبی رواناب مشاهداتی و محاسباتی دوره تست شش ایستگاه مورد مطالعه به ترتیب در شکل‌های (۹) تا (۱۴) نشان داده شده است.



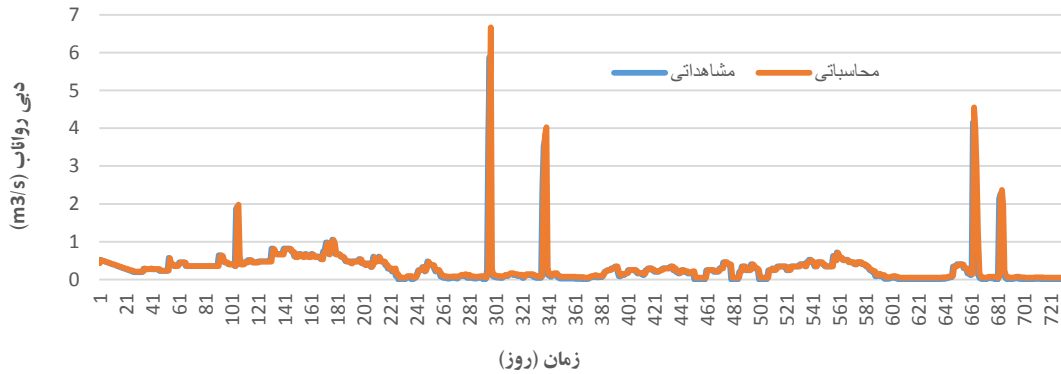
شکل ۹: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه تازه‌کند برای دوره تست

Figure 9: The time series graph of observed and calculated runoff discharge with GEP model at Tazekand station for the test period



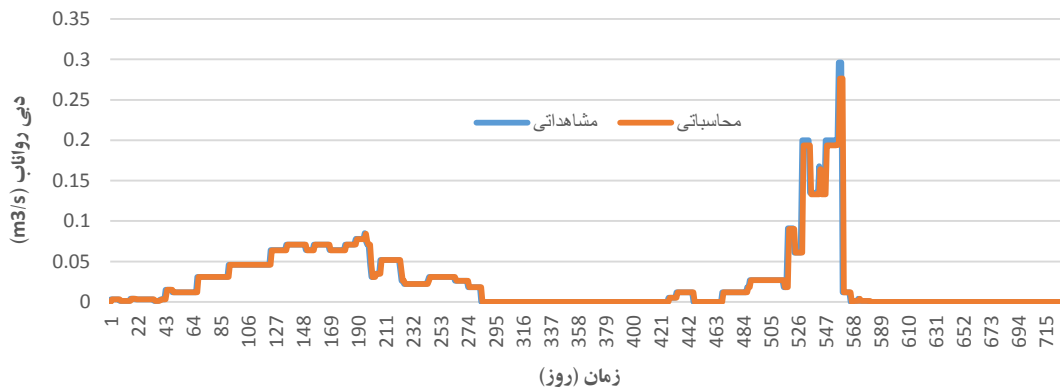
شکل ۱۰: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه رواسجان برای دوره تست

Figure 10: The time series graph of observed and calculated runoff discharge with GEP model at Ravasjan station for the test period



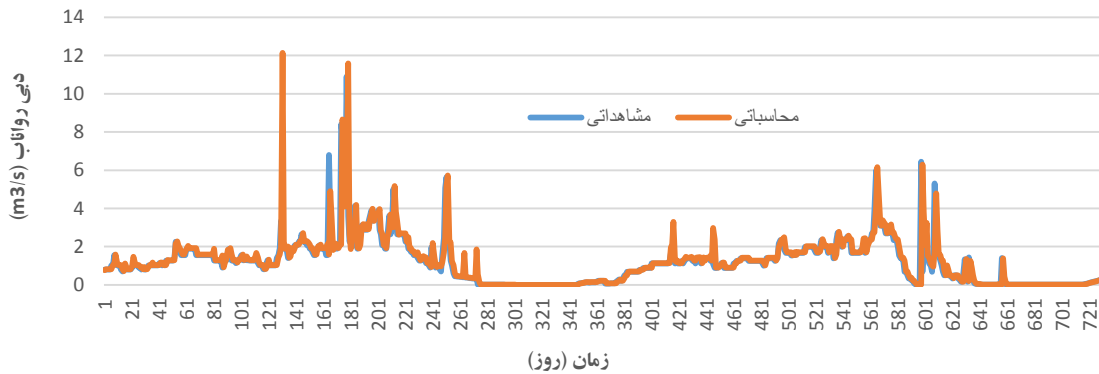
شکل ۱۱: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه اشدلق برای دوره تست

Figure 11: The time series graph of observed and calculated runoff discharge with GEP model at Oushdilaq station for the test period



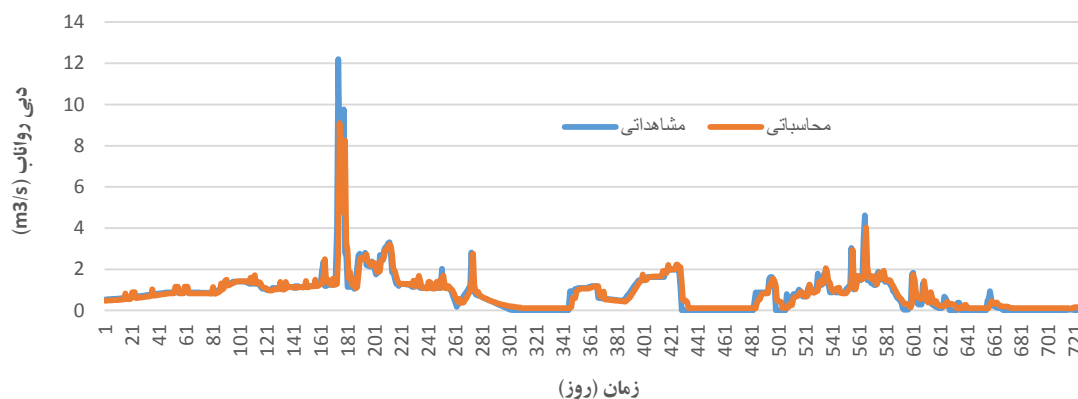
شکل ۱۲: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه برمیس برای دوره تست

Figure 12: The time series graph of observed and calculated runoff discharge with GEP model at Bermis station for the test period



شکل ۱۳: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه اورنگ برای دوره تست

Figure 13: The time series graph of observed and calculated runoff discharge with GEP model at Owrang station for the test period



شکل ۱۴: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل GEP در ایستگاه کاسین برای دوره تست

Figure 14: The time series graph of observed and calculated runoff discharge with GEP model at Kasin station for the test period

ج) شبیه‌سازی با مدل شبکه عصبی مصنوعی (ANN)

در ساختار شبکه عصبی مصنوعی برای شبیه‌سازی بارش-رواناب حوضه آبریز اهرچای از شبکه عصبی پرسپترون سه لایه، دارای تابع تانژانت سیگموئید و محرک خطی برای نگاشت اطلاعات به ترتیب از لایه ورودی به لایه پنهان و از لایه پنهان به لایه خروجی با الگوریتم آموزش پس انتشار لونیگ-مارکواریت استفاده شده است. تعداد نرون‌های لایه مخفی نیز با سعی و خطا بدست آمده است. بدین صورت که برای لایه مخفی تعداد نرون‌های ۱ تا ۲۰ در نظر گرفته شده و با توجه به معیارهای ارزیابی دوره تست، ساختار مناسب انتخاب گردیده است. نتایج شبیه‌سازی با تعداد نرون‌های ۱ تا ۲۰ برای ترکیب‌های مختلف ورودی Total، PMI و PC در شش ایستگاه بدست آمد. پس از انتخاب ساختار مناسب، به بررسی نتایج سه ترکیب ورودی برای هر کدام از ایستگاه‌ها پرداخته شد. نتایج حاصل از بهترین مدل‌های شبیه‌سازی ANN برای شش ایستگاه تازه‌کند، رواسجان، اشدلق، برمیس، اورنگ و کاسین در جدول (۶) نشان داده شده است.

نتایج مرحله تست مدل‌های ANN حاکی از این است که در ایستگاه‌های اشدلق، برمیس، اورنگ و کاسین استفاده از کل ورودی‌های پیشنهادی (Total) بهترین نتایج را به همراه داشته و تنها در ایستگاه‌های تازه‌کند و رواسجان انتخاب ترکیب ورودی با معیار PC نتایج بهتری به همراه داشته است. نکته قابل توجه اینکه در هیچ یک از ایستگاه‌ها ترکیب ورودی با معیار PMI بهترین مدل نبوده است. لذا بر خلاف مدل‌های SVM و GEP، به طور کلی استفاده از روش‌های PC و PMI در ANN باعث بهبود نتایج نشده و استفاده از ترکیب Total نتایج بهتری به همراه داشته است.



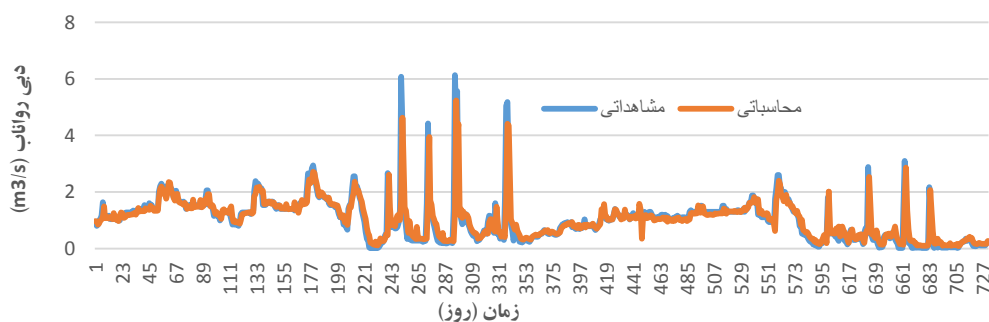
پس از بررسی نتایج حالت های مختلف ورودی و انتخاب بهترین مدل ANN هر ایستگاه، به ترسیم نمودارهای سری زمانی دبی-رواناب مشاهداتی و محاسباتی دوره تست شش ایستگاه مورد مطالعه به ترتیب در شکل های (۱۵) تا (۲۰) نشان داده شده است.

در نهایت با مقایسه ساختار بهینه مدل های مختلف مشخص شد که مدل های مورد نظر برای هر شش ایستگاه از دقت قابل قبولی در شبیه سازی رواناب حوضه آبریز اهرچای برخوردارند. با توجه به نتایج حاصل، مدل ANN در سه ایستگاه تازه کند، اشدلق و کاسین، مدل GEP در دو ایستگاه رواسجان و برمیس و مدل SVM در ایستگاه اورنگ نتایج بهتری را در شبیه سازی رواناب حوضه آبریز اهرچای داشته اند. در مجموع شبکه عصبی مصنوعی در شبیه سازی رواناب حوضه آبریز اهرچای یا بهترین مدل بوده و یا به بهترین مدل نزدیک می باشد و می توان از آن در شبیه سازی رواناب حوضه آبریز اهرچای استفاده نمود.

جدول ۶- نتایج حاصل از مدل ANN با ترکیب ورودی های PC، PMI و Total

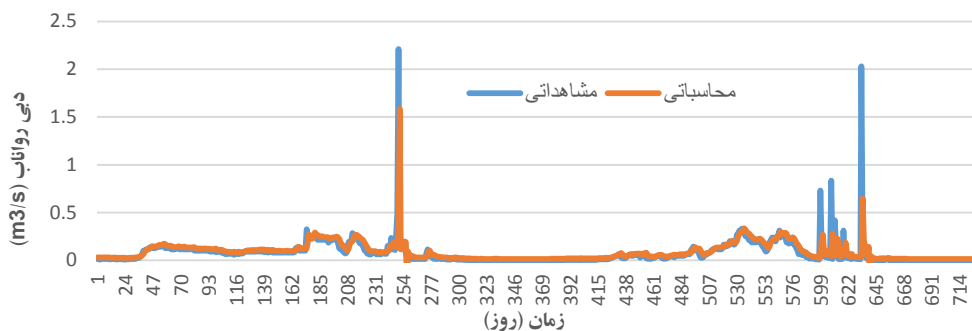
Table 6- The results of ANN model with the combination of PC, PMI and Total inputs

ایستگاه	ورودی	ساختار	آموزش			تست		
			E	RMSE	R <sup>2</sup>	E	RMSE	R <sup>2</sup>
تازه کند	PC	۱-۷-۱۹	۰/۵۹	۱/۴۰	۰/۵۹	۰/۶۹	۰/۴۲	۰/۶۹
رواسجان	PC	۱-۵-۱۹	۰/۸۵	۰/۳۳	۰/۸۵	۰/۲۰	۰/۱۲	۰/۱۳
اشدلق	Total	۱-۱۲-۱۹	۰/۷۱	۲/۴۵	۰/۷۰	۰/۴۶	۰/۳۵	۰/۴۰
برمیس	Total	۱-۲-۱۹	۰/۶۲	۰/۰۶	۰/۶۲	۰/۹۰	۰/۰۱	۰/۸۹
اورنگ	Total	۱-۳-۱۹	۰/۸۸	۰/۷۴	۰/۸۸	۰/۷۱	۰/۶۳	۰/۶۹
کاسین	Total	۱-۱-۱۹	۰/۵۷	۱/۶۶	۰/۵۷	۰/۷۱	۰/۵۱	۰/۷۱



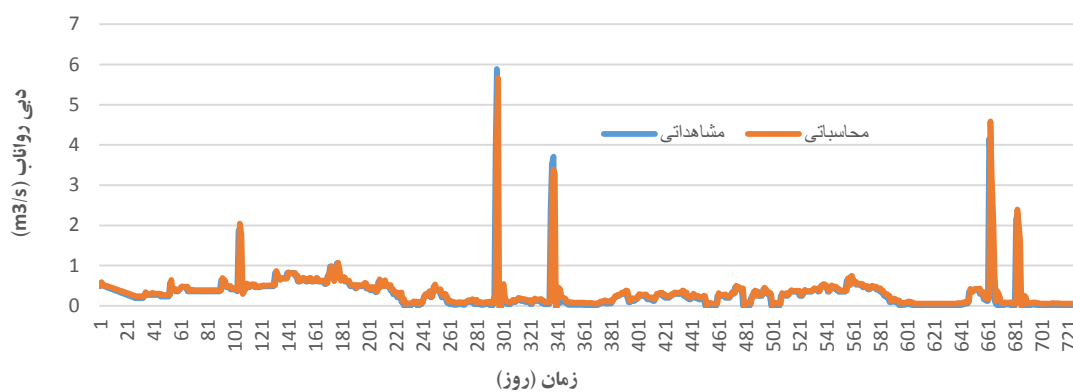
شکل ۱۵: نمودار سری زمانی دبی رواناب مشاهداتی و محاسباتی با مدل ANN در ایستگاه تازه کند برای دوره تست

Figure 15: The time series graph of observed and calculated runoff discharge with ANN model at Tazekand station for the test period



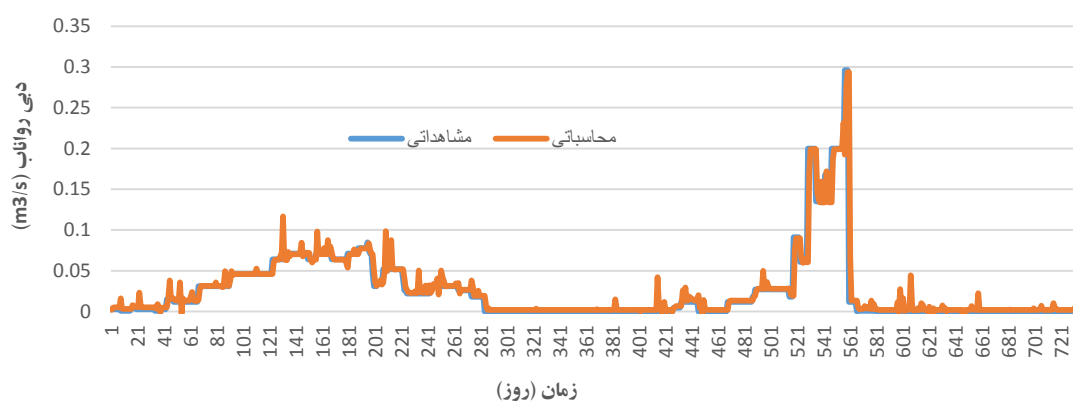
شکل ۱۶: نمودار سری زمانی دبی رواناب مشاهده‌ای و محاسباتی با مدل ANN در ایستگاه رواسجان برای دوره تست

Figure 16: The time series graph of observed and calculated runoff discharge with ANN model at Ravasjan station for the test period



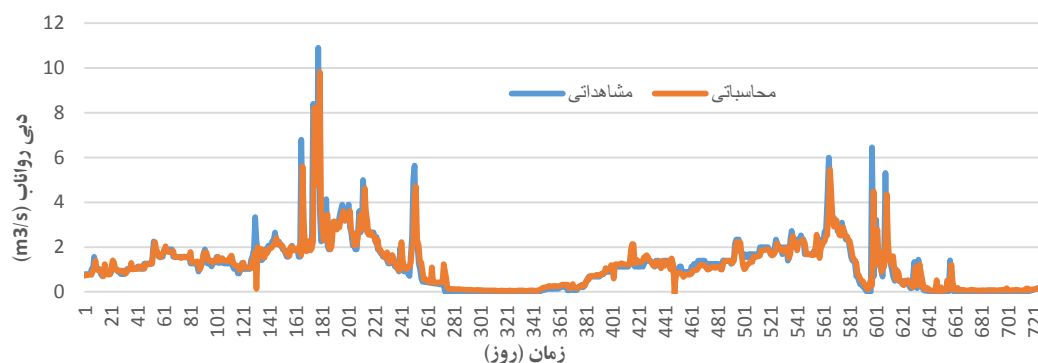
شکل ۱۷: نمودار سری زمانی دبی رواناب مشاهده‌ای و محاسباتی با مدل ANN در ایستگاه اشدلق برای دوره تست

Figure 17: The time series graph of observed and calculated runoff discharge with ANN model at Oushdilaq station for the test period



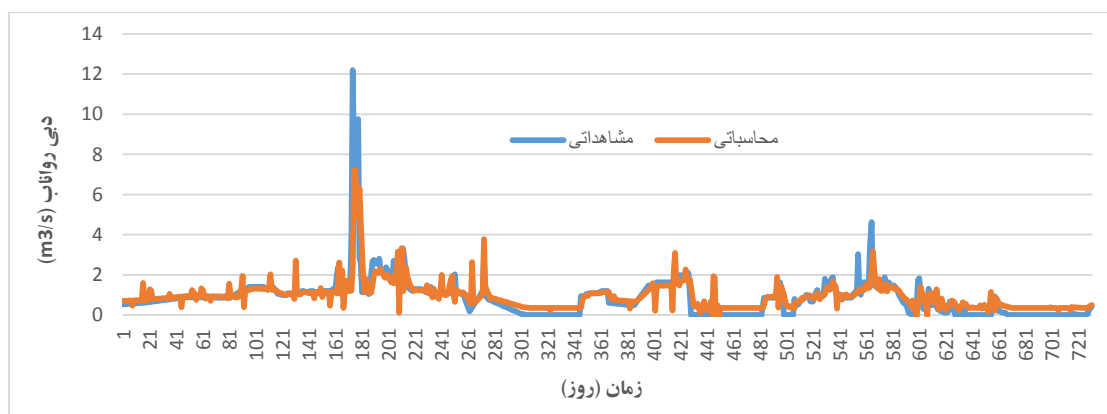
شکل ۱۸: نمودار سری زمانی دبی رواناب مشاهده‌ای و محاسباتی با مدل ANN در ایستگاه برمیس برای دوره تست

Figure 18: The time series graph of observed and calculated runoff discharge with ANN model at Bermis station for the test period



شکل ۱۹: نمودار سری زمانی دبی رواناب مشاهده‌ای و محاسباتی با مدل ANN در ایستگاه اورنگ برای دوره تست

Figure 19: The time series graph of observed and calculated runoff discharge with ANN model at Owrang station for the test period



شکل ۲۰: نمودار سری زمانی دبی رواناب مشاهده‌ای و محاسباتی با مدل ANN در ایستگاه کاسین برای دوره تست

Figure 20: The time series graph of observed and calculated runoff discharge with ANN model at Kasin station for the test period

### نتیجه گیری

در این پژوهش به منظور توسعه مدل‌های شبیه‌سازی بارش-رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری تازه‌کند، رواسجان، اشدلق، برمیس، اورنگ و کاسین با استفاده از مدل‌های ماشین بردار پشتیبان، برنامه ریزی بیان ژن و شبکه عصبی مصنوعی به تخمین داده‌های رواناب پرداخته و عملکرد مدل‌های مذکور مورد بررسی قرار گرفته‌است. جهت تعیین ترکیب ورودی مدل‌ها در ابتدا لیستی از متغیرهای مستقل مرتبط با رواناب هر ایستگاه تهیه شده و سپس به انتخاب ورودی‌های مناسب از میان لیست متغیرها پرداخته شده است. دو معیار همبستگی خطی پیرسون (PC) و اطلاعات متقابل جزئی (PMI) در انتخاب ورودی‌های مناسب بکار گرفته شده و ترکیب‌های ورودی بدست آمده با استفاده از هر معیار در شبیه‌سازی بارش-رواناب حوضه آبریز اهرچای در محل ایستگاه‌های هیدرومتری مورد بررسی قرار گرفته است. علاوه بر دو ترکیب مذکور، مدل‌سازی به کمک کلیه متغیرهای مستقل مرتبط نیز انجام شده است (Total). بدین ترتیب برای هر ایستگاه از سه ترکیب ورودی استفاده شده است. نتایج

حاصل از شبیه‌سازی بارش-رواناب شش ایستگاه هیدرومتری واقع در حوضه آبریز اهرچای حاکی از دقت کافی مدل‌های مورد نظر و برتری نسبی مدل شبکه عصبی مصنوعی می‌باشد. در تعیین متغیرهای ورودی نیز معیار PC دارای بهترین نتایج بوده و به طور کلی بهترین معیار تعیین متغیرهای ورودی می‌باشد. همچنین نتایج حاصل مطابقت مناسبی با نتایج پژوهش (Baba Ali and Dehghani (2019 دارد و در هر دو پژوهش، برتری روش شبکه‌های عصبی مصنوعی نسبت به سایر روش‌ها نتیجه‌گیری شده است که می‌توان گفت این برتری ناشی از ساختار و سرعت بالای پردازش شبکه‌های عصبی مصنوعی، توانایی یادگیری الگو و تعمیم آن پس از یادگیری و همچنین انعطاف‌پذیری این روش در مواجهه با مشکلات و خطاهای ناخواسته می‌باشد.

## References

- Abdollahzadeh, A., Ownegh, M., Sadoddin, A., & Mostafazadeh, R. (2016). "Constraints to residential land use development arising from flood and runoff coefficient in a land use planning framework, case study: Ziarat Watershed, Golestan Province". *Watershed Engineering and Management*, 8 (2), 221-235. <https://doi.org/10.22092/ijwmse.2016.106462>.
- Baba Ali, H. R., & Dehghani, R. (2019). "Evaluating the Performance of Wavelet Neural Network Models in Estimation of Daily Discharge". *Journal of Irrigation Sciences and Engineering*, 42 (3), 105-116. [In Persian]. <https://doi.org/10.22055/JISE.2017.22047.1580>.
- Ghorbani, M. A., Dinpazhoh, Y., & Moayeri, M. M. (2019). "Uncertainty reduction of conceptual rainfall-runoff models using Joint ensemble Kalman filter (Case study: Behesht-Abad catchment)". *Iranian Water Research Journal*, 12 (4), 123-130. [In Persian]. Retrieved from [https://iwrij.sku.ac.ir/article\\_10618.html?lang=en](https://iwrij.sku.ac.ir/article_10618.html?lang=en).
- Lee, W. K., & Resdi, T. A. T. (2016). "Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman Catchment, Terengganu, Malaysia". *Hydrological Sciences Journal*, 61 (2), 2930-2945. <https://doi.org/10.1080/02626667.2016.1174333>.
- Liu, S., & Shi, H. (2019). "A recursive approach to long-term prediction of monthly precipitation using genetic programming". *Water Resources Management*, 33 (3), 1103-1121. <https://doi.org/10.1007/s11269-018-2169-0>.
- Mohammadi, F., Fakherifard, A., Ghorbani, M. A., Dinpazhoh, Y., & Shahmorad, S. (2018). "Development of Cross wavelet- kalman filter isochrones lines model to analyze compound rainfall-runoff events". *Journal of Water and Irrigation Management*, 8 (1), 39-53. [In Persian]. <https://doi.org/10.22059/jwim.2018.249040.583>.
- Mohammadi, F., Fakherifard, A., Ghorbani, M. A., Dinpazhoh, Y., & Shahmorad, S. (2019). "Simulation of rainfall - runoff events by applying phase differences diagrams and correcting effective rainfall components". *Journal of Watershed Management Research*, 20, 97-108. [In Persian]. <http://doi.org/10.29252/jwimr.10.20.97>.
- Mostafazadeh, R., Sadeghi, S. H. R., & Sadoddin, A. (2015). "Modelling effects of land use type and spatial pattern on flow hydrograph variations". *Iranian Journal of Watershed Management Science and Engineering*, 31, 51-58. [In Persian]. Retrieved from [https://jwmsei.ir/browse.php?sid=1&slc\\_lang=fa&mag\\_limit\\_start=5](https://jwmsei.ir/browse.php?sid=1&slc_lang=fa&mag_limit_start=5).
- Nikpour, M. R., & Abdollahi Asadabadi, S. (2020). "Monthly precipitation prediction of Ardabil province using ANN and WANN models". *Journal of Irrigation and Water Engineering*, 10 (4), 241-257. [In Persian]. <https://doi.org/10.22125/iwe.2020.110094>.
- Nikpour, M. R., Sanikhani, H., Mahmodi Babelan, S., & Nastarani Amuqin, S. (2019). "Daily rainfall – runoff modeling of Darreh-Rud river in Ardabil province, Iran". *Journal of Irrigation Sciences and Engineering*, 41 (4), 133-146. [In Persian]. <https://doi.org/10.22055/jise.2017.21053.1508>.
- Nourani, V., & Komasi, M. (2013). "A geomorphology-based ANFIS model for multistation modeling of rainfall–runoff process". *Journal of Hydrology*, 490 (9), 41-55. <https://doi.org/10.1016/j.jhydrol.2013.03.024>.
- Nourani, V., Sattari, M. T., & Molajou, A. (2017). "A hybrid decision tree association rules approach for long-term precipitation forecasting". *Journal of Water and Irrigation Management*, 6 (2), 331-346. [In Persian]. <https://doi.org/10.22059/jwim.2017.63786>.

- Sadeghi, S. H. R., & Mostafazadeh, R. (2016). "Triple diagram models for changeability evaluation of precipitation and flow discharge for suspended sediment load in different time scales". *Environmental Earth Sciences*, 75 (9), 843-856. <https://doi.org/10.1007/s12665-016-5621-6>.
- Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. (2020). "Rainfall forecasting through ANN and SVM in Bolangir Watershed, India". *Smart Intelligent Computing and Applications*. 159 (5), 767-774. [https://doi.org/10.1007/978-981-13-9282-5\\_74](https://doi.org/10.1007/978-981-13-9282-5_74).
- Shafeizadeh, M., Fathian, H., & Nikbakht Shahbazi, A.R. (2019). "Continuous rainfall-runoff simulation by artificial neural networks based on selection of effective input variables using partial mutual information (PMI) algorithm". *Iranian Water Research Journal*, 15 (2), 144-161.[In Persian]. <https://doi.org/20.1001.1.17352347.1398.15.2.12.1>.
- Sofian, I. M., Afandi, A. K., Iskandar, I., & Apriani, Y. (2018). "Monthly rainfall prediction based on artificial neural networks with backpropagation and radial basis function". *International Journal of Advances in Intelligent Informatics*, 4 (2), 154-166. <https://doi.org/10.26555/ijain.v4i2.208>.
- Tetzlaff, D., Soulsby, C., & Birkel, C. (2010). "Hydrological connectivity and microbiological fluxes in montane catchments: The role of seasonality and climatic variability". *Hydrological Processes*, 24 (2), 1231-1239. <https://doi.org/10.1002/hyp.7680>.
- Yadav, A. K., Chandola, V. K., Singh, A., & Singh, B. P. (2020). "Rainfall-runoff modelling using artificial neural networks (ANN) model". *International Journal of Current Microbiology and Applied Sciences*, 9 (3), 127-135. <https://doi.org/10.20546/ijcmas.2020.903.016>.
- Yaseen, Z. M., Ebtehaj, L., Kim, S., Sanikhani, H., Asadi, H., Ghareb, M. L., Bonakdari, H., Mohtar, W. H. M., Al-Ansari, N., & Shahid, S. (2019). "Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis". *Water (MDPI)*, 502 (11), 1-23. <https://doi.org/10.3390/w11030502>.

## Simulation of the Rainfall-runoff Process in the Drainage Basin of Aharchay Using data Mining Techniques

Ali Ebrahimzadeh<sup>1</sup>, Bita Bagheri Seygalani<sup>2</sup>, Vahid Nourani<sup>3</sup>

1- Assistant Professor, Department of Civil Engineering, Islamic Azad University of Marand, Marand, Iran

2- Assistant Professor, Department of Architecture, Islamic Azad University of Miyaneh, Miyaneh, Iran

3- Professor, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

### Abstract

#### Introduction

Rainfall-runoff simulation is an effective step in the management of water resources, watersheds, and floodwater control deficiency crises. The inherent complexity of the rainfall-runoff process, temporal and spatial variations, and influential factors make it difficult to perform simulations using physical or hydrologic models. Thus, it is necessary to select a model to predict the runoff of rainfall at an acceptable accuracy by using influential factors. To obtain the overall rainfall-runoff pattern, it is crucial to incorporate the influential watershed data. In recent years, metaheuristic algorithms, e.g., support vector machines (SVMs), gene expression programming (GEP), and artificial neural networks (ANNs) have been increasingly employed in hydrologic studies, particularly in the investigation of phenomena with no certain relationships. Such models are adopted in a wide range of hydrologic applications. The present study evaluates the SVM, GEP, and ANN approaches in the simulation of the rainfall-runoff process for Aharchay Watershed, Iran, at the hydrometric stations of Tazekand, Ravasjan, Oushdilaq, Bermis, Owrang, and Kasin.

#### Materials and methods

Aharchay Watershed lies in Northwest Iran and covers a significant portion of Aras Basin. It plays a key role in the development of the region as it supplies drinking, agricultural, and industrial water demands. Aharchay Watershed is a major sub-basin of the Aras Basin, and its main river is the Aharchay River. The Aharchay River joins the Gharesou River in the outlet of the watershed to form the Darrehrood River. This study develops rainfall-runoff simulation models for Aharchay Watershed at the hydrometric stations of Tazekand, Ravasjan, Oushdilaq, Bermis, Owrang, and Kasin by using the SVM, GEP, and ANN approaches, estimating runoff data. These three approaches are evaluated in terms of performance. To determine a combination of inputs for the models, a list of runoff-related independent variables of each station was created. Then, suitable inputs were selected from the listed variables by using the Pearson correlation (PC) and partial mutual information (PMI). The input combinations obtained from these two criteria were evaluated for rainfall-runoff simulation at each hydrometric station. Also, modeling was performed using the related independent variables (total efficiency). Therefore, three input combinations were employed for each station. To evaluate model performance, the coefficient of determination ( $R^2$ ), root-mean-square error (RMSE), and the Nash-Sutcliffe coefficient (E) were utilized.

#### Discussion and conclusion

This study employed the SVM, GEP, and ANN models to simulate the rainfall-runoff process of Aharchay Watershed, Iran, at the hydrometric stations of Tazekand, Ravasjan, Oushdilaq, Bermis, Owrang, and Kasin. Furthermore, the PCI, PMI, and total criteria were employed to identify suitable combinations of inputs. The accuracy of the models was examined using the coefficient of determination ( $R^2$ ), root-mean-square-error (RMSE), and the Nash-Sutcliffe coefficient (E) to identify the best model. The results revealed that the proposed models were acceptable accurate in watershed runoff simulation. The ANN model outperformed the two other models at the Tazekand, Oushdilaq, and Kasin Station. Also, the GEP model had better performance at the Ravasjan and Barmis Stations, while the SVM model yielded better results at the Owrang Station. Overall, it is concluded that the ANN approach is either the best model or almost the best one in simulating the runoff of Aharchay Watershed. Thus, it can be employed for the runoff simulation of Aharchay Watershed at acceptable accuracy.

**Keywords:** Artificial neural network, Gene expression programming, Hydrometric station, Support vector machine