علوم و مهندسی آبیاری
جلد ۴۰، شماره‌ی٤، زمستان ۹٦، ص. ۱۰۷-۹۱

بررسی عملکرد مدل‌های سری زمانی و مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن خودهمبسته در شبیه‌سازی رواناب ماهانه (مطالعه موردی: حوضه خرخره‌چای)

محمد عیسی‌زاده[1*]، حجت احمدزاده[2]، محمدعلی قربانی[3] و محمدحسن فاضلی‌فرد[4]

۱*- نویسنده مسئول: دانش آموخته کارشناسی ارشد مهندسی منابع آب، دانشکده کشاورزی، دانشگاه تبریز.mohammadisazade@gmail.com
۲- دانشجوی دکتری مهندسی منابع آب، دانشکده کشاورزی، دانشگاه تبریز.
۳- دانشیار گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز.
۴- دانشجوی دکتری مهندسی منابع آب دانشگاه تبریز، تبریز، ایران.

تاریخ دریافت: ۹۴/۱۲/۱ تاریخ پذیرش: ۹۵/۴/۱۴

چکیده

افزایش دقت تخمین رواناب در حوضه‌های فاقد داده‌های هواشناسی نقش مهمی در مدیریت صحیح منابع آب این حوضه‌ها دارد. در این راستا شبیه‌سازی خود همبسته می‌تواند مفید واقع گردد. در مطالعه حاضر، کارایی سه مدل داده‌کاوی شامل شبکه عصبی مصنوعی، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان در کنار مدل‌های سری زمانی برای پیش‌بینی رواناب ماهانه در حوضه خرخره‌چای مورد ارزیابی قرار گرفت. برای این منظور، در ابتدا ترکیب‌های مختلف ورودی به مدل‌ها که باید در تخمین جریان رودخانه در نظر گرفته شوند، تعیین شدند. همچنین تأثیر ساختار مختلف مدل‌ها بر شبیه‌سازی جریان از طریق به کار گیری آنها و مقایسه نتایج به‌دست آمده، مورد بررسی قرار گرفت. در این راستا، مدل‌های خطی خود همبسته، خود همبسته با میانگین متحرک و خود همبسته میانگین متحرک یکپارچه، شبکه عصبی پرسپترون چند لایه، برنامه‌ریزی بیان ژن با چهار و سیزده تابع ریاضی و مدل ماشین بردار پشتیبان با سه تابع کرنل به منظور شبیه‌سازی خودهمبسته جریان ماهانه (طی دوره ۹۰- ۱۳٦۷) به کار گرفته شدند. نتایج نشان داد مدل پرسپترون چند لایه (۵،۳) با مقادیر ضریب همبستگی و جذر میانگین مربعات خطا به‌ترتیب برابر با ۰/۸٤ و ۸/۲۱ متر مکعب بر ثانیه در دوره واسنجی و مقادیر ۰/۸٦ و ۵/٦٦ متر مکعب بر ثانیه در دوره صحت‌سنجی بیشترین دقت را در شبیه‌سازی رواناب ماهانه نسبت به سایر مدل‌ها داشتند.

کلید واژه‌ها: شبیه‌سازی خودهمبسته رواناب ، مدل‌های سری زمانی، پرسپترون چند لایه، حوضه خرخره‌چای.

مقدمه

در حال حاضر، مدل‌های مختلفی در شبیه‌سازی رواناب حوضه‌ها مورد استفاده قرار می‌گیرند که کارایی و همچنین موارد استفاده از آنها براساس ساختار حوضه، داده‌های در دسترس و مورد نیاز و همچنین هدف شبیه‌سازی از همدیگر متفاوت است. استفاده از مدل‌هایی که مبنای فیزیکی داشته و به متغیرهای هواشناسی و هیدرولوژیکی بیشتری نیاز دارند، برای شبیه‌سازی رواناب در حوضه‌هایی که دارای آمار و اطلاعات کافی نبوده و یا فقط دارای داده‌های جریان رودخانه هستند، مقدور نیست. شبیه‌سازی خودهمبسته جریان رودخانه که در آن داده‌های تأخیر یافته جریان رودخانه تنها داده‌های ورودی به مدل را تشکیل می‌دهند از رویکردهای راه‌گشا در این گونه حوضه‌ها است. مدل‌های سری زمانی به عنوان یکی از مدل‌های آماری از مهم‌ترین و معمول‌ترین ابزارهای مورد استفاده برای این منظور هستند (باکس و همکاران[1] ، ۱۹۹۴). علاوه بر این مدل‌ها، در دهه‌های اخیر مدل‌های داده‌کاوی کاربرد گسترده‌ای در زمینه شبیه‌سازی و پیش‌بینی جریان رودخانه‌ها پیدا کرده‌اند (سام‌سودین و همکاران[2]، ۲۰۱۱). این مدل‌های متکی به داده، ساختار ریاضی مشخصی داشته و بدون اینکه فرآیند فیزیکی حاکم بر پدیده را شبیه‌سازی نمایند، توانایی تشخیص روابط پیچیده غیرخطی بین داده‌های ورودی و خروجی را دارا هستند (نایاک و همکاران[3]، ۲۰۰۴). ویژگی‌های مذکور این مدل‌ها قابلیت استفاده از آنها را برای شبیه‌سازی خودهمبسته رواناب ممکن ساخته است.

مدل سری زمانی از دو بخش اصلی، شامل مؤلفه جبری و مؤلفه تصادفی تشکیل شده است که مؤلفه جبری مدل با استفاده از ارقام مشاهداتی و بخش تصادفی با استفاده از روش‌های مختلف استوکاستیک به دست می‌آید. بنابراین ساختار مدل‌های سری زمانی می‌تواند با ساختار سری‌های هیدرولوژیکی در صورت انتخاب درست مدل و محاسبات صحیح آن سازگاری و مطابقت ویژه‌ای داشته باشد (سالاس[4]، ۱۹۹۳). شبکه‌های عصبی مصنوعی به عنوان یکی از روش‌های جعبه سیاه این قابلیت را دارند که روابط منطقی بین پارامترهای ورودی و خروجی یک پدیده را بیان کنند. این شبکه‌ها بر مبنای مقایسه‌ی بین خروجی شبکه و هدف، تعدیل می‌شوند و تا زمانی که خروجی شبکه با خروجی هدف تطبیق پیدا کند، این کار ادامه می‌یابد (کوچک‌زاده و بهمنی[5]، ۲۰۰۶).

1- Box *et al.*
2- Samsudin *et al.*
3- Nayak *et al.*
4- Salas
5- Kouchakzadeh and Bahmani

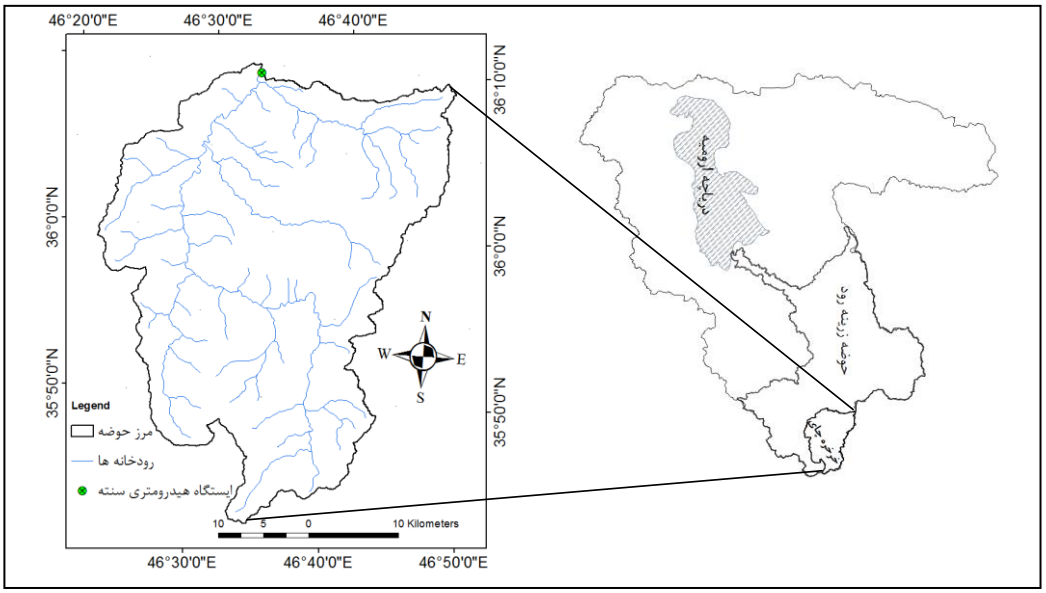
مدل برنامه‌ریزی بیان ژن اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای اندازه‌گیری و مقایسه روش‌های مختلف حل، در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، روش حل مناسب را ارائه می‌نمایند (بوریلیو همکاران[1] ، ۲۰۰۶).

ماشین‌های بردار پشتیبان، سیستم‌های یادگیری خاصی هستند که از یک فضای فرضیه به نام فضای مشخصه[2]، شامل توابع خطی با بعد زیاد استفاده کرده و می‌توانند برای پیش‌بینی (دسته‌بندی و رگرسیون) مورد استفاده قرار گیرند. این سیستم‌ها توسط الگوریتم یادگیری که بر اساس تئوری بهینه‌سازی استوار است، واسنجی داده می‌شوند (وپنیک و کورتیس[3]، ۱۹۹۵).

تاکنون تحقیقات زیادی در مورد ارزیابی و همچنین مقایسه عملکرد هر یک از این مدل‌ها در شبیه‌سازی متغیرهای هیدرولوژیکی انجام گرفته است. جین و کومار[4](۲۰۰۷) کارآیی مدل‌های شبکه عصبی مصنوعی و مدل‌های سری زمانی را در پیش‌بینی مقادیر جریان ماهانه رودخانه کلرادو تحت تأثیر سه سناریو مورد بررسی قرار دادند که در هر سه سناریو دقت مدل شبکه عصبی بهتر از مدل خودهمبسته بوده است. در تحقیقی دیگر، کیشی و همکاران[5] (۲۰۱۲) دقت مدل‌های نروفازی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان را در پیش‌بینی رواناب روزانه در دو ایستگاه آب‌سنجی واقع در شمال غرب ترکیه مورد مقایسه قرار دادند. نتایج تحقیق آنان نشان داد که مدل‌های شبکه عصبی مصنوعی و نروفازی به‌ترتیب بالاترین دقت را در پیش‌بینی رواناب روزانه در ایستگاه‌های اول و دوم دارا هستند. همچنین کالته[6] همکاران (۲۰۱۳) ترکیب مدل‌های رگرسیون بردار پشتیبان (SVR[7]) و شبکه عصبی مصنوعی با تبدیل موجک را برای پیش‌بینی رواناب ماهانه به‌کار گرفتند و به این نتیجه رسیدند که هر دو مدل رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی در ترکیب با تبدیل موجک کارایی بهتری نسبت به مدل‌های رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی تنها دارند. ترزی و ایرجین[8] (۲۰۱۴) مدل‌های خود همبسته، برنامه‌ریزی بیان ژن، شبکه عصبی پایه شعاعی[9]، شبکه عصبی پیشرو[10] و نروفازی را برای پیش‌بینی رواناب ماهانه رودخانه کیزیلیرماک[11] در ترکیه مورد استفاده قرار دادند. نتایج نشان داد که مدل خود همبسته (۲) از بین مدل‌های به‌کار گرفته شده در تحقیق عملکرد بهتری را در پیش‌بینی جریان داشت. در تحقیقی دیگر آوچی[12] (۲۰۱۴) دقت مدل‌های شبکه عصبی پیشرو، شبکه عصبی پایه شعاعی، شبکه‌های عصبی رگرسیونی و رگرسیون چندگانه را در پیش‌بینی جریان ماهانه رودخانه‌های زاب بالادست و پایین دست در شمال عراق مورد مقایسه قرار دادند و به این نتیجه رسیدند که مدل‌های شبکه عصبی مصنوعی عملکرد بهتری از مدل رگرسیون چندگانه در پیش‌بینی جریان دارند. آدامویسکی[13] (۲۰۱۳) مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی را برای شبیه‌سازی بارش–رواناب در یک حوضه کوهستانی با داده‌های محدود در یوتارانچال هندوستان به کار گرفت. نتایج حاکی از توانایی مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی در پیش‌بینی رواناب مستقیم، جریان پایه و جریان کل این منطقه کوهستانی و پر از تپه بود. کاکائی لفدانی و همکاران[14] (۲۰۱۳) سه مدل ماشین بردار پشتیبان و نرو فازی[15] و مدل هیدرولوژیکی مفهومی مایک نسخه ۱۱ را در شبیه‌سازی جریان روزانه حوضه اسکندری به کار گرفتند. نتایج نشان داد که هر سه مدل دارای عملکرد مناسب در منطقه مورد مطالعه هستند. در نهایت مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی بالاترین ضریب همبستگی و کمترین میانگین مربعات خطا را در شبیه‌سازی جریان رودخانه نسبت به دو مدل دیگر از خود نشان داد.

انتخاب مدلی که میزان جریان رودخانه را با دقت بالا و خطای کمتر شبیه‌سازی نماید، می‌تواند کمک شایانی در مدیریت صحیح طرح‌های منابع آبی و افزایش عملکرد این طرح‌ها داشته باشد. با توجه به آنچه آمد، تحقیق حاضر سعی دارد تا کارایی مدل‌های شبکه عصبی مصنوعی، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان را در کنار مدل‌های سری زمانی در شبیه‌سازی خودهمبسته رواناب ماهانه حوضه رودخانه خرخره‌چای مورد ارزیابی قرار دهد. با توجه به اینکه ترکیب ورودی‌ها و نوع ساختار به کار گرفته شده برای هر یک از این مدل‌ها تأثیر چشمگیری در نتایج شبیه‌سازی و عملکرد آنها دارد؛ بنابراین در این تحقیق تلاش شد کارایی مدل‌های مذکور به ازای تأخیرهای یک تا پانزده ماه رواناب و همچنین انواع ساختارهای مختلف مدل‌ها مورد ارزیابی قرار گیرد. برای این منظور مدل‌های خطی خود همبسته، خود همبسته با میانگین متحرک و خود همبسته میانگین متحرک یکپارچه (به عنوان مدل‌های سری زمانی)، مدل شبکه عصبی پرسپترون چند لایه[16]، مدل برنامه‌ریزی بیان ژن با چهار و سیزده تابع ریاضی و مدل ماشین بردار پشتیبان با سه تابع کرنل پایه شعاعی[17]، چند جمله‌ای درجه d و کرنل خطی به ازای تأخیرهای یک تا پانزده ماهه جریان برای شبیه‌سازی ماهانه رواناب به‌کار گرفته شدند.

1- Borelli *et al.*
2- Feature Space
3- Vapnik and Cortes
4- Jain and Kumar
5- Kişi *et al.*
6- Kalteh
7- Support Vector Regression
8- Terzi and Ergin
9- Radial Basis Function Network
10- Feed-Forward Neural Networks
11- Kizilirmak
12- Awchi
1- Adamowski
14- Kakaei Lafadani *et al.*
15- Adaptive Neuro-Fuzzy Inference System
16- Multi Layer Perceptron
17 -Radial Basis Function



شکل ۱- موقعیت حوضه خرخره‌چای و ایستگاه آب‌سنجی سنته

منطقه مورد مطالعه و داده‌ها

محدوده مورد مطالعه در این تحقیق، حوضه رودخانه خرخره‌چای از شاخه‌های اصلی حوضه زرینه‌رود است که در طول جغرافیایی '۲۲ °۴۶ تا '۴۸ °۴۶ شرقی و عرض جغرافیایی '۴۲ °۳۵ تا '۱۰ °۳۶ شمالی در استان آذربایجان‌غربی قرار دارد (شکل ۱). این حوضه سهم قابل توجهی در حجم آب ورودی به سد بوکان و در نتیجه تأمین نیازهای آبی پایین دست این سد دارد. مساحت این حوضه در محل ایستگاه آب‌سنجی سنته بالغ بر ۱۲۳۲ کیلومتر مربع بوده و متوسط ارتفاع آن از سطح دریا ۲۰۳۵ متر است. همچنین متوسط بارش سالانه این حوضه در حدود ۴۰۲ میلی‌متر است.

داده‌های مورد استفاده شامل جریان ماهانه مشاهداتی در ایستگاه آب‌سنجی سنته طی دوره ۱۳۶۷ الی ۱۳۹۰ است. به طوری‌که داده‌های دوره آماری (بهمن ۱۳۶۷- بهمن ۱۳۸۴) به عنوان داده‌های واسنجی و دوره (اسفند ۱۳۸۴- شهریور ۱۳۹۰) به عنوان داده‌های صحت‌سنجی هر یک از مدل‌ها انتخاب گردیدند. به عبارت دیگر ۷۵ درصد از داده‌های مشاهداتی برای واسنجی و ۲۵ درصد دیگر برای صحت‌سنجی مدل‌ها در نظر گرفته شدند.

روش تحقیق

نرمال‌سازی داده‌ها

لازمه استفاده از مدل‌های سری زمانی تبعیت نمودن داده‌های مورد استفاده از توزیع نرمال است (عادلی و همکاران، ۱۳۹۳). به علت تبعیت نکردن داده‌های دبی مشاهداتی ماهانه ایستگاه سنته از توزیع نرمال، از روش‌های متعددی همچون باکس- کاکس[1]، تبدیل جانستون[2] و توابع لگاریتمی، نمایی، توانی، رادیکالی و معکوس هر یک از آنها برای نرمال‌سازی داده‌های جریان استفاده گردید. با به کارگیری هر یک از این روش‌ها بر روی داده‌های جریان و انجام آزمون‌های کلموگروف- اسمیرنوف و شاپیرو-ویلک[3] (شاپیرو و ویلک، ۱۹۶۵) در سطح اطمینان ۰/۹۵ روی آنها، این نتیجه حاصل شد که هیچ یک از روش‌های مذکور در نرمال‌سازی داده‌های جریان کارساز نیستند. در نهایت، در تحقیق حاضر از روابط یک تا سه که حاصل تجربه نویسندگان مقاله است، برای نرمال‌سازی داده‌ها استفاده گردید. در این روابط n: تعداد سال‌ها، Q_{ij}: میزان جریان مشاهداتی در ماه jام از سال iام، $\overline{Q}_j$: میانگین دراز مدت جریان در ماه jام، SD_W: انحراف معیار متغیر W_{ij} و K_{ij}: جریان نرمال شده در ماه jام از سال iام هستند. بعد از نرمال‌سازی داده‌ها، شاخص‌های آماری میانگین، مقادیر حداقل و حداکثر، انحراف معیار، چولگی و کشیدگی محاسبه گردیده و آزمون‌های کلموگروف- اسمیرنوف و شاپیرو- ویلک در سطح اطمینان ۰/۹۵ برای داده‌های مشاهداتی اولیه و نرمال شده انجام گرفت که نتایج آنها در جدول (۲) آورده شده است:

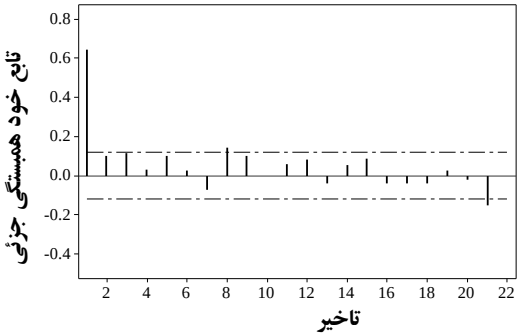
$$P_{ij} = Ln(Q_{ij} + 0.052) \qquad \begin{matrix} i = 1:n \\ j = 1:12 \end{matrix} \qquad (۱)$$

$$W_{ij} = P_{ij} - \overline{Q}_j \qquad (۲)$$

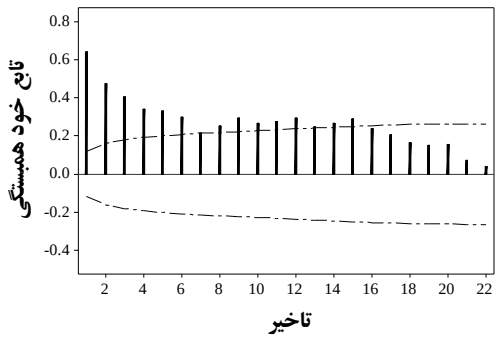
$$K_{ij} = \frac{W_{ij}}{SD_w} \qquad (۳)$$

1- Box-Cox
2- Johnson-Transformation

3- Shapiro and Wilk



شکل ۲- توابع الف) خودهمبستگی



ب) خودهمبستگی جزئی داده‌های جریان ماهانه نرمال ایستگاه آب‌سنجی سنته

تعیین تعداد ورودی مدل‌ها

توابع خودهمبستگی[1] و خودهمبستگی جزئی[2] از جمله معیارهای مطرح در تعیین تعداد ورودی‌های موثر در شبیه‌سازی متغیر مورد نظر هستند (یانگ و همکاران[3]، ۲۰۰۶؛ جیان و همکاران[4]، ۲۰۰۶). در تحقیق حاضر نیز از این توابع برای تعیین تعداد تأخیرهای لازم در سری جریان ماهانه که باید در ارزیابی کارایی مدل‌های ماشین بردار پشتیبان، شبکه عصبی مصنوعی، برنامه‌ریزی بیان ژن و سری‌های زمانی مورد بررسی قرار گیرند، استفاده گردید. با استفاده از توابع مذکور می‌توان میزان همبستگی جریان ماهانه با تأخیرهای ماهانه مختلف را مشخص نموده و از این طریق به‌طور تقریبی تعداد و درجه تأخیرها که بر جریان تاثیر معنی‌دار دارند را تعیین کرد (یانگ و همکاران، ۲۰۰۶؛ جیان و همکاران، ۲۰۰۶). بنابراین در این تحقیق، با استفاده از نمودارهای خودهمبستگی و خودهمبستگی جزئی (شکل ۲) حداکثر تعداد تأخیر ممکن و معنی‌دار برای شبیه‌سازی جریان ایستگاه آب‌سنجی سنته بدست آمده است.

پس از بررسی تابع‌های خودهمبستگی و خودهمبستگی جزئی مربوط به داده‌های جریان ماهانه نرمال شده معلوم گردید که براساس نمودار خودهمبستگی جزئی، اولین جریان تأخیر یافته (درجه تأخیر برابر با یک) بیشترین تاثیر را در برآورد جریان دارد و همچنین بر اساس نمودار خودهمبستگی، جریان با تأخیرهای یک تا پانزده ماهه خارج از باند استقلال قرار گرفته‌اند. پس می‌توان نتیجه گرفت که حداکثر تعداد تأخیرهای جریان ماهانه لازم برای لحاظ نمودن در شبیه‌سازی جریان (تعداد متغیرهای ورودی لازم به مدل‌ها) برابر با پانزده ماه است. بنابراین، در این تحقیق از جریان به ازای یک تا پانزده تاخیر (پانزده ترکیب مختلف) به عنوان ورودی مدل‌ها برای پیش‌بینی جریان ماهانه ایستگاه سنته استفاده شده است.

مدل ماشین بردار پشتیبان

اولین کاربرد این روش در مسائل آب توسط دیباک و همکاران[5] (۲۰۰۱) با شبیه‌سازی بارش- رواناب ارائه شد. ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه-سازی مقید است که از اصل استقرای کمینه سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. در مدل رگرسیون بردار پشتیبان تابعی مرتبط با متغیر وابسته y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند f(x) به علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود:

$$f(x) = W^T.\phi(x) + b \qquad (۴)$$

$$y = f(x) + noise \qquad (۵)$$

چنانچه W: بردار ضرایب، b: ثابت مشخصه‌های تابع رگرسیونی و $\phi(x)$: نیز تابع کرنل باشد، آنگاه هدف پیدا کردن شکل تابعی برای f(x) است. این مهم با واسنجی مدل ماشین بردار پشتیبان توسط مجموعه‌ای از نمونه‌ها (مجموعه واسنجی) محقق می‌شود. برای محاسبه W و b لازم است تابع خطا (رابطه ۶) در مدل ماشین بردار پشتیبان رگرسیونی با در نظر گرفتن شرایط مندرج (قیود) در رابطه (۷) بهینه شود:

$$W^T.\phi(x_i) + b - y_i \le \varepsilon + \varepsilon_i^* \qquad (۶)$$

1- Autocorrelation Function
2- Partial Autocorrelation Function
3- Yang *et al.*
4- Jian *et al.*
5- Dibike *et al.*

$$\begin{aligned}&\frac{1}{2}W^T w + C\sum_{i=1}^{N}\varepsilon_i + C\sum_{i=1}^{N}\varepsilon_i^* y_i \\ &-W^T.\phi(x_i) - b \le \varepsilon + \varepsilon_i \\ &\varepsilon_i, \varepsilon_i^* \ge 0, i = 1,2,...,N\end{aligned} \qquad (۷)$$

در روابط بالا، C: عددی صحیح و مثبت است که عامل تعیین جریمه در هنگام رخ دادن خطای واسنجی مدل است، $\phi(x)$: تابع کرنل، N: تعداد نمونه‌ها و دو مشخصه ε_i^* و ε_i: متغیرهای کمبود هستند. در نهایت، تابع رگرسیون بردار پشتیبان را می‌توان به شکل زیر باز نویسی کرد:

$$f(x) = \sum_{i=1}^{N}\overline{\alpha_i}.\phi(x_i)^T.\phi(x) + b \qquad (۸)$$

در رابطه (۸)، $\overline{\alpha_i}$: میانگین ضرایب لاگرانژ است. محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل رگرسیون بردار پشتیبان انتخاب یک تابع کرنل است.

ماشین بردار پشتیبان معمولا با سه تابع کرنل پایه شعاعی، چند جمله‌ای درجه d و خطی مورد استفاده قرار می‌گیرد که روابط مورد استفاده در هر کدام از آنها به ترتیب در زیر آورده شده است (کاوزوگلو و کولکیسین[1]، ۲۰۰۹؛ گواو همکاران[2] ۲۰۰۸).

$$k(x, x_i) = \exp\left(-\|x - x_i\|^2 / \sigma^2\right) \qquad (۹)$$

$$k(x, x_i) = \left(t + \langle x, x_i \rangle\right)^d \qquad (۱۰)$$

$$k(x, x_i) = x.x_i \qquad (۱۱)$$

لازم به ذکر است فرآیند محاسبات ماشین بردار پشتیبان، بر اساس کدنویسی در محیط متلب انجام و پارامترهای توابع کرنل از طریق سعی و خطا بهینه شدند. از آنجا که در تحقیق حاضر از داده-های تأخیر یافته رواناب به عنوان ورودی مدل ماشین بردار پشتیبان در شبیه‌سازی جریان استفاده شده است؛ بنابراین به این مدل عنوان مدل ماشین بردار پشتیبان خودهمبسته داده شده است.

مدل شبکه عصبی مصنوعی با آموزش لونبرگ- مارکوارت

می‌توان گفت که حدود ۹۰ درصد از مدل‌های شبکه عصبی که در مسائل هیدرولوژی به‌کار گرفته شده‌اند از نوع پیشرو هستند (کولیبا و همکاران[3] ، ۲۰۰۰) که مدل شبکه عصبی پرسپترون چند لایه با الگوریتم آموزش لونبرگ–مارکوارت[4] از مهم‌ترین آنها به‌شمار می‌رود.

در شبکه‌های پیشرو، اتصالات میان گره‌هایی وجود دارد که در لایه‌های متفاوت قرار دارند. همچنین ورودی‌ها در لایه ورودی شبکه نمایش داده شده و عمل تحریک از ورودی به سمت خروجی شبکه شروع می‌شود (بی‌نام، ۲۰۰۰). در این تحقیق، به علت آنکه از داده‌های تأخیر یافته رواناب به عنوان ورودی مدل شبکه عصبی پرسپترون چند لایه در شبیه‌سازی رواناب استفاده شده است، بنابراین به این مدل عنوان مدل شبکه عصبی مصنوعی خودهمبسته داده شده است (بنی‌حبیب و ولی پور، ۱۳۸۸). لازم به ذکر است که در تحقیق حاضر از تعداد نرون‌های یک تا بیست برای لایه پنهان شبکه استفاده گردید تا از این طریق بتوان تاثیر ساختار مختلف شبکه بر عملکرد آن در شبیه‌سازی رواناب را مورد ارزیابی قرار داد، به این ترتیب ۳۰۰ نوع مدل شبکه عصبی مصنوعی خودهمبسته ایجاد شده و عملکرد هر کدام از آنها مورد ارزیابی قرار گرفت. معماری معمول این شبکه متشکل از سه بخش لایه ورودی، پنهان و خروجی است. از آنجا که متداول‌ترین نوع توابع محرک تابع تانژانت سیگموئید است، بنابراین در این مطالعه از آن برای نگاشت اطلاعات از لایه ورودی به لایه پنهان و از تابع محرک خطی برای نگاشت اطلاعات از لایه پنهان به لایه خروجی استفاده شد. لازم به ذکر است فرآیند محاسبات شبکه عصبی بر اساس کدنویسی در محیط متلب انجام شده است.

مدل برنامه‌ریزی بیان ژن

این مدل تعمیم یافته الگوریتم ژنتیک[5] است که برای اولین بار توسط کزا[6] (۱۹۹۲) بر اساس تئوری داروین ارائه شد. به این ترتیب که جمعیتی در جهت تکامل به صورت انتخابی، جمعیت نامناسب را رها کرده و فرزندانی اصلاح شده ایجاد می‌کنند. برنامه‌ریزی بیان ژن بر خلاف الگوریتم ژنتیک روی ساختار درختی فرمول‌ها به جای سلسله ارقامِ دودوئی عمل می‌کند. ساختارهای درختی از مجموعه توابع (تابع‌های ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند (قربانی و صالحی، ۱۳۹۰). به منظور بررسی بیشتر علاوه بر مجموعه چهار تابع اصلی (F1)، تابع‌های ریاضی پیش فرض برنامه (F2) نیز برای شبیه‌سازی رواناب به‌کار گرفته شدند. در روابط دوازده و سیزده به‌ترتیب مجموعه تابع‌های F1 و F2 نشان داده شده‌اند (سلطانی و همکاران، ۱۳۸۹؛ مسعودی و همکاران، ۱۳۹۱):

$$F1: (\pm, \times, \div) \qquad (۱۲)$$

$$\begin{aligned}F2: &(\pm, \times, \div, \sqrt{Q}, Exp(Q), Ln(Q), Q^2 \\ &, Q^3, Q^{1/3}, Sin(Q), Cos(Q), Arc\tan(Q))\end{aligned} \qquad (۱۳)$$

1- Kavzoglu and Colkesen
2- Guo et al.
3- Coulibaly *et al.*
4- Levenberg-Marquardt

5- Genetic Algorithm
6- Koza

جدول ۱- مشخصات مدل‌های سری زمانی مورد استفاده در تحقیق

مدل سری زمانی		توابع
(خود همبسته)	(۱۴)	$AR(p): X_t = \mu + \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu) + \varepsilon_t$
(خود همبسته با میانگین متحرک)	(۱۵)	$ARMA(p,q): Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$
	(۱۶)	$Z_t = X_t - \mu$
(خود همبسته میانگین متحرک یکپارچه)	(۱۷)	$u_t = \sum_{k=1}^{p} \phi_k u_{t-k} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$
	(۱۸)	$X_t = u_t + u_{t-1} + u_{t-2} + \ldots$

جدول ۲- نتایج آزمون‌های کلموگروف- اسمیرنوف و شاپیرو- ویلک برای داده‌های مشاهداتی اولیه و نرمال شده.

	جریان متوسط(مترمکعب بر ثانیه)	جریان حداقل(مترمکعب بر ثانیه)	جریان حداکثر(مترمکعب بر ثانیه)	انحراف معیار	چولگی	کشیدگی	کلموگروف- اسمیرنوف: درجه آزادی	کلموگروف- اسمیرنوف: آماره	کلموگروف- اسمیرنوف: سطح معنی داری	شاپیرو- ویلک: آماره	شاپیرو- ویلک: سطح معنی داری
داده‌های مشاهداتی اولیه	۸/۳۹	۰/۰۰	۸۹/۴۳	۱۳/۸۳	۲/۶۲	۱۳/۵۵	۲۷۲	۰/۲۷	۰/۰۰	۰/۶۵	۰/۰۰
داده‌های نرمال شده	۰/۰۰	-۲/۸۶	۲/۸۷	۱/۰۰	۰/۰۰	۸/۳۵	۲۷۲	۰/۰۴	۰/۲۰	۰/۹۹	۰/۲۷
داده‌های واسنجی	۰/۰۶	-۲/۸۶	۲/۸۷	۱/۰۱	-۰/۰۷	۰/۴۵	۲۰۵	۰/۰۵	۰/۲۰	۰/۹۹	۰/۲۱
داده‌های صحت سنجی	-۰/۱۷	-۲/۸۶	۲/۳۶	۰/۹۶	۰/۱۸	۰/۴۷	۶۷	۰/۰۵	۰/۲۰	۰/۹۸	۰/۵۹

همچنین در این تحقیق برای ساخت درخت تجزیه و تولید یک جمعیت اولیه از راه حل‌هایی استفاده شد که این راه حل‌ها از ترکیب تصادفی مجموعه توابع و ترمینال‌ها ایجاد می‌شوند. نحوه و دلیل انتخاب این نوع تابع‌ها با توجه به مطالعات خطیبی و همکاران[۱] (۲۰۱۳) و قربانی و همکاران[۲] (۲۰۱۲) صورت گرفت. لازم به ذکر است فرآیند محاسبات مدل برنامه‌ریزی بیان ژن با استفاده از نرم افزار GeneXproTools انجام گرفته است.

مدل‌های سری زمانی

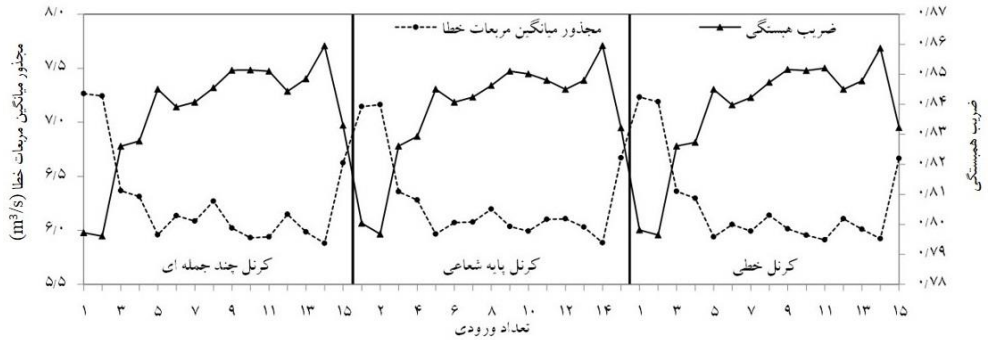
شبیه‌سازی سری‌های زمانی از جمله رویکردهای مطرح در زمینه پیش‌بینی و تولید داده‌های جدید می‌باشد که در هیدرولوژی از اوایل دهه‌ی ۱۹۶۰ توسط توماس[۳]، فیرینگ[۴] و یوجویچ[۵] آغاز گردیده و در دهه ۱۹۷۰ توسط باکس[۶] و جنکینس[۷] توسعه یافته است (باکس و همکاران، ۱۹۹۴). در پیش‌بینی با دقت بالا، افق زمانی مشخص و محدود بوده در حالی که در تولید داده‌ها، افق زمانی خاصی را نمی‌توان در نظر گرفت (باکس و همکاران، ۱۹۹۴).

در این تحقیق از مدل‌های خود همبسته، خود همبسته با میانگین متحرک و خود همبسته میانگین متحرک یکپارچه برای پیش‌بینی سری‌های زمانی استفاده گردیده که در جدول (۱) به‌طور مختصر به روابط هر کدام از آنها اشاره شده است (ژانگ[۸]، ۲۰۰۳؛ وی[۹] ۲۰۰۶):

در این روابط X_t و Z_t: سری‌های زمانی وابسته، μ: میانگین سری زمانی داده‌ها، ϕ_j: ضرایب مدل AR و ε_t: سری مستقل با توزیع نرمال، میانگین صفر و انحراف معیار σ_ε، θ_i و ϕ_i: ضرایب مدل ARMA، θ_j و ϕ_k: ضرایب مدل خود همبسته میانگین متحرک یکپارچه، n: تعداد داده‌ها، ρ_j: همبستگی جریان با تاخیر j، p و q: مرتبه مدل‌ها و σ^2: واریانس داده‌ها می‌باشند. همچنین سری u_t توسط تفاضل‌گیری dام از سری X_t حاصل می‌شود.

لازم به ذکر است، دقت روش حداکثر درست‌نمایی از روش گشتاورها در تخمین پارامترهای مدل‌های فوق بیشتر است (کارآموز و عراقی‌نژاد، ۱۳۸۴؛ صفوی، ۱۳۸۸). بنابراین در تحقیق حاضر از این روش برای تخمین پارامترها استفاده شد. همچنین فرآیند محاسبات مدل‌های سری زمانی از طریق کدنویسی در محیط متلب انجام گرفت.

1 - Khatibi *et al.*
2 - Ghorbani *et al.*
3- Thomas
4- Fiering
5- Yevjevich
6- Box
7- Jenkins
8- Zhang
9- Wei



شکل ۳- روند تغییرات ضریب همبستگی و ریشه میانگین مربعات خطا با توجه به افزایش تعداد ورودی‌ها و به ازای سه تابع کرنل

معیارهای ارزیابی عملکرد مدل‌ها

با استفاده از روش‌های مختلفی می‌توان عملکرد مدل‌ها را مورد مقایسه قرار داد که یکی از آنها استفاده از معیارهای ارزیابی است. از جمله معیارهای ارزیابی پر کاربرد در علوم مهندسی آب، ضریب همبستگی[1] و ریشه میانگین مربعات خطا[2] هستند که در این تحقیق نیز از این معیارها برای بررسی کارایی هر یک از مدل‌ها استفاده شد.

مراحل شبیه‌سازی در هر یک از مدل‌ها

در تحقیق حاضر پس از نرمال‌سازی داده‌های جریان، در ابتدا هر کدام از مدل‌های مورد استفاده به ازای پانزده ترکیب ورودی، با توجه به داده‌های دوره واسنجی (جدول ۲) آموزش داده شدند. سپس با استفاده از مدل‌های واسنجی شده، پیش‌بینی یک ماه بعد جریان رودخانه در دوره واسنجی برای ایستگاه سنته صورت گرفت. لازم به ذکر است که این مراحل در مدل ماشین بردار پشتیبان به ازای هر سه تابع کرنل (۴۵ مدل)، برای مدل شبکه عصبی مصنوعی به ازای تعداد نرون‌های یک تا بیست در لایه پنهان (۳۰۰ مدل)، در مدل برنامه‌ریزی بیان ژن برای دو مجموعه تابع F1 و F2 (۳۰ مدل) و برای مدل خود همبسته میانگین متحرک یکپارچه به ازای میانگین متحرک و تفاضل‌های مختلف (۸۱ مدل) انجام گرفت. همچنین با استفاده از معکوس روشی که برای نرمال سازی داده‌ها توضیح داده شد، جریان شبیه‌سازی شده نرمال به جریان شبیه‌سازی شده واقعی تبدیل گردید. سپس مقادیر شاخص‌های آماری ذکر شده برای ارزیابی عملکرد مدل‌ها در مراحل واسنجی و صحت‌سنجی محاسبه شد. در ادامه با استفاده از این شاخص‌ها بهترین مدل برای شبیه‌سازی جریان ماهانه ایستگاه سنته انتخاب گردید.

نتایج و بحث

پس از بررسی نرمال سازی داده‌ها، نتایج مربوط به هر یک از مدل‌های ماشین بردار پشتیبان، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن خودهمبسته به ترتیب در زیر آورده شده است.

نرمال‌سازی داده‌های جریان

به علت نرمال نبودن توزیع احتمالاتی داده‌های ماهانه جریان مشاهداتی، نرمال‌سازی داده‌ها براساس روشی که در بخش نرمال‌سازی داده‌ها این مقاله توضیح داده خواهد شد صورت گرفته است. در ادامه شاخص‌های آماری میانگین، مقدار حداقل، مقدار حداکثر، انحراف معیار، چولگی و کشیدگی محاسبه گردیده و آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک در سطح اطمینان ۰/۹۵ برای داده‌های مشاهداتی اولیه و نرمال شده انجام گرفت که نتایج آنها در جدول (۲) آورده شده است.

لازم به ذکر است چنانچه طی آزمون‌های مذکور سطح معنی‌داری بیشتر از پنج درصد به‌دست آید، داده‌ها با سطح اطمینان ۰/۹۵ از توزیع نرمال تبعیت می‌کنند. بر این اساس با توجه به نتایج آمده در جدول (۲) می‌توان دریافت که توزیع احتمالاتی داده‌های جریان ماهانه مشاهداتی اولیه در ابتدا نرمال نبوده، ولی پس از اعمال معادلات بخش نرمال‌سازی داده‌ها روی آنها، توزیع داده‌ها به نرمال تبدیل شده است.

شبیه‌سازی رواناب بامدل ماشین بردارپشتیبان خودهمبسته

در جدول (۳) بهترین نتیجه مربوط به هر تابع کرنل مدل ماشین بردار پشتیبان نشان داده شده است. مطابق با این جدول مدل ماشین بردار پشتیبان (۴,۱۴) با تابع کرنل چند جمله‌ای (چهار درجه تابع چند جمله‌ای و چهارده تعداد متغیر ورودی هستند) با مقادیر شاخص‌های ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۶ و ۵/۸۸ مترمکعب بر ثانیه در دوره صحت‌سنجی نسبت به مدل‌های ماشین بردار پشتیبان دیگر بهترین کارایی را داشته است. شبیه‌سازی‌ها نشان داد که مدل‌های ماشین بردار پشتیبان با توابع کرنل پایه شعاعی و چند جمله‌ای تقریباً از دقت یکسانی در تخمین رواناب برخوردارند و همچنین مدل ماشین بردار پشتیبان با تابع کرنل خطی نسبت به دیگر مدل‌های ماشین بردار پشتیبان عملکرد ضعیف‌تری دارد. همچنین شکل (۳) مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا را به ازای هر ۱۵

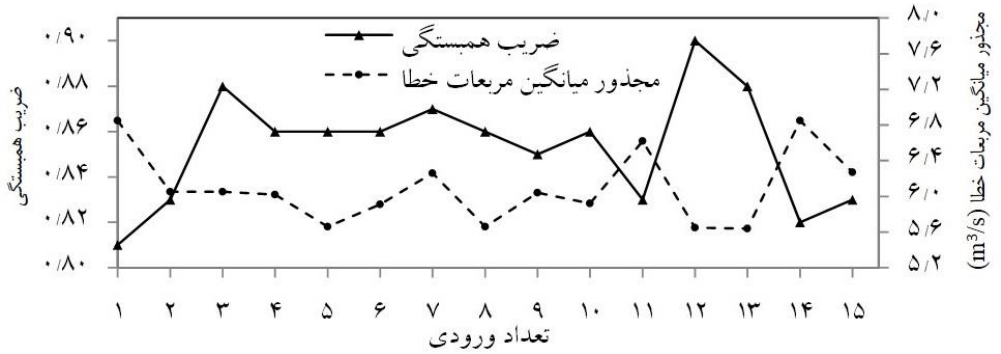
1- Correlation Coefficient
2- Root Mean Square Error

ورودی مختلف برای هر سه تابع کرنل چندجمله‌ای، پایه شعاعی و خطی در دوره صحت‌سنجی نشان می‌دهد.

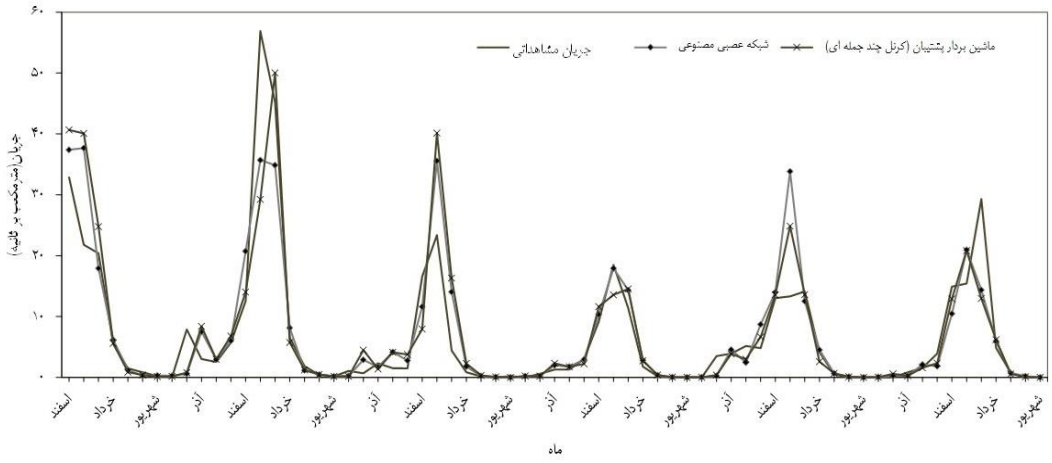
مطابق با این شکل با افزایش تعداد ورودی‌ها به مدل ماشین بردار پشتیبان با تابع کرنل چند جمله‌ای درجه چهار، مقادیر شاخص‌های ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب روند صعودی و نزولی داشته و به ازای ۱۴ ورودی شاخص‌های مذکور به‌ترتیب به بیشترین و کمترین مقدار خود می‌رسند. در شکل (۵) جریان شبیه‌سازی شده توسط مدل ماشین بردار پشتیبان با تابع کرنل چند جمله‌ای درجه چهار بازای چهارده ورودی به همراه جریان مشاهداتی برای دوره صحت‌سنجی نشان داده شده است. در اکثر تحقیقات صورت گرفته در زمینه شبیه-سازی‌های هیدرولوژیکی عمدتاً از ماشین بردار پشتیبان با تابع کرنل پایه شعاعی استفاده می‌شود (اسکندری و همکاران، ۱۳۸۹)، اما برتری تابع کرنل چندجمله‌ای در تحقیق حاضر و مطالعاتی از قبیل ازانی و همکاران (۱۳۹۳)، لزوم بررسی کارایی توابع کرنل مختلف را نشان می‌دهد.

جدول ۳- ارزیابی عملکرد مدل ماشین بردار پشتیبان در تخمین جریان ماهانه ایستگاه آب‌سنجی سنته.

	دوره واسنجی			دوره صحت سنجی		مقادیر بهینه پارامترها			
تابع کرنل	بهترین تعداد ورودی	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	c	σ	d	t	ε
چند جمله‌ای	۱۴	۰/۸۱	۸/۸۲	۰/۸۶	۵/۸۸	۱۵/۹۶	-	۴	۳۰۴/۹۴	۱/۷
پایه شعاعی	۱۴	۰/۸۲	۸/۷۷	۰/۸۶	۵/۸۹	۱۱۶/۶۶	۱۵۹۸/۹۱	-	-	۱/۴۴
خطی	۱۱	۰/۸۰	۹/۱۴	۰/۸۵	۵/۹۱	۰/۰۷	–	-	-	۰/۴۴



شکل ٤- روند تغییرات مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا مدل‌های شبکه عصبی مصنوعی منتخب، با توجه به تعداد ورودی‌های مختلف



شکل ۵- سری زمانی جریان ماهانه مشاهداتی و شبیه‌سازی شده ایستگاه سنته توسط مدل‌های منتخب ماشین بردار پشتیبان و شبکه عصبی مصنوعی خودهمبسته در دوره صحت‌سنجی (اسفند ۱۳۸٤ تا شهریور ۱۳۹۰)

مطابق با این شکل با افزایش تعداد ورودی‌ها به مدل ماشین بردار پشتیبان با تابع کرنل چند جمله‌ای درجه چهار، مقادیر شاخص‌های ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب روند صعودی و نزولی داشته و به ازای چهارده ورودی شاخص‌های مذکور به ترتیب به بیشترین و کمترین مقدار خود می‌رسند. در شکل (۵) جریان شبیه‌سازی شده توسط مدل ماشین بردار پشتیبان با تابع کرنل چند جمله‌ای درجه چهار بازای چهارده ورودی به همراه جریان مشاهداتی برای دوره صحت‌سنجی نشان داده شده است. در اکثر تحقیقات صورت گرفته در زمینه شبیه‌سازی‌های هیدرولوژیکی عمدتاً از ماشین بردار پشتیبان با تابع کرنل پایه شعاعی استفاده می‌شود (اسکندری و همکاران، ۱۳۹۱)، اما برتری تابع کرنل چندجمله‌ای در تحقیق حاضر و مطالعاتی از قبیل ازانی و همکاران (۱۳۹۳)، لزوم بررسی کارایی توابع کرنل مختلف را نشان می‌دهد.

شبیه‌سازی جریان با مدل شبکه عصبی مصنوعی خودهمبسته

در شکل (۴) روند تغییرات مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا مدل شبکه عصبی مصنوعی خودهمبسته به ازای بهترین تعداد نرون برای هر یک از تأخیرهای یک تا پانزده ماه جریان برای دوره صحت‌سنجی نشان داده شده است.

با توجه به شکل (۴) مدل خودهمبسته شبکه عصبی مصنوعی (۵،۳) (مدل با پنج متغیر ورودی و تعداد سه نرون در لایه پنهان) با مقادیر شاخص‌های ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۶ و ۵/۶۶ متر مکعب بر ثانیه در دوره صحت‌سنجی نسبت به مدل‌های شبکه عصبی مصنوعی دیگر از کارایی و دقت بیشتری در شبیه‌سازی رواناب حوضه برخوردار است. در شکل (۵) جریان ماهانه مشاهداتی و شبیه‌سازی شده توسط مدل شبکه عصبی مصنوعی (۵،۳) برای دوره صحت‌سنجی نشان داده شده است.

شبیه‌سازی رواناب با مدل برنامه‌ریزی بیان ژن خودهمبسته

در این تحقیق استفاده از مدل برنامه‌ریزی بیان ژن نیز به دلیل توانایی انتخاب متغیرهای مؤثر در مدل‌سازی، حذف متغیرهایی با تأثیر کمتر و همچنین قابلیت ارائه رابطه صریح برای تخمین جریان مد نظر قرار گرفت. در جدول (۴) مقادیر پارامترهای مورد استفاده در مدل برنامه‌ریزی بیان ژن ارائه شده است. در شکل (۶) روند تغییرات مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا در مدل برنامه‌ریزی بیان ژن خودهمبسته به ازای دو تابع F1 و F2، برای هر یک از تأخیرهای یک تا پانزده ماه جریان برای دوره صحت‌سنجی نشان داده شده است. در جدول (۵) بهترین نتیجه حاصل از مدل برنامه‌ریزی بیان ژن خودهمبسته با هر یک از دو تابع F1 و F2 و نشان داده شده است. نتایج نشان داد که مدل برنامه‌ریزی بیان ژن ($F1,5$) (مدل با ۵ ورودی و مجموعه تابع F1) با مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۴ و ۶/۲۳ متر مکعب بر ثانیه در مرحله صحت‌سنجی نسبت به سایر ۲۹ مدل برنامه‌ریزی بیان ژن دیگر از دقت بیشتری در شبیه‌سازی رواناب برخوردار است. در شکل (۷) سری زمانی ماهانه رواناب مشاهداتی و شبیه‌سازی شده توسط مدل برنامه‌ریزی بیان ژن ($F1,5$) برای دوره صحت‌سنجی نشان داده شده است.

نتایج مدل‌های سری زمانی

برای حصول اطمینان از نتایج توابع خودهمبستگی و خودهمبستگی جزئی مدل‌های سری زمانی خود همبسته(p) با تأخیرهای یک تا پانزده، مدل خودهمبسته با میانگین متحرک (p,q) با مقادیر مختلف p و q و مدل خودهمبسته میانگین متحرک یکپارچه (p,d,q) با تأخیر، میانگین متحرک و تفاضل‌های مختلف به‌کار گرفته شدند. بر این اساس جریان رودخانه با ۸۱ مدل مختلف سری زمانی شبیه‌سازی شده و عملکرد هر یک از آنها در دوره‌های واسنجی و صحت‌سنجی با استفاده از شاخص‌های آماری ضریب همبستگی و ریشه میانگین مربعات خطا مورد ارزیابی قرار گرفت. قابل ذکر است شاخص‌های آماری آورده شده در جدول‌ها پس از بازگردانی داده‌های جریان نرمال شبیه‌سازی شده به داده‌های جریان واقعی شبیه‌سازی شده محاسبه شده‌اند.

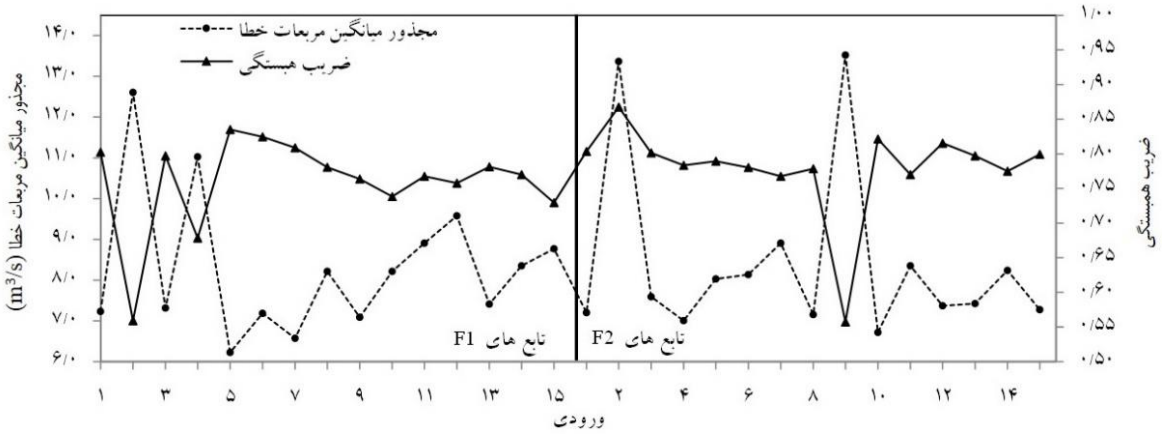
نتایج نشان داد که مدل خودهمبسته با میانگین متحرک (۶،۲) از بین ۸۱ مدل سری زمانی به‌کار گرفته شده با مقادیر شاخص‌های آماری ضریب همبستگی و ریشه میانگین مربعات خطا به ترتیب با ۰/۸۲ و ۶/۴۷ متر مکعب بر ثانیه برای دوره صحت‌سنجی، بهترین مدل سری زمانی برای شبیه‌سازی جریان ماهانه در ایستگاه سنته است. با مقایسه مقادیر شاخص‌های آماری مدل‌های خودهمبسته (۱) و خودهمبسته با میانگین متحرک (۶،۲) در دوره صحت‌سنجی می توان به این نتیجه رسید که مدل خودهمبسته (۱) با مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۷۶ و ۷/۹۶ متر مکعب بر ثانیه با یک ورودی از دقت قابل قبولی در شبیه‌سازی جریان برخوردار است، این موضوع نشان از توانایی تابع خود همبستگی جزئی در تعیین موثرترین ورودی‌ها در شبیه‌سازی دارد.

جدول ٤- مقادیر پارامترهای مورد استفاده با استفاده از روش برنامه‌ریزی بیان ژن

تنظیمات کلی		تابع‌های ژنتیکی	
تعداد کروموزوم‌ها	۳۰	نرخ جهش	۰/۰۴۴
اندازه رأس	۸	نرخ وارون سازی	۰/۱
تعداد ژن‌ها در هر کروموزوم	۳	نرخ ترانهش درج متوالی	۰/۱
تعداد جمعیت تولیدی	۳۰۰۰۰	نرخ ترانهش ریشه درج متوالی	۰/۱
تابع پیوند	جمع (+)	نرخ ترکیب تک نقطه‌ایی	۰/۳
		نرخ ترکیب دو نقطه‌ایی	۰/۳

جدول ٥- نتایج مدل برنامه‌ریزی بیان ژن خودهمبسته با استفاده از دو مجموعه تابع ریاضی F1 و F2.

مدل	واسنجی		صحت سنجی	
	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)
برنامه‌ریزی بیان ژن (F1,5)	۰/۷۷	۹/۷۰	۰/۸۴	۶/۲۳
برنامه‌ریزی بیان ژن (F2,۱۰)	۰/۸۴	۸/۳۰	۰/۸۲	۶/۷۲



شکل ٦- روند تغییر مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا مدل برنامه‌ریزی بیان ژن برای هر یک از تابع‌های عملگر F1 و F2 با توجه به تعداد ورودی‌های مختلف

عملکرد و روابط بدست آمده برای مدل‌های خودهمبسته (۱) و خودهمبسته با میانگین متحرک (۲،۶) به‌ترتیب در جدول (۶) و رابطه‌های (۱۹) و (۲۰) آورده شده‌اند:

$$Q_t = 0.02 + 0.73Q_{t-1} + \varepsilon_t \qquad (۱۹)$$

$$Q_t = 0.03 - 0.07Q_{t-1} - 0.43Q_{t-2} + 0.64Q_{t-3} + 0.05Q_{t-4} + 0.15Q_{t-5} + 0.13Q_{t-6} + 0.75\varepsilon_{t-1} + 0.95\varepsilon_{t-2} + \varepsilon_t \qquad (۲۰)$$

در این روابط Q_t: جریان تخمین زده شده، Q_{t-1}، Q_{t-2}، .... و Q_{t-6} به‌ترتیب جریان ماهانه با تأخیرهای یک تا شش ماه و ε_t، ε_{t-1} و ε_{t-2} به‌ترتیب نوفه سفید با تأخیرهای یک تا سه ماه هستند.

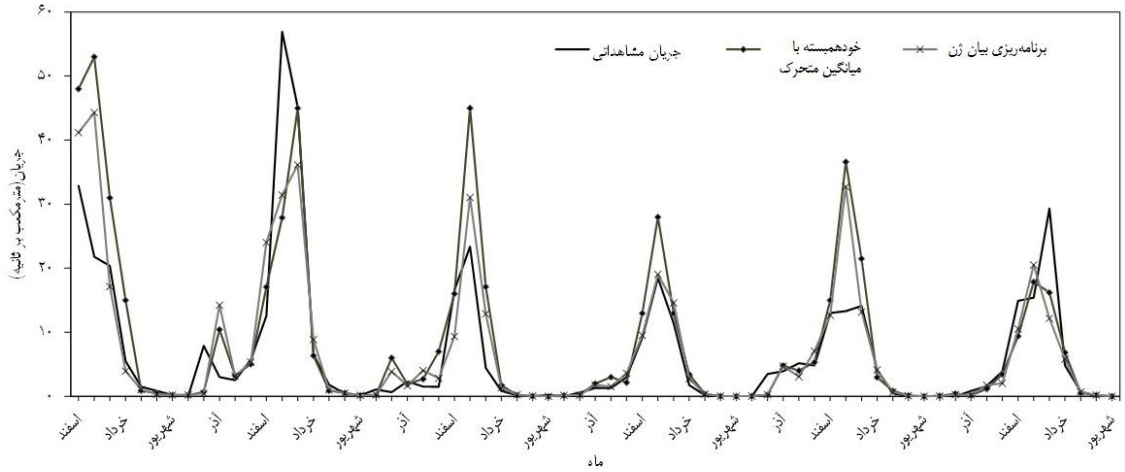
در شکل (۷) مقادیر مشاهداتی و شبیه‌سازی شده رواناب ماهانه توسط مدل خودهمبسته با میانگین متحرک (۲،۶) برای دوره صحت‌سنجی نشان شده است.

مقایسه عملکرد مدل‌های سری زمانی، شبکه عصبی مصنوعی، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان

براساس نتایج بخش‌های قبل، بهترین نتیجه به‌دست آمده به ازای هر یک از چهار مدل به‌کار گرفته شده در جدول (۷) آمده است.

جدول ٦- ارزیابی عملکرد مدل‌های خودهمبسته (۱) و خودهمبسته با میانگین متحرک (٦،۲) در تخمین جریان ماهانه ایستگاه سنته

مدل	دوره واسنجی		دوره صحت‌سنجی	
	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)
خودهمبسته (۱)	۰/۸۲	۸/۴۱	۰/۷۶	۷/۹۵
خودهمبسته با میانگین متحرک (۶،۲)	۰/۷۷	۹/۸۲	۰/۸۲	۶/۴۷



شکل ۷- جریان ماهانه مشاهداتی و شبیه‌سازی شده ایستگاه سنته توسط مدل‌های منتخب برنامه‌ریزی بیان ژن و خود همبسته میانگین متحرک در دوره صحت‌سنجی (اسفند ۱۳۸٤ تا شهریور ۱۳۹۰

جدول ۷- ارزیابی کارایی مدل‌های خودهمبسته با میانگین متحرک (٦،۲)، ماشین بردار پشتیبان، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن در شبیه‌سازی جریان ماهانه ایستگاه سنته

مدل	دوره واسنجی		دوره صحت‌سنجی	
	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)
خودهمبسته با میانگین متحرک (۶،۲)	۷۷/۰	۸۲/۹	۸۲/۰	۴۷/۶
ماشین بردار پشتیبان (۴،۱۴)	۸۱/۰	۸۲/۸	۸۶/۰	۸۸/۵
شبکه عصبی مصنوعی (۵،۳)	۰/۸۴	۸/۲۱	۰/۸۶	۵/۶۶
برنامه‌ریزی بیان ژن (F1,۵)	۰/۷۷	۹/۷۰	۰/۸۴	۶/۲۳

براساس نتایج آمده در جدول (۷) مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و خودهمبسته با میانگین متحرک (۶،۲) به لحاظ دقت در شبیه‌سازی جریان به‌ترتیب در رتبه‌های اول تا چهارم قرار دارند. در نهایت مدل شبکه عصبی مصنوعی با پنج ورودی به عنوان بهترین مدل از بین ۱۷۱ مدل مورد ارزیابی قرار گرفته در این تحقیق انتخاب گردید. تحقیقاتی از قبیل جین و همکاران[1] (۲۰۰۶) و محرم پور و همکاران[2] (۲۰۱۳) نشان دادند که مدل ماشین بردار پشتیبان با توابع کرنل مختلف نسبت به مدل‌های سری زمانی خودهمبسته با میانگین متحرک و خود همبسته میانگین متحرک یکپارچه عملکرد بهتری در شبیه‌سازی جریان رودخانه‌ها از خود نشان می‌دهد که این موضوع با نتایج تحقیق حاضر مطابقت دارد.

مطالعات هوانگ و چان‌هیلتون[3] (۲۰۰۴) و همچنین جین و کومار[4] (۲۰۰۷) نیز همانند نتایج این مطالعه مدل شبکه عصبی مصنوعی را بهتر از مدل‌های سری زمانی در شبیه‌سازی جریان رودخانه‌ها دانسته‌اند.

1- Jian *et al.*
2- Moharrampour *et al.*
3- Huang and Chan Hilton
4- Jain and Kumar

جدول ۸- ارزیابی دقت مدل‌های خودهمبسته با میانگین متحرک، ماشین بردار پشتیبان، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن در تخمین مقادیر جریان در سه بازه مختلف

	$۳۰ < Q \leq ۶۰$		$۱۰ < Q \leq ۳۰$		$۰ \leq Q \leq ۱۰$	
	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی	ریشه میانگین مربعات خطا (متر مکعب بر ثانیه)	ضریب همبستگی
ماشین بردار پشتیبان	۱۶/۸۳	-۰/۵۴	۹/۵۲	۰/۳۸	۲/۳۸	۰/۶۸
خودهمبسته با میانگین متحرک	۱۸/۹۰	-۰/۹۲	۱۳/۸۱	۰/۴۱	۳/۰۲	۰/۶۶
برنامه‌ریزی بیان ژن	۱۶/۳۱	-۰/۹۹	۱۰/۷۰	۰/۱۸	۲/۴۷	۰/۶۶
شبکه عصبی مصنوعی	۱۶/۷۰	-۰/۴۶	۹/۱۵	۰/۲۳	۲/۰۹	۰/۷۴

عملکرد مدل‌های منتخب در سه بازه مختلف جریان

هر یک از چهار مدل منتخب مذکور (جدول ۷) در تخمین جریان در سه محدوده جریان کوچک‌تر از ۱۰، ۱۰ تا ۳۰ و بزرگ‌تر از ۳۰ مترمکعب بر ثانیه عملکرد متفاوتی از خود نشان داده‌اند که نتایج مربوط به آنها در جدول (۸) آورده شده است. بر اساس مقادیر شاخص‌های آماری ضریب همبستگی و ریشه میانگین مربعات خطا آورده شده در جدول (۸)، مدل شبکه عصبی مصنوعی با پنج ورودی در محدوده اول و دوم و مدل برنامه‌ریزی بیان ژن به ازای پنج ورودی با چهار تابع اصلی در محدوده سوم از دقت بیشتری نسبت به سایر مدل‌ها در شبیه‌سازی جریان ماهانه برخوردار هستند.

نتیجه‌گیری

در تحقیق حاضر سعی گردید عملکرد مدل‌های سری زمانی، شبکه عصبی مصنوعی، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان در شبیه‌سازی خودهمبسته جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته مورد بررسی قرار گرفته و دقت این مدل‌ها در برآورد جریان ماهانه این ایستگاه مقایسه شوند. در این راستا ابتدا توزیع احتمالاتی داده‌های جریان مشاهداتی به نرمال تبدیل گردیده و سپس آزمون‌های کلموگروف- اسمیرنوف و شاپیرو- ویلک در سطح اطمینان ۰/۹۵ برای آنها انجام گرفت. در ادامه برای لحاظ نمودن تاثیر جریان با تأخیرهای ماهانه مختلف و همچنین ساختار به‌کار رفته در مدل‌ها بر عملکرد آنها در شبیه‌سازی جریان، کارایی مدل‌های شبکه عصبی، ماشین بردار پشتیبان با توابع کرنل مختلف، برنامه‌ریزی بیان ژن با دو سری تابع (F1 و F2) و مدل‌های سری زمانی به ازای تأخیرهای یک تا پانزده جریان ماهانه مورد ارزیابی قرار گرفت و نتایج زیر بدست آمد.

مدل شبکه عصبی مصنوعی با پنج ورودی با مقادیر شاخص‌های آماری ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۶، ۵/۶۶ متر مکعب بر ثانیه در دوره صحت‌سنجی از بین سایر مدل‌های ارزیابی شده شبکه عصبی مصنوعی، ماشین بردار پشتیبان، خود همبسته میانگین متحرک یکپارچه و برنامه‌ریزی بیان ژن به عنوان بهترین مدل برای شبیه‌سازی جریان ماهانه حوضه خرخره‌چای تشخیص داده شد. همچنین نتایج نشان داد که مدل ماشین بردار پشتیبان با تابع کرنل چندجمله‌ای به ازای چهارده ورودی با مقادیر شاخص‌های ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۶ و ۵/۸۸ متر مکعب بر ثانیه در دوره صحت‌سنجی نسبت به سایر مدل‌های ماشین بردار پشتیبان با توابع کرنل خطی و پایه شعاعی بهترین کارایی را دارا است. مدل برنامه‌ریزی بیان ژن با مجموعه تابع F1 و پنج ورودی با مقادیر شاخص‌های آماری ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۴ و ۶/۲۳ متر مکعب بر ثانیه، نسبت به سایر مدل‌های برنامه‌ریزی بیان ژن ارزیابی شده نتایج بهتری را به‌دست آورده است.

براساس نتایج تحقیق حاضر، مدل خودهمبسته با میانگین متحرک (۶،۲) با مقادیر ضریب همبستگی و ریشه میانگین مربعات خطا به‌ترتیب برابر با ۰/۸۲ و ۶/۴۷ متر مکعب بر ثانیه در دوره صحت‌سنجی، بهترین مدل از بین مدل‌های سری زمانی به‌کار گرفته شده با تأخیر، میانگین متحرک و تفاضل‌های مختلف در شبیه‌سازی جریان ماهانه حوضه خرخره‌چای تشخیص داده شد.

مقایسه مقادیر جریان شبیه‌سازی شده در سه محدوده کوچک‌تر از ۱۰، بین ۱۰ تا ۳۰ و بزرگتر از ۳۰ متر مکعب بر ثانیه با مقادیر جریان مشاهداتی آنها در بازه‌های مذکور نشان داد که مدل شبکه عصبی مصنوعی با پنج ورودی در محدوده اول و دوم و مدل برنامه‌ریزی بیان ژن به ازای پنج ورودی با چهار تابع اصلی در محدوده سوم دارای خطای برآورد کمتری نسبت به مدل‌های ماشین بردار پشتیبان و خودهمبسته با میانگین متحرک (۶،۲) هستند.

ضعف مدل‌ها در شبیه‌سازی جریان‌های حداکثر می‌تواند به این علت باشد که در شبیه‌سازی خودهمبسته جریان ماهانه رودخانه‌ها، جریان در هر ماه تابعی از جریان ماه‌های قبلی می‌باشد، لذا بسیاری از عوامل مؤثر در شبیه‌سازی جریان از جمله دما، بارش، ذوب ناگهانی برف و ... که موجب کاهش و یا افزایش ناگهانی جریان رودخانه می‌شوند، در محاسبات وارد نمی‌شوند. لذا در شبیه‌سازی خودهمبسته جریان احتمال تخمین کمتر و یا زیادتر جریان در برخی از ماه‌ها همواره وجود دارد. همچنین در برخی موارد، این موضوع

یکی از دلایل اصلی نزدیکی همزمان جریان شبیه‌سازی شده مدل‌ها با یکدیگر و فاصله آنها با مقدار واقعی جریان می‌باشد.

در شبیه‌سازی رواناب این موضوع بسیار حائز اهمیت است که خطای تخمین‌ها به یک سو سوق پیدا نکنند. به‌عبارت دیگر اگر مدلی تمام جریان‌ها را کمتر و یا بیشتر از مقدار واقعی جریان مشاهداتی تخمین بزند، حاکی از ضعیف شدید شبیه‌سازی‌ها می‌باشد. در شبیه‌سازی هر یک از مدل‌های این تحقیق نیز تخمین زیاد و یا کم پیوسته وجود ندارد.

مقایسه عملکرد مدل‌های مورد بررسی نشان داد که مدل شبکه عصبی مصنوعی با ۵ ورودی از دقت و کارایی بیشتری در شبیه‌سازی جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته نسبت به مدل‌های سری زمانی (خود همبسته، خود همبسته با میانگین متحرک و خود همبسته میانگین متحرک یکپارچه)، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان برخوردار است. از دیگر مزایای شبکه عصبی مصنوعی سرعت شبیه‌سازی و دقت بیشتر این مدل در شبیه‌سازی جریان رودخانه نسبت به سایر مدل‌های به‌کار رفته است.

منابع

۱- ازانی، ع.، فاضلی‌فرد، م. ح. و م. ع. قربانی. ۱۳۹۳. شبیه‌سازی سطح آب دریاچه ارومیه با استفاده از ماشین‌های بردار پشتیبان و شبکه عصبی مصنوعی. سیزدهمین کنفرانس هیدرولیک ایران، دانشگاه تبریز.

۲- اسکندری، ع.، نوری، ر.، معراجی، ع. و ا. کیاقادی. ۱۳۹۱. توسعه مدلی مناسب بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی به هنگام اکسیژن خواهی بیو شیمیایی ۵ روزه. مجله محیط شناسی، ۳۸(۶۱): ۸۲-۷۱.

۳- بنی حبیب، م. ا. و ولی پور. م .۱۳۸۷. ارزیابی مقایسه‌ای مدل‌های آرما، آریما و مدل خودهمبسته شبکه عصبی مصنوعی در پیش بینی جریان ورودی به مخزن سد دز. اولین کنفرانس بین المللی بحران آب، دانشگاه زابل.

۴- سلطانی، ع.، قربانی، م. ع.، فاخری‌فرد، ا.، دربندی، ص. و د. فرسادی‌زاده. ۱۳۸۹. برنامه‌ریزی ژنتیک و کاربرد آن در مدل‌سازی فرآیند بارش- رواناب. دانش آب و خاک، ۲۰(۴) : ۷۱-۶۱.

۵- صفوی، ح. ر. ۱۳۸۸. هیدرولوژی مهندسی. انتشارات ارکان دانش، ۶۶۳-۶۱۱.

۶- عادلی، ع.، فتحی مقدم، م. و س. ح. موسوی جهرمی. ۱۳۹۳. استفاده از مدل استوکستیک برای تولید سری‌های زمانی مصنوعی و پیش‌بینی جریان ورودی، مطالعه موردی: مخزن سد تالوگ، استان خوزستان. فصلنامه بین المللی پژوهشی تحلیل منابع آب و توسعه، ۲(۵): ۱۳-۱.

۷- قربانی، م. ع. و ا. صالحی. ۱۳۹۰. استفاده از برنامه‌ریزی بیان ژن در بررسی تغییرات داده‌های کیفی آب زیرزمینی با نوسانات سطح آب در دشت برخوار اصفهان. ششمین کنگره ملی مهندسی عمران، دانشگاه سمنان.

۸- کارآموز، م. و ش. عراقی نژاد. ۱۳۸۴. هیدرولوژی پیشرفته. انتشارات دانشگاه صنعتی امیرکبیر، ۳۱۰-۲۷۱.

۹- مسعودی، ا.، پارسامهر، پ.، سلماسی، ف. و س. پوراسکندر. ۱۳۹۱. تخمین ضریب دبی در سرریزهای لبه پهن مرکب با استفاده از رگرسیون، برنامه‌ریزی ژنتیک و شبکه عصبی. آب وخاک (علوم و صنایع کشاورزی)، (۴)۲۶: ۹۴۲-۹۳۳.

10- Adamowski, J. 2013. Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India. Versita, 45(1), 71-83.

11- Anonymos. 2000. Artificial neural networks in hydrology, I: preliminary concepts. Task Committee on Application of Artificial Neural Networks in Hydrology. Journal of Hydrologic Engineering, 5(2): 115-123.

12- Awchi, T.A. 2014. River discharges forecasting in northern Iraq using different ANN techniques. Water Resources Management, 28(3): 801–814.

13- Borelli, A., De Falco, I., Della Cioppa, A., Nicodemi, M. and G.Trautteur. 2006. Performance of genetic programming to extract the trend in noisy data series. Physica A: Statistical Mechanics and Its Applications, 370(1): 104-108.

14- Box, G.E.P., Jenkins, G.M. and G.C. Reinsel. 1994. Time series analysis: Forecasting and control. Third edition, Prentice Hall.

15- Coulibaly, P., Anctil, F. and B. Bobée. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology, 230(3-4): 244-257.

16- Dibike, Y., Velickov, S., Solomatine, D. and M. Abbott. 2001. Model induction with of support vector machines: Introduction and applications. Journal of Computing in Civil Engineering, 15(3): 208- 216.

17- Ghorbani, M.A., Khatibi, R., Asadi, H. and P. Yousefi. 2012. Inter-comparison of an evolutionary programming model of suspended sediment time-series whit other local model. INTECH. doi. org/10.5772/47801: 255-282.

18- Guo, B., Gunn, S.R., Damper, R.I. and J.D.B. Nelson. 2008. Customizing kernel functions for SVM-based hyperspectral image classification. IEEE Transactions on Image Processing, 17(4): 622-629.

19- Huang, W. and B.X. Chan-Hilton. 2004. Forecasting flows in Apalachicola river usingneural networks. Hydrological Processes, 18(13): 2545-2564.

20- Jain, A. and A.M. Kumar. 2007. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing Journal, 7(2): 585-592.

21- Jian, Y., Cheng, C.T. and K.W. Chau. 2006. Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal, 51(4): 599-612.

22- Kakaei Lafadani, E., Moghaddam Nia, A., Ahmadi, A., Jajarmizadeh, M. and M. Ghafari. 2013. Stream flow simulation using SVM, ANFIS and NAM models (A case study). Caspian Journal of Applied Sciences Reaserch, 2(4): 86-93.

23- Kalteh, A.M. 2013. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Computers and Geosciences, 54: 1–8.

24- Kavzoglu, T. and I. Colkesen. 2009. A kernel functions analysis for support vector machines for land coverclassification. International Journal of Applied Earth Observation and Geoinformation, 11(5): 352-359.

25- Khatibi, R., Naghipour, L., Ghorbani, M.A. and M.T. Aalami. 2013. Predictability of relative humidity by two artificial intelligence techniques using noisy data from two Californian gauging stations. Neural Computing and Applications, 23(7-8): 2241-2252.

26- Kisi, O., Moghaddam Nia, A., Ghafari Gosheh, M., Jamalizadeh Tajabadi, M.R. and A. Ahmadi. 2012. Intermittent streamflow forecasting by using several data driven techniques. Water Resources Management, 26(2): 457–474.

27- Kouchakzadeh, M. and A. Bahmani. 2006. Assessment of artificial neural networks revenuein reducing reqired parameters for estimation of reference evapotranspiration. Journal of Agricultural Sciences, 11(4): 87-97.

28- Koza, J.R. 1992. Genetic programming: On the programming of computers by means of natural selection. Cambridge, MA, MIT Press.

29- Moharrampour, M., Mehrabi, A., Hajikandi, H. and S. Sohrabi. 2013. Comparison of support vector machines (SVM) and autoregressive integrated moving average (ARIMA) in daily flow forecasting. Journal of River Engineering, 1(1): 34-45.

30- Nayak, P.C., Sudheer, K.P., Rangan, D.M. and K.S. Ramasastri. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology, 291(2-1): 52-66.

31- Salas, J.D. 1993. Analysis and modeling of hydrological time series. In: Handbook of Hydrology, edited by David R, Maidment, McGraw-Hill, New York, 19(72): 1-19.

32- Samsudin, R., Saad, P. and A. Shabri. 2011. River flow time series using least squares support vector machines. Hydrology and Earth System Sciences, 15: 1835-1852.

33- Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika, 52(3-4): 591-611.

34- Terzi, O. and G. Ergin. 2014. Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. Neural Computing and Applications, 25(1): 179-188.

35- Vapnik, V.N. and C. Cortes. 1995. Support vector networks. Machine Learning, 20: 273-297.

36- Wei, W.W.S. 2006. Time series analysis: univariate and multivariate methods. (Second edition), Greg Tobin Publisher, ISBN 0-321-32216-9.

37- Yang, K., Shan, G. and L. Zhao. 2006. Correlation coefficient method for support vector machine input samples. International Conference on Machine Learning and Cybernetics, PP. 2857-2861.

38- Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neuro Computing, 50: 159-175.

Shahid Chamran University of Ahvaz
Irrigation Sciences and Engineering **Vol. 40, No. 4, Winter 2018, p. 91-107**

EXTENDED ABSTRACT

An Assessment of Time Series and Autoregressive Artificial Neural Network Models, Support Vector Machine and Gene Expression Programming Models Performance in Monthly River Flow Simulation (Case Study: Kherkherechi River Basin)

M. Isazadeh[1*], H. Ahmadzadeh[2], M. A. Ghorbani[3] and M. H. Fazelifard[4]

1*- Corresponding Author, M.Sc. of Water Resources Engineering, Faculty of Agriculture, University of Tabriz. (*mohammadisazade@gmail.com*)
2- Ph.D. Student of Water Resources Engineering, Faculty of Agriculture, University of Tabriz.
3- Associate Professor, Water Engineering Department, Faculty of Agriculture, University of Tabriz.
4- Ph.D. Student of Water Resources Engineering, Faculty of Agriculture, University of Tabriz.

Received: 20 February 2016 Accepted:4 July 2016

Keywords: Autoregressive Runoff simulation, Time series models, Multilayer perceptron, Kherkherehchai basin.

Introduction

Selecting a model that simulate the runoff with high accuracy and less error, can be helpful in favorable management of water resources plans and increasing the performance of these plans. Also, increasing the accuracy of runoff simulation in the basins with no meteorological data, is of great significance in efficient management of water resources in these basins.

Methodology

For the purpose of this study,autoregressive simulation can be useful. In this paper, the performance of three data-driven models, namely, Artificial Neural Networks (ANN), Gene Expression Programming (GEP) and Support Vector Machine (SVM) as well as time series models were evaluated to forecast the monthly river flow in the Kherkherehchai river basin. The study area in this paper is Kherkherehchai river basin as one of the Zarrinehrud river sub-basin. This basin has significant contribution in the Bukan reservoir inflow and the water demands supply downstream of this reservoir. The data used in this case is monthly river flow in the Santeh hydrometric station during 1988 to 2011. Therefore, 70 percent of data was selected as calibration data and 30 percent of data was selected as validation data in each model. In this paper, Auto Correlation and Partial Auto Correlation Functions (ACF and PACF) were used to determine the maximum number of delays required in monthly flow modeling. Then, various combinations of model inputs that should be considered in the river flow estimation, were specified. Also, the effects of different models structure on river flow simulation were verified by applying them and comparing their results. In this regard, linear models of autoregressive, autoregressive moving average, autoregressive integrated moving average, multilayer perceptron, Gene Expression Programming (GEP) model with four and thirteen mathematical functions and support vector machine model with three kernel functions were applied for autoregressive simulation of monthly river flow. Correlation ccoefficient and Root Mean Square Errors were used as the criteria for evaluating the performance of each models.

Results and Discussion

Investigating the Auto Correlation Functions and Partial Auto Correlation Functions of normalized monthly flow showed that the first delay of the flow (delay degree of 1) has the most impact on flow estimation based on the Partial Auto Correlation Functions diagram. Also based on the Auto Correlation Functions diagram, maximum required number of delays of monthly flow for the runoff simulation is 15 months.

Support Vector Machine model (4,14) with polynomial kernel function (4 and 14 are degree of polynomial function and input variable numbers, respectively) with Correlation Coefficient and Root Mean Square Errors values of 0.86 ,and 5.88 (m^3/sec) has the best performance compared to other Support Vector Machine models in the validation period. Simulations showed that the Support Vector Machine models with kernel functions of radial basis and polynomial have the same accuracy in runoff estimation, and also the Support Vector Machine model with linear kernel function has poor performance compared to the other Support Vector Machine models.

The results showed that the Gene Expression Programming model (F1,5)(F1,5 are function set F1 and input variable numbers, respectively) with Correlation Coefficient and Root Mean Square Errors values of 0.84, and 6.23 (m^3/sec) has high accuracy compared to 29 other Gene Expression Programming models in runoff simulation during the validation period.

The results indicated that the Autoregressive Moving Average (ARMA) (6,2) model with Correlation Coefficient and Root Mean Square Errors values of 0.82, and 6.47 (m^3/sec) is the best time series model for simulating monthly flow at Senteh station between 81 time series models during the validation period.

Comparing the evaluation criteria values of the Autoregressive (AR)(1) and ARMA (6,2) it can be concluded that AR (1) model with Correlation Coefficient and Root Mean Square Errors values of 0.76, and 7.96 (m^3/sec) has an acceptable accuracy in monthly flow simulations. This denotesthe ability of the Partial Auto Correlation Functions to determine the most effective inputs in the simulation.

The results illustrated that multilayer perceptron (5,3) (5,3 are input variable numbers and number of neuron in the Artificial Neural Network models, respectively)

with Correlation Coefficient and Root Mean Square Errors values of 0.84, and 8.21(m^3/sec) in training period and 0.86, and 5.66 (m^3/sec) in the validation period respectively has the highest accuracy in estimation of monthly river flow compared to other considered models.

The Models have different performance in the flow estimation in three ranges of less than 10, 10 to 30, and more than 30 (m^3/sec). The results showed that the artificial neural network model with five inputs in the first and second ranges and Gene Expression Programming model with five inputs in the third ranges have more accuracy than the other models in the monthly flow simulation.

Conclusions

Comparison of the performances of the studied models showed that the Artificial Neural Networks has a better performance compared to the other models in the monthly flow simulation of the Kherkherehchai river basin in the Santeh station. The other advantage of artificial neural network is the high simulation speed and the accuracy of this model in stream flow simulation.