نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و چهارم، شماره دوم، ۱۳۹۶
http://jwsc.gau.ac.ir

بررسی عملکرد مدل برنامه‌ریزی بیان ژن با روش‌های پیش‌پردازش داده‌ها جهت مدل‌سازی جریان رودخانه

*ابازر سُلگی[1]، حیدر زارعی[2] و محمدرضا گلابی[1]

[1]دانشجوی دکتری گروه مهندسی منابع آب، دانشگاه شهید چمران اهواز، [2]استادیار گروه هیدرولوژی و منابع آب، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۹۵/٤/۲ ؛ تاریخ پذیرش: ۹٦/٤/۲۸

چکیده

سابقه و هدف: نیاز روزافزون به آب سبب گردیده است که برنامه‌ریزی‌های مدیریتی به‌منظور کنترل مصرف آب در آینده از اهمیت بیش‌تری برخوردار باشد. با پیش‌بینی جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، می‌توان حوادث طبیعی مانند سیل و خشکسالی را نیز پیش‌بینی و مهار نمود. به همین دلیل برآورد صحیح و دقیق جریان رودخانه با استفاده از مدل‌های مختلف یکی از موضوعاتی است که در منابع آب مورد بررسی پژوهشگران می‌باشد. مدل‌های هوشمند جهت پیش‌بینی جریان رودخانه توسط پژوهشگران مختلف به‌کار رفته‌اند. یکی از این مدل‌ها که عملکرد خوبی از خود نشان داده است مدل برنامه‌ریزی بیان ژن می‌باشد. اخیراً شیوه استفاده از مدل‌های هوشمند به‌صورت ترکیبی مورد پذیرش قرار گرفته است که جهت انجام این کار معمولاً از تبدیل موجک استفاده می‌شود.

مواد و روش‌ها: در این مطالعه از مدل برنامه‌ریزی بیان ژن (GEP) برای مدل‌سازی جریان در مقیاس‌های روزانه و ماهانه در رودخانه گاماسیاب استفاده شد. برای این منظور از داده‌های بارش، دما، تبخیر و جریان رودخانه گاماسیاب در ایستگاه وراینه با یک دوره آماری ٤۳ ساله (۱۳۹۰-۱۳٤۸) استفاده شد. برای افزایش عملکرد مدل از دو روش پیش‌پردازش داده‌ها یعنی تبدیل موجک (Wavelet Transform) و تجزیه به مؤلفه‌های اصلی (PCA) استفاده شد. بدین‌صورت که سیگنال اولیه هر یک از پارامترهای ورودی با استفاده از تبدیل موجک تجزیه شد. سپس برای مشخص کردن زیرسیگنال‌های مهم از روش تجزیه به مؤلفه‌های اصلی استفاده شده و زیرسیگنال‌های مهم به‌عنوان ورودی به مدل برنامه‌ریزی بیان ژن وارد شد تا مدل ترکیبی برنامه‌ریزی بیان ژن- موجک (WGEP) حاصل گردید.

یافته‌ها: بررسی ساختارهای مختلف برای مدل برنامه‌ریزی بیان ژن نشان داد که عملکرد مدل در دوره روزانه خوب بوده ولی در دوره ماهانه عملکرد کاهش یافته است. مقایسه مدل ترکیبی برنامه‌ریزی بیان ژن- موجک با مدل برنامه‌ریزی بیان ژن نشان داد که عملکرد مدل ترکیبی در هر دو دوره زمانی روزانه و ماهانه از مدل ساده بهتر بوده است. دلیل این امر به‌خاطر پیش پردازشی است که روی داده‌ها پیاده شده بود. این در حالی است که نتایج مدل ترکیبی در دوره روزانه حدود ٤ درصد و در دوره ماهانه ۲۳ درصد ضریب تعیین مدل را افزایش داد. همچنین با توجه به تعداد زیاد زیرسیگنال‌ها به‌کار بردن روش PCA باعث افزایش سرعت اجرای برنامه شد.

* مسئول مکاتبه: a-solgi@phdstu.scu.ac.ir

نتیجه‌گیری: استفاده از روش‌های پیش‌پردازش داده‌ها باعث افزایش عملکرد مدل شده است و استفاده از روش PCA به‌عنوان یک ابزار کمکی برای تبدیل موجک موجب افزایش سرعت و دقت مدل شده است. به‌طورکلی نتایج این مطالعه نشان داد که می‌توان از ترکیب مدل برنامه‌ریزی بیان ژن با تبدیل موجک به‌عنوان یک ابزار مناسب برای مدل‌سازی و پیش‌بینی جریان رودخانه گاماسیاب بهره برد.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، پیش‌پردازش داده‌ها، تبدیل موجک، روش PCA، مدل‌سازی جریان

مقدمه

با توجه به محدودیت منابع آب شیرین قابل استحصال، پیش‌بینی هرچه دقیق‌تر جریان و تغییرات آن در طول رودخانه از ارکان اساسی برنامه‌ریزی و مدیریت آب‌های سطحی است. با پیش‌بینی جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، می‌توان حوادث طبیعی مانند سیل و خشکسالی را نیز پیش‌بینی و مهار نمود. این مسأله به‌عنوان یکی از چالش‌های مدیریت منابع آب در دهه‌های اخیر مطرح بوده است. روش‌های مختلفی تاکنون توسط پژوهشگران در سراسر جریان برای میل به این هدف مورد استفاده قرار گرفته است که عمدتاً شامل روش‌های رگرسیونی، مدل‌های مفهومی و روش‌های پیچیده‌تر بر پایه شبکه‌های عصبی مصنوعی و نرو فازی می‌باشد.

کاربرد تجزیه و تحلیل موجک در هیدرولوژی موضوعی است که در قرن بیست و یکم مورد توجه قرار گرفت. شاید بتوان گفت که ناکان (۱۹۹۹) از نخستین کسانی بود که از تجزیه و تحلیل موجک برای مشخص کردن تغییرات زمانی بارش و رواناب و روابط آن‌ها بهره جست (۱۰). در سال‌های بعد دمیانوف و همکاران (۲۰۰۱) از ترکیب تجزیه و تحلیل موجک و ابزار زمین‌آمار (کریجینگ) برای بررسی تغییرات مکانی بارش، سود جستند و نتایج را با مدل ترکیبی شبکه عصبی مصنوعی و کریجینگ مقایسه کردند (۳). همچنین جایاواردن و همکاران (۲۰۰٤) از تجزیه کردن موجک به همراه مدل مارکف برای شبیه‌سازی بارش روزانه حوضه فوریا[1] در تایلند استفاده کردند (٦). نورانی و همکاران (۲۰۱٤) کاربرد مدل‌های ترکیبی هوش مصنوعی- موجک را مورد بررسی قرار دادند. نتایج این بررسی نشان داد که علت مورد توجه قرار گرفتن تبدیلات موجک، به‌خاطر فایده و سودمندی تجزیه و تحلیل تفکیک‌پذیری چندجانبه، حذف اختلالات مربوط به یک امواج الکتریکی یا الکترومغناطیسی[2] و نیز قابلیت قدرتمند هوش مصنوعی در بهینه‌سازی، تطبیق‌پذیری و تخمین فرآیندهاست (۱۱).

برنامه‌ریزی بیان ژن[3] شاخه‌ای از الگوریتم‌های تکاملی است که توانایی مدل‌سازی فرآیندهای غیرخطی و پویا را دارد. در زمینه استفاده از GEP مطالعاتی انجام شده است که در ادامه به تعدادی از آن‌ها اشاره می‌شود: کایزی و همکاران (۲۰۱۱) یک مدل ترکیبی برنامه‌ریزی بیان ژن- موجک را برای پیش‌بینی کوتاه مدت و بلندمدت دمای هوا ارائه دادند. این مطالعه با استفاده از داده‌های روزانه و ماهانه دما در ایستگاه‌های مهاباد و ارومیه در ایران صورت پذیرفته است. نتایج نشان داد مدل ترکیبی WGEP عملکرد بهتری نسبت به مدل GEP دارد (۸). مقایسه برنامه‌ریزی بیان ژن با ANFIS جهت پیش‌بینی کوتاه‌مدت نوسانات عمق آب توسط شیری و کایزی (۲۰۱۱) صورت پذیرفته

1- Chao Phvraya
2- Signal
3- Gene expression programming

است. نتایج نشان داد که هر دو مدل می‌توانند به‌عنوان یک ابزار دقیق برای پیش‌بینی نوسانات عمق آب استفاده شوند. اما مدل GEP دارای فرمول ساختاری ساده و راحت‌تری نسبت به مدل ANFIS می‌باشد (۱۵).

شوئیب و همکاران (۲۰۱۵) پیش‌بینی رواناب را با استفاده از مدل ترکیبی برنامه‌ریزی بیان ژن- موجک[1] (WGEP) برای ٤ حوضه مختلف از سراسر جهان با استفاده از داده‌های بارش- رواناب انجام دادند. ده تابع موجک مادر مختلف مورد بررسی قرار گرفت. نتایج نشان داد که مدل ترکیبی WGEP با استفاده از تابع موجک Dmey دارای عملکرد بهتری نسبت به مدل GEP بوده است (١٦). کریمی و همکاران (۲۰۱۵) پیش‌بینی کوتاه‌مدت و بلندمدت جریان رودخانه فلیوس[2] در ترکیه را با استفاده از مدل ترکیبی برنامه‌ریزی بیان ژن- موجک مورد بررسی قرار دادند. نتایج نشان داد که مدل ترکیبی عملکرد بهتری نسبت به مدل GEP دارد. آن‌ها همچنین از مدل‌هایARMA[3]، ANN و ANFIS برای مقایسه استفاده کردند. نتایج نشان داد که عملکرد مدل ترکیبی از بقیه مدل‌ها بهتر بوده است (۷).

با توجه به اهمیت پیش‌بینی‌های کوتاه‌مدت در مهندسی منابع آب و ویژگی غیرخطی و ناایستای سری زمانی جریان روزانه و ماهانه، استفاده از یک ابزار با دقت مناسب یکی از موضوعاتی است که همیشه مورد نظر پژوهشگران می‌باشد. بنابراین در این مطالعه در جهت نیل به این هدف، سعی شده است از یک ابزار به نام تبدیل موجک که قابلیت خوبی در هیدرولوژی از خود نشان داده است به صورت ترکیب با مدل برنامه‌ریزی بیان ژن استفاده نموده و با مقایسه با مدل برنامه‌ریزی بیان ژن هم میزان تأثیر این ابزار را مورد بررسی قرار داد و هم بتوان در جهت پیش‌بینی جریان رودخانه گاماسیاب که منبع مهمی از نظر شرب و کشاورزی شهرستان نهاوند می‌باشد گام مؤثری برداشت. یکی دیگر از کارهایی که در این مطالعه استفاده شده است مربوط به روش PCA می‌باشد و میزان تأثیر استفاده از این ابزار هم مورد بررسی قرار گرفته است.

مواد و روش‌ها

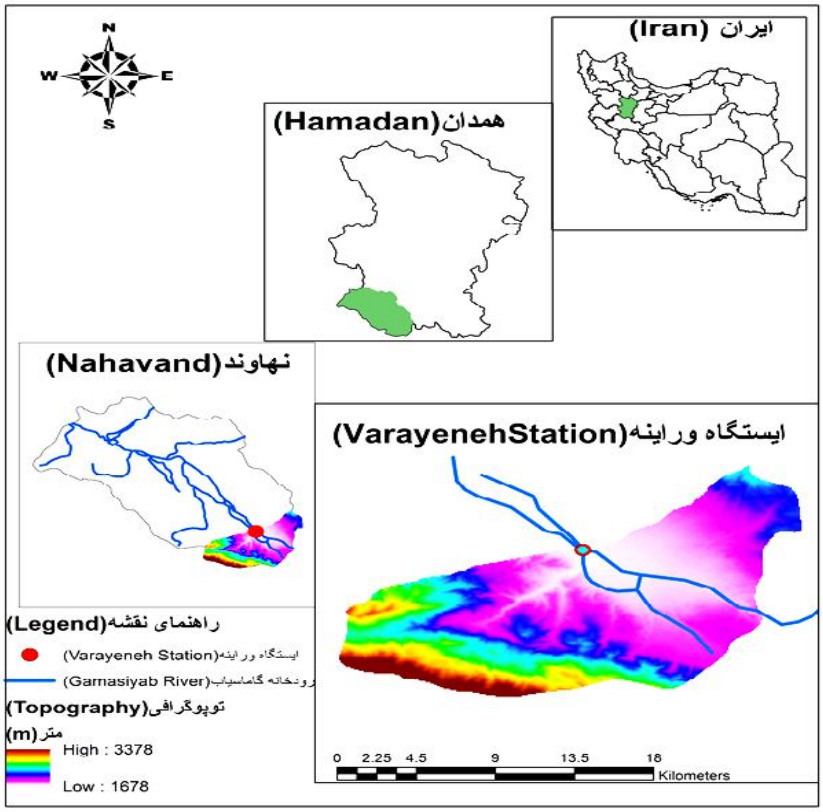
منطقه مورد مطالعه: رودخانه گاماسیاب در غرب کشور، در محدوده استان‌های همدان، کرمانشاه و لرستان واقع شده است. رودخانه گاماسیاب از چشمه‌های کارستی گاماسیاب در ۲۰ کیلومتری جنوب‌غربی شهر نهاوند و در فاصله اندکی از جاده ارتباطی نهاوند به نورآباد لرستان، از ارتفاع ۱۸٦۰ متری از محلی به نام کوه سنگ سوراخ سرچشمه می‌گیرد. حوضه آبریز در این پژوهش به بررسی بخشی از حوضه آبریز رودخانه گاماسیاب از قسمت ابتدا تا نقطه ایستگاه وراینه پرداخته شده است. ایستگاه وراینه در موقعیت جغرافیایی ٤٨ درجه و ٢٤ دقیقه و ١٥ ثانیه طول شرقی و ٣٤ درجه و ٠٤ دقیقه و ٣٢ ثانیه عرض شمالی قرار دارد. این ایستگاه در سال ١٣٤٨ تأسیس شده است و دارای ارتفاع ۱۷۹۵ متر از سطح دریا با میانگین بارش سالانه درازمدت ۵۲۱ میلی‌متر می‌باشد. در این پژوهش از داده‌های بارش، دما، تبخیر و جریان رودخانه گاماسیاب در یک دوره ٤٣ ساله (١٣٩٠-١٣٤٨) در ایستگاه وراینه استفاده شده است (جدول ۱). شکل ۱ موقعیت ایستگاه وراینه را نشان می‌دهد.

1- Wavelet-GEP
2- Filyos
3- Auto regressive moving average

جدول ۱- متغیرهای اقلیمی ایستگاه ورانیه.

Table 1. Climatic variables of Varayeneh station.

مقیاس زمانی Time scale	متغیر اقلیمی Climatic Variable	واحد اندازه‌گیری Unit of Measurement	میانگین Average	حداقل Minimum	حداکثر Maximum
روزانه Daily	جریان Flow	مترمکعب در ثانیه (m^3/s)	3.78	0.67	22.86
	بارش Precipitation	میلی‌متر (mm)	1.43	0.0	96
	دما Temperature	درجه سلسیوس (°C)	11.09	-34	46
	تبخیر Evaporation	میلی‌متر (mm)	4.94	0.0	24.5
ماهانه Monthly	جریان Flow	مترمکعب در ثانیه (m^3/s)	3.76	0.97	12.93
	بارش Precipitation	میلی‌متر (mm)	43.8	0.0	266
	دما Temperature	درجه سلسیوس (°C)	10	-9.10	26.5
	تبخیر Evaporation	میلی‌متر (mm)	150.2	0.0	409.9



شکل ۱- موقعیت ایستگاه ورانیه در شهرستان نهاوند، استان همدان و ایران.

Figure 1. Location of Varayeneh station and Gamaisiab River in Nahavand, Hamedan, Iran.

آماده‌سازی اطلاعات: به‌علت این‌که وارد کردن داده‌ها به‌صورت خام باعث کاهش سرعت و دقت مدل می‌شود. از روش استانداردسازی (نرمال‌سازی) داده‌ها استفاده شده است. با روش استانداردسازی هر عدد تبدیل به عددی بین صفر تا ۱ می‌شود (۱۳). با توجه به پیشنهاد سلگی (۲۰۱٤) از رابطه زیر استفاده شده است (۱۷).

$$y = 0.5 + (0.5 \times \left(\frac{x-\bar{x}}{x_{max}-x_{min}}\right)) \qquad (۱)$$

که در آن، X داده موردنظر، $\overline{X}$ میانگین داده‌ها، X_{max} حداکثر داده‌ها، X_{min} حداقل داده‌ها و y داده نرمال شده می‌باشد. سپس ۷۵ درصد از داده‌ها برای آموزش[1] و ۲۵ درصد برای آزمون[2] مورد استفاده قرار گرفت. ابتدا داده‌های روزانه از سازمان آب منطقه‌ای همدان دریافت شد، با توجه به این‌که داده‌ها به‌صورت کامل موجود بودند بررسی داده‌ها فقط از نظر داده‌های پرت مورد بررسی قرار گرفته و مشکلات موجود برطرف گردید.

تبدیل موجک: به هر کمیت متغیر در زمان یا مکان که قابل اندازه‌گیری باشد، امواج الکتریکی یا الکترومغناطیس (سیگنال) گویند. برای تحلیل سیگنال‌ها، مبدل‌های ریاضی مورد استفاده قرار می‌گیرد تا بتوان اطلاعاتی را که از سیگنال‌های خام به آسانی قابل دسترس نیست، به‌دست آورد. تبدیل موجک یکی از تبدیل‌های ریاضی کارآمد در زمینه پردازش سیگنال است. تبدیل موجک تبدیلی است که سیگنال را به یک مجموعه از توابع اساسی سیگنال تجزیه می‌کند. در حقیقت یک مجموعه تابع اساسی از تأخیر و تغییر در مقیاس موجک مادر به‌دست می‌آید (۱٤). مزیت مهم تبدیل موجکی این است که اطلاعات زمان و فرکانس را به‌طور مؤثری از سیگنال متغیر در زمان استخراج می‌کند. $\psi(t)$ تابع موجک مادر است اگر دارای سه مشخصه تعداد نوسان محدود، بازگشت سریع به صفر در هر دو جهت مثبت و منفی در دامنه خود و میانگین صفر باشد که این‌ها شرط مقبولیت[3] نامیده می‌شود (این سه ویژگی شرط لازم برای این است که تابعی بتواند به‌عنوان تبدیل موجکی عمل کند). تابع موجک $\psi(t)$ به شکل ریاضی زیر تعریف می‌شود (۹).

$$\int_{-\infty}^{+\infty} \psi(t)d(t) = 0 \qquad (۲)$$

$\psi_{(a,b)}(t)$ با استفاده از تأخیر و تغییر مقیاس موجک مادر از رابطه زیر حاصل می‌گردد.

$$\psi_{(a,b)}(t) = |a|^{-0.5}\psi\left(\frac{t-b}{a}\right), \quad a \in R, b \in R, a \neq 0 \qquad (۳)$$

توابع مورد استفاده در تحلیل با دو عمل ریاضی انتقال[4] و مقیاس[5] در طول سیگنال مورد تحلیل، تغییر اندازه و محل می‌یابند و در نهایت ضرایب موجک در هر نقطه از سیگنال (b) و برای هر مقدار از مقیاس (a) با رابطه زیر قابل محاسبه است:

$$\Psi_{a,b}(\mathrm{t}) = \frac{1}{\sqrt{a}}\Psi(\frac{t-b}{a}) \qquad (٤)$$

$$CWT(a,b) = Wf(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\Psi\left(\frac{t-b}{a}\right)dt = \int_{-\infty}^{+\infty} f(t)\Psi_{a,b}dt \qquad (۵)$$

در تحلیل سیگنال از فرم دیگری از WT با نام Discrete Wavelet Transform که به اختصار DWT گفته می‌شود، نیز استفاده می‌شود. در DWT

1- Train
2- Test
3- Admissibility
4- Translation
5- Dilation

پارامترهای انتقال و مقیاس به‌طور غیرپیوسته انتخاب می‌شوند، به‌طوری‌که:

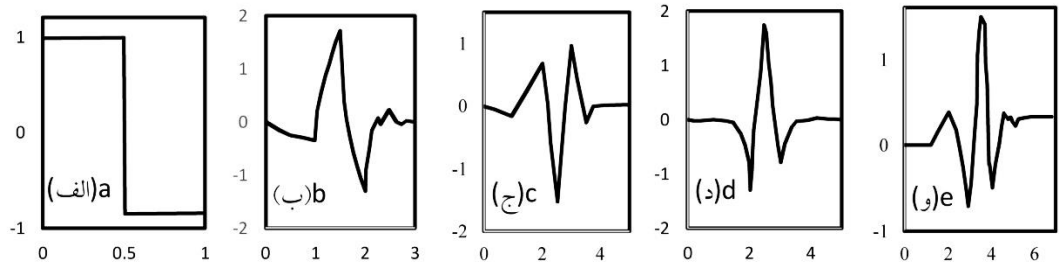
$$a = 2^{-j} , b = 2^{-j}k \qquad (۶)$$

که در آن، j و k اعداد صحیح هستند. در نتیجه با جایگذاری به‌جای a و b رابطه زیر حاصل می‌شود.

$$\Psi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k) \qquad (۷)$$

با توجه به رابطه مشخص‌کننده یک تابع موجک، می‌توان دریافت که توابع بسیاری وجود دارند که دارای این ویژگی‌ها باشند. در سال‌های اخیر، تعداد زیادی توابع موجک بسط داده شده‌اند که در موارد متعددی قابلیت‌های گوناگونی نشان داده‌اند و هر کدام در شاخه‌ای کاربرد گسترده‌تری یافته‌اند.

در این پژوهش با توجه به آزمایش توابع موجک مختلف و توجه به سری زمانی آن‌ها ۵ تابع موجک که در شکل ۲ نشان داده شده‌اند انتخاب شدند.



شکل ۲- الف) موجک هار، ب) موجک Db2، ج) موجک Sym3، د) موجک Coif1 و و) موجک Db4.

Figure 2. a) Haar wavelet, b) Db2 wavelet, c) Sym3 wavelet, d) Coif1 wavelet and e) Db4 wavelet.

برنامه‌ریزی بیان ژن (GEP): برنامه‌ریزی بیان ژن که در ادامه سیر تکاملی مدل‌های هوشمند به‌وجود آمده است جزء روش‌های الگوریتم گردشی محسوب می‌شود که مبنای تمامی آن‌ها براساس نظریه تکامل داروین استوار است (۲). مزیت برنامه‌ریزی بیان ژن نسبت به مدل‌های دیگر از جمله شبکه عصبی مصنوعی این است که در برنامه‌ریزی بیان ژن، ابتدا ساختار (متغیرهای ورودی، هدف و مجموع توابع) تعریف شده و سپس ساختار بهینه مدل و ضرایب طی فرایند آموزش تعیین می‌شوند، در حالی‌که در شبکه‌های عصبی، ابتدا باید ساختار تعیین شده، فقط ضرایب مدل طی فرآیند آموزش حاصل می‌شوند. همچنین این الگوریتم به‌طور خودکار می‌تواند متغیرهای ورودی که در مدل بیش‌ترین تأثیر را دارند، انتخاب کند. فرآیند گام به گام حل یک مسأله با استفاده از برنامه‌ریزی بیان ژن متشکل از ۵ مرحله به‌شرح زیر می‌باشد (٤): ۱- انتخاب مجموعه ترمینال: که همان متغیرهای مستقل مسأله و متغیرهای حالت سامانه می‌باشد. انتخاب تابع برازش در این مرحله صورت می‌گیرد که معمولاً از جذر میانگین مربعات خطا استفاده می‌شود. ۲- انتخاب مجموعه توابع: که شامل عملگرهای ریاضی، توابع آزمون و توابع بولی می‌باشد. عملگرهای ریاضی شامل ۱۰ عملگر ضرب، تقسیم، جمع، تفریق، جذر، لگاریتم، مجذور، مکعب و ... می‌باشد که سه نوع جمع، تفریق و ضرب بیش‌ترین استفاده را دارند. ۳- شاخص اندازه‌گیری دقت مدل که بر مبنای آن می‌توان مشخص کرد که توانایی مدل در حل یک مسأله خاص تا چه اندازه می‌باشد. ٤- مؤلفه‌های کنترل: مقادیر مؤلفه‌های عددی و متغیرهای کیفی که برای کنترل اجرای برنامه استفاده می‌شود. تعداد داده‌های بخش آموزش، تعداد داده‌های بخش آزمون، تعداد کروموزوم‌ها، اندازه

سر[1]، تعداد ژن‌ها، انتخاب عملگر پیوند که با چهار گزینه جمع، تفریق، ضرب و تقسیم قابل تنظیم هستند. ۵- معیارهای توقف برنامه: معیاری برای حصول نتیجه و توقف اجرای برنامه می‌باشد، مثل میزان جمعیت تولید شده، ماکزیمم انطباق و هماهنگی و ....

تجزیه به مؤلفه‌های اصلی (PCA)[2]: تجزیه به مؤلفه‌های اصلی نوعی از تجزیه و تحلیل آماری است که تعداد کم‌تری از عوامل را به نام مؤلفه‌های اصلی از میان عوامل اولیه گزینش می‌کند، به‌طوری‌که تعدادی از اطلاعات کم اهمیت حذف می‌شود. در صورتی که آماره آزمون کیسر- مایر- آلکین KMO[3] مربوط به این روش کم‌تر از ۰/۵ باشد، داده‌ها برای تجزیه و تحلیل عوامل اصلی مناسب نخواهد بود و اگر مقدار آن بین ۰/۵ تا ۰/۶۹ باشد باید با احتیاط بیش‌تر به تجزیه و تحلیل عوامل پرداخت. اما در صورتی‌که مقدار آن بزرگ‌تر از ۰/۷ باشد همبستگی‌های موجود در بین داده‌ها برای تجزیه و تحلیل مناسب خواهد بود (۵). شاخص KMO مورد استفاده در دامنه صفر تا یک قرار دارد. این فاکتور به کمک ضرایب همبستگی ساده و ضرایب همبستگی جزئی[4] از رابطه ۸ به‌دست می‌آید که در این رابطه r_{ij} ضریب همبستگی بین متغیرهای i و j، a_{ij} ضریب همبستگی جزیی (که برآورد ضریب همبستگی جملات خطا) بین i و j با شرط ثابت بودن سایر متغیرها می‌باشد (۵). جهت کسب اطلاعات بیش‌تر در این زمینه به منبع (۱ و ۵) مراجعه شود.

$$KMO = \frac{\sum\sum {r_{ij}}^2}{\sum\sum {r_{ij}}^2 + \sum\sum {a_{ij}}^2} \qquad (۸)$$

ترکیب‌های مختلف جهت مدل‌سازی: در این مطالعه برای اجرای مدل GEP از برنامه GeneXpro Tools(v5) استفاده شده است. در این برنامه انتخاب جمعیت‌های اولیه که همان الگوهای ورودی می‌باشد از اهمیت بالایی برخوردار است. با توجه به این‌که در این مطالعه علاوه بر پارامترهای مختلف ورودی (بارش، دما و تبخیر) توالی جریان روزهای قبل در پیش‌بینی جریان مدنظر بوده است از ترکیب‌های مختلف مطابق جدول ۲ برای مدل GEP استفاده شد. در این جدول پارامترهای Q_t، P_t، T_t و E_t به‌ترتیب جریان، بارش، دما و تبخیر، Q_{t-1}، P_{t-1}، T_{t-1} و E_{t-1} به‌ترتیب جریان، بارش، دما و تبخیر در یک دوره زمانی گذشته، Q_{t-2}، P_{t-2}، T_{t-2} و E_{t-2} به‌ترتیب جریان، بارش، دما و تبخیر در دو دوره زمانی گذشته و Q_{t+1} جریان در دوره آتی می‌باشد.

جدول ۲- ترکیب‌های مختلف مدل برنامه‌ریزی بیان ژن.

Table 2. Combinations details of GEP model.

ترکیب Combination	ورودی Input	خروجی Output
1	Q_{t-1}, Q_t	Q_{t+1}
2	Q_t, P_t	Q_{t+1}
3	$P_{t-1}, P_t, Q_{t-1}, Q_t$	Q_{t+1}
4	P_t, Q_t, E_t, T_t	Q_{t+1}
5	$P_{t-1}, P_t, Q_{t-1}, Q_t, E_{t-1}, E_t, T_{t-1}, T_t$	Q_{t+1}
6	$P_{t-2}, P_{t-1}, P_t, Q_{t-2}, Q_{t-1}, Q_t, E_{t-2}, E_{t-1}, E_t, T_t, T_{t-1}, T_t$	Q_{t+1}

1- Head
2- Principal component analysis
3- Kaiser-Meyer-Olkin
4- Partial correlation coefficients

معیارهای ارزیابی مدل: در این پژوهش برای ارزیابی مدل‌های مختلف از معیارهای زیر استفاده شد.

۱- جذر میانگین مربعات خطا (RMSE)[1]:

$$RMSE = \sqrt{\frac{\sum(Q_{obs} - Q_{pre})^2}{n}} \quad (۹)$$

۲- ضریب تعیین (R^2)[2]:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Q_{obs} - Q_{pre})^2}{\sum_{i=1}^{N}(Q_{obs} - \overline{Q})^2} \quad (۱۰)$$

۳- معیار اطلاعاتی آکائیک (AIC)[3]:

$$AIC = m \times \ln(RMSE) + 2(Npar) \quad (۱۱)$$

که در آن‌ها، پارامترها عبارتند از: n تعداد داده‌ها، Q_o داده‌های مشاهداتی، $\overline{Q}$ میانگین داده‌های مشاهداتی، Q_i داده‌های محاسباتی، m تعداد پارامترها، Npar تعداد داده‌های آموزش‌دیده.

نتایج و بحث

برای اجرای مدل GEP ساختارهای مختلف برای هر یک از ترکیب‌های مختلف مورد بررسی قرار گرفت. در جدول ۳ بهترین ساختار همراه با نتایج حاصل از آن در دو مقیاس زمانی روزانه و ماهانه ارائه شده است.

جدول ۳- بهترین ساختار در ترکیب‌های مدل GEP.

Table 3. The best structure in GEP model combinations.

ترکیب Combination	مقیاس زمانی Time Scale	جذر میانگین مربعات خطا RMSE		ضریب تعیین R^2	
		آموزش Train	آزمون Test	آموزش Train	آزمون Test
1	Daily روزانه	0.0134	0.0130	0.94	0.92
	Monthly ماهانه	0.0527	0.0550	0.71	0.51
2	Daily روزانه	0.0140	0.0132	0.94	0.92
	Monthly ماهانه	0.0526	0.0502	0.71	0.58
3	Daily روزانه	0.0130	0.0126	0.95	0.92
	Monthly ماهانه	0.0512	0.0542	0.73	0.57
4	Daily روزانه	0.0130	0.0127	0.95	0.92
	Monthly ماهانه	0.0513	0.0533	0.72	0.56
5	Daily روزانه	0.0129	0.0125	0.95	0.92
	Monthly ماهانه	0.0433	0.0503	0.80	0.64
6	Daily روزانه	0.0133	0.0135	0.94	0.91
	Monthly ماهانه	0.0494	0.0537	0.75	0.59

1- Root mean square error
2- Coefficient of determination
3- Akaike information criterion

بر اساس این جدول مشاهده می‌شود که در هر دو مقیاس زمانی روزانه و ماهانه ترکیب ۵ بهتر از بقیه ترکیب‌های دیگر بوده است. این یعنی این‌که استفاده از پارامترهای دما و تبخیر علاوه بر پارامترهای بارش و جریان عملکرد مدل را بهبود بخشیده است. با توجه به این جدول مشخص می‌شود که عملکرد مدل در دوره روزانه (R^2=0.92) بسیار بهتر از دوره ماهانه (R^2=0.64) بوده است. با مقایسه مقادیر جذر میانگین مربعات هم این نتیجه تأیید می‌گردد. همچنین با توجه به این جدول و مقایسه نتایج بین ترکیب ۵ و ٦ مشخص شد که افزایش تعداد گام‌های زمانی بیش‌تر از یک گام نتیجه معکوس داشته و باعث کاهش عملکرد مدل می‌شود.

برای اجرای مدل ترکیبی علاوه بر تشخیص تابع موجک باید سطح تجزیه نیز مشخص شود. در دوره ماهانه با پیشنهاد نورانی و همکاران (۲۰۰۹) برای تعیین درجه تجزیه از رابطه زیر استفاده شد (۱۲).

$$L=Int[\log(N)] \qquad (۱۲)$$

که در آن، L سطح تجزیه پیشنهادی، N تعداد سری زمانی می‌باشد. در این مطالعه با تعداد داده ٥١٦ سطح تجزیه برابر ۲ به‌دست آمد که برای دقت بیش‌تر سطح تجزیه ۱ تا ٤ مورد بررسی قرار گرفت. همچنین برای دوره روزانه از درجه تجزیه ۵ تا ۹ استفاده شده است. بعد از تجزیه نمودن سیگنال اصلی پارامترهای ورودی پژوهش، با تعیین زیرسیگنال‌های مهم با روش PCA، از آن‌ها به‌عنوان ورودی به مدل GEP وارد شد تا مدل ترکیبی WGEP حاصل گردید.

نتایج مربوط به روش PCA: بعد از اینکه از تبدیل موجک استفاده شد به‌دلیل این‌که تعداد زیرسیگنال‌های متفاوتی با توجه به نوع تابع موجک و سطح تجزیه حاصل گردید، زیرسیگنال‌های مهم[1] به کمک روش PCA تعیین شد. با توجه به به‌دست آمدن مقدار آماره آزمون KMO برابر ۰/٦۰۷ (جدول ٤)، امکان استفاده از PCA بر متغیرهای مورد استفاده قابل تأیید می‌باشد. برای اجرای PCA پس از استاندارد کردن متغیرهای ورودی، ماتریس متقارن همبستگی R از مرتبه ٤۰ (معادل با بیش‌ترین تعداد ورودی‌ها در دوره روزانه) تشکیل شد. سپس ٤۰ مقدار ویژه و به‌ازای هر مقدار ویژه ٤۰ بردار ویژه حاصل گردید که با استفاده از آن‌ها ٤۰ مؤلفه یعنی به تعداد متغیرهای ورودی حاصل گردید. مشخصات این مؤلفه‌ها در جدول ۵ ارائه شده است. در جدول ۵ اطلاعات مربوط به ارزش هر مؤلفه و درصد پراکندگی از متغیرهای اولیه که توسط هر مؤلفه نشان می‌دهد، ارائه شده است. برای دوره روزانه از بین ٤۰ مؤلفه ورودی با توجه به شکل ۳، ۲۰ مؤلفه اول حدود ۹۵ درصد از پراکندگی و اطلاعات متغیرهای اصلی را نشان می‌دهد. این در حالی است که ۱۲ مؤلفه اول حدود ۸۹ درصد از پراکندگی و اطلاعات متغیرهای اصلی را نشان می‌دهد، بنابراین در دوره روزانه از ۲۰ مؤلفه اول استفاده شد.

1- Main sub-signals

جدول ٤- مقادیر آماره آزمون در روش PCA.

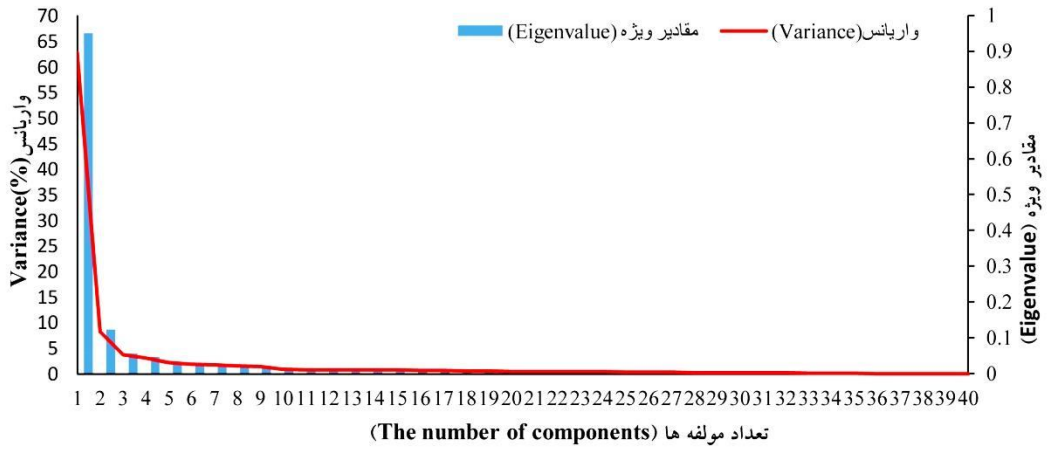
Table 4. The value of the test statistic in PCA method.

0.607	مقدار آماره KMO Kaiser-Meyer-Olkin Measure of Sampling Adequacy	
58878.050	کای اسکوئر Approx. Chi-Square	آزمون کرویت بارتلت Bartlett's Test of Sphericity
780	درجه آزادی (Df)	
0.000	معنی‌داری (Sig.)	

جدول ٥- درصد اطلاعاتی از متغیرهای اولیه که توسط هر مؤلفه بیان می‌شود.

Table 5. The percentage of the primary variables which has been expressed by every component.

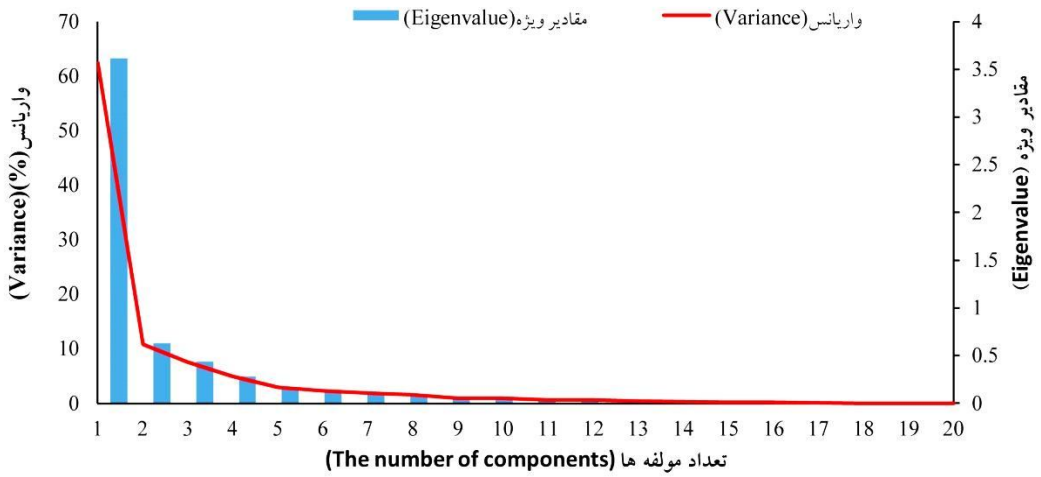
واریانس تجمعی (٪) Cumulate Variance	واریانس (٪) Variance	مقادیر ویژه Eigenvalue	متغیر Variable	واریانس تجمعی (٪) Cumulate Variance	واریانس (%) Variance	مقادیر ویژه Eigenvalue	متغیر Variable
95.57	0.5007	0.0076	21	62.89	62.89	0.9509	1
96.05	0.4814	0.0073	22	71.13	8.233	0.1245	2
96.48	0.4330	0.0065	23	74.90	3.770	0.0570	3
96.92	0.4314	0.0065	24	78.05	3.156	0.0477	4
97.34	0.4199	0.0063	25	80.25	2.195	0.0332	5
97.73	0.3898	0.0059	26	82.15	1.902	0.0288	6
98.09	0.3674	0.0056	27	83.96	1.812	0.0274	7
98.38	0.2884	0.0044	28	85.53	1.570	0.0237	8
98.67	0.2849	0.0043	29	87.03	1.499	0.0227	9
98.91	0.2474	0.0038	30	87.98	0.950	0.0144	10
99.14	0.2307	0.0035	31	88.83	0.848	0.0128	11
99.36	0.2179	0.0033	32	89.66	0.836	0.0126	12
99.56	0.2012	0.0030	33	90.46	0.796	0.0120	13
99.73	0.1643	0.0025	34	91.25	0.790	0.0119	14
99.87	0.1452	0.0022	35	92.03	0.782	0.0118	15
99.94	0.0704	0.0011	36	92.72	0.686	0.0104	16
99.98	0.0335	0.0005	37	93.37	0.652	0.0099	17
99.99	0.0190	0.0003	38	93.99	0.618	0.0094	18
99.99	0.0028	0.0000	39	94.56	0.568	0.0086	19
100	0.0019	0.0000	40	95.07	0.513	0.0078	20



شکل ۳- نمودار اسکری مقادیر ویژه در برابر درصد واریانس و تعداد مؤلفه‌ها (دوره روزانه).

Figure 3. Scree plot of Eigenvalues vs. components along with percentage variances vs. components (daily).

بررسی نتایج در دوره ماهانه نشان داد که مؤلفه اول بیش از ٦٢ درصد اطلاعات متغیرهای اولیه را شامل می‌شود. همچنین ملاحظه شد که ١٢ مؤلفه اول نزدیک به ٩٩ درصد کل پراکندگی و اطلاعات متغیرهای اصلی را بیان می‌کنند. در حالی‌که ٨ مؤلفه اول حدود ٩٥ کل پراکندگی و اطلاعات متغیرهای اصلی را بیان می‌کنند. پس تصمیم بر استفاده از ١٢ مؤلفه اول در این مطالعه برای دوره ماهانه گرفته شد. در شکل ٤ هم مقادیر ویژه و درصد واریانس هر کدام از آن‌ها به‌صورت نموداری نشان داده شده است که نتایج بالا را تأیید می‌کند.



شکل ٤- نمودار اسکری مقادیر ویژه در برابر درصد واریانس و تعداد مؤلفه‌ها (دوره ماهانه).

Figure 4. Scree plot of Eigenvalues vs. components along with percentage variances vs. components (monthly period).

نتایج حاصل از بررسی ساختارهای مختلف مدل WGEP به‌ازای توابع موجک و سطوح تجزیه مختلف در دوره روزانه در جدول ٦ و در دوره ماهانه در جدول ٧ ارائه شده است.

جدول ٦- نتایج مدل WGEP در دوره روزانه به‌ازای توابع موجک در سطوح مختلف.

Table 6. Result of WGEP model in daily period with wavelet functions at different levels.

تابع موجک Wavelet Function	سطح Level	جذر میانگین مربعات خطا RMSE		ضریب تعیین R^2	
		آموزش Train	آزمون Test	آموزش Train	آزمون Test
Coif1	5	0.0110	0.0111	0.96	0.94
Coif1	6	0.0109	0.0112	0.96	0.94
Coif1	7	0.0145	0.0157	0.93	0.88
Coif1	8	0.0133	0.0125	0.94	0.92
Coif1	9	0.0137	0.0128	0.94	0.92
Sym3	5	0.0107	0.0105	0.96	0.95
Sym3	6	0.0102	0.0105	0.97	0.95
Sym3	7	0.0124	0.0125	0.95	0.92
Sym3	8	0.0081	0.0088	0.98	0.96
Sym3	9	0.0158	0.0156	0.92	0.88

ادامه جدول ٦-

Continue Table 6.

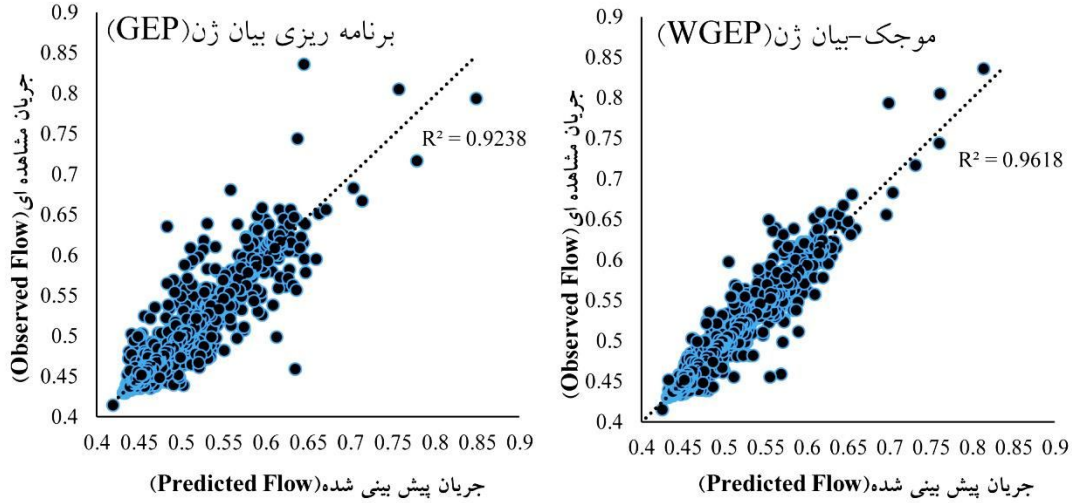
ضریب تعیین R^2		جذر میانگین مربعات خطا RMSE		سطح Level	تابع موجک Wavelet Function
آزمون Test	آموزش Train	آزمون Test	آموزش Train		
0.93	0.96	0.0120	0.0109	5	Haar
0.93	0.96	0.0121	0.0117	6	Haar
0.89	0.93	0.0151	0.0150	7	Haar
0.94	0.97	0.0115	0.0101	8	Haar
0.86	0.91	0.0166	0.0173	9	Haar
0.95	0.96	0.0103	0.0110	5	Db2
0.95	0.97	0.0103	0.0103	6	Db2
0.91	0.94	0.0134	0.0135	7	Db2
0.95	0.97	0.0101	0.0101	8	Db2
0.94	0.95	0.0113	0.0120	9	Db2
0.94	0.96	0.0109	0.0106	5	Db4
0.94	0.97	0.0109	0.0104	6	Db4
0.96	0.98	0.0095	0.0073	7	Db4
0.91	0.94	0.0145	0.0141	8	Db4
0.94	0.96	0.0111	0.0108	9	Db4

جدول ۷- نتایج مدل WGEP در دوره ماهانه به‌ازای توابع موجک در سطوح مختلف.

Table 7. Result of WGEP model in monthly period with wavelet functions at different levels.

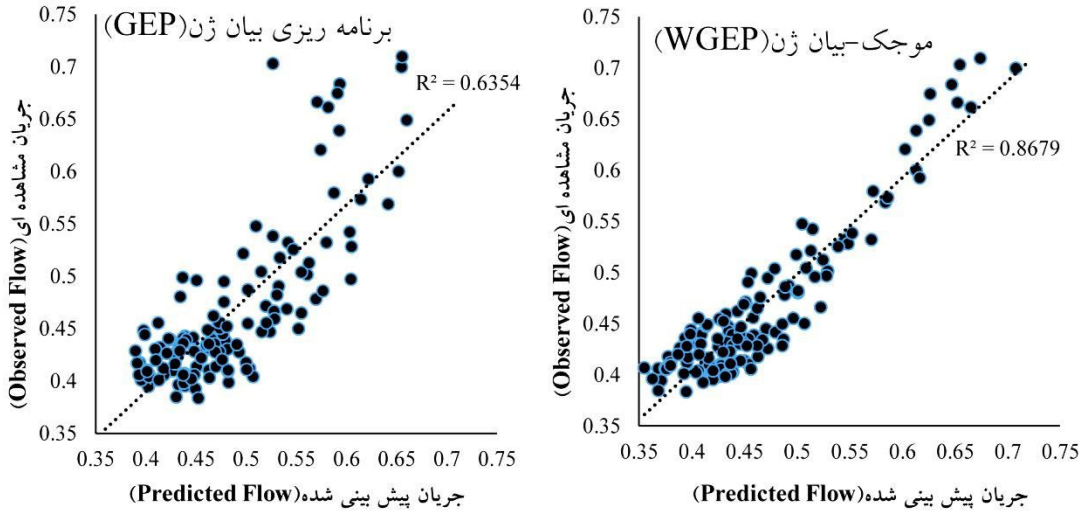
ضریب تعیین R^2		جذر میانگین مربعات خطا RMSE		سطح Level	موجک مادر Mother Wavelet
آزمون Test	آموزش Train	آزمون Test	آموزش Train		
0.65	0.93	0.0452	0.0260	1	Coif1
0.70	0.93	0.0423	0.0258	2	Coif1
0.76	0.93	0.0373	0.0259	3	Coif1
0.68	0.92	0.0432	0.0286	4	Coif1
0.74	0.87	0.0391	0.0357	1	Sym3
0.82	0.96	0.0336	0.0196	2	Sym3
0.83	0.93	0.0317	0.0259	3	Sym3
0.77	0.91	0.0376	0.0298	4	Sym3
0.58	0.90	0.0521	0.0302	1	Haar
0.73	0.90	0.0393	0.0302	2	Haar
0.65	0.90	0.0450	0.0303	3	Haar
0.75	0.92	0.0380	0.0279	4	Haar
0.70	0.96	0.0448	0.0192	1	Db2
0.60	0.93	0.0505	0.0258	2	Db2
0.70	0.93	0.0452	0.0259	3	Db2
0.73	0.90	0.0396	0.0303	4	Db2
0.85	0.94	0.0296	0.0249	1	Db4
0.87	0.96	0.0278	0.0191	2	Db4
0.78	0.96	0.0356	0.0192	3	Db4
0.79	0.95	0.0358	0.0228	4	Db4

براساس جدول ٦ تابع موجک Sym3 در سطح تجزیه ٨ دارای بهترین عملکرد بوده است. این ساختار دارای ضریب تعیین ٠/٩٦ و جذر میانگین مربعات خطای آزمون ٠/٠٠٨٨ می‌باشد. همچنین براساس جدول ٧ تابع موجک Db4 در سطح ٢ در دوره ماهانه دارای ضریب تعیین ٠/٨٧ و جذر میانگین مربعات خطای آزمون ٠/٠٢٧٨ می‌باشد. در شکل‌های ٥ و ٦ مقادیر برآوردی برای مدل‌های مختلف در برابر مقادیر مشاهداتی (به‌صورت استانداردشده) در دوره‌های روزانه و ماهانه نشان داده شده است. مقایسه بین دو مدل GEP و WGEP نشان از برتری مدل ترکیبی WGEP داشته یعنی این‌که استفاده از تبدیل موجک باعث افزایش عملکرد مدل GEP شده است. در شکل ٦ عملکرد هر یک از مدل‌ها در برابر مقادیر مشاهداتی رسم شده است. در این شکل نمودارها برای دوره ماهانه و براساس مقادیر استانداردشده می‌باشد، با مقایسه مدل WGEP با مدل GEP مشاهده می‌شود که مدل ترکیبیWGEP عملکرد بسیار بهتری داشته و در حدود ٢٣ درصد افزایش عملکرد را در برداشته است و این به‌دلیل استفاده از تبدیل موجک و پیش‌پردازشی است که روی داده‌ها صورت گرفته است. پس می‌توان نتیجه گرفت که استفاده از تبدیل موجک در دوره ماهانه به‌طور خیلی قابل آشکار باعث افزایش ضریب تعیین مدل‌ها شده است. در جدول ٨ مقایسه مدل‌های استفاده شده در دو دوره زمانی روزانه و ماهانه براساس معیارهای مختلف ارزیابی مورد بررسی قرار گرفته‌اند.



شکل ٥- مقادیر جریان مشاهده‌ای در برابر مقادیر برآوردی (استانداردشده) برای مدل‌های استفاده شده در دوره روزانه.

Figure 5. The predicted flow versus observed flow (standardized) for the used models in daily period.



شکل ٦- مقادیر جریان مشاهده‌ای در برابر مقادیر برآوردی (استانداردشده) برای مدل‌های استفاده شده در دوره ماهانه.

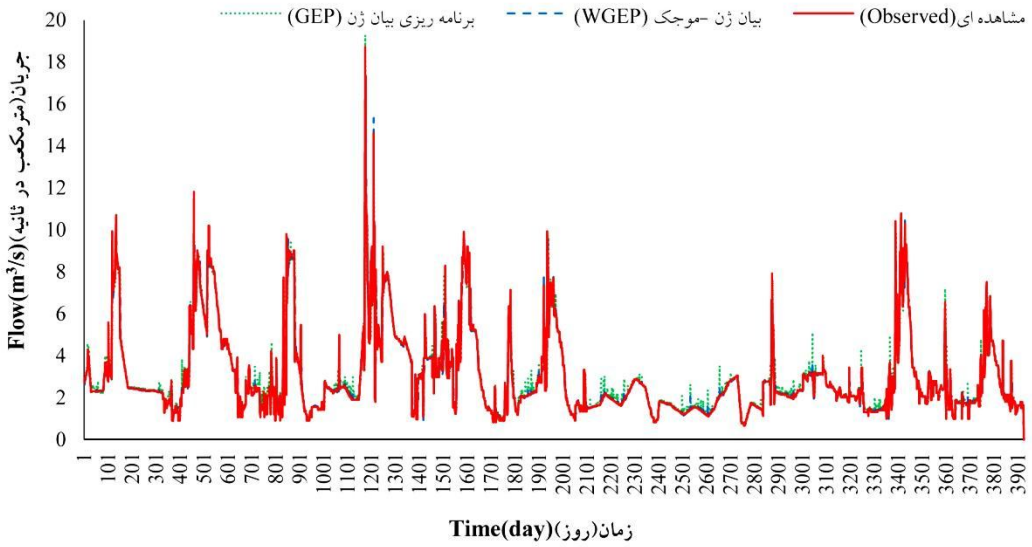
Figure 6. The predicted flow versus observed flow (standardized) for the used models in monthly period.

جدول ۸- مقایسه مدل‌های استفاده شده در این پژوهش.

Table 8. Comparison the models used in this study.

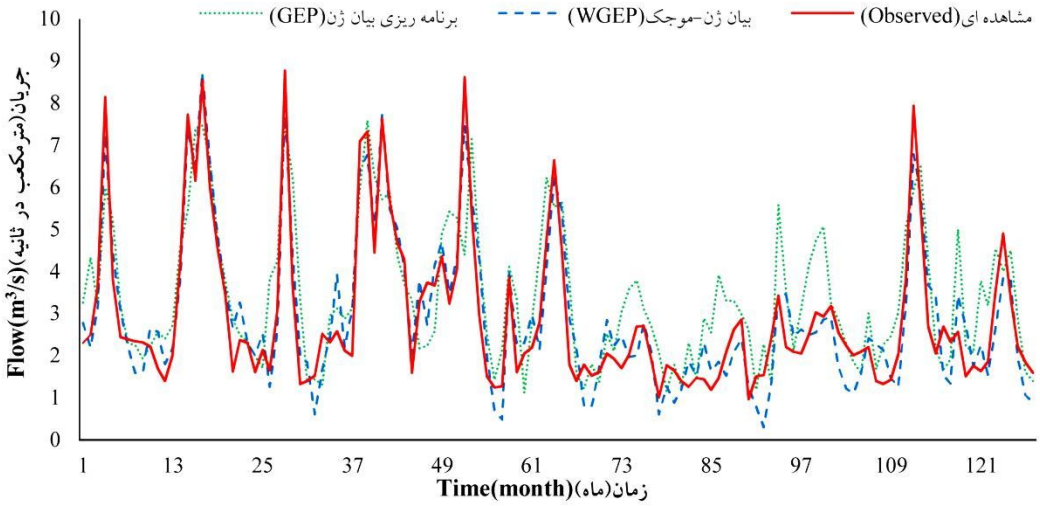
معیار اطلاعاتی آکائیک AIC		جذر میانگین مربعات خطا RMSE		ضریب تعیین R^2		مقیاس زمانی Time Scale	نوع مدل Model Type
آزمون Test	آموزش Train	آزمون Test	آموزش Train	آزمون Test	آموزش Train		
7834.47	23538.61	0.0125	0.0129	0.92	0.95	روزانه Daily	برنامه‌ریزی بیان ژن GEP
244.04	759.44	0.0503	0.0433	0.64	0.80	ماهانه Monthly	
7833.06	23536.74	0.0088	0.0081	0.96	0.98	روزانه Daily	برنامه‌ریزی بیان ژن- موجک WGEP
241.67	756.17	0.0278	0.0191	0.87	0.96	ماهانه Monthly	

بر اساس این جدول و براساس ضریب تعیین در هر دو دوره روزانه و ماهانه مدل WGEP با ضریب تعیین ۰/۹٦ و ۰/۸۷ دارای بهترین عملکرد بوده است. بر اساس معیارهای جذر میانگین مربعات خطا و معیار اطلاعاتی آکائیک مدل WGEP در دو دوره روزانه و ماهانه دارای جذر میانگین مربعات خطا و معیار آکائیک کم‌تری بوده پس مدل برتر شناخته می‌شود. در شکل‌های ۷ و ۸ مقایسه بین مدل‌های مختلف استفاده شده در این پژوهش در دوره‌های روزانه و ماهانه ارائه شده است. با توجه به شکل ۷ از آن‌جایی‌که در دوره روزانه تعداد داده‌ها زیاد بوده عملکرد مدل‌ها به هم نزدیک‌تر بوده ولی مدل ترکیبی WGEP دارای مقادیر نزدیک‌تری به مقادیر مشاهداتی می‌باشد. با توجه به شکل ۸ که مقایسه بین مدل‌ها در حالت ماهانه می‌باشد تفاوت بین مدل‌ها بسیار واضح‌تر بوده و عملکرد خوب مدل ترکیبی در برآورد مقادیر حداقل و حداکثری قابل تأیید می‌باشد. نتایج این پژوهش با نتایج کریمی و همکاران (۲۰۱۵) مبنی بر عملکرد بهتر مدل WGEP نسبت به مدلGEP در پیش‌بینی جریان رودخانه منطبق می‌باشد.



شکل ۷- مقایسه مدل‌های استفاده شده در این پژوهش در دوره روزانه.

Figure 7. Comparison of the models used in this study in daily period.



شکل ۸- مقایسه مدل‌های استفاده شده در این پژوهش در دوره ماهانه.

Figure 8. Comparison of the models used in this study in monthly period.

نتیجه‌گیری

در این مطالعه از مدل برنامه‌ریزی بیان ژن برای مدل‌سازی جریان روزانه و ماهانه رودخانه گاماسیاب نهاوند استفاده شد. نتایج نشان داد که عملکرد مدل برنامه‌ریزی بیان ژن در دوره روزانه مناسب بوده ولی در دوره ماهانه عملکرد مدل، نسبت به دوره روزانه کاهش یافته است. برای بهبود نتایج مدل، از تبدیل موجک و روش PCA استفاده شد. نتایج نشان داد که استفاده از این دو روش باعث بهتر شدن عملکرد مدل در دوره روزانه و ماهانه شده است. در دوره روزانه به‌دلیل زیاد بودن داده‌ها افزایش عملکرد مدل ناچیز و حدود ٤ درصد بوده ولی در دوره ماهانه افزایش

عملکرد بیش‌تر بوده و به ۲۳ درصد رسید. به‌طورکلی استفاده از روش‌های پیش‌پردازش داده‌ها باعث افزایش عملکرد مدل‌ها شده و ترکیب مدل برنامه‌ریزی بیان ژن با تبدیل موجک ابزار مناسبی برای مدل‌سازی و پیش‌بینی جریان رودخانه گاماسیاب می‌باشد.

در کارهای مشابه مربوط به پیش‌بینی جریان رودخانه، می‌توان از این روش‌های پیش‌پردازش داده‌ها استفاده نمود چون این ابزارها باعث افزایش عملکرد مدل می‌شوند و می‌توان مدل‌سازی و پیش‌بینی دقیق‌تری داشت.

منابع

1.Cattel, R.B. 1996. The scree test for the number of the factor. Multivariate Behavioral Research. 1: 245-276.

2.Danandehmehr, A., and Majdzadeh Tabatabai, M.R. 2010. I Prediction of Daily Discharge Trend of River Flow Based on Genetic Programming. J. Water Soil (Iran). 24: 2. 325-333. (In Persian)

3.Demyanov, V., Soltani, S., Kanevski, M., Conu, S., Maignan, M., Savelieva, E., Timonin, V., and Pisaren, K.V. 2001. Wavelet analysis residual kriging Vs. neural network residual kriging. Stochastic Env. Res. Risk Ass. 15: 18-32.

4.Ferreira, C. 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problem. Complex Systems. 13: 87-129.

5.Hutcheson, G., and Nick, S. 1999. The multivariate social scientist: Introductory statistics using generalized linear models. Thousand Oaks, CA, Sage Publications.

6.Jayawardena, A.W., Xu, P., and Tsang, F.L.L. 2004. Rainfall predication by wavelet decomposition. Proceedings of the 2nd Asia Pacific Association of Hydrology and Water Resources Conference, volume II, 5-8, July 2004, Singapore, Pp: 11-19.

7.Karimi, S., Shiri, J., Kisi, O., and Shiri, A.A. 2015. Short-term and long-term streamflow prediction by using 'wavelet–gene expression' programming approach. ISH J. Hydraul. Engin. Pp: 1-15.

8.Kisi, O., Shiri, J., and Nazemi, A.H. 2011. A Wavelet-Genetic Programming Model for Predicting Short-Term and Long-Term Air Temperatures. J. Civil Engin. Urbanism. 1: 1. 25-37.

9.Mallat, S.G. 1998. A wavelet tour of signal processing, San Diego.

10.Nakken, M. 1999. Wavelet analysis of rainfall–runoff variability isolating climatic from anthropogenic patterns. Environmental Modelling & Software. 14: 4. 283-295.

11.Nourani, V., Hosseini Baghanam, A., Adamowski, J., and Kisi, O. 2014. Applications of hybrid Wavelet-Artificial Intelligence models in hydrology, A review. J. Hydrol. 514: 358-377.

12.Nourani, V., Komasi, M., and Mano, A. 2009. A Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling. Water Resour. Manage. 23: 2877-2894.

13.Riad, S., Mania, J., Bouchaou, L., and Najjar, Y. 2004. Rainfall-runoff model usingan artificial neural network approach. Mathematical and Computer Modelling. 40: 7-8. 839-846.

14.Shafaei, M., Fakheri Fard, A., Darbandi, S., and Ghorbani, M.A. 2014. Prediction Daily Flow of Vanyar Station Using ANN and Wavelet Hybrid Procedure. J. Irrig. Water Engin. 4: 24. 113-129. (In Persian)

15.Shiri, J., and Kişi, Ö. 2011. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Computers & Geosciences. 37: 10. 1692-1701.

16.Shoaib, M., Shamseldin, A.Y., Melville, B.W., and Khan, M.M. 2015. Runoff Forecasting using HybridWavelet Gene Expression Programming (WGEP) Approach. J. Hydrol. 527: 326-344.

17.Solgi, A. 2014. Stream flow forecasting using combined Neural Network Wavelet model and comparsion with Adaptive Neuro Fuzzy Inference System and Artificial Neural Network methods (Case study: Gamasyab River, Nahavand). M.Sc. Thesis, Shahid Chamran University of Ahvaz, Iran, 164p. (In Persian)

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 24(2), 2017
http://jwsc.gau.ac.ir

Performance assessment of gene expression programming model using data preprocessing methods to modeling river flow

***A. Solgi[1], H. Zarei[2] and M.R. Golabi[1]**

[1]Ph.D. Student, Dept. of Water Resources Engineering, Shahid Chamran University of Ahvaz,
[2]Assistant Prof., Dept. of Hydrology and Water Resources, Shahid Chamran University of Ahvaz
Received: 06/22/2016; Accepted: 07/19/2017

Abstract

Background and Objectives: An increasing need to water causes the importance of planning management in order to control water consumption in the future. River flow prediction, in addition to the management of water resources, can predict natural disasters such as flood and drought. Therefore, an accurate estimation of river flow using different models is an issue which has been considered by different water resource researchers. Intelligent models have been used to predict river flow. One of these models, which have shown appropriate performance, is Gene Expression Programming (GEP). A use of intelligent models in combinations has been lately accepted and for this purpose, the wavelet transform is usually used.

Materials and Methods: In this study, the GEP model was used for modeling flow in the daily and monthly scale in Gamasiyab River. For this purpose, data of precipitation, temperature, evaporation and flow Gamasiyab River in Varayeneh Station was used during the period from 1970 to 2012. To increase the accuracy of the model, two methods of data pre-process, called Wavelet transform and principal components analysis (PCA) and were used in such a way that the primary signal of each input parameter was decomposed using the wavelet transform. Then, to determine main sub-signals, the principal components analysis was used and main sub-signals as inputs were entered into the GEP model to produce Wavelet-Gene Expression Programming (WGEP).

Results: Detection of different structures of the GEP model showed that the performance of the model was good on the daily scale, but in the monthly scale, the performance was reduced. The comparison of the WGEP model with The GEP model showed that the performance of the hybrid model in both of the daily and monthly scale was better than the simple model. It's because of a pre-process which was done on data. The results of the hybrid model, based on the coefficient determination, was increased by 4% on the daily scale and by 23% in the monthly scale. Also, regarding too many sub-signals, using the Principal Components Analysis increased the speed of running.

Conclusion: Using pre-process of data has increased the performance of the model and using the PCA, as an auxiliary tool for the wavelet transform, increased the speed and accuracy of the model. Totally, the results showed that it's possible to use the GEP model with the wavelet transform as a suitable tool for modeling and predicting the flow of Gamasiyab River.

Keywords: Data pre-processing, Flow modeling, Gene expression programming, Wavelet transform, PCA method

* Corresponding Author; Email: a-solgi@phdstu.scu.ac.ir