

## بهبود عملکرد مدل برنامه‌نویسی بیان ژن با استفاده از تبدیل موجک برای تخمین بارش درازمدت شهر رشت

محمد معین فلاحی، بهروز یعقوبی\*، فریبرز یوسفوند و سعید شعبانلو<sup>۱</sup>

(تاریخ دریافت: ۱۳۹۸/۳/۲۸؛ تاریخ پذیرش: ۱۳۹۸/۹/۲۶)

### چکیده

بارندگی مهم‌ترین منبع تأمین آب شرب و کشاورزی ساکنین نواحی مختلف کره زمین محسوب می‌شود. بنابراین شبیه‌سازی و تخمین این پدیده هیدرولوژیکی از اهمیت بسزایی برخوردار است. در این مطالعه برای اولین بار، بارش درازمدت شهر رشت در طی یک دوره ۶۲ ساله از ۱۹۵۶ تا ۲۰۱۷ به صورت ماهانه توسط یک مدل هوش مصنوعی ترکیبی بهینه‌یافته، شبیه‌سازی شد. برای توسعه مدل هوش مصنوعی ترکیبی (WGEP)، مدل برنامه‌نویسی بیان ژن (GEP) و تبدیل موجک (Wavelet transform) ترکیب شدند. در ابتدا، تأخیرهای مؤثر داده‌های سری زمانی با استفاده از تابع خودهمبستگی شناسایی شدند و با استفاده از آنها برای هر یک از مدل‌های GEP و WGEP هشت مدل متفاوت تعریف شد. سپس، نتایج مدل‌های GEP تجزیه و تحلیل شدند و مدل برتر GEP و مؤثرترین تأخیرها معرفی شدند. مقادیر شاخص عملکرد (VAF)، ضریب همبستگی (R) و شاخص پراکندگی (SI) برای مدل برتر GEP به ترتیب مساوی با ۰/۷۶۵، ۰/۵۰۸ و ۰/۷۰۹ محاسبه شدند. علاوه بر این، تأخیرهای شماره (t-1)، (t-2)، (t-3) و (t-12) مؤثرترین تأخیرها بودند. در ادامه، اعضای مختلف موجک‌های مادر نیز بررسی شدند که موجک مادر demy به عنوان بهینه‌ترین انتخاب شد. همچنین، تجزیه و تحلیل نتایج مدل‌های ترکیبی نشان داد که تبدیل موجک عملکرد مدل برنامه‌نویسی بیان ژن را به شکل قابل ملاحظه‌ای بهبود بخشید. استفاده از این موجک باعث افزایش سه برابری شاخص عملکرد مدل WGEP برتر شد. علاوه بر این، شاخص‌های آماری R و MARE برای مدل WGEP برتر به ترتیب مساوی با ۰/۹۳۵ و ۰/۸۶۲ به دست آمدند. همچنین مقادیر SI، VAF و ضریب نش‌ساکلیف برای این مدل به ترتیب برابر با ۰/۲۹۶، ۰/۳۹۴ و ۰/۸۵۸ تخمین زده شدند. نتایج این مطالعه نشان داد که تبدیل موجک عملکرد مدل برنامه‌نویسی بیان ژن را به شکل قابل توجهی افزایش می‌دهد و پیشنهاد می‌شود تبدیل موجک برای بهبود عملکرد سایر الگوریتم‌های هوش مصنوعی در مباحث هیدرولوژیکی مورد استفاده قرار گیرد.

واژه‌های کلیدی: بارندگی، تحلیل حساسیت، شبیه‌سازی، الگوریتم تکاملی، بهینه‌سازی

۱. گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران

\*: مسئول مکاتبات: پست الکترونیکی: behrouz.yaghoubi.h@gmail.com

## مقدمه

در سال‌های اخیر به دلیل تغییر اقلیم و گرمایش کره زمین، مقدار و الگوی بارش در نقاط مختلف دست‌خوش تغییرات محسوسی شده است. لازم به ذکر است که این تغییرات در نواحی خشک و نیمه‌خشک مانند کشور ایران به دلیل خشکسالی‌های پیاپی از اهمیت بسزایی برخوردار است. به‌طور کلی تخمین و شبیه‌سازی بارش در بازه‌های زمانی مختلف در نواحی گوناگون انجام شده است. لازم به ذکر است که در میان مدل‌های عددی مختلف، مدل‌های هوش مصنوعی از جایگاه خاصی برخوردارند. این مدل‌ها دقت بالایی برای تخمین تابع هدف دارند. همچنین، آنها از انعطاف‌پذیری قابل قبولی نیز بهره می‌گیرند و توانایی شبیه‌سازی پدیده‌های پیچیده و غیرخطی را دارند. بنابراین، استفاده از این مدل‌های عددی روز به روز در حال گسترش است که لوری و سچی (۱۰) مقدار بارش - رواناب حوضه جزیره ساردینا در غرب کشور ایتالیا واقع در دریای مدیترانه را با استفاده از یک مدل عصبی شبیه‌سازی کردند. آنها برای طراحی مدل عصبی خود، یک لایه ورودی و دو لایه پنهان در نظر گرفتند و از تابع فعال‌سازی سیگموئید نیز بهره بردند. آنها طی فرایند آموزش و آزمون مدل عصبی گفته شده بیان کردند که این مدل توانایی مناسبی در تخمین پدیده‌های هیدرولوژیکی مانند بارش و رواناب داشت. در ادامه، ساویچ و همکاران (۱۶) توسط شبکه عصبی مصنوعی و برنامه‌نویسی ژنتیک مقادیر بارش در کشور اسکاتلند را پیش‌بینی کردند. آنها نشان دادند که برنامه‌نویسی ژنتیک توانایی شبیه‌سازی مقادیر بارش برای تأخیرهای بیشتر از سه روز را داشت و به‌طور کلی روش‌های هوش مصنوعی دقت بالایی برای تخمین و مدل‌سازی بارندگی دارند. همچنین، ونگ و همکاران (۱۷) دو مدل هوش مصنوعی از قبیل یک مدل شبکه عصبی مصنوعی و یک مدل فازی را برای شبیه‌سازی مقادیر بارش در یک ناحیه مشخص در کشور ایتالیا به کار گرفتند. آنها حدوداً ۲۰ درصد مقادیر مشاهداتی را برای آموزش مدل‌های عددی و ۸۰ درصد باقیمانده را نیز برای صحت‌سنجی نتایج

استفاده کردند. با تجزیه و تحلیل نتایج نشان داده شد که مدل فازی عملکرد بهتری برای تخمین تابع هدف داشت و این مدل مقادیر تابع هدف را با دقت قابل قبولی پیش‌بینی کرد. ناصری و همکاران (۱۴) شبکه عصبی مصنوعی را توسط الگوریتم ژنتیک برای تخمین مقادیر بارش در یک حوضه واقع در کشور استرالیا بهینه‌سازی کردند. آنها همچنین تأخیرهای مختلف را برای توسعه مدل‌های بیشتر استفاده کردند و در انتها مدل برتر که دارای درصد خطای کمتر و همبستگی بالایی بود را معرفی کردند. علاوه بر این، آنها اذعان کردند که تکنیک‌های هوش مصنوعی ابزاری کارآمد برای شبیه‌سازی بارش است. ناگامولا و همکاران (۱۳) تکنیک‌های مختلف شبکه عصبی مصنوعی را برای شبیه‌سازی مقدار بارندگی در کشور سریلانکا استفاده کردند. آنها با ارزیابی عملکرد مدل‌های به کار گرفته شده نشان دادند که مدل عصبی (Ensemble Neural Network) مقادیر بارش‌ها را با دقت بیشتری تخمین زد. همچنین، تجزیه و تحلیل نتایج مدل‌های هوش مصنوعی به کار رفته توسط این محققین نشان داد که این تکنیک‌ها می‌توانند ابزاری با دقت و ارزان برای پیش‌بینی بارش در نظر گرفته شوند. هاردوینایتو و آپاسا (۸) بارش ماهانه در یک بازه زمانی ۲۲ ساله یک حوضه واقع در کشور اندونزی را توسط شبکه عصبی مصنوعی شبیه‌سازی کردند. آنها در انتها بیان کردند که مدل هوش مصنوعی گفته شده مقادیر تابع هدف را با دقت قابل قبولی تخمین زد و شبکه عصبی مصنوعی یک مدل عددی قابل اعتماد برای تخمین مقادیر بارش در بازه‌های زمانی بلندمدت است. علاوه بر این، دابرال و موری (۴) داده‌های سری زمانی بارش در یک بازه ۱۴ ساله حوضه واقع در شرق کشور هندوستان را توسط مدل ساریما (SARIMA) در بازه‌های زمانی روزانه، هفتگی و ماهانه تخمین زدند. آنها تأخیرهای مؤثر را نیز شناسایی کردند و نشان دادند که مدل گفته شده مقادیر بارش‌ها را با دقت خوبی شبیه‌سازی کرد. لازم به ذکر است که آنها بیان کردند این مدل عددی می‌تواند مقادیر بارش‌ها را در بازه‌های زمانی متفاوت پیش‌بینی کند. همچنین، ژانگ و همکاران (۲) با استفاده از یک

موجک در محیط نرم‌افزار متلب اجرا شده و با ترکیب این دو کد کامپیوتری، مدل هیبریدی WGEP ارائه می‌شود.

در ارتباط با نوآوری مقاله لازم به ذکر است که پژوهش حاضر از دو منظر دارای نوآوری است؛ (۱) تاکنون مدل هیبریدی موجک- برنامه‌نویسی بیان ژن برای شبیه‌سازی پدیده بارندگی استفاده نشده است و این مدل برای اولین بار در مدل‌سازی این پدیده هیدرولوژیکی مورد استفاده قرار می‌گیرد. (۲) تاکنون بارش درازمدت شهر رشت در یک دوره ۶۲ ساله توسط یک مدل هوش مصنوعی ترکیبی پیش‌بینی نشده است.

در ابتدا با استفاده از تابع خودهمبستگی تأخیرهای مؤثر مربوط به داده‌های سری زمانی شناسایی خواهند شد. در ادامه، هشت مدل GEP و WGEP توسط این تأخیرها تعریف می‌شوند. علاوه بر این، اعضای مختلف موجک مادر نیز ارزیابی شده و بهترین موجک مادر نیز انتخاب خواهد شد. با تجزیه و تحلیل نتایج مدل‌سازی‌ها، مدل‌های GEP و WGEP برتر به همراه مؤثرترین تأخیرهای ورودی معرفی خواهند شد. لازم به ذکر است که تبدیل موجک دقت مدل برنامه‌نویسی بیان ژن را به شکل قابل ملاحظه‌ای افزایش خواهد داد.

## مواد و روش‌ها

### ناحیه مورد مطالعه

شهر رشت در ۴۹ درجه و ۳۶ دقیقه طول شرقی و ۳۷ درجه و ۱۶ دقیقه عرض شمالی واقع شده است. همچنین شهر رشت با مساحت ۱۸۰ کیلومتر مربع در زمینی مسطح و هموار به ارتفاع به‌طور میانگین ۵ متر از سطح آب‌های آزاد قرار دارد. رشته کوه البرز در قسمت جنوبی شهر رشت واقع شده است. رشت از شمال به دریای خزر و مرداب انزلی، از غرب به رودخانه پسپخان و شفت، صومعه سرا و فومن، از جنوب به بخش سنگر و شهرستان رودبار و از شرق به کوچصفهان و آستانه اشرفیه محدود است. بارندگی‌ها در رشت بر پایه رژیم بارشی سواحل

مدل نرو- فازی تحت عنوان شبکه استنتاج فازی وابسته به خود مقدار بارش- رواناب را برای حوضه مختلف در کشورهای چین، سوئد و استرالیا مدل‌سازی کردند. آنها نشان دادند که مدل نرو- فازی مذکور از انعطاف مناسبی برای شبیه‌سازی بارندگی برخوردار است و این پدیده هیدرولوژیکی به راحتی و با دقت مناسبی توسط این تکنیک هوش مصنوعی شبیه‌سازی می‌شود. سپس، مهر و همکاران (۱۱) یک مدل ترکیبی بهینه‌یافته را برای تخمین مقادیر بارش درازمدت شهر تبریز و ارومیه توسعه دادند. آنها برای بهینه‌سازی مدل هوش مصنوعی رگرسیون بردار پشتیبان از الگوریتم کرم شب‌تاب بهره گرفتند و نشان دادند که مدل هوش مصنوعی گفته شده مقادیر بارش‌ها را با دقت قابل قبولی تخمین می‌زند. همان‌طور که ملاحظه می‌شود اکثر محققین اذعان کرده‌اند که مدل‌های هوش مصنوعی توانایی قابل قبولی برای تخمین و شبیه‌سازی پدیده بارندگی دارند.

لازم به ذکر است که مدل‌سازی پدیده بارش در بازه‌های زمانی درازمدت قادر خواهد بود که دید جامع‌تری به محققین و مهندسين هیدرولوژی و محیط زیست برای بررسی الگوی‌های بارشی ارائه دهد. در مقابل، مدل‌سازی پدیده بارندگی توسط الگوریتم‌های مختلف هوش مصنوعی روز به روز در حال گسترش است زیرا این مدل‌ها بسیار منعطف بوده و توانایی شبیه‌سازی پدیده‌های پیچیده و غیرخطی را دارند. علاوه بر این، مدل‌های هوش مصنوعی دارای دقت مناسبی هستند و استفاده از آنها باعث صرفه- جویی در زمان و هزینه‌های مربوط به مطالعات میدانی خواهد شد.

بنابراین در این مطالعه، برای اولین بار، مقادیر بارندگی شهر رشت در یک بازه زمانی ۶۲ ساله از سال ۱۹۵۶ تا ۲۰۱۷ به صورت ماهانه توسط یک مدل هوش مصنوعی ترکیبی بهینه‌سازی شده، شبیه‌سازی خواهد شد. برای تولید مدل هوش مصنوعی ترکیبی که به اختصار با WGEP نشان داده می‌شود، از مدل برنامه‌نویسی بیان ژن (GEP) و تبدیل موجک (wavelet transform) استفاده می‌شود. به عبارت دیگر، کد کامپیوتری برنامه‌نویسی بیان ژن و تبدیل

جدول ۱. مشخصات هواشناسی شهر رشت

پارامتر	میانگین سالانه	میانگین حداکثر	میانگین حداقل	میانگین سالانه	حداکثر بارندگی	حداقل دمای	حداکثر دمای
مقدار	۱۵/۹ سانتی‌گراد	۲۰/۶ سانتی‌گراد	۱۱/۳ سانتی‌گراد	۱۳۵۹ میلی‌متر	۱۷۰ میلی‌متر	۱۹ - سانتی‌گراد	۴۰ درجه سانتی‌گراد
هیدرولوژیکی	دمای هوا	سالانه دما	سالانه دما	بارندگی	۲۴ ساعته	گزارش شده	گزارش شده

گرفته شود، تابع شایستگی مربوط به برنامه  $i$ ام به صورت زیر محاسبه می‌شود:

$$f_i = \frac{1000}{1 + RRSE_i} \quad (1)$$

با توجه به اینکه مقدار  $RRSE_i$  می‌تواند از صفر تا بی‌نهایت باشد، مقدار تابع شایستگی در دامنه صفر تا ۱۰۰۰ قرار دارد.

تابع شایستگی  $RRSE_i$  از رابطه زیر محاسبه می‌شود:

$$RRSE_j = \sqrt{\frac{\sum_{i=1}^n (p_{ij} - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (2)$$

که  $P_{ij}$  مقادیر پیش‌بینی شده با استفاده از برنامه  $j$ ام برای شایستگی  $i$ ام،  $O_i$  مقادیر مشاهده شده،  $\bar{O}$  و  $n$  مقدار متوسط مقادیر مشاهده شده و تعداد نمونه‌هاست. برخلاف الگوریتم ژنتیک نشان داده شده در برنامه‌نویسی ژنتیک، برنامه‌نویسی بیان ژن به منظور نشان دادن نتایج، از یک رشته با طول ثابت استفاده می‌کند که آن را به صورت درخت تجزیه با اندازه و شکل‌های مختلف، ارائه می‌دهد. طول ژن در برنامه‌نویسی بیان ژن از رابطه زیر محاسبه می‌شود:

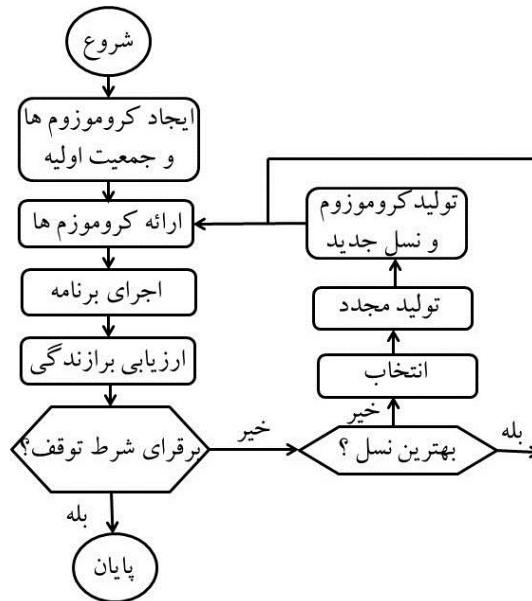
$$l = h + t = h + h \times (n - 1) + 1 \quad (3)$$

که در آن  $l$  طول ژن،  $h$  سر،  $t$  دم و  $n$  تعداد آرگومان‌ها در یک تابع ریاضی که دارای بیشترین آگومان هستند. یکی از مزیت‌های روش برنامه‌نویسی بیان ژن این است که از تنوع ژنتیکی بسیار ساده که به صورت اپراتورهای ژنتیکی در سطح کروموزوم‌های مختلف کار می‌کند، برخوردار است. علوی و گندمی (۱) بیان کردند که از دیگر ویژگی‌های منحصر به فرد برنامه‌نویسی بیان ژن طبیعت چندژنی است که برنامه‌های پیچیده را با استفاده از چند زیربرنامه مختلف، ارائه می‌دهد. روند کار برنامه‌نویسی عبارات ژنتیکی بدین‌گونه است که

دریای خزر، بیشتر متأثر از سامانه‌های پرفشار شمالی است. در این مطالعه، بارندگی ماهانه شهر رشت از سال ۱۹۵۶ تا ۲۰۱۷ در یک بازه زمانی ۶۲ ساله برای اعتبارسنجی نتایج مدل‌های هوش مصنوعی استفاده می‌شود. مشخصات هواشناسی شهر رشت در جدول ۱ ارائه شده است.

#### برنامه‌نویسی بیان ژن (Gene Expression Programming (GEP))

برنامه‌نویسی ژنتیک (Genetic programming (GP)) یکی از روش‌های محاسبات نرم است که به منظور حل مسائل مختلف از اصول انتخاب طبیعی داروین استفاده می‌کند. کوزا (۹) با توسعه الگوریتم ژنتیک، برنامه‌نویسی ژنتیک را ارائه داد. تفاوت اصلی بین الگوریتم ژنتیک و برنامه‌نویسی ژنتیک در نشان دادن نتایج است. برنامه‌نویسی بیان ژن یک گونه خطی از برنامه‌نویسی ژنتیک است که توسط فری‌پرا (۶) ارائه شده است. با توجه به اینکه برنامه‌نویسی ژنتیک، ژنوتیپ و فنوتیپ در یک سیستم تکرار شونده ساده با یکدیگر ترکیب شده‌اند اما برنامه‌نویسی بیان ژن یک سیستم ژنوتیپ/فنوتیپ تکامل‌یافته است که معمولاً ژنوتیپ به‌طور کامل از فنوتیپ مجزا شده است. بنابراین سیستم ژنوتیپ/فنوتیپ تکامل‌یافته برنامه‌نویسی بیان ژن سبب برتری با فاکتور ۱۰۰ تا ۶۰۰۰۰ برابر نسبت به سیستم برنامه‌نویسی ژنتیک است. گندمی و همکاران (۷) نشان دادند که برنامه‌نویسی بیان ژن شامل ۵ بخش اصلی است: تابع شایستگی (fitness function)، مجموعه توابع (function set)، پارامترهای کنترل، مجموعه ترمینال‌ها (terminal set) و شرایط پایان یافتن است. اگر تابع شایستگی مورد استفاده در این مطالعه به صورت ریشه مربعات نسبی خطا ((root relative squared error (RRSE)) در نظر



شکل ۱. ساختار کلی برنامه‌نویسی بیان ژن

فضای تناوب-زمان از هم جدا می‌کند (۵). تبدیل موجک همانند تبدیل فوریه، سری زمانی را به صورت ترکیب خطی از چند تابع اصلی، در نظر می‌گیرد. یکی از مشخصات اصلی تبدیل موجک این است که این تبدیل توانایی به دست آوردن اطلاعات در زمان، تناوب و موقعیت را به طور همزمان دارد. کوهن و کواچویچ (۳) نشان دادند که در تبدیل موجک گسسته تنها یک زیرمجموعه از مقیاس‌ها و موقعیت‌ها به منظور انجام محاسبات، انتخاب می‌شود. تبدیل موجک گسسته یکی سیگنال را به مجموعه ای از توابع، تجزیه می‌کند:

$$\psi_{i,l}(x) = 2^{i/2} \psi_{i,l}(2^i x - l) \quad (4)$$

که  $\psi_{i,l}(x)$  با استفاده از یک موجک دیگر  $\psi(x)$  که با استفاده از  $l$  ترجمه شده است و توسط  $i$  گسترش یافته است، تولید می‌شود. موجک مادر باید شرایط زیر را اکتفا کند:

$$\int \psi(x) dx = 0 \quad (5)$$

موجک گسسته یکی سیگنال به صورت زیر محاسبه می‌شود:

$$c_{i,l} = \int_{-\infty}^{+\infty} f(x) \psi_{i,l}^*(x) dx \quad (6)$$

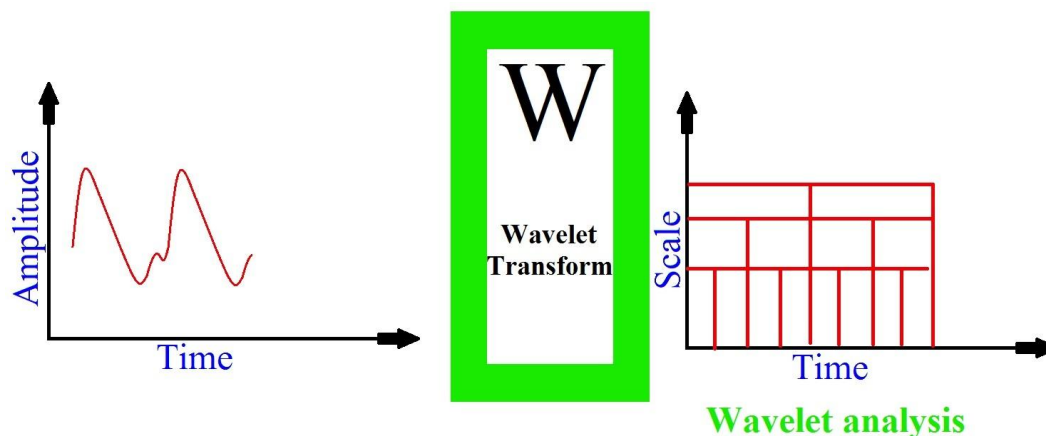
که:

$$f(x) = \sum_{i,l} c_{i,l} \psi_{i,l}(x) dx \quad (7)$$

ابتدا توابع مورد نیاز برای ایجاد مدل و مجموعه ترمینال انتخاب می‌شوند. در مرحله بعد مجموعه داده‌های موجود به منظور برآورد پارامتر مورد نظر و مقایسه آن با مقدار واقعی، فراخوانی می‌شود. سپس به منظور ارائه جمعیت اولیه به صورت تصادفی، کروموزوم‌ها تولید می‌شوند. در مرحله بعد برای جمعیت تولید شده با استفاده از کروموزوم‌های موجود، برنامه اجرا می‌شود و برازندگی تابع هدف بررسی می‌شود. اگر به شرایط توقف برنامه ختم شود، برنامه را متوقف می‌کنیم در غیر این صورت با استفاده از کروموزوم‌های جدید که به وسیله عملگرهای ژنتیکی اصلاح شده‌اند و جمعیت جدید، دوباره تابع هدف ارزیابی می‌شود. این عمل تا زمانی ادامه می‌یابد که شرایط توقف برنامه مهیا شود. شکل ۱ روند مدل‌سازی برنامه‌نویسی بیان ژن را به صورت نمادین نشان می‌دهد.

### تبدیل موجک

میستی و همکاران (۱۲) بیان کردند که موجک تبدیلی است که برای تجزیه سری‌های زمانی مورد استفاده قرار می‌گیرد. این روش یک آنالیز طیفی مستقل از زمان است که به منظور توصیف زمان-مقیاس فرایندها و روابط آنها، سری‌های زمانی را در



شکل ۲. نحوه عملکرد تبدیل موجک

مختلف بر پیش‌بینی بارش در شهر رشت، ارائه می‌شود. برای ارائه این مدل‌ها، از تابع خودهمبستگی (Autocorrelation Function) استفاده می‌شود (شکل ۳). به عبارت دیگر، داده‌های سری زمانی دارای ۱۲ تأخیر متناظر با دوازده ماه سال هستند که در هر سال به صورت متوالی بارش‌ها به صورت تناوبی تکرار می‌شوند. با استفاده از تابع خودهمبستگی مؤثرترین تأخیرها شناسایی می‌شوند زیرا برخی از تأخیرها تأثیر قابل توجهی روی سیکل بارش دارند و در مقابل برخی تأثیر قابل ملاحظه‌ای ندارند. با توجه به تأثیر قابل توجه تأخیرهای ۱، ۲ و ۳ در این نمودار و همچنین دوره‌ای بودن پارامتر بارش، ترکیبات مختلفی از تأخیرهای شماره ۱، ۲، ۳ و ۱۲ به شرح زیر ارائه می‌شود:

$h(t) = f(h(t-1))$	M1
$h(t) = f(h(t-1), h(t-2))$	M2
$h(t) = f(h(t-1), h(t-3))$	M3
$h(t) = f(h(t-1), h(t-12))$	M4
$h(t) = f(h(t-1), h(t-2), h(t-3))$	M5
$h(t) = f(h(t-1), h(t-2), h(t-12))$	M6
$h(t) = f(h(t-1), h(t-3), h(t-12))$	M7
$h(t) = f(h(t-1), h(t-2), h(t-3), h(t-12))$	M8

حال که نمونه‌های آموزشی به همراه مدل‌های مختلف تعیین شدند، باید نوع موجک مادر و سطح تجزیه سری‌های زمانی نیز مشخص شوند. از نکات اساسی در انتخاب موجک مادر، نوع سری زمانی و طبیعت رخداد مورد نظر (در این مطالعه

که سیگنالی است که شکل ضرایب تابع مقیاس را  $\phi(x)$  تخمین می‌زند و موجک مادر  $\phi(x)$  به شکل زیر ارائه می‌شود:

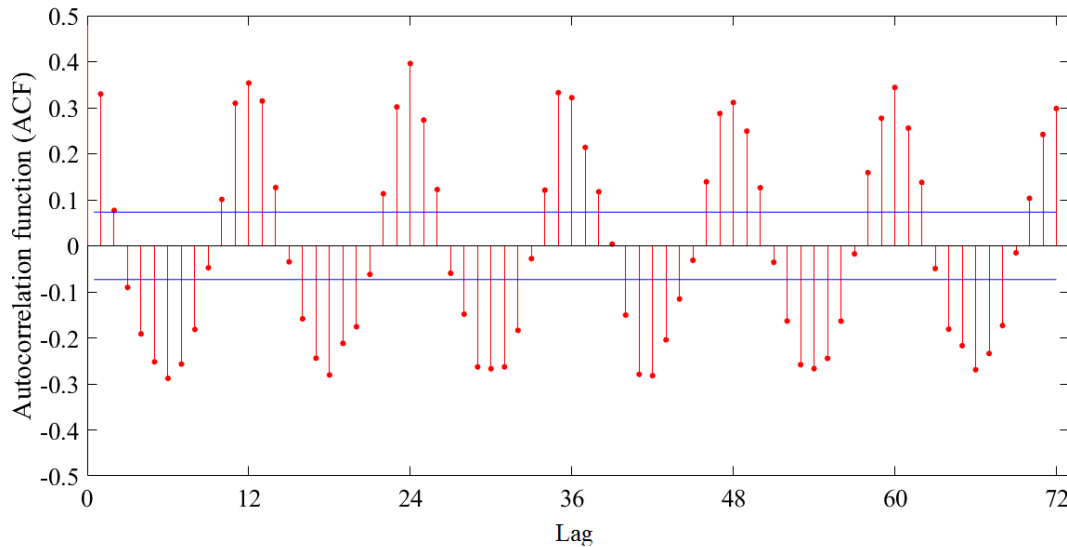
$$\phi(x) = \sqrt{2} \sum k_0(n) \phi(2x - n) \quad (8)$$

$$\psi(x) = \sqrt{2} \sum k_1(n) \phi(2x - n) \quad (9)$$

در اینجا  $k_1 = (-1)^n h_0(1-n)$  است. با توجه به نوع موجک با مشخصات مختلف، مجموعه‌های مختلفی از ضرایب  $k_0(n)$  در نظر گرفته می‌شود. در شکل ۲ نحوه عملکرد تبدیل موجک به تصویر کشیده شده است.

#### پیش‌بینی بارش با استفاده از WGEP

برای تولید مدل ترکیبی موجک- برنامه‌نویسی بیان ژن، کد کامپیوتری برنامه‌نویسی بیان ژن (GEP) و تبدیل موجک (W) در محیط نرم‌افزار متلب اجرا شدند و با ترکیب این دو کد کامپیوتری، مدل هیبریدی WGEP ارائه شد. برای مدل‌سازی مقدار بارش در ماه‌های مختلف مربوط به شهر رشت با استفاده از روش WGEP، ابتدا داده‌ها باید تقسیم‌بندی شوند. مجموع داده‌های مورد استفاده در این مطالعه برابر با ۷۴۴ ماه پشت سر هم از سال ۱۹۵۶ تا ۲۰۱۷ است. تعداد ۵۰۴ ماه اول برای آموزش مدل انتخاب می‌شود و باقیمانده ماه‌ها که ۲۰ سال را دربرمی‌گیرد (۲۰۱۷-۱۹۹۸) برای بررسی عملکرد WGEP به عنوان داده‌های آزمون در نظر گرفته می‌شوند. در ادامه، مدل‌های مختلفی برای بررسی اثر تأخیر زمانی‌های



شکل ۳. دیاگرام ACF برای داده‌های بخش آموزش مدل

$$R = \frac{\sum_{i=1}^n (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (F_i - \bar{F})^2 \sum_{i=1}^n (O_i - \bar{O})^2}} \quad (8)$$

$$VAF = \left( 1 - \frac{\text{var}(F_i - O_i)}{\text{var}(F_i)} \right) \times 100 \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2} \quad (10)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (11)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|F_i - O_i|}{O_i} \right) \quad (12)$$

$$NSC = 1 - \frac{\sum_{i=1}^n (O_i - F_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (13)$$

در اینجا  $O_i$  مقادیر مشاهداتی،  $F_i$  مقادیر پیش‌بینی شده به‌وسیله مدل‌های عددی،  $\bar{O}$  میانگین مقادیر مشاهداتی و  $n$  برابر تعداد مقادیر مشاهداتی هستند. لازم به‌ذکر است که نتایج مدل‌های هوش مصنوعی برای شرایط آزمون مورد ارزیابی قرار می‌گیرد. در ادامه، ابتدا دقت مدل‌های GEP مختلف بررسی می‌شوند. در ادامه، کلیه اعضای خانواده‌های موجک نیز مورد ارزیابی قرار گرفته و بهینه‌ترین عضو شناسایی خواهد شد. در

بارندگی) است. در واقع، تابع موجک باید به‌نحوی انتخاب شود که از لحاظ شکل هندسی به‌خوبی بر داده‌های سری زمانی مورد بررسی منطبق شود تا عمل نگاشت را با دقت بالاتری انجام دهد. توابع موجک مادر مورد استفاده در این مطالعه به‌شرح db, haar, sym, coif, bior, dmey و همکاران (۱۵) بیان کردند که برای تعیین سطح تجزیه از رابطه زیر استفاده می‌شود:

$$l = \text{int}[\log(N)] \quad (7)$$

که  $l$  برابر است با سطح تجزیه،  $N$  تعداد نمونه‌ها در حالت آموزش و  $\text{int}$  قسمت صحیح مقدار  $l$  است. با توجه به اینکه مقدار  $N$  برابر با ۵۰۴ است، بنابراین سطح تجزیه برابر با ۲ در نظر گرفته می‌شود.

#### معیارهای بررسی دقت مدل‌های عددی

در این مطالعه برای بررسی عملکرد و دقت مدل‌های عددی از شاخص‌های آماری مختلفی از جمله ضریب همبستگی ( $R$ )، شاخص عملکرد ( $VAF$ )، خطای مربعات میانگین ریشه ( $RMSE$ )، شاخص پراکندگی ( $SI$ )، درصد خطای نسبی میانگین ( $MARE$ ) و ضریب بهره‌وری نش ساتکلیف ( $NSC$ ) به‌صورت زیر استفاده می‌شود:

GEP6 مقادیر پارامترهای RMSE و VAF به ترتیب برابر با ۷۶/۰۹۰ و ۲۵/۵۲۰ بودند. این درحالی است که شاخص‌های آماری ضریب همبستگی و شاخص پراکندگی برای این مدل نیز به ترتیب مساوی با ۰/۵۰۶ و ۰/۷۰۹ محاسبه شدند. لازم به ذکر است که مدل GEP6 مقادیر بارش‌ها را برحسب تأخیرهای شماره (t-1)، (t-2) و (t-12) شبیه‌سازی کرد. باید خاطر نشان شد که مقادیر شاخص‌های MARE، R و NSC برای مدل GEP7 نیز به ترتیب برابر با ۲/۹۴۲، ۰/۴۷۷ و ۲/۳۷۴- تخمین زده شدند. علاوه بر این، مقادیر شاخص عملکرد و پراکندگی برای همین مدل هوش مصنوعی به ترتیب برابر با ۲۲/۷۳۹ و ۰/۷۲۴ به دست آمدند. لازم به یادآوری است که مدل GEP7 تابعی از تأخیرهای (t-1)، (t-3) و (t-12) بود. برای مدل GEP8 نیز مقادیر شاخص‌های VAF، SI و R به ترتیب مساوی با ۲۵/۷۶۵، ۰/۷۰۹ و ۰/۵۰۸ محاسبه شدند. با توجه به نتایج کلیه مدل‌های برنامه‌نویسی بیان ژن، مدل GEP8 مقادیر بارش‌ها را با دقت بیشتری در مقایسه با سایر مدل‌های GEP شبیه‌سازی کرد. برای این مدل، سایر شاخص‌ها از قبیل MARE، RMSE و NSC به ترتیب برابر با ۲/۷۸۰، ۷۶/۰۱۰ و ۲/۰۹۷- محاسبه شدند. این مدل مقادیر تابع هدف را با استفاده از تأخیرهای شماره (t-1)، (t-2)، (t-3) و (t-12) مدل‌سازی کرد.

بنابراین، بر اساس نتایج کلیه مدل‌های GEP، مدل GEP8 به‌عنوان مدل برتر معرفی شد. علاوه بر این، تحلیل حساسیت نشان داد که تأخیرهای شماره (t-1)، (t-2)، (t-3) و (t-12) نیز به‌عنوان مؤثرترین تأخیرها شناسایی شدند. در شکل ۴ نتایج شاخص‌های آماری محاسبه شده برای مدل‌های GEP مختلف قابل مشاهده است. نتایج شاخص‌های آماری مختلف برای مدل‌های GEP در جدول ۲ قابل مشاهده است. همچنین، نمودار پراکندگی و مقایسه میان بارش‌های مشاهده‌ای و مدل‌سازی شده برای مدل برتر (GEP8) در شکل ۵ به‌تصویر کشیده شده است.

انتها، مدل‌های ترکیبی، WGEP، مورد ارزیابی قرار می‌گیرند و بهترین مدل هوش مصنوعی ترکیبی و مؤثرترین تأخیرها معرفی می‌شوند.

## نتایج و بحث

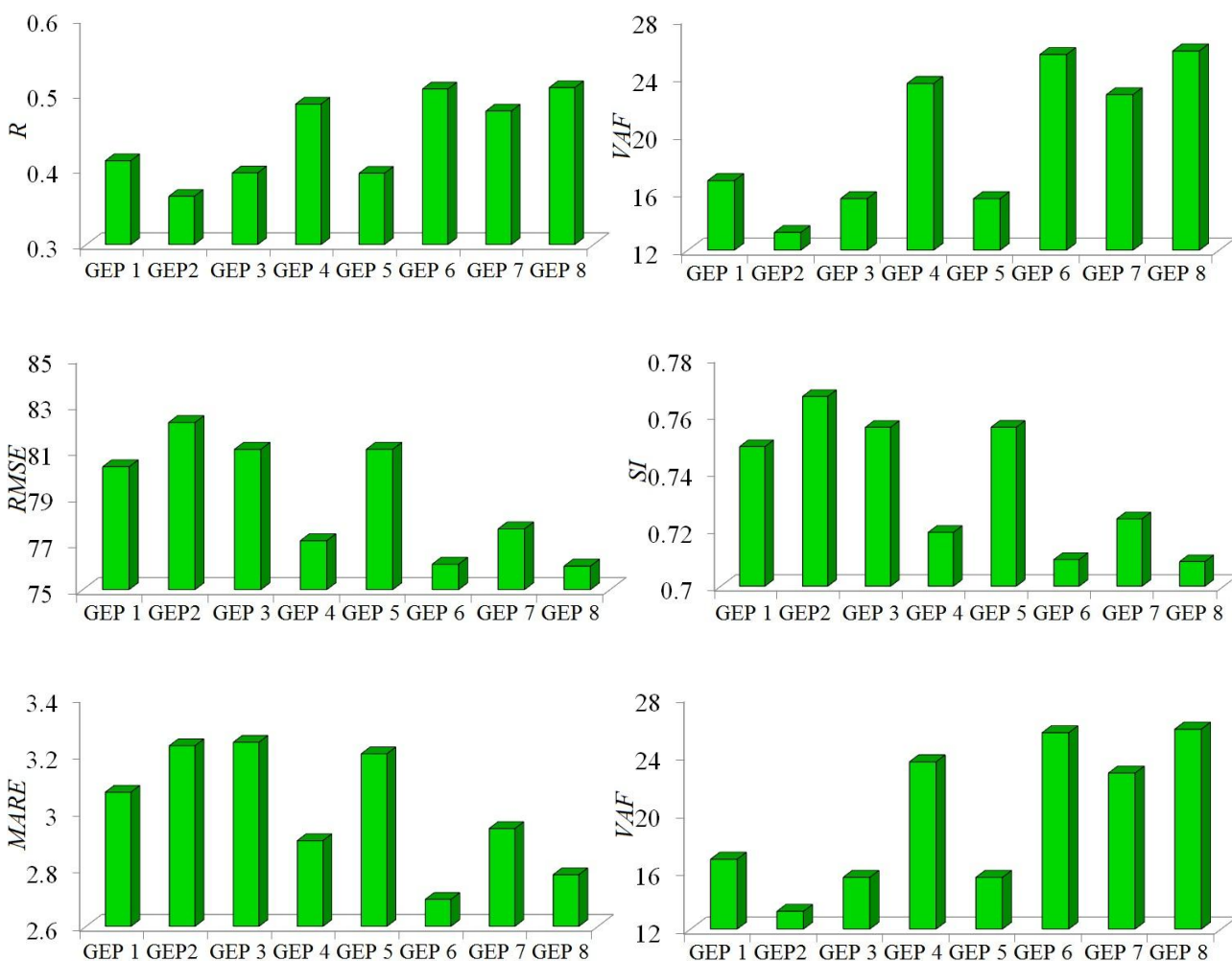
### مدل‌های GEP

در این قسمت دقت مدل‌های GEP مورد بررسی قرار می‌گیرد. مدل 1 GEP تابعی از تأخیر (t-1) بوده و مقادیر R، RMSE و MARE برای آن به ترتیب برابر با ۰/۴۱۱، ۸۰/۲۹۹، ۳/۰۶۹ محاسبه شدند. برای GEP1 مقادیر شاخص‌های آماری VAF و SI به ترتیب مساوی با ۱۶/۸۰۶ و ۰/۷۴۹ به دست آمدند. این درحالی است که مدل GEP2 مقادیر بارش‌ها را برحسب تأخیرهای (t-1) و (t-2) تخمین زد. برای مدل مذکور مقادیر ضریب همبستگی، شاخص عملکرد و شاخص پراکندگی به ترتیب مساوی ۰/۳۶۳، ۱۳/۲۱۲ و ۰/۷۶۶ به دست آمدند. لازم به ذکر است که در میان کلیه مدل‌های GEP، مدل GEP2 کمترین همبستگی را با مقادیر مشاهده‌ای داشت. همان‌گونه که در قسمت قبل بیان شد، مدل GEP3 مقادیر بارندگی را با استفاده از تأخیرهای (t-1) و (t-3) تخمین زد و شاخص‌های آماری R، NSC و MARE برای این مدل به ترتیب مساوی با ۰/۳۹۵، ۳-/۹۰۱ و ۳/۲۲۳ به دست آمدند. در مقابل، برای مدل GEP4 مقادیر شاخص عملکرد و ضریب همبستگی به ترتیب برابر با ۲۳/۵۱۹ و ۰/۴۸۶ تخمین زده شدند. این مدل مقادیر تابع هدف را توسط تأخیرهای (t-1) و (t-12) شبیه‌سازی کرد. همچنین، برای این مدل مقادیر RMSE و SI به ترتیب مساوی با ۷۷/۱۰۹ و ۰/۷۱۹ محاسبه شدند. بر اساس نتایج شبیه‌سازی‌ها، مقادیر R، NSC و MARE برای مدل GEP5 به ترتیب مساوی ۰/۳۹۴، ۴-/۳۹۳ و ۳/۲۰۳ تخمین زده شدند. علاوه بر این، برای مدل مذکور مقادیر شاخص عملکرد و پراکندگی به ترتیب مساوی با ۱۵/۵۵۱ و ۰/۷۵۶ بودند. این درحالی بود که این مدل تابعی از تأخیرهای شماره (t-1)، (t-2) و (t-3) در نظر گرفته شد. علاوه بر این، برای مد g

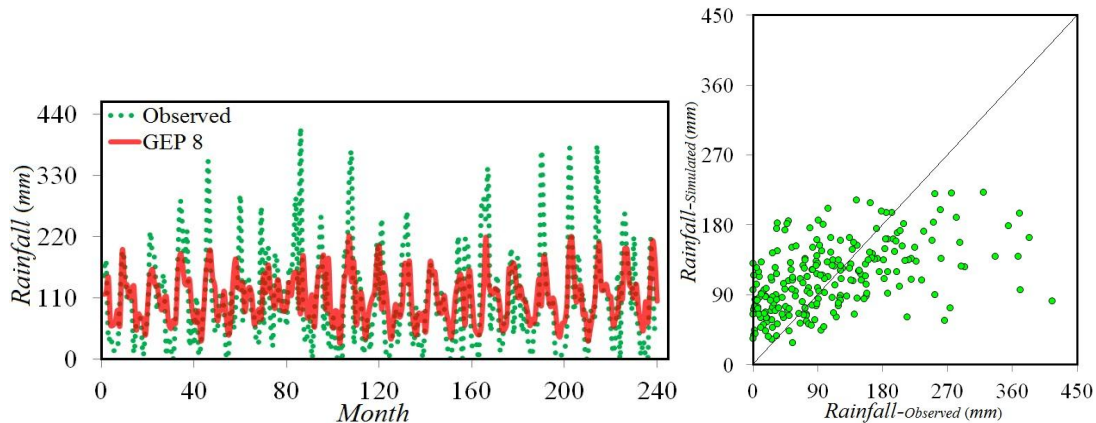


جدول ۲. نتایج شاخص‌های آماری مختلف برای مدل‌های GEP

R	VAF	RMSE	SI	MARE	NSC	
۰/۴۱۱	۱۶/۸۰۶	۸۰/۲۹۹	۰/۷۴۹	۳/۰۶۹	-۴/۷۰۵	GEP1
۰/۳۶۴	۱۳/۲۱۲	۸۲/۲۱۰	۰/۷۶۶	۳/۲۳۱	-۴/۹۳۹	GEP2
۰/۳۹۵	۱۵/۵۵۹	۸۱/۰۵۴	۰/۷۵۶	۳/۲۴۳	-۳/۹۰۱	GEP3
۰/۴۸۶	۲۳/۵۱۹	۷۷/۱۰۹	۰/۷۱۹	۲/۸۹۹	-۲/۷۴۴	GEP4
۰/۳۹۴	۱۵/۵۵۱	۸۱/۰۵۶	۰/۷۵۶	۳/۲۰۳	-۴/۳۹۳	GEP5
۰/۵۰۶	۲۵/۵۲۰	۷۶/۰۹۰	۰/۷۰۹	۲/۶۹۴	-۲/۳۰۲	GEP6
۰/۴۷۷	۲۲/۷۳۹	۷۷/۶۱۶	۰/۷۲۴	۲/۹۴۲	-۲/۳۷۴	GEP7
۰/۵۰۸	۲۵/۷۶۵	۷۶/۰۱۰	۰/۷۰۹	۲/۷۸۰	-۲/۰۹۷	GEP8



شکل ۴. نتایج شاخص‌های آماری مختلف برای مدل‌های GEP



شکل ۵. نمودار پراکندگی و مقایسه بارش‌های مشاهداتی و مدل‌سازی شده توسط مدل GEP8

### خانواده‌های موجک

در ادامه، خانواده‌های مختلف تبدیل موجک مورد ارزیابی قرار می‌گیرند. به‌طور کلی، تبدیل موجک دارای خانواده‌هایی با عنوان داوچیچ (db)، هار (haar)، سایملتس (sym)، کویفلتس (coif)، بایورثوگونال (bior) و دی‌میر (dmey) است. در این بخش، عملکرد این خانواده‌ها مورد سنجش قرار می‌گیرند و مقادیر بارش شبیه‌سازی شده توسط این خانواده‌ها بررسی می‌شوند. در شکل ۶ شاخص‌های آماری محاسبه شده برای این اعضای خانواده‌های تبدیل موجک نشان داده شده است. با بررسی نتایج کلیه خانواده‌ها مشخص شد که موجک مادر دی‌میر (dmey) مقادیر تابع هدف که همان مقادیر بارش است را با دقت بیشتری در مقایسه با سایر موجک‌های مادر تخمین زد. مقادیر RMSE، MARE و NSC برای این موجک مادر به ترتیب مساوی با ۳۱/۷۹۹، ۰/۸۶۲ و ۰/۸۵۸ محاسبه شدند. علاوه بر این، مقادیر شاخص عملکرد، ضریب همبستگی و شاخص پراکندگی برای dmey به ترتیب برابر با ۸۷/۳۹۴، ۰/۹۳۵ و ۰/۲۹۶ تخمین زده شدند. این در حالی است که مقادیر RMSE و NSC برای موجک مادر haar به ترتیب مساوی با ۶۳/۸۹۹ و ۰/۱۳۸ به دست آمدند. همچنین، مقادیر شاخص‌های آماری R، MARE و SI برای عضو برتر خانواده موجک مادر db که db2 بود تقریباً مساوی با ۰/۸۰۶، ۱/۴۶۱ و ۰/۴۹۱ تخمین زده شدند. با بررسی کلیه اعضای خانواده‌های

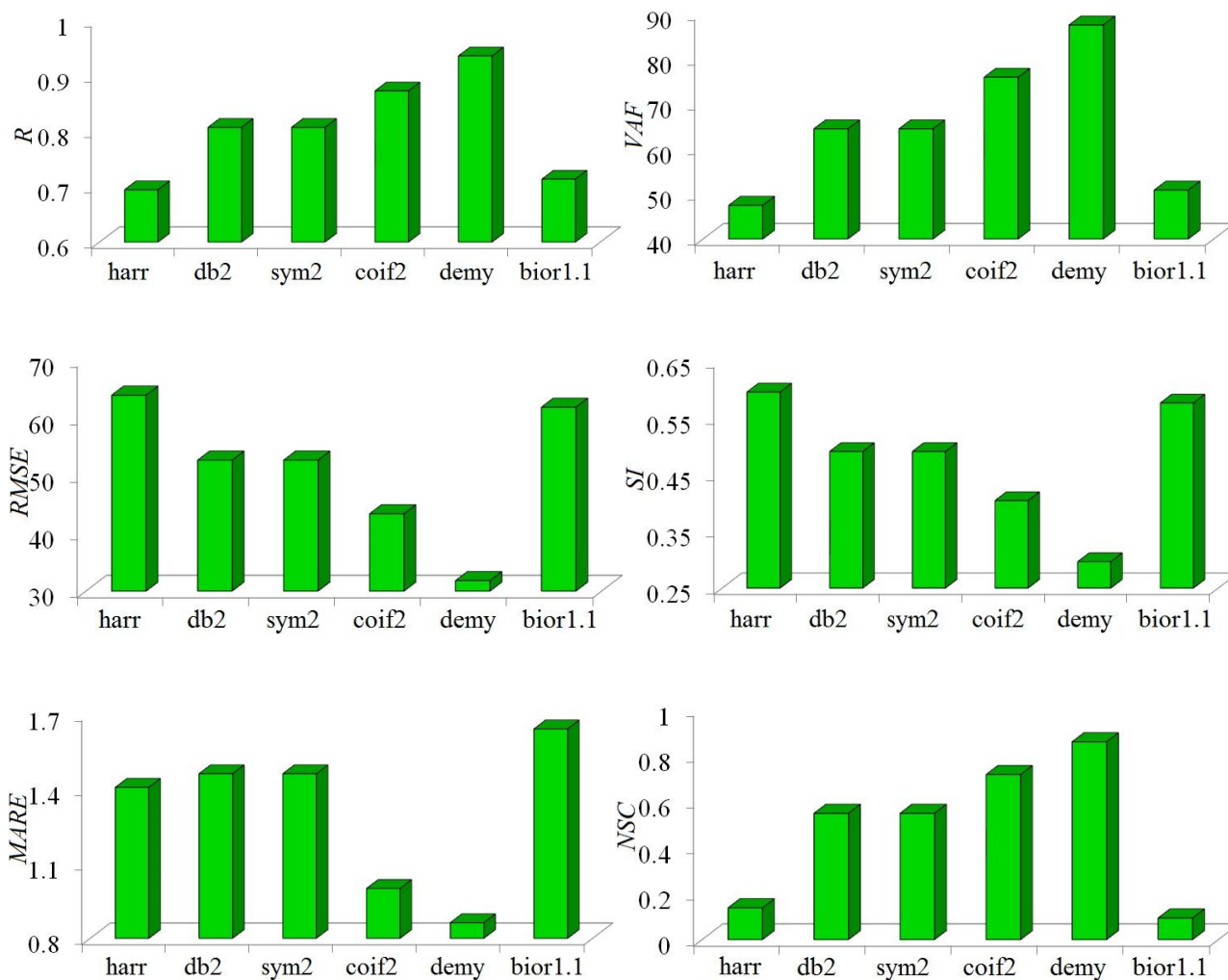
موجک مادر مشخص شد که sym2، coif2 و bior1.1 به ترتیب دقیق‌ترین اعضای موجک‌های مادر خانواده‌های sym، coif و bior بودند. مقادیر شاخص‌های آماری شاخص پراکندگی و عملکرد برای sym2 به ترتیب مساوی با ۰/۴۹۱ و ۶۴/۳۷۸ بودند. این در حالی بود که مقادیر RMSE و NSC برای coif2 نیز به ترتیب برابر با ۴۳/۳۶۲ و ۰/۷۱۶ محاسبه شدند. همچنین، مقادیر شاخص‌های R، VAF و SI برای bior1.1 به ترتیب مساوی با ۰/۷۱۳، ۵۰/۸۱۵ و ۰/۵۷۶ تخمین زده شدند. نتایج شاخص‌های آماری مختلف برای خانواده‌های موجک نیز در جدول ۳ مرتب شده است. بنابراین همان‌گونه که ملاحظه شد، موجک مادر demy به‌عنوان بهینه‌ترین موجک مادر برای شبیه‌سازی مقادیر بارش‌ها شناسایی شد و در ادامه این موجک مادر با مدل برنامه‌نویسی بیان ژن (GEP) ترکیب خواهد شد.

### مدل‌های WGEP

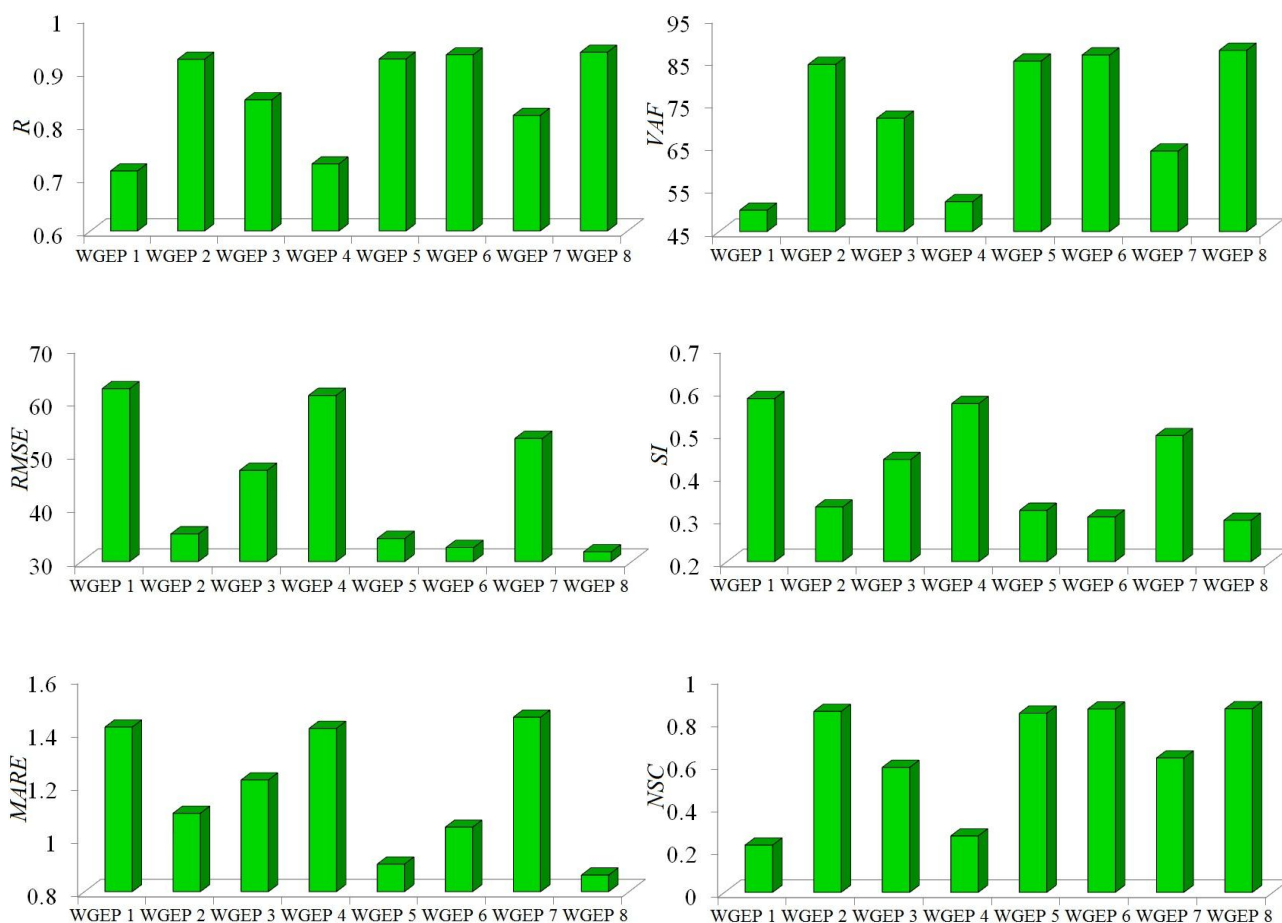
در این قسمت به ارزیابی عملکرد مدل‌های ترکیبی WGEP پرداخته می‌شود. در شکل ۷ نتایج شاخص‌های آماری مختلف برای کلیه مدل‌های ترکیبی WGEP به تصویر کشیده شده است. بررسی عملکرد مدل‌های ترکیبی نشان داد که تبدیل موجک به‌شکل قابل توجهی عملکرد مدل برنامه‌نویسی بیان ژن را بهبود بخشید و بهینه‌سازی این مدل هوش مصنوعی را به‌درستی انجام داده است. مقادیر شاخص‌های آماری MARE، RMSE و NSC

جدول ۳. نتایج شاخص‌های آماری مختلف برای خانواده‌های موجک

R	VAF	RMSE	SI	MARE	NSC	
۰/۶۹۴	۴۷/۴۳۹	۶۳/۸۹۹	۰/۵۹۶	۱/۴۰۶	۰/۱۳۸	haar
۰/۸۰۶	۶۴/۳۷۸	۵۲/۶۴۹	۰/۴۹۱	۱/۴۶۱	۰/۵۴۸	db2
۰/۸۰۶	۶۴/۳۷۸	۵۲/۶۴۹	۰/۴۹۱	۱/۴۶۱	۰/۵۴۸	sym2
۰/۸۷۲	۷۵/۸۰۰	۴۳/۳۶۲	۰/۴۰۴	۱/۰۰۰	۰/۷۱۶	Coif2
۰/۹۳۵	۸۷/۳۹۴	۳۱/۷۹۹	۰/۲۹۶	۰/۸۶۲	۰/۸۵۸	dmey
۰/۷۱۳	۵۰/۸۱۵	۶۱/۸۲۶	۰/۵۷۶	۱/۶۴۱	۰/۰۹۴	bior1.1



شکل ۶. نتایج شاخص‌های آماری مختلف برای خانواده‌های موجک



شکل ۷. نتایج شاخص‌های آماری مختلف برای مدل‌های WGEP

برای مدل WGEP1 به ترتیب مساوی با  $۱/۴۱۸$ ،  $۶۲/۳۵۴$  و  $۰/۲۲۱$  محاسبه شدند. این درحالی است که مقادیر شاخص عملکرد و ضریب همبستگی برای این مدل ترکیبی به ترتیب مساوی با  $۴۹/۹۷۱$  و  $۰/۵۸۱$  محاسبه شدند. بنابراین شاخص عملکرد مدل WGEP1 در مقایسه با مدل GEP1 تقریباً سه برابر افزایش یافت. این درحالی بود که مقادیر پارامترهای R، SI و MARE برای مدل WGEP2 به ترتیب مساوی با  $۰/۳۲۸$ ،  $۰/۹۲۲$  و  $۱/۰۹۴$  تخمین زده شدند. همچنین سایر شاخص‌های آماری از قبیل RMSE و NSC برای این مدل ترکیبی به ترتیب مساوی با  $۳۵/۱۶۶$  و  $۰/۸۴۶$  بودند. لازم به ذکر است که شاخص عملکرد برای WGEP2 برابر با  $۸۴/۰۸۵$  بود که در مقایسه با مدل GEP2 تقریباً شش برابر افزایش یافت. همچنین، مقادیر RMSE و MARE برای مدل WGEP3 به ترتیب مساوی با  $۴۷/۰۶۰$  و  $۱/۲۱۹$  به دست آمدند. علاوه بر این، بررسی‌ها نشان داد که شاخص عملکرد برای این مدل در مقایسه با GEP3 حدوداً پنج برابر بیشتر محاسبه شد. در مقابل برای مدل WGEP4 مقادیر شاخص‌های آماری R و NSC، SI به ترتیب برابر با  $۰/۵۶۹$ ،  $۰/۲۶۴$  و  $۰/۷۲۶$  به دست آمدند. همچنین، مقادیر پارامترهای RMSE و MARE برای این مدل ترکیبی تقریباً برابر با  $۶۱/۰۸۲$  و  $۱/۴۱۲$  تخمین زده شدند. باید خاطر نشان کرد که شاخص عملکرد برای WGEP4 برابر با  $۵۱/۹۷۳$  بود که در مقایسه با مدل ساده GEP4 حدوداً دو برابر افزایش نشان داد. برای مدل WGEP5 مقادیر شاخص‌های آماری R، RMSE و SI به ترتیب مساوی با  $۰/۹۲۲$ ،  $۳۴/۳۱۶$  و  $۰/۳۲۰$  محاسبه شدند. این درحالی بود که سایر شاخص‌های آماری از قبیل NSC و

برای مدل WGEP1 به ترتیب مساوی با  $۱/۴۱۸$ ،  $۶۲/۳۵۴$  و  $۰/۲۲۱$  محاسبه شدند. این درحالی است که مقادیر شاخص عملکرد و ضریب همبستگی برای این مدل ترکیبی به ترتیب مساوی با  $۴۹/۹۷۱$  و  $۰/۵۸۱$  محاسبه شدند. بنابراین شاخص عملکرد مدل WGEP1 در مقایسه با مدل GEP1 تقریباً سه برابر افزایش یافت. این درحالی بود که مقادیر پارامترهای R، SI و MARE برای مدل WGEP2 به ترتیب مساوی با  $۰/۳۲۸$ ،  $۰/۹۲۲$  و  $۱/۰۹۴$  تخمین زده شدند. همچنین سایر شاخص‌های آماری از قبیل RMSE و NSC برای این مدل ترکیبی به ترتیب مساوی با  $۳۵/۱۶۶$  و  $۰/۸۴۶$  بودند. لازم به ذکر است که شاخص عملکرد برای WGEP2 برابر با  $۸۴/۰۸۵$  بود که در مقایسه با مدل GEP2 تقریباً شش برابر افزایش یافت. همچنین، مقادیر RMSE و MARE برای مدل

جدول ۴. نتایج شاخص‌های آماری مختلف برای خانواده‌های موجک

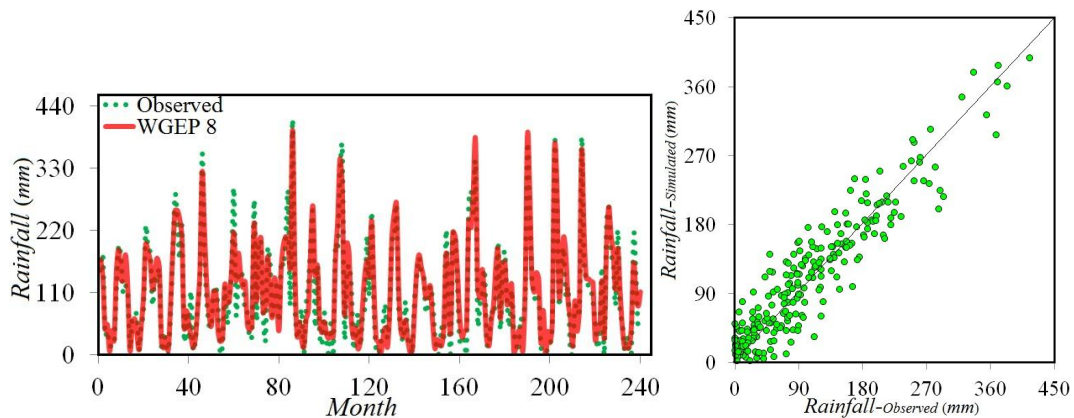
R	VAF	RMSE	SI	MARE	NSC	
۰/۷۱۲	۴۹/۹۷۱	۶۲/۳۵۴	۰/۵۸۱	۱/۴۱۸	۰/۲۲۱	WGEP1
۰/۹۲۲	۸۴/۰۸۵	۳۵/۱۶۶	۰/۳۲۸	۱/۰۹۴	۰/۸۴۶	WGEP2
۰/۸۴۶	۷۱/۵۰۲	۴۷/۰۶۰	۰/۴۳۹	۱/۲۱۹	۰/۵۸۴	WGEP3
۰/۷۲۶	۵۱/۹۷۳	۶۱/۰۸۲	۰/۵۶۹	۱/۴۱۳	۰/۲۶۴	WGEP4
۰/۹۲۲	۸۴/۸۷۸	۳۴/۳۱۶	۰/۳۲۰	۰/۹۰۳	۰/۸۳۸	WGEP5
۰/۹۳۰	۸۶/۳۰۶	۳۲/۶۱۸	۰/۳۰۴	۱/۰۴۲	۰/۸۵۷	WGEP6
۰/۸۱۷	۶۳/۸۱۸	۵۳/۰۴۱	۰/۴۹۴	۱/۴۵۵	۰/۶۲۸	WGEP7
۰/۹۳۵	۸۷/۳۹۴	۳۱/۷۹۹	۰/۲۹۶	۰/۸۶۲	۰/۸۵۸	WGEP8

بنابراین همان‌طور که مشاهده شد، تبدیل موجک عملکرد مدل برنامه‌نویسی بیان ژن را به شکل قابل ملاحظه‌ای افزایش داد. علاوه بر این، در میان کلیه مدل‌های هوش مصنوعی مدل WGEP8 مقادیر بارندگی ماهانه درازمدت شهر رشت را با دقت قابل قبولی شبیه‌سازی کرد. همچنین، نتایج تحلیل حساسیت نشان داد که تأخیرهای شماره (t-1)، (t-2)، (t-3) و (t-12) مؤثرترین تأخیرها شناسایی شدند. نتایج شاخص‌های آماری مختلف برای مدل‌های WGEP در جدول ۴ دیده می‌شود. در شکل ۸، نمودار پراکندگی مدل برتر هوش مصنوعی یا همان WGEP8 به همراه مقایسه مقادیر بارش‌های مشاهده‌ای و شبیه‌سازی شده به تصویر کشیده شده است.

### نتیجه‌گیری کلی

در این مطالعه، برای اولین بار، یک مدل نوین هوش مصنوعی بهینه‌سازی شده که از ترکیب مدل برنامه‌نویسی بیان ژن و تبدیل موجک توسعه یافت، معرفی شد. سپس بارش درازمدت شهر رشت در طی یک دوره ۶۲ ساله از سال ۱۹۵۶ تا ۲۰۱۷ به صورت ماهانه توسط این مدل هوش مصنوعی تخمین زده شد. برای این کار، در ابتدا، داده‌ها به دو دسته آموزش (۲۲ سال) و آزمون (۲۰ سال) تقسیم‌بندی شدند. در ادامه، تأخیرهای

MARE به ترتیب برابر با ۰/۸۳۸ و ۰/۹۰۳ به دست آمدند. بررسی این مدل ترکیبی نشان داد که در مقایسه با مدل ساده GEP5 شاخص عملکرد حدوداً پنج برابر با افزایش همراه شد. همچنین، مقادیر شاخص‌های آماری از قبیل ضریب همبستگی و شاخص پراکندگی برای مدل WGEP6 به ترتیب مساوی با ۰/۹۳۰ و ۰/۳۰۴ بودند. این درحالی بود که مقادیر پارامترهای RMSE و NSC برای این مدل نیز به ترتیب برابر با ۳۲/۶۱۸ و ۰/۸۵۷ به دست آمدند. شاخص عملکرد برای WGEP6 نیز تقریباً برابر با ۸۶/۳۰۶ محاسبه شد که در مقایسه با GEP6 حدوداً سه برابر شد. نتایج مدل ترکیبی WGEP7 نشان داد که شاخص‌های آماری RMSE و SI به ترتیب مساوی با ۷۷/۶۱۶ و ۰/۷۲۴ بودند. همچنین، مقادیر NSC، R و MARE برای مدل WGEP7 به ترتیب مساوی با ۰/۶۲۸، ۰/۸۱۷ و ۱/۴۵۵ تخمین زده شدند. این درحالی بود که شاخص عملکرد WGEP7 در مقایسه با GEP7 حدوداً ۲/۸ برابر با افزایش همراه شد. در مقابل، برای مدل WGEP8 نیز مقادیر RMSE، MARE و SI به ترتیب برابر با ۳۱/۷۹۹، ۰/۸۶۲ و ۰/۲۹۶ به دست آمدند. لازم به ذکر است که شاخص‌های آماری NSC و R برای WGEP8 به ترتیب مساوی با ۰/۸۵۸ و ۰/۹۳۵ بودند و شاخص عملکرد آن در مقایسه با GEP8 حدوداً سه برابر افزایش یافت.



شکل ۸. نمودار پراکندگی و مقایسه بارش‌های مشاهداتی و مدل‌سازی شده توسط مدل WGE8

و مدل‌سازی پدیده بارندگی از اهمیت بسزایی برخوردار است که در این مطالعه با استفاده از یک مدل تکاملی (GEP) و یک مدل ترکیبی (WGE8) شبیه‌سازی شد. مهم‌ترین نتایج این مطالعه حاکی از توانایی قابل قبول مدل ترکیبی در تخمین پدیده بارش بود. به عبارت دیگر تبدیل موجک نقش خود را به عنوان ابزاری که برای تجزیه و تحلیل داده‌های سری زمانی و افزایش دقت مدل هوش مصنوعی تکاملی به خوبی ایفا کرد. پیشنهاد می‌شود مدل‌های هوش مصنوعی بیشتری به همراه تبدیل موجک و یا سایر ابزارهای مشابه ترکیب شده و این پدیده مهم را تخمین و پیش‌بینی کنند زیرا شبیه‌سازی بارندگی در ایران از اهمیت بسزایی برخوردار است. در سال‌های اخیر الگوی بارش در نواحی مختلف این کشور خشک و نیمه‌خشک دست‌خوش تغییرات قابل توجهی شده است. همچنین در نواحی شمالی ایران از جمله استان‌های شمالی مانند گیلان، مازندران و گلستان، بارندگی مهم‌ترین منبع تأمین آب برای مصارف مختلف محسوب می‌شود که اهمیت این موضوع را دوچندان می‌کند.

تأثیرگذار داده‌های مربوط به سری زمانی شامل تأخیرهای شماره  $(t-1)$ ،  $(t-2)$ ،  $(t-3)$  و  $(t-12)$  با استفاده از تابع خودهمبستگی معرفی شدند. همچنین، اعضای مختلف موج‌های مادر بررسی شدند و در انتها بهترین آنها شامل موجک مادر demy انتخاب شد. با استفاده از این تأخیرها و موجک مادر بهینه شده، برای هر یک از مدل‌های برنامه‌نویسی بیان ژن (GEP) و ترکیبی (WGE8) هشت مدل مختلف توسعه داده شد. با تجزیه و تحلیل نتایج مدل‌سازی‌ها، مدل‌های برتر و مؤثرترین تأخیرهای ورودی شناسایی شدند. علاوه بر این، بررسی‌ها نشان داد که تبدیل موجک عملکرد مدل برنامه‌نویسی بیان ژن را به شکل قابل توجهی افزایش داد و این مدل هوش مصنوعی ترکیبی مقادیر بارندگی درازمدت شهر رشت را با دقت قابل قبولی شبیه‌سازی کرد. شاخص عملکرد (VAF) برای مدل برتر GEP مساوی با  $۰/۷۶۵$  و برای مدل برتر WGE8 برابر با  $۸۷/۳۹۴$  محاسبه شدند که نشان‌دهنده بهبود عملکرد مدل ترکیبی در مقایسه با مدل تکاملی ساده GEP بود. برای مدل ترکیبی برتر مقادیر R، SI و NSC به ترتیب مساوی با  $۰/۹۳۵$ ،  $۰/۲۹۶$  و  $۰/۸۵۸$  تخمین زده شدند. به طور کلی تخمین

#### منابع مورد استفاده

- Alavi, A. H. and A. H. Gandomi. 2011. A robust data mining approach for formulation of geotechnical engineering systems. *Engineering Computations* 28(3): 242-274.

2. Chang, T. K., A. Talei, C. Quek and V. R. Pauwels. 2018. Rainfall-runoff modelling using a self-reliant fuzzy inference network with flexible structure. *Journal of Hydrology* 564: 1179-1193.
3. Cohen, A. and J. Kovacevic. 1996. Wavelets: the mathematical background. *Proceedings of the IEEE* 84: 514-522.
4. Dabral, P. P. and M. Z. Murry. 2017. Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes* 4(2): 399-419.
5. Daubechies, I. 1990. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory* 36(5): 961-1005.
6. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems* 13(2): 87-129.
7. Gandomi, A. H., A. H. Alavi, M. R. Mirzahosseini and F. Moqhadas Nejad. 2011. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *Journal of Materials in Civil Engineering* 23(3): 248-263.
8. Hardwinarto, S. and M. Aipassa. 2015. Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan-Indonesia. *Procedia Computer Science* 59: 142-151.
9. Koza, J. 1992. Genetic Programming, on the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, MA.
10. Lorrai, M. and G. M. Sechi. 1995. Neural nets for modelling rainfall-runoff transformations. *Water Resources Management* 9(4): 299-313.
11. Mehr, A. D., V. Nourani, V. K. Khosrowshahi and M. A. Ghorbani. 2019. A hybrid support vector regression-firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology* 16(1): 335-346.
12. Misiti, M., Y. Misiti, G. Oppenheim and J. M. Poggi. 1996. Wavelet Toolbox for Use with Matlab. The Mathworks, Inc.: Natick, Massachusetts, USA.
13. Nagahamulla, H. R., U. R. Ratnayake and A. Ratnaweera. 2012. An ensemble of artificial neural networks in rainfall forecasting. In International Conference on Advances in ICT for Emerging Regions (ICTer2012), IEEE.
14. Nasser, M., K. Asghari and M. J. Abedini. 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Systems With Applications* 35(3): 1415-1421.
15. Nourani, V., M. T. Alami and M. H. Aminfar. 2009. Combined neural-wavelet model for prediction of Ligvanchayi watershed precipitation. *Engineering Applications of Artificial Intelligence* 22: 466-472.
16. Savic, D. A., G. A. Walters and J. W. Davidson. 1999. A genetic programming approach to rainfall-runoff modelling. *Water Resources Management* 13(3): 219-231.
17. Wong, K. W., P. M. Wong, T. D. Gedeon and C. C. Fung. 2003. Rainfall prediction model using soft computing technique. *Soft Computing* 7(6): 434-438.

## Improvement of Gene Expression Programming Model Performance using Wavelet Transform for the Estimation of Long-Term Rainfall in Rasht City

M. M. Fallahi, B. Yaghoubi\*, F. Yosefvand and S. Shabanlou<sup>1</sup>

(Received: June 18-2019; Accepted: December 17-2019)

### Abstract

Rainfall may be considered as the most important source of drinking water and watering land in different areas all over the world. Therefore, simulation and estimation of the hydrological phenomenon is of paramount importance. In this study, for the first time, the long-term rainfall in Rasht city was simulated using an optimum hybrid artificial intelligence (AI) model over a 62 year period from 1956 to 2017. The gene expression programming (GEP) and wavelet transform (WT) were combined to develop the hybrid AI model (WGEP). Firstly, the most effective lags of time series data were identified by means of the autocorrelation function (ACF); then eight various GEP and WGEP models were defined. Next, the GEP models were analyzed and the superior GEP model as well as the most influenced lags was detected. For instance, the variance accounting for (VAF), correlation coefficient (R) and scatter index (SI) for the superior GEP model was calculated to be 0.765, 0.508 and 0.709, respectively. Additionally, lags (t-1), (t-2), (t-3) and (t-12) were the most influenced. Then, the different mother wavelets were examined, indicating that the demy mother wavelet was the most optimal one. Moreover, analyzing the numerical simulations showed that the mother wavelet enhanced the performance of the GEP model significantly. For example, the VAF index for the superior WGEP model was increased almost three times after using the mother wavelet. Furthermore, the R and MARE statistical indices for the WGEP model were computed to be 0.935 and 0.862, respectively.

**Keywords:** Long-term rainfall; Rasht city; Simulation; Wavelet transform; Gene expression programming; Optimization

---

1. Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.

\*: Corresponding author, Email: behrouz.yaghoubi.h@gmail.com