

بهینه‌سازی مدل برنامه‌ریزی بیان ژن توسط تبدیل موجک برای شبیه‌سازی بارش درازمدت شهر انزلی

فرشاد حیاتی، احمد رجبی*، محمد علی ایزدبخش و سعید شعبانلو^۱

(تاریخ دریافت: ۱۳۹۸/۱۱/۲۰؛ تاریخ پذیرش: ۱۳۹۹/۵/۷)

چکیده

تخمین و شبیه‌سازی روند بارندگی در نواحی مختلف جهان به دلیل خشکسالی و تغییر اقلیم از اهمیت فراوانی برخوردار است. در این مطالعه، یک مدل هوش مصنوعی ترکیبی برنامه‌ریزی بیان ژن-موجک (WGEP) برای مدل‌سازی بارندگی درازمدت ۶۷ ساله شهر انزلی برای اولین بار توسعه داده شد. این مدل از ترکیب تبدیل موجک (Wavelet) و برنامه‌ریزی بیان ژن (GEP) به دست آمد. در ابتدا، بهینه‌ترین عضو خانواده تبدیل موجک معرفی شد. سپس با تجزیه و تحلیل نتایج مدل‌سازی، دقیق‌ترین تابع اتصال و برازش برای مدل برنامه‌ریزی بیان ژن به دست آمد. در ادامه، با استفاده از تابع خودهمبستگی و خودهمبستگی نسبی و تأخیرهای مختلف، ۱۵ مدل WGEP توسعه داده شد. مدل‌های WGEP برای بازه‌های زمانی ۳۷، ۲۰ و ۱۰ ساله به ترتیب آموزش، آزمون و صحت‌سنجی شدند. همچنین، با انجام تحلیل حساسیت، مدل برتر و مؤثرترین تأخیرها برای شبیه‌سازی بارش درازمدت شناسایی شدند. مدل برتر مقادیر تابع هدف را با دقت بالایی تخمین زد. به عنوان مثال، مقادیر ضریب همبستگی و شاخص پراکندگی برای این مدل در شرایط صحت‌سنجی به ترتیب برابر با ۰/۹۴۶ و ۰/۳۱۰ محاسبه شدند. علاوه بر این، تأخیرهای شماره ۱، ۲، ۴ و ۱۲ به عنوان مؤثرترین تأخیرها در مدل‌سازی بارش توسط مدل ترکیبی معرفی شدند. همچنین، نتایج مدل برتر ترکیبی با مدل برنامه‌نویسی بیان ژن مقایسه شد که مدل ترکیبی دقت بیشتری داشت.

واژه‌های کلیدی: مدل ترکیبی، بارندگی، سری زمانی، شبیه‌سازی، تحلیل حساسیت

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

*: مسئول مکاتبات: پست الکترونیکی: ahmad.rajabi1974@gmail.com

مقدمه

در دهه‌های اخیر، به‌علت تغییر اقلیم و گرمایش کره زمین میزان بارش در نواحی مختلف جهان دچار تغییرات محسوس شده است. الگوی این تغییرات در نواحی خشک و نیمه‌خشک جهان از جمله فلات قاره ایران از اهمیت بسزایی برخوردار است. علاوه بر این، به‌دلیل اهمیت موضوع، مطالعات میدانی و عددی فراوانی روی نحوه تغییرات بارش در نواحی مختلف جهان انجام شده است.

به‌عنوان مثال، ناصری و همکاران (۱۵)، هارشانی و همکاران (۱۴)، کومار و جها (۱۰) و مکانیک و همکاران (۱۳) میزان بارش در نواحی مختلف را توسط الگوریتم‌های هوش مصنوعی تخمین زدند. در ادامه، اکرمی و همکاران (۱) با استفاده از تبدیل موجک داده‌های سری زمانی بارش را تجزیه کردند و دو مدل موجک- شبکه عصبی مصنوعی و موجک- انفیس را توسعه دادند. آنها بیان کردند که مدل‌های ترکیبی عملکرد بهتری در شبیه‌سازی مقادیر بارش دارد. گروسمن و مورلت (۷) با استفاده از تکنیک شبکه عصبی مصنوعی پس‌خور مقدار بارش ماهانه در شرق اندونزی را شبیه‌سازی کردند. آنها پارامتر خطای مربع میانگین را برای کنترل دقت مدل‌سازی به کار گرفتند. پورنومو و همکاران (۱۷) برای تخمین نرخ بارش ماهانه در اندونزی دو مدل شبکه عصبی مصنوعی با معماری مختلف را تعریف کردند. آنها بیان کردند که دقت یکی از مدل‌های هوش مصنوعی تقریباً ۹۸ درصد و دیگری حدود ۷۵ درصد بود که بهترین عملکرد هر دو مدل هنگامی که نوسانات بارش کمینه است به‌دست آمد. دانلادی و همکاران (۵) با استفاده از شبکه انفیس تغییرات بارش را در یک دوره زمانی کوتاه‌مدت تخمین زدند. آنها مقادیر بارش را تابعی از دما و رطوبت نسبی فرض کردند و با استفاده از این دو پارامتر ورودی مقادیر تابع هدف را مدل‌سازی کردند. علاوه بر این، خلیلی و همکاران (۹) مقدار بارش شهر مشهد را در یک بازه زمانی ۵۳ ساله با استفاده از شبکه عصبی مصنوعی شبیه‌سازی کردند. آنها بیان کردند که مدل هوش مصنوعی مذکور مقدار

بارش ماهانه را با دقت قابل قبولی شبیه‌سازی کرده است، به عنوان مثال، مقدار ضریب همبستگی مدل برتر مساوی با ۰/۹۲ محاسبه شد. در ادامه، قربانی و همکاران (۷) با ترکیب تبدیل موجک با شبکه عصبی مصنوعی و ماشین بردار پشتیبان، دو مدل ترکیبی هوش مصنوعی توسعه دادند. آنها با تجزیه و تحلیل نتایج مدل‌سازی بارش- رواناب حوضه آبریز بالیخ- لوچای در جنوب غربی شهرستان اردبیل نشان دادند که مدل‌های ترکیبی از دقت بیشتری برخوردار بودند. باباعلی و دهقانی (۲) مقادیر بارش ماهانه در یک بازه زمانی ۱۰ ساله برای حوضه آبریز دشت کاکارضا استان لرستان را با استفاده از مدل برنامه‌نویسی بیان ژن تخمین زدند. آنها نشان دادند که ضریب همبستگی مدل عددی مذکور برابر با ۰/۹۷۸ بود بنابراین این مدل توانایی خوبی در شبیه‌سازی مقدار بارندگی داشت.

همچنین، چانگ و همکاران (۴) با استفاده از یک مدل نرو فازی تحت عنوان شبکه استنتاج فازی وابسته به خود مقدار بارش- رواناب را برای حوضه‌های 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. سپس، دانش‌مهر و همکاران (۱۲) 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.

بارندگی یکی از مهم‌ترین منابع تأمین آب در کشور خشک و نیمه‌خشک ایران محسوب می‌شود و در سال‌های اخیر به‌دلیل خشکسالی‌های پیاپی و کمبود منابع آب شرب، شبیه‌سازی الگوی بارش در نواحی مختلف این کشور از اهمیت بسزایی برخوردار بوده است. در مقابل، مدل‌های هوش مصنوعی کاربرد گسترده‌ای در مدل‌سازی پدیده‌های هیدرولوژیکی از قبیل بارندگی دارند و روز به روز به گستردگی این کاربرد افزوده می‌شود. لازم به ذکر است که افزایش دقت مدل‌های هوش

۳) در هر تولید، مراحل زیر برای انتخاب یک جمعیت جدید دنبال می‌شود:

الف) یکی از عملگرهای تلاقی ژنی، جهش و انتقال انتخاب می‌شوند.

ب) تعداد مناسبی از افراد جمعیت حاضر انتخاب می‌شوند.

ج) از عملگر انتخابی برای تولید فرزند استفاده می‌شود.

د) فرزند یاد شده جدید در یک جمعیت جدید وارد می‌شود.

ه) مدل مورد نظر با استفاده از برازش‌های مختلف مورد ارزیابی واقع می‌شود.

۴) گام سوم تا رسیدن به حداکثر تعداد تولید، تکرار خواهد شد.

در این روش در ابتدای فرایند هیچ‌گونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مؤلفه‌های آن است. کاربرد برنامه‌ریزی بیان ژن در مدل‌سازی بارندگی در شکل‌های ۱ تا ۳ نشان داده شده است. با فرض h_1, h_2, h_3 به‌عنوان سه ورودی در شبیه‌سازی میزان بارندگی، ساختار درختی مربوط به برنامه ساده $(3+h_1)/(-h_2(h_3)^{0.5})$ در شکل ۱ ارائه شده است. لازم به ذکر است که مدل برنامه‌نویسی بیان ژن دارای توابع برازش مختلف شامل توابع (RRSE (Root Relative Squared Error), MAE (Mean Absolute Error) و (RMSE (Root Mean Square Error) هستند که به‌صورت تفصیلی در بخش نتایج مورد بررسی قرار می‌گیرند.

حال با توجه به مراحل چهارگانه فرایند برنامه‌ریزی بیان ژن، روند حل برنامه نشان داده شده در شکل ۱، بدین صورت ارائه می‌شود:

تولید نسل: حال یک جمعیت از درخت‌های نشان‌دهنده برنامه به‌صورت اولیه تشکیل شده است و فرایند ژنتیک بر این درختان عمل می‌کند تا افراد جامعه با کمک سری پایانه‌ها (T) و سری توابع (F) تولید شوند. بر این اساس برای شکل ۱ می‌توان نوشت:

$$F \subseteq \{-, +, \sqrt{\quad}, / \} \quad (1)$$

$$T \subseteq \{h_1, h_2, h_3, 3\} \quad (2)$$

به‌منظور تولید یک درخت تصادفی بایستی به‌نحوی به

مصنوعی هدف دیگری است که با استفاده از ابزارهای مختلف از قبیل تبدیل موجک صورت می‌گیرد. بنابراین در این مطالعه برای اولین بار، بارش ماهانه درازمدت ۶۷ ساله شهرستان بندر انزلی توسط یک مدل ترکیبی موجک- برنامه‌نویسی بیان ژن (WGEP) شبیه‌سازی می‌شود. به‌عبارت دیگر، تبدیل موجک و مدل برنامه‌نویسی بیان ژن با هم ترکیب شده و مقادیر تابع هدف برای یک بازه زمانی ۳۷ ساله آموزش داده می‌شوند (۵۵ درصد داده‌ها)، ۲۰ ساله آزمون می‌شوند (۳۰ درصد داده‌ها) و ۱۰ ساله (۱۵ درصد داده‌ها) صحت‌سنجی می‌شوند. سپس با تجزیه و تحلیل نتایج مدل‌سازی، مدل برتر و مؤثرترین تأخیرها معرفی خواهند شد.

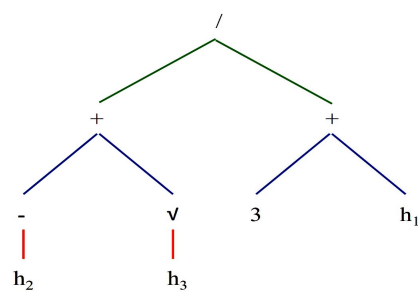
مواد و روش‌ها

برنامه‌ریزی بیان ژن (GEP)

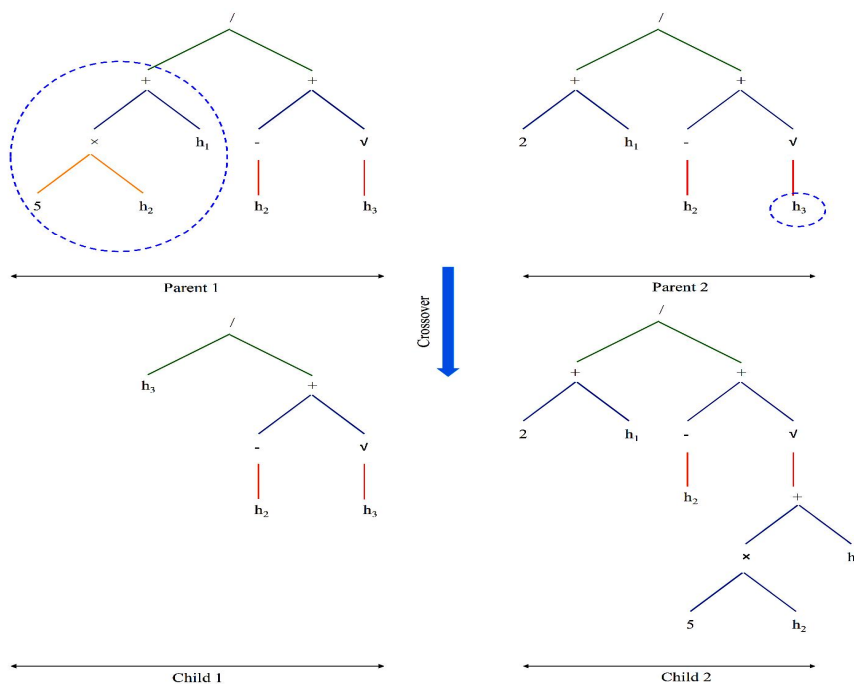
این روش جزء الگوریتم‌های گردش محسوب می‌شود که مبنای تمامی آنها بر اساس نظریه تکامل داروین استوار است. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی کرده است و سپس تابع یاد شده را برای اندازه‌گیری و مقایسه روش‌های مختلف حل، در یک فرایند گام به گام تصحیح ساختار داده‌ها، به‌کار می‌گیرند و درنهایت، روش حل مناسب را ارائه می‌کنند. برنامه‌ریزی بیان ژن (۶) ترکیب و توسعه یافته دو روش الگوریتم ژنتیک و برنامه‌ریزی ژنتیک است. در این روش، کروموزوم‌ها که خطی بوده و دارای طول ثابت هستند، مشابه الگوریتم ژنتیک که ساختاری شاخه‌ای با اندازه متفاوت دارد، ترکیب می‌شوند. هر شاخه مجموعه ساختارهای شاخه‌ای از یک مجموعه پایانه‌ها (متغیرهای مسئله) و مجموعه توابع (عملگرهای اصلی) است (۳). فرایند گام به گام برنامه‌ریزی بیان ژن به‌صورت مراحل زیر است:

۱) یک جمعیت اولیه از توابع مرکب نشان‌دهنده مدل‌های شبیه‌سازی، به‌صورت تصادفی در نظر گرفته می‌شود.

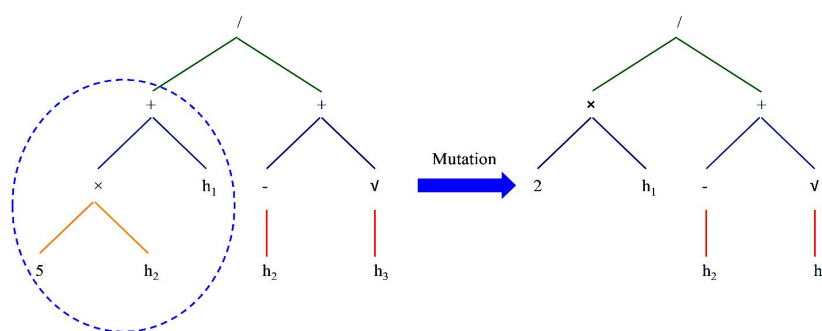
۲) هر یک از افراد جمعیت مذکور با استفاده از توابع برازش، مورد ارزیابی قرار می‌گیرند.



شکل ۱. نمونه‌ای از ساختار درختی در GEP



شکل ۲. نمونه‌ای از عملگر تقاطع در GEP



شکل ۳. نمونه‌ای از عملگر جهش در GEP

والد) شده و سپس زیر درخت‌های حاصل شده از دو والد تعویض می‌شوند و بدین ترتیب یک برنامه جدید نظیر آنچه در شکل ۲ دیده می‌شود، به دست می‌آید.

گزینه‌های تصادفی از مجموعه اقدام کرد تا تمامی شاخه‌ها به پایانه‌ها منتهی شوند. تلافی: از درون برنامه اخیر اقدام به انتخاب دو گره تصادفی

$$\int_{-\infty}^{+\infty} \frac{|\psi(x)|}{|\omega|^2} d\omega < +\infty \quad (3)$$

این شرط به‌عنوان شرط پذیرفتگی برای موجک $\psi(x)$ شناخته می‌شود. رابطه فوق را می‌توان به‌صورت زیر بازنویسی کرد:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (4)$$

این ویژگی تابع با میانگین صفر، چنان محدود کننده نبوده و توابع بسیاری را می‌توان بر اساس آن تابع موجک نامید. $\psi(x)$ تابع موجک مادر است که توابع مورد استفاده در تحلیل، با دو عمل ریاضی انتقال و مقیاس در طول سیگنال مورد تحلیل، تغییر اندازه و تغییر محل می‌یابند.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (5)$$

در نهایت ضرایب موجک در هر نقطه از سیگنال (b) و برای هر مقدار از مقیاس (a) با رابطه زیر قابل محاسبه هستند:

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) f(t) dt \quad (6)$$

که a کار مقیاس و b کار انتقال تابع را انجام می‌دهند. به‌ازای مقادیر مختلف a و b مقدار T به‌دست می‌آید به طوری که هر موقع T بیشترین مقدار مثبت را داشته باشد، بیشترین انطباق را دارد. به‌ازای T برابر با صفر انطباقی وجود ندارد و به‌ازای T های منفی، انطباق عکس یا بیشترین تفاوت وجود دارد.

ترکیب برنامه‌ریزی بیان ژن و تبدیل موجک (WGEP)

به‌منظور مدل‌سازی GEP با تبدیل موجک، ابتدا داده‌ها باید به دسته‌های مختلف تقسیم شوند. در ابتدا نیاز هست که ورودی مدل شناسایی شود. بدین منظور در این مطالعه از تابع خود همبستگی (ACF) استفاده می‌شود. شکل ACF مربوط به داده های آموزشی در این مطالعه، در شکل ۴ ارائه شده است. با توجه به شکل مشاهده می‌شود که تأخیرهای ۱، ۲، ۴، و ۵ دارای تأثیر قابل توجهی در مدل‌سازی هستند. علاوه بر این، تأخیر زمانی ۱۲، که نشان‌دهنده دوره است، نیز دارای تأثیر قابل

جهش: یک زیر درخت به‌طور تصادفی با یک زیر درخت دیگر جابه‌جا می‌شود (شکل ۳).

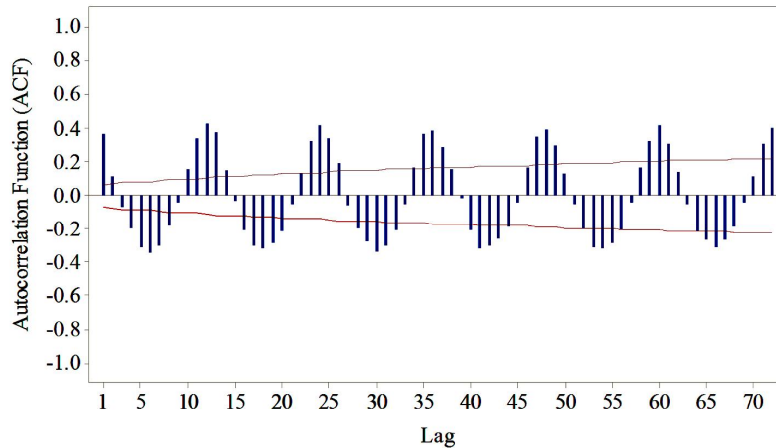
تولید مثل: این مرحله به معنی اجرای مجدد برنامه است و این امر در صورتی انجام می‌پذیرد که معیارهای برازش، صحت برنامه را تأیید کند.

روش GEP ویرایش جدید از برنامه‌ریزی ژنتیک است که به استنتاج برنامه‌های رایانه‌ای با اندازه‌ها و شکل‌های مختلف می‌پردازد. یکی از نقاط قوت GEP آن است که معیار تنوع ژنتیکی بسیار ساده بود و لذا عملگرهای ژنتیک در سطح کروموزومی عمل می‌کند. همچنین یکی دیگر از نقاط قوت این روش، عبارت از طبیعت منحصر به‌فرد چند ژنه آن است که زمینه ارزیابی مدل‌های پیچیده‌ای را که شامل چندین زیر مدل هستند، فراهم می‌آورد (۳).

تبدیل موجک

تئوری موجک یکی از روش‌های علم ریاضی است که ایده اصلی آن برگرفته از تبدیل فوریه است که در قرن نوزدهم مطرح شده است. مفهوم کلی موجک‌ها به‌صورت تئوری کنونی توسط مورلت و تیمی در مرکز تحقیقات فیزیک نظری مارسل زیر نظر آلکس گراسمن در فرانسه ارائه شد (۷). موجک‌ها توابع ریاضی هستند که شکل مقیاس زمان از سری‌های زمانی و روابط آنها برای آنالیز سری‌های زمانی که شامل متغیرها و غیرثابت‌ها است را ارائه می‌دهد. تحلیل موجکی استفاده از فاصله‌های زمانی طولانی مدت را برای اطلاعات دارای فرکانس پایین و تناوب‌های کوتاه‌تر را برای اطلاعات دارای فرکانس بالا ارائه می‌دهد. تحلیل موجکی قادر به نمایش جنبه‌های مختلف داده‌های متفاوت، نقاط شکست و ناپیوستگی‌ها است که ممکن است دیگر روش‌های آنالیز سیگنال آنها را نشان ندهند. تابع موجک، تابعی است که دو ویژگی مهم نوسانی بودن و کوتاه مدت بودن را دارا است.

تابع $\psi(x)$ تابع موجک است اگر و فقط اگر تبدیل فوریه آن، شرط زیر را اکتفا کند (۱۱):



شکل ۴. نمودار ACF برای داده‌های بخش آموزش مدل

سری زمانی منطبق شود، بهتر می‌تواند عمل نگاشت را انجام دهد و نتایج بهتری داشته باشد. توابع موجک مادر مورد استفاده در این مطالعه به شرح داوپیچیز (db)، هار (haar)، سایملتس (sym)، کوئیفلتس (coif)، بایورثوگونال (bior) و دی میسر (dmey) است.

علاوه بر آن، برای تعیین سطح تجزیه از رابطه زیر (۱۶) استفاده می‌شود:

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

که l برابر است با سطح تجزیه، n تعداد نمونه‌ها در حالت آموزش و int قسمت صحیح مقدار l است.

ناحیه مورد مطالعه

بندر انزلی، شهری واقع در مرکز شهرستان انزلی و غرب استان گیلان است. مسافت این شهر تا مرکز استان گیلان یعنی شهر رشت در حدود ۴۰ کیلومتر است. بندر انزلی دارای هوای مرطوب بوده و نیز یکی از پربارش‌ترین شهرهای ایران است. میانگین سالانه دمای هوا در بندر انزلی ۱۶/۲ درجه سانتی‌گراد است. میانگین بیشینه سالانه دمای هوا ۱۹/۲ و میانگین کمینه سالانه آن ۱۳/۲ درجه است. اختلاف بیشینه و کمینه دمای سالانه ۵/۱ درجه است. میانگین سالانه رطوبت نسبی هوا در بندر انزلی ۸۴ درصد است. میانگین رطوبت نسبی در زمستان ۸۶/۷ درصد، در بهار ۸۲/۷، در تابستان ۸۰ و در پاییز ۸۶/۷

توجهی است. بنابراین با در نظر گرفتن تأخیرهای ۱ و ۱۲ (که نسبت به سایر تأخیرها بیشترین اثر را دارند) در تمامی مدل‌ها، مدل‌های زیر پیشنهاد می‌شوند:

M1	$h(t) = f(h(t-1), h(t-12))$
M2	$h(t) = f(h(t-1), h(t-2), h(t-12))$
M3	$h(t) = f(h(t-1), h(t-4), h(t-12))$
M4	$h(t) = f(h(t-1), h(t-5), h(t-12))$
M5	$h(t) = f(h(t-1), h(t-6), h(t-12))$
M6	$h(t) = f(h(t-1), h(t-2), h(t-4), h(t-12))$
M7	$h(t) = f(h(t-1), h(t-2), h(t-5), h(t-12))$
M8	$h(t) = f(h(t-1), h(t-2), h(t-6), h(t-12))$
M9	$h(t) = f(h(t-1), h(t-4), h(t-5), h(t-12))$
M10	$h(t) = f(h(t-1), h(t-4), h(t-6), h(t-12))$
M11	$h(t) = f(h(t-1), h(t-5), h(t-6), h(t-12))$
M12	$h(t) = f(h(t-1), h(t-2), h(t-4), h(t-5), h(t-12))$
M13	$h(t) = f(h(t-1), h(t-2), h(t-4), h(t-6), h(t-12))$
M14	$h(t) = f(h(t-1), h(t-4), h(t-5), h(t-6), h(t-12))$
M15	$h(t) = f(h(t-1), h(t-2), h(t-4), h(t-5), h(t-6), h(t-12))$

پس از انتخاب داده‌های آموزشی و تعریف مدل‌های مختلف، دو گام اساسی دیگر پیش از انجام تحلیل وجود دارد، تعیین تابع موجک و سطح تجزیه. در استفاده از تبدیل موجک‌ها، تعیین نوع موجک و سطح بهینه تجزیه بسیار مهم است. یکی از نکات مهم و اساسی در انتخاب موجک مادر، طبیعت رخداده پدیده و نوع سری زمانی آن است. بنابراین الگوهایی از توابع موجک مادر که بتواند به لحاظ شکل هندسی به خوبی بر منحنی

جدول ۱. شاخص‌های آماری مختلف برای بارش ۶۷ ساله شهر انزلی از سال ۱۹۵۱ تا ۲۰۱۷

حالت	واریانس (میلی‌متر)	انحراف معیار (میلی‌متر)	میانگین (میلی‌متر)	کمینه (میلی‌متر)	بیشینه (میلی‌متر)
آموزش	۲۴۵۵۰/۴۳	۱۵۶/۶۸۵۸	۱۵۹/۲۱۶۹	۰	۹۶۰
آزمون	۱۷۵۶۶/۳۲	۱۳۲/۵۳۸	۱۴۰/۶۲۳۳	۰	۷۴۲/۹
اعتبار سنجی	۲۰۶۴۳/۳۵	۱۴۳/۶۷۷۹	۱۵۱/۲۵۳۳	۰	۶۲۱/۴

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

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

در اینجا O_i مقادیر مشاهداتی، F_i مقادیر شبیه‌سازی شده به وسیله مدل‌های عددی، \bar{O} میانگین مقادیر مشاهداتی و n برابر تعداد مقادیر مشاهداتی هستند. در این مطالعه ۵۵ درصد مقادیر مشاهداتی برای آموزش مدل‌های هوش مصنوعی، ۳۰ درصد برای آزمون و ۱۵ درصد باقیمانده برای صحت‌سنجی استفاده می‌شود.

درصد است. میانگین سالانه بارندگی در بندر انزلی ۱۸۵۳/۵ میلی‌متر است. کم‌باران‌ترین سال طی ۵۰ سال اخیر در انزلی سال ۱۹۷۱ با ۱۱۵۶/۲ میلی‌متر و پر باران‌ترین سال طی این مدت سال ۱۹۵۷ با ۳۰۲۰/۲ میلی‌متر بارندگی است. میانگین بارندگی فصلی این شهر در فصل بهار ۱۶۰/۹، در تابستان ۴۳۴/۳، در پاییز ۸۵۲/۶ و در زمستان ۴۰۵/۷ میلی‌متر است. همچنین، تعداد روزهای بارندگی در بندر انزلی ۱۴۸ روز است. حداکثر بارندگی در ۲۴ ساعت در بندر انزلی به میزان ۳۵۳ میلی‌متر در تاریخ ۱۷ سپتامبر ۱۹۶۲ ثبت شده است. در جدول ۱ شاخص‌های آماری مختلف برای بارش ۶۷ ساله از سال ۱۹۵۱ تا ۲۰۱۷ مرتب شده است.

نتایج و بحث

بررسی خانواده‌های موجک

در ابتدا به ارزیابی خانواده‌های مختلف تبدیل موجک پرداخته می‌شود. به‌طور کلی، تبدیل موجک دارای خانواده‌هایی با عنوان داوپیچ (db)، هار (haar)، سایملتس (sym)، کوئیفلتس (coif)، بایورثوگونال (bior) و دی میر (dmey) است که هر کدام از این خانواده‌ها به‌ترتیب دارای ده، یک، هفت، پنج، سه و یک عضو هستند. بر اساس تجزیه و تحلیل کلیه اعضای خانواده‌های مختلف تبدیل موجک مشخص شد که تنها عضو خانواده dmey دارای بیشترین دقت و کمترین مقدار خطا در مقایسه با سایر اعضای خانواده‌های موجک است. در شکل ۵ نتایج شاخص‌های آماری برای اعضای برتر خانواده‌های مختلف موجک قابل مشاهده است. به‌عنوان مثال برای خانواده داوپیچ، db10 به‌عنوان عضو برتر شناسایی شد. برای این عضو خانواده

معیارهای بررسی دقت مدل‌های عددی

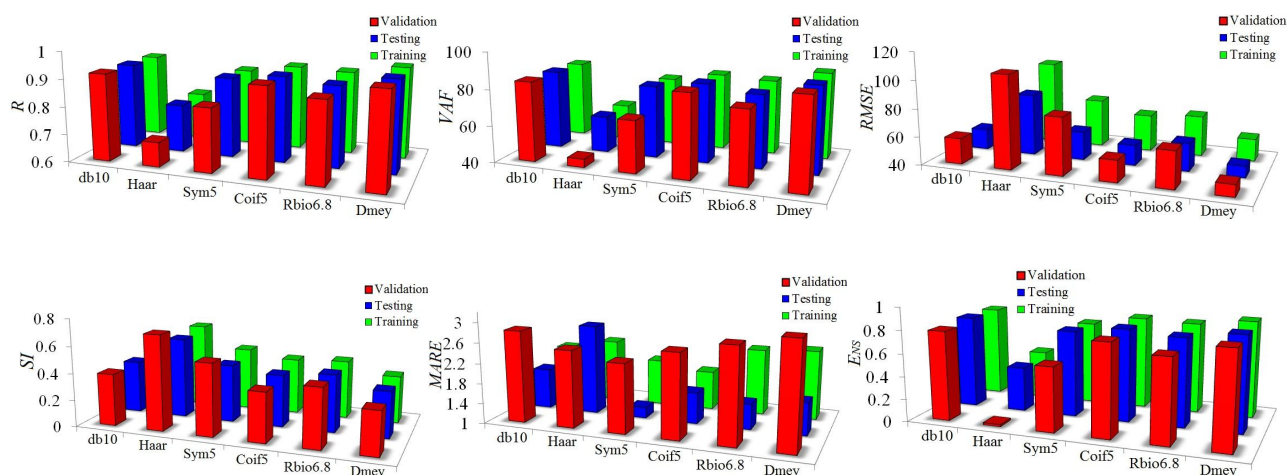
در مطالعه حاضر به‌منظور ارزیابی دقت مدل‌های عددی معرفی شده از شاخص‌های آماری ضریب همبستگی (R)، ضریب تغییرات (VAF)، خطای مربعات میانگین ریشه (RMSE)، شاخص پراکندگی (SI)، درصد خطای نسبی میانگین (MARE) و ضریب بهره‌وری نش ساتکلیف (ENS) به‌صورت زیر استفاده می‌شود:

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

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

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

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

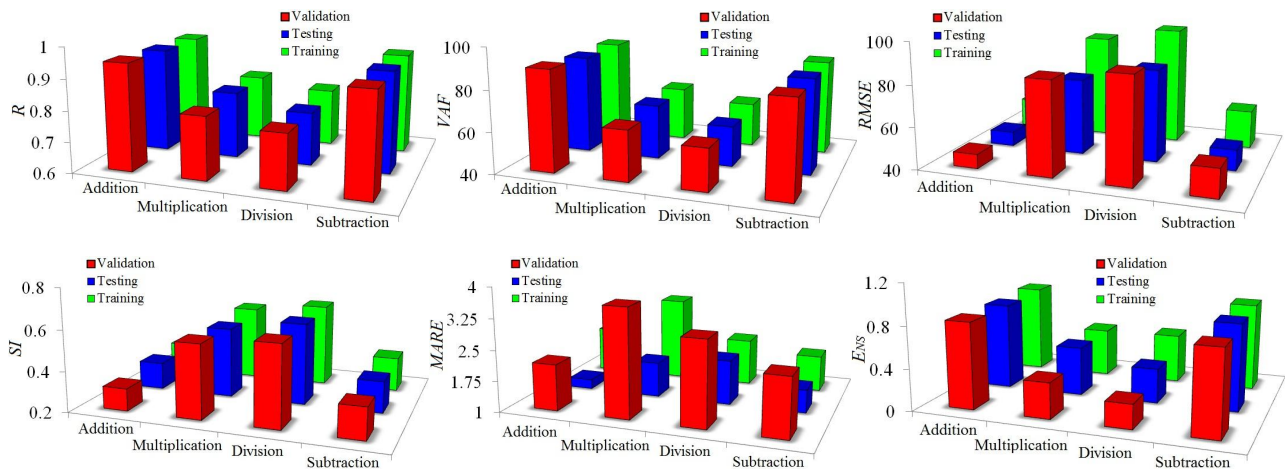


شکل ۵. نتایج شاخص‌های آماری برای اعضای برتر خانواده‌های مختلف موجک در وضعیت‌های آموزش، آزمون و صحت‌سنجی

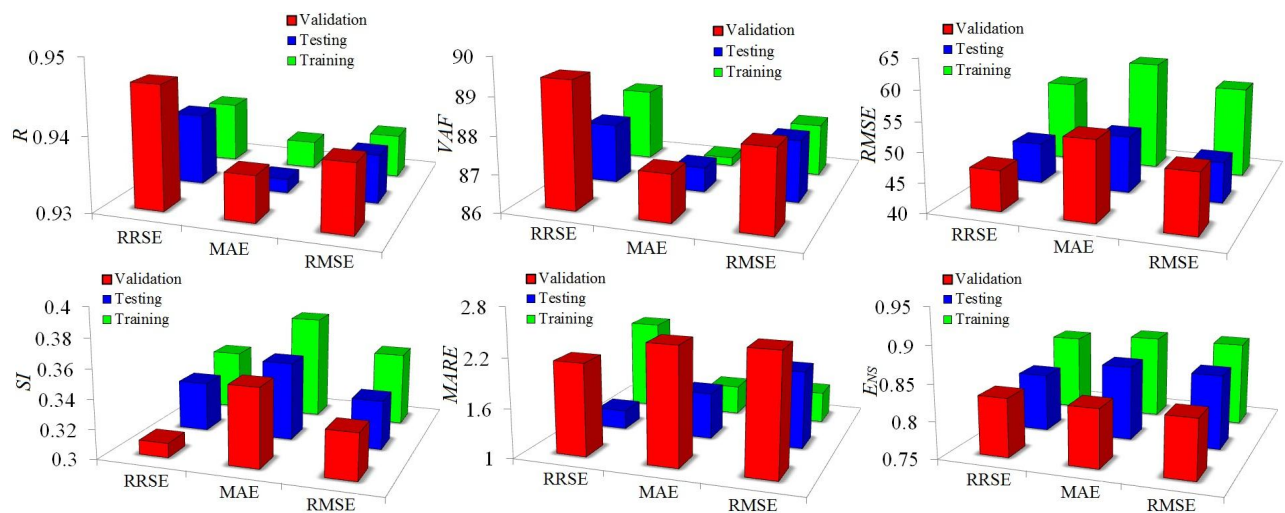
بررسی توابع اتصال مدل برنامه‌ریزی بیان ژن

مدل برنامه‌نویسی بیان ژن دارای چهار تابع اتصال با عنوان‌های جمع، ضرب، تقسیم و تفریق است که بهترین آنها باید برای مدل‌سازی بارش مورد استفاده قرار گیرد. در ادامه به بررسی این توابع اتصال پرداخته می‌شود که مقایسه شاخص‌های آماری مختلف برای آنها در شکل ۶ نشان داده شده است. بر اساس نتایج مدل‌سازی با استفاده از توابع اتصال مختلف، مقادیر شاخص‌های آماری R، VAF و SI برای تابع Addition در شرایط صحت‌سنجی به ترتیب مساوی با ۰/۹۴۶، ۰/۳۸۷ و ۰/۳۱۰ تخمین زده شدند. علاوه بر این، مقادیر RMSE، MARE و ENS برای این تابع اتصال در وضعیت آزمون به ترتیب برابر با ۴۶/۹۴۷، ۱/۱۲۵ و ۰/۸۲۹ بودند. در مقابل، برای توابع Multiplication، Division و Subtraction در حالت آزمون مقدار ضریب بهره‌وری نش‌ساتکلیف به ترتیب مساوی با ۰/۴۶۵، ۰/۴۷۲ و ۰/۸۴۵ تخمین زده شدند. بنابراین با بررسی نتایج توابع اتصال مختلف مشاهده شد که تابع اتصال Addition در هر سه شرایط آموزش، آزمون و صحت‌سنجی دارای کمترین مقدار خطا و بیشترین دقت بود که در ادامه این تابع اتصال برای مدل‌سازی بارش استفاده می‌شود. در این مطالعه، تابع اتصال انتخاب شده توسط یک پروسه سعی و خطا انتخاب شد هر چند که سایر پژوهشگران نیز به استفاده گسترده

داوپیچیز مقادیر RMSE و SI در شرایط آزمون به ترتیب مساوی با ۵۴/۷۱۸ و ۰/۳۸۹ تخمین زده شدند. همچنین برای تنها عضو خانواده هار نیز مقادیر ضریب همبستگی و نش ساتکلیف در وضعیت صحت‌سنجی به ترتیب برابر ۰/۶۸۹ و ۰/۰۱۶ محاسبه شدند. علاوه بر این، Sym 5 دقیق‌ترین عضو خانواده سایملتس به دست آمد که برای آن مقادیر شاخص‌های آماری VAF، MARE و ENS در حالت آموزش به ترتیب مساوی با ۷۷/۴۱۳، ۱/۹۲۴ و ۰/۷۳۵ بودند. برای خانواده‌های کویفلتس و بایورثوگونال نیز اعضای Coif5 و Rbio6.8 به ترتیب به عنوان عضو برتر انتخاب شدند. برای این اعضا مقدار شاخص پراکندگی در وضعیت صحت‌سنجی به ترتیب برابر با ۰/۳۶۷ و ۰/۴۳۷ تخمین زده شدند. لازم به ذکر است که تنها عضو خانواده دی میر به عنوان دقیق‌ترین عضو خانواده های مختلف موجک شناسایی شد. این عضو مقادیر بارش را با کمترین مقدار خطا و بالاترین مقدار همبستگی شبیه‌سازی کرد. برای این عضو خانواده دی میر (Dmey) مقادیر ضریب همبستگی، شاخص پراکندگی و نش ساتکلیف در وضعیت صحت‌سنجی به ترتیب مساوی با ۰/۹۴۲، ۰/۳۲۰ و ۰/۸۳۵ محاسبه شدند. بنابراین در ادامه Dmey به عنوان عضو برتر خانواده‌های مختلف موجک برای شبیه‌سازی بارش مورد استفاده قرار می‌گیرد.



شکل ۶. مقایسه شاخص‌های آماری برای توابع اتصال مختلف مدل برنامه نویسی بیان ژن



شکل ۷. مقایسه شاخص‌های آماری برای توابع برازش مختلف مدل برنامه نویسی بیان ژن

محاسبه شدند. این در حالی است که مقادیر RMSE و MARE و R برای وضعیت صحت‌سنجی تابع برازش RRSE به‌ترتیب مساوی با ۴۶/۸۳۹، ۲/۱۳۰ و ۰/۹۴۶ بودند. در مقابل، مقدار ضریب بهره‌وری نش‌ساتکلیف برای توابع برازش MAE و RMSE در شرایط آموزش به‌ترتیب برابر با ۰/۸۶۵ و ۰/۸۶۶ به‌دست آمدند. بنابراین، برای شبیه‌سازی بارش درازمدت شهر انزلی توسط مدل برنامه‌ریزی بیان ژن از تابع برازش RRSE بهره گرفته می‌شود. نتایج شاخص‌های آماری محاسبه شده برای توابع برازش مختلف مدل برنامه‌ریزی بیان ژن در هر سه وضعیت آموزش، آزمون و صحت‌سنجی در شکل ۷ قابل مشاهده است.

آن در مقایسه با سایر توابع اتصال اشاره کرده‌اند (۶).

بررسی توابع برازش مدل برنامه‌ریزی بیان ژن

در این بخش به ارزیابی توابع برازش مختلف مدل برنامه‌نویسی بیان ژن که شامل توابع RRSE، MAE و RMSE هستند پرداخته می‌شود. با توجه به نتایج این توابع برازش، تابع RRSE دارای بیشترین همبستگی و کمترین مقدار خطا در هر سه شرایط آموزش، آزمون و صحت‌سنجی بود. به‌عنوان مثال، مقدار شاخص‌های آماری VAF، SI و ENS در حالت آزمون برای تابع برازش RRSE به‌ترتیب برابر با ۰/۳۳۴، ۸۷/۶۰۸ و ۰/۸۲۹

جدول ۲. خلاصه پارامترهای اولیه مورد استفاده در مدل برنامه‌ریزی بیان ژن

پارامتر	مقدار
تعداد کروموزوم	۱۰۰
تعداد سر	۱۰
تعداد ژن‌ها	۵
تابع اتصال	Addition
تابع برازش	RRSE
جهش	۰/۰۴۴
واژگونی	۰/۱۵
انتقال IS	۰/۱۵
انتقال RIS	۰/۱۵
نو ترکیبی یک نقطه	۰/۷۵
نو ترکیبی دو نقطه	۰/۸۵
نو ترکیبی ژن	۰/۱۵
انتقال ژن	۰/۱۵

در این مطالعه، تعداد کروموزوم‌ها در دامنه ۳۰ تا ۱۵۰ بررسی و مقدار ۱۰۰ به‌عنوان بهترین مقدار انتخاب شد. علاوه بر این تعداد سرها نیز در دامنه ۳ تا ۱۵ مورد ارزیابی قرار گرفت که بهترین مقدار برابر با ۱۰ تعریف شد. همچنین مقادیر جهش و انتقال ژن نیز با روند سعی و خطا به‌دست آمدند. در جدول ۲ خلاصه پارامترهای اولیه مورد استفاده در مدل برنامه‌ریزی بیان ژن مرتب شده است.

مقایسه مدل‌های WGEP

در این قسمت به انجام تحلیل حساسیت مدل‌های مختلف WGEP پرداخته می‌شود. لازم به ذکر است که ۱۵ مدل ترکیبی WGEP با استفاده از تأخیرهای مختلف تعریف شده‌اند. نتایج شاخص‌های آماری مختلف برای مدل برتر WGEP در جدول ۳ مرتب شده است. بر اساس نتایج مدل‌سازی، مقدار ضریب همبستگی و شاخص پراکندگی برای مدل WGEP 1 در شرایط آموزش به‌ترتیب مساوی با ۰/۷۳۷ و ۰/۶۶۵ به‌دست آمدند. این مدل مقادیر تابع هدف را برحسب تأخیرهای شماره ۱ و ۱۲

تخمین زد. علاوه بر این، مدل WGEP 2 مقدار بارندگی درازمدت شهر انزلی را با استفاده از تأخیرهای شماره ۱، ۲ و ۱۲ شبیه‌سازی کرد. این مدل تابعی از تأخیرهای شماره ۱، ۴ و ۱۲ بود. همچنین، مدل WGEP 4 مقدار تابع هدف را با استفاده از تأخیرهای ۱، ۵ و ۱۲ مدل‌سازی کرد. برای شرایط آموزش WGEP 4، مقادیر SI، R، MARE و به‌ترتیب مساوی با ۰/۸۱۵، ۰/۵۶۹ و ۲/۶۸۹ به‌دست آمدند. مدل WGEP 5 مقدار بارندگی درازمدت را برحسب تأخیرهای شماره ۱، ۶ و ۱۲ محاسبه کرد. همچنین مقادیر VAF و RMSE در شرایط آموزش مدل WGEP 6 به‌ترتیب برابر با ۸۸/۰۰۶ و ۵۴/۲۳۹ محاسبه شدند. مدل WGEP 6 مقادیر بارش را با ترکیبی از تأخیرهای شماره ۱، ۲، ۴ و ۱۲ شبیه‌سازی کرد. برای مدل WGEP 7 در وضعیت آزمون مقادیر R و ENS به‌ترتیب برابر با ۰/۹۲۰ و ۰/۸۴۱ بودند. لازم به ذکر است که شاخص آماری ضریب بهره‌وری نش‌ساتکلیف برای مدل‌های WGEP 11 و WGEP 12 در وضعیت صحت‌سنجی به‌ترتیب برابر با ۰/۲۰۹ و ۰/۸۲۳ تخمین زده شدند. با توجه به نحوه ترکیب تأخیرها، مدل WGEP 11 تابع

جدول ۳. نتایج شاخص‌های آماری برای مدل برتر

مدل	تأخیر	وضعیت	R	VAF	RMSE	SI	MARE	ENS
GEP 1	۱ و ۱۲	آموزش	۰/۷۳۷	۵۴/۰۲۲	۱۰۵/۶۰۴	۰/۶۶۵	۲/۷۴۰	۰/۲۶۹
		آزمون	۰/۷۴۲	۵۴/۹۱۷	۸۸/۸۳۰	۰/۶۳۲	۱/۵۶۹	۰/۲۷۷
		صحت‌سنجی	۰/۷۲۰	۵۰/۶۸۹	۱۰۰/۵۱۱	۰/۶۶۵	۲/۶۲۵	۰/۰۷۵
GEP 2	۱ و ۲ و ۱۲	آموزش	۰/۹۳۰	۸۶/۳۴۵	۵۷/۷۷۸	۰/۳۶۴	۱/۹۹۴	۰/۸۴۹
		آزمون	۰/۹۲۸	۸۶/۰۹۴	۴۹/۸۷۵	۰/۳۵۵	۲/۰۲۳	۰/۸۳۱
		صحت‌سنجی	۰/۹۳۶	۸۷/۶۸۶	۵۰/۹۵۵	۰/۳۳۷	۲/۸۶۲	۰/۸۲۴
GEP 3	۱ و ۴ و ۱۲	آموزش	۰/۷۹۹	۶۳/۸۱۳	۹۳/۶۹۳	۰/۵۹۰	۲/۰۸۳	۰/۴۴۶
		آزمون	۰/۸۱۰	۶۵/۵۷۷	۷۷/۶۳۸	۰/۵۲۲	۲/۶۰۲	۰/۴۷۸
		صحت‌سنجی	۰/۷۶۰	۵۷/۲۴۷	۹۳/۸۶۲	۰/۶۲۱	۳/۳۷۷	۰/۲۳۷
GEP 4	۱ و ۵ و ۱۲	آموزش	۰/۸۱۵	۶۶/۳۶۲	۹۰/۳۵۰	۰/۵۶۹	۲/۶۸۹	۰/۵۱۴
		آزمون	۰/۷۹۷	۶۳/۴۹۳	۷۹/۹۲۹	۰/۵۶۸	۱/۵۲۴	۰/۴۵۶
		صحت‌سنجی	۰/۷۷۱	۵۹/۰۱۹	۹۱/۷۹۲	۰/۶۰۷	۳/۱۴۴	۰/۲۸۳
GEP 5	۱ و ۶ و ۱۲	آموزش	۰/۷۴۶	۵۵/۵۲۶	۱۰۳/۹۳۲	۰/۶۵۵	۲/۱۵۴	۰/۲۸۴
		آزمون	۰/۷۷۱	۵۹/۴۴۸	۸۴/۲۸۴	۰/۵۹۹	۱/۵۲۵	۰/۳۳۱
		صحت‌سنجی	۰/۷۳۷	۵۳/۶۲۷	۹۷/۵۴۰	۰/۶۴۵	۲/۹۰۷	۰/۱۰۴
GEP 6	۱ و ۲ و ۴ و ۱۲	آموزش	۰/۹۳۸	۸۸/۰۰۶	۵۴/۲۳۹	۰/۳۴۲	۲/۱۴۰	۰/۸۵۷
		آزمون	۰/۹۴۰	۸۷/۶۰۸	۴۶/۹۴۷	۰/۳۳۴	۱/۲۳۵	۰/۸۳۰
		صحت‌سنجی	۰/۹۴۶	۸۹/۳۸۷	۴۶/۸۳۹	۰/۳۱۰	۲/۱۳۰	۰/۸۳۰
GEP 7	۱ و ۲ و ۵ و ۱۲	آموزش	۰/۹۲۰	۸۲/۹۲۳	۶۵/۷۱۲	۰/۴۱۴	۱/۳۶۸	۰/۸۴۱
		آزمون	۰/۹۱۲	۸۲/۱۵۲	۵۷/۱۴۶	۰/۴۰۶	۱/۲۴۷	۰/۸۱۸
		صحت‌سنجی	۰/۹۲۸	۸۴/۹۵۴	۵۷/۷۶۴	۰/۳۸۲	۱/۸۶۱	۰/۸۱۴
GEP 8	۱ و ۲ و ۶ و ۱۲	آموزش	۰/۹۲۰	۸۲/۷۹۲	۶۵/۹۹۱	۰/۴۱۶	۱/۳۵۴	۰/۸۳۹
		آزمون	۰/۹۱۰	۸۲/۰۰۹	۵۷/۵۰۵	۰/۴۰۹	۱/۲۷۴	۰/۸۱۳
		صحت‌سنجی	۰/۹۲۵	۸۴/۵۳۶	۵۸/۶۷۷	۰/۳۸۸	۱/۹۴۱	۰/۸۰۵
GEP 9	۱ و ۴ و ۵ و ۱۲	آموزش	۰/۷۴۷	۵۵/۴۵۸	۸۸/۳۰۴	۰/۶۲۸	۲/۵۶۶	۰/۳۰۶
		آزمون	۰/۷۵۰	۵۵/۹۶۵	۱۰۳/۳۸۰	۰/۶۵۱	۲/۲۵۶	۰/۳۲۷
		صحت‌سنجی	۰/۷۱۴	۴۹/۷۰۹	۱۰۱/۶۳۲	۰/۶۷۲	۲/۸۹۴	۰/۰۸۰
GEP 10	۱ و ۴ و ۶ و ۱۲	آموزش	۰/۷۸۱	۶۰/۴۴۷	۹۸/۴۹۷	۰/۶۲۰	۲/۲۷۷	۰/۴۵۲
		آزمون	۰/۷۷۰	۵۸/۶۷۰	۸۵/۶۰۱	۰/۶۰۹	۲/۱۸۱	۰/۴۲۷
		صحت‌سنجی	۰/۷۳۴	۵۱/۷۷۶	۹۹/۷۶۸	۰/۶۶۰	۲/۹۳۵	۰/۲۲۲
GEP 11	۱ و ۵ و ۶ و ۱۲	آموزش	۰/۸۴۰	۷۰/۳۸۷	۸۴/۷۳۷	۰/۵۳۴	۲/۷۷۲	۰/۵۴۲
		آزمون	۰/۸۲۸	۶۷/۹۴۴	۷۴/۸۸۴	۰/۵۳۳	۳/۶۵۰	۰/۴۲۷
		صحت‌سنجی	۰/۷۸۹	۶۲/۱۸۹	۸۸/۰۱۱	۰/۵۸۲	۵/۱۳۷	۰/۲۰۹
GEP 12	۱ و ۲ و ۴ و ۵ و ۱۲	آموزش	۰/۹۳۱	۸۵/۵۹۵	۶۰/۵۲۸	۰/۳۸۱	۱/۲۹۰	۰/۸۶۰
		آزمون	۰/۹۲۴	۸۴/۹۴۹	۵۲/۶۵۵	۰/۳۷۴	۱/۴۷۴	۰/۸۳۸
		صحت‌سنجی	۰/۹۳۱	۸۶/۳۷۸	۵۵/۰۳۹	۰/۳۶۴	۲/۵۰۵	۰/۸۲۳

ادامه جدول ۳. نتایج شاخص‌های آماری برای مدل برتر

مدل	تأخیر	وضعیت	R	VAF	RMSE	SI	MARE	ENS
GEP 13	۱۲ و ۶ و ۴ و ۲ و ۱	آموزش	۰/۹۳۱	۸۵/۴۸۹	۶۰/۴۳۰	۰/۳۸۱	۱/۲۴۷	۰/۸۶۱
		آزمون	۰/۹۲۴	۸۴/۹۳۱	۵۲/۳۲۵	۰/۳۷۲	۱/۴۶۵	۰/۸۴۳
		صحت‌سنجی	۰/۹۳۳	۸۶/۵۳۸	۵۴/۳۱۷	۰/۳۵۹	۲/۳۹۴	۰/۸۳۱
GEP 14	۱۲ و ۶ و ۵ و ۴ و ۱	آموزش	۰/۸۲۴	۶۷/۵۸۹	۸۸/۷۷۸	۰/۵۵۹	۲/۱۰۸	۰/۵۸۵
		آزمون	۰/۸۱۹	۶۶/۳۵۵	۷۶/۷۹۰	۰/۵۴۶	۲/۳۹۹	۰/۵۸۳
		صحت‌سنجی	۰/۷۶۸	۵۶/۷۶۹	۹۴/۳۸۴	۰/۶۲۴	۳/۴۲۷	۰/۳۷۰
GEP 15	۱۲ و ۶ و ۵ و ۴ و ۲ و ۱	آموزش	۰/۹۳۵	۸۷/۳۵۸	۵۵/۴۹۵	۰/۳۵۰	۲/۱۱۲	۰/۸۵۹
		آزمون	۰/۹۳۲	۸۶/۷۲۳	۴۸/۵۶۷	۰/۳۴۵	۲/۰۰۴	۰/۸۳۵
		صحت‌سنجی	۰/۹۳۷	۸۷/۸۲۶	۵۰/۴۱۵	۰/۳۳۳	۲/۴۴۷	۰/۸۲۲

مقایسه مدل‌های GEP با WGEP

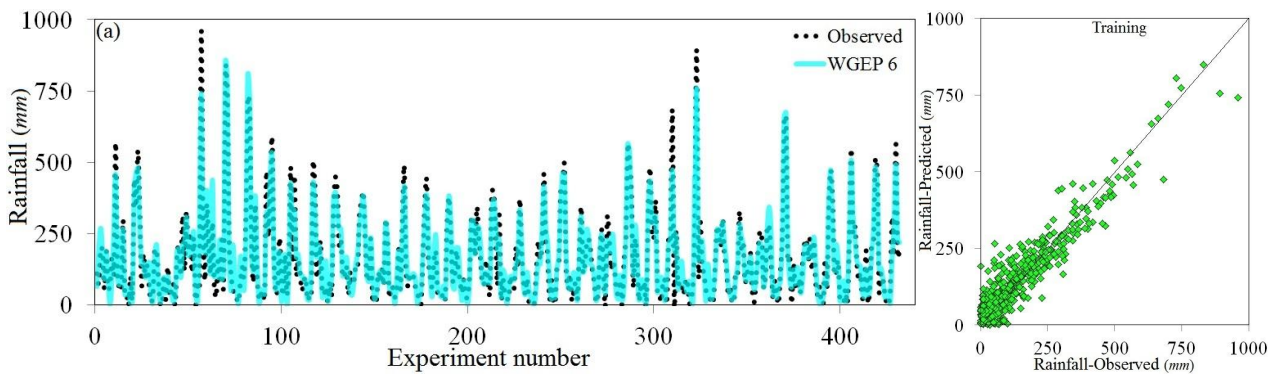
در این قسمت، نتایج مدل برتر ترکیبی (WGEP 6) با مدل برتر GEP نیز مقایسه می‌شود. مدل برتر GEP مدل متناظر WGEP است. به عبارت دیگر، مقادیر تابع هدف توسط این مدل با ترکیبی از تأخیرهای ۱، ۲، ۴ و ۱۲ شبیه‌سازی شد. در شکل ۹ مقایسه نتایج مدل‌های GEP و WGEP برتر با مقادیر مشاهداتی در وضعیت‌های آموزش، آزمون و صحت‌سنجی قابل مشاهده است. با توجه به نتایج مدل‌سازی، تبدیل مویک، دقت مدل‌سازی برنامه نویسی بیان ژن را به شکل قابل ملاحظه‌ای افزایش داد. به عنوان مثال، شاخص آماری MARE محاسبه شده برای WGEP در وضعیت آموزش تقریباً ۶۲ درصد کمتر از مقدار پارامتر مذکور برای مدل GEP بود. این درحالی است که برای شرایط صحت‌سنجی، مقادیر ضریب همبستگی و شاخص پراکندگی مدل WGEP به ترتیب تقریباً ۷۲ درصد و ۶۱ درصد در مقایسه با مدل GEP بهبود یافت. بنابراین همان‌گونه که مشاهده شد، مدل ترکیبی دارای دقت بیشتری در مقایسه با مدل ساده برنامه‌نویسی بیان ژن بود.

نتیجه‌گیری

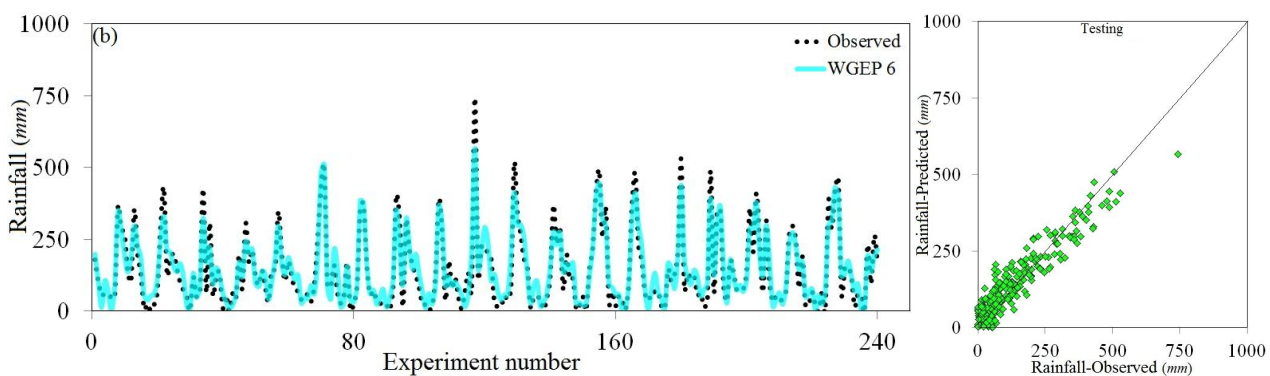
در این مطالعه، با استفاده از تبدیل مویک (WT)، مدل برنامه‌ریزی بیان ژن (GEP) برای تخمین بارش درازمدت ۶۷ ساله شهر انزلی بهینه‌سازی شد. در ابتدا، با بررسی نتایج مدل‌سازی برای اعضای مختلف خانواده مویک بهینه‌ترین

تأخیرهای ۱، ۲، ۴، ۵ و ۱۲ و WGEP تابع ۱، ۲، ۴، ۵ و ۱۲ بود. در شرایط آموزش، آزمون و صحت‌سنجی مدل WGEP 13 مقدار VAF به ترتیب مساوی با ۸۵/۴۸۹، ۸۴/۹۳۱ و ۸۶/۵۳۸ بود. لازم به یادآوری است که این مدل تابعی از تأخیرهای شماره ۱، ۲، ۴، ۶ و ۱۲ بود. همچنین برای شرایط آموزش مدل WGEP 14 مقادیر R، SI و RMSE به ترتیب برابر با ۰/۸۲۴، ۰/۵۵۹ و ۸۸/۷۷۸ محاسبه شدند. این مدل مقدار بارندگی را با استفاده از تأخیرهای شماره ۱، ۴، ۵، ۶ و ۱۲ تخمین زد. این در حالی است که مدل WGEP 15 مقدار تابع هدف را برحسب تأخیرهای ۱، ۲، ۴، ۵، ۶ و ۱۲ مدل‌سازی کرد. لازم به یادآوری است که ترکیب تأخیرهای شماره ۱، ۴، ۵ و ۱۲ باعث کاهش عملکرد مدل هوش مصنوعی شد (مدل WGEP 9).

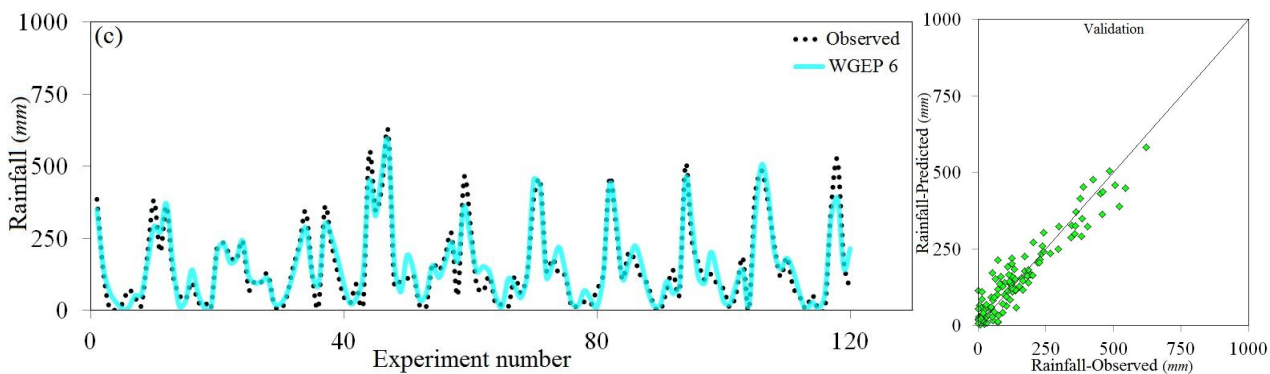
بنابراین، با تجزیه و تحلیل نتایج مدل‌های پانزده‌گانه، مدل WGEP 6 دارای کمترین مقدار خطا (شاخص پراکندگی در شرایط صحت‌سنجی مساوی با ۰/۳۱۰) و بیشترین همبستگی (ضریب همبستگی در وضعیت آزمون برابر با ۰/۹۴۰) است. در شکل ۸ مقایسه مقادیر بارش شبیه‌سازی شده با نتایج مشاهداتی و نمودارهای پراکندگی برای مدل WGEP 6 در سه وضعیت آموزش، آزمون و صحت‌سنجی به تصویر کشیده شده است. با توجه به تجزیه و تحلیل نتایج مدل‌سازی، مدل WGEP 6 به عنوان مدل برتر و تأخیرهای شماره ۱، ۲، ۴ و ۱۲ نیز به عنوان مؤثرترین تأخیرها شناسایی شدند.



(الف)

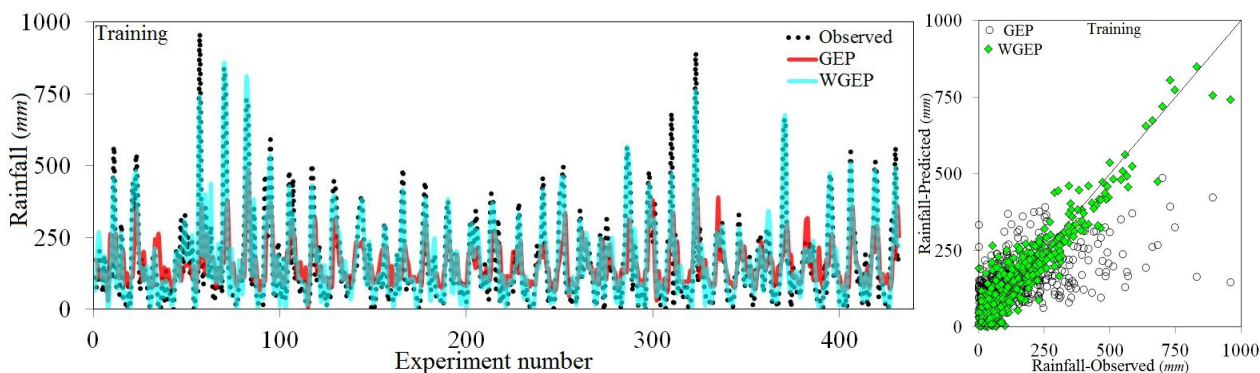


(ب)

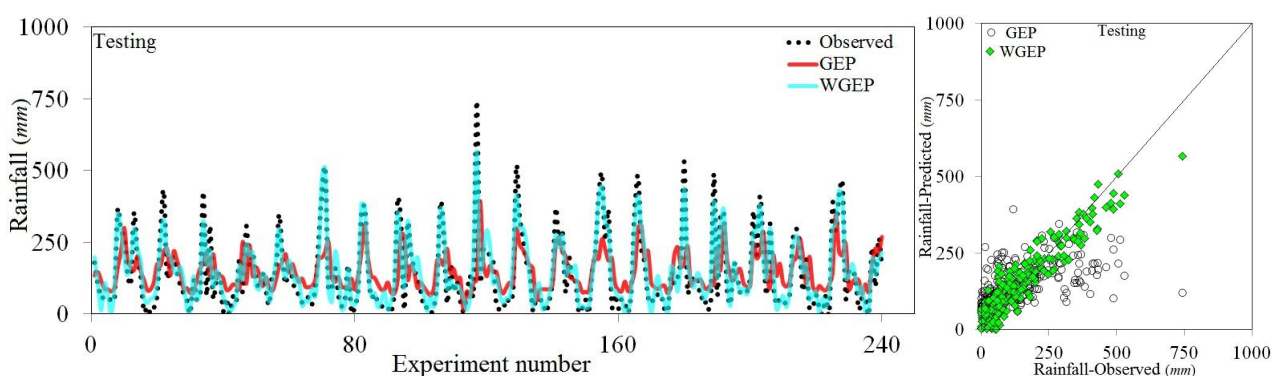


(ج)

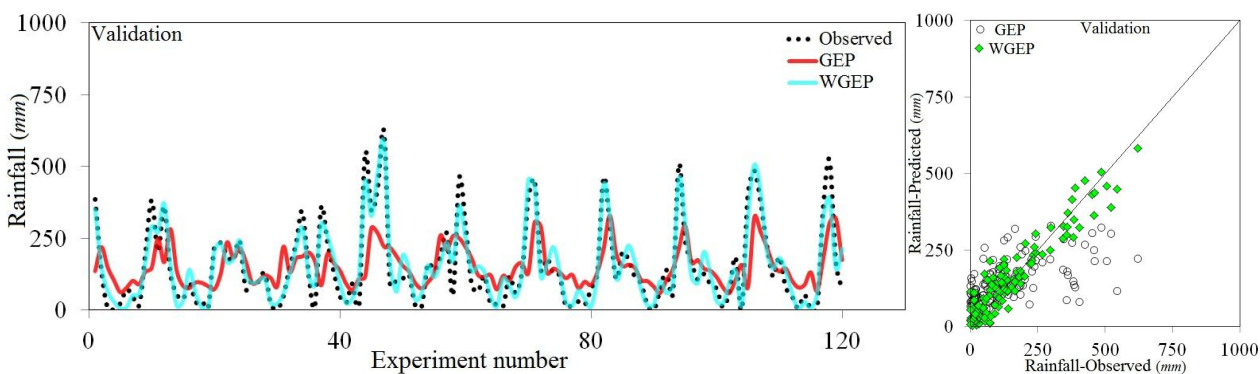
شکل ۸. مقایسه بارش‌های شبیه‌سازی شده و مشاهداتی (سمت چپ) به همراه نمودارهای پراکندگی (سمت راست) برای مدل WGEP 6 در وضعیت‌های: الف) آموزش، ب) آزمون و ج) صحت‌سنجی



(الف)



(ب)



(ج)

شکل ۹. مقایسه نتایج مدل‌های GEP و WGEP برتر با مقادیر مشاهداتی در وضعیت‌های: الف) آموزش، ب) آزمون و ج) صحت‌سنجی

خودهمبستگی نسبی و تأخیرهای مختلف، ۱۵ مدل ترکیبی (WGEP) مختلف توسعه داده شد. با انجام تحلیل حساسیت برای مدل‌های پانزده‌گانه WGEP، مدل برتر و تأخیرهای مؤثر

عضو انتخاب شد. سپس با تجزیه و تحلیل نتایج توابع اتصال و برازش متفاوت مدل برنامه‌ریزی بیان ژن، دقیق‌ترین توابع به‌دست آمد. در ادامه، به‌وسیله تابع خودهمبستگی،

به‌عنوان مؤثرترین تأخیرها معرفی شدند. علاوه بر این، نتایج مدل ترکیبی برتر با مدل برنامه‌ریزی بیان ژن مقایسه و نشان داده شد که مدل ترکیبی دارای دقت بیشتری بود.

معرفی شدند. مدل برتر دارای دقت مناسبی بود به‌گونه‌ای که مقدار ضریب همبستگی در سه وضعیت آموزش، آزمون و صحت‌سنجی برای این مدل به ترتیب برابر با ۰/۹۳۸، ۰/۹۴۰ و ۰/۹۴۶ به دست آمدند. همچنین، تأخیرهای شماره ۱، ۲، ۴ و ۱۲

منابع مورد استفاده

1. Akrami S. A., V. Nourani and S. J. S. Hakim. 2014. Development of nonlinear model based on wavelet-ANFIS for rainfall forecasting at Klang Gates Dam. *Water Resources Management* 28(10): 2999-3018.
2. Baba Ali H., and R. Dehghani. 2017. Comparison of intelligent models in estimating monthly rainfall of Kaka Reza Basin. *Ecology* 4(1): 1-11. (in Farsi).
3. Borelli, A., I. De Falco, C. A. Della, M. Nicodemi and G. Trautteur. 2006. Performance of genetic programming to extract the trend in noisy data series. *Physica A* 370: 104-108.
4. 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.
5. Danladi, A., M. Stephen, B. M. Aliyu, G. K. Gaya, N. W. Silikwa and Y. Machael. 2018. Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term. *Alexandria Engineering Journal* 57(2): 1157-1162.
6. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems* 13: 87-129.
7. Ghorbani, M. E., E. Azani and S. Mahmoudi Vanelia. 2015. Modeling of precipitation-runoff using hybrid intelligent models (technical note). *Iran Water Resource Studies* 11(2): 146-150. (in Farsi).
8. Grossmann, A. and J. Morlet. 1984. Decomposition of hardy function into square integrable wavelets of constant shape. *SIMA Journal on Mathematical Analysis (SIMA)* 5: 723-736.
9. Khalili, N., S. R. Khodashenas, K. Davari and M. Mousavi Bayegi. 2010. Prediction of daily precipitation using artificial intelligence networks, Case study: Mashhad synoptic station. *Watershed Studies* 23(4): 7-15. (in Farsi).
10. Kumar, N. and G. K. Jha. 2013. A time series ann approach for weather forecasting. *International Journal of Control Theory and Computer Modelling (IJCTCM)* 3(1):19-25.
11. Mallat, S. G. 1998. *A Wavelet Tour of Signal Processing*. 3rd Edition: Elsevier.
12. 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.
13. Mekanik, F., M. A. Imteaz, S. Gato-Trinidad and A. Elmahdi. 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology* 503: 11-21.
14. Nagahamulla, H. R., U. R. Ratnayake and A. Ratnaweera. 2012. An ensemble of artificial neural networks in rainfall forecasting. In *Advances in ICT for Emerging Regions (ICTer)*, International Conference on (176-181). IEEE.
15. 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.
16. Nourani, V., A. Hosseini Baghanam, J. Adamowski and O. Kisi. 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: A review. *Journal of Hydrology* 514: 358-377.
17. Purnomo, H. D., K. D. Hartomo and S. Y. J. Prasetyo. 2017. Artificial neural network for monthly rainfall rate prediction. IOP Conference Series: Materials Science and Engineering. IOP Publishing.

Optimization of Gene Expression Programming Model using Wavelet Transform for Simulating Long-term Rainfall in Anzali City

F. Hayati, A. Rajabi*, M. A. Izadbakhsh and S. Shabanlou¹

(Received: February 9-2020 ; Accepted: July 28-2020)

Abstract

Due to drought and climate change, estimation and prediction of rainfall is quite important in various areas all over the world. In this study, a novel artificial intelligence (AI) technique (WGEP) was developed to model long-term rainfall (67 years period) in Anzali city for the first time. This model was combined using Wavelet Transform (WT) and Gene Expression Programming (GEP) model. Firstly, the most optimized member of wavelet families was chosen. Then, by analyzing the numerical models, the most accurate linking function and fitness function were selected for the GEP model. Next, using the autocorrelation function (ACF), the partial autocorrelation function (PACF) and different lags, 15 WGEP models were introduced. The GEP models were trained, tested and validated in 37, 20- and 10-years periods, respectively. Also, using sensitivity analysis, the superior model and the most effective lags for estimating long-term rainfall were identified. The superior model estimated the target function with high accuracy. For instance, correlation coefficient and scatter index for this model were 0.946 and 0.310, respectively. Additionally, lags 1, 2, 4 and 12 were proposed as the most effective lags for simulating rainfall using hybrid model. Furthermore, results of the superior hybrid model were compared with GEP model that the hybrid model had more accuracy.

Keywords: Hybrid model, Rainfall, Time series, Simulation, Sensitivity analysis

1. Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.
Corresponding author, Email: ahmad.rajabi1974@gmail.com