

Research Paper

Estimation of Long-Term Rainfall in Babolsar City by Using the Optimized Gene Expression Programming

Yosef Esmaeli¹, Ahmad Rajabi², Fariborz Yosefvand³, Saeid Shabanlou⁴

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

² Associate Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran, CORRESPONDING AUTHOR: EMAIL: ahmad.rajabi1974@gmail.com

³ Assistant Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran

⁴ Associate Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran



10.22125/IWE.2022.162632

Received:
April 26, 2020
Accepted:
January 20, 2021
Available online:
December 11, 2022

Keywords:
Rainfall of Babolsar city, Time series, Gene expression programming, Wavelet transform, Modeling

Abstract

Estimation and simulation of precipitation is considered as one of the most issues in the field of hydrology. In this study, for the first time, the long-term rainfall in Babolsar city during a 68 years period from 1951 to 2019 was predicted by using an optimized hybrid artificial intelligence (AI) technique. To end this, the gene expression programming (GEP) model was combined with the wavelet transform. To training the AI models, 70% of the observed values were utilized and 30% of these values were used to testing those models. Additionally, the autocorrelation function (ACF) was applied to identify the most influential lags and then six GEP models were developed by means of these detected lags. The number of optimized genes was selected to be four. In addition, the Multiplication function was introduced as the best linking function of the GEP model. The superior GEP model was introduced through a sensitivity analysis that the correlation coefficient (R) and scatter index (SI) for this model were calculated to be 0.571 and 0.792, respectively. The (t-1), (t-2), (t-12), and (t-36) time series lags were introduced as the most effective input lags. The coif was detected as the best mother wavelet to simulate the target function. The hybrid WGEP model simulated the values of rainfall with acceptable accuracy. In the other words, the wavelet transform enhanced the performance of the GEP model significantly. For instance, the value of variance accounted for (VAF) for the GEP and WGEP models were respectively computed to be 31.710 and 82.064

1. Introduction

Rainfall is a major component of the water cycle, and also is responsible for precipitation of fresh water on the Earth. Estimation and simulation of precipitation is considered as one of the most issues in the field of hydrology.

2. Materials and Methods

In this study, for the first time, the long-term rainfall in Babolsar city during a 68 years period from 1951 to 2019 was predicted by using an optimized hybrid artificial intelligence (AI) technique. It is

worth mentioning that Babolsar is located on the southern coast of the Caspian Sea. Since the Babol river passes through the city, both the bank of the river and the beach of the sea. Babolsar has a Mediterranean climate with hot, steamy, but mostly dry summers and cool winters. Rainfall may occur at any time of the year but is heaviest in autumn and winter. Babolsar is leased on average 884.6 mm of rainfall per year, or 73.7 mm per month. The driest weather is in May when an average of 19.3 mm of rainfall occurs, whereas the wettest weather is in October as an average of 163.2 mm of rainfall happens. To end this, the gene expression programming (GEP) model was combined with the wavelet transform. The GEP is an evolutionary algorithm that produces computer programs or models. These computer programs are complex tree structures that learn and adapt by changing their sizes, shapes, and composition, much like a living organism. And like living organisms, the computer programs of GEP are also encoded in simple linear chromosomes of fixed length. Thus, GEP is a genotype–phenotype system, benefiting from a simple genome to keep and transmit the genetic information and a complex phenotype to explore the environment and adapt to it (Ferreira 2001). Moreover, in mathematics terminology, a wavelet series is a representation of a square-integrable (real- or complex-valued) function by a certain orthonormal series generated by a wavelet. This article provides a formal, mathematical definition of an orthonormal wavelet and of the integral wavelet transform. A time series is a series of data points indexed (or listed or graphed) in time order (Wang and Ding, 2003). Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus, it is a sequence of discrete-time data. To training the AI models, 70% of the observed values were utilized and 30% of these values were used to testing those models. Additionally, the autocorrelation function (ACF) was applied to identify the most influential lags. The autocorrelation function (ACF) defines how data points in a time series are related, on average, to the preceding data points. In other words, it measures the self-similarity of the signal over different delay times. In the next step, six GEP models were developed by means of these detected lags.

3. Results

The number of optimized genes was selected to be four. In addition, the Multiplication function was introduced as the best linking function of the GEP model. The superior GEP model was introduced through a sensitivity analysis that the correlation coefficient (R) and scatter index (SI) for this model were calculated to be 0.571 and 0.792, respectively. The (t-1), (t-2), (t-12), and (t-36) time series lags were introduced as the most effective input lags. The coif was detected as the best mother wavelet to simulate the target function.

4. Discussion and Conclusion

The hybrid WGEP model simulated the values of rainfall with acceptable accuracy. In the other words, the wavelet transform enhanced the performance of the GEP model significantly. For instance, the value of variance accounted for (VAF) for the GEP and WGEP models were respectively computed to be 31.710 and 82.064.

5. Six important references

- 1) 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.
- 2) Dabral, P. P. and M. Z. Murry. 2017. Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes*, 4(2): 399–419.
- 3) Ferreira, C. 2001. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. *Complex Systems*, 13(2): 87–129.
- 4) Hardwinarto, S. and M. Aipassa. 2015. Rainfall monthly prediction based on artificial neural network: A case study in Tenggara Station, East Kalimantan–Indonesia. *Procedia Computer Science*, 59: 142–151.

- 5) 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.

- 6) Wang, W. and J. Ding. 2003. Wavelet network model and its application to the prediction of the hydrology. *Nat. Sci.* 1 (1): 67–71.

Conflict of Interest

Authors declared no conflict of interest.

Acknowledgments

We are grateful to



تخمین بارش درازمدت شهر بابلسر با استفاده از برنامه‌ریزی بیان ژن بهینه‌یافته

یوسف اسماعیلی^۱، احمد رجبی^۲، فریبرز یوسفوند^۳، سعید شهبانلو^۴

تاریخ ارسال: ۱۳۹۹/۰۲/۰۷

تاریخ پذیرش: ۱۳۹۹/۱۱/۰۱

مقاله پژوهشی

چکیده

تخمین و شبیه‌سازی بارندگی یکی از مهمترین حوزه‌های علم هیدرولوژی محسوب می‌شود. در این مطالعه، برای اولین بار بارش دراز مدت شهر بابلسر در یک بازه ۶۸ ساله از سال ۱۹۵۱ تا ۲۰۱۹ به صورت ماهانه توسط یک مدل هوش مصنوعی ترکیبی بهینه‌یافته پیش‌بینی شد. برای انجام اینکار، مدل برنامه‌ریزی بیان ژن (GEP) با تبدیل موجک (WT) ترکیب شدند. برای آموزش مدل‌های هوش مصنوعی از ۷۰ درصد مقادیر مشاهداتی و برای آزمون آنها از ۳۰ درصد باقیمانده استفاده گردید. همچنین، با استفاده از تابع خود همبستگی (ACF) تاخیرهای موثر داده‌های سری زمانی شناسایی شدند و با استفاده از آنها، شش مدل GEP مختلف توسعه داده شدند. تعداد ژن‌های بهینه‌سازی شده مدل GEP مساوی با چهار انتخاب گردید. علاوه بر این تابع Multiplication به‌عنوان بهترین تابع اتصال مدل GEP معرفی گردید. مدل برتر GEP با انجام یک تحلیل حساسیت معرفی شد که مقادیر ضریب همبستگی (R) و شاخص پراکندگی (SI) آن به ترتیب مساوی با ۰/۵۷۱ و ۰/۷۹۲ محاسبه شدند. تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-24) به‌عنوان موثرترین تاخیرهای داده‌های سری زمانی معرفی شدند. لازم به ذکر است که در بین موجک‌های مادر مختلف، coif به‌عنوان موجک مادر برتر معرفی شد و سپس با مدل GEP ترکیب گردید. مدل ترکیبی WGEP مقادیر بارش‌ها را با دقت قابل قبولی شبیه‌سازی کرد. به عبارت دیگر، تبدیل موجک دقت مدل سازی را به شکل محسوسی بهبود بخشید. به‌عنوان مثال مقدار شاخص عملکرد (VAF) برای مدل GEP و WGEP به ترتیب مساوی با ۳۱/۷۱۰ و ۸۲/۰۶۴ بودند.

واژه‌های کلیدی: بارش شهر بابلسر، برنامه‌ریزی بیان ژن، تبدیل موجک، سری زمانی، مدل سازی

^۱ دانشجوی دکتری منابع آب، گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران، ۰۹۱۸۳۵۷۸۹۲۷،

yfesmaelie@gmail.com

^۲ دانشیار گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران، ۰۹۱۸۸۳۲۵۲۳۷، ahmad.rajabi1974@gmail.com

(نویسنده مسئول)

^۳ استادیار گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران، ۰۹۱۸۳۳۶۳۲۴۱، fariborzosefvand@gmail.com

^۴ دانشیار گروه مهندسی آب، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران، ۰۹۱۸۸۱۲۸۲۲۵، saeid.shabanlou@gmail.com

مقدمه

بارندگی مهمترین پدیده هیدرولوژیکی است که افزایش یا کاهش آن در طول دوره‌های مختلف می‌تواند باعث تغییرات محسوس در اکوسیستم منطقه و تغییر جمعیت شود. در سال‌های اخیر، به دلیل تغییرات اقلیمی میزان بارش در نواحی مختلف کشور ایران دستخوش تغییرات قابل توجهی شده است. بنابراین تخمین و مدل سازی بارندگی در این کشور خشک و نیمه خشک از اهمیت بسزایی برخوردار است. علاوه بر این، مدل‌های هوش مصنوعی به شکل گسترده‌ای برای شبیه‌سازی پدیده‌های مختلف از قبیل بارندگی مورد استفاده قرار گرفته‌اند. به عنوان مثال، Lorrai and Sechi (1995)، Savic et al. (1999) و Nasser et al. (2008) مقادیر بارندگی در نقاط مختلف کره زمین را توسط مدل‌های هوش مصنوعی شبیه‌سازی نمودند. علاوه بر این، Nagahamulla et al. (2012) تکنیک‌های مختلف شبکه عصبی مصنوعی را برای شبیه‌سازی مقدار بارندگی در کشور سریلانکا استفاده نمودند. آنها با ارزیابی عملکرد مدل‌های بکار گرفته شده نشان دادند که مدل عصبی ENN^۱ مقادیر بارندگی‌ها را با دقت بهتری تخمین زد. سپس، Hardwinarto and Aipassa (2015) بارنگی ماهانه در یک بازه ۲۲ ساله یک ناحیه واقع در کشور اندونزی را توسط شبکه عصبی مدل سازی نمودند. آنها در انتها بیان نمودند که مدل هوش مصنوعی مقادیر تابع هدف را با دقت قابل قبولی تخمین زد. Shoaib et al. (2015) مدل برنامه‌نویسی بیان ژن را با تبدیل موجک ترکیب کردند و مقادیر بارندگی‌ها در حوضه‌های آبریز مختلف در سراسر دنیا شبیه‌سازی کردند. آنها نتایج مدل ترکیبی را با مدل برنامه‌نویسی بیان ژن نیز مقایسه کردند که مدل ترکیبی از عملکرد بهتری برخوردار بود. علاوه بر این، Dabral and Murry (2017) داده‌های سری زمانی بارش در یک بازه ۱۴ ساله حوضه واقع در شرق کشور هندوستان را توسط مدل ساریما^۲

در بازه‌های زمانی روزانه، هفتگی و ماهانه تخمین زدند. آنها تاخیرهای موثر را نیز شناسایی نمودند و نشان دادند که مدل مذکور مقادیر بارش‌ها را با دقت خوبی شبیه سازی کرد. همچنین، Chang et al. (2018) با استفاده از یک مدل نرو- فازی تحت عنوان شبکه استنتاج فازی وابسته به خود مقدار بارش- رواناب را برای حوزه مختلف در کشورهای چین، سوئد و استرالیا مدل سازی نمودند. آنها نشان دادند که مدل نرو- فازی مذکور از انعطاف مناسبی برای شبیه‌سازی بارندگی برخوردار است. سپس، دانش Mehr et al. (2019) یک مدل ترکیبی بهینه یافته را برای تخمین مقادیر بارش دراز مدت شهر تبریز و ارومیه توسعه دادند. آنها برای بهینه سازی مدل هوش مصنوعی رگرسیون بردار پشتیبان از الگوریتم کرم شب‌تاب بهره گرفتند. علاوه بر این، غفاری و وفاخواه (۱۳۹۲) فرآیند بارش- رواناب به صورت روزانه، در حوزه آبخیز حاجی- قوشان را با استفاده از شبکه‌های عصبی مصنوعی و مدل انفیس شبیه‌سازی نمودند. دهقانی و همکاران (۱۳۹۵) جریان روزانه در حوزه آبخیز کسلیان را با استفاده از شبکه عصبی مصنوعی و شبکه عصبی- فازی تطبیقی شبیه‌سازی کردند. علاوه بر این، آنها تأخیرهای داده‌های سری زمانی را با استفاده از نمودار خودهمبستگی جزئی شناسایی کردند. در این مطالعه نشان داده شد که شبکه عصبی- فازی تطبیقی مقادیر تابع هدف را با دقت بهتری پیش‌بینی نمود. بااعلی و دهقانی (۱۳۹۶) 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 فلاحی و همکاران (۱۳۹۹) عملکرد مدل برنامه‌نویسی بیان ژن را با استفاده از تبدیل موجک جهت تخمین مقادیر بارش شهر رشت توسعه

¹ Ensemble Neural Network

² SARIMA



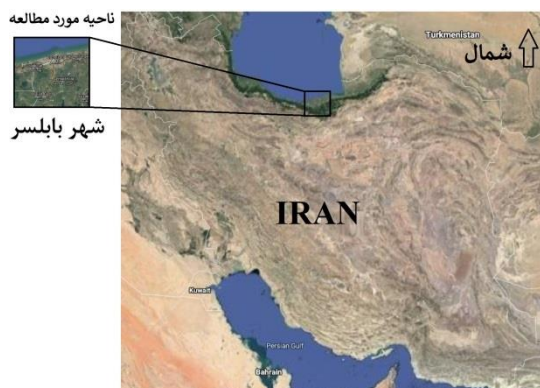
مواد و روش‌ها

ناحیه مورد مطالعه

در این مطالعه برای آموزش و آزمون مدل‌های هوش مصنوعی از مقادیر بارش‌های گزارش شده توسط سازمان هواشناسی کشوری شهر بابلسر در یک بازه زمانی ۶۸ ساله از سال ۱۹۵۱ تا ۲۰۱۹ استفاده شد. برای آزمون مدل‌های هوش مصنوعی از ۴۹ سال این داده‌های سری زمانی استفاده خواهد شد و ۱۹ سال باقی‌مانده برای آزمون این مدل‌ها به کار گرفته می‌شوند. شهر بابلسر در مدار معتدل گرم قرار گرفته و ضریب اعتدال این شهر بسیار معتدل است. این ناحیه با میانگین سالیانه دمای ۱۸/۴ درجه سانتیگراد، بارش سالیانه ۷۹۱ میلی‌متر و ۷۹ درصد رطوبت یکی از خوش آب و هواترین مناطق کشور به‌شمار می‌آید. از نظر توپوگرافی سراسر بابلسر در بخش جلگه‌ای و در موقعیت ساحلی قرار دارد و از لحاظ ناهمواری عوارض برجسته توپوگرافیک در آن دیده نمی‌شود. تقریباً هموار بوده و با شیب ملایمی از سوی رشته کوه‌های البرز به سوی دریای خزر امتداد دارد. بلندای شهر ۲۱ متر پایین‌تر از سطح دریاهای آزاد و ۷ متر بالاتر از سطح دریای مازندران قرار دارد. طول جغرافیایی شهرستان بابلسر ۵۲ درجه و ۳۹ دقیقه و عرض آن ۳۶ درجه و ۴۲ دقیقه است. تصویر ناحیه مورد مطالعه در شکل ۱ قابل مشاهده است. لازم به ذکر است که طبق آمار ثبت شده در سازمان هواشناسی کشوری، ایستگاه بابلسر یکی از کاملترین ایستگاه‌ها اندازه‌گیری بارش در کل کشور محسوب می‌شود. بنابراین، داده‌های این ایستگاه برای مطالعه حاضر استفاده شد.

دادند. نتایج مدل‌های هوش مصنوعی در یک بازه دراز مدت مورد صحت‌سنجی قرار گرفت. مدل مذکور در هر دو شرایط آموزش و آزمون عملکرد مناسبی از خود نشان داد. علاوه بر این، مطالعات بررسی شده در مقدمه نشان می‌دهد که محققین مختلف تنها از داده‌های یک ایستگاه جهت اعتبارسنجی مدل هوش مصنوعی خود استفاده کرده‌اند. علاوه بر این، داده‌های بارندگی در سایر اقلیم‌ها دارای نواقص و کمبودهایی از قبیل نبود بارش در بخش قابل توجهی از ماه‌های سال و وجود بازه‌های زمانی کوتاه‌تر، بودند. بنابراین نویسندگان تصمیم گرفتند که در مطالعه حاضر از مقادیر بارش‌های درازمدت شهر بابلسر جهت صحت‌سنجی مدل هوش مصنوعی استفاده کنند.

لازم به ذکر است که به دلیل تغییرات اقلیمی و به تبع آن افزایش چشمگیر میزان باندگی در نقاط مختلف جهان علی‌الخصوص در کشور خشک و نیمه خشک ایران، مدل‌سازی این پدیده از اهمیت بسزایی برخوردار است. علاوه بر این، مدل‌های هوش مصنوعی دارای مزایایی فراوانی از قبیل سرعت بالا، دقت قابل قبول و ارزان بودن ابزارهای بسیار کاربردی جهت شبیه‌سازی و تخمین الگوی بارش محسوب می‌شوند. در این مطالعه، برای اولین بار یک مدل ترکیبی بهینه یافته هوش مصنوعی تحت عنوان "موجک-برنامه‌ریزی بیان ژن" جهت مدل‌سازی بارندگی دراز مدت ۶۸ ساله از سال ۱۹۵۱ تا ۲۰۱۹ شهر بابلسر استان مازندران ارائه گردید. این مدل ترکیبی توانایی مناسبی جهت پیش‌بینی بارش درازمدت از خود نشان داد. لازم به ذکر است که تبدیل موجک دقت مدل برنامه‌ریزی بیان ژن را به شکل قابل توجهی افزایش داد.



شکل (۱) تصویر ناحیه مورد مطالعه

(۱) ناحیه سر مربوط به یک ژن می‌تواند نشانه‌های مختلفی را داشته باشد بطوریکه نودهای مربوط به توابع و ترمینال در درخت‌های اصطلاحات استنتاج شده را نشان دهد اما ناحیه دم تنها شامل ترمینال‌ها می‌باشد. نشانه "Q" ریشه دوم، "M" ماکزیمم دو مقدار، "*" ضرب و "+" جمع است. علاوه بر این، نشانه‌های "a" و "b" نشان دهنده ثابت‌های عددی یا متغیرهای برآورد شده هستند (نودهای ترمینال).

(۲) طول ناحیه سر (h) در یک کروموزوم به اختیاز انتخاب می‌شود و با استفاده از آن، طول دم (t) به صورت $t = h(n-1) + 1$ محاسبه می‌شود. بطوریکه n تعداد آرگومان‌هایی است که یک تابع با تاخیر از پیش تعیین شده می‌تواند حمل کند، می‌باشد. h=6 و n برابر تعداد آرگومان‌های استفاده شده در تابع M است. بنابراین مقدار $t=7$ بدست می‌آید. بنابراین، شکل و اندازه یک درخت اصطلاحات به تعداد و نوع توابع نشانه‌هایی که در سر مورد استفاده قرار می‌گیرند، وابسته است. اگر سر پر باشد از اصطلاحات ریاضی، هر یک از آن‌ها دارای دو آرگومان باشد، همه نواحی که کدگذاری نشده‌اند، کدگذاری می‌شوند. اندازه درخت اصطلاحات برابر با ماکزیمم مقدار آن، ۱۳، می‌شود.

(۳) کروموزوم‌هایی که بیش از یک ژن دارند، اغلب به منظور ایجاد گوناگونی در افراد مورد استفاده قرار می‌گیرند. هر ژن کروموزوم نشان دهنده یک زیر مجموعه از درخت اصطلاحات مربوطه است. باید توجه

برنامه ریزی بیان ژن^۱

برنامه ریزی بیان ژن (Ferreira, 2001) یک تکنیک جدید در زمینه هوش مصنوعی تکاملی است که از برنامه نویسی ژنتیک، گرفته شده است. برنامه ریزی بیان ژن یک سیستم فنوتیپ-ژنوتیپ است که از عملگرهای ژنتیکی مختلفی استفاده می‌کند. در این سیستم، ژنوتیپ مربوط به هر فرد یک رشته از نمادهای خطی را نشان می‌دهد که تحت عنوان کروموزوم شناخته می‌شوند و ممکن است دارای تعداد مختلفی ژن با طول ثابت (یک یا بیشتر) باشد. همچنین رشته‌های غیرخطی تحت فنوتیپ‌ها به صورت درخت‌های اصطلاحات مطابق شکل (۱-ب) ارائه می‌دهد. علاوه بر اینکه برنامه نویسی بیان ژن، درخت‌های اصطلاحات را به راحتی و به طور مستقیم از رشته‌های خطی دستکاری شده می‌خواند، بلکه توانایی ارائه یک رابطه ریاضی را بدون از دست دادن پیچیدگی‌های محاسباتی را داراست. روند انتقال اطلاعات ژنتیکی از یک ژن در یک درخت اصطلاحات با استفاده از سیستم ترجمه انجام می‌شود بطوریکه به منظور جلوگیری از شکل‌گیری درخت‌های اصطلاحات غیر مجاز و اطمینان از اینکه تمام پیام‌های احتمالی و تغییرات ایجاد شده در ژن توانایی ترجمه به درخت‌های اصطلاحات را داشته باشد. قوانین مورد استفاده را می‌توان بدین شرح ارائه داد:

¹ Gene Expression Programming (GEP)

$$\int_{-\infty}^{+\infty} \psi_{a,b}(t) dt = 0 \quad (1)$$

که $\psi_{a,b}(t)$ به صورت زیر محاسبه می شود:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), a, b \in R, a \neq 0 \quad (2)$$

که a و b فاکتورهای فرکانس و زمان (به ترتیب) هستند، R دامنه اعداد واقعی و $\psi_{a,b}(t)$ موجک متوالی است. برای یک سیگنال مانند $f(t) \in L^2(R)$ ، تبدیل موجک متوالی آن $f(t)$ به شکل زیر ارائه می شود:

$$W_{\psi} f(a,b) = |a|^{-1/2} \int_R f(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (3)$$

که $\bar{\psi}(t)$ بیانگر توابع پیچیده $\psi(t)$ است. با توجه به رابطه فوق، واضح است که تبدیل موجک به چندین تابع $f(t)$ در سطوح مختلف، تجزیه می شود. در واقع، ماهیت تبدیل موجک فیلتر کردن موج $f(t)$ با استفاده از فیلترهای مختلف است.

با در نظر گرفتن $a = a_0^j$ ($a_0 > 1$)، $b = kb_0 a_0^j$ ($b_0 \in R$) و z و k به عنوان اعداد مثبت، تبدیل موجک گسسته به صورت زیر تعریف می شود:

$$W_{\psi} f(j,k) = a_0^{-j/2} \int_R f(t) \bar{\psi}(a_0^{-j}t - kb_0) dt \quad (4)$$

با در نظر گرفتن a_0 و b_0 برابر با ۲ و ۱، رابطه فوق به شکل تبدیل موجک باینری به شکل زیر، بازنویسی می شود:

$$W_{\psi} f(j,k) = 2^{-j/2} \int_R f(t) \bar{\psi}(2^{-j}t - k) dt \quad (5)$$

مشخصات سری زمانی اصلی در دامنه زمان (b) یا (k) و فرکانس (a یا j) در زمان مشابه به صورت $W_{\psi} f(j,k)$ و $W_{\psi} f(a,b)$ (به ترتیب)، ارائه می شود. اگر تجزیه فرکانس مربوط به تبدیلی موجک کم باشد اما تجزیه دامن هزمان زیاد باشد، مقادیر پارامترهای a یا z کوچک می باشد به طور عکس، زیاد بودن مقدار تجزیه فرکانس و کم بودن تجزیه دامنه زمان منتج به بزرگ بودن مقادیر a یا z میشود (وانگ و دینگ، ۲۰۰۳). تابع تبدیل موجک گسسته برای یک سری زمانه گسسته به صورت $f(t)$ (عدد صحیح است) که

کرد که مقدار شایستگی هر زیر درخت اصطلاحات در طول دوره تکاملی ارزیابی می شود.

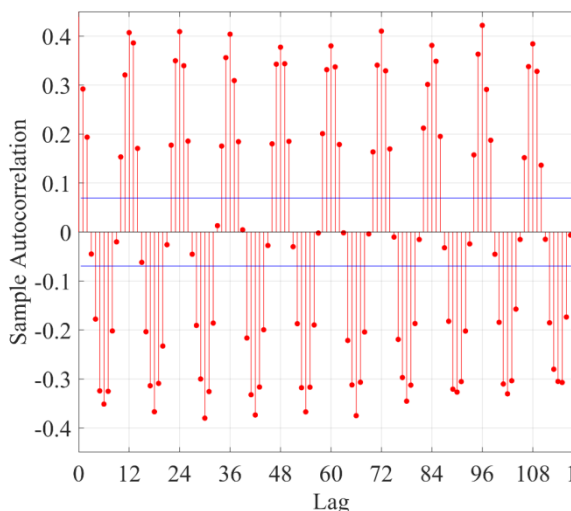
۴) تولید مجدد ژنوتیپ مربوط به یک فرد که دارای بهترین میزان شایستگی در میان عمخیرهای ژنتیکی است، یک قسمت مهم در فرایند تکاملی به حساب می آید. با توجه به ساختار رشته های خطی مربوط به ساختارهای ژنتیکی، امکان ویرایش و اصلاح آن ها بر اساس فنوتیپ با استفاده از عمخیرهای ژنتیکی مستقیماً به فنوتیپ ترجمه می شوند، وجود دارد.

در مدل سازی مسائل مختلف با استفاده از برنامه ریزی بیان ژن به منظور افزایش مقدار شایستگی، دارای یک روند کلی است که به طور خلاصه بیان می شود. همانند سایر روش های تکاملی، اولین مرحله از مدلسازی، ایجاد تصادفی یک جمعیت اولیه از افراد است (ژنوتیپ). در ادامه هر یک از این افراد، که با استفاده از سیستم ترجمه به یک مدل رگرسیونی ترجمه شده اند (ژنوتیپ)، به صورت ترکیبی از توابع از پیش تعریف شده جهت مدل سازی و اصطلاحات محلی می تواند متغیرهای پیش بینی و ثابت های مربوطه را تشکیل دهد. احتمال وجود هر فرد در نسل بعدی به مقدار تابع ارزیابی شایستگی بستگی دارد. نسل های دوم و پس از آن، در طول فرایند "تولید مجدد" تولید می شوند به طوری که در این فرایند یک فرد با یک مقدار شایستگی بالا، دارای بالاترین شانسی برای انتخاب در نسل بعدی و انتقال مشخصات ژنتیکی آن به فرزندان است. پروسه "تولید مجدد" و کنترل مقدار تابع شایستگی تا رسیدن به جواب بهینه یا شرایط توقف مدل، ادامه می یابد. فلوجارت روش برنامه ریزی بیان ژن به صورت زیر می باشد.

تبدیل موجک^۱

تابع موجک که تحت عنوان mother wavelet ($\psi(t)$) شناخته می شود، به صورت زیر تعریف می شود:

¹ Wavelet transform



شکل (۲): تابع خود همبستگی مربوط به داده های بارش ماهانه شهر بابل

پس از تعیین مدل های مختلف، نوع موجک مادر و همچنین سطح تجزیه، تعریف شود. در این مطالعه، شش نوع موجک مادر شامل Hahr (haar), Coiflets, Symlets (sym), Daubechies (db), Dmeyer (dmey) و BiorSplines (bior), (coif) می باشد. پس از انتخاب موجک مادر، نوبت به تعیین سطح تجزیه می رسد. به منظور تعیین سطح تجزیه برای استفاده از موجک مادر، در این مطالعه از رابطه زیر استفاده می شود (Nourani et al., 2014):

$$DL = \text{int} [Ln(Z)] \quad (8)$$

در این رابطه، Z تعداد نمونه های مورد استفاده در آموزش مدل، int جز صحیح مقدار بدست آمده برای سطح تجزیه (DL) است. البته باید توجه داشت که استفاده از سطح تجزیه بالا، هر چند که ممکن است دقت مدلسازی را افزایش دهد، اما پیچیدگی قابل توجه مدل را به همراه خواهد داشت. بنابراین، نیازی نیست که همیشه بالاترین سطح تجزیه مجاز را مورد استفاده قرار دهیم و بعضاً با استفاده از سطح تجزیه های پایین نیز می توان به نتایج قابل قبولی دست یافت. شکل ۳ چگونگی تجزیه توسط تبدیل موجک برای سطوح مختلف را نشان می دهد. اگر هر یک از

در گام های زمانی مختلف اتفاق می افتد، به صورت زیر است:

$$W_{\psi} f(j, k) = 2^{-j/2} \sum_{t=0}^{N-1} f(t) \psi(2^{-j} t - k) \quad (6)$$

که $W_{\psi} f(j, k)$ بیانگر ضرایب موجک برای تابع موجک گسسته با مقایس های $a=2^j$ و $b=2^j k$ است.

روش ترکیبی WGEP

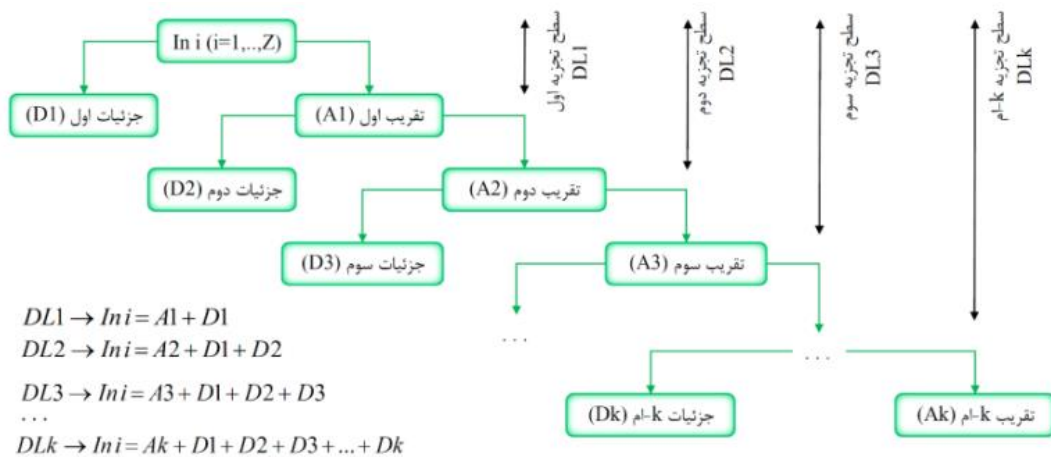
در این مدل ترکیبی، جهت پیش بینی سری زمانی بارندگی ماهانه در بابل، از تریب روش های برنامه ریزی بیان ژن (GEP) و تبدیل موجک (Wavelet) استفاده می شود. روند مدلسازی در این روش، با تعریف پارامترهای ورودی، آغاز می شود. جهت تعیین موثرترین تاخیر ها، از تابع خود همبستگی (ACF^1) استفاده می شود. با رسم این تابع برای داده های بارندگی ماهانه در شهر بابل (شکل ۲)، مهمترین تاخیر ها به صورت تاخیر های ۱، ۲ می باشند. علاوه بر این، با توجه به مشاهده روند ماهانه تغییرات، اثر تاخیر ۱۲ نیز در نظر گرفته می شود. علاوه بر این، اثر تاخیر های فصلی بیش از ۱۲، مانند ۲۴ و ۳۶ به صورت مدل های ۵ و ۶، در نظر گرفته می شود. بنابراین، مدل های تعریف شده برای این مطالعه، به شرح زیر می باشد:

$$\begin{aligned} M1: R(t) &= f(R(t-1), R(t-2)) \\ M2: R(t) &= f(R(t-1), R(t-12)) \\ M3: R(t) &= f(R(t-2), R(t-12)) \\ M4: R(t) &= f(R(t-1), R(t-2), R(t-12)) \\ M5: R(t) &= f(R(t-1), R(t-2), R(t-12), R(t-24)) \\ M6: R(t) &= f(R(t-1), R(t-2), R(t-12), R(t-36)) \end{aligned} \quad (7)$$

¹ Autocorrelation function

ورودی i -ام در سطوح تجزیه مختلف، در شکل ۳ ارائه شده است. استفاده از تبدیل موجک گسسته منجر به تولید دو مجموعه توابع می‌شود به طوریکه سری زمانی اصلی به سری های تقریب که روند اصلی سری را نشان می‌دهد و جزئیات تجزیه می‌شوند.

ورودی های مدل را به صورت In_i در نظر بگیریم، تجزیه این ورودی در سطوح مختلف، منجر به چند برابر شدن ورودی های مدل می‌شود. اما همانطور که گفته شد، در نظر گرفتن یک الی دو سطح، نه تنها مدل را پیچیده نمی‌کند، بلکه توانایی پیش بینی و تعمیم پذیری آن را نیز افزایش می‌دهد. تجزیه



شکل (۳): چگونگی تجزیه توسط تبدیل موجک برای سطوح مختلف

$$SI = \frac{RMSE}{\bar{O}} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |F_i - O_i| \quad (13)$$

در اینجا O_i مقادیر مشاهداتی، F_i مقادیر تراز پیش‌بینی شده به وسیله مدل‌های عددی، \bar{O} میانگین مقادیر مشاهداتی و n برابر تعداد مقادیر مشاهداتی است. نزدیک بودن مقدار R به یک به معنای بالا بودن همبستگی مدل هوش مصنوعی و نزدیک بودن VAF به عدد صد به معنای عملکرد مناسب مدل مورد نظر است. نزدیک بودن شاخص‌های $RMSE$ ، SI و MAE به عدد صفر نشان دهنده دقت بالای مدل عددی است. در این مطالعه، ۷۰ درصد مقادیر مشاهداتی بارش برای آموزش مدل‌های هوش مصنوعی و ۳۰ درصد باقیمانده برای آزمون مدل‌های مذکور استفاده

معیارهای بررسی دقت مدل‌های عددی

در مطالعه حاضر به منظور ارزیابی دقت مدل‌های عددی معرفی شده از شاخص‌های آماری ضریب همبستگی (R)، شاخص عملکرد (VAF)، خطای مربعات میانگین ریشه ($RMSE$)، شاخص پراکندگی (SI) و خطای مطلق میانگین (MAE) به صورت زیر استفاده می‌گردد:

$$R = \frac{\sum_{i=1}^n (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (F_i - \bar{F})^2 \sum_{i=1}^n (O_i - \bar{O})^2}} \quad (9)$$

$$VAF = \left(1 - \frac{\text{var}(F_i - O_i)}{\text{var}(F_i)} \right) \times 100 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2} \quad (11)$$

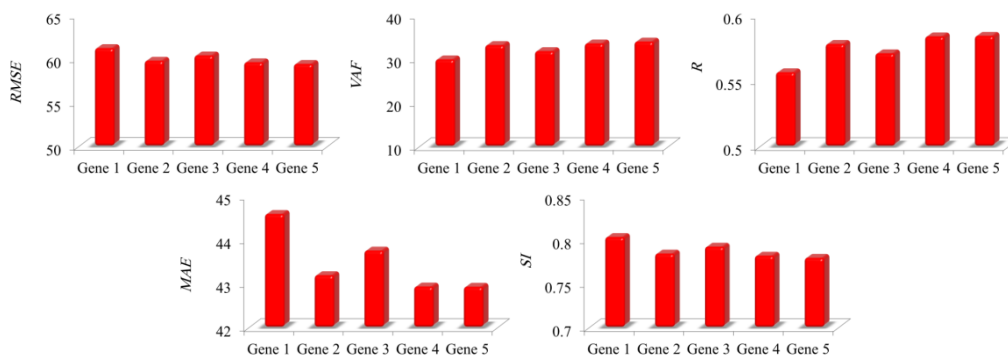
بهینه‌سازی تعداد ژن‌ها

در این بخش به بررسی تعداد ژن‌ها پرداخته می‌شود. نتایج مقایسه کلیه شاخص‌های آماری مختلف برای تعداد ژن‌ها در شکل ۴ نشان داده شده است. بر اساس نتایج مدل‌سازی، هنگامی که تعداد ژن‌ها مساوی با یک است دقت مدل عددی بیشتر از سایر حالت‌ها است. به عنوان مثال هنگامی که تعداد ژن‌ها برابر یک است مقادیر R ، MAE و VAF به ترتیب برابر با ۰/۵۵۶، ۴۴/۵۸ و ۲۹/۸۲۹ بود. لازم به ذکر است که شاخص‌های آماری SI و $RMSE$ برای این تعداد ژن به ترتیب مساوی ۰/۸۰۳ و ۶۱/۱۹۱ محاسبه شدند. همچنین مقایسه بارش شبیه‌سازی شده توسط مدل برنامه نویسی بیان ژن هنگامی که تعداد ژن‌ها مساوی با ۴ است و نمودار پراکندگی آن در شکل ۵ به تصویر کشیده شده است. در این شرایط دقت مدل عددی در حد قابل قبولی بود، به عبارت دیگر خطای مدل عددی به کمترین مقدار خود و همبستگی آن به بیشترین مقدار خود رسید. بنابراین تعداد ژن‌ها مساوی با ۴ فرض می‌شود. برای مدلی که تعداد ژن‌های آن مساوی با ۴ است، مقادیر R ، VAF و SI به ترتیب مساوی با ۰/۵۸۴، ۳۳/۴۶۸ و ۰/۷۸۲ بدست آمدند. مقایسه بارش‌های شبیه‌سازی شده توسط مدلی که دارای چهار ژن است با مقادیر بارندگی‌های مشاهداتی در قالب شکل ۵ به-تصویر کشیده شده است

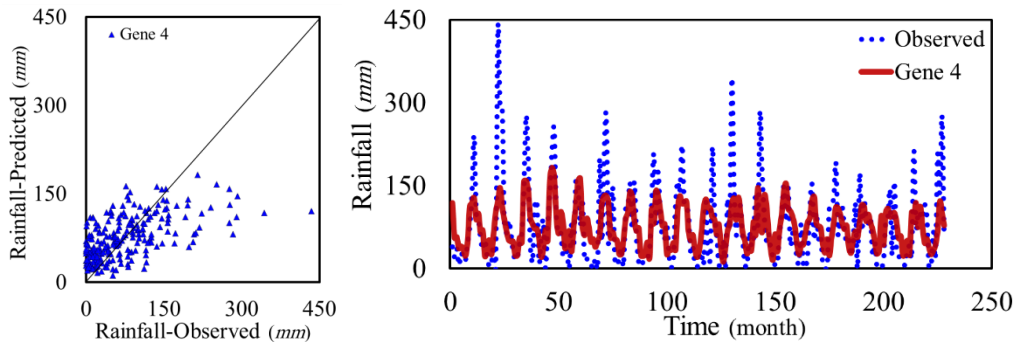
می‌شود. لازم به ذکر است که در ادامه کلیه نتایج ارائه شده برای وضعیت آزمون مدل‌های هوش مصنوعی ارائه می‌شوند.

نتایج و بحث

در این بخش در ابتدا تعداد ژن‌های مدل برنامه‌ریزی بیان ژن بهینه‌سازی می‌شود و بهترین تعداد برای این پارامتر بدست می‌آید. سپس توابع اتصال مختلف مدل برنامه‌ریزی بیان ژن مورد ارزیابی قرار می‌گیرند. در ادامه، شش مدل GEP با استفاده از تاخیرهای مختلف داده‌های سری زمانی توسعه داده می‌شوند و بهترین مدل با انجام یک تحلیل حساسیت شناسایی شده و موثرترین تاخیرها نیز معرفی می‌گردند. لازم به ذکر است که برای تولید مدل ترکیبی "موجک-برنامه-ریزی بیان ژن" ($WGEP$) از ترکیب تبدیل موجک و مدل بهینه‌سازی شده GEP استفاده می‌شود. شایان ذکر است که در ادامه موجک‌های مادر مختلف مورد ارزیابی قرار خواهند گرفت و بهترین موجک مادر انتخاب شده و با مدل GEP برتر ترکیب شده و مدل ترکیبی بهینه یافته ارائه خواهد شد.



شکل (۴): شاخص‌های آماری محاسبه شده برای ژن‌های مختلف

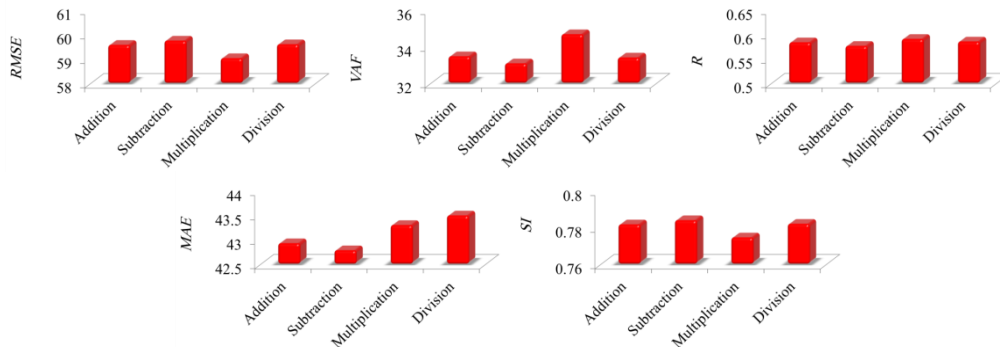


شکل (۵): نتایج بارش‌های شبیه‌سازی شده توسط تعداد ژن بهینه‌سازی شده (تعداد ژن برابر با چهار)

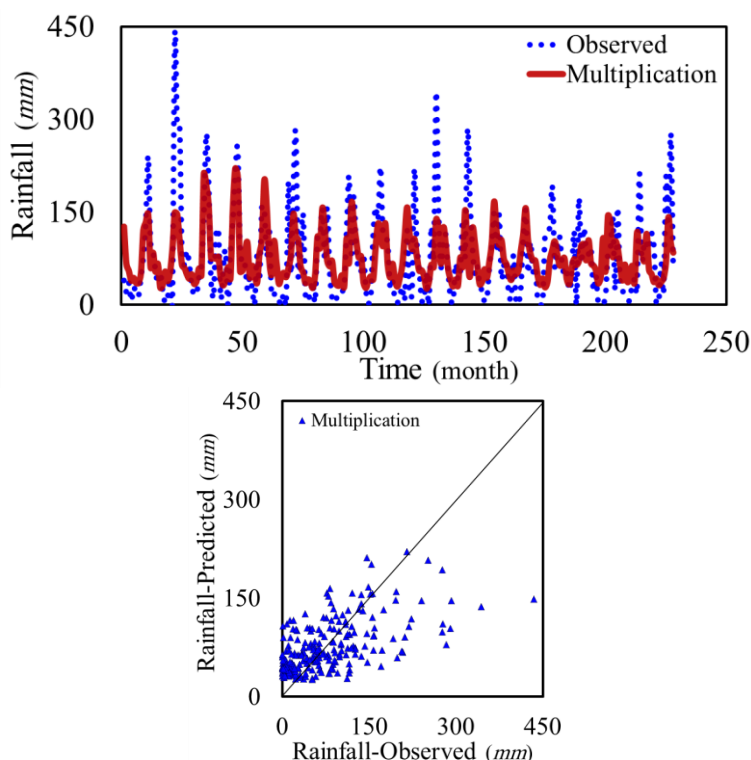
ترتیب مساوی با $۵۹/۷۴۴$ و $۵۹/۰۲۸$ محاسبه شدند. لازم به ذکر است که برای تابع اتصال addition مقادیر ضریب همبستگی و شاخص پراکندگی به ترتیب مساوی با $۰/۷۸۲$ و $۰/۵۸۴$ تخمین زده شدند. همچنین نمودارهای پراکندگی و مقایسه نتایج تابع اتصال برتر در اینجا تابع multiplication است با مقادیر مشاهداتی در شکل ۷ قابل مشاهده است. بنابراین با توجه به بررسی توابع اتصال، در ادامه از تابع اتصال multiplication برای تخمین مقادیر بارندگی توسط مدل هوش مصنوعی ترکیبی استفاده می‌شود.

انتخاب تابع اتصال

در قسمت بعدی به بهینه‌سازی تابع اتصال مدل برنامه‌ریزی بیان ژن پرداخته می‌شود. در شکل ۶ نتایج شاخص‌های آماری مختلف برای توابع اتصال addition، subtraction، division و multiplication به تصویر کشیده شده است. به عنوان مثال مقدار شاخص‌های آماری SI ، R و VAF برای تابع اتصال division به ترتیب مساوی با $۰/۷۸۲$ ، $۰/۵۸۵$ و $۳۳/۴۰۱$ بدست آمدند. همچنین مقدار $RMSE$ برای توابع اتصال subtraction و multiplication به



شکل (۶): نتایج شاخص‌های آماری محاسبه شده برای توابع اتصال مختلف



شکل (۷): نتایج بارش‌های شبیه‌سازی شده توسط تابع اتصال برتر

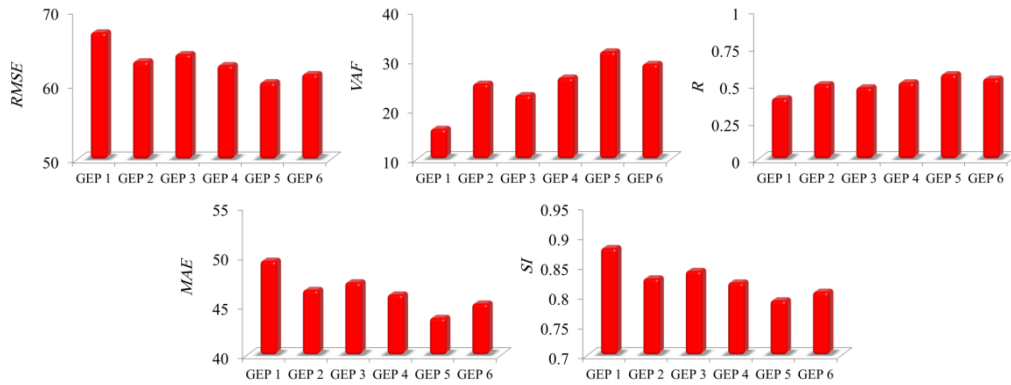
بین کلیه مدل‌های هوش مصنوعی، مدل شماره یک (GEP 1) دارای بیشترین خطا و کمترین دقت است. علاوه بر این، مدل GEP 2 تابعی از تاخیرهای (t-1) و (t-12) است. برای این مدل مقادیر ضریب همبستگی و شاخص پراکندگی به ترتیب مساوی ۰/۵۰۳ و ۰/۸۸۰ است. همچنین مقدار RMSE برای GEP 2 مساوی ۶۷/۰۴۷ محاسبه گردیده است. برای مدل GEP 3 مقادیر بارندگی‌ها بر حسب (t-2) و (t-12) مدل سازی می‌کند. برای این مدل مقدار شاخص R مساوی ۰/۴۸۲ است. همچنین مقادیر RMSE و MAE برای مدل GEP 3 به ترتیب مساوی ۶۴/۱۲۶ و ۴۷/۳۳۸ تخمین زده شده است. برای مدل GEP 4 مقادیر SI و RMSE به ترتیب برابر ۰/۸۲۲ و ۶۲/۶۴۴ محاسبه شده است. علاوه بر این، مقدار ضریب همبستگی برای این مدل مساوی ۰/۵۱۵ محاسبه شده است. این مدل مقادیر بارش‌ها را بر حسب تاخیرهای شماره (t-1)، (t-2) و (t-12) مدل سازی می‌کند. لازم به ذکر است که مدل GEP 5 مقادیر تابع هدف را بر حسب تاخیرهای شماره

مدل‌های برنامه‌ریزی بیان ژن (GEP)

در مطالعه حاضر، جهت صحت سنجی مدل‌های هوش مصنوعی، نتایج این مدل‌ها در وضعیت آزمون با استفاده از شاخص‌های آماری مختلف مورد ارزیابی قرار گرفتند. به بیان دیگر، با توجه به شاخص‌های آماری استفاده شده، مدل‌ها از لحاظ دقت و همبستگی مورد تجزیه و تحلیل قرار گرفتند. در این مطالعه شش مدل مختلف GEP برای تخمین ضریب دبی معرفی می‌شود. در شکل ۸ نتایج شاخص‌های آماری محاسبه شده برای مدل‌های GEP 1 تا GEP 5 نشان داده شده است. همچنین در شکل ۹ مقایسه نتایج مدل برتر GEP با مقادیر مشاهداتی و نمودارهای پراکندگی آنها به تصویر کشیده شده است. در مدل GEP 1 مقادیر بارش را بر حسب تاخیرهای شماره (t-1) و (t-2) پیش‌بینی می‌کند. برای مدل GEP 1 مقادیر RMSE، MAE و ضریب همبستگی به ترتیب مساوی ۶۷/۰۴۷، ۴۹/۵۱۵ و ۰/۴۰۹ محاسبه گردیده است. همچنین مقدار VAF برای این مدل مساوی ۱۶/۰۸۵ بدست آمده است. در

۰/۵۴۱ بودند. شایان ذکر است که مدل GEP 6 مقادیر بارندگی‌ها را بر حسب تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-36) تخمین می‌زند. بنابراین با توجه به نتایج مدل‌سازی‌ها، مدل شماره 5 GEP دارای بیشترین دقت بود. این مدل مقادیر تابع هدف را بر حسب تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-24) تخمین زد. لازم به ذکر است که تاخیرهای داده‌های سری زمانی شماره (t-1)، (t-2)، (t-12) و (t-24) موثرترین تاخیرها معرفی شدند

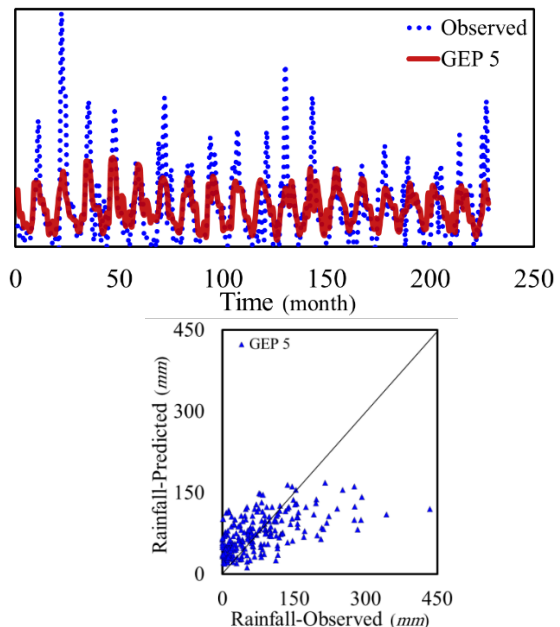
تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-24) تخمین می‌زند. همچنین، مقادیر R، VAF و SI برای این مدل هوش مصنوعی به ترتیب مساوی با ۰/۵۷۱، ۳۱/۷۱۰ و ۰/۷۹۲ بدست آمدند. علاوه بر این، مقادیر شاخص‌های آماری RMSE و MAE برای ۶۰/۳۵۴ و ۴۳/۷۵۰ تخمین زده شدند. برای مدل GEP 6 نیز مقادیر VAF، SI و MAE به ترتیب برابر با ۲۹/۱۹۸، ۰/۸۰۶ و ۴۵/۱۹۲ محاسبه شدند. این در حالی است که برای این مدل مقادیر RMSE و R به ترتیب برابر با ۶۱/۴۵۳ و



شکل (۸): شاخص‌های آماری محاسبه شده برای مدل‌های مختلف GEP

موجک‌های مادر

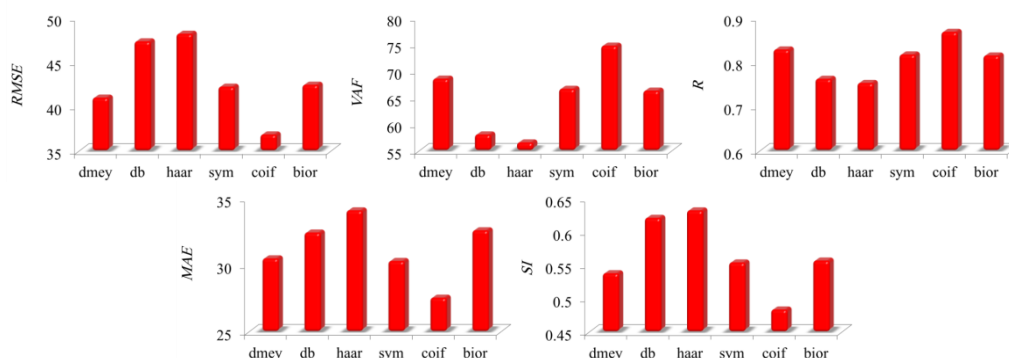
در این قسمت به ارزیابی دقت خانواده‌های مختلف تبدیل موجک پرداخته می‌شود. به بیان دیگر، تبدیل موجک دارای خانواده‌هایی با عنوان داوبیچیز (db)، هار (haar)، سایملتس (sym)، کویفلتس (coif)، باپورثوگنال (bior) و دی میبر (dmey) می‌باشد. با توجه به تجزیه و تحلیل کلیه اعضای خانواده‌های مختلف تبدیل موجک مشخص شد که خانواده کویفلتس (coif) دارای بیشترین دقت و کمترین مقدار خطا در مقایسه با سایر اعضای خانواده‌های موجک بود. در شکل ۱۰ نتایج شاخص‌های آماری برای اعضای برتر خانواده‌های مختلف موجک قابل مشاهده است. به‌عنوان مثال برای موجک مادر haar شاخص‌های آماری R، MAE و SI به ترتیب برابر با ۰/۷۵۲، ۲۷/۵۴۶ و ۰/۴۸۳ بودند. همچنین، برای موجک مادر db مقادیر RMSE، MAE و به ترتیب مساوی با ۴۷/۳۱۶ و ۳۲/۴۳۴ بدست آمدند. علاوه بر این، برای موجک مادر sym مقادیر



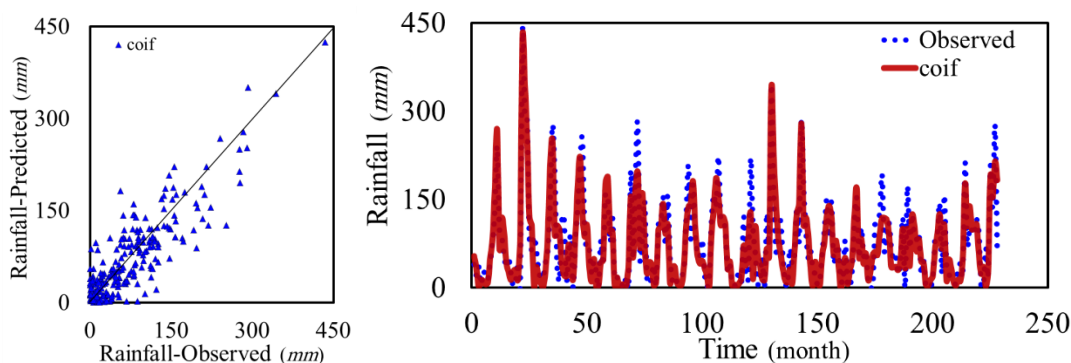
شکل (۹): نتایج بارش‌های شبیه‌سازی شده توسط مدل

برتر GEP 5

SI و MAE به ترتیب برابر با ۰/۸۱۴، ۰/۵۵۷ و ۳۲/۶۰۸ بودند. بنابراین، با توجه به نتایج کلیه اعضای خانواده-های موجک، عضو coif به عنوان دقیق ترین و بهینه ترین عضو خانواده های موجک معرفی شد و در ادامه مطالعه این عضو خانواده موجک برای شبیه سازی بارش دراز مدت شهر بابلسر مورد استفاده قرار خواهد گرفت. همچنین مقایسه مقادیر بارش های شبیه سازی شده توسط موجک های مادر مختلف با مقادیر مشاهداتی در شکل ۱۱ نشان داده شده است.



شکل (۱۰): شاخص های آماری محاسبه شده برای موجک های مادر مختلف



شکل (۱۱): نتایج بارش های شبیه سازی شده توسط موجک مادر برتر (coif)

ترکیبی WGEP به ترتیب مساوی با ۰/۳۷۴، ۰/۹۰۶ و ۸۲/۰۶۳ محاسبه شدند. علاوه بر این، مقادیر MAE و RMSE برای این مدل ترکیبی به ترتیب مساوی با ۱۸/۳۸۴ و ۲۱/۴۹۳ تخمین زده شده است. نشان شبیه سازی ها نشان داد که مدل ترکیبی موجک- برنامه نویسی بیان ژن مقادیر بارندگی های شهر بابلسر با همبستگی مناسب (شاخص R)، عملکرد قابل قبول (شاخص VAF) و دقت بالا (شاخص های $RMSE$ ، SI و MAE) شبیه سازی نمود.

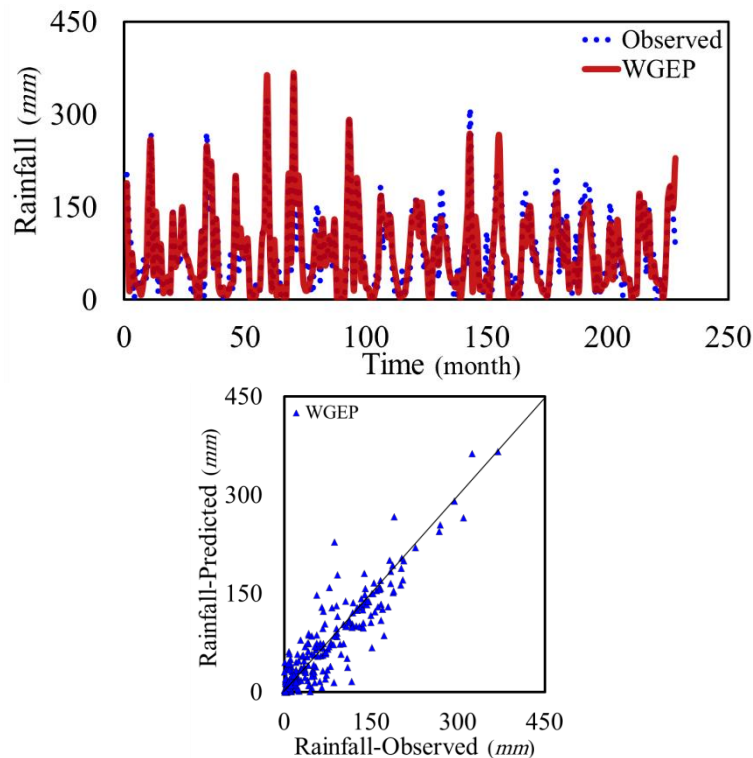
$RMSE$ و R ، VAF به ترتیب مساوی با ۰/۸۱۶، ۰/۵۹۴ و ۴۲/۲۱۷ تخمین زده شدند. برای موجک coif، شاخص عملکرد و ضریب همبستگی به ترتیب برابر با ۷۴/۶۹۵ و ۰/۸۶۷ محاسبه شدند. بر اساس نتایج شبیه سازی مقدار بارندگی، در میان کلیه اعضای مختلف خانواده موجک عضو coif به عنوان بهترین عضو این خانواده معرفی شد. برای آن شاخص های آماری MAE و $RMSE$ نیز به ترتیب برابر با ۳۰/۳۰۳ و ۴۲/۲۱۷ تخمین زده شدند. لازم به ذکر است که برای bior مقادیر شاخص های R ،

مدل ترکیبی WGEP

در این قسمت به ارزیابی توانایی بررسی دقت مدل هوش مصنوعی ترکیبی بهینه یافته تحت عنوان WGEP پرداخته می شود. در جدول ۱ نتایج شاخص های آماری محاسبه شده برای این مدل مرتب شده است. علاوه بر این، شکل ۱۲ نتایج بارش های شبیه سازی شده توسط مدل ترکیبی WGEP را نشان می دهد. با توجه به نتایج شبیه سازی ها، مقادیر شاخص پراکندگی، ضریب همبستگی و شاخص عملکرد مدل

جدول (۱): نتایج شاخص‌های آماری محاسبه شده برای مدل WGEP

	<i>R</i>	<i>VAF</i>	<i>RMSE</i>	<i>SI</i>	<i>MAE</i>
WGEP	۰/۹۰۶	۸۲/۰۶۳	۲۱/۴۹۳	۰/۳۷۴	۱۸/۳۸۴



شکل (۱۲): نتایج بارش‌های شبیه‌سازی شده توسط مدل ترکیبی WGEP

تحلیل حساسیت نشان داد که تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-24) مهمترین و تاثیرگذارترین تاخیر-های داده‌های سری زمانی بودند. مدل ترکیبی برتر مقادیر بارش درازمدت شهر بابلسر را با عملکرد قابل قبولی پیش‌بینی کرد.

نتیجه‌گیری

در این مطالعه، بارش دراز مدت شهر بابلسر در یک بازه ۶۸ (از سال ۱۹۵۱ تا ۲۰۱۹) توسط یک مدل هوش مصنوعی ترکیبی بهینه‌سازی شده شبیه‌سازی شد. برای اینکار، مدل برنامه ریزی بیان ژن (GEP) و تبدیل موجک (wavelet transform) با هم ترکیب شدند.

بحث

در این مطالعه، یک مدل بهینه‌سازی شده جهت تخمین مقادیر بارندگی‌های درازمدت شهر بابلسر توسعه داده شد. به عبارت دیگر، موجک مادر بهینه، تعداد بهینه ژن-ها و تابع اتصال بهینه مدل GEP با اجرای یک روند سعی و خطا شناسایی شدند. تجزیه و تحلیل نتایج شبیه‌سازی‌ها در مطالعه حاضر نشان داد که تبدیل موجک عملکرد مدل برنامه‌نویسی بیان ژن را به شکل قابل توجهی بهبود بخشید. به عبارت دیگر، مدل ترکیبی برنامه‌نویسی بیان ژن (WGEP) در مقایسه با مدل ساده GEP از دقت بیشتر و همبستگی بالاتری با مقادیر مشاهداتی برخوردار بود. علاوه بر این، نتایج

GEP 5 به عنوان مدل برتر معرفی شد. برای این مدل مقادیر R ، SI و MAE به ترتیب مساوی با ۰/۵۷۱، ۰/۷۹۲ و ۴۳/۷۵۰ محاسبه شدند. این در حالی بود که تاخیرهای شماره (t-1)، (t-2)، (t-12) و (t-24) به عنوان موثرترین تاخیرهای داده‌های سری زمانی شهر بابل معرفی شدند. سپس، موجک‌های مادر مختلف مورد ارزیابی قرار گرفتند که $coif$ به عنوان بهترین موجک مادر انتخاب شد. نتایج شبیه‌سازی‌ها نشان داد که تبدیل موجک دقت مدل‌سازی را به شکل قابل توجهی افزایش داد. به عنوان مثال، مقدار شاخص عملکرد (VAF) مدل‌های GEP برتر و WGEP به ترتیب مساوی با ۳۱/۷۱۰ و ۸۲/۰۶۴ بدست آمدند

لازم به ذکر است که ۷۰ درصد مقادیر بارش‌های مشاهداتی برای آموزش مدل‌های هوش مصنوعی و ۳۰ درصد باقیمانده برای آزمون این مدل‌ها استفاده گردید. همچنین، با استفاده از تابع خود همبستگی (ACF) تاخیرهای تاثرگذار داده‌های سری زمانی شناسایی شدند و توسط آنها شش مدل GEP مختلف توسعه داده شد. علاوه بر این، با استفاده از یک روند سعی و خطا، تعداد ژن‌ها و توابع اتصال مدل برنامه‌ریزی بیان ژن بهینه‌سازی شدند. به عبارت دیگر، تعداد بهینه ژن‌ها مساوی با چهار انتخاب گردید و تابع اتصال Multiplication نیز به عنوان بهترین تابع اتصال معرفی گردید. با اجرای یک تحلیل حساسیت، مدل

منابع

- باباعلی، ح. و ر. دهقانی. ۱۳۹۶. مقایسه مدل‌های هوشمند در تخمین بارش ماهانه حوضه کاکارضا. اکولوژی، سال ۴، شماره ۱، ص ۱۱-۱.
- دهقانی، ر. م. علی قربانی، م. تشنه لب، ا. ریخته گر غیاثی و ا. اسدی. ۱۳۹۴. مقایسه و ارزیابی مدل‌های شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی در تخمین بده جریان؛ مطالعه موردی حوضه صوفی چای. مجله آبیاری و آب ایران، سال ۵، شماره ۴، ص ۸۵-۶۵.
- دهقانی، ن. م. وفاخواه و ع. بهره مند. ۱۳۹۵. مدل سازی بارش-رواناب با استفاده از شبکه عصبی مصنوعی و شبکه فازی-عصبی تطبیقی در حوزه آبخیز کسلیان. پژوهشنامه مدیریت حوزه آبخیز، سال هفتم، شماره ۱۳، ص ۱۳۷-۱۲۸.
- غفاری، غ. م. وفاخواه. ۱۳۹۲. شبیه سازی فرآیند بارش-رواناب با استفاده از شبکه عصبی مصنوعی و سیستم فازی-عصبی تطبیقی (مطالعه موردی: حوزه آبخیز حاجی قوشان). پژوهشنامه مدیریت حوزه آبخیز، سال ۴، شماره ۱۰، ص ۱۲۰-۱۳۶.
- فلاحی م م، یعقوبی ب، یوسفوند ف، شعبانلو س. بهبود عملکرد مدل برنامه‌نویسی بیان ژن با استفاده از تبدیل موجک برای تخمین بارش درازمدت شهر رشت. مجله علوم آب و خاک. ۱۳۹۹. ۲۴ (۳): ص ۱۶-۱.
- محمدرضاپور، ا. ۱۳۹۶. پیش بینی تبخیر- تعرق پتانسیل ماهانه با استفاده از مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی ژنتیک و سیستم استنتاج عصبی - فازی. مجله آبیاری و آب ایران، سال ۷، شماره ۳، ص ۱۵۰-۱۳۵.
- 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.
- Dabral, P. P. and M. Z. Murry. 2017. Modelling and forecasting of rainfall time series using SARIMA. Environmental Processes, 4(2): 399-419.
- Ferreira, C. 2001. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex Systems, 13(2): 87-129.
- Hardwinarto, S. and M. Aipassa. 2015. Rainfall monthly prediction based on artificial neural network: A case study in Tenggara Station, East Kalimantan-Indonesia. Procedia Computer Science, 59: 142-151.



- Lorrai, M. and G. M. Sechi. 1995. Neural nets for modelling rainfall-runoff transformations. *Water resources management*, 9(4): 299-313.
- 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.
- 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)* (pp. 176-181). IEEE.
- Nasseri, 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.
- 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.
- 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.
- Shoaib, M., Shamseldin, A. Y., Melville, B. W., and Khan, M. M. 2015. Runoff forecasting using hybrid wavelet gene expression programming (WGEP) approach. *Journal of Hydrology*, 527: 326-344.
- Wang, W. and J. Ding. 2003. Wavelet network model and its application to the prediction of the hydrology. *Nat. Sci.* 1 (1): 67–71.
- 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.



Estimation of Long-Term Rainfall in Babolsar City by Using the Optimized Gene Expression Programming

Yosef Esmaeli¹, Ahmad Rajabi², Fariborz Yosefvand³, Saeid Shabanlou⁴

Abstract

Estimation and simulation of precipitation is considered as one of the most issues in the field of hydrology. In this study, for the first time, the long-term rainfall in Babolsar city during a 68 years period from 1951 to 2019 was predicted by using an optimized hybrid artificial intelligence (AI) technique. To end this, the gene expression programming (GEP) model was combined with the wavelet transform. To training the AI models, 70% of the observed values were utilized and 30% of these values were used to testing those models. Additionally, the autocorrelation function (ACF) was applied to identify the most influential lags and then six GEP models were developed by means of these detected lags. The number of optimized genes was selected to be four. In addition, the Multiplication function was introduced as the best linking function of the GEP model. The superior GEP model was introduced through a sensitivity analysis that the correlation coefficient (R) and scatter index (SI) for this model were calculated to be 0.571 and 0.792, respectively. The (t-1), (t-2), (t-12), and (t-36) time series lags were introduced as the most effective input lags. The coif was detected as the best mother wavelet to simulate the target function. The hybrid WGEP model simulated the values of rainfall with acceptable accuracy. In the other words, the wavelet transform enhanced the performance of the GEP model significantly. For instance, the value of variance accounted for (VAF) for the GEP and WGEP models were respectively computed to be 31.710 and 82.064

Keywords: Rainfall of Babolsar city, Time series, Gene expression programming, Wavelet transform, modeling

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

² Associate Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran, CORRESPONDING AUTHOR: EMAIL: ahmad.rajabi1974@gmail.com

³ Assistant Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran

⁴ Associate Professor, Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran