

Streamflow prediction based on hybrid Empirical Mode Decomposition and artificial intelligence methods

Amir Amirashayeri¹, Javad Behmanesh^{*2}, Nasrin Fathollahzadeh Attar³

¹PhD student, Water Engineering Department, Urmia University, Urmia, Iran

²Professor of Water Engineering, Department of Water Engineering, Urmia University, Urmia, Iran

³Ph.D. Graduated, Department of Water Engineering, Urmia University, Urmia, Iran

Corresponding Author's E-mail: j.behmanesh@urmia.ac.ir

Received: 16 Feb. 2022

Accepted: 15 July 2022



Homepage: ijwer.uoz.ac.ir

Abstract: The correct and accurate estimation of the river flow using different models is a significant issue in water resources research. In this research, two hydrometric stations of Sari-Qomish and Nizam-Abad, located in West Azerbaijan province, were used to estimate the daily flow of the Zarrineh Rud River accurately. To reach this aim, the Empirical Mode Decomposition (EMD) preprocessing algorithm was used to deal with the complexity and instability of time series data. EMD is a data analysis method for extracting signals in data generation through non-linear and non-stationary operations. In this research, the Gene Expression Programming model and Artificial Neural Network model methods were used. The results of the study showed that the performance of the Gene Expression Programming model was equal to and sometimes less than the performance of the Artificial Neural Network. However, the combination of the two mentioned models with the EMD technique increased the accuracy of the model and reduced the error in simulating the river flow in Sari-Qomish and Nizam-Abad stations.

Keywords: Streamflow, Prediction, Gene Expression model, Empirical Mode Decomposition, Artificial Neural Network.

Introduction: Streamflow prediction has been one of the most critical challenges in water resources management in recent decades, so researchers have used different methods to do so. Typically, physics-based numerical models are used for modeling or better modeling in order to estimate streamflow (Partington et al., 2012). In recent years, the use of artificial intelligence methods such as Artificial Neural Network (ANN), Support Vector Machines (SVM), and Gene Expression Programming (GEP) to model and solve problems in water engineering due to their considerable benefits, is current (Rezaie-Balf et al., 2019). The Empirical Mode Decomposition (EMD) was proposed as a noise-based data analysis method by (Huang et al., 1998), which separates the higher frequency input series into the solved (frequency) components with lower frequency. Amirashayeri et al. Studied three stations in West Azerbaijan to predict daily reference transpiration evaporation with ANN models, Tree models, and a combination of these models with Empirical Mode Decomposition (EMD). The results showed that the Empirical Mode Decomposition hybrid method with the Tree model was the best model (Amirashayeri et al., 2020). The purpose of this study was to provide an accurate prediction model for streamflow at Sari-Qomish and Nizam-Abad stations on the Zarrineh Rud River in the Urmia Lake basin, using hybrid Gene Expression Programming (GEP) models and Artificial Neural Network (ANN) models with the Empirical Mode Decomposition (EMD) to have a more accurate combination for daily flow in the desired stations.

Methodology: In this study, to accurately estimate the daily streamflow of the Zarrineh Rud River, two hydrometric stations of Sari-Qomish and Nizam-Abad located in West Azerbaijan province were used. Recently, well-known AI-based intelligent models, along with hybrid models, have been accepted in all predictions. In this study, the Empirical Mode Decomposition (EMD) algorithm was used to extract signals in the information generated by nonlinear and non-static operations to deal with the complexity and instability of time series data. In this study, The Gene Expression Programming (GEP) model, which is part of the circulatory algorithm, and The Artificial Neural Network model were used. ANN model simulates the ability of the human brain to find patterns

and learn from trial and error. The daily streamflow of the Zarrineh Rud River in West Azerbaijan province was used during a 26-year period, investigated, and then estimated. The accuracy of the models is evaluated by three indexes. These three indexes are Mean Absolute Error (*MAE*), Root Mean Squared Error (*RMSE*) and Correlation Coefficient (R^2).

Results and Discussion: The results showed that the performance of the Gene Expression Programming model was equal to and sometimes lower than the performance of the Artificial Neural Network (ANN). However, the combination of the two models with the EMD algorithm increased the model's accuracy and reduced the error in simulating streamflow in Sari-Qomish and Nezam-Abad stations. The results showed that in the test phase, combining the GEP method with EMD improved the coefficient of correlation by 7.29% and 2.15% for the Sari-Qomish and Nezam-Abad stations, respectively. Also, in the test phase, a combination of the ANN method and EMD improved the coefficient of explanation by 3.06% and 3.09% for the Sari-Qomish and Nezam-Abad stations, respectively.

Conclusion: Since river flow forecasting plays a vital role in resource planning and management, the correct and accurate estimation of the river flow using different intelligent artificial models is one of the issues that researchers in water resources are investigating. In this research, the daily flow data of two stations of Sari-Qomish and Nizam-Abad in a period of 26 years (1990 to 2016) and up to three-time delays were applied for estimating the river flow through prediction models, Artificial Neural Network (ANN) model, and Gene Expression Programming (GEP). The preprocessing method of the Empirical Mode Decomposition (EMD) algorithm was used to increase the accuracy of the two models. The results showed that pre-processing data using the experimental mode analysis algorithm increases the accuracy of the Gene Expression Programming model by 29.7% and 2.15% for the Sari-Qomish and Nizam-Abad stations, respectively. Furthermore, the proposed combination method, for example, at the Sari-Qomish station has far less error than the Nizam-Abad station in daily discharge forecasting, and it increases the accuracy of the indexes using the Artificial Neural Network model.

© 2024 University of Zabol, Zabol, Iran.



This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (<http://creativecommons.org/licenses/by/4.0/>)

پیش‌بینی جریان رودخانه با استفاده از روش ترکیبی هوشمند مصنوعی بر پایه الگوریتم پیش‌پردازنده تجزیه مد تجربی

امین امیر عشایری^۱، جواد بهمنش^{۲*}، نسرین فتح‌اله زاده عطاری^۳

^۱ دانشجوی دوره دکتری مهندسی آب، دانشگاه ارومیه، ارومیه، ایران

^۲ استاد، گروه مهندسی آب، دانشگاه ارومیه، ارومیه، ایران

^۳ دانش‌آموخته دوره دکتری، گروه مهندسی آب، دانشگاه ارومیه، ارومیه، ایران

پست الکترونیکی نگارنده مسئول: j.behmanesh@urmia.ac.ir

وب‌گاه نشریه: ijwer.uoz.ac.ir



تاریخ دریافت: ۱۴۰۰/۱۱/۲۷

تاریخ پذیرش: ۱۴۰۱/۰۴/۲۸

چکیده: نیاز روزافزون به آب سبب گردیده است برنامه‌ریزی‌های مدیریتی به‌منظور کنترل مصرف آب در آینده، از اهمیت روزافزونی برخوردار شود. به همین دلیل، برآورد صحیح و دقیق جریان رودخانه با استفاده از مدل‌های مختلف، یکی از موضوعاتی است که پژوهشگران در منابع آب به بررسی آن می‌پردازند. در این تحقیق جهت برآورد دقیق مقدار روزانه جریان رودخانه زربینه‌رود، از دو ایستگاه هیدرومتری ساری-قمیش و نظام‌آباد واقع در استان آذربایجان غربی استفاده شد. اخیراً شیوه استفاده از مدل‌های هوشمند به‌صورت ترکیبی مورد پذیرش قرار گرفته است. در این تحقیق، برای مقابله با پیچیدگی و ناپایداری داده‌های سری زمانی از الگوریتم پیش‌پردازش تجزیه مد تجربی^۱ (EMD) استفاده شد. EMD یک روش تجزیه اطلاعات برای استخراج سیگنال‌ها در اطلاعات تولیدشده توسط عملیات غیرخطی و غیرایستا است. در این تحقیق، از روش مدل برنامه‌ریزی بیان ژن و از مدل شبکه عصبی مصنوعی استفاده شد. نتایج تحقیق نشان داد که مدل برنامه‌ریزی بیان ژن، عملکردی برابر و گاهی کمتر از عملکرد شبکه عصبی مصنوعی دارد. باین‌حال ترکیب دو مدل مذکور با فن (EMD) باعث افزایش دقت مدل و کاهش خطا در شبیه‌سازی جریان رودخانه در ایستگاه‌های ساری-قمیش و نظام‌آباد شد. نتایج نشان داد که در مرحله آزمون، ترکیب روش GEP^۲ با EMD ارتقا ضریب تبیین به میزان ۷/۲۹٪ و ۲/۱۵٪ به ترتیب برای ایستگاه ساری-قمیش و نظام‌آباد را در پی دارد. همچنین در مرحله آزمون، ترکیب روش ANN^۳ با EMD باعث ارتقا ضریب تبیین به میزان ۳/۰۶٪ و ۳/۰۹٪ به ترتیب برای ایستگاه ساری-قمیش و نظام‌آباد گردید.

کلیدواژگان: جریان رودخانه، پیش‌بینی، مدل برنامه‌ریزی بیان ژن، الگوریتم تجزیه مد تجربی، مدل شبکه عصبی مصنوعی

^۱ Empirical Mode Decomposition

^۲ Gene Expression Programming

^۳ Artificial Neural Network

۱- مقدمه

یک سامانه هشدار زمان واقعی سیلاب در حوضه‌های شهری با استفاده از روش ANN و GEP مدل‌سازی کردند. آن‌ها نشان دادند که تجزیه و تحلیل حوضه‌های رودخانه‌ای با روش ANN و GEP نسبت به روش‌های تصادفی از قبیل هیدروگراف واحد به بهبود مدل‌سازی بارش-رواناب کمک می‌کند (Dorado et al., 2003). دستورانی و همکاران به بررسی کارایی الگوهای ANN و ANFIS² در الگوبندی بارش-رواناب در حوضه آبخیز سد زاینده‌رود پرداختند. مشاهده شد که هر دو الگو با به‌کارگیری پارامترهای ورودی و ساختار مناسب، نتایج قابل قبولی ارائه می‌دهند (Dastorani et al., 2012). نایاک و همکاران به بررسی عملکرد ANFIS در پیش‌بینی رواناب پرداختند. نتایج به‌دست‌آمده نشان داد که الگوی ANFIS در مقایسه با ANN و سری‌های زمانی به لحاظ سرعت محاسبات، خطای پیش‌بینی، کارایی و تخمین مقدار حداکثر رواناب، عملکرد بهتری دارد (Nayak et al., 2004).

شبیه‌سازی بار معلق در رودخانه تانگو واقع در ایالات متحده آمریکا حاکی از آن بود که روش GEP عملکرد بهتری نسبت به روش‌های منحنی سنج و رگرسیون خطی چندگانه دارد (Aytek et al., 2008). مقایسه توانایی دو روش ANN و GEP در پیش‌بینی جریان روزانه رودخانه شویل‌کیل در آمریکا نشان داد که روش ANN از دقت بالاتری نسبت به روش GEP برخوردار است؛ هرچند که روش GEP در برآورد میزان دبی دقت بالاتری دارد. (Guvén., 2009).

شوئیب و همکاران پیش‌بینی رواناب با به‌کارگیری مدل ترکیبی برنامه‌ریزی بیان ژن-موجک (WGEP³) را برای ۴ حوضه مختلف از سراسر جهان، با استفاده از داده‌های بارش-رواناب انجام دادند. آن‌ها ۱۰ تابع موجک مادر مختلف را بررسی کردند و نشان دادند که مدل ترکیبی WGEP با استفاده از تابع موجک Dmey⁴ دارای عملکرد بهتری نسبت به روش مدل منفرد GEP است (Shoabit et al., 2015). رضایی و همکاران نیز برای پیش‌بینی مواد جامد محلول در آب رودخانه کشکان از روش‌های برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان (SVM) و شبکه بیزین استفاده

ایران کشوری با متوسط ریزش‌های جوی ۲۵۰ میلی‌متر است که در زمره مناطق خشک و نیمه‌خشک جهان طبقه‌بندی می‌گردد (Taheri et al., 2019). محدودیت منابع آب و نبود آب با کیفیت مناسب از عمده‌ترین مشکلات کشاورزی در کشور است؛ بنابراین یافتن راه‌کارهایی برای غلبه بر کمبود آب در این مناطق همواره از اولویت‌های مطالعاتی و تحقیقاتی بوده است (Seifi et al., 2010). پیش‌بینی دبی جریان، یکی از چالش‌های مهم مدیریت منابع آب در دهه‌های اخیر، سبب شده است تا محققان روش‌های متفاوتی را برای این امر به‌کارگیرند. پیش‌بینی دقیق جریان (ماهانه یا روزانه) برای بهره‌برداری، برنامه‌ریزی و توزیع امکانات نیروگاه‌های آبی از اهمیت بالایی برخوردار است. همچنین، برای مدیریت بهینه منابع آب، پیش‌بینی جریان آب به‌ویژه پیش‌بینی‌های بلندمدت اهمیتی دوچندان دارد. تولید برق آبی، بهره‌برداری بهینه از مخزن، تخصیص منطقی منابع آب و حفاظت از محیط‌زیست برای تصمیم‌گیری‌ها نیز نیازمند پیش‌بینی رواناب هستند (Guo et al., 2011; Wang et al., 2009).

به‌طورمعمول، برای ارزیابی مدل‌ها یا شناخت مدل بهتر جهت پیش‌بینی جریان از مدل‌های عددی مبتنی بر فیزیک استفاده می‌شود (Partington et al., 2012). رفتار سری‌های رواناب رودخانه‌ها می‌توانند غیرخطی باشد، لذا استفاده از الگوهای خطی دقت لازم و کافی را ندارند و با توجه به اهمیت پیش‌بینی جریان آب رودخانه از نقطه‌نظرهای مختلف، یافتن روشی مناسب در این خصوص موردتوجه است.

در سال‌های اخیر، استفاده از روش‌های هوش مصنوعی نظیر شبکه عصبی مصنوعی (ANN)، ماشین‌های بردار پشتیبان (SVM¹) و برنامه‌ریزی بیان ژن (GEP) برای مدل‌سازی و حل مسائل در مهندسی آب (دارای داده‌های هیدرولوژیکی کافی) فراگیر شده است (Rezaie-Balf et al., 2019). دوردو و همکاران فرآیند بارش-رواناب را باهدف طراحی

³ Wavelet Transform-GEP⁴ Discrete Meyer Wavelet¹ Support Vector Machine² Adaptive Neuro-Fuzzy Inference System

۱ به‌عنوان ورودی‌های خود بهره می‌گیرد. خطی و همکاران این روش را به‌عنوان راهی برای به‌حداکثر رساندن اطلاعات استخراج‌شده و یا افزایش دقت مدل‌های محلی از طریق انتخاب تعداد محدودی از مدل‌ها و بدون آزمون همه مدل‌های موجود، ارائه کردند (Khatibi et al., 2020). بررسی دو رویکرد متفاوت تصمیم‌گیری چندمعیاره (MCDM³) و هوش مصنوعی (AI⁴) که در مطالعات هیدرولوژی و محیطی کاربرد هستند، ریسک مسائل آب-های زیرزمینی را برجسته می‌کند. MCDM موضوع را به‌عنوان یک مسئله تصمیم‌گیری فرموله می‌کند و هدف آن برآورد وزن معیارهای مختلف در یک روش از پیش تعریف‌شده است. موضوعات مختلفی با فن MCDM مورد مطالعه قرار گرفته است؛ مانند جلوگیری از آلاینده‌های زیرزمینی توسط فرآیند تحلیل سلسله مراتبی (Jhariya et al., 2019)، فرآیند شبکه تحلیلی (Garewal et al., 2015)، نمایه‌سازی پتانسیل آب‌های زیرزمینی با تجزیه و تحلیل تصمیم‌گیری چندمعیاره (Andualem et al., 2019)، و خطر آب زیرزمینی برای نفوذ آب شور توسط مدل فازی (Sadeghfam et al., 2020).

یکی از نقاط ضعف مدل‌های هوشمند، عدم در نظر گرفتن خاصیت تصادفی در فرآیند مدل‌سازی است، چراکه برای دستیابی به بهترین عملکرد، از روابط ریاضی مشخص استفاده می‌نمایند. از طرف دیگر ترکیب مدل‌های پیش‌پردازش با حذف کردن نویز و خطاهای مربوط به داده‌ها می‌تواند اطلاعات مفیدی را برای مدل‌های هوشمند فراهم نمایند.

هدف از انجام این تحقیق، ارائه مدلی دقیق جهت پیش‌بینی دبی روزانه در دو ایستگاه ساری-قمیش و نظام‌آباد، به‌ترتیب در محدوده صاین‌قلعه و میان‌دوآب واقع بر روی رودخانه زرینه‌رود در حوضه دریاچه ارومیه، در استان آذربایجان غربی است. در این راستا، مدل بیان ژن (GEP) و مدل شبکه عصبی مصنوعی (ANN) با روش پیش‌پردازنده تجزیه مدّ تجربی، ترکیب شده و با مقایسه پاسخ‌های به‌دست‌آمده، مدل برتر تعیین می‌گردد.

کردند. آن‌ها بیان نمودند که هر سه مدل، دارای دقت قابل قبول برای پیش‌بینی کیفیت آب هستند (Rezaei et al., 2019).

از آنجاکه، داده‌های هیدرولوژیکی نظیر دبی به‌صورت کلی غیرخطی و غیرثابت^۱ هستند، به‌کارگیری مستقیم آن‌ها جهت کالیبره کردن مدل‌های هوشمند، ممکن است دقت لازم را جهت پیش‌بینی دبی نداشته باشد. از این‌رو، استفاده از فن‌های پیش‌پردازش عمدتاً باعث بهبود عملکرد مدل‌ها می‌گردد.

پژوهشگران از این فن‌های پیش‌پردازش برای استخراج ویژگی‌های داده‌های سری زمانی مانند سطح آب زیرزمینی (Rezaei-Bif et al., 2017 a)، تبخیر (Ghaemi et al., 2019)، شاخص تشعشع خورشید (Deo et al., 2017)، سرعت باد (Hung et al., 2019)، شاخص خشک‌سالی (Ali et al., 2019) و پارامترهای کیفیت آب (Fijani et al., 2019) استفاده کرده‌اند.

روش تجزیه مدّ تجربی (EMD) به‌عنوان یک روش تجزیه داده‌ها مبتنی بر نویز توسط هوانگ و همکاران ارائه شد که سری ورودی فرکانس بالاتر را به مولفه‌های (سیگنال) با فرکانس پایین‌تر تفکیک می‌کند (Huang et al., 1998). امیر عشایری و همکاران به بررسی ۳ ایستگاه آذربایجان غربی جهت پیش‌بینی تبخیر و تعرق مرجع روزانه با مدل‌های ANN، مدل درخت و ترکیبی از این دو با روش پیش‌پردازش مدّ تجربی پرداختند. نتایج حاکی از آن بود که روش هیبریدی پیش‌پردازنده مدّ تجربی با مدل درخت، از بیشترین کارایی برخوردار است (Amirashayeri et al., 2020).

روش‌های مدل‌سازی چندگانه فراگیر IMM² از طریق فرمول‌بندی استراتژی‌هایی برای تقویت همبستگی بین پیش‌بینی‌کننده‌های مقیاس بزرگ، در ۲ سطح اجرا می‌شود. در سطح ۱ تعدادی از مدل‌های در دسترس، با استفاده از پیش‌بینی‌کننده‌ها تولید می‌گردد و در سطح ۲ مدل دیگری ساخته می‌شود که از پیش‌بینی‌های مدل‌های سطح

³ Multiple Criteria Decision Making

⁴ Artificial Intelligence

¹ Non-Stationary

² Inclusive Multiple Modeling

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

وسعت حوضه آبریز دریاچه ارومیه ۵۲۳۵۵ کیلومتر مربع است که از این مقدار حدود ۵۸۲۲ کیلومتر مربع وسعت خود دریاچه مستقیماً با ارتفاع آب دریاچه ارتباط دارد. تراز کف دریاچه از سطح آزاد دریاها ۱۲۶۸ متر، متوسط درجه حرارت سالیانه آن ۱۱ درجه سانتی‌گراد و میانگین سالانه بارندگی در این منطقه در بازه ۱۸۰ تا ۵۰۰ میلی‌متر متغیر است (Goodarzi et al., 2016). رودخانه‌های عمده حوضه، شامل نازلوچای، باراندوزچای، آجی‌چای، زرینه‌رود، سیمینه‌رود، مهابادچای، زولاچای و مجموعه‌ای دیگر از رودخانه‌های بخش غربی دریاچه ارومیه است (Kakahaji et al., 2019). زرینه‌رود در شمال غربی ایران، در جلگه جنوبی دریاچه ارومیه قرار دارد. این رود با طول ۳۰۲ کیلومتر، اندازه حوضه ۱۱۸۵۰ کیلومتر مربع و دبی ۱۸۱۳ متر مکعب بر ثانیه از رودهای پرآب شمال غرب ایران به-شمار می‌رود.

در این تحقیق، دو ایستگاه هیدرومتریک ساری-قمیش و نظام‌آباد که دارای داده‌های غنی جریان رودخانه در منطقه هستند، جهت تخمین و مدل‌سازی دبی روزانه مورد استفاده قرار گرفتند. شکل ۱ موقعیت این دو ایستگاه را نشان می‌دهد. دبی روزانه رودخانه زرینه‌رود در دوره آماری ۱۳۹۵-۱۳۶۹ (۲۶ سال) در محل ایستگاه‌های هیدرومتری ساری-قمیش و نظام‌آباد مورد استفاده قرار گرفت. نواقص آماری موجود در داده‌های مربوطه، با استفاده از روش‌های همبستگی، نظیر میانگین‌گیری تکمیل شد. در جدول ۱ مشخصات آماری سری جریان روزانه بیان شده است.

۲-۲- الگوریتم تجزیه مُد تجربی (EMD)

تجزیه مُد تجربی (EMD) یک روش تجزیه اطلاعات برای استخراج سیگنال‌ها در اطلاعات تولیدشده توسط عملیات غیرخطی و غیرایستا است که شامل جداسازی محلی و کامل برگرفته از اطلاعات یک سیگنال در نوسان‌های سریع و آهسته است. EMD یک سری زمانی را به تعداد محدودی

از حالت‌های نوسانی به نام توابع مُد ذاتی^۱ (IMF) بسط می‌دهد. روش تجزیه مُد تجربی توسط هوانگ و همکاران در سال ۱۹۹۸ ابداع شد. نتایج تجزیه EMD می‌تواند داده‌های موردنیاز برای تکمیل تبدیل هیلبرت را فراهم نماید. در موارد زیادی EMD با تبدیل هیلبرت ترکیب شده و به نام تبدیل هیلبرت-هوانگ (HHT) شناخته می‌شود. پس از مطالعه مفهوم فرکانس آبی در تبدیل هیلبرت، هوانگ مفهوم تازه‌ای از تابع مُد ذاتی (IMF) را که مولفه اساسی هر مجموعه داده پیچیده است، ارائه نمود. در ادامه آنها روش تازه‌ای به نام تجزیه مُد تجربی را ایجاد نمودند تا یک سیگنال را به مجموعه‌ای از توابع مُد ذاتی تجزیه نماید. برخلاف روش‌های تجزیه سنتی، مانند تبدیل فوریه و موجک، این روش جدید، داده-وفقی است. مزیت اصلی EMD نسبت به روش‌های دیگر، امکان تجزیه خودکار به صورت کاملاً داده-وفقی می‌باشد و نیازی به یک مجموعه توابع مفروض، همانند تبدیل موجک ندارد. روش EMD در حوزه‌های مختلف پردازش سیگنال همچون کاربردهای اقتصادی، دینامیک سیالات، مهندسی اقیانوس، و تحلیل سری‌های زمانی الکترومغناطیس استفاده شده است (Huang et al., 1998).

الگوریتم EMD تلاش می‌کند تقریباً هر سیگنالی را به مجموعه‌ای متناهی از توابع، تجزیه نماید و تبدیل هیلبرت نیز مقادیر فرکانس لحظه‌ای فیزیکی را می‌دهد. این توابع به نام توابع مُد ذاتی نامیده می‌شوند. الگوریتم از یک روند غربال تکرارشونده استفاده می‌کند که با موفقیت می‌تواند میانگین محلی را از یک سیگنال کم نماید. روند غربال‌گری در ادامه بیان می‌شود (Cheng et al., 2019):

- ◆ اکثرم محلی سیگنال (ماکزیمم و مینیمم) تعیین می‌شود،
- ◆ مقدار ماکزیمم به یک تابع درون‌یابی متصل شده و پوش بالایی سیگنال را می‌سازد،
- ◆ مقدار مینیمم به یک تابع درون‌یابی متصل شده و پوش پایینی سیگنال را می‌سازد،
- ◆ میانگین محلی به عنوان نصف اختلاف بین پوش بالایی

¹ Intrinsic Mode Function

و پایینی محاسبه می‌شود،

♦ میانگین محلی از سیگنال کم می‌شود و

♦ این روند برای باقیمانده داده‌ها تکرار می‌شود.

جدول ۱ مشخصات آماری سری جریان روزانه رودخانه زرنه رود در ایستگاه‌های ساری-قمیش و نظام‌آباد

Table 1 Statistical characteristics of Zarrineh Rud River daily flow series in Sari-Qomish and Nezam-Abad stations

River	Hydrometric Station	Statistical Course (Year)	Average Flow (m^3/s)	Standard Deviation (m^3/s)	Max Flow (m^3/s)	Min Flow (m^3/s)
Zarrineh Rud River	Sari-Qomish	1369-1395	59.88	65	923	0
Zarrineh Rud River	Nezam-Abad	1369 -1395	116.5	63	994.8	0.011

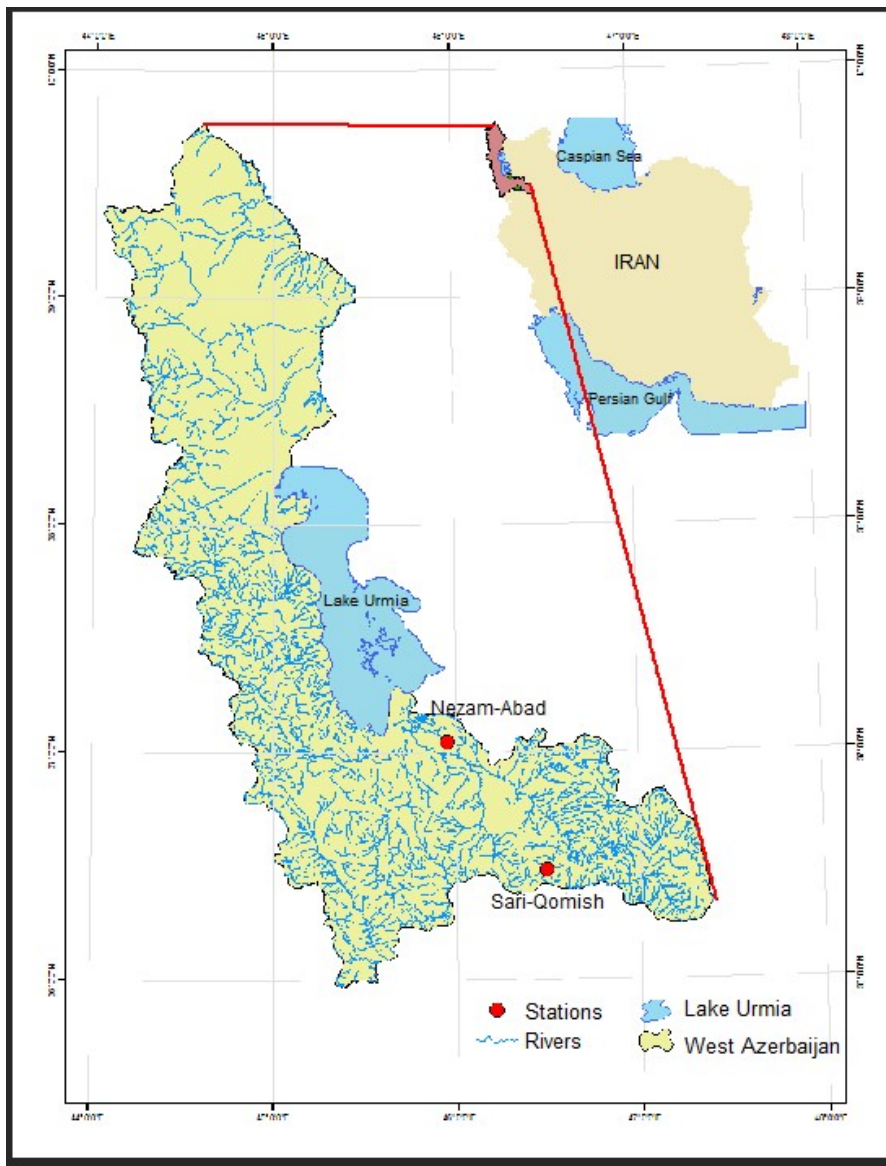


Fig. 1 Geographical location of the studied hydrometric stations

شکل ۱ موقعیت جغرافیایی ایستگاه‌های هیدرومتری مورد مطالعه

در فرکانس است و با استفاده از یک روش غربال‌گری محاسبه می‌شود. برای انجام آن، یک IMF باید دو شرط را ارضا کند (Huang., 1998): تعداد اکسترمم N_e (ماکزیمم و مینیمم محلی) و تعداد عبور از صفرها N_z باید برابر باشند یا حداکثر یکی اختلاف داشته باشند:

$$(N_z - 1) \leq N_e \leq (N_z + 1) \quad (1)$$

در هر نقطه، مقدار میانگین پوشی که به‌وسیله ماکزیمم محلی و پوشی که به‌وسیله مینیمم محلی به‌وجود می‌آید، صفر است. برنامه‌ریزی بیان ژن (GEP) اولین بار توسط فریرا در سال ۱۹۹۹ معرفی گردید. برنامه‌ریزی بیان ژن شبیه به الگوریتم (GA^1) و برنامه‌ریزی ژنتیک (GP^2) است (Ferreira C., 2001). برنامه‌ریزی بیان ژن که در ادامه سیر تکاملی مدل‌های هوشمند به‌وجود آمده است، جزء روش-های الگوریتم گردشی محسوب می‌شود که مبنای تمامی آنها براساس نظریه تکاملی داروین استوار است (Danandehmehr et al., 2010). مزیت برنامه‌ریزی بیان ژن نسبت به مدل‌های دیگر از جمله شبکه عصبی مصنوعی این است که در برنامه‌ریزی بیان ژن، ابتدا ساختار (متغیرهای ورودی، هدف و مجموع توابع) تعریف و سپس ساختار بهینه مدل و ضرایب، طی فرآیند آموزش تعیین می‌شوند، درحالی‌که در شبکه‌های عصبی، ابتدا باید ساختار تعیین شود و فقط ضرایب مدل طی فرآیند آموزش، به‌دست می‌آیند. همچنین این الگوریتم به‌طور خودکار می‌تواند متغیرهای ورودی که در مدل، بیشترین تاثیر را دارند، انتخاب کند.

فرآیند گام‌به‌گام حل یک مساله با استفاده از برنامه‌ریزی بیان ژن، متشکل از ۵ مرحله است (Ferreira, C., 2001):

۱. انتخاب مجموعه پایانه،
۲. انتخاب مجموعه توابع،
۳. شاخص‌های اندازه‌گیری دقت مدل،
۴. مولفه‌های کنترل،
۵. معیارهای توقف برنامه.

در شکل ۲ مثالی از یک تابع مُد ذاتی (IMF) با پوش‌های مربوطه آورده شده است. روند غربال‌گری تا زمانی که سیگنال به تعریف یک IMF برسد، تکرار می‌شود. سپس، IMF از سیگنال اصلی، کسر شده و روند غربال‌گری بر باقیمانده تکرار می‌شود. این کار تا زمانی که باقیمانده، یک تابع یکنواخت شود، ادامه می‌یابد. آخرین IMF استخراج‌شده، کم‌ترین مولفه فرکانسی است. در شکل ۲ منحنی سبز، یک تابع مُد ذاتی (IMF) را به همراه پوش‌های مربوطه نشان می‌دهد. با اعمال EMD، یک سیگنال می‌تواند به مجموعه‌ای از توابع تک‌مولفه با نام توابع مُد ذاتی (IMFs) تجزیه شود (Al-Musaylh et al., 2018).

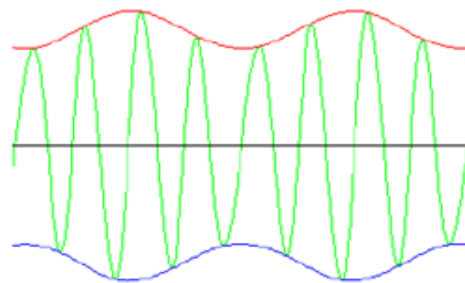


Fig. 2 An example of an intrinsic mode function (IMF), (green curve) with corresponding caps
شکل ۲ مثالی از یک تابع مُد ذاتی (IMF)، (منحنی سبز) با پوش‌های مربوطه

ایده اصلی در پس‌زمینه IMF، تفکیک اطلاعات به بخش میانگین محلی با تغییرات آرام و بخش نوسانی متقارن با تغییرات سریع می‌باشد که بخش آخر همان IMF است و میانگین محلی با عنوان باقیمانده تعریف می‌شود. این باقیمانده به‌عنوان ورودی برای تجزیه بیشتر به‌کار می‌رود. این روند تا آنجا که هیچ نوسان دیگری نباشد، تکرار می‌شود. به‌عنوان یک حالت نوسانی، دامنه و فرکانس یک IMF با زمان به‌نحوی تغییر خواهد کرد که به‌طور محلی متقارن و تک مولفه باشد. بنابراین، فرکانس لحظه‌ای می‌تواند به‌طریقی معنادار محاسبه شود (Zhou et al., 2019). از این‌رو، IMFها فرکانس‌هایی در محدوده بین بالاترین و پایین‌ترین سیگنال را شامل می‌شوند که به‌عنوان سیگنال‌های مدوله‌شده دامنه و فرکانس (AM-FM) موجود هستند؛ AM پوش را حمل می‌کند و FM تغییر دامنه ثابت

² Genetic Programming

¹ Genetic Algorithm

فعال‌سازی) را به‌وسیله یک تابع فعال‌سازی بیرون می‌دهد. تانژانت هیپربولیک (رابطه (۲))، تابع لجستیک (رابطه (۳)) و توابع فعال‌سازی (رابطه (۴))، روابطی هستند که به‌طور مشترک به‌کار گرفته می‌شوند. خروجی (y_j) ، یک نورون z در لایه نامین لایه، براساس معادلات زیر محاسبه شده است. اطلاعات بیشتر در منبع (Kumar et al., 2011) یافت می‌شود.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

$$y_j = f(l_j) \quad (3)$$

$$l_j = \sum_{i=1}^n w_{ij}x_i + b_i \quad (4)$$

که l_j مقدار فعال‌سازی نورون z از لایه نامین w_{ij} وزن نامین ورودی و نورون z در لایه، x_i مقدار نامین ورودی، b_i نامین جمله بایاس، y_j خروجی نورون z و $f(x)$ تابع فعال‌سازی هستند.

در این تحقیق، تعداد لایه‌های پنهان و تعداد نورون‌ها در لایه پنهان، بر مبنای مقایسه عملکرد شبکه‌ها انتخاب شد. برای تشکیل شبکه، از تابع آموزش Levenberg-Marquardt و تابع‌های انتقال purelin و logsigmoid استفاده گردید. بهترین معیار برای مقایسه عملکرد شبکه‌های ایجاد شده برای مرحله ارزیابی آموزش، ریشه میانگین مربعات خطا است (Kumar et al., 2002).

۲-۴- معیار ارزیابی عملکرد

برای مقایسه عملکرد مدل‌های ارائه‌شده جهت تخمین جریان روزانه در ایستگاه‌های موردنظر در مرحله آزمون، شاخص‌های آماری شامل ضریب تبیین (R^2)، ریشه میانگین مربعات خطا ($RMSE$) و ریشه مربعات خطا (MAE) به‌صورت زیر محاسبه می‌گردند:

$$R^2 = \left[\frac{\sum_{i=1}^M (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^M (O_i - \bar{O})^2 \sum_{i=1}^M (P_i - \bar{P})^2}} \right]^2 \quad (5)$$

برنامه‌ریزی بیان ژن همانند الگوریتم ژنتیک و برنامه‌ریزی ژنتیک است که از جمعیتی از افراد استفاده کرده، آن‌ها را براساس برازندگی انتخاب و تغییرات ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می‌کند. تفاوت اساسی میان سه الگوریتم یادشده مربوط به ماهیت افراد آنهاست. در الگوریتم ژنتیک، افراد به‌صورت رشته‌های خطی با طول ثابت (کروموزوم‌ها) هستند، درحالی‌که در برنامه‌ریزی ژنتیک، افراد به‌صورت نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (درختان تجزیه) می‌باشند. این در حالی است که در GEP، افراد در ابتدا به‌صورت رشته‌های خطی با طول ثابت کدگذاری می‌شوند (مشابه با الگوریتم ژنتیک) و سپس، به شکل نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (مشابه با برنامه‌ریزی ژنتیک) بیان می‌شوند. تولید جمعیت اولیه از راه‌حل‌ها، اولین مرحله در روش GEP است. در این تحقیق، از نرم‌افزار GeneXproTools 0.4 به‌منظور اعمال روش GEP بر روی داده‌ها استفاده گردید.

۲-۳- شبکه عصبی مصنوعی

یک شبکه عصبی مصنوعی، سیستم پردازش اطلاعات است که توانایی مغز انسان را جهت یافتن الگوها و یادگیری از آزمون و خطا شبیه‌سازی می‌کند و عموماً شامل مجموعه عناصر پردازشی به نام سلول‌های عصبی یا نورون‌ها است که به‌وسیله سیناپس‌های وزندار به یکدیگر متصل هستند (Patil et al., 2016).

دستگاه‌های عصبی، ترکیبی از ۳ لایه هستند (Landeras et al., 2008):

۱. لایه‌های ورودی که داده‌ها به شبکه عصبی مصنوعی معرفی می‌شوند،
۲. لایه‌های مخفی که داده‌ها در آن پردازش می‌شوند و
۳. یک لایه خروجی که برابر لایه ورودی است و نتایج در آن تولید و خارج می‌شوند.

در یک ANN هر نورون، ورودی‌های وزنی را دریافت می‌کند که می‌تواند دربرگیرنده متغیرهای ورودی یا خروجی‌های نورون‌های لایه‌های قبلی باشد. سپس ورودی‌ها را با یکدیگر جمع کرده و یک جمله بایاس به آن افزوده و نتیجه (مقدار

میانگین‌گیری تکمیل شد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^M (P_i - O_i)^2}{M}} \quad (6)$$

$$MAE = \frac{|P_i - O_i|}{M} \quad (7)$$

۲-۶- توسعه مدل‌ها برای پیش‌بینی جریان رودخانه

ابزارهای مهم در تشخیص مدل، تابع‌های خودهمبستگی و خودهمبستگی جزئی هستند. رسم این نمودارها در تعیین نوع و مرتبه فرآیند، مفید است که در میانگین و واریانس داده‌ها، مورد ارزیابی قرار می‌گیرند. تابع خودهمبستگی (ACF^1) یکی ابزارهای بسیار مهم برای امتحان وابستگی داده‌ها است. این تابع، همبستگی میان مشاهدات را در فواصل مختلف، اندازه می‌گیرد و جهت بررسی یک سری زمانی واحد در قلمرو زمان به کار می‌رود. این تابع اغلب اطلاعاتی از الگوی احتمالی که داده‌ها را تولید می‌کند، به ما ارائه می‌دهد که از آن به منظور تشخیص و برازش مدل استوکاستیکی مناسب برای داده‌ها استفاده می‌شود. رفتار تابع خودهمبستگی جزئی ($PACF^2$) در نمودار همبستگی، یکی از مهم‌ترین معیارها برای تخمین الگوی سری زمانی است. در صورت ناپیوستگی، ابتدا سری موردنظر با استفاده از سری تفاضلی مناسب و تبدیل داده‌ها از روش باکس-کاکس در میانگین و واریانس پایدار، سپس سری ایستا می‌گردد. بنابراین، در این مرحله با آنالیز واریانس داده‌های تفاضل‌شده و همچنین با استفاده از نمودار $PACF$ ، تعداد تاخیرهای وارد شده به داده‌ها جهت پیش‌بینی با استفاده از روش برنامه‌ریزی بیان ژن و جهت آماده شدن داده‌ها برای ارزیابی، مورد استفاده قرار می‌گیرند. برای انتخاب داده‌های ورودی به مدل‌ها از ۳ سناریو به‌عنوان گام‌های تاخیر برای ورودی مدل استفاده شد که در جدول ۲ نشان داده شده‌اند.

با استفاده از معیارهای انتخاب‌شده، سناریو نوع ۳ خطای کمتر و دقت بیشتری را به‌عنوان داده‌های ورودی به مدل، به ما ارائه داد. که در این سناریوها $q(x)$ مقدار جریان روزانه و $q(x-1)$ تاخیر زمانی گام اول به‌عنوان ورودی به مدل، $q(x-2)$ تاخیر زمانی گام دوم به‌عنوان ورودی به مدل و $q(x-3)$ تاخیر زمانی گام سوم به‌عنوان ورودی به مدل هستند. برای ارزیابی بیشتر هر یک از روش‌های مستقل و ترکیبی پیشنهادی، ایستگاه نظام‌آباد در نظر گرفته شد.

که O_i مقدارهای مشاهده‌ای، P_i مقدارهای پیش‌بینی، \bar{O} میانگین داده‌های مشاهده‌ای، \bar{P} میانگین داده‌های پیش‌بینی و M تعداد داده‌ها هستند. R^2 نشانگر ضریب تبیین بوده که هر چه این مقدار به ۱ نزدیک‌تر باشد، مدل صحیح‌تر و منطبق‌تر است. $RMSE$ برای اندازه‌گیری دقت پیش‌بینی مدل استفاده می‌شود که با جذر گرفتن از صفر یک مقدار مثبت را نشان می‌دهد. محدوده $RMSE$ از صفر تا یک عدد مثبت بزرگ ادامه می‌یابد که واحد آن از جنس متغیر خروجی مشاهداتی است. مقادیر بالای R^2 و مقادیر کم $RMSE$ نشان‌دهنده عملکرد بالای مدل است. همین‌طور، MAE مقدار میانگین مثبت خطاها در یک مجموعه از مقادیر پیش‌بینی‌شده را محاسبه می‌کند و واحد آن از جنس پارامتر خروجی مشاهداتی است. مقدار این شاخص آماری معمولاً مشابه مقدار $RMSE$ می‌باشد.

۲-۵- داده‌های ورودی به مدل‌ها

انتخاب متغیرهای ورودی مناسب برای مدل‌های هوشمند مصنوعی و روش‌های داده‌کاوی، در عملکرد هر یک از مدل‌ها بسیار اثرگذار است. از این‌رو، براساس مشاهدات تجربی و مطالعات سایر محققین، می‌توان از میان متغیرهای ورودی موثر در پیش‌بینی جریان رودخانه، از گام‌های تاخیر برای ورودی مدل استفاده نمود که اثرگذاری قابل‌توجهی بر خروجی مدل‌ها دارند. در بسیاری از موارد داده‌های ثبت‌شده برای دبی جریان، توزیعی نامتقارن دارند و در برخی موارد (مانند وجود روند، خاصیت تناوبی و ...) ناپیوستگی می‌باشند. بنابراین، الزام است که داده‌ها قبل از مدل‌سازی، نرمال شده و از ایستایی بودن آنها اطمینان حاصل شود. داده‌های مربوط به جریان رودخانه در ایستگاه‌های ساری-قمیش و نظام‌آباد در دوره آماری ۱۳۹۵-۱۳۶۹ جمع‌آوری گشت. نکته قابل‌توجه این است که نواقص آماری موجود در داده‌های مربوط، با استفاده از روش‌های همبستگی، نظیر

² Partial Autocorrelation Function

¹ Autocorrelation Function

افزایش و شاخص ریشه میانگین مربعات خطا نیز به میزان $(m^3/s) 0.002/0$ کاهش یابد. ترکیب روش پیش‌پردازش-کننده EMD با روش ANN نیز منجر به افزایش ضریب تبیین به میزان ۳ درصد و کاهش شاخص ریشه میانگین مربعات خطا به میزان $(m^3/s) 0.01/0$ می‌گردد که حاکی از بهبود دقت این دو مدل با روش پیش‌پردازنده EMD است. به‌کارگیری این روش افزایش چشمگیر دقت در شاخص‌های ریشه میانگین مربعات خطا و خطای مطلق در مدل شبکه هوشمند مصنوعی را در دربی داشت. در واقع نتایج حاکی از سازگاری این روش پیش‌پردازنده با هر دو مدل مذکور است.

جدول ۳ نتایج ارزیابی عملکرد مدل‌های پیشنهادی برای ایستگاه ساری‌قمیش

Table 3 Performance of proposed models at validation phases for Sari-Qomish station

Models	R^2	RMSE (m^3/s)	MAE (m^3/s)
GEP	0.89	0.016	0.003
EMD-GEP	0.96	0.002	0.0004
ANN	0.95	24.22	25.75
EMD-ANN	0.98	0.01	0.45

شکل‌های ۴ و ۵، نمودارهای پراکندگی مقادیر دبی رودخانه مشاهداتی و پیش‌بینی نظیر آن توسط مدل پیشنهادی GEP و EMD-FEP را در مرحله آزمون نشان می‌دهد. همان‌طور که در شکل ۵ مشاهده می‌شود، داده‌ها در اطراف خط ۴۵ درجه متمرکز شده که نشان‌دهنده همبستگی مقادیر واقعی و محاسباتی مدل می‌باشد. در شکل ۴ این متمرکز شدن به نسبت، کمتر است. یکی از ویژگی‌های اصلی نمودار پراکندگی، بررسی پدیده بیش‌برآورد و کم-برآورد^۱ است که به معنی میزان تخمین رو به بالا یا رو به پایین داده‌های دبی است. همان‌گونه که در این شکل مشخص است، نمودار مربوط به EMD-GEP همبستگی بیشتری نسبت به نمودار مربوط به GEP دارد که بیان‌گر عملکرد مطلوب روش EMD در افزایش دقت مدل است.

همچنین نتایج بیان‌گر عملکرد بهتر روش‌های شبکه عصبی مصنوعی و بیان ژن پس از ترکیب با EMD است.

جدول ۲ سناریوهای فرض‌شده جهت انتخاب بهترین سناریو

Table 2 Assumed scenarios to select the best scenario

(1)	$q(x) = q(x - 1)$
(2)	$q(x) = q(x - 1), q(x - 2)$
(3)	$q(x) = q(x - 1), q(x - 2), q(x - 3)$

۳- نتایج و بحث

مقدار جریان رودخانه در مقیاس روزانه برای هر یک از دو ایستگاه ساری-قمیش و نظام‌آباد در شکل ۳ نشان داده شده است. مشاهده می‌شود که مقدار جریان رودخانه برای هر دو ایستگاه با توجه به موقعیت مکانی جغرافیایی، دارای محدوده تقریباً یکسانی است.

همان‌گونه که در جدول ۳ برای ایستگاه ساری-قمیش نشان داده شده است. شاخص آماری (R^2)، در مرحله آزمون برای مدل بیان ژن و شبکه عصبی مصنوعی به ترتیب ۰/۸۹ و ۰/۹۵ به دست آمد که تفاوتی قابل توجه در پیش‌بینی جریان رودخانه داشته‌اند و بیانگر دقت بیشتر مدل شبکه عصبی مصنوعی است. در مقابل، مقدار شاخص آماری RMSE به ترتیب $(m^3/s) 0.016/0$ و $(m^3/s) 24.22/0$ به دست آمده است که مقدار بالای RMSE بیانگر دقت نه‌چندان خوب این معیار است. بدین ترتیب، شاخص‌های آماری، نشان می‌دهد هر دو مدل، با آموزش مطلوبی مواجه نبوده‌اند. از طرف دیگر، MAE به ترتیب برای مدل‌های مذکور $(m^3/s) 0.003/0$ و $(m^3/s) 25.75/0$ تعیین گردید که بیانگر دقت نه‌چندان خوب مدل شبکه عصبی مصنوعی برای ارزیابی دبی جریان است (شکل ۳). در حقیقت برای بالا بردن دقت پیش‌بینی در این ایستگاه، ترکیب کردن روش پیش‌پردازنده EMD با روش GEP باعث می‌شود که ضریب تبیین به میزان ۷/۲۹ درصد

¹ Over Estimation and Under Estimation

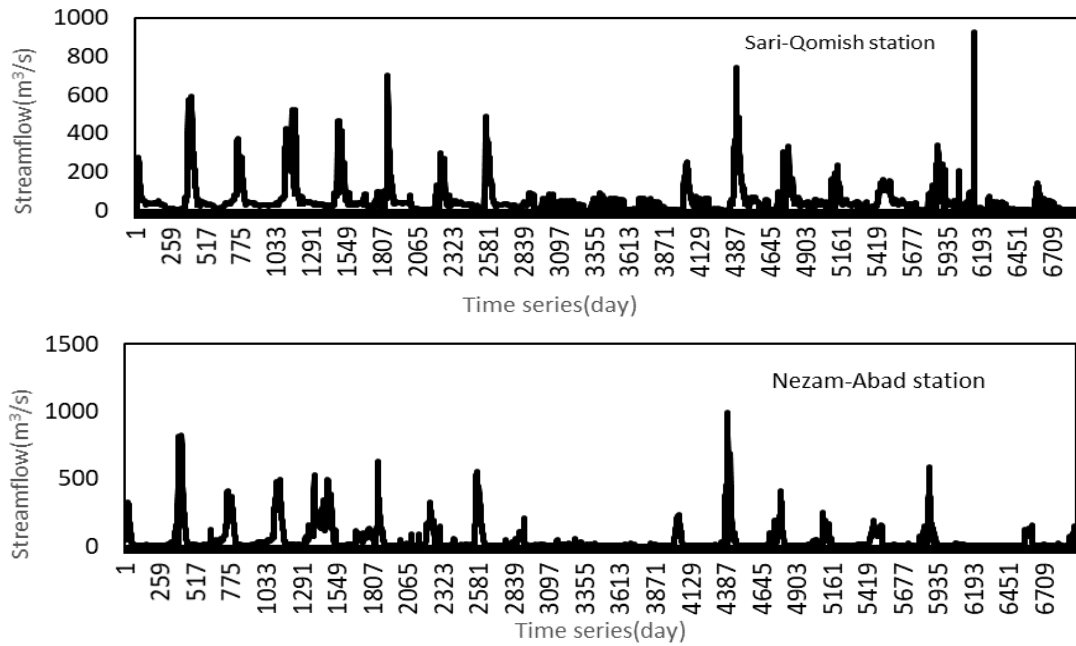


Fig. 3 Daily flow time series plot in Sari-Qomish and Nezam-Abad hydrometric stations
شکل ۳ سری زمانی دبی روزانه در ایستگاه‌های هیدرومتریک ساری‌قمیش و نظام‌آباد

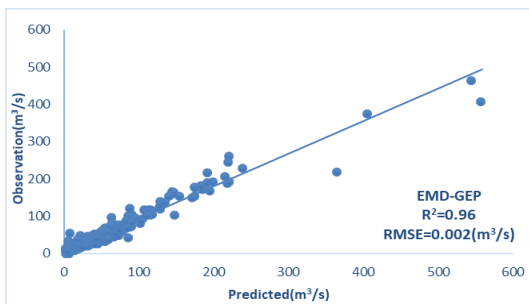


Fig. 5 Scatter plots of the proposed EMD-GEP model in the validation phase for Sari-Qomish station
شکل ۵ نمودار پراکندگی مدل پیشنهادی EMD-GEP در مرحله آزمون برای ایستگاه ساری‌قمیش

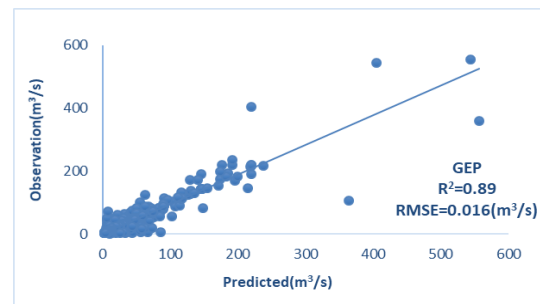


Fig. 4 Scatter plots of the proposed GEP model in the validation phase for Sari-Qomish station
شکل ۴ نمودار پراکندگی مدل پیشنهادی GEP در مرحله آزمون برای ایستگاه ساری‌قمیش

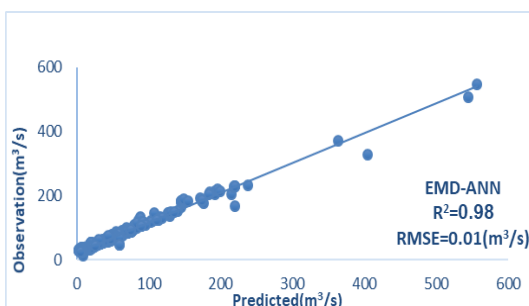


Fig. 7 Scatter plots of the proposed EMD-ANN model in the validation phase for Sari-Qomish station
شکل ۷ نمودار پراکندگی مدل پیشنهادی EMD-ANN در مرحله آزمون برای ایستگاه ساری‌قمیش

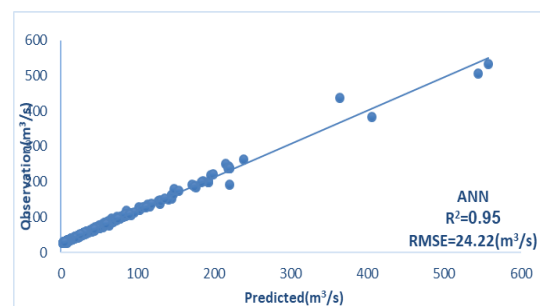


Fig. 6 Scatter plots of the proposed ANN model in the validation phase for Sari-Qomish station
شکل ۶ نمودار پراکندگی مدل پیشنهادی ANN در مرحله آزمون برای ایستگاه ساری‌قمیش

مرحله آزمون توانست دقت مدل بیان ژن را از مقدار ۰/۹۱ به ۰/۹۳ افزایش دهد؛ همچنین مقدار ریشه میانگین مربعات خطا را از $0.28 \text{ (m}^3/\text{s)}$ به $0.05 \text{ (m}^3/\text{s)}$ کاهش داد که در جدول ۴ نشان داده شده است.

روش‌های منفرد مذکور مانند ایستگاه قبلی عملکرد نسبتاً ضعیف‌تری نسبت به روش‌های هوشمند مصنوعی ترکیب‌شده داشتند. از این رو می‌تواند به‌عنوان ابزاری معتبر در داده‌کاوی و پژوهش‌های هوش مصنوعی در زمینه مهندسی آب مورد استفاده قرار گیرند. همچنین نتایج بیان-گر عملکرد بهتر روش EMD-ANN نسبت به روش منفرد ANN است. این روش، در مرحله آزمون توانست دقت مدل را از ۰/۹۴ به ۰/۹۷ افزایش دهد. همچنین مانند ایستگاه قبلی باعث کاهش مطلوب شاخص‌های ریشه میانگین مربعات خطا و خطای مطلق شد به طوری که مقدار $RMSE$ را از مقدار $3.92 \text{ (m}^3/\text{s)}$ به $0.03 \text{ (m}^3/\text{s)}$ کاهش داد و باعث بهبود دقت پیش‌بینی در این ایستگاه شد.

جدول ۴ نتایج ارزیابی عملکرد مدل‌های پیشنهادی برای ایستگاه نظام‌آباد

Table 4 Performance of proposed models at validation phases for Nezam-Abad station

Models	R^2	$RMSE$ (m^3/s)	MAE (m^3/s)
GEP	0.91	0.028	0.006
EMD-GEP	0.93	0.005	0.0001
ANN	0.94	3.92	11.79
EMD-ANN	0.97	0.003	0.14

شکل‌های ۶ و ۷ نیز نمودارهای پراکندگی مقادیر دبی رودخانه مشاهداتی و پیش‌بینی نظیر آن توسط مدل پیشنهادی ANN و EMD-ANN را در مرحله آزمون نشان می‌دهند. همان‌طور که در شکل ۷ مشاهده می‌گردد، داده‌ها در اطراف خط ۴۵ درجه متمرکز شده‌اند که نشان از همبستگی مقادیر واقعی و محاسباتی مدل دارد. در شکل ۶ این متمرکز شدن به نسبت کمتر است و نمودار مربوط به EMD-ANN همبستگی بیشتری نسبت به نمودار مربوط به ANN دارد. همچنین مدل شبکه عصبی مصنوعی نسبت به مدل بیان ژن، برای ایستگاه ساری-قمیش، همبستگی بیشتری برای ترکیب شدن با روش پیش‌پردازنده EMD از خود نشان داد (شکل‌های ۴ تا ۷). نمودار سری زمانی جریان روزانه مشاهداتی و جریان پیش‌بینی‌شده با الگوی برتر مدل ترکیب‌شده، برای دو مدل ANN و GEP در مرحله آزمون در شکل ۸ ارائه شده است. مشاهده می‌شود، در بیشتر موارد مقادیر پیش‌بینی‌شده ترکیب‌شده با دو مدل ANN و GEP با مقادیر مشاهداتی، روند یکسانی دارند که نشان از دقت بیشتر مدل‌های ترکیب‌شده نسبت به مدل‌های منفرد در پیش‌بینی روزانه مقادیر جریان ایستگاه ساری-قمیش دارد.

برای ارزیابی بیشتر هر یک از روش‌های مستقل و ترکیبی پیشنهادی، ایستگاه نظام‌آباد نیز در نظر گرفته شد که نتایج نشان‌دهنده عملکرد بهتر روش ترکیب EMD با روش‌های شبکه عصبی مصنوعی و بیان ژن است. الگوریتم EMD در

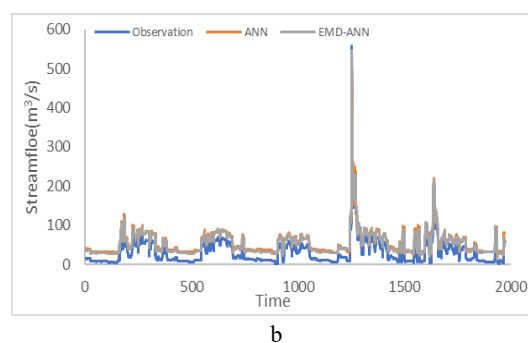
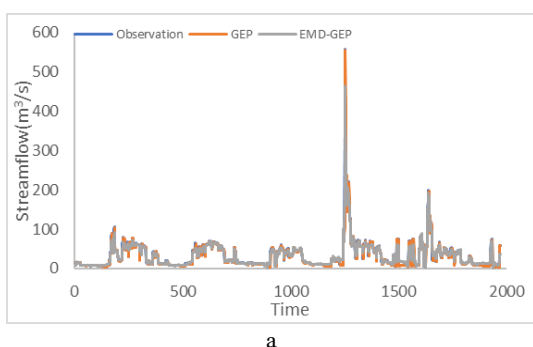


Fig. 8 Time series plot of the validation phase of the observed and estimated values at Sari-Qomish station. **a)** GEP and EMD-GEP models, **b)** ANN and EMD-ANN models.

شکل ۸ الف) مدل‌های GEP و EMD-GEP و ب) مدل‌های ANN و EMD-ANN.

بالای مدل ترکیب‌شده به نسبت مدل منفرد است.

نمودار سری زمانی جریان روزانه مشاهداتی و جریان پیش‌بینی‌شده با الگوی برتر مدل ترکیب‌شده برای دو مدل ANN و GEP در مرحله آزمون در شکل ۱۳ ارائه شده است. بر این اساس مشاهده می‌شود که در بیشتر موارد مقادیر پیش‌بینی‌شده ترکیب‌شده با دو مدل ANN و GEP با مقادیر مشاهداتی، روندی یکسانی دارند که نشان از دقت قابل قبول مدل‌های ترکیب‌شده نسبت به مدل‌های منفرد در پیش‌بینی روزانه مقادیر جریان ایستگاه نظام‌آباد دارد (شکل‌های ۹ تا ۱۳).

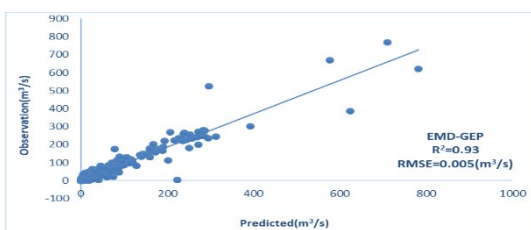


Fig. 10 Scatter plots of the proposed EMD-GEP model in the validation phase for the Nezam-Abad station
شکل ۱۰ نمودار پراکندگی مدل پیشنهادی EMD-GEP در مرحله آزمون برای ایستگاه نظام‌آباد

نمودار پراکندگی مقادیر پیش‌بینی مدل بیان ژن و ترکیب مدل برای ایستگاه نظام‌آباد در مرحله آزمون نیز در شکل-های ۹ و ۱۰ نشان داده شده است. نتایج بر همبستگی بهتر و تمرکز بیشتر داده‌ها حول خط ۴۵ درجه در مدل ترکیب‌شده EMD-GEP دلالت دارد. همانند ایستگاه قبلی روش ترکیب‌شده باعث همبستگی بیشتر داده‌ها شد. شکل-های ۱۱ و ۱۲ نیز نمودارهای پراکندگی مقادیر دبی رودخانه مشاهداتی و پیش‌بینی نظیر آن توسط مدل پیشنهادی ANN و EMD-ANN را در مرحله آزمون نشان می‌دهد که در روش ترکیب‌شده به نسبت روش منفرد، همبستگی داده‌ها روی خط ۴۵ درجه بیشتر است که حاکی از دقت

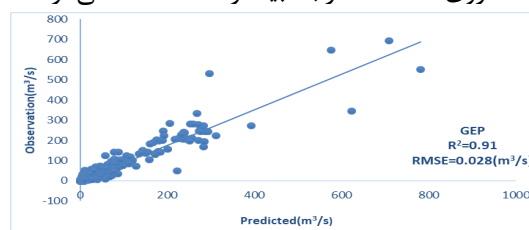


Fig. 9 Scatter plots of the proposed GEP model in the validation phase for the Nezam-Abad station
شکل ۹ نمودار پراکندگی مدل پیشنهادی GEP در مرحله آزمون برای ایستگاه نظام‌آباد

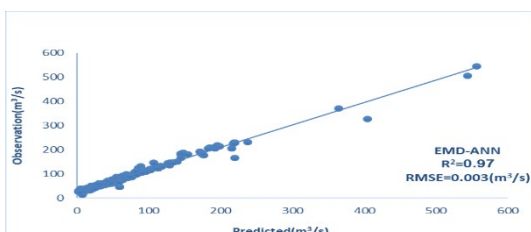


Fig. 12 Scatter plots of the proposed EMD-ANN model in the validation phase for the Nezam-Abad station
شکل ۱۲ نمودار پراکندگی مدل پیشنهادی EMD-ANN در مرحله آزمون برای ایستگاه نظام‌آباد

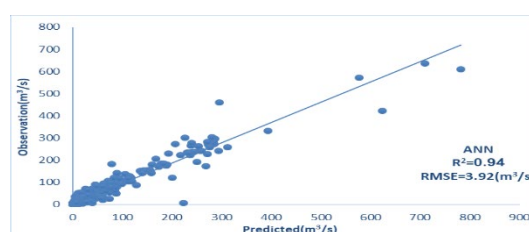


Fig. 11 Scatter plots of the proposed ANN model in the validation phase for the Nezam-Abad station
شکل ۱۱ نمودار پراکندگی مدل پیشنهادی ANN در مرحله آزمون برای ایستگاه نظام‌آباد

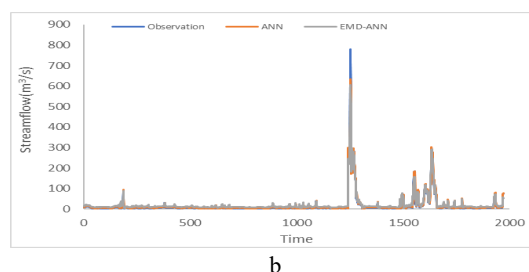
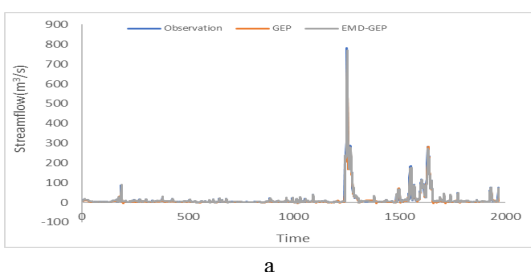


Fig. 13 Time series plot of the validation phase of the observed and estimated values at Nezam-Abad station. **a)** GEP and EMD-GEP models, **b)** ANN and EMD-ANN models.

شکل ۱۳ نمودار سری زمانی مرحله آزمون مقادیر مشاهداتی و پیش‌بینی‌شده در ایستگاه نظام‌آباد. الف) مدل GEP و EMD-GEP و ب) مدل ANN و EMD-ANN

به‌عنوان ورودی مدل‌های پیش‌بینی مدل شبکه عصبی مصنوعی (ANN) و برنامه‌ریزی بیان ژن (GEP) انتخاب و از روش پیش‌پردازنده الگوریتم تجزیه مُد تجربی (EMD) جهت افزایش دقت دو مدل، استفاده شد. نتایج نشان داد که با پیش‌پردازش کردن داده‌ها با استفاده از الگوریتم تجزیه مُد تجربی، دقت مدل برنامه‌ریزی بیان ژن را در ضریب تبیین، به‌ترتیب به میزان ۷/۲۹٪ و ۲/۱۵٪ برای ایستگاه‌های ساری-قمیش و نظام‌آباد در مرحله آزمون افزایش داد. همچنین این روش ترکیبی پیشنهادی به‌عنوان مثال در ایستگاه ساری-قمیش خطای به‌مراتب کمتری نسبت به ایستگاه نظام‌آباد در پیش‌بینی دبی روزانه داشت. علاوه بر این، مدل ترکیبی برای مدل شبکه عصبی مصنوعی باعث افزایش بیشتر دقت ترم شاخص‌های ریشه مربعات خطا و خطای مطلق نسبت به روش برنامه‌ریزی بیان ژن شد.

۵- فهرست نشانه‌ها

b_i	نامین جمله بایاس
$f(x)$	تابع فعال‌سازی
I_k	مقدار فعال‌سازی نورون
M	تعداد داده‌ها
N_z	تعداد عبور از صفرها
N_e	تعداد اکسترمم
O_i	مقدار مشاهده‌ای
\bar{O}	میانگین داده‌های مشاهده‌ای
P_i	مقدار پیش‌بینی
\bar{P}	میانگین داده‌های پیش‌بینی
W_{ij}	وزن نامین ورودی و نورون z در لایه
X_i	مقدار نامین ورودی
y_j	خروجی نورون z

۳-۱- آزمون t-test برای آزمون همبستگی باقیمانده

برای مدل ترکیبی GEP-EMD، حاصل داده‌های واقعی و مدل به‌منظور داده‌های باقیمانده برای ایستگاه ساری-قمیش به‌دست آمد. برای انجام آزمون، ۲ فرض H_0 و H_1 به‌ترتیب برای فرض صفر و فرض مقابل در نظر گرفته شد. سپس با استفاده از آزمون t-test و نرم‌افزار SPSS آزمون همگنی انجام گرفت. با توجه به ۲ فرض در نظر گرفته‌شده و مقدار ضریب سطح بامعنایی^۱ ۰/۰۰۱ حاصل از نرم‌افزار SPSS و با توجه به $\text{sig} < ۰/۰۵$ ، فرض صفر رد خواهد شد و فرض مقابل که همان همبستگی داده‌ها است، اثبات شد. نتایج نشان داد که برای ایستگاه ساری-قمیش همبستگی واریانس در داده‌های موردبررسی، وجود دارد.

۴- نتیجه‌گیری

روندپایی جریان در رودخانه‌ها یکی از موضوعات حیاتی در مدیریت منابع آب‌های سطحی جهت اتخاذ تدابیر مناسب در مواقع سیلاب و بروز خشک‌سالی‌ها است. از آنجاکه پیش‌بینی جریان رودخانه نقش مهمی در برنامه‌ریزی و مدیریت منابع آب دارد، می‌توان با پیش‌بینی جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، حوادث طبیعی نظیر سیل و خشک‌سالی را نیز پیش‌بینی و مهار نمود.

از این رو برآورد صحیح و دقیق جریان رودخانه با استفاده از مدل‌های مختلف هوشمند مصنوعی یکی از موضوعاتی است که در منابع آب، موردتوجه پژوهشگران است. هدف از این تحقیق، افزایش دقت پیش‌بینی جریان رودخانه زرینه-رود از دو ایستگاه ساری-قمیش و نظام‌آباد با استفاده از ترکیب کردن روش پیش‌پردازش‌کننده تجزیه مُد تجربی است. بدین منظور، داده‌های جریان روزانه دو ایستگاه در یک دوره ۲۶ ساله (۱۳۶۹ تا ۱۳۹۵) و تا سه تاخیر زمانی

References

منبع‌ها

Factors (MLR-PCA) in Predicting Reference Evapotranspiration (Case Study: Kerman Station). J. Water and Soil. 24(6). (In Persian).

Seifi, A. Mirltefi, S.M. and Riyahi, H. (2010). Development of a Combined Model of Multiple Regression Analysis of Major Components and

¹ Significant Level

- Ghaemi, A., Rezaie-Balf, M., Adamowski, J., Kisi, O., & Quilty, J. (2019). On the applicability of maximum overlap, the discrete wavelet transform is integrated with the MARS and M5 model tree for monthly pan evaporation prediction: *Agricultural and Forest Meteorology*, 278, 107647.
- Huang, Y., Yang, L., Liu, S., & Wang, G. (2019). Multi-Step Wind Speed Forecasting Based On Ensemble Empirical Mode Decomposition, Long Short Term Memory Network and Error Correction Strategy. *Energies*, 12(10), 1822.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903-995.
- Kakahaji, H., Banadaki, H. D., Kakahaji, A., & Kakahaji, A. (2013). Prediction of Urmia Lake water-level fluctuations by using analytical, linear statistics, and intelligent methods. *Water resources management*, 27(13), 4469-4492.
- Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017a). Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of Hydrology*, 553, 356-373.
- Rezaie-Balf, M., Kisi, O., & Chua, L. H. (2019). Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. *Hydrology Research*, 50(2), 498-516.
- Sadeghfam, S.; Khatibi, R.; Daneshfaraz, R.; Rashidi, H.B. Transforming vulnerability indexing for saltwater intrusion into risk indexing through a fuzzy catastrophe scheme. *Water Resour. Manag.* 2020, 34, 175–194.
- Taheri, M., Emadzadeh, M., Gholizadeh, M., Tajrishi, M., Ahmadi, M., & Moradi, M. (2019). Investigating the temporal and spatial variations of water consumption in Urmia Lake River Basin considering the climate and anthropogenic effects on the agriculture in the basin. *Agricultural water management*, 213, 782-791.
- Zhou, L., Meng, Y., & Abbaspour, K. C. (2019). A new framework for multi-site stochastic rainfall generator based on empirical orthogonal function analysis and Hilbert-Huang transform. *Journal of*
- Goodarzi, M. Sallahi, B. and Hosieni, S.A (2016). Investigation of Climate Change on Surface Runoff Changes (Case Study: Urmia Lake Watershed). *J. Echo Hydrology*. 2 (2). (In Persian).
- Dastorani, M.T. Sharifi darani, H. Talebi, H. and Moghannia, H. (2012). The efficiency of artificial neural networks and adaptive neural-fuzzy inference system in rainfall-runoff modeling in Zayandehrud watershed. *J. Water and Wastewater*. 114-125:(80)22. (In Persian).
- Amirashayeri, A. Behmanesh, J. Verdinejad, V.R. and Fathollahzadeh Attar, N. (2020). Prediction of Daily Reference Evapotranspiration Using Hybrid Artificial Intelligence Method Based on Empirical Mode Decomposition. *Iranian Journal of Irrigation and Drainage*. 14(4), 1376-1390. (In Persian).
- Al-Musaylh, M. S., Deo, R. C., Li, Y., & Adamowski, J. F. (2018). Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. *Applied Energy*, 217, 422-439.
- Ali, M., & Prasad, R. (2019). Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*, 104, 281-295.
- Cheng, Y., Wang, Z., Chen, B., Zhang, W., & Huang, G. (2019). An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis. *ISA transactions*.
- Deo, R. C., Downs, N., Parisi, A. V., Adamowski, J. F., & Quilty, J. M. (2017). Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environmental Research*, 155, 141-166.
- Fijani, E., Barzegar, R., Deo, R., Tziritis, E., & Konstantinos, S. (2019). Design and implementation of a hybrid model based on a two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Science of the total environment*, 648, 839-853.

- Guven, A. 2009. Linear genetic programming for time-series modeling of daily flow rate. *Journal of Earth System Science*, 118(2): 157-173.
- Rezaei, E., B. Shahinejad and H. Yonesi. 2019. Analysis and evaluation of effective parameters on the amount of total dissolved solids in rivers. *Watershed Engineering and Management*, 11(1): 147-165. (In Persian).
- Kumar, M., Raghuvanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, 128(4), 224-233.
- Jhariya, D.C.; Kumar, T.; Pandey, H.K.; Kumar, S.; Kumar, D.; Gautam, A.K.; Kishore, N. Assessment of groundwater vulnerability to pollution by modified DRASTIC model and analytic hierarchy process. *Environ. Earth Sci.* 2019, 78, 1–20.
- Landeras, G., Ortiz-Barredo, A., López, J.J., 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agric. Water Manag.* 95, 553–565. doi:10.1016/j.agwat.2007.12.011
- Garewal, S.K.; Vasudeo, A.D.; Landge, V.S.; Ghare, A.D. A GIS-based Modified DRASTIC (ANP) method for assessment of groundwater vulnerability: A case study of Nagpur city, India. *Water Qual. Res.* 2015, 52, 121–135.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex System*, 13(2): 87–129.
- Khatibi, R., Ghorbani, M. A., Naghshara, S., Aydin, H. & Karimi, V. 2020 Introducing a framework for 'inclusive multiple modeling' with critical views on modeling practices – applications to modeling water levels of the Caspian Sea and Lakes Urmia and Van. *Journal of Hydrology* 587, 124923.
- Danandehmehr A., Olliaie E., Ghorbani M.A. 2010. Suspended sediment load prediction based on river discharge and genetic programming method. *Watershed Management Researches Journal (Pajouhesh & Sazandegi)*, 88:44-54. (In Persian with English abstract).
- Hydrology, 575, 730-742.
- Shoaib, M., Shamseldin, A.Y., Melville, B.W., and Khan, M.M. 2015. Runoff Forecasting using HybridWavelet Gene Expression Programming (WGEP) Approach. *J. Hydrol.* 527: 326-344.
- Nayak P.C., Sudheer K.P., Rangan D.M., and Ramasastry K.S. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1):52-66.
- Aytek A., and Kisi, O. 2008. A genetic programming approach to suspended sediment modeling. *Journal of Hydrology*, 351: 288-298.
- Guo, J., Zhou, J., Qin, H., Zou, Q., Li, Q., 2011. Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38(10): 13073-13081.
- Partington, D. et al., 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. *Journal of Hydrology*, 458-459(Supplement C): 28-39.
- Wang, W.-C., Chau, K.-W., Cheng, C.-T., Qiu, L., 2009. A comparison of the performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374(3): 294-306.
- Andualem, T.G.; Demeke, G.G. Groundwater potential assessment using GIS and remote sensing: A case study of Guna tana Landscape, upper blue Nile Basin, Ethiopia. *J. Hydrol. Reg. Stud.* 2019, 24, 100610.
- Patil, A.P., Deka, P.C., 2016. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* 121, 385–392. doi:10.1016/j.compag.2016.01.016.
- Kumar, M., Raghuvanshi, N.S., Singh, R., 2011. Artificial neural networks approach in evapotranspiration modeling: a review. *Irrigation Sci.* 29, 11-25. doi:10.1007/s00271-010-0230-8.
- Dorado J., J.R. Rabunal, A. Pazos, D. Rivero, A. Santos and J. Puertas. 2003. Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. *Applied Artificial Intelligence*, 17: 329-343.