نشریه علمی
علوم و مهندسی آبخیزداری ایران
Iran-Watershed Management Science & Engineering
Vol. 19, No. 70, Fall 2025
سال نوزدهم- شماره ۷۰- پاییز ۱۴۰۴

Daily Runoff Prediction Using Data Mining Models

Negin Rashidi [1], Vahid Moosavi [2*], Mehdi Vafakhah [3]

Received:07-10-2025, Revised: 25-10-2025, Accepted: 15-11-2025, Published: 17-11-2025
http://dx.doi.org/10.22034.19.70.5

Extended Abstract

Introduction

The rainfall–runoff process is one of the most complex and important components of the hydrological cycle, serving as a foundation for water resources management, flood forecasting, and watershed planning. Accurate estimation of runoff plays a critical role in optimizing the use of water resources, particularly in regions where hydrometric data are limited. However, due to the nonlinear and dynamic interactions between rainfall, temperature, soil moisture, and other climatic variables, physically based hydrological models often fail to adequately represent the real-world behavior of watersheds. In recent years, artificial intelligence (AI) and machine learning (ML) methods have emerged as powerful alternatives, offering flexible, data-driven frameworks capable of learning complex nonlinear relationships without requiring explicit physical equations. Among these approaches, the Group Method of Data Handling (GMDH), Random Forest (RF), and Adaptive Neuro-Fuzzy Inference System (ANFIS) models have shown great promise for hydrological applications. Previous studies have demonstrated the effectiveness of these models, but their performance may vary depending on watershed characteristics and data conditions. Therefore, the present study aimed to evaluate and compare the predictive performance of the GMDH, RF, and ANFIS models in simulating daily runoff in two Iranian watersheds i.e., Taleghan and Telvar, representing different topographic and climatic settings, in order to identify the most accurate and computationally efficient model for daily runoff forecasting.

Materials and Methods

The study was conducted in two Iranian watersheds: Taleghan in Alborz Province and Telvar in Kurdistan Province. The Taleghan basin covers about 790 km², with elevations ranging from 1,707 to 4,370 m and a mean elevation of 2,734 m. The Telvar basin extends over roughly 2,490 km², with elevations between 1,280 and 2,880 m and an average elevation of 1,927 m. Daily meteorological and hydrological data were obtained from the Regional Water Companies and the Iran Meteorological Organization for a continuous 13-year period. Input variables included 24-hour rainfall, air temperature, relative humidity, mean wind speed, total snow precipitation, antecedent precipitation index (API), and antecedent discharge index (ADI). The next-day discharge was used as the target variable. Preprocessing steps consisted of filling missing values, normalizing the dataset, and

1. M.Sc. Student in Watershed Sciences and Engineering, Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Negin.rashidi@modares.ac.ir.
2. Assistant Professor (Corresponding Author), Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Noor, Iran. Email: v.moosavi@modares.ac.ir.
3. Professor, Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, vafakhah@modares.ac.ir.

dividing it into 70% for training and 30% for testing. Identical subsets were applied to all models to ensure consistent comparison. Three prediction models were implemented in MATLAB. The GMDH model, based on the Kolmogorov–Gabor polynomial, automatically generated nonlinear polynomial relationships between inputs and output while iteratively selecting the optimal structure. The ANFIS model combined neural-network learning with fuzzy logic using a Sugeno-type inference system and a hybrid least-squares/backpropagation training algorithm. The RF model employed an ensemble of decision trees, each trained on bootstrap samples and random subsets of predictors, with final predictions obtained by averaging all trees. Model performance was evaluated using the coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), and root mean square error (RMSE).

Results and Discussion

The modeling results showed that all three AI-based models performed well in simulating daily runoff, but their accuracy differed between the two watersheds. In the Taleghan basin, the GMDH model achieved the highest accuracy, with $R^2 = 0.8845$, NSE = 0.8836, and RMSE = 4.09. The RF model ranked second with $R^2 = 0.8801$, NSE = 0.8798, and RMSE = 4.16, showing only a slight decline relative to GMDH. ANFIS produced the weakest results in Taleghan ($R^2 = 0.8710$, NSE = 0.8698, RMSE = 4.32), likely due to its sensitivity to membership function settings and the noise in the dataset. In the Telvar watershed, the performance pattern changed slightly. GMDH again provided the best results (RMSE = 0.3631, NSE = 0.9396, $R^2 = 0.9409$), but ANFIS closely matched it with RMSE = 0.3640, NSE = 0.9392, and $R^2 = 0.9406$. RF ranked last, achieving RMSE = 0.4017, NSE = 0.9260, and $R^2 = 0.9273$. The lower variability in Telvar likely enabled GMDH and ANFIS to capture stable nonlinear relationships more effectively. Comparing the best and weakest models showed that GMDH improved RMSE by about 9.6% in Telvar but only 1.9% in Taleghan. Overall, model superiority appears to depend on watershed characteristics, with GMDH excelling in stable conditions and RF performing relatively well in more variable settings.

Conclusion

The overall comparison shows that none of the models is universally superior; their effectiveness depends on watershed characteristics and data quality. The GMDH model demonstrated consistently high accuracy in both basins, particularly where hydrological conditions were more stable and correlations among variables were stronger. The RF model showed notable robustness, performing nearly as well as GMDH in the more complex Taleghan watershed, where data variability and noise were higher. ANFIS, although capable of modeling nonlinear relationships, required careful tuning of fuzzy rules and high-quality inputs, making it less reliable under variable conditions. These outcomes are consistent with earlier studies, such as those by Shortridge et al. (2016) and Mosavi et al. (2020), which emphasize that model selection should be adapted to local hydrological behavior rather than generalized across regions. Comparing performance across two different basins provided valuable insight into model transferability and highlighted the importance of matching model structure to watershed dynamics. Overall, the study confirms the potential of AI-based data-driven models for daily runoff prediction and their usefulness as complements to traditional hydrological approaches. GMDH achieved the best overall performance, while RF offered strong adaptability in noisy environments. ANFIS performed well when conditions were stable but was more sensitive to data inconsistencies. The findings suggest that site-specific model selection and the development of hybrid or ensemble methods combining GMDH, RF, and ANFIS could further improve predictive reliability. Accurate runoff forecasting using these techniques can support flood warning, reservoir operation, and water management in data-limited mountainous regions.

Keywords: *Artificial Intelligence, Data-Driven Model, Machine Learning, Modeling, Runoff*
Article Type: Research Article

Acknowledgement

This research is in the form of a master's thesis by the first author of the article and is financially supported by Tarbiat Modares

University. The authors also express their gratitude and appreciation to the Faculty of Natural Resources and Marine Sciences, Tarbiat Modares Pardis Noor.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the writing and publication of the contents and results of this research.

Data Availability Statement

The data and results used in this research will be made available upon reasonable request and correspondence with the corresponding author.

Authors' Contribution

Negin Rashidi**:** Data collection and preprocessing, computations, writing, drafting, and finalizing the manuscript. Vahid Moosavi: Research idea, guidance, and supervision of the research process, result validation, data analysis, manuscript editing, and revision.

Mehdi Vafakhah: Guidance on the research process, manuscript editing, and revision.

Citation: Rashidi N, Moosavi V, Vafakhah M, Daily Runoff Prediction using Data Mining Models. jwmseir 2025; 19 (70): 73-90

Iran-Watershed Management Science & Engineering, Year 2025, Vol 19, No 70, PP 73-90

Publisher: Watershed Management Society of Iran

پیش بینی رواناب روزانه با استفاده از مدل‌های یادگیری ماشین

نگین رشیدی[1] ، وحید موسوی[2] مهدی وفاخواه[3]

تاریخ دریافت: ۱۴۰۴/۰۷/۱۵، تاریخ داوری:۱۴۰۴/۰۸/۰۳، تاریخ پذیرش: ۱۴۰۴/۰۸/۲۴، تاریخ انتشار: ۱۴۰۴/ ۰۸/ ۲۶
http://dx.doi.org/10.22034.19.70.5

چکیده

مدل‌سازی فرآیند بارش-رواناب، یکی از روش‌های رایج در برآورد رواناب بوده و ابزاری کارآمد برای تحلیل فرآیندهای هیدرولوژیکی، ارزیابی منابع آب و مدیریت بهینه آبخیزها به‌شمار می‌رود. از این رو، بهره‌گیری از روش‌هایی که در عین پویایی، دارای ساختار مفهومی روشن، قابلیت توسعه و کاربرد آسان باشند، ضرورت دارد. در این تحقیق، کارایی سه مدل هوش‌مصنوعی جنگل تصادفی (RF)، برخورد گروهی با داده‌ها (GMDH) و استنتاج فازی-عصبی تطبیقی (ANFIS) برای پیش‌بینی رواناب روزانه در دو آبخیز تلوار و طالقان مورد ارزیابی قرار گرفت. داده‌های مورد استفاده شامل بارش ۲٤ ساعته، دما، رطوبت نسبی، میانگین سرعت باد، میانگین رطوبت نسبی، میزان بارش برف کلی، و نیز شاخص بارش پیشین و شاخص دبی پیشین برای دوره آماری ۱۳ ساله در مقیاس زمانی روزانه جمع‌آوری و محاسبه شد. نتایج حاصل از مدل‌سازی برای آبخیز طالقان نشان داد که مدل GMDH با بیشترین مقادیر ضریب تعیین (R^2= ۰/۸۸٤٥) و معیار نش-ساتکلیف (NSE= ۰/۸۸۳٦) و همچنین کمترین ریشه میانگین مربعات خطا، با (RMSE= ٤/۰۹) بالاترین دقت و کمترین خطا را در شبیه‌سازی رواناب ارائه داد. مدلRF با مقادیر(R^2=/۸۸۰۱ ،NSE=/ ۸۷۹۸،۰ ،۰ ٤/۱٦ =RMSE) در رتبه دوم قرار گرفت و توانست نتایج نسبتاً قابل قبولی تولید کند، هرچند دقت آن اندکی کمتر از مدل GMDH بود. در مقابل، مدل ANFIS با R^2= ۰/۸۷۱۰ ، NSE= ۰/۸٦۱۰ وRMSE= ٤/۳۲ ضعیف‌ترین عملکرد را در میان مدل‌های مورد بررسی داشت. و همچنین نتایج برای آبخیز تلوار نشان داد که مدل GMDH با مقادیر RMSE= ۰/ ۳٦۳۱ وNSE= ۰/ ۹۳۹ و مدل RF با مقادیر NSE= ۰/۹۲٦۰ و RMSE= ۰/٤۰۱۷ و مدل ANFIS با مقادیر RMSE= ۰/۳٦٤۰ و NSE= ۰/۹۳۹۲ بیانگر آن است که مدل‌های مورد بررسی از لحاظ عملکرد به این ترتیب قرار گرفته‌اند: RF،GMDH وANFIS. نتایج نشان داد که هر سه مدل عملکرد کلی نسبتاً مطلوبی در شبیه‌سازی تغییرات رواناب داشتند که نشان می‌دهد استفاده از مدل‌های هوش مصنوعی برای تخمین رواناب سطحی ابزار مناسبی است.

کلیدواژه‌ها: هوش مصنوعی، رواناب ،مدل داده‌محور، مدل‌سازی ، یادگیری ماشین

نوع مقاله: پژوهشی

استناد: رشیدی نگین، موسوی وحید، وفاخواه مهدی. پیش‌بینی رواناب روزانه با استفاده از روش های یادگیری ماشین. علوم ومهندسی آبخیزداری ایران. ۱٤۰٤؛ ۱۹(۷۰): ۷۳- ۹۰

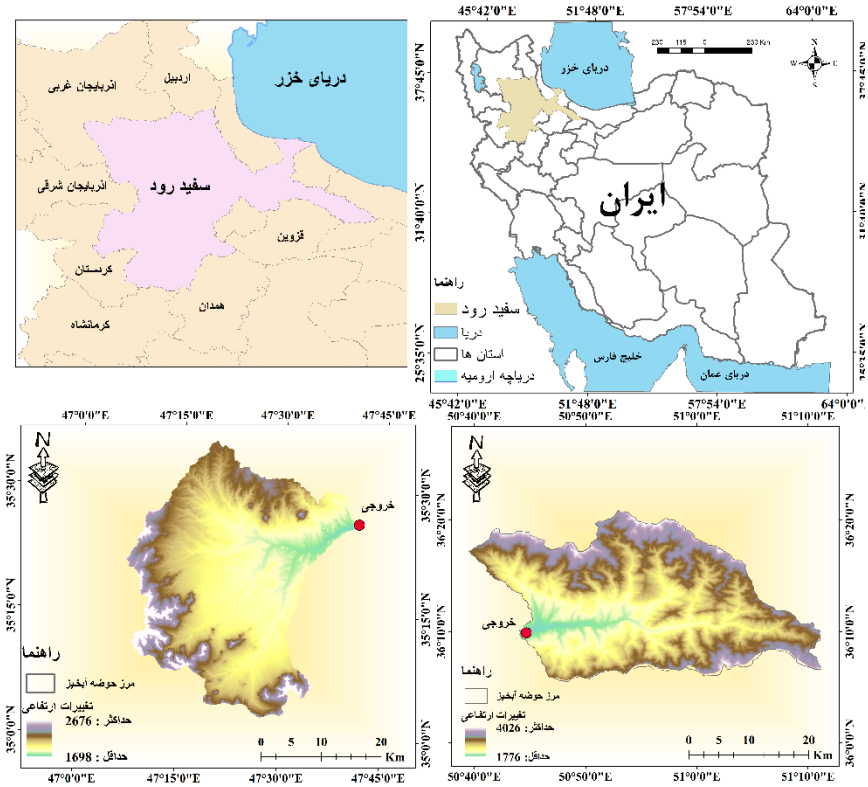
۱- دانشجوی کارشناسی ارشد علوم و مهندسی آبخیزداری، گروه مهندسی ابخیزداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، نور، ایران.
۲- دانشیار(نویسنده مسئول)، گروه مهندسی آبخیزداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، نور، ایران. پست الکترونیکی: v.moosavi@modares.ac.ir
۳. استاد گروه مهندسی آبخیزداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، نور، ایران.

مقدمه

بارش به‌عنوان یکی از اجزای کلیدی چرخه هیدرولوژیک، محرک اصلی در فرآیند تولید رواناب محسوب می‌شود. رواناب ناشی از بارش، پتانسیل آن را دارد که به‌طور مستقیم در مدیریت و بهره‌برداری بهینه از منابع آب برای تأمین نیازهای آبی مورد استفاده قرار گیرد [۲۸،۱]. با این حال، طراحی سازه‌های منابع آب نیازمند داده‌های رواناب بلندمدت است، اما در کشورهای در حال توسعه به دلیل هزینه‌های بالای احداث و نگهداری ایستگاه‌های اندازه‌گیری دبی، دسترسی به این داده‌ها همواره با چالش‌هایی مواجه است [۲۳]. درک و پیش‌بینی فرآیندهای تولید رواناب و انتقال آن به نقطه خروجی آبخیز، یکی از چالش‌های اساسی در علم هیدرولوژی محسوب می‌شود [۵]. تحلیل رواناب در مقیاس‌های زمانی روزانه و مکانی یک حوزه آبخیز، فرآیندی پیچیده است که با روابط غیرخطی و عوامل متعددی مانند نفوذ، دما و تبخیر در ارتباط است [۳]. پیش‌بینی دقیق دبی روزانه از اهمیت ویژه‌ای برخوردار است، چرا که این اطلاعات به مدیران منابع آب کمک می‌کند تا تصمیمات بهتری در زمینه تخصیص و مصرف منابع آبی اتخاذ کنند. به همین دلیل، پژوهشگران به طور مستمر به دنبال روش‌ها و مدل‌های نوین و دقیق‌تری برای پیش‌بینی دبی و رواناب هستند. با این حال، وجود چالش‌هایی همچون تأثیر نفوذ، دما، تبخیر و نیز ماهیت غیرخطی و پیچیده این فرآیندها، فرایند مدل‌سازی را با دشواری‌هایی روبه‌رو کرده است. دستیابی به درکی صحیح از آبخیز و فرآیندهای حاکم بر آن می‌تواند نقش مؤثری در بهبود دقت مدل‌سازی فرآیند بارش–رواناب ایفا کند[٤]. به همین دلیل، مدل‌سازی هیدرولوژیکی می‌تواند به عنوان راه‌حلی برای این مشکل عمل کند، زیرا با ایجاد ارتباط میان فرآیندهای مختلف هیدرولوژیکی، امکان پیش‌بینی رواناب فراهم می‌شود [۲٦]. مدل‌ها به عنوان ابزارهای اصلی، امکان شبیه‌سازی و پیش‌بینی تغییرات چرخه آب در طبیعت را فراهم کرده و به ارزیابی مؤلفه‌های بیلان آب کمک می‌کنند. به عبارت دیگر، یکی از بهترین روش‌ها برای درک و تحلیل سامانه‌های هیدرولوژیک و منابع آبی، استفاده از مدل‌های هیدرولوژیک است. در این راستا مدل‌های هوش مصنوعی به عنوان یکی از مهم‌ترین روش‌های پیش‌بینی رواناب مورد توجه قرار گرفتند. در زمینه هیدرولوژی، از یادگیری ماشین برای درک بهتر پیچیدگی‌های هیدرولوژیکی استفاده شده است. در نتیجه، استفاده از رویکردهای مبتنی بر یادگیری ماشین برای شبیه‌سازی هیدرولوژیکی، به ویژه مدل‌سازی رواناب، در سال‌های اخیر به روشی پرطرفدار تبدیل شده است [۱۵]. هوش مصنوعی به عنوان یکی از مؤثرترین رویکردهای نوین، نقش مهمی در مدل‌سازی و پیش‌بینی رواناب ایفا می‌کند و به سامانه‌هایی اطلاق می‌شوند که قادرند با الهام از توانایی‌های انسانی مانند یادگیری، استدلال و تصمیم‌گیری، داده‌ها را تحلیل کرده و الگوهای پیچیده را شناسایی کنند. مدل‌های شبکه عصبی مصنوعی با استفاده از حداقل داده‌های ورودی، قادرند روابط میان متغیرها و ساختارسامانه را تحلیل کنند. این مدل‌ها از پتانسیل بالایی در انجام وظایفی مانند پیش‌بینی و دسته‌بندی اطلاعات برخوردارند و از طریق ساختار غیرخطی خود، تعاملات پیچیده میان متغیرها را به‌خوبی بازنمایی می‌کنند. مدل‌های RF[2]، GMDH[1] و ANFIS[3] از جمله روش‌های پیشرفته یادگیری ماشین و هوش مصنوعی هستند که به طور گسترده‌ای برای مدل‌سازی و پیش‌بینی رواناب در حوزه هیدرولوژی استفاده می‌شوند. این مدل‌ها با توانایی شناسایی روابط غیرخطی و پیچیده بین داده‌ها، دقت بالایی در تخمین رواناب ارائه می‌دهند.

در سال‌های اخیر، مدل‌های هوش مصنوعی و یادگیری ماشین به طور گسترده برای پیش‌بینی رواناب روزانه استفاده شده‌اند Moosavi و همکاران [١٦] مدلی تلفیقی بر پایه GMDH و تبدیل موجک ارائه کردند و نشان دادند که استفاده از موجک و بسته موجک دقت مدل را افزایش می‌دهد، هرچند روش دوم هزینه محاسباتی بیشتری دارد. در ادامه، Moosavi و همکاران [١٧] با بهره‌گیری از روش تاگوچی، اهمیت عوامل مؤثر بر پیش‌بینی رواناب را بررسی کرده و دریافتند که ترکیب GMDH و موجک با داده‌های کامل و دوره ۲۰ ساله دقیق‌ترین نتایج را ارائه می‌دهد. مطالعات دیگر نیز نشان‌دهنده توانمندی مدل‌های مختلف بوده‌اند پورنعمت‌رودسری و همکاران [۲۰] در مطالعه‌ای نشان دادند که هر دو مدل GMDH و ANN در پیش‌بینی رواناب مؤثرند اما GMDH عملکرد بهتری دارد Pour .و همکاران[۱۹] نیز در پژوهش خود بیان کردند که GMDH در مقایسه با GEP در همه حوضه‌ها نتایج رضایت‌بخش‌تری داشته است Qiao .و همکاران [۲۱] مدلی ترکیبی مبتنی بر IAO، TCN و RF ارائه دادند که در شبیه‌سازی بارش–رواناب موفق بود. همچنین Gaertner، [۹]، با استفاده از مدل RF رواناب منطقه آتلانتیک میانی را برای سال‌های ۲۰۲۲–۲۰۳۱ پیش‌بینی کرد. Liu و همکاران[۱۳] سهم تغییرات اقلیمی و انسانی را با مدل RF بررسی کردند. Vilasec و همکاران] ۲۷ [با استفاده از روش SHAP مهم‌ترین متغیرهای مؤثر در مدل‌های RF را شناسایی کردند. در حوزه مدل‌های Remesan، ANFIS و همکاران[۲۲] عملکرد بالای ANFIS را در پیش‌بینی رواناب نشان دادند. همچنین، اسماعیلی گیساوندانی و همکاران[۸] و کیا و همکاران [۱۲] دقت بالاتر ANFIS نسبت به مدل‌های رگرسیونی را گزارش کردند. با توجه به این مطالعات، پژوهش حاضر به مقایسه سه مدل GMDH، RF و ANFIS در دو حوزه آبخیز طالقان و تلوار می‌پردازد تا با در نظر گرفتن دقت و هزینه محاسباتی، بهترین مدل برای پیش‌بینی رواناب روزانه انتخاب شود. انتخاب دو حوزه مورد مطالعه به این دلیل صورت گرفته است که از نظر ویژگی‌های توپوگرافی، اقلیمی و شرایط آب‌وهوایی تفاوت قابل‌توجهی با یکدیگر دارند. این تفاوت‌ها امکان بررسی عملکرد مدل در شرایط متنوع را فراهم می‌کند و در نتیجه، ارزیابی دقت و پایداری مدل در موقعیت‌های مختلف ممکن می‌شود.

1. Group Method of Data Handling
2. Random forest
3. Adaptive neuro fuzzy inference system



شکل-۱ موقعیت مکانی حوزه‌های آبخیز مورد مطالعه

Fig 1. Location of the study Area

اگرچه مدل‌های هوش مصنوعی مانند GMDH، RF وANFIS پیش‌تر در مدل‌سازی رواناب به‌کار رفته‌اند، اما بیشتر مطالعات گذشته صرفاً به ارزیابی عملکرد یک مدل در یک حوزه آبخیز خاص محدود بوده و درک جامعی از میزان پایداری و کارایی مدل‌ها در شرایط مختلف اقلیمی و هیدرولوژیکی ارائه نکرده‌اند. در شرایط کنونی که تغییرات اقلیمی و رشد نامتوازن الگوهای بارش، پیش‌بینی دقیق رواناب را برای مدیریت لحظه‌ای منابع آب و کاهش خطر سیلاب ضروری کرده است، نیاز به رویکردی تطبیقی و مقایسه‌ای میان مدل‌های یادگیری ماشین بیش از پیش احساس می‌شود. از این‌رو، نوآوری اصلی پژوهش حاضر در ارزیابی هم‌زمان و نظام‌مند سه مدل پیشرفته هوش مصنوعی GMDH، RF و ANFIS در دو حوزه آبخیز با ویژگی‌های اقلیمی و هیدرولوژیکی متمایز (تلوار و طالقان) نهفته است. این مقایسه تطبیقی به درک عمیق‌تری از حساسیت مدل‌ها نسبت به تفاوت‌های مکانی و داده‌ای منجر می‌شود و راهکارهایی عملی برای به‌کارگیری مدل بهینه در شرایط واقعی مدیریت منابع آب ارائه می‌دهد. با وجود پیشرفت‌های قابل‌توجه در استفاده از مدل‌های هوش مصنوعی در هیدرولوژی، یک خلأ در ادبیات علمی از نظر بررسی تطبیقی مدل‌های مختلف در مقیاس روزانه و در حوضه‌هایی با تفاوت‌های بارز اقلیمی و هیدرولوژیکی وجود دارد. اغلب پژوهش‌ها، پایداری عملکرد مدل‌ها در برابر تغییرات مکانی و داده‌ای را نادیده گرفته‌اند، در نتیجه، انتخاب مدل بهینه برای شرایط متنوع محیطی همچنان با عدم قطعیت همراه است.

مواد و روش ها

منطقه مورد مطالعه

این تحقیق در دو حوزه آبخیز تلوار و طالقان انجام شده است. آبخیز طالقان با مساحت حدود ۷۹۰ کیلومتر مربع واقع در رشته کوه البرز با عرض جغرافیایی °۳۶ و '۱۱ تا °۳۶ و '۵۰ شمالی و طول جغرافیایی °۵۰ و '۴۵ تا °۵۱ و '۱۱ شرقی قرار گرفته و حداکثر و حداقل ارتفاع آن به ترتیب ۴۳۷۰ و ۱۷۰۷متر از سطح دریا بوده و ارتفاع متوسط منطقه نیز ۲۷۳۴ متر از سطح دریا است. آبخیز طالقان در رشته‌کوه البرز و در حدود ۹۰ کیلومتری شمال‌غرب تهران واقع شده است. این حوضه دارای توپوگرافی متنوعی شامل دره‌های عمیق و شیب‌های تند است که این ویژگی‌ها موجب پیچیدگی‌های هیدرولوژیکی آن می‌شود. آبخیز رودخانه تلوار در استان کردستان با مساحت ۲۴۹۰ کیلومترمربع در موقعیت جغرافیایی °۴۵ و ‹۴۷ تا ° ۴۷ و‹۰۶ طول شرقی °۳۵ و‹۰۳ تا°۳۵ و ‹۳۵ عرض شمالی قرار دارد. حداقل ارتفاع این حوضه ۱۲۸۰ متر از سطح دریا است که در ایستگاه هیدرومتری تلوار–حسن‌خان در خروجی حوضه واقع شده است. حداکثر ارتفاع آن ۲۸۸۰ متر و ارتفاع متوسط ۱۹۲۷ متر است. آبخیز تلوار در شهرستان‌های قروه و دهگلان در استان کردستان واقع شده است. آبخیز تلوار دارای توپوگرافی کوهستانی تا تپه‌ماهور است که بیشترین ارتفاع آن در بخش‌های جنوبی و جنوب‌غربی حوضه قرار دارد[۱۸]. شکل (۱) مناطق مورد مطالعه در این تحقیق را نشان می‌دهد.

روش انجام پژوهش

داده‌های ورودی مدل بر اساس ارتباط فیزیکی و هیدرولوژیکی آن‌ها با تولید رواناب شامل داده‌های هواشناسی از قبیل بارش، دما، دبی، دبی با تاخیر زمانی، میانگین سرعت باد، میانگین رطوبت نسبی، میانگین ابرناکی، شاخص بارش‌های پیشین [1] (API) شاخص دبی‌های پیشین [2](ADI) و داده هدف دبی برای یک روز بعد انتخاب شده‌اند تا رفتار سیستم به‌طور جامع شبیه‌سازی شود. به صورت روزانه برای یک دوره ۱۳ ساله از (شرکت‌های سهامی آب منطقه‌ای استان‌ها و سازمان هواشناسی کشور) اخذ و نواقص آماری احتمالی رفع و مرتب شد. در این دوره آماری (۲۰۲۱-۲۰۰۸)، ۷۰ درصد داده‌ها به منظور آموزش مدل‌ها و ۳۰ درصد باقی‌مانده برای آزمون استفاده شدند. با توجه به هدف مقایسه عملکرد چند مدل مختلف، داده‌های بخش آموزش و آزمون برای همه مدل‌ها به صورت یکسان انتخاب شدند تا شرایط مقایسه منصفانه و قابل اعتماد فراهم شود. سپس فرآیند مدل‌سازی با استفاده از مدل‌های جنگل تصادفی، برخورد گروهی با داده‌ها و استنتاج فازی-عصبی تطبیقی انجام شد و صحت و دقت مدل ارائه شده با استفاده از نمایه‌های آماری RMSE (ریشه میانگین مربعات خطا)، NSE (ضریب نش-ساتکلیف)، R^2 (ضریب تعیین) مورد ارزیابی قرارگرفت و در نهایت به مقایسه این سه مدل پرداخته شد. تمامی بخش‌های تحقیق با کدنویسی در زبان برنامه‌نویسی متلب[3] صورت پذیرفت. در ادامه توضیحاتی در خصوص این مدل‌ها ارائه شده است.

به منظور محاسبه شاخص‌های بارش پیشین (API) و دبی پیشین (ADI) از رابطه(۱و۲) استفاده شد.

$$API = \sum_{t=-1}^{-i} P_t K^{-t} \quad (۱)$$

$$ADI = \sum_{t=-1}^{-i} D_t K^{-t} \quad (۲)$$

مدل کنترل گروهی داده‌ها

الگوریتم GMDH به‌عنوان یکی از روش‌های مبتنی بر شبکه‌های عصبی مصنوعی شناخته می‌شود که ساختاری چندلایه و سلسله مراتبی دارد. یکی از روش های یادگیری ماشین و هوش مصنوعی است و ساختاری شبیه شبکه‌های عصبی دارد اما برخلاف آن‌ها، ساختار مدل به‌صورت پویا و طی فرآیندی انتخاب می‌شود که در آن ترکیب‌های مختلف متغیرهای ورودی ارزیابی شده و بهترین ترکیب‌ها بر اساس معیارهای خطا یا دقت پیش‌بینی انتخاب می‌شوند. این الگوریتم به‌صورت خودسازمان‌ده طراحی شده است و هدف آن مدل‌سازی ارتباط بین متغیرهای ورودی و خروجی در سیستم‌های غیرخطی و پیچیده است. در این ساختار، هر لایه از شبکه شامل نورون‌هایی است که به‌طور خودکار ترکیب‌های مختلفی از متغیرهای ورودی را بررسی می‌کنند. این ویژگی، الگوریتم GMDH را به ابزاری قدرتمند برای تحلیل داده‌های چندمتغیره تبدیل کرده است. یک مجموعه داده شامل چندین متغیر ورودی و یک متغیر خروجی ($X_i, Y_i = 1, 2, \ldots, M$) است. هر نرون در شبکه، تابعی از ترکیب چند متغیر ورودی است که معمولاً به‌صورت یک چندجمله‌ای غیرخطی (غالباً درجه دوم) مدل‌سازی می‌شود. این توابع با استفاده از تحلیل رگرسیون و با هدف کمینه‌سازی خطای پیش‌بینی تخمین زده می‌شوند شناخته‌شده‌ترین فرم‌های تابع به‌کار رفته در GMDH، چندجمله‌ای Kolmogorov-Gabor است که رابطه بین ورودی‌ها و خروجی‌ها را به صورت معادله(۳) بیان می‌کند. y، x و a به‌ترتیب خروجی شبکه، بردار ورودی و بردار ضریب در چندجمله‌ای Kolmogorov-Gabor هستند. در این الگوریتم، مدل‌های مختلفی از ترکیب ورودی‌ها ساخته شده و عملکرد آن‌ها از طریق داده‌های اعتبارسنجی ارزیابی می‌شود. در هر لایه، تنها مدل‌هایی که عملکرد مناسب‌تری در پیش‌بینی دارند، انتخاب و به لایه بعدی منتقل می‌شوند. به‌عبارتی، الگوریتم GMDH با استفاده از یک سازوکار انتخابگر، مدل‌های ضعیف را حذف کرده و ترکیب‌هایی با دقت بالاتر را در سطوح بالاتر مورد استفاده قرار می‌دهد. این فرایند تا زمانی ادامه می‌یابد که افزودن لایه‌های جدید منجر به بهبود قابل‌توجهی در دقت پیش‌بینی نشود. در نتیجه، ساختار نهایی مدل شامل تعداد بهینه‌ای از لایه‌ها و نرون‌ها خواهد بود که به‌صورت خودکار توسط الگوریتم تعیین می‌شوند[۲].

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} x_i x_j + \quad (۳)$$

مدل سامانه استنتاج عصبی-فازی تطبیقی

مدل ANFIS که از ترکیب منطق فازی و شبکه‌های عصبی مصنوعی حاصل شده است، یکی از ابزارهای قدرتمند در مدل‌سازی پدیده‌های پیچیده و غیرخطی در سامانه‌های طبیعی و انسانی به شمار می‌رود. هدف این مدل بهره‌گیری از توانایی شبکه‌های عصبی در یادگیری و از قابلیت سامانه‌های فازی در استنتاج تقریبی است. در واقع، ANFIS ساختار یک سیستم فازی نوع سوگنو (Sugeno-type) را با الگوریتم‌های یادگیری شبکه عصبی ترکیب می‌کند تا بتواند روابط ورودی و خروجی داده‌ها را با دقت بالا مدل‌سازی کند. در مدل ANFIS، قواعد فازی به‌صورت (اگر-آنگاه) تعریف شده‌اند و از توابع عضویت برای نمایش عدم قطعیت در داده‌ها استفاده می‌شود. این مدل از دو مرحله آموزش[4] بهره می‌برد، مرحله اول با استفاده از روش پس‌انتشار خطا (Backpropagation) برای تنظیم توابع عضویت، و مرحله دوم با الگوریتم کم‌ترین مربعات برای تعیین پارامترهای خروجی انجام می‌شود. این ترکیب باعث افزایش دقت، انعطاف‌پذیری و توانایی تعمیم‌دهی مدل می‌شود[۱۱].

1. Antecedent Precipitation Index
2. Antecedent Discharge Index
3. Matlab
4. Training

معماری این سامانه شامل پنج لایه است که لایه‌های اول و چهارم به ترتیب شامل پارامترهای مقدماتی و نتیجه‌ای هستند که با استفاده از روش‌های یادگیری ترکیبی به‌روزرسانی می‌شوند. ANFIS از یک رویکرد یادگیری ترکیبی استفاده می‌کند که شامل تخمین‌گر حداقل مربعات برای به‌روزرسانی پارامترهای رو به جلو و انتشار خطا به عقب برای انتشار خطای رو به عقب است. در حالی که ANFIS مزایای قابل‌توجهی در مدل‌سازی سامانه‌های پیچیده ارائه می‌دهد، توجه به این نکته مهم است که عملکرد سامانه به شدت به کیفیت داده‌های ورودی و تنظیمات اولیه قوانین فازی و توابع عضویت بستگی دارد. به عبارتی، لایه اول، لایه گره‌های ورودی است؛ هر گره در این لایه X یا Y ورودی‌های غیرفازی به گره i و A_1 و B_1 برچسب‌های زبانی‌ای هستند که به‌ترتیب با توابع عضویت μ_{A_i} و μ_{B_i} با استفاده از رابطه (۳و٤) مشخص می‌شوند.

$$O_{1,i} = \mu_{A_i}(X) for \quad i = 1,2 \tag{٤}$$

$$O_{1,i} = \mu_{B_{i-2}}(Y) \; for \quad i = 3,4 \tag{٥}$$

با توجه به رابطه (٦) در لایه دوم، هر گره سیگنال ورودی خود را ضرب می‌کند و خروجی، حاصل‌ضرب تمام سیگنال‌های ورودی است.

$$O_{2,i} = W_i = \mu_{A_i}(X) \times \mu_{B_{i-2}}(Y), \quad i = 1,2 \tag{٦}$$

در لایه سوم از طریق رابطه (۷) گره i ام که با N نام‌گذاری شده است، نسبت‌های نرمال‌شده مربوط به قوانین را محاسبه می‌کند.

$$O_{3,i} = \overline{W} = \frac{W_i}{W_i + W_2}, \quad i = 1,2 \tag{۷}$$

لایه چهارم: گره i در این لایه، مشارکت قانون i ام را به سمت خروجی مدل، با استفاده از معادله (۸) محاسبه می‌کند.

$$O_{4,i} = \overline{W}_i F = \overline{W}_i (p_i + q_i + r_i), \quad i = 1,2 \tag{۸}$$

که در آن $\overline{W}_i$ خروجی لایه سه بوده و $\{p_i, q_i, r_i\}$ مجموعه پارامترها هستند. لایه پنجم: تنها گره این لایه، خروجی کلی ANFIS را به‌عنوان جمع‌بندی تمام ورودی‌های مدل که غیر فازی شده‌اند، مطابق رابطه (۹) محاسبه می‌کند [۱۰].

$$O_{5,i} \sum_{i=1}^{4} \overline{W}_i \, f \frac{\sum_{i=1}^{4} W_i f}{\sum_{i=1}^{4} W_i} \tag{۹}$$

برای توابع عضویت ورودی در مدل ANFIS ، چندین نوع تابع پرکاربرد به‌طور هم‌زمان مورد آزمایش قرار گرفتند تا تأثیر شکل توابع بر دقت مدل شناسایی شود. این توابع شامل مثلثی[1] ، ترپزی[2]، گوسی[3] و سایه‌دار[4] بودند که هر یک ویژگی‌های متفاوتی در بازنمایی عدم قطعیت و توزیع داده‌ها دارند. توابع مثلثی و ترپزی به دلیل ساختار ساده و محاسبات سریع، اغلب برای داده‌هایی با تغییرات خطی یا تقریبی مناسب هستند، در حالی که توابع گوسی و سایه‌دار می‌توانند رفتارهای نرم و پیوسته متغیرها را بهتر مدل‌سازی کنند و انعطاف بیشتری در آموزش شبکه عصبی فراهم می‌آورند. در فرآیند آماده‌سازی مدل، هر نوع تابع عضویت با تعداد قوانین فازی مختلف و پارامترهای متغیر ورودی ترکیب شد و مدل‌های حاصل با استفاده از داده‌های بخش آموزش و معیارهای آماری استاندارد شامل ضریب تعیین (R^2) ، شاخص نش-ساتکلیف (NSE) و ریشه میانگین مربعات خطا (RMSE) ارزیابی شدند. این رویکرد امکان مقایسه دقیق عملکرد توابع عضویت مختلف را فراهم کرده و به انتخاب بهترین ترکیب تابع و پارامترها در مراحل بعدی آموزش کمک می‌کند، بدون اینکه در این بخش نتیجه نهایی تعیین شود.

مدل جنگل تصادفی

الگوریتم جنگل تصادفی[5]، یکی از الگوریتم‌های ترکیبی مهم در حوزه یادگیری ماشین است[١٤]. این الگوریتم، یک روش جمعی درختی است که با میانگین‌گیری پیش‌بینی‌های درخت تصمیم به پیش‌بینی متغیر وابسته می‌پردازد. هر درخت با استفاده از نمونه bootstrap از داده‌های آموزشی و زیرمجموعه تصادفی از متغیرهای پیش‌بینی‌کننده آموزش داده می‌شود. هر گره درخت نمایانگر یک انتخاب است و گره برگ نمایانگر تصمیم است. جنگل تصادفی قادر به شناسایی روابط خطی و غیرخطی برای طبقه‌بندی و رگرسیون است[٦]. این روش با استفاده از تعداد زیادی درخت تصمیم، داده‌ها را با سرعت بالا دسته‌بندی کرده و به‌عنوان یک مدل غیرخطی در خانواده روش‌های دسته‌بندی شناخته می‌شود [۲۵]. تعداد پیش‌فرض درختان مدل برای هر جنگل برابر با ۱۰۰درخت= (n_estimators 100)در نظر گرفته شد و عملکرد مدل بر اساس آزمون حساسیت به تعداد درختان مورد بررسی قرار گرفت تا اطمینان حاصل شود افزایش بیشتر تعداد درختان باعث بهبود قابل‌توجهی در دقت نمی‌شود. علاوه بر این، هایپرپارامترهای مهم مانند تعداد ویژگی‌های انتخابی در هر تقسیم و حداکثر عمق درخت‌ها با استفاده از روش جستجوی شبکه‌ای و اعتبارسنجی متقابل تنظیم شدند تا بهترین ترکیب برای داده‌های هر حوزه آبخیز انتخاب شود. خروجی نهایی جنگل تصادفی به‌صورت میانگین پیش‌بینی تمام درخت‌ها محاسبه شد. این ساختار باعث می‌شود مدل هم‌زمان از قابلیت تعمیم‌دهی بالا برخوردار باشد و هم در برابر نوسانات داده‌ها و نویز مقاوم باشد. عملکرد مدل در پیش‌بینی دبی روزانه با استفاده از معیارهای آماری R^2 ، NSE و RMSE ارزیابی شد و مقایسه با سایر مدل‌ها انجام گرفت.

1. Triangular
2. Trapezoidal
3. Gaussian
4. Bell-shaped
5. Random Forest

نتایج

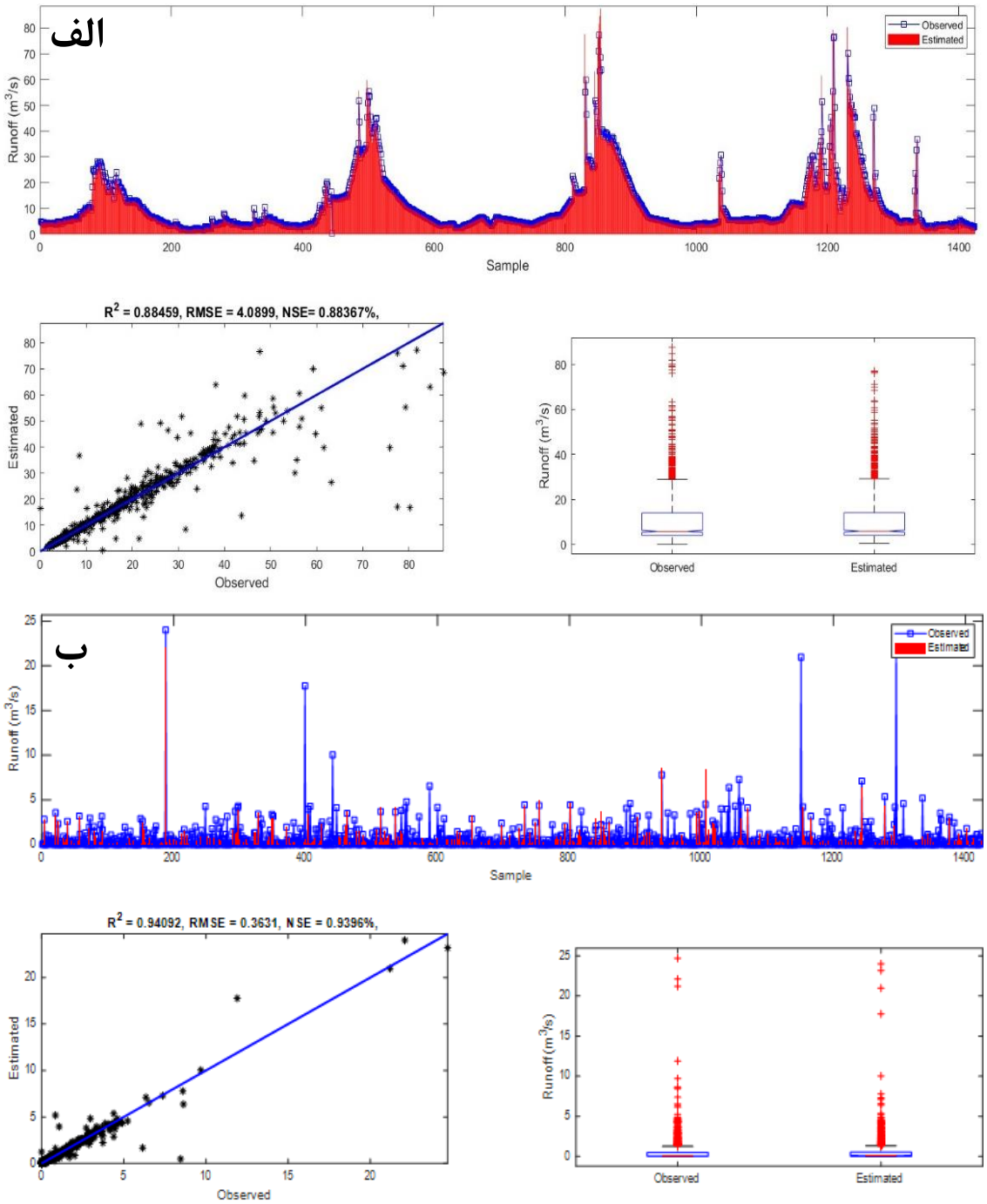
جدول ۱، شاخص‌های آماری مربوط به متغیرهای اقلیمی و هیدرولوژیکی آبخیز طالقان و تلوار را نشان می‌دهد. همان‌طور که مشاهده می‌شود، در آبخیز طالقان میانگین بارندگی سالانه در این آبخیز برابر با ۶۲۳/۸۳ میلی‌متر بوده که با ضریب تغییرات ۰/۲۶ نشان‌دهنده نوسانات نسبتاً متوسط بارندگی در دوره مورد بررسی است. میانگین دمای سالانه ۱۲/۴۹ درجه سانتی‌گراد بوده و با ضریب تغییرات پایین ۰/۰۵ حاکی از پایداری نسبی دما در منطقه است. سرعت باد با میانگین ۱/۸۳ متر بر ثانیه و ضریب تغییرات ۰/۰۹ تغییرات کمی را نشان می‌دهد. میانگین رطوبت نسبی ۵۱/۴۳ درصد است که با توجه به انحراف معیار پایین، شرایط نسبتاً یکنواخت رطوبتی در محدوده مورد مطالعه را نشان می‌دهد. همچنین میانگین دبی جریان ۱۱/۲۹ مترمکعب بر ثانیه بوده و ضریب تغییرات ۰/۲۰ بیانگر تغییرپذیری متوسط جریان در خروجی حوضه است. در آبخیز تلوار با توجه به شاخص‌های آماری متغیرهای اقلیمی و هیدرولوژیکی میانگین بارندگی سالانه این آبخیز۳۴ /۳۰۲ میلی‌متر است که با ضریب تغییرات ۰/۳۴ نشان‌دهنده نوسانات قابل توجه در میزان بارش‌های سالانه است. میانگین دمای سالانه برابر ۱۳/۸۷ درجه سانتی‌گراد بوده و با توجه به ضریب تغییرات ۰/۱۴، دما از پایداری نسبی متوسطی برخوردار است. سرعت باد با میانگین ۲/۹۹ متر بر ثانیه و ضریب تغییرات ۱۷/ ۰نشان‌دهنده تغییرات نه‌چندان زیاد در الگوی باد منطقه است. میانگین رطوبت نسبی ۴۶/۰۵ درصد بوده که مقدار ضریب تغییرات ۰/۱۱، یکنواختی نسبی شرایط رطوبتی را در محدوده مورد مطالعه نشان می‌دهد. در مقابل، میانگین دبی جریان سطحی ۰/۵۸ مترمکعب بر ثانیه و ضریب تغییرات بالا بیانگر تغییرپذیری بسیار زیاد جریان‌های خروجی از آبخیز است؛ موضوعی که می‌تواند ناشی از تغییرات قابل توجه بارش، نفوذپذیری خاک، و ویژگی‌های فیزیوگرافی منطقه باشد.

با توجه به شکل (۲، ۳ و۴) ، نمودار (الف) نتایج بخش آزمون برای مدل‌سازی رواناب روزانه در حوضه طالقان و نمودار (ب) نتایج مدلسازی رواناب روزانه در حوضه تلوار برای مدل GMDH،RF،ANFIS را نشان می‌دهد. مدل GMDH با بالاترین مقدار R^2 و بیشترین NSE و همچنین کمترین مقدار RMSE در مقایسه با سایر مدل‌ها، بیشترین دقت و کمترین خطا را در پیش‌بینی داده‌ها نشان داده است. این نتایج حاکی از توان بالای این مدل در شناسایی الگوهای داده و تولید برآوردهای قابل اطمینان است.

جدول 1- ویژگی‌های آماری متغیرهای هیدرواقلیمی مناطق مورد مطالعه

Table 1. Statistical characteristics of the hydro-climatic variables of the studied regions

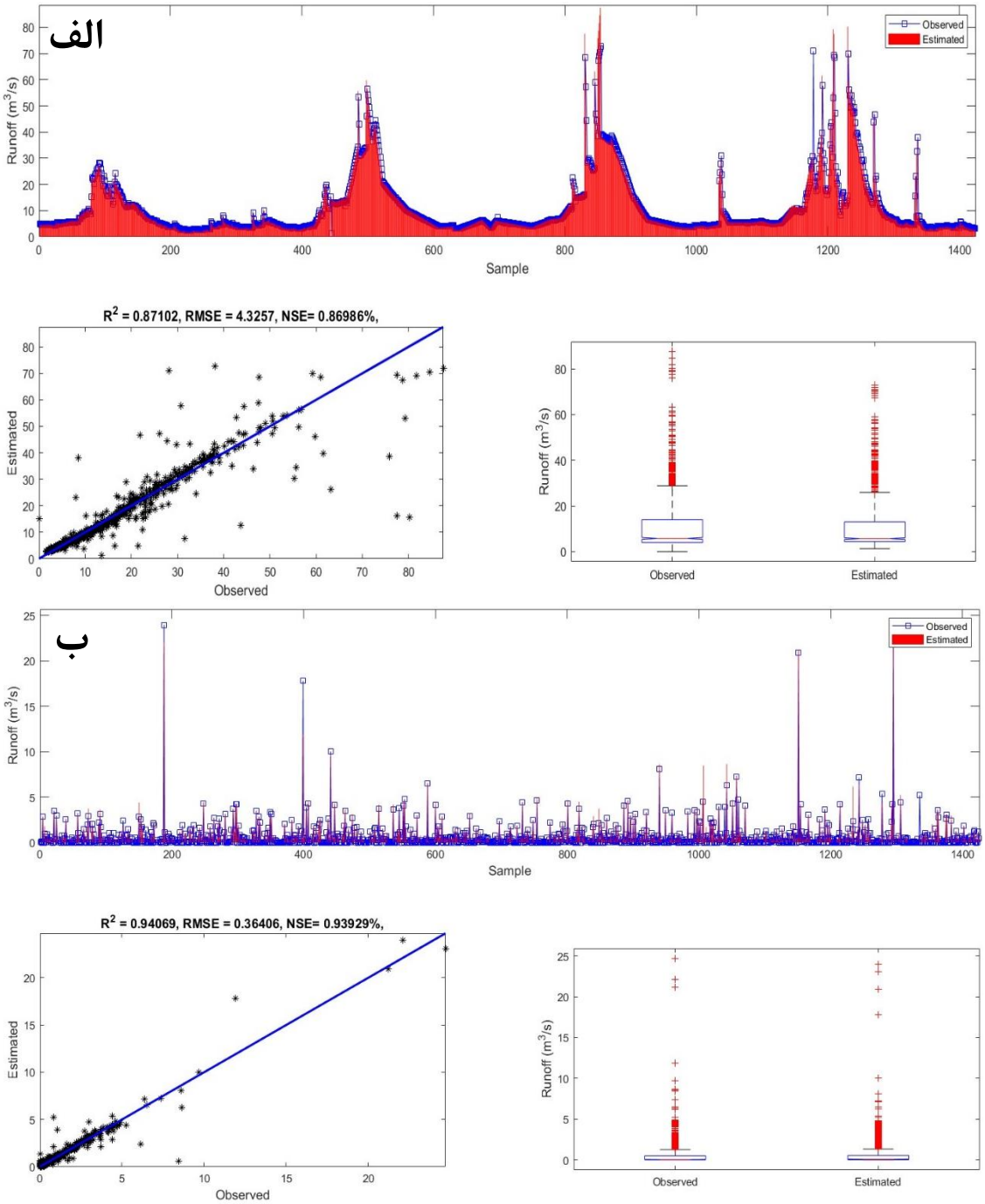
Coefficient of variation (ضریب تغییرات)	Maximum (بیشینه)	Minimum (کمینه)	Standard Deviation (انحراف معیار)	Median (میانه)	Mean (میانگین)	Variable (متغیر)	حوضه Watershed
0/26	951/86	431/55	165/22	569/64	623/83	باران Rainfall (mm)	طالقان Taleghan
0/05	14/18	11/69	0/64	12/29	12/49	دما Temperature (°C)	
0/09	2/04	1/48	0/18	1/84	1/83	سرعت باد Wind speed (m/s)	
0/05	56/66	45/66	2/80	51/93	51/43	رطوبت نسبی Relative humidity (%)	
0/20	15/79	7/35	2/32	11/25	11/29	دبی Discharge (m^3/s)	
0/34	479/24	103/24	103/43	308/64	302/34	باران Rainfall (mm)	تلوار Telvar
0/14	17/69	8/49	1/94	14/01	13/87	دما Temperature (°C)	
0/17	3/73	2/08	0/53	3/01	2/99	سرعت باد Wind speed (m/s)	
0/11	52/15	33/78	5/07	47/37	46/05	رطوبت نسبی Relative humidity (%)	
0/68	1/60	0/12	0/40	0/48	0/58	دبی Discharge (m^3/s)	



شکل ۲- نتایج مدل GMDH، برای مرحله آزمون، الف) آبخیز طالقان، ب) آبخیز تلوار

Fig 2. GMDH model results for the testing phase: (a) Taleghan watershed (b) Telvar watershed.

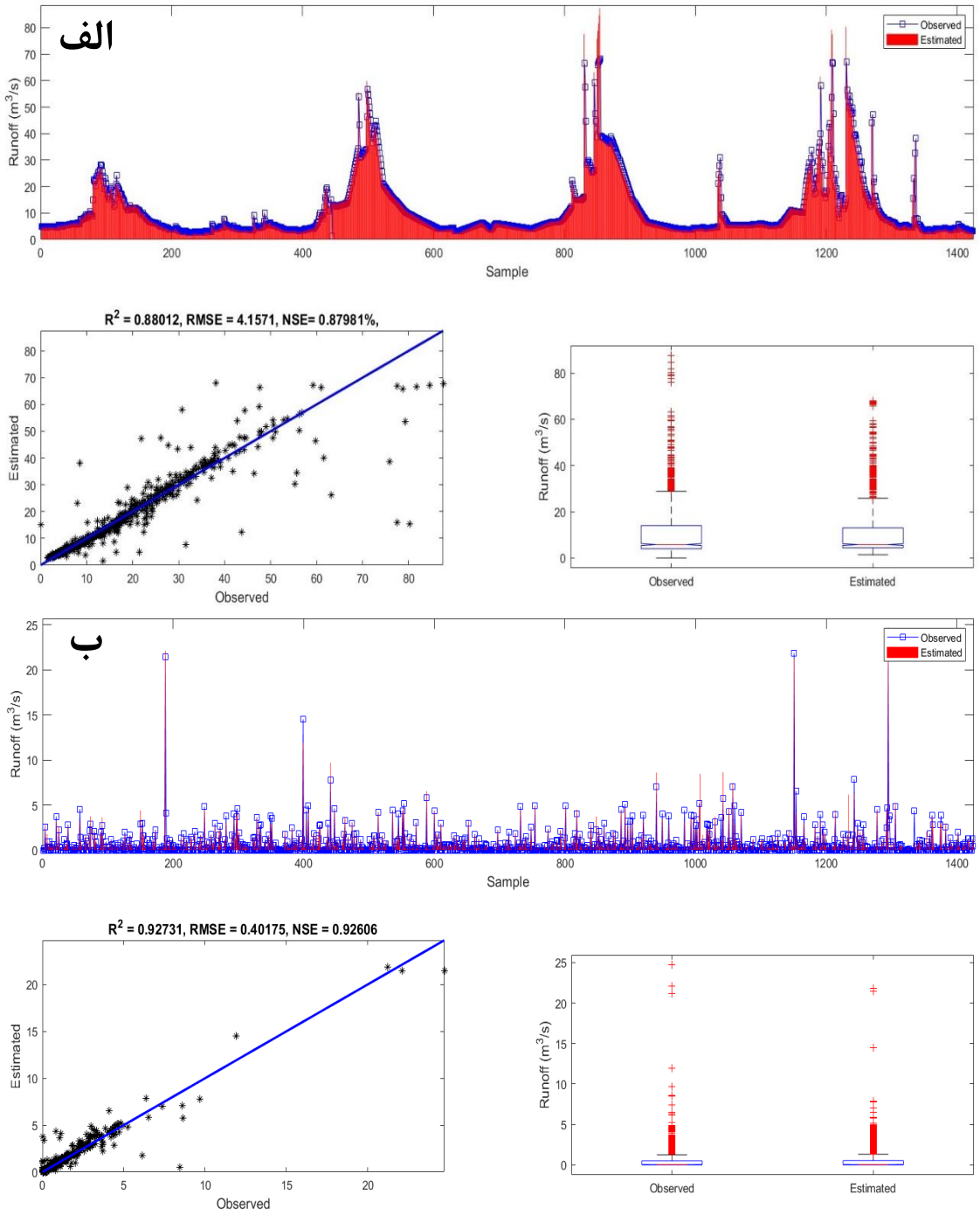
باتوجه به نمودارهای ارائه شده و نیز جدول ۲، نتایج این مطالعه نشان می‌دهد که عملکرد مدل‌های شبیه‌سازی رواناب به‌شدت به ویژگی‌های هیدرولوژیکی هر حوضه و ساختار داده‌های موجود وابسته است. مقایسه بین دو حوضه تلوار و طالقان به‌خوبی نشان می‌دهد که یک مدل نمی‌تواند در تمام شرایط برتر باشد و میزان دقت مدل‌ها به شرایط خاص حوضه بستگی دارد. باتوجه به شکل (۵) در حوضه تلوار، مدل‌های GMDH و ANFIS عملکرد بسیار نزدیکی داشته‌اند. مقدار R^2 برای GMDH برابر ۰/۹۴۰۹ و برای ANFIS برابر ۰/۹۴۰۶ به‌دست آمد که نشان‌دهنده توانایی تقریباً یکسان این دو مدل در تبیین واریانس داده‌هاست. همچنین مقادیر NSE به‌ترتیب ۰/۹۳۹۶ برای GMDH و ۰/۹۳۹۲ برای ANFIS گزارش شد که بیانگر قدرت بالای این دو مدل در بازتولید رفتار واقعی جریان رواناب است. در مقابل، مدل RF با مقادیر R^2 معادل ۰/۹۲۷۳ و NSE معادل ۰/۹۲۶۰ اندکی ضعیف‌تر عمل کرد. تفاوت



شکل۳- نتایج مدل ANFIS، برای مرحله آزمون، الف) آبخیز طالقان، ب) آبخیز تلوار

Fig 3. Results of the ANFIS model, for the test phase (a) Taleghan watershed (b) Talvar watershed

RMSEنیز این موضوع را تأیید می‌کند، به‌طوری‌که RMSE برای GMDH برابر ۰/۳۶۳۱، برای ANFIS برابر ۰/۳٦٤۰ و برایRF معادل ۰/٤۰۱۷ بود. این اختلاف نشان می‌دهد که دقت پیش‌بینی RF نسبت به دو مدل دیگر کاهش یافته و میزان خطای آن در شبیه‌سازی رواناب بیشتر است. برتری GMDH نسبت به ANFIS نیز هرچند بسیار جزئی است، اما نشان‌دهنده انعطاف‌پذیری بیشتر این مدل در استخراج روابط غیرخطی بین داده‌ها و پاسخ هیدرولوژیکی حوضه تلوار است. در حوضه طالقان، نتایج الگوی متفاوتی را نشان داد. در این حوضه مدل GMDH با R^2 معادل ۰/۸۸٤۵، NSE معادل ۰/۸۸۳٦ و RMSE معادل ٤/۰۹ بهترین عملکرد را داشته است. مدل RFبا مقادیر R^2 برابر ۰/۸۸۰۱، NSE معادل ۰/۸۷۹۸، وRMSE برابر ٤/۱٦عملکردی نزدیک بهGMDH نشان داد و فاصله چندانی



شکل٤- نتایج مدل RF برای مرحله آزمون، الف) آبخیز طالقان، ب) آبخیز تلوار

Fig4. RF model results for the test phase (a) Taleghan watershed (b) Talvar watershed

از آن نداشت. در حالی که مدل ANFIS با R^2 برابر ۰/۸۷۱۰، NSE برابر ۰/۸۶۹۸و RMSE معادل ٤/۳۲ نتایج ضعیف‌تری ارائه داد. به عبارت دیگر، در این حوضه مدل GMDH تنها با برتری اندکی نسبت به RF در جایگاه نخست قرار گرفت و این امر نشان می‌دهد که ساختار داده‌ها و ویژگی‌های هیدرولوژیکی حوضه طالقان موجب شده است که توانایی مدل RF در بازنمایی پویایی رواناب به GMDH نزدیک شود. ضعف ANFIS در این حوضه احتمالاً ناشی از حساسیت این مدل به انتخاب پارامترها و درجه پیچیدگی بالاتر آن است که در صورت وجود نویز یا تغییرات شدید در داده‌ها ممکن است منجر به کاهش دقت پیش‌بینی شود.

جدول2- نتایج مدل‌های مورد استفاده در مناطق مورد مطالعه

Table 2. Results of the models used in the studied regions

حوضه Watershed	Model مدل	Phase مرحله	Metric آماره	Value مقدار
طالقان Taleghan	GMDH	Training آموزش	R^2	0/930
			RMSE	3/291
			NSE	0/931
		Test آزمون	R2	0/884
			RMSE	4/089
			NSE	0/883
	ANFIS	Training آموزش	R2	0/922
			RMSE	3/430
			NSE	0/920
		Test آزمون	R2	0/871
			RMSE	4/325
			NSE	0/869
	RF	Training آموزش	R2	0/920
			RMSE	3/401
			NSE	0/923
		Test آزمون	R2	0/880
			RMSE	4/157
			NSE	0/879
تلوار Telvar	GMDH	Training آموزش	R^2	0/901
			RMSE	0/432
			NSE	0/891
		Test آزمون	R^2	0/940
			RMSE	0/363
			NSE	0/939
	ANFIS	Training آموزش	R^2	0/910
			RMSE	0/451
			NSE	0/912
		Test آزمون	R^2	0/940
			RMSE	0/364
			NSE	0/939
	RF	Training آموزش	R^2	0/891
			RMSE	0/482
			NSE	0/904
		Test آزمون	R^2	0/927
			RMSE	0/401
			NSE	0/926

ارزیابی عملکرد مدل‌های پیش‌بینی جریان رودخانه‌ای در دو حوضه طالقان و تلوار نشان‌دهنده تفاوت‌هایی در قابلیت‌های مدل‌های GMDH، ANFIS و RF است. تمرکز اصلی بر مرحله آزمون قرار گرفته تا اعتبار مدل‌ها در شرایط واقعی و خارج از داده‌های آموزشی بررسی شود. این رویکرد اجازه می‌دهد تا از بیش‌برازش جلوگیری شود و قابلیت تعمیم‌پذیری مدل‌ها ارزیابی شود. به طور کلی، نتایج آزمون حاکی از آن است که مدل GMDH در هر دو حوضه عملکرد برتر را از خود نشان داده، در حالی که ANFISو RF بسته به ویژگی‌های حوضه، جایگاه‌های متفاوتی داشته‌اند. این یافته‌ها می‌توانند راهنمایی برای انتخاب مدل‌های مناسب در مدیریت منابع آب و پیش‌بینی هیدرولوژیکی باشند، به ویژه در مناطق کوهستانی ایران که با تغییرات اقلیمی و نوسانات بارش مواجه هستند.

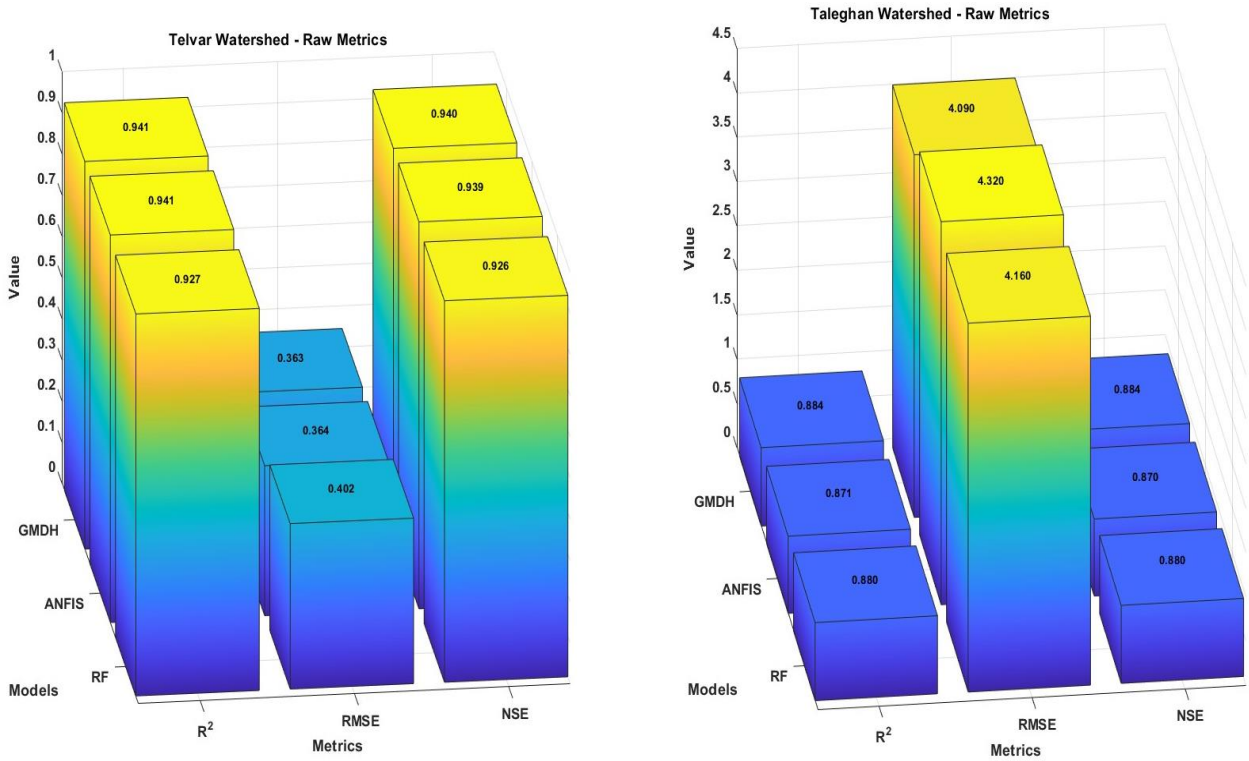
در حوضه طالقان، که با ویژگی‌های توپوگرافی پیچیده و داده‌های پراکنده همراه است، ترتیب عملکرد مدل‌ها بر اساس معیارهای آزمون به این شرح است که GMDH برتر از RF و سپس ANFIS. مدل GMDH با دقت بالاتر و خطای کمتر، توانسته الگوهای غیرخطی جریان را بهتر مدل‌سازی کند، که این امر احتمالاً به دلیل ساختار خودسازمان‌دهنده آن است که اجازه می‌دهد روابط پیچیده بین متغیرهای ورودی مانند بارش، تبخیر و دما را بدون نیاز به فرضیات پیشین استخراج کند. در مقابل، RF که بر پایه درخت‌های تصمیم‌گیری عمل می‌کند، عملکردی نزدیک به GMDH داشته اما کمی ضعیف‌تر ظاهر شده، شاید به خاطر حساسیت بیشتر به نویز داده‌ها در این حوضه. ANFIS، به عنوان یک سیستم فازی-عصبی، کمترین عملکرد را نشان داده که می‌تواند ناشی از نیاز به تنظیم دقیق پارامترهای عضویت فازی باشد، که در داده‌های محدود آزمون، منجر به تعمیم‌پذیری پایین‌تر شده است. این ترتیب نشان می‌دهد که در حوضه‌هایی با داده‌های ناهمگن، مدل‌های داده‌محور مانند GMDHمی‌توانند گزینه‌ای مطمئن‌تر باشند، زیرا انعطاف‌پذیری بیشتری در برابر تغییرات محیطی ارائه می‌دهند.

در حوضه تلوار، که ممکن است با الگوهای جریان پایدارتر و داده‌های همگن‌تر همراه باشد، ترتیب عملکرد مدل‌ها متفاوت است: GMDH همچنان برتر، اما ANFIS بسیار نزدیک به آن عمل کرده و RF در جایگاه سوم قرار گرفته. GMDH با حاشیه‌ای اندک بر ANFIS غلبه کرده، که این برتری احتمالاً به دلیل توانایی بهتر در مدیریت روابط چندمتغیره است. ANFIS در این حوضه عملکردی چشمگیر داشته، که می‌تواند به دلیل ترکیب منطق فازی با شبکه‌های عصبی باشد، که اجازه می‌دهد عدم قطعیت‌های ذاتی داده‌های هیدرولوژیکی را بهتر مدیریت کند. این مدل در مقایسه با حوضه طالقان، بهبود قابل‌توجهی نشان داده، که حاکی از وابستگی عملکرد به ویژگی‌های محلی حوضه است. RF، هرچند قابل‌قبول، ضعیف‌ترین مدل بوده که ممکن است به دلیل تمرکز بیش از حد بر میانگین‌گیری درخت‌ها باشد، که در داده‌های آزمون با واریانس پایین، منجر به دقت کمتر شده است. این یافته‌ها تأکید می‌کنند که انتخاب مدل باید بر اساس ویژگی‌های خاص هر حوضه صورت گیرد؛ برای مثال، در حوضه‌هایی با داده‌های نسبتاً یکنواخت مانند تلوار، ANFIS می‌تواند جایگزین مناسبی برای GMDH باشد.

به طور کلی، برتری GMDH در هر دو حوضه نشان‌دهنده پتانسیل بالای مدل‌های خودسازمان‌دهنده در پیش‌بینی جریان رودخانه‌ای است. این مدل توانسته با کمترین خطا، الگوهای فصلی و غیرفصلی را پیش‌بینی کند، که برای برنامه‌ریزی سدسازی و مدیریت سیلاب حیاتی است. با این حال، عملکرد نزدیک ANFIS در تلوار پیشنهاد

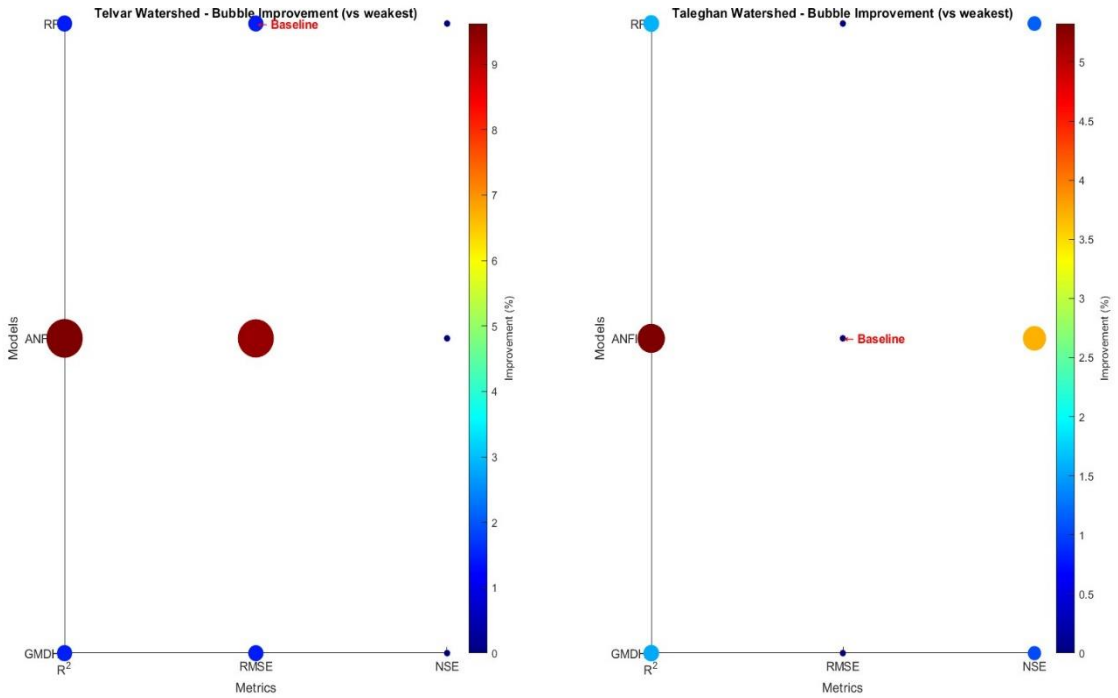
می‌کند که ترکیبی از مدل‌ها مانند هیبریدیGMDH-ANFIS می‌تواند دقت را افزایش دهد.RF، هرچند انعطاف‌پذیر، نیاز به داده‌های بیشتر برای آموزش دارد تا در آزمون‌ها بهتر عمل کند. محدودیت‌های مطالعه شامل عدم در نظر گرفتن متغیرهای خارجی مانند تغییرات کاربری زمین است، که می‌تواند در تحقیقات آتی بررسی شود. مقدار RMSE در حوضه تلوار به‌مراتب کوچک‌تر از طالقان است، زیرا مقیاس دبی جریان در تلوار بسیار پایین‌تر است. میانگین دبی حوضه تلوار در حد چند مترمکعب بر ثانیه است، در حالی که طالقان، دبی‌های بالاتری (ده‌ها مترمکعب بر ثانیه) دارد. بنابراین، خطای مطلق (RMSE) به‌طور طبیعی در مقیاس کوچک‌تر تلوار کاهش می‌یابد، حتی اگر خطای نسبی مشابه باشد.

شکل (٦) با استفاده از نمودار حبابی، میزان بهبود نسبی بهترین مدل هر حوضه نسبت به ضعیف‌ترین مدل را بر اساس سه معیار آماری R^2، RMSE و NSE نشان می‌دهد. در حوضه تلوار، مدل GMDHبا بزرگ‌ترین و قرمزترین حباب‌ها نمایش داده شده است، زیرا نسبت به RF (به‌عنوان ضعیف‌ترین مدل) بهبود داشته است؛ به‌طور خاص، مقدار R^2 حدود ١/٣٦ درصد، مقدار NSE حدود ١/٤٦ درصد افزایش یافته و RMSE نیز نزدیک به ٩/٦ درصد کاهش یافته است. این مقادیر بیانگر برتری GMDH در این حوضه هستند. در مقابل،ANFIS تنها اندکی ضعیف‌تر از GMDH بوده و عملکردی تقریباً هم‌سطح ارائه داده است. در حوضه طالقان نیز GMDH به‌عنوان بهترین مدل شناسایی شد، اما فاصله آن با RF نسبت به حوضه تلوار بسیار کمتر بود. بهبود R^2 تنها حدود ٠/٥درصد، بهبود NSE در حدود ٠/٤٣ درصد و کاهشRMSE نزدیک به ١/٩ درصد بوده است. این اختلاف اندک به‌خوبی نشان می‌دهد که در شرایط داده‌ای و هیدرولوژیکی این حوضه،RF توانسته است عملکردی تقریباً مشابه با GMDH ارائه دهد. از طرف دیگر، ANFIS در مقایسه با GMDH حدود ١/٣٥ درصد کاهش در R^2، حدود ١/٣٨ درصد کاهش در NSE و نزدیک به ٥/٣ درصد افزایش در RMSE داشته است که نشان‌دهنده ضعف نسبی آن در این حوضه است. این یافته‌ها بیانگر آن هستند که هیچ مدلی به‌صورت مطلق برتر نیست و انتخاب مدل به ماهیت داده‌ها و شرایط هیدرولوژیکی بستگی دارد. در مجموع، مدل GMDH در هر دو حوضه بهترین عملکرد را از خود نشان داده است، اما میزان برتری آن در حوضه تلوار چشم‌گیرتر از طالقان بوده است. علت اصلی این موضوع می‌تواند به ویژگی‌های داده‌ای نظیر میزان همبستگی ورودی‌ها، پایداری الگوهای بارش-رواناب و سطح پیچیدگی پویا هیدرولوژیکی مربوط باشد. به‌طور کلی، نتایج این مطالعه نشان می‌دهد که استفاده از مدل‌های یادگیری ماشین و مقایسه میان آن‌ها می‌تواند نقش کلیدی در بهبود دقت پیش‌بینی رواناب داشته باشد. چنین مقایسه‌هایی نه‌تنها به انتخاب مدل مناسب برای هر حوضه کمک می‌کند، بلکه می‌تواند در مدیریت و برنامه‌ریزی منابع آب و کاهش خطر ناشی از عدم قطعیت‌های هیدرولوژیکی تأثیرگذار باشد.



شکل ٥- نتایج مدل‌های GMDH، ANFIS و RF در دو حوضه مورد مطالعه

Fig 5. Results of GMDH, ANFIS and RF models in the two watersheds



شکل ٦- میزان بهبود نسبی بهترین مدل نسبت به سایر مدل‌ها

Fig 6. Relative improvement of the best model compared to other models

بحث و نتیجه‌گیری

نتایج این پژوهش نشان داد که عملکرد مدل‌های داده‌محور در شبیه‌سازی رواناب، تحت تأثیر ویژگی‌های هیدرولوژیکی حوضه و ساختار داده‌های ورودی قرار دارد. اگرچه مدل GMDH در هر دو حوضه مطالعه‌شده توانست نسبت به سایر مدل‌ها نتایج بهتری ارائه دهد، اما میزان این برتری در حوضه‌ها یکسان نبود. در حوضه‌ای با داده‌های منظم‌تر و همبستگی بالاتر، این مدل توانست به‌طور محسوسی از رقبا پیشی بگیرد، در حالی که در حوضه‌ای با پویایی پیچیده‌تر، اختلاف عملکرد آن با مدل RF بسیار کمتر بود. این یافته بیانگر آن است که نمی‌توان یک مدل را به عنوان بهترین گزینه در همه شرایط معرفی کرد و انتخاب مدل باید به صورت محلی و مبتنی بر شرایط خاص هر حوضه صورت گیرد. مقایسه بین مدل‌ها نشان داد که الگوریتم‌های مبتنی بر شبکه‌های خودسازمان‌ده مانند GMDH در شرایطی که داده‌ها الگوهای نسبتاً پایداری دارند، قادر به استخراج روابط غیرخطی پیچیده بوده و عملکردی قابل اعتماد ارائه می‌دهند. در مقابل، مدل RF به دلیل ساختار درختی و توانایی کاهش واریانس از طریق تجمیع، در شرایطی که داده‌ها با نویز یا تغییرات بیشتری همراه هستند، می‌تواند دقتی نزدیک به بهترین مدل‌ها داشته باشد. این موضوع در مطالعات مشابه نیز گزارش شده است؛ برای مثال، مطابق نتایج Shortridge و همکاران [٢٤]، مدل‌های مبتنی بر جنگل تصادفی در پیش‌بینی رواناب و دبی سیلاب در حوضه‌های پیچیده عملکردی پایدار و رقابتی داشته‌اند.

همچنین، مرور انجام‌شده توسط Mosavi و همکاران[٢١] نشان می‌دهد که مدل‌های داده‌محور مختلف بسته به اقلیم، مقیاس مکانی و کیفیت داده‌ها عملکرد متفاوتی دارند و هیچ مدلی به‌صورت مطلق برتر نیست. از سوی دیگر، مدل ANFIS در این مطالعه عملکرد ضعیف‌تری نسبت به دو مدل دیگر در یکی از حوضه‌ها داشت. علت احتمالی این امر حساسیت بالای این مدل به انتخاب تعداد توابع عضویت و ساختار قواعد فازی است که در صورت وجود نویز یا تغییرپذیری شدید داده‌ها می‌تواند دقت مدل را کاهش دهد. این نتیجه با یافته‌های Karimi و همکاران[١١]همسو است که نشان داده‌اند استفاده از ANFIS بدون تنظیم دقیق پارامترها می‌تواند منجر به کاهش عملکرد در شبیه‌سازی رواناب شود. بنابراین، گرچه ANFIS در شرایطی با داده‌های منظم‌تر قادر به ارائه نتایج دقیق است، اما در محیط‌های داده‌ای پیچیده‌تر ممکن است نیازمند اصلاحات یا ترکیب با سایر روش‌ها باشد .به طور کلی، نتایج این پژوهش هم‌راستا با مطالعات پیشین تأکید می‌کند که انتخاب مدل مناسب باید مبتنی بر آزمون‌های تجربی در همان حوضه انجام شود، نه صرفاً بر اساس نتایج مطالعات دیگر. از این منظر، پژوهش Emami و Koch [٧] نیز اشاره می‌کند که حتی مدل‌های مشابه در دو حوضه با ویژگی‌های متفاوت، عملکردی یکسان نخواهند داشت و باید با رویکرد مقایسه‌ای مورد ارزیابی قرار گیرند. کاربردهای عملی این یافته‌ها برای مدیریت منابع آب بسیار مهم است. داشتن مدلی دقیق و سازگار با شرایط هر حوضه می‌تواند در طراحی سامانه‌های هشدار سیلاب،

مدیریت مخازن سدها و تخصیص بهینه منابع آب نقشی کلیدی ایفا کند. همچنین، نتایج نشان می‌دهد که توسعه مدل‌های هیبریدی و ترکیبی، که بتوانند نقاط قوت روش‌های مختلف را هم‌زمان به‌کار گیرند، می‌تواند دقت پیش‌بینی رواناب را بیش از پیش افزایش دهد. در مجموع، این پژوهش نشان می‌دهد که مدل‌های یادگیری ماشین ابزارهای ارزشمندی برای شبیه‌سازی رواناب هستند، اما انتخاب مدل مناسب به شرایط داده‌ای و هیدرولوژیکی وابسته است. مقایسه و ترکیب مدل‌ها نه‌تنها می‌تواند در بهبود دقت پیش‌بینی مؤثر باشد، بلکه در کاهش خطر ناشی از عدم قطعیت‌های هیدرولوژیکی و ارتقای برنامه‌ریزی منابع آب نیز اهمیت ویژه‌ای دارد. مطالعه حاضر نشان داد که مدل‌های مبتنی بر یادگیری ماشین، شامل GMDH، ANFIS و جنگل تصادفی، توانایی مناسبی در پیش‌بینی رواناب روزانه دارند و انتخاب بهترین مدل به ویژگی‌های هیدرولوژیکی هر حوزه آبخیز و کیفیت داده‌ها بستگی دارد. مدلGMDH عملکرد بالاتری در هر دو حوزه مورد مطالعه از خود نشان داد، اما در حوزه‌های با پیچیدگی هیدرولوژیکی بالاتر، جنگل تصادفی نیز دقت قابل توجهی ارائه داد. از منظر کاربرد عملی، این مدل‌ها می‌توانند به طور مؤثر در مدیریت منابع آب ایران استفاده شوند. بهبود پیش‌بینی سیلاب‌های روزانه و کوتاه‌مدت در مناطق کوهستانی که داده‌های هیدرومتری محدود دارند، بهینه‌سازی عملکرد سدها و مخازن با استفاده از پیش‌بینی دقیق دبی ورودی، توسعه سیستم‌های هشدار سریع برای کاهش خسارات ناشی از رواناب شدید، کمک به برنامه‌ریزی تخصیص بهینه آب برای کشاورزی و شرب با پیش‌بینی دقیق تغییرات روزانه جریان‌ها. با این حال، این مطالعه دارای محدودیت‌هایی نیز است که باید در کاربردهای عملی مدنظر قرار گیرد: عدم اعتبارسنجی خارجی با داده‌های دیگر حوزه‌های آبخیز، که ممکن است دقت مدل در شرایط جدید متفاوت باشد، حساسیت مدل‌ها به کیفیت و پوشش داده‌های ورودی، به‌ویژه برای ANFIS که نیازمند داده‌های کم‌نویز و پارامترهای بهینه اولیه است. محدودیت زمانی و مکانی داده‌های مورد استفاده که ممکن است برخی رفتارهای هیدرولوژیکی نادر یا شدید را به‌طور کامل منعکس نکند. در مجموع، استفاده هوشمندانه از مدل‌های یادگیری ماشین می‌تواند به افزایش دقت پیش‌بینی رواناب و کاهش عدم قطعیت‌های هیدرولوژیکی در ایران کمک کند، اما توصیه می‌شود قبل از به‌کارگیری عملی، اعتبارسنجی مدل در حوزه‌های جدید و با داده‌های بلندمدت و متنوع انجام شود.

سپاسگزاری

این پژوهش بر مبنای پایان‌نامه کارشناسی ارشد نویسنده اول و با حمایت مالی دانشگاه تربیت مدرس انجام گرفته است. نویسندگان بر خود لازم می‌دانند از دانشکده منابع طبیعی و علوم دریایی پردیس نور دانشگاه تربیت مدرس به‌سبب پشتیبانی‌های علمی و فراهم‌سازی بسترهای لازم جهت پیشبرد این تحقیق صمیمانه قدردانی نمایند.

تضاد منافع نویسندگان

بدین‌وسیله نویسندگان تصریح می‌نمایند که در زمینه انجام پژوهش حاضر و انتشار نتایج آن، هیچ‌گونه تضاد منافع احتمالی وجود نداشته است.

دسترسی به داده‌ها

اطلاعات و داده‌های مرتبط با این پژوهش، در صورت ارائه درخواست منطقی، از طریق نویسنده مسئول در اختیار پژوهشگران علاقه‌مند قرار خواهد گرفت.

مشارکت نویسندگان

نگین رشیدی: جمع‌آوری داده‌ها، پردازش داده‌ها، انجام محاسبات،نگارش، تدوین و تنظیم مقاله. وحید موسوی: راهنمایی و نظارت بر روند اجرای پژوهش، کنترل نتایج،ویرایش و بازبینی مقاله. مهدی وفاخواه:راهنمایی بر روند پژوهش،ویرایش و بازبینی مقاله

منابع مورد استفاده

1.Abrahart, R. J., & See, L. (2000). Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrologicalprocesses,14(1112),21572172.https://doi:10.1002/10991085(20000815/30)14:11/123.0.CO;2-S.

2.Agarwal, A. (2008). Low Power Neural Network Training Using A GMDH Type Algorithm. Journal of Computer Science, Informatics& Electrical engineering, https://doi:10.5391/IJFIS.2005.5.1.052.

3.Ahli, H., Merabtene, T., & Seddique, M. (2021, 7-10 Dec. 2021). Optimization of a Conceptual Rainfall-Runoff Model using Evolutionary Computing methods. 2021 14th International Conference on Developments in eSystems Engineering (DeSE). https://doi:10.1109/DeSE54285.2021.9719369.

4.Ahmadi F (۲۰۱۹) Evaluation of the performance of support vector machine methods and adaptive neural-fuzzy inference system in predicting the monthly flow of rivers Case study: Nazlo and Caesar rivers. Research on Water and Soil Resources of Iran (Iranian Sciences and Agriculture) 51(3):673-https://doi:10.22059/ijswr.2019.290994.668356. (In Persian)

5.Besaw, L. E., Rizzo, D. M., Bierman, P. R., & Hackett, W. R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. Journal of Hydrology, 386(1), 27-37. https://doi:10.1016/j.jhydrol.2010.02.037.

6.Breiman, L. (1996). Bagging predictors Machine Learning 24 (2), 123-140 (1996) 10.1023. A*: 1018054314350*. https://doi.

16.Moosavi, V., Gheisoori Fard, Z., & Vafakhah, M. (2022). Which one is more important in daily runoff forecasting using data driven models: Input data, model type, preprocessing or data length? Journal of Hydrology, 606, 127429. https://doi.org/10.1016/j.jhydrol.2022.127429.

17.Moosavi, V., Talebi, A., & Hadian, M. R. (2017). Development of a Hybrid Wavelet Packet- Group Method of Data Handling (WPGMDH) Model for Runoff Forecasting. Water Resources Management, 31(1), 43-59. https://doi.org/10.1007/s11269-016-1507-3.

18.Naserabadi, F., Rezavi, R., & Zakerinia, M. (2020). Hydrological modeling of the Telvar watershed under climate change impacts. *Iranian Water Research Journal*, 16(37), 37–49. https://civilica.com/doc/1330618/.(In Persian).

19.Pour, S. H., Shahid, S., & Sammen, S. S. (2023). Runoff modeling using group method of data handling and gene expression programming. In Handbook of Hydroinformatics (pp. 353-377). Elsevier. https://doi.org/10.1016/B978-0-12-821962-1.00009-X.

20.Pournemat Roodsari, A., Ghaderi, K., & Karimi Goughari, Sh. (2014). Rainfall–runoff process modeling using Group Method of Data Handling (GMDH) and Artificial Neural Network (ANN) in the Polrud watershed. *Watershed Management Research Journal*, 5(10). (In Persian). http://jwmr.sanru.ac.ir/article-1-414-fa.html

21. Qiao, X., Peng, T., Sun, N., Zhang, C., Liu, Q., Zhang, Y., ... & Nazir, M. S. (2023). Metaheuristic evolutionary deep learning model based on temporal convolutional network, improved aquila optimizer and random forest for rainfall-runoff simulation and multi-step runoff prediction. Expert Systems with Applications, 229,120616.https://doi.org/10.1016/j.eswa.2023.120616.

22.Remesan, R., Shamim, M. A., Han, D., & Mathew, J. (2008, October). ANFIS and NNARX based rainfall-runoff modeling. In 2008 IEEE International Conference on Systems, Man and Cybernetics (pp.14541459).IEEE.https://doi.org/10.1109/ICSMC.2008.4811490.

23.Seo, Y., S. Kim. V.P. Singh. 2018. Machine learning models coupled with variational mode decomposition: A new approach for modeling daily rainfall-runoff. Atmosphere 9(7): 251. https://doi.org/10.3390/atmos9070251.

24.Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences, 20*(7), 2611–2628. https://doi.org/10.5194/hess-20-2611-2016. 25.Siasar, org/10.1007/BF00058655.

7.Emami, F., & Koch, M. (2018). Modeling rainfall–runoff process using data-driven models vs. conceptual hydrological model. *Physics and Chemistry of the Earth, Parts A/B/C, 105*, 212–221. https://doi.org/10.1016/j.pce.2018.03.012.

8.Esmaeili Gisavandani, H., Akhoundali, A. M., Zarei, H., & Taghian, M. (2017). Evaluation of the capability of Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), and Regression models in regional flood analysis. *Journal of Water and Soil Conservation Research*, 24(3) https://doi.org/10.22069/jwfst.2017.11413.2581.(In Persian).

9.Gaertner, B. (2024). Geospatial patterns in runoff projections using random forest based forecasting of time-series data for the mid-Atlantic region of the United States. Science of the Total Environment, 912,169211.https://doi.org/10.1016/j.scitotenv.2023.169211.

10.Karbasii, M. (2016). Weekly reference evapotranspiration prediction using a hybrid wavelet–adaptive neuro-fuzzy inference system (Wavelet–ANFIS) model. *Journal of Water Research in Agriculture*,30(1),7387.https://doi.org/10.22092/jwra.2016.106203. (In Persian).

11.Karimi, M., Kisi, O., Shiri, J., & Zounemat-Kermani, M. (2022). Comparative assessment of ANFIS, ANN, and regression-based models for streamflow forecasting. *Water Resources Management, 36*(7), 2341–2356. https://doi.org/10.1007/s11269-022-03075-1.

12.Kia, A., Emadi, A., & Gholami Sefidkouhi, M. A. (2019). Rainfall–runoff modeling using Adaptive Neuro-Fuzzy Inference System (ANFIS) and multiple linear regression (MLR). *Scientific-Research Journal of Irrigation and Water Engineering in Iran*, 9(3). https://sid.ir/paper/۲۴۷۲۵۰/fa. (In Persian).

13.Liu, X., Zhang, X., Kong, X., & Shen, Y. J. (2022). Random forest model has the potential for runoff simulation and attribution. *Water*, *14*(13),2053.https://doi.org/10.3390/w14132053.

14.Miao T.Y., Wang, M. (2015). Susceptibility Analysis of Earthquake-Induced Landslide Using Random Forest Method, International Conference on Computer Information Systems and Industrial Applications (CISIA), 771-775. https://doi.org/10.2991/cisia-15.2015.209.

15.Mohammadi, B. (2021). A review on the applications of machine learning for runoff model. Sustainable Water Resources Management, 7(6), 98. https://doi.org/10.1007/s40899-021-00584-y.

H., & Honar, T. (2019). Application of Support Vector Machine (SVM), CHAID, and Random Forest (RF) models in estimating daily reference evapotranspiration in the north of Sistan and Baluchestan Province. *Iranian Journal of Irrigation and Drainage*, 13(2), 378–388. (In Persian). https://sid.ir/paper/۱۳۱۷۵۳/fa.

26.Solomatine, D.P., A. Ostfeld. 2008. Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics10(1):322.https://doi.org/10.2166/hydro.2008.015.

27.Vilaseca, F., Castro, A., Chreties, C., & Gorgoglione, A. (2023). Assessing influential rainfall–runoff variables to simulate daily streamflow using random forest. Hydrological Sciences Journal, 68(12),17381753.https://doi.org/10.1080/02626667.2023.2232356.

28.Yerel, S., & Ankara, H. (2012). Application of multivariate statistical techniques in the assessment of water quality in Sakarya River, Turkey. Journal of the Geological Society of India, 79(1), 89-93. https://doi.org/10.1007/s12594-011-0121-5.