

Technical Note

Prediction of Monthly Streamflow Using Data-Driven Models

A. Ahani¹ and M. Shourian²

Abstract

In recent years, data-driven modeling techniques have gained numerous applications in hydrology and water resources studies. River runoff estimation and forecasting is one of the research fields in which these techniques have several applications. In the current study, four data-driven modeling techniques of multiple linear regression, K-nearest neighbors, artificial neural networks, and adaptive neuro-fuzzy inference systems have been used to form runoff forecasting models and then their results have been evaluated. Also, effects of using some different scenarios to select predictor variables have been studied. It was evident from the results that using flow data related to one or two months ago in the predictor variables dataset can improve the accuracy of the results. In addition, comparison of general performances of the modeling techniques showed superiority of KNN models results among the studied models. The selected KNN model presented best performance with a linear correlation coefficient equal to 0.84 between observed flow data and predicted values and a RMSE equal to 2.64.

Keywords: Streamflow Forecasting, Data-Driven Modeling, MLR, KNN, ANN, ANFIS.

Received: March 11, 2016

Accepted: July 28, 2016

یادداشت فنی

پیش‌بینی جریان ماهیانه رودخانه با استفاده از مدل‌های داده‌مبنا

علی آهنی^۱ و مجتبی شوریان^{۲*}

چکیده

در سال‌های اخیر، تکنیک‌های مدل‌سازی داده‌مبنا کاربردهای فراوانی در مطالعات هیدرولوژی و مهندسی منابع آب یافته‌اند. توسعه مدل‌های برآورد یا پیش‌بینی رواناب رودخانه، یکی از زمینه‌های مطالعاتی است که این تکنیک‌ها در آن کاربرد زیادی دارند. در مطالعه حاضر، چهار تکنیک مدل‌سازی داده‌مبنا، شامل رگرسیون خطی چندگانه، K نزدیک‌ترین همسایه، شبکه‌های عصبی مصنوعی و سیستم‌های استنتاج عصبی - فازی تطبیقی به‌منظور تشکیل مدل‌های پیش‌بینی رواناب مورد استفاده قرار گرفته و نتایج حاصل بررسی شده است. همچنین تأثیر انتخاب چند سناریوی مختلف در انتخاب متغیرهای ورودی ارزیابی شده است. نتایج به دست آمده حاکی از آن است که استفاده از داده‌های جریان ماه‌های قبل در مجموعه داده‌های مورد استفاده جهت پیش‌بینی می‌تواند سبب بهبود دقت نتایج مدل‌ها شود. به‌علاوه، مقایسه عملکرد کلی تکنیک‌های مدل‌سازی، بیانگر برتری نتایج حاصل از به‌کارگیری تکنیک KNN نسبت به سایر تکنیک‌ها می‌باشد. در میان مدل‌های برگزیده تکنیک‌های مختلف نیز، مدل برگزیده KNN برای حالت استفاده از داده‌های جریان با ضریب همبستگی خطی ۰/۸۴ بین داده‌های مشاهداتی جریان و پیش‌بینی‌های مدل و مقدار شاخص خطای RMSE برابر ۲/۶۴ بهترین عملکرد را به نمایش گذاشت.

کلمات کلیدی: پیش‌بینی جریان رودخانه، مدل‌سازی داده‌مبنا، رگرسیون خطی چندگانه، K نزدیک‌ترین همسایه، شبکه عصبی مصنوعی، سیستم استنتاج عصبی - فازی.

تاریخ دریافت مقاله: ۹۴/۱۲/۲۱

تاریخ پذیرش مقاله: ۹۵/۵/۷

1- PhD Student, Department of Civil, Faculty of Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran.

2-Assistant Professor, Department of Civil, Faculty of Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran, Email: m_shourian@sbu.ac.ir.

*- Corresponding Author

۱- دانشجوی دکتری، دانشکده مهندسی عمران، آب و محیط زیست، دانشگاه شهید بهشتی

۲- استادیار دانشکده مهندسی عمران، آب و محیط زیست، دانشگاه شهید بهشتی

*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان پاییز ۱۳۹۶ امکانپذیر است.

۱- مقدمه

سپس داده‌های مربوط به ۱۰ سال پایانی برای ارزیابی کیفیت پیش‌بینی‌های مدل‌های ایجاد شده، مورد استفاده قرار گرفته‌اند.

۲-۲- انتخاب متغیرها، تکنیک‌های مدل‌سازی و مشخصات آن‌ها

در مدل‌سازی‌های مطالعه‌ی حاضر، جریان متوسط ماهانه به‌عنوان متغیر وابسته در نظر گرفته شد و بر این اساس چندین سناریو برای تشکیل مدل‌های پیش‌بینی رواناب اتخاذ شد. در سناریوی نخست از سری زمانی جریان‌های متوسط ماهانه به‌عنوان متغیر وابسته و از سری‌های زمانی دمای متوسط ماهانه و بارندگی متوسط ماهانه به‌عنوان متغیرهای ورودی برای تشکیل یک مدل کلی بر اساس تمام داده‌ها استفاده شد. در سناریوی دوم، از داده‌های جریان، دما و بارندگی متوسط ماهانه هر ماه تنها برای تشکیل مدل پیش‌بینی رواناب در همان ماه استفاده شد. در نتیجه در این حالت، با استفاده از هر روش مدل‌سازی، دوازده مدل برای دوازده ماه تشکیل شد. در سناریوی سوم، همانند سناریوی دوم برای پیش‌بینی جریان متوسط ماهانه هر یک از دوازده ماه یک مدل تشکیل شد، با این تفاوت که دما و بارندگی متوسط ماهانه یک ماه قبل نیز در کنار دما و بارندگی متوسط آن ماه به‌عنوان متغیرهای ورودی قرار گرفتند. در سناریوی چهارم، داده‌های دما و بارندگی متوسط دو ماه قبل و یک ماه قبل نیز در کنار دما و بارندگی متوسط هر ماه به‌عنوان متغیرهای ورودی به‌کار گرفته شدند. در سناریوی پنجم مشابه سناریوی سوم عمل شد، با این تفاوت که داده‌های جریان متوسط ماهانه یک ماه قبل نیز به متغیرهای ورودی مدلهای افزوده شدند. در سناریوی ششم، داده‌های جریان متوسط ماهانه یک و دو ماه قبل نیز به متغیرهای ورودی مدل‌ها در سناریوی چهارم افزوده شدند. شایان ذکر است که برای کاهش اثر اختلاف ابعاد و واریانس داده‌های مربوط به متغیرهای مختلف، از روش استانداردسازی استفاده شد.

در این مطالعه، برای یافتن تعداد بهینه K در مدل KNN، مقدار آن از ۱ تا ۲۵ تغییر کرده است. همچنین از یک شبکه‌ی عصبی پرسپترون چند لایه^۶ MLP با یک لایه پنهان برای مدل‌سازی استفاده شده است که تعداد نورون‌های لایه پنهان برای یافتن عملکرد بهینه شبکه از ۱ تا ۴۰ تغییر داده شده است. الگوریتم لونیگ - مارکورات برای آموزش شبکه به‌کار گرفته شده است. به علاوه در مورد مدل ANFIS، از تعداد دو تابع عضویت هم‌نوع برای هر ورودی استفاده شد.

مدل‌های پیش‌بینی رواناب مجموعه‌ای از روش‌های بسیار پرکاربرد در زمینه هیدرولوژی و مدیریت منابع آب هستند. یکی از انواع مشهور این مدل‌ها، مدل‌های بارندگی - رواناب هستند که در آن‌ها با استفاده از روش‌های مختلف، پیش‌بینی یا برآورد مقدار رواناب بر اساس داده‌های بارندگی صورت می‌گیرد.

به‌کارگیری روش‌های مدل‌سازی داده مینا^۱ (DDM) بخش قابل توجهی از فعالیت‌ها و پژوهش‌های انجام شده در زمینه مدل‌های پیش‌بینی رواناب را به خود اختصاص داده است. عملکرد این مدل‌ها مبتنی بر یافتن رابطه‌ای بین داده‌های ورودی و خروجی یک سیستم، بدون درک صریحی از فرآیندهای فیزیکی روی داده در آن سیستم است. البته دانش فرد متخصص در مورد فرآیندهای فیزیکی مرتبط با سیستم می‌تواند به بهبود انتخاب متغیرها و نتایج حاصل از به کارگیری مدل منتهی شود. در جدول ۱، تعدادی از مطالعات انجام گرفته در زمینه‌ی استفاده از انواع جدیدتر مدل‌های داده مینا در موضوعات مرتبط با هیدرولوژی و مهندسی منابع آب، به ویژه تشکیل مدل‌های پیش‌بینی رواناب، طبقه‌بندی و ارائه شده‌اند.

در این مطالعه، چهار روش مدل‌سازی داده مینا شامل رگرسیون خطی چندگانه^۲ (MLR)، K نزدیک‌ترین همسایه^۳ (KNN)، شبکه‌های عصبی مصنوعی^۴ (ANN) و سیستم‌های استنتاج عصبی - فازی تطبیقی^۵ (ANFIS) برای تشکیل مدل‌های پیش‌بینی رواناب در یک حوضه به‌کار گرفته شده‌اند. هدف از انجام مطالعه‌ی حاضر، بررسی و مقایسه توانایی این نوع مدل‌ها در پیش‌بینی مقدار رواناب و همچنین آزمون تأثیر انتخاب داده‌های ورودی مختلف در کیفیت پیش‌بینی‌های صورت گرفته است.

۲- روش تحقیق

۲-۱- ناحیه‌ی مورد مطالعه و داده‌های مورد استفاده

در مطالعه‌ی حاضر از داده‌های جریان متوسط ماهانه، بارندگی متوسط ماهانه و دمای متوسط ماهانه ایستگاه (زیرحوضه) سراب هنده در حوضه آبریز دریاچه‌ی نمک استفاده شده است. موقعیت ایستگاه سراب هنده در حوضه آبریز دریاچه نمک در شکل ۱ نشان داده شده است. طول بازه زمانی داده‌های مورد استفاده برابر ۳۵ سال و از سال آبی ۵۱-۱۳۵۰ تا سال آبی ۸۵-۱۳۸۴ است. از داده‌های مربوط به ۲۵ سال نخست برای مدل‌سازی استفاده شده است و

Table 1. Some of studies related to application of the data-driven modeling for stream flow prediction

جدول ۱- برخی مطالعات مربوط به کاربرد مدل‌های داده مبنا در پیش‌بینی رواناب رودخانه

Data-Driven Modeling Technique	Studies
KNN	Karlsson and Yakowitz (1987); Brath et al. (2002); Mehrotra and Sharma (2006); Praire et al. (2006); Solomatine et al. (2008); Elshorbage et al. (2010a,b); Salas and Lee (2010); Lee and Ouarda (2011); Jiang et al. (2013); Silva-Ramirez et al. (2015)
ANN	Minns and Hall (1996); Abrahart and See (2000); Abrahart (2003); Huang et al. (2004); Jeong and Kim (2005); Parasuraman and Elshorbagy (2007); Solomatine et al. (2008); Elshorbagy et al. (2010a); Nourani et al. (2013)
ANFIS	Chen et al. (2006); Aqil et al. (2007); Mukerji et al (2009); Pramanik and Panda (2009); Talei et al. (2010); Ghose et al. (2013); Talei et al. (2013)

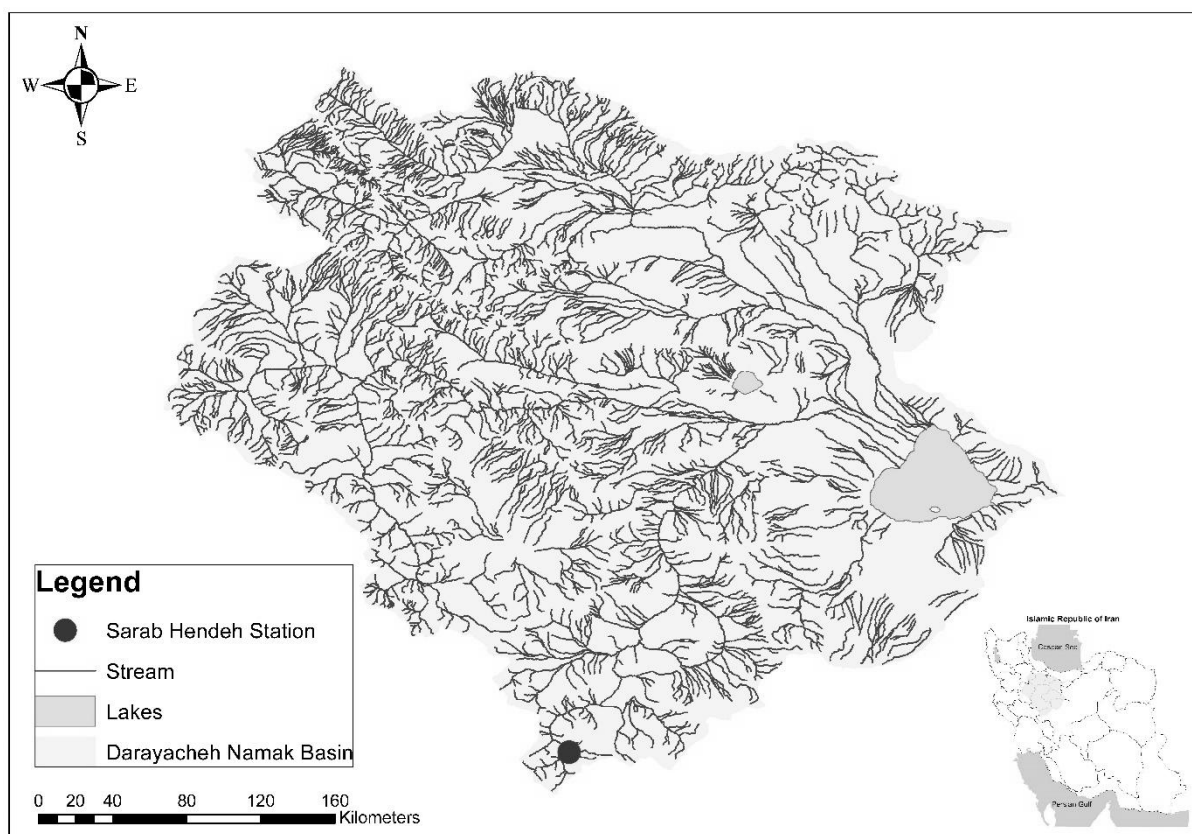


Fig. 1. Geographical location of the Sarab-Hendeh hydrometric station in the Daryacheh-Namak basin

شکل ۱- موقعیت جغرافیایی ایستگاه سراب هنده در حوضه آبریز دریاچه نمک

۳- نتایج و بحث

چهار روش مدل‌سازی داده مبنا برای تشکیل مدل‌های پیش‌بینی رواناب بر اساس شش سناریوی مختلف تعیین متغیرهای ورودی به کار گرفته شدند. در جدول ۲، نتایج حاصل از به‌کارگیری رگرسیون خطی چندگانه برای سناریوهای مختلف انتخاب متغیرهای ورودی، از نظر نزدیکی مقادیر پیش‌بینی شده و مقادیر مشاهداتی برای مجموعه داده‌های ۱۰ سال پایانی مقایسه شده‌اند که بالاترین مقدار ضریب همبستگی خطی (R) مربوط به سناریوی پنجم است که در آن از

بدین منظور توابع عضویت مثلثی، دوزنقه‌ای، زنگوله‌ای شکل تعمیم یافته، گاوسی، ترکیب گاوسی، تابع عضویت مرکب از اختلاف دو تابع سیگموئیدی، تابع عضویت حاصل از ضرب دو تابع سیگموئیدی شکل و تابع عضویت II شکل به کار گرفته شدند.

این توابع در ادامه متن به ترتیب با trimf، trapmf، gbellmf، gauss2mf، gaussmf، dsigmf، psigmf و pimf نشان داده می‌شوند.

داده‌های بارندگی و دمای هر ماه و بارندگی، دما و جریان ماه قبل آن، برای پیش‌بینی جریان در آن ماه استفاده شده است.

پنهان با استفاده از متغیرهای ورودی سناریوی پنجم، بهترین نتایج را در میان این دسته از مدل‌ها حاصل می‌کند.

شکل ۲ مقادیر ضریب همبستگی خطی بین داده‌های مشاهداتی رواناب و مقادیر پیش‌بینی شده با استفاده از تکنیک KNN را برای سناریوهای مختلف انتخاب متغیرهای ورودی نشان می‌دهد که در آن‌ها بالاترین مقدار ضریب همبستگی بین داده‌های مشاهداتی رواناب و مقادیر پیش‌بینی شده مدل‌ها، در اکثر موارد مربوط به سناریوی پنجم است. به ازای برخی مقادیر K نیز سناریوی ششم، نتایج بهتری را ارائه می‌کند. بررسی مقادیر این شاخص نشان می‌دهد که مدل KNN با تعداد ۵ همسایه برای سناریوی پنجم می‌تواند به عنوان گزینه بهینه در میان مدل‌های KNN شناخته شود.

شکل ۴ مقادیر R را برای مدل‌های ANFIS با توابع عضویت مختلف برای شش سناریوی انتخاب متغیرهای ورودی، نشان می‌دهد. بررسی نتایج منعکس شده در این شکل نشان می‌دهد که برای سه سناریوی نخست انتخاب متغیرهای ورودی، استفاده از توابع عضویت مثلثی بهترین نتایج را حاصل می‌کند، در حالی که برای سناریوی چهارم تابع ترکیب گاوسی و برای دو سناریوی پنجم و ششم، تابع زنگوله‌ای شکل تعمیم یافته بهترین انتخاب برای حصول پیش‌بینی‌های نزدیک به مقادیر مشاهداتی است. آن‌چنان که مشاهده می‌شود با انتخاب توابع عضویت مختلف، مقادیر پیش‌بینی شده مربوط به سناریوی ششم، همواره بالاترین مقادیر R را به نمایش می‌گذارند.

در شکل ۳ تغییرات مقادیر شاخص R برای مدل‌های ANN نشان داده شده است. نتایج مندرج در شکل ۳ حاکی از آن است که افزودن متغیر جریان به متغیرهای ورودی مدل‌های پیش‌بینی بر اساس داده‌های تا یک ماه و دو ماه قبل، موجب نزدیک شدن مقادیر پیش‌بینی‌های مدل‌ها به داده‌های مشاهداتی می‌شود. در مجموع، ارزیابی کلیه مدل‌های ANN تشکیل شده بر مبنای شش سناریوی مختلف، حاکی از آن است که استفاده از مدل MLP با ۱ نورون لایه

پس از بررسی نتایج حاصل از به کارگیری چهار تکنیک مختلف مدل‌سازی داده مینا و شناسایی گزینه بهینه مربوط به هر یک از این تکنیک‌ها، عملکرد این گزینه‌های بهینه بر اساس شاخص ضریب همبستگی خطی (R)، شاخص جذر میانگین مربعات خطا (RMSE) و شاخص پراکنش (SI) مورد مقایسه قرار گرفت.

Table 2. The values of correlation coefficient for the results of MLR models

جدول ۲- مقادیر شاخص همبستگی نتایج مدل‌های MLR

Scenario	1	2	3	4	5	6
R	0.46	0.64	0.62	0.55	0.72	0.57

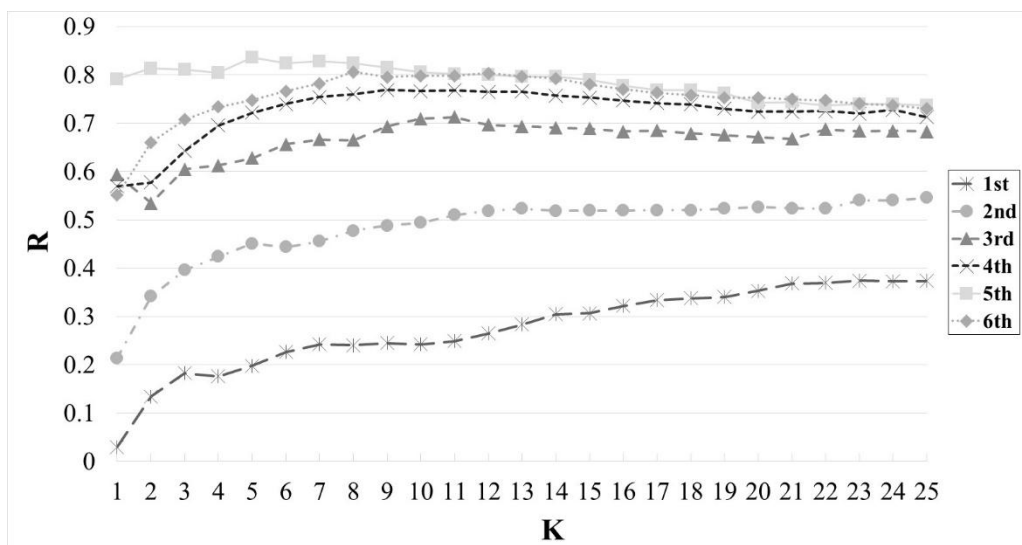


Fig. 2. Linear correlation coefficient between observed data and results of KNN models for 1 to 25 nearest neighbors

شکل ۲- ضریب همبستگی خطی بین مقادیر مشاهداتی و نتایج مدل‌های KNN برای ۱ تا ۲۵ نزدیک‌ترین همسایه

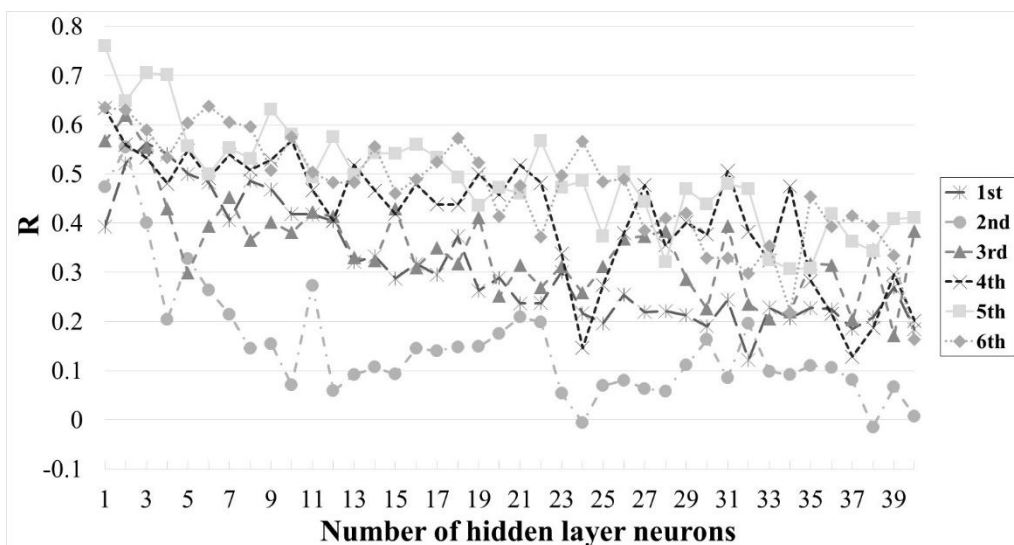


Fig. 3. Linear correlation coefficient between observed data and results of the ANN models for 1 to 40 hidden layer neurons

شکل ۳- ضریب همبستگی خطی بین مقادیر مشاهداتی و نتایج مدل‌های ANN برای ۱ تا ۴۰ نورون لایه پنهان

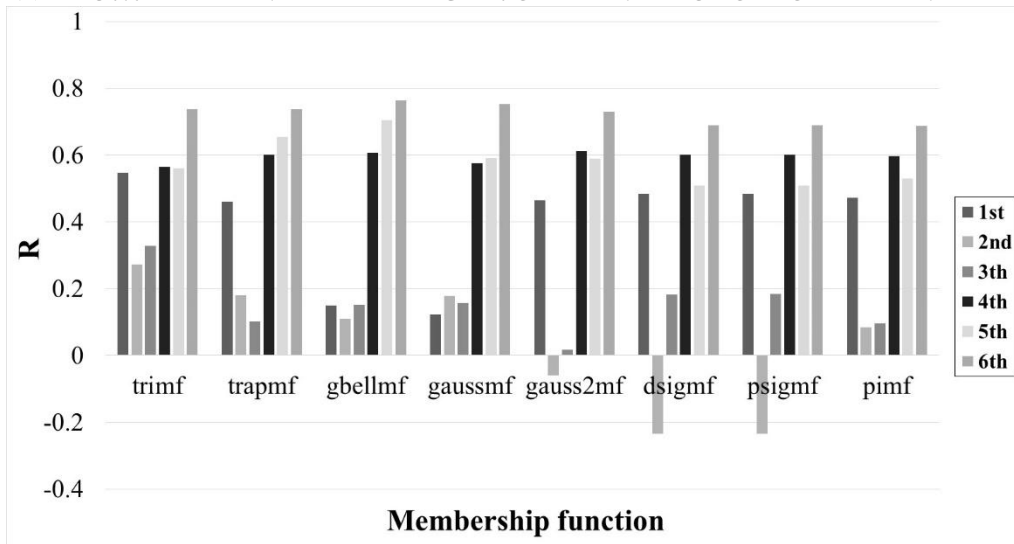


Fig. 4. Linear correlation coefficient between observed data and results of the ANFIS models for the different membership functions

شکل ۴- ضریب همبستگی خطی بین مقادیر مشاهداتی و نتایج مدل‌های ANFIS برای توابع عضویت مختلف

اساس، در میان چهار مدل بهینه، مدل برگزیده KNN با بالاترین مقدار R و پایین‌ترین مقادیر شاخص‌های خطای RMSE و SI، بهترین عملکرد را به نمایش گذاشت و در مقابل مدل برگزیده ANN ضعیف‌ترین عملکرد را نسبت به سایر مدل‌های بهینه ارائه کرد. شایان ذکر است که نتایج مدل ANN در دفعات مختلف مدل‌سازی می‌تواند تا حدودی دچار تغییر شود و در موارد قابل توجهی می‌تواند عملکردی بهتر از مدل رگرسیون خطی را ارائه نماید.

شکل‌های ۵ و ۶ مربوط به مقایسه عملکرد این چهار گزینه بهینه در پیش‌بینی مقادیر رواناب بر اساس شاخص‌های مذکور است. آن چنان که در شکل ۵ مشاهده می‌شود، مقدار ضریب همبستگی خطی (R) بین نتایج مدل‌های برگزیده KNN، ANFIS، MLR و ANN و داده‌های مشاهداتی جریان به ترتیب برابر ۰/۸۴، ۰/۷۶، ۰/۷۲ و ۰/۷۰ محاسبه شد. همچنین مقادیر شاخص خطای RMSE برای این مدل‌ها به ترتیب برابر ۲/۶۴، ۳/۰۲، ۳/۲۶ و ۳/۹۷ و مقادیر شاخص SI برابر ۹۸/۵۱، ۱۱۲/۷۱، ۱۲۱/۴۳ و ۱۴۸/۰۱ محاسبه شد. بر این

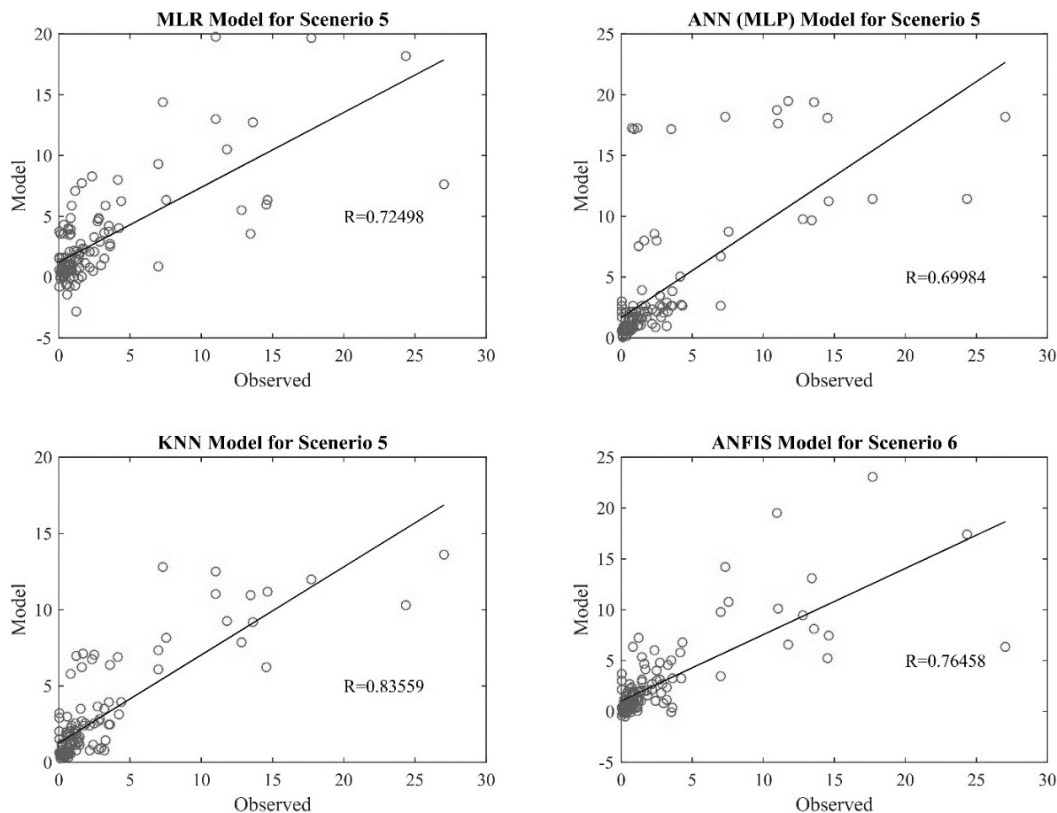


Fig. 5. Linear correlation coefficient between observed data and the results of the selected models of four data-driven modeling techniques

شکل ۵- ضرایب همبستگی خطی بین داده‌های مشاهداتی و نتایج مدل‌های برگزیده چهار تکنیک مدل‌سازی داده مینا

دست آمده توسط Solomatine et al. (2008) مطابقت دارد. در هر دو مورد مذکور، مدل‌های ANN نسبت به مدل‌های حاصل از به کارگیری سایر تکنیک‌ها، نتایج ضعیف‌تری را حاصل کردند. البته باید توجه داشت که نتایج مدل‌های ANN به دلیل ساختار و ماهیت آن‌ها تغییرپذیری قابل توجهی دارد و عوامل مختلفی چون مقدار و کیفیت داده‌های آموزشی، معماری شبکه شامل تعداد لایه‌های پنهان و نورون‌ها و الگوریتم آموزش شبکه می‌توانند در کیفیت نتایج مؤثر باشند. بر اساس نتایج به دست آمده، به نظر می‌رسد در مواردی که مدل‌سازی روابط پیچیده و غیرصریح بین متغیرهای ورودی و متغیر پاسخ مد نظر باشد، در صورت وجود یک مجموعه داده‌ی آموزشی مناسب، روش‌های یادگیری نمونه‌مینا (IBL) مانند KNN دارای توانایی بالایی در این زمینه هستند.

پی‌نوشت‌ها

- 1-Data-Driven Modeling (DDM)
- 2-Multiple Linear Regression (MLR)
- 3-K-Nearest Neighbors (KNN)
- 4-Artificial Neural Network (ANN)
- 5-Adaptive Neuro-Fuzzy Inference System (ANFIS)
- 6-Multi-Layer Perceptron (MLP)
- 7-Scatter Index (SI)

۴- جمع‌بندی

در این مطالعه چهار تکنیک مدل‌سازی داده مینا شامل MLR، ANN، KNN و ANFIS به منظور تشکیل مدل‌های پیش‌بینی رواناب مورد استفاده و بررسی قرار گرفتند و برای انتخاب متغیرهای ورودی هر یک از این مدل‌ها نیز شش سناریوی مختلف اتخاذ شد. نتایج به دست آمده حاکی از آن است که در مورد هر چهار تکنیک مدل‌سازی مورد مطالعه، برای مدل‌سازی و پیش‌بینی مقادیر جریان یا رواناب یک ماه، استفاده از سناریوهای پنجم و ششم که شامل استفاده از داده‌های جریان یک و دو گام ماهانه پیشین به همراه داده‌های بارندگی و دما هستند، در اغلب موارد می‌تواند کیفیت پیش‌بینی‌ها را نسبت به حالتی که مدل‌سازی تنها با استفاده از داده‌های بارندگی و دما انجام شود، بهبود بخشد.

در مورد عملکرد تکنیک‌های مدل‌سازی، چه در مواردی که از داده‌های جریان یک یا دو ماه قبل در میان متغیرهای ورودی استفاده شد و چه در مواردی که این داده‌ها مورد استفاده قرار نگرفتند، هم میانگین نتایج مدل‌های KNN و هم نتایج مدل‌های برگزیده KNN نسبت به نتایج سایر روش‌ها برتری داشتند. این نتیجه با نتایج به

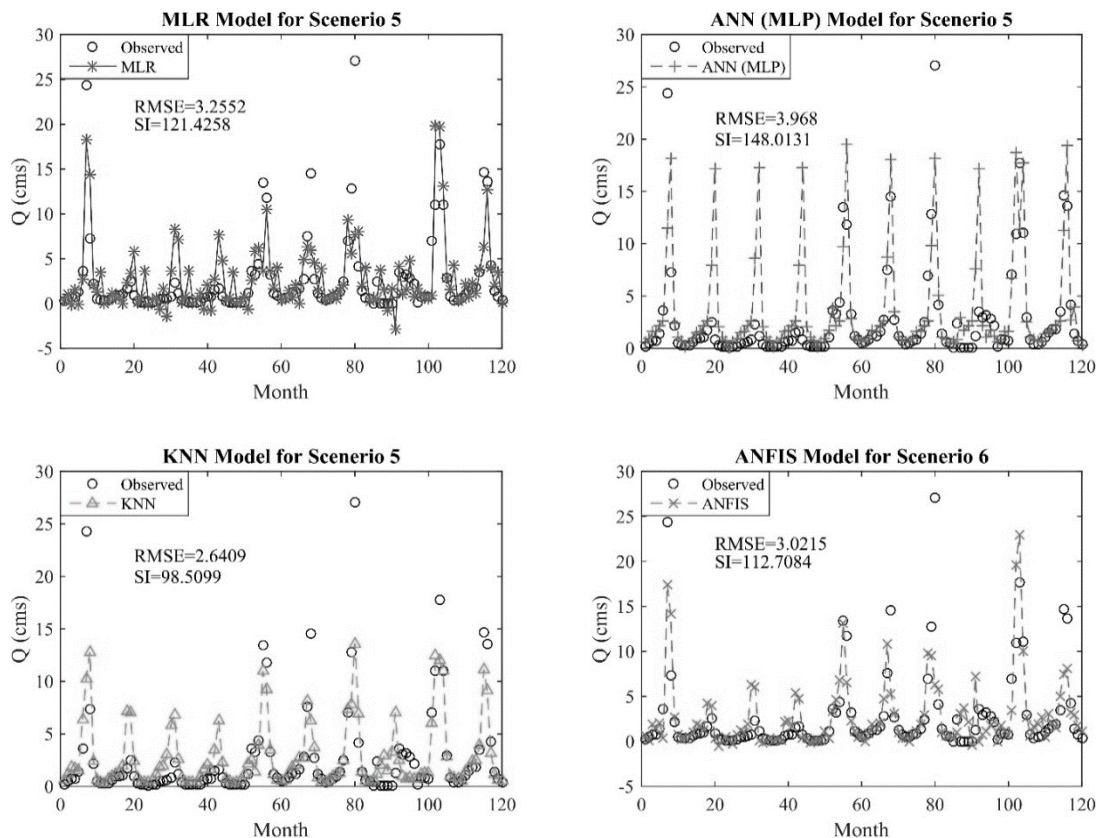


Fig. 6. Time series of the observed data and the forecasted values by the selected models of four data-driven modeling techniques

شکل ۶- سری‌های زمانی داده‌های مشاهداتی و پیش‌بینی‌های مدل‌های برگزیده چهار تکنیک مدل‌سازی داده‌مبنا

neuro-fuzzy network. *Hydrological Processes* 20:1525-1540

Elshorbagy A, Corzo G, Srinivasulu S, Solomatine DP (2010a) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. *Hydrology and Earth System Sciences* 14:1931-1941

Elshorbagy A, Corzo G, Srinivasulu S, Solomatine DP (2010b) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Hydrology and Earth System Sciences* 14:1943-1961

Ghose DK, Panda SS, Swain PC (2013) Prediction and optimization of runoff via ANFIS and GA. *Alexandria Engineering Journal* 59:209-220

Huang W, Xu B, Chan-Hilton A (2004) Forecasting flows in Apalachicola River using neural networks. *Hydrological Processes*, 18:2545-2564

Jeong DI, Kim YO (2005) Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological Processes* 19:3819-3835

۶- مراجع

Abrahart RJ, See L (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes* 14:2157-2172

Abrahart RJ (2003) Neural network rainfall-runoff forecasting based on continuous resampling. *Journal of Hydroinformatics* 5(1):51-61

Aqil M, Kita I, Yano A, Nishiyama S (2007) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behavior of runoff. *Journal of Hydrology* 337:22-34

Brath A, Montanari A, Toth E (2002) Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences* 6(4):627-640

Chen SH., Lin YH, Chang LC, Chang FJ (2006) The strategy of building a flood forecast model by

- Prairie JR, Rajagopalan B, Fulp TJ, Zagona EA, (2006) Modified K-NN model for stochastic streamflow simulation. *Journal of Hydrologic Engineering*, 11:371-378
- Pramanik A, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. *Hydrological Sciences Journal* 54(2):247-260
- Salas JD, Lee T (2010) Nonparametric simulation of single-site seasonal streamflows. *Journal of Hydrologic Engineering* 15:284-296
- Shamseldin AY, O'Connor KM (1996) A nearest neighbor linear perturbation model for river flow forecasting. *Journal of Hydrology* 179:353-375
- Silva-Ramírez EL, Pino-Mejías R, López-Coello M (2015) Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbors for monotone patterns. *Applied Soft Computing* 29:65-74
- Solomatine DP, Maskey M, Shrestha DL (2008) Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes* 22:257-287
- Talei A, Chua LHC, Wong TSW (2010) Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall-runoff modeling. *Journal of Hydrology* 391:248-262
- Talei A, Chua LHC, Quek C, Jansson PE (2013) Runoff forecasting using a Takagi-Sugeno neuro-fuzzy model with online learning. *Journal of Hydrology* 488:17-32
- Jiang Z, Wang HY, Song WW (2013) Discharge estimation based on machine learning. *Water Science and Engineering*, 6(2):145-152
- Karlsson M, Yakowitz S (1987) Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. *Water Resources Research* 23(7):1300-1308
- Lee T, Ouarda TBMJ (2011) Identification of model order and number of neighbors for k-nearest neighbor resampling. *Journal of Hydrology* 404:136-145
- Mehrotra R, Sharma A (2006) Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbor approach. *Advances in Water Resources* 29:987-999
- Minns AW, Hall MJ (1996) Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal* 41(3):399-417
- Mukerji A, Chatterjee C, Raghuwanshi NS (2009). Flood forecasting using ANN, Neuro-Fuzzy, and Neuro-GA models. *Journal of Hydrologic Engineering* 14:647-652
- Nourani V, Hosseini Baghanam A, Adamowski J, Gebremichael M (2013) Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. *Journal of Hydrology*, 476:228-243
- Parasuraman K, Elshorbagy A (2007) Cluster-based hydrologic prediction using genetic algorithm-trained neural networks. *Journal of Hydrologic Engineering* 12:52-62