

## مدل‌سازی جریان رودخانه باراندوزچای با استفاده از روش نزدیکترین-K همسایه و روش‌های هوشمند

هادی ثانی خانی<sup>۱\*</sup>، یعقوب دین پژوه<sup>۲</sup> و محمد علی قربانی<sup>۳</sup>

تاریخ دریافت: ۹۲/۱۱/۱۹ تاریخ پذیرش: ۹۳/۰۵/۲۱

<sup>۱</sup> دانشجوی دکتری منابع آب، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

<sup>۲</sup> دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

\* مسئول مکاتبات، پست الکترونیکی: [hsanikhani12@gmail.com](mailto:hsanikhani12@gmail.com)

### چکیده

پیش‌بینی دقیق جریان رودخانه در طراحی، بهره‌برداری و برنامه‌ریزی منابع آب از اهمیت بسزایی برخوردار است. در این مطالعه، عملکرد روش‌های ناپارامتری نزدیکترین همسایه، فازی-عصبی تطبیقی و روش رگرسیون بردار پشتیبان در پیش‌بینی جریان رودخانه ارزیابی شده است. برای مدل‌سازی از داده‌های جریان ۳۶ ساله ایستگاه هیدرومتری دیزج واقع بر رودخانه باراندوزچای (در مقیاس زمانی ماهانه) استفاده گردید. ترکیبات مختلفی از داده‌های ثبت شده به عنوان الگوی ورودی جهت پیش‌بینی دبی جریان استفاده شد. نتایج به دست آمده حاکی از عملکرد قابل قبول روش‌های مورد استفاده در پیش‌بینی مقدار جریان ماهانه بودند. با اضافه نمودن ضریب فصلی جریان به الگوی ورودی مدل، عملکرد مدل‌های هوشمند در پیش‌بینی به صورت قابل ملاحظه‌ای افزایش یافت. به طور کلی، مدل رگرسیون بردار پشتیبان با بکارگیری مناسب‌ترین الگوی ورودی به عنوان بهترین روش انتخاب شد. مقادیر سه شاخص ارزیابی شامل ضریب همبستگی، ریشه میانگین مربعات خطا و متوسط قدر مطلق خطای نسبی به ترتیب برابر ۰/۸۸، ۳/۶۳ مترمکعب بر ثانیه و ۷۸/۴۵ بود. همچنین ارزیابی عملکرد مدل‌ها در پیش‌بینی مقادیر دبی جریان نشان داد که در موارد با جریان زیاد، همه مدل‌های مورد استفاده، دبی جریان را کمتر از مقدار مشاهداتی تخمین می‌زنند.

واژه‌های کلیدی: پیش‌بینی جریان، رگرسیون بردار پشتیبان، رودخانه باراندوزچای، فازی-عصبی، نزدیکترین همسایه

## Baranduz-Chay River Flow Modeling Using the K-Nearest Neighbor and Intelligent Methods

H Sanikhani<sup>1\*</sup>, Y Dinpashoh<sup>2</sup> and MA Ghorbani<sup>3</sup>

Received: 8 February 2014 Accepted: 12 August 2014

<sup>1-</sup> Ph.D. Student of Water Resources, Dept. of Water Engin., Faculty of Agric. Engin., Univ. of Tabriz, Iran

<sup>2,3-</sup> Assoc. Prof., Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Iran

\* Corresponding Author Email: hsanikhani12@gmail.com

### Abstract

River flow accurate forecasting is so important in design, operation and planning of water resources systems. In this study, the performances of the non-parametric nearest neighbor method, adaptive neuro- fuzzy (ANFIS) and support vector regression (SVR) approaches were evaluated for streamflow forecasting. In order to derive the model, monthly streamflow observations of 36 years time period at Dizaj hydrometric station located in the Baranduz-Chay River (in monthly time scale) were used. Different combinations of the recorded data were used as the input pattern of streamflow forecasting. The results indicated that all of the applied models had reasonable performances in prediction of the monthly river flow. By adding the seasonality coefficient of streamflow to input pattern, performances of the intelligent models increased considerably. In general, the SVR model using the suitable input pattern was selected as the best method. The values of the three different evaluation criteria, namely correlation coefficient (R), root mean square error (RMSE), mean absolute relative error (MARE) were equal to 0.88, 3.63 (m<sup>3</sup>/s) and 78.45, respectively. Furthermore, evaluation of the performances of the models for streamflow forecasting revealed that in the cases of high discharges all of the models underestimated the streamflow comparing with the observed values.

**Keywords:** Baranduz-Chay river, Nearest neighbor, Neuro-fuzzy, Streamflow prediction, Support vector regression

آتی منابع آب می‌باشد (کیشی ۲۰۰۸). از جمله مدل‌های مرسوم جهت پیش‌بینی جریان ماهانه، مدل خودهمبسته میانگین متحرک<sup>۱</sup> می‌باشد که این مدل‌ها به دلیل

مقدمه

پیش‌بینی جریان رودخانه به منظور طراحی، بهره‌برداری و برنامه‌ریزی منابع آب ضروری به نظر می‌رسد. در اغلب موارد، هدف از انجام پیش‌بینی، بهینه‌سازی مصارف آب و برنامه‌ریزی جهت توسعه‌های

<sup>1</sup> Autoregressive moving average

حاکمی از عملکرد قابل قبول K-NN در مقایسه با سایر روش‌ها بود.

درایران نیز مطالعاتی در خصوص بکارگیری روش K-NN در زمینه‌های مختلف انجام شده است. سلاجقه و همکاران (۱۳۸۸) از روش K-NN برای مدل‌سازی دبی جریان رودخانه کرج استفاده کردند. مقایسه نتایج با روش سری زمانی، حاکمی از عملکرد بهتر سری‌های زمانی در پیش‌بینی دبی جریان بود. عزمی و عراقی نژاد (۱۳۹۱) از روش K-NN به منظور پیش‌بینی جریان رودخانه در حوضه بالادست سد زاینده‌رود استفاده کردند و این روش را برای سری‌های تاریخی بلندمدت مناسب دانستند.

در سال‌های اخیر استفاده از روش‌های هوشمند مورد توجه بسیاری از محققان از جمله محققان منابع آب و هیدرولوژی بوده است. در این خصوص، سامانه استنتاج فازی-عصبی تطبیقی<sup>۴</sup> (ANFIS) و ماشین‌های بردار پشتیبان<sup>۵</sup> (SVM) از جمله روش‌هایی هستند که در مدل‌سازی پدیده‌های غیرخطی کاربرد فراوانی دارد. در این میان مدل SVM اغلب برای طبقه‌بندی داده‌ها و یا پیش‌بینی مورد استفاده قرار می‌گیرد. شکلی از SVM را که جهت پیش‌بینی و شبیه‌سازی مورد استفاده قرار می‌گیرد، رگرسیون بردار پشتیبان<sup>۶</sup> (SVR) می‌نامند. در ادامه به تعدادی از تحقیقات انجام شده در خصوص به کارگیری مدل‌های ANFIS و SVM در زمینه منابع آب و هیدرولوژی اشاره می‌گردد.

چانگ و چن (۲۰۰۱) از مدل ANFIS به منظور پیش‌بینی جریان رودخانه‌داکا در مرکز تایلند استفاده نمودند. آن‌ها جهت مدل‌سازی از داده‌های بارش و جریان استفاده کردند. نتایج بدست آمده حاکمی از عملکرد بهتر روش ANFIS در مقایسه با مدل‌های سری زمانی بود. نایاک و همکاران (۲۰۰۴) عملکرد روش ANFIS را در مدل‌سازی سری زمانی هیدرولوژیکی مورد بررسی قرار دادند و از آن برای مدل‌بندی جریان آب رودخانه‌ای در هندوستان بهره گرفتند. مقایسه نتایج

برقراری رابطه خطی بین متغیرها از عدم قطعیت بالایی برخوردارند. ماهیت جریان رودخانه اغلب متغیر بوده و لذا مدل‌های خطی نظیر مدل‌های خودهمبسته کارایی لازم را در این خصوص نداشته و لازم است تا از مدل‌های غیرخطی استفاده شود (مایر و دندی ۲۰۰۰). استفاده از روش ناپارامتری نزدیکترین همسایگی<sup>۱</sup> (K-NN) جهت تشخیص الگو و پیش‌بینی همواره مورد توجه محققین بوده و مطالعات انجام شده حاکمی از کارایی بالای این روش می‌باشد (داساراسی ۱۹۹۱).

کمبر و همکاران (۱۹۹۳) از روش K-NN جهت پیش‌بینی دبی جریان روزانه رودخانه‌ای در کانادا در یک دوره آماری ۷۰ ساله استفاده نمودند. آن‌ها در مطالعات خود یک روش K-NN سه پارامتری به صورت دینامیک و غیرخطی ارائه و ارتباط میزان خطای پیش‌بینی را با طول داده‌های تاریخی مورد بررسی قرار دادند. مقایسه نتایج این روش با نتایج مدل خود همبسته میانگین متحرک (ARMA)، حاکمی از عملکرد بهتر روش K-NN می‌باشد. شارما و لال (۱۹۹۹) یک روش ناپارامتری برگرفته از الگوریتم K-NN را برای شبیه‌سازی بارش روزانه ارائه دادند. روش پیشنهادی برای یک دوره ۱۲۳ ساله بارش روزانه در سیدنی استرالیا مورد بررسی قرار گرفت و عملکرد آن موفقیت آمیز گزارش شد.

پیتروفسکی و همکاران (۲۰۰۶) عملکرد روش K-NN و چندین نوع از شبکه‌های عصبی مصنوعی<sup>۲</sup> (ANN) را در پیش‌بینی سیلاب‌های ناگهانی در یک حوضه کوهستانی مورد مطالعه قرار دادند. نتایج بدست آمده حاکمی از عملکرد بهتر روش ANN از نوع تابع پایه شعاعی<sup>۳</sup> در مقایسه با سایر روش‌ها در پیش‌بینی دبی اوج سیلاب‌های ۳ و ۶ ساعته بود. وئو و چائو (۲۰۱۰) عملکرد مدل‌های مختلف شامل مدل ARMA، ANN و K-NN را در پیش‌بینی جریان ماهانه در چندین رودخانه از کشور چین مورد بررسی قرار دادند. نتایج بدست آمده

<sup>۴</sup> Adaptive neuro-fuzzy inference system

<sup>۵</sup> Support vector machines

<sup>۶</sup> Support vector regression

<sup>۱</sup> K-nearest neighbor

<sup>۲</sup> Artificial neural networks

<sup>۳</sup> Radial basis function

مدل در پیش‌بینی مقادیر دبی جریان در بازه‌های زمانی کوتاه مدت و طولانی مدت عملکرد قابل قبولی دارند ولیکن مقایسه عملکرد مدل‌ها نشان داد که مدل ترکیبی موجک- فازی- عصبی عملکرد بهتری در مقایسه با مدل‌های دیگر داشته است.

کیشی و سیمین (۲۰۱۱) از مدل ترکیبی موجک و SVR جهت پیش‌بینی جریان ماهانه رودخانه استفاده کردند. نتایج بدست آمده حاکی از عملکرد بهتر مدل ترکیبی در مقایسه با مدل SVR بود. لوهانی و همکاران (۲۰۱۲) به مقایسه عملکرد مدل‌های ANFIS، ANN و خودهمبسته در پیش‌بینی جریان ماهانه ورودی به مخزن سدی در هندوستان پرداختند. یافته‌های آن‌ها حاکی از برتری مدل ANFIS در مقایسه با مدل‌های ANN و خودهمبسته به ازای کلیه الگوهای مورد استفاده به عنوان ورودی مدل بود.

با بررسی منابع انجام شده می‌توان نتیجه گرفت که استفاده از روش‌های مدل‌سازی غیرخطی جهت پیش‌بینی مقدار جریان رودخانه از عملکرد بالایی برخوردارند. در این راستا روش K-NN به دلیل عملکرد قابل قبول و سادگی اجرای آن و روش‌های ANFIS و SVR بدلیل کارایی بالا می‌توانند مورد استفاده قرار گیرند. بنابراین هدف از این تحقیق، بررسی عملکرد روش آماری ناپارامتری K-NN و روش‌های هوشمند ANFIS و SVR در تخمین جریان رودخانه باراندوزچای می‌باشد. همچنین در این تحقیق به بررسی اثرات ضریب فصلی جریان در پیش‌بینی‌های انجام شده پرداخته شده است.

### مواد و روش‌ها

#### حوضه آبریز و داده‌ها

حوضه باراندوزچای با مساحت ۱۲۰۳ کیلومترمربع در شمال‌غرب کشور بین دریاچه ارومیه و مرز ایران و کشورهای عراق و ترکیه واقع شده است. گستردگی جغرافیایی این حوضه از ۴۴° ۴۵' تا ۴۵° ۱۴' طول شرقی و ۳۷° ۰۶' تا ۳۷° ۲۹' عرض شمالی است. طول آبراهه اصلی ۷۵ کیلومتر، حداکثر ارتفاع این

مدل ANFIS با ANN و مدل‌های مربوط به سری‌های زمانی نشان داد که مدل ANFIS به لحاظ سرعت محاسباتی، خطای پیش‌بینی و قدرت تخمین مقادیر اوج جریان عملکرد بهتری دارد. موسوی و همکاران (۲۰۰۷) عملکرد مدل‌های رگرسیون کمترین مربعات، رگرسیون فازی و ANFIS را جهت استخراج قوانین بهره‌برداری مخازن مورد مقایسه قرار دادند. نتایج بدست آمده نشان داد که مدل ANFIS در بهینه‌سازی استوکاستیکی صریح در بازه‌های زمانی متوسط، مفید بوده و قادر است مولفه‌های اصلی سیستم را از مجموعه داده‌های ورودی- خروجی تولیدی، استخراج و قوانین بهره‌برداری عمومی مخزن را ارائه کند.

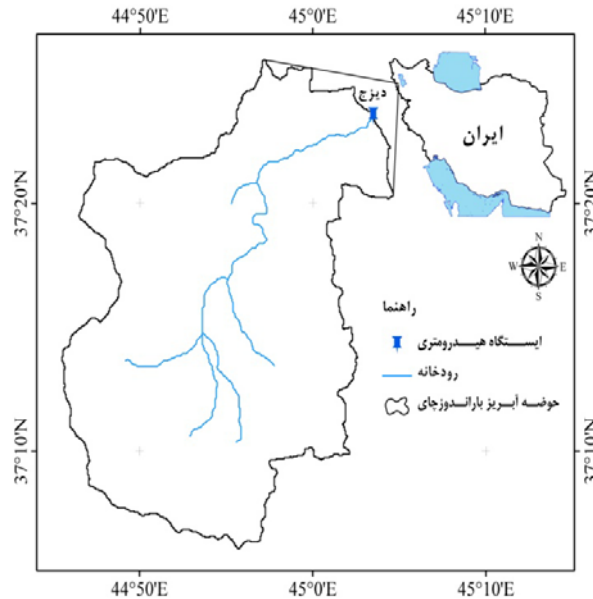
فیرات و گونگور (۲۰۰۷) به تخمین جریان رودخانه با استفاده از ANFIS در حوضه‌ای در ترکیه پرداختند. آن‌ها نتایج مدل ANFIS را در سری داده‌های آموزش و آزمون با داده‌های مشاهداتی مقایسه نموده و عملکرد مدل را مورد ارزیابی قرار دادند و نتیجه گرفتند که مدل ANFIS می‌تواند به طور موفقیت آمیزی جهت پیش‌بینی دبی جریان آب رودخانه به کار گرفته شود. وانگ و همکاران (۲۰۰۹) به مقایسه عملکرد مدل ARMA و چندین مدل هوشمند شامل ANN، ANFIS، برنامه‌ریزی ژنتیک<sup>۱</sup> (GP) و SVM جهت پیش‌بینی سری زمانی دبی جریان در مقیاس ماهانه پرداختند. نتایج آن‌ها نشان داد که به ازای داده‌های مورد استفاده در مراحل آموزش و آزمون، بهترین عملکرد با توجه به معیارهای مختلف ارزیابی خطا به مدل‌های ANFIS، GP و SVM اختصاص دارد.

نورانی و همکاران (۲۰۱۱) عملکرد دو مدل ترکیبی هوشمند را در مدل‌سازی فرآیند بارش- رواناب در دو حوضه از استان آذربایجان شرقی مورد مطالعه قرار دادند. مدل‌های بکار گرفته شده شامل مدل خودهمبسته میانگین متحرک فصلی با ورودی برون- زاد<sup>۲</sup> (SARIMAX) در ترکیب با ANN و مدل ترکیبی موجک- فازی- عصبی بود. نتایج نشان داد که هر دو

<sup>۱</sup>Genetic programming

<sup>۲</sup>Seasonal Auto Regressive Integrated Moving Average with Exogenous input

تغذیه کننده دریاچه از جمله رودخانه باراندوزچای لازم و ضروری است. شکل ۱ موقعیت جغرافیایی منطقه مورد مطالعه را نشان می‌دهد.



شکل ۱- موقعیت جغرافیایی حوضه آبریز باراندوزچای و ایستگاه دیزج در منطقه مورد مطالعه.

تابع هزینه<sup>۱</sup> زیر را روی بردار سری‌های زمانی  $D_N$  (داده‌های آموزشی) کمینه می‌کند:

$$C(y^{NP}) = \frac{\sum_{i=1}^N (y_i - y^{NP})^2 K(x_i, x_0, b)}{N} \quad [1]$$

در این رابطه، متغیر خروجی، با  $y_i$  و متغیرهای ورودی به صورت بردار  $x_i$  نشان داده می‌شوند. متغیرهای ورودی را می‌توان بارش، جریان، جریان با تاخیرهای مختلف و یا هر متغیر دیگر، طوری که ترکیبی از این متغیرها در یک یا چند ایستگاه است، در نظر گرفت.  $x_0$  مشخص کننده نقطه پرسش<sup>۲</sup> بوده و از مجموعه داده‌های آزمون در فضای ورودی انتخاب می‌شود. مقادیر متغیر خروجی با توجه به  $x_0$  محاسبه می‌شود (پانده و همکاران ۲۰۰۹). مشابه  $D_N$  مجموعه دیگری تحت عنوان داده‌های آزمون یا  $D_M$  در نظر گرفته می‌شود، طوری که این داده‌ها هیچ عضو مشترکی با داده‌های مربوط به بخش آموزشی یا  $D_N$  ندارند و نقطه

حوضه ۳۵۰۰ متر و حداقل ارتفاع در نقطه خروجی حوضه معادل ۱۲۵۰ متر از سطح آزاد دریاها می‌باشد. با توجه به وضعیت بحرانی منابع آب و خشکسالی در دریاچه ارومیه، توجه به سرشاخه‌های

در این تحقیق، از داده‌های جریان ماهانه در ایستگاه دیزج واقع بر رودخانه باراندوزچای در خلال سالهای ۱۳۵۳ تا ۱۳۸۸ (۳۳۲ ماه) استفاده شد. داده‌های مورد استفاده از سازمان آب منطقه‌ای استان آذربایجان غربی تهیه شدند. دلیل انتخاب این ایستگاه داشتن رژیم طبیعی در محدوده مورد مطالعه و همچنین داشتن داده‌های با کیفیت و طول آماری مناسب در منطقه بود.

#### روش نزدیکترین همسایگی

روش K-NN از توابع رگرسیونی ناپارامتری محسوب می‌شود. اگر یک سری زمانی هیدرولوژیکی معین نظیر  $\{y_i, x_i\} \in R_+^1 \times R_+^1, i = 1, \dots, N$  در دست باشد، یک تخمین‌گر رگرسیونی ناپارامتری،  $y^{NP}$

<sup>۱</sup> Cost function

<sup>۲</sup> Query Point

که اصطلاحاً تابع عضویت<sup>۱</sup> نامیده می‌شود. پارامترهای توابع عضویت با استفاده از الگوریتم آموزش و بر اساس دو روش زیر تعیین می‌شوند:

(۱) روش پس انتشار خطا برای همه پارامترهای توابع عضویت.

(۲) روش ترکیبی شامل پس انتشار خطا برای پارامترهای توابع عضویت ورودی و روش حداقل مربعات برای پارامترهای توابع عضویت خروجی.

به‌طور کلی در جعبه ابزار منطق فازی دو نوع سامانه استنتاج فازی موسوم به الف) سامانه استنتاجی نوع ممدانی<sup>۲</sup> و ب) سامانه استنتاجی نوع سوگنو<sup>۳</sup> وجود دارد. تفاوت اصلی آنها در نوع تابع عضویت است، به‌گونه‌ای که تابع عضویت خروجی در روش ممدانی غیرخطی است ولی در روش سوگنو خطی یا ثابت است. اگر فرض کنیم یک سامانه استنتاج فازی دارای دو ورودی  $x$  و  $y$  و یک خروجی  $z$  باشد، مدل فازی مرتبه اول سوگنو با دو قانون فازی اگر-آنگاه به صورت زیر بیان می‌شود:

قانون ۱- اگر  $x$  برابر  $A_1$  و  $y$  برابر  $B_1$  باشد آنگاه:

$$z_1 = p_1x + q_1y + r_1$$

قانون ۲- اگر  $x$  برابر  $A_2$  و  $y$  برابر  $B_2$  باشد آنگاه:

$$z_2 = p_2x + q_2y + r_2$$

در شکل ۲ مدل فازی مرتبه اول نوع سوگنو با دو ورودی ( $x$  و  $y$ )، یک خروجی  $f$ ، به همراه دو قانون فازی و همچنین مدل ANFIS معادل آن نشان داده شده است.

مدل‌بندی در محیط MATLAB با استفاده از قابلیت‌های نوار ابزار مربوط به ANFIS، کدنویسی شده است. لازم به توضیح است که ANFIS برای طبقه‌بندی

پرسش نیز از بین این داده‌ها (مجموعه داده‌های تست) انتخاب می‌شود و متناظر با متغیر مستقل مربوط به بردار  $X$  می‌باشد. همچنین متغیر وابسته  $y$  در مجموعه داده‌های آزمون جهت تست تخمین‌گر با توجه به اطلاعات داده‌های بخش آموزش یا  $D_N$  استفاده می‌شود. تابع  $K$  نیز معرف وزن یا تابع کرنل می‌باشد که مقدار آن با توجه به فاصله اقلیدسی هر نقطه در بخش آموزش از نقطه پرسش در بخش آزمون طبق رابطه ۲ بدست می‌آید:

$$K(x_i, x_o, b) = \begin{cases} 1 & \text{if } \|x_i - x_o\| \leq b \\ 0 & \text{if } \|x_i - x_o\| > b \end{cases} \quad [2]$$

در این رابطه  $\| \cdot \|$  نشان دهنده فاصله اقلیدسی و  $b$  شعاع همسایگی است. کمینه کردن معادله (۱) با توجه به مقدار پارامتر  $y^{NP}$  و کرنل تعریف شده در رابطه ۲ انجام می‌شود. برای تخمین  $y^{nn}$  از رابطه زیر استفاده می‌شود:

$$y_{pred}^{nn} = \frac{\sum_{i \in I_{nn}} y_i}{|I_{nn}|} \quad [3]$$

که در آن  $I_{nn}$  مجموعه‌ای است که عضوهای آن، اندیس آن تعداد از داده‌های مشاهداتی است که در داخل دایره-ای به شعاع  $b$  از نقطه پرسش قرار دارند و  $|I_{nn}|$  تعداد عضوهای مربوط به مجموعه  $I_{nn}$  می‌باشد یعنی:

$$|I_{nn}| = \{i: \|x_i - x_o\| \leq b\} \quad [4]$$

کارایی عملکرد این روش، به انتخاب پارامترهای  $b$  (شعاع همسایگی) و  $l$  (تعداد تاخیرها در پارامترهای ورودی) بستگی دارد (پانده و همکاران ۲۰۰۹).

#### سیستم استنتاج فازی-عصبی تطبیقی

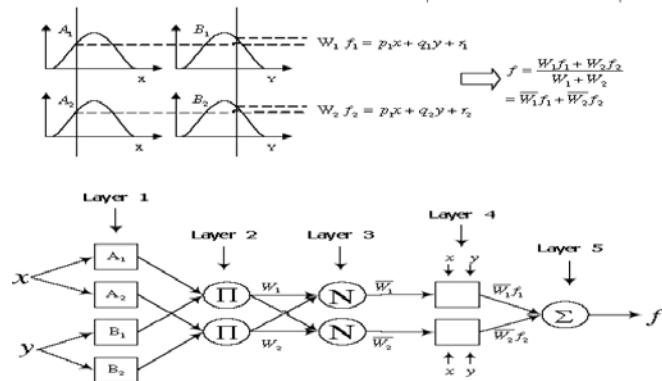
سیستم استنتاج فازی-عصبی تطبیقی (ANFIS) برای اولین بار توسط جانگ (۱۹۹۳) معرفی شد. این سامانه ابزاری کارآمد جهت تخمین توابع پیوسته حقیقی در دامنه‌ای محدود می‌باشد که به ازای هر درجه‌ای از صحت قابل استفاده است. بخش فازی مدل ANFIS رابطه‌ای بین متغیرهای ورودی و خروجی برقرار نموده

<sup>1</sup>Membership Function

<sup>2</sup>Mamdani

<sup>3</sup>Sugeno

داده‌ها از دو روش افراز شبکه و دسته‌بندی تفریقی استفاده می‌کند.



شکل ۲- مدل فازی مرتبه اول سوگنو با دو قانون فازی و مدل ANFIS معادل آن (ژانگ ۱۹۹۳).

و از یک توزیع احتمالاتی نامعین با تابع  $P(\vec{x}, y)$  گرفته شده‌اند و رابطه زیر برای آن‌ها وجود دارد:

$$F = \{f|f(\vec{x}) = (\vec{w}, \vec{x}) + B, \vec{w} \in R^n, R^n \rightarrow R\} \quad [5]$$

که در این رابطه بردار  $\vec{w}$  و ضریب  $B$  با توجه به داده‌های ورودی تخمین زده می‌شوند. با توجه به رابطه ارائه شده، مهمترین مساله، یافتن تابعی مانند  $f(\vec{x}) \in F$  است که تابع ریسک بشرح زیر را کمینه می‌کند:

$$R[f(\vec{x})] = \int L(y - f(\vec{x}), x) dP(\vec{x}, y) \quad [6]$$

که در این رابطه  $L$  تابع زیان<sup>۱</sup> بوده که جهت تعیین اختلاف بین مقدار خروجی  $y$  و مقادیر تخمینی  $f(\vec{x})$  می‌باشد. از آنجا که تابع توزیع احتمالاتی  $P(\vec{x}, y)$  نامعلوم می‌باشد، لذا نمی‌توان مستقیماً  $R[f(\vec{x})]$  را حداقل نمود، اما تابع ریسک تجربی به شرح زیر محاسبه می‌گردد:

$$R_{\text{emp}}[f(\vec{x})] = \frac{1}{N} \sum_{i=1}^N L(y - f(\vec{x}_i)) \quad [7]$$

حداقل نمودن تابع ریسک تجربی مطابق رابطه ۷ بدون استفاده از ابزاری جهت کنترل یا تنظیم ساختار قابل توصیه نمی‌باشد. بنابراین، تابع ریسک تنظیم شده با حداقل شیب در میان توابع موجود جهت کمینه نمودن تابع ریسک تجربی بشرح زیر به کار گرفته شد:

$$R_{\text{reg}}[f(\vec{x})] = R_{\text{emp}}[f(\vec{x})] + \gamma \|\vec{w}\|^2 \quad [8]$$

عملکرد این دو روش در مدل ANFIS مشابه بوده و در پاره‌ای از موارد عملکرد روش دسته‌بندی تفریقی مناسب‌تر گزارش شده است (کوبانر ۲۰۱۱). به همین دلیل، در این تحقیق از روش دسته‌بندی تفریقی جهت مدل‌سازی در مدل ANFIS استفاده شد. از مهمترین پارامترها جهت بکارگیری این روش، تعیین شعاع دسته‌بندی می‌باشد. انتخاب مقداری کوچک برای آن، موجب کوچک شدن بیش از حد دسته‌ها در فضای داده‌ها و به تبع آن موجب افزایش قوانین فازی می‌گردد. بنابراین، انتخاب صحیح شعاع موثر جهت دسته‌بندی داده‌ها ضروری است. پس از تعیین شعاع موثر، باید قوانین فازی و توابع عضویت تعیین گردند. سپس روش مربعات خطا برای تعیین نتیجه توابع عضویت خروجی و سیستم انطباقی فازی به کار گرفته می‌شود (چیو ۱۹۹۷).

#### ماشین‌های بردار پشتیبان

ایده اصلی SVM که به عنوان فرآیندهای رگرسیونی و طبقه‌بندی شناخته می‌شوند، توسط واپنیک (۱۹۹۵) مطرح گردید. SVR جهت توصیف رگرسیون با استفاده از SVM می‌باشد. در تخمین رگرسیونی با روش SVR تلاش بر این است تا وابستگی بین مجموعه داده‌های ورودی  $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_1\}$  از فضای  $R^n$  و مقادیر خروجی  $Y = \{\vec{y}_1, \vec{y}_2, \dots, \vec{y}_1\}$  (که در آن  $y_i \in R$ ) از طریق تابع  $f(\vec{x})$  تخمین زده شود. در این روش، فرض بر این است که نمونه‌ها به صورت مستقل

<sup>۱</sup> Loss function

پردازش روی داده‌ها قابل بررسی است). بنابراین، ترم اریب حذف می‌گردد (واپنیک ۱۹۹۵). تابع کرنل قابلیت انجام عملیات را در فضای ورودی دارد. بنابراین، محصول درونی در فضای مشخصه دارای یک کرنل معادل در فضای ورودی می‌باشد. به طور کلی، توابع کرنل مورد استفاده در روش SVR توابع چند جمله‌ای، پایه شعاعی گوسی<sup>۲</sup> (GRBF)، پایه شعاعی نمایی، پرسپترون چند لایه و اسپیلاین می‌باشند. GRBF از پر کاربردترین توابع در قالب توابع کرنل می‌باشند که رابطه مربوط به آن به صورت زیر می‌باشد:

$$K(x, x_i) = \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right) \quad [۱۲]$$

که در آن  $\sigma$  انحراف معیار میزان اغتشاشات گوسی<sup>۳</sup> می‌باشد.

#### ارزیابی عملکرد مدل‌ها

جهت آزمون عملکرد مدل‌های مورد استفاده در این مطالعه از سه شاخص آماری شامل ضریب همبستگی (R)، ریشه میانگین مربعات خطا (RMSE) و متوسط قدرمطلق خطای نسبی (MARE) مطابق روابط زیر استفاده گردید:

$$R = \sqrt{1 - \frac{\sum_{i=1}^N (O_i - S_i)^2}{\sum_{i=1}^N O_i^2 + \sum_{i=1}^N S_i^2}} \quad [۱۳]$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - S_i)^2} \quad [۱۴]$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|O_i - S_i|}{O_i} * 100\right) \quad [۱۵]$$

در روابط فوق، N تعداد داده‌ها،  $O_i$  مقدار مشاهداتی برای جریان رودخانه و  $S_i$  مقدار پیش‌بینی شده برای جریان رودخانه توسط مدل در گام زمانی  $t$ ام می‌باشد.

#### نتایج و بحث

در این مطالعه، از داده‌های ۲۶ سال اول (۳۱۲ ماه) به منظور آموزش و از ده سال بقیه (۱۲۰ ماه) برای آزمون مدل‌ها استفاده شد. پارامترهای آماری

که در این رابطه  $\gamma$  یک ضریب ثابت ( $\gamma \geq 0$ ) می‌باشد. عبارت اضافه شده در رابطه بالا موجب کاهش فضای مدل و در نتیجه سبب کنترل پیچیدگی روند حل مساله می‌گردد. به همین دلیل، می‌توان رابطه زیر را ارائه نمود:

$$R_{reg}[f(\vec{x})] = C_c \sum_{x_i \in X} l_\varepsilon(y_i - f(\vec{x})) + \frac{1}{2} \gamma \|\vec{w}\|^2 \quad [۹]$$

که در این رابطه  $C_c$  یک ضریب ثابت مثبت (پارامتر اضافه شده جهت کنترل ظرفیت) می‌باشد که از قبل انتخاب می‌گردد. ثابت  $C_c$  که توازن بین خطای تقریب و بردار وزنی رگرسیون ( $\|\vec{w}\|$ ) را تحت تاثیر قرار می‌دهد، یک پارامتر طراحی می‌باشد (سیمن ۲۰۰۸). در چنین شرایطی، تابع زیان که تابع زیان غیرحساس نامیده می‌شود، دارای این مزیت است که دیگر نیازی به کلیه داده‌های ورودی جهت توصیف بردار رگرسیون نمی‌باشد و می‌تواند به صورت زیر ارائه گردد:

$$l_\varepsilon(y_i - f(\vec{x})) = \begin{cases} 0 & \text{for } (y_i - f(\vec{x})) < \varepsilon \\ (y_i - f(\vec{x})) & \text{otherwise} \end{cases} \quad [۱۰]$$

این تابع هنگامی که با عبارت اصلاحی ( $\gamma \|\vec{w}\|^2$ ) ترکیب گردد، به صورت تخمین‌گر اریب عمل می‌کند. اگر اختلاف بین پیش‌بینی انجام شده ( $f(\vec{x})$ ) و مقدار اندازه‌گیری شده ( $y_i$ ) کمتر از  $\varepsilon$  باشد، زیان برابر صفر می‌شود. انتخاب مقدار  $\varepsilon$  آسان‌تر از انتخاب  $C_c$  می‌باشد و به صورت درصد مطلوب مقادیر خروجی توصیف می‌شود. بنابراین، تابع رگرسیونی غیرخطی توسط تابعی که کمینه‌کننده رابطه ۹ (با توجه به رابطه ۱۰) است، به صورت زیر ارائه شد:

$$f(x) = \frac{1}{N} \sum_{i=1}^N (\alpha_i^* - \alpha_i) K(x, x_i) + B \quad [۱۱]$$

که در آن  $\alpha_i$  و  $\alpha_i^*$  شمارنده‌های لاگرانژ<sup>۱</sup>، عبارت اریب و تابع کرنل می‌باشد. اغلب فرض می‌شود که داده‌ها دارای میانگین صفر می‌باشند (این مساله با پیش

<sup>۲</sup> Gaussian radial basis function

<sup>۳</sup> Gaussian noise level

<sup>۱</sup> Lagrange multipliers



مقدار نظیر در دوره آموزش بیشتر است. این واقعیت جهت برون‌یابی داده‌های دوره آزمون به منظور ارزیابی بهتر مدل‌های مورد استفاده در این تحقیق حائز اهمیت خاصی است (کیشی ۲۰۰۷).

ماهانه داده‌های جریان در ایستگاه دیزج در جدول ۱ درج شده است. با توجه به جدول مذکور برای داده‌های جریان در ایستگاه دیزج، گرچه حداقل دبی جریان در دوره آزمون از حداقل مقدار آن در دوره آموزش مدل کمتر است، ولی حداکثر مقدار جریان در دوره آزمون از

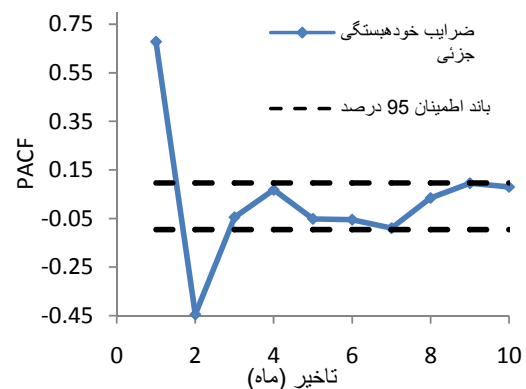
جدول ۱- مقادیر پارامترهای آماری داده‌های ماهانه جریان رودخانه باراندوزچای در ایستگاه دیزج (۱۳۵۳-۱۳۸۸).

نام دوره	دوره آماری	میانگین	حداکثر	حداقل	انحراف معیار	ضریب تغییرات	ضریب چولگی
		(m <sup>3</sup> /s)	(m <sup>3</sup> /s)	(m <sup>3</sup> /s)	(m <sup>3</sup> /s)		(m <sup>3</sup> /s)
آموزش	۱۳۷۸-۱۳۵۳	۸/۶۰	۴۳/۸۸	-۰/۱۴	۸/۱۳	۰/۹۴	۱/۸۰
آزمون	۱۳۷۹-۱۳۸۸	۶/۱۲	۴۵/۱۳	-۰/۰۲	۷/۶۲	۱/۲۴	۲/۵۸

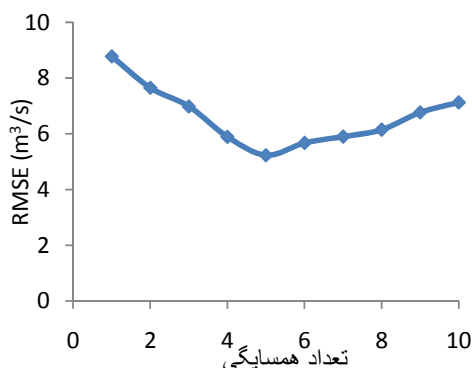
با توجه به شکل ۳ مشاهده می‌شود که مقادیر جریان در ماه جاری به مقادیر پیشین جریان وابستگی داشته و این وابستگی به ازای تاخیر برابر ۲ ماه وجود دارد. بنابراین، جهت اطمینان بیشتر در فرآیند مدل‌سازی از مقادیر جریان تا ۳ ماه پیشین نیز به عنوان ورودی مدل‌ها استفاده گردید.

در گام اول عملکرد روش K-NN در پیش‌بینی جریان مورد بررسی قرار گرفت. به طوری که قبلاً اشاره شد، در این روش پیش‌بینی با توجه به تعداد تاخیرها و همچنین شعاع همسایگی انجام می‌شود. البته انتخاب تعداد منطقی همسایگی‌ها نیز از پارامترهای مهم در پیش‌بینی، توسط این روش می‌باشد. با استفاده از سعی و خطا و ملاحظه عملکرد مدل در این تحقیق بهترین پیش‌بینی به ازای تاخیر برابر ۳ گام زمانی و شعاع همسایگی برابر ۰/۴۸ بدست آمد. در شکل ۴، تغییرات شاخص آماری RMSE نسبت به تعداد همسایگی (k) نشان داده شده است. با توجه به این شکل می‌توان استنباط کرد که بهینه‌ترین تعداد همسایگی برای مطالعه حاضر برابر ۵ می‌باشد، زیرا در این حالت کمترین مقدار RMSE حاصل شده است.

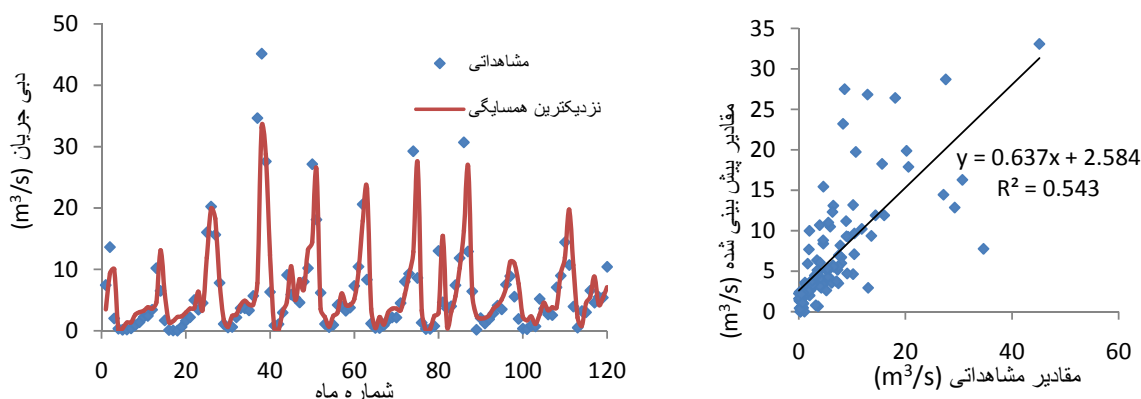
جهت پیش‌بینی جریان در ماه جاری از داده‌های جریان در ماه‌های قبل به عنوان ورودی در مدل‌های مورد بررسی استفاده گردید. جهت تعیین ترکیبات ورودی‌های مختلف در مدل‌ها، از آزمون‌های آماری رگرسیون و تعیین ضرایب خودهمبستگی (ACF) و خودهمبستگی جزئی (PACF) سری‌های زمانی جریان در ایستگاه مورد مطالعه استفاده شد. نمودار ضرایب خود همبستگی جزئی برای ایستگاه دیزج در شکل ۳ نشان داده شده است.



شکل ۳- مقادیر ضرایب خودهمبستگی جزئی برای سری جریان رودخانه باراندوزچای در مقیاس ماهانه.



شکل ۴- تغییرات مقدار شاخص ریشه میانگین مربعات خطا به ازای تعداد همسایگی انتخاب شده در این مطالعه.



شکل ۵- مقادیر مشاهداتی و پیش‌بینی شده با مدل K-NN در ایستگاه دیزج در رودخانه باراندوزچای ارومیه.

شد. در این بخش اثر فصلی و تناوبی جریان با اضافه کردن ضریب  $\alpha$  مورد بررسی قرار گرفت. مقدار ضریب  $\alpha$  بین ۱ و ۱۲ می‌باشد و مقدار آن برای ماه فرودین برابر ۱ و برای ماه اسفند ۱۲ در نظر گرفته شد. پارامترهای اصلی مربوط به ساختار مدل ANFIS از نوع دسته‌بندی تفریقی شامل شعاع دسته بندی می‌باشد که مقدار آن بین ۰ و ۱ در نظر گرفته می‌شود. در این تحقیق ساختار اصلی مدل ANFIS و مقدار شعاع دسته‌بندی به ازای ورودی‌های مختلف با آزمون و خطا بدست آمد. ساختار و شاخص‌های ارزیابی مدل برای مدل ANFIS در دوره آزمون برای داده‌های جریان در ایستگاه دیزج در جدول ۲ ارائه شده است. در این جدول نشان دهنده سری دبی جریان با  $i$  ماه تاخیر زمانی می‌باشد.

در شکل ۵، تغییرات و پراکنندگی داده‌های مشاهداتی و پیش‌بینی شده جریان توسط مدل نزدیکترین همسایگی به ازای بهینه‌ترین ساختار مدل نشان داده شده است. در این حالت مقادیر شاخص‌های آماری شامل R، RMSE و MARE برای مدل نزدیکترین همسایگی به ترتیب برابر ۰/۷۴، ۵/۲۳ و ۲۲۹ بدست آمدند.

با توجه به شکل ۵ می‌توان دریافت که عملکرد مدل در پیش‌بینی مقادیر حداقل جریان به مراتب بهتر از پیش‌بینی مقادیر حداکثر جریان بوده است که این مساله ناشی از وجود مقادیر دبی متناظر با مقادیر حداقل جریان در همسایگی نقطه مورد پیش‌بینی می‌تواند باشد که در عملکرد مدل تاثیر زیادی دارد.

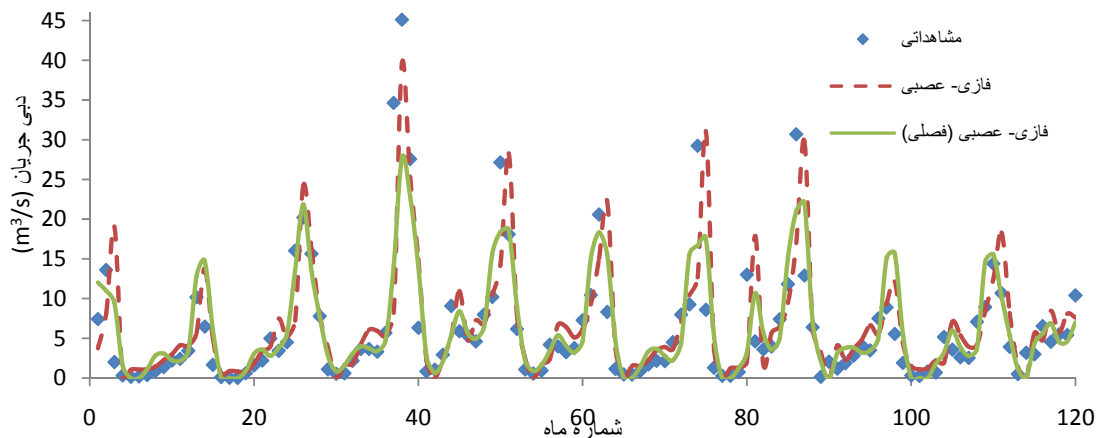
در گام بعدی، عملکرد مدل ANFIS در پیش‌بینی جریان ماهانه به ازای الگوهای ورودی مختلف ارزیابی

جدول ۲- ساختار مدل ANFIS و عملکرد آن به ازای الگوهای مختلف مورد استفاده در این مطالعه.

شماره الگو	الگوی ورودی بکار گرفته شده	شعاع کلاستر	R	RMSE (m <sup>3</sup> /s)	MARE
۱	Q(t-1)	۰/۱۹	۰/۶۲	۵/۹۹	۲۸۰/۲۵
۲	Q(t-1), $\alpha$	۰/۳۹	۰/۸۳	۴/۳۴	۹۶/۳۶
۳	Q(t-1), Q(t-2)	۰/۱۶	۰/۷۱	۵/۶۲	۱۶۰/۶۶
۴	Q(t-1), Q(t-2), $\alpha$	۰/۳۰	۰/۸۴	۴/۲۶	۸۰/۲۳
۵	Q(t-1), Q(t-2), Q(t-3)	۰/۲۹	۰/۷۱	۵/۴۹	۲۵۵/۲۶
۶	Q(t-1), Q(t-2), Q(t-3), $\alpha$	۰/۹	۰/۸۳	۴/۳۳	۲۲۱/۳۱

باتوجه به نتایج ارائه شده در جدول ۲، می‌توان نتیجه گرفت که بهترین الگوی ورودی، الگوی شماره ۴ می‌باشد که در آن از ورودی‌های  $Q(t-1)$ ,  $Q(t-2)$ ,  $\alpha$  استفاده شده است. در شکل ۶ نیز مقادیر مشاهداتی و شبیه‌سازی شده جریان توسط مدل ANFIS به ازای الگوهای شماره ۳ و ۴ نشان داده شده است.

برای تعیین اثرات پریودیک و فصلی جریان‌های ماهانه در پیش‌بینی، ضریب  $\alpha$  به الگوهای ورودی شماره ۱، ۳ و ۵ اضافه شد. با توجه به جدول ۲ می‌توان نتیجه گرفت که با اضافه کردن ضریب  $\alpha$  مقدار خطای مدل‌سازی به طور قابل توجهی کاهش یافته است.



شکل ۶- مقادیر مشاهداتی و شبیه‌سازی شده جریان با مدل ANFIS در ایستگاه دیزج، رودخانه باراندوزچای ارومیه.

جدول ۳- مقادیر پارامترهای مدل SVR به ازای الگوهای ورودی مختلف مورد استفاده در این مطالعه.

شماره الگو	الگوی بکار گرفته شده	$C_c$	$\varepsilon$	$\sigma$
۱	Q(t-1)	۱۰۰	۰/۰۰۱	۰/۴
۲	Q(t-1), $\alpha$	۱۰	۰/۰۰۰۱	۰/۹
۳	Q(t-1), Q(t-2)	۱	۰/۰۱	۰/۳
۴	Q(t-1), Q(t-2), $\alpha$	۱۰۰۰	۰/۰۰۱	۰/۱
۵	Q(t-1), Q(t-2), Q(t-3)	۱۰۰	۰/۰۱	۱/۵
۶	Q(t-1), Q(t-2), Q(t-3), $\alpha$	۱۰	۰/۰۰۱	۰/۶

توجه: علائم  $C_c$ ,  $\varepsilon$  و  $\sigma$  در جدول به ترتیب معرف ضریب ثابت، اختلاف بین مقادیر پیش‌بینی و مشاهداتی و انحراف معیار مقادیر اغتشاشات گوسی هستند.

جدول ۴- مقادیر شاخص‌های آماری به ازای الگوهای ورودی مختلف در مدل SVR.

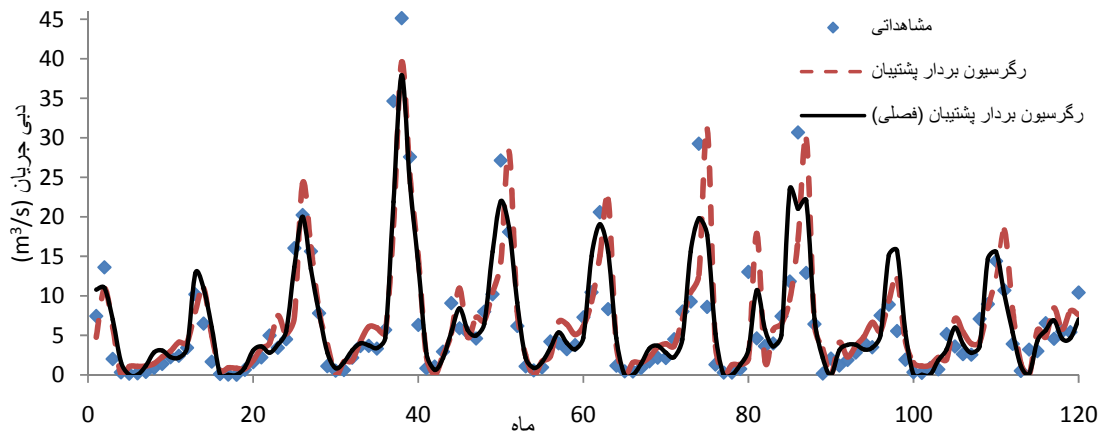
شماره الگو	الگوی بکار گرفته شده	R	RMSE (m <sup>3</sup> /s)	MARE
۱	Q(t-1)	۰/۶۷	۵/۰۱	۲۲۰/۴۱
۲	Q(t-1), $\alpha$	۰/۸۴	۴/۱۵	۸۶/۲۶
۳	Q(t-1), Q(t-2)	۰/۷۴	۵/۱۲	۱۴۰/۲۵
۴	Q(t-1), Q(t-2), $\alpha$	۰/۸۸	۳/۶۳	۷۸/۴۵
۵	Q(t-1), Q(t-2), Q(t-3)	۰/۷۳	۵/۱۹	۲۰۵/۱۸
۶	Q(t-1), Q(t-2), Q(t-3), $\alpha$	۰/۸۲	۴/۲۴	۱۸۶/۶۸

بررسی دقت مدل‌های مورد استفاده در پیش‌بینی مقادیر حداکثر جریان در ایستگاه دیزج پرداخته می‌شود. در جدول ۴ مقادیر مشاهداتی حداکثر جریان و مقادیر پیش‌بینی شده با مدل‌های مختلف و دقت پیش‌بینی مدل‌ها ارائه شده است. قابل ذکر است که داده‌های انتخابی از مجموعه داده‌های مرحله آزمون مدل بوده‌اند و نتایج ارائه شده برای مدل‌های مختلف به ازای بهینه‌ترین الگو و ساختار ارائه شده‌اند (درصد خطای نسبی نیز از رابطه  $(\frac{S_i - O_i}{O_i} * 100)$  محاسبه شد). همچنین در شکل ۸ تغییرات روند پیش‌بینی‌های انجام شده برای مقادیر حداکثر جریان با استفاده از مدل‌های مختلف در مقایسه با مقادیر مشاهداتی نشان داده شده است.

با توجه به نتایج ارائه شده در این بخش می‌توان دید که به طور کلی در پیش‌بینی‌های انجام شده با مدل‌های مختلف، مقادیر حداکثر جریان در مقایسه با مقادیر مشاهداتی کمتر برآورد شده‌اند. شاید یکی از دلایل ضعف مدل‌ها در پیش‌بینی مقادیر حداکثر ماهیت تصادفی مقادیر حداکثر باشد. برای نمونه در پیش‌بینی حداکثر مقدار جریان مشاهداتی برابر ۴۵/۱۳ مترمکعب بر ثانیه، مدل‌های نزدیکترین همسایگی، فازی-عصبی و رگرسیون بردار پشتیبان به ترتیب به میزان ۲۶، ۳۸ و ۱۶ درصد کمتر برآورد نموده‌اند. بنابراین، میتوان نتیجه گرفت که در حالت کلی عملکرد مدل رگرسیون بردار پشتیبان در مقایسه با مدل‌های دیگر بهتر بوده است.

در این مدل نیز اضافه کردن ضریب فصلی جریان موجب بهبود عملکرد مدل در پیش‌بینی جریان گردیده است. البته قابل ذکر است که استفاده از ضریب فصلی برای جریانات ماهانه کاربرد دارد و در این راستا لازم است قبل از به کارگیری مدل، پیش‌پردازشی بر روی داده‌ها (جهت بررسی فصلی یا غیرفصلی بودن آن‌ها) صورت گیرد. در مدل SVR بکارگیری ضریب فصلی (الگوی شماره ۴) در مقایسه با الگویی که در آن از ضریب فصلی استفاده نشده بود (الگوی شماره ۳) مقدار شاخص‌های آماری خطا شامل R، RMSE و MARE را به ترتیب به میزان ۱۹، ۲۹ و ۴۴ درصد بهبود داد که این مقدار در پیش‌بینی قابل توجه است.

در شکل ۷ نیز مقادیر مشاهداتی و شبیه‌سازی شده جریان توسط مدل SVR برای الگوهای ۳ و ۴ ارائه شده است. نمودارهای نشان داده شده در شکل ۷ نتایج بدست آمده در جدول ۴ را تأیید می‌کند. بنابراین، می‌توان نتیجه گرفت که عملکرد مدل SVR در پیش‌بینی جریان ماهانه در مقایسه با مدل‌های ANFIS و K-NN بهتر بوده است. همچنین قابل ذکر است که بیشترین خطای پیش‌بینی در مدل‌های مورد بررسی، مربوط به پیش‌بینی مقادیر حداکثر جریان می‌باشد. با توجه به اهمیت مقادیر حداکثر جریان جهت بکارگیری در سیستم‌های هشدار سیل و طراحی سازه‌های آبی و مطالعات منابع آب، در این بخش از تحقیق به

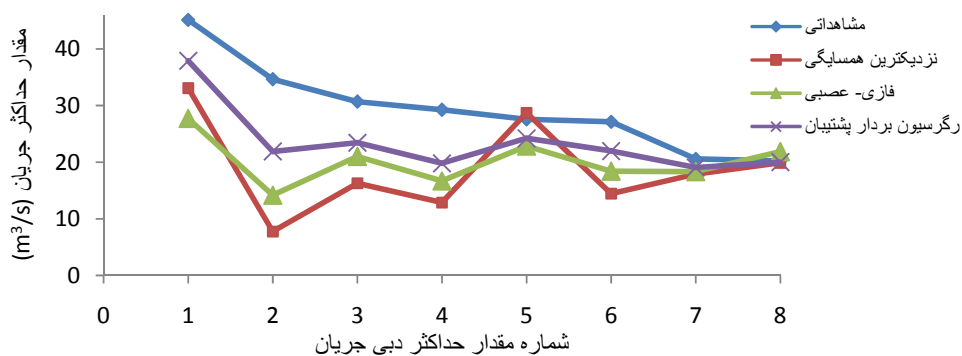


شکل ۷- مقادیر مشاهداتی و شبیه‌سازی شده جریان با مدل SVR در ایستگاه دیزج، رودخانه باراندوزچای ارومیه.

جدول ۵- پیش‌بینی مقادیر حداکثر دبی جریان رودخانه باراندوزچای با مدل‌های مورد استفاده و خطای نسبی متناظر با آن‌ها.

شماره ردیف	مقدار حداکثر جریان (m³/s)	تاریخ رویداد	K-NN (m³/s)	ANFIS (m³/s)	SVR (m³/s)	خطای نسبی K-NN	خطای نسبی ANFIS	خطای نسبی SVR
۱	۴۵/۱۳	آبان ۸۱	۳۳/۰۹	۲۷/۷۳	۳۷/۹۱	-۲۶	-۳۸	-۱۶
۲	۳۴/۶۵	مهر ۸۱	۷/۷۶	۱۴/۱۶	۲۱/۸۹	-۷۷	-۵۹	-۳۶
۳	۳۰/۶۹	آبان ۸۵	۱۶/۲۸	۲۱/۰۱	۲۳/۴۵	-۴۶	-۳۱	-۲۳
۴	۲۹/۲۵	آبان ۸۴	۱۲/۸۶	۱۶/۶۹	۱۹/۸۱	-۵۶	-۴۲	-۳۲
۵	۲۷/۵۷	آذر ۸۱	۲۸/۷۰	۲۲/۸۴	۲۴/۲۱	۴	-۱۷	-۱۲
۶	۲۷/۱۴	آبان ۸۲	۱۴/۴۵	۱۸/۸۵	۲۱/۹۸	-۴۶	-۳۲	-۱۹
۷	۲۰/۵۸	آبان ۸۳	۱۷/۸۹	۱۸/۳۳	۱۹/۰۶	-۱۳	-۱۰	-۷
۸	۲۰/۲۲	آبان ۸۰	۱۹/۸۸	۲۱/۸۷	۲۰/۰۱	-۱	۸	-۱
۱۸						۳۲	۲۷	۱۸

متوسط قدر مطلق خطا



شکل ۸- تخمین مقادیر حداکثر دبی جریان رودخانه باراندوزچای با استفاده مدل‌های مختلف در دوره آزمون.

پیش‌بینی قابل قبولی در مقایسه با سایر روش‌ها ارائه نموده و مقدار آن منطبق بر مقدار مشاهداتی می‌باشد که دلیل آن می‌تواند وجود نقاط همسایگی متناظر با این دبی در مجموعه داده‌های مورد بررسی باشد. نتایج

اگرچه به صورت موردی در پاره‌ای از موارد تخمین روش نزدیکترین همسایگی در مقایسه با دو روش دیگر تطابق بیشتری با مقادیر مشاهداتی داشته است. به ازای حداکثر جریان برابر ۲۷/۵۷ روش K-NN

مقایسه با الگوی شماره ۳ مقدار RMSE در مدل‌های ANFIS و SVR به ترتیب به میزان ۲۴ و ۲۹ درصد بهبود یافت. با توجه به اهمیت مقادیر حداکثر جریان در مهندسی منابع آب، در بخشی دیگر از این تحقیق، به بررسی عملکرد مدل‌های مورد بررسی در پیش‌بینی مقادیر حداکثر جریان پرداخته شد. در این خصوص، مدل SVR با داشتن متوسط قدر مطلق خطا برابر ۱۸ درصد بهترین عملکرد را در پیش‌بینی مقادیر حداکثر جریان داشت و مدل‌های ANFIS و K-NN در اولویت‌های بعدی قرار گرفتند.

در تعمیم نتایج به دست آمده در این تحقیق باید نکاتی را مدنظر قرار داد. با توجه به وجود گرمایش جهانی و اینکه بخشی از دبی جریان در حوضه ناشی از ذوب برف است (که در سال‌های آتی مقدار انباشت برف در حوزه کمتر از سال‌های قبل خواهد بود)، این واقعیت در حافظه سیستم تاثیر گذاشته و در نتیجه روی دقت مدل بسط داده شده موثر خواهد بود. بنابراین، به نظر می‌رسد چنین مطالعاتی می‌تواند در یک دوره معین (مثلاً ۵ ساله) تجدید گردد. از دیگر نکات قابل توجه در این تحقیق این بود که کلیه مدل‌های مورد بررسی در پیش‌بینی مقادیر حداکثر جریان، تخمین کمتری نسبت به مقایسه‌های مشاهده‌ای داشتند و عملکرد آن‌ها در پیش‌بینی مقادیر حداقل و متوسط جریان به مراتب قابل قبول‌تر بود.

بررسی عملکرد روش‌های مورد استفاده در این تحقیق جهت پیش‌بینی‌های مکانی نظیر استفاده از داده‌های ایستگاه‌های مجاور جهت پیش‌بینی میزان جریان در یک ایستگاه معین می‌تواند موضوع تحقیقات آتی باشد که این مساله در حوضه‌های فاقد آمار اهمیت فراوانی دارد.

بدست آمده در این تحقیق با مطالعه پانده و همکاران (۲۰۰۹) مطابقت دارد اگرچه در مطالعه آن‌ها از سری داده‌های روزانه مقادیر جریان در پیش‌بینی استفاده شده بود، اما آن‌ها نیز در مطالعه خود به کارایی کمتر روش K-NN در پیش‌بینی مقادیر حداکثر جریان اشاره نموده‌اند. همچنین مقادیر پارامترهای بدست آمده برای مدل SVR و شاخص‌های آماری خطا در این تحقیق در تطابق با نتایج کیشی و سیمن (۲۰۱۱) می‌باشد. اگرچه جهت مقایسه دقیق‌تر لازم است تا خصوصیات آماری مقادیر جریان و مشخصات حوضه‌های مورد بررسی یکسان باشد.

### نتیجه‌گیری کلی

در این مطالعه به بررسی عملکرد روش‌های ANFIS، SVR و K-NN در پیش‌بینی جریان ماهانه رودخانه پرداخته شد. در این راستا از داده‌های دبی ماهانه رودخانه باراندوزچای در ایستگاه دیزج در شمال‌غرب کشور استفاده شد. از الگوهای ورودی مختلف شامل مقادیر دبی‌های پیشین جهت مدل‌سازی دبی استفاده شد. نتایج به دست آمده حاکی از عملکرد قابل قبول روش‌های مورد استفاده در پیش‌بینی مقادیر جریان ماهانه بود. مقدار ضرایب R و RMSE برای روش K-NN در حالت بهینه به ترتیب برابر ۰/۷۳ و ۵/۲۳ بدست آمدند. همچنین در این تحقیق به بررسی اثرات ضریب فصلی در عملکرد مدل‌های ANFIS و SVR پرداخته شد. نتایج نشان داد که اضافه نمودن ضریب فصلی به عنوان یکی از متغیرها در الگوهای ورودی موجب بهبود عملکرد مدل‌ها می‌گردد.

در مدل‌های هوشمند بهترین الگوی ورودی الگوی شماره ۴ بود که در آن از متغیرهای دبی پیشین با دو گام تاخیر و ضریب فصلی استفاده شده بود. با اضافه نمودن ضریب فصلی (الگوی شماره ۴) در

### منابع مورد استفاده

- سلاجقه ع، غلامی ح و فتح آبادی ا، ۱۳۸۸. پیش‌بینی دبی جریان با استفاده از روش نزدیکترین همسایه. صفحه‌های ۱۵۱ تا ۱۵۸. پنجمین همایش علوم و مهندسی آب‌خیزداری ایران، ۳-۲ اردیبهشت، انجمن آب‌خیزداری، کرج.
- عزمی م و عراقی نژاد ش، ۱۳۹۱. توسعه روش رگرسیون K- نزدیکترین همسایگی در پیش‌بینی جریان رودخانه. نشریه آب و فاضلاب، شماره ۲، صفحه‌های ۱۰۸-۱۱۹.

- Chang FJ, and Chen YC, 2001. Counter propagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology* 245:153-164.
- Chiu S, 1997. Extracting Fuzzy Rules From Data for Function Approximation and Pattern Classification. Pp. 149-162. In: Dubois D, Prade H and Yager R (eds). *Fuzzy Information Engineering: A Guided Tour of Applications*. Springer, Berlin.
- Cimen M, 2008. Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal* 53 (3): 656-666.
- Cobaner M, 2011. Evapotranspiration estimation by two different neuro-fuzzy inference systems. *Journal of Hydrology* 398: 292-302.
- Dasarathy BV, 1991. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, Los Alamitos, CA.
- Firat M and Gungor M, 2007. River flow estimation using adaptive neuro fuzzy inference system. *Mathematics and Computers in Simulation* 75(3): 87-96.
- Jang JSR, 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23(3): 665-685.
- Kember G, Flower AC and Holubeshen J, 1993. Forecasting river flow using nonlinear dynamics. *Stochastic Hydrology and Hydraulics* 7(3): 205-212.
- Kisi O, 2007. Streamflow forecasting using different artificial neural network algorithms. *ASCE Journal of Hydrologic Engineering* 12 (5):532-539.
- Kisi O, 2008. River flow forecasting and estimation using different artificial neural network techniques. *Hydrology Research* 39(1): 27-40.
- Kisi O and Cimen M, 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology* 399(1): 132-140.
- Lohani AK, Kumar R and Singh RD, 2012. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology* 442-443: 23-35.
- Maier HR and Dandy G, 2000. Neural networks for prediction and forecasting of water resources variables: are view of modelling issues and applications. *Environmental Modelling & Software* 15(10):1-124.
- Mousavia SJ, Ponnambalam K and Karray F, 2007. Inferring operating rules for reservoir operations using fuzzy regression and ANFIS. *Fuzzy Sets and Systems* 158: 1064-1082.
- Nayak PC, Sudheer KP, Rangan DM and Ramasastri KS, 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* 29: 52-66.
- Nourani V, Kisi O and Komasi M, 2011. Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *Journal of Hydrology* 402: 41-59.
- Pande S, McKee M and Bastidas LA, 2009. Complexity-based robust hydrologic prediction. *Water Resources Research* 45: W10406.
- Piotrowski A, Napiorkowski JJ and Rowinski PM, 2006. Flash-flood forecasting by means of neural networks and nearest neighbour approach- a comparative study. *Nonlinear Processes in Geophysics* 13(4): 443-448.
- Sharma A and Lall U, 1999. A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation* 48: 361-371.
- Vapnik V, 1995. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, USA.
- Wang WC, Chau KW, Cheng CT and Qiu L, 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* 374: 294-306.
- Wu CL and Chau KW, 2010. Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence* 23:1350-1367.