

بررسی دقت مدل‌های ANFIS، SVM و GP در مدل‌سازی مقادیر دبی جریان رودخانه

محمد ناظری تهرودی^۱، سید رضا هاشمی^{۲*}، فرشاد احمدی^۳، زهرا ناظری تهرودی^۴

۱. دانشجوی دکتری منابع آب، دانشکده کشاورزی، دانشگاه بیرجند

۲. استادیار دانشکده کشاورزی، دانشگاه بیرجند

۳. دانشجوی دکتری منابع آب، دانشکده علوم آب، دانشگاه شهید چمران اهواز

۴. دانشجوی دکتری آبخیزداری، دانشکده کشاورزی و منابع طبیعی، دانشگاه کاشان

(تاریخ دریافت ۱۳۹۵/۰۸/۲۹؛ تاریخ تصویب ۱۳۹۵/۰۹/۲۹)

چکیده

پیش‌بینی دقیق جریان رودخانه‌ها در مدیریت بهینه منابع آب‌های سطحی اهمیت به‌سزایی دارد. یافتن مدل مناسب برای پیش‌بینی دقیق این پارامتر یکی از راه‌های مهم اقدامات در شبیه‌سازی و پیش‌بینی است. در این مطالعه سه مدل ANFIS، SVM و GP برای مدل‌سازی دبی ماهانه رودخانه نازلوچای در محل ایستگاه هیدرومتری تپیک واقع در غرب دریاچه ارومیه تحت تأثیر بارش حوضه رودخانه مطالعه شده بررسی و مقایسه شد. در همه روش‌های یادشده الگوهای M1 تا M5 داده‌های دبی جریان با تأخیر یک تا پنج و الگوهای M6 تا M10 الگوی ترکیبی با داده‌های بارش و دبی و با تأخیرهای یک تا پنج ماه بررسی شدند. برای بررسی مقادیر خطای ناشی از مدل‌سازی از سه روش ضریب تبیین، مجذور میانگین مربعات خطا و معیار کارایی مدل استفاده شد. نتایج بررسی دقت و میزان خطای مدل‌ها نشان داد الگوی ترکیبی فقط در مدل SVM بهترین نتیجه را داده است و در دو مدل GP و ANFIS الگوهای تک‌سری بهترین نتیجه را ارائه کردند. از بین سه مدل بررسی شده، مدل ANFIS با الگوی ورودی چهار و پنج تأخیر بهترین نتیجه را داد. به‌طور کلی، نتایج نشان داد با به‌کارگیری مدل ANFIS در مدل‌سازی دبی جریان ماهانه رودخانه نازلوچای، خطای مدل نسبت به دو مدل GP و SVM به ترتیب حدود ۲۳ و ۳ درصد (در واحد دبی جریان) کاهش و دقت مدل نیز نسبت به دو مدل GP و SVM به ترتیب حدود ۱۰ و ۴ درصد افزایش می‌یابد.

کلیدواژگان: الگوریتم ژنتیک، دبی جریان، دریاچه ارومیه، ماشین بردار پشتیبان.

مقدمه

ANFIS^۳، برنامه‌ریزی ژنتیک و SVM^۴ برای پیش‌بینی سری زمانی دبی جریان در مقیاس ماهانه پرداختند. نتایج آنها نشان داد به‌ازای داده‌های استفاده‌شده در مراحل آموزش و آزمون، بهترین عملکرد با توجه به معیارهای مختلف ارزیابی خطا به مدل‌های ANFIS، GP و SVM اختصاص دارد. گسوون (۲۰۰۹) برای پیش‌بینی دبی جریان روزانه رودخانه شویل کیل در ایالات متحده، روش برنامه‌ریزی ژنتیک خطی را به کار برد و نتایج به‌دست‌آمده را با دو الگو از روش‌های شبکه عصبی مقایسه کرد. وی نشان داد هر دو روش نتایج قابل قبولی داشته‌اند، اما روش برنامه‌ریزی ژنتیک خطی دقت بیشتری نسبت به روش‌های شبکه عصبی داشت. قربانی و همکارانش (۲۰۱۰) عملکرد سه روش برنامه‌ریزی ژنتیک، شبکه عصبی مصنوعی و نروفازی^۵ را در روندیابی سیلاب رودخانه قزل ایرماق ترکیه ارزیابی کردند. نتایج مطالعه آنها نشان داد از بین سه روش یادشده، مدل برنامه‌ریزی ژنتیک با دقت بیشتری هیدروگراف خروجی را شبیه‌سازی می‌کند. ظهیری و عظمت‌الله (۲۰۱۴) برای پیش‌بینی دبی جریان در مقاطع مرکب از دو روش برنامه‌ریزی ژنتیک خطی و مدل درختی M5 استفاده کردند. نتایج نشان داد هرچند هر دو مدل دقت زیادی برای پیش‌بینی جریان داشتند، اما دقت روش برنامه‌ریزی ژنتیک خطی بیشتر از مدل درختی M5 بود. ماشین بردار پشتیبان نیز یکی از روش‌های یادگیری تحت نظارت^۶ است که هم برای دسته‌بندی و هم رگرسیون قابل استفاده است. این روش توسط وپنیک (۱۹۹۸) بر پایه تئوری یادگیری آماری^۷ بنا نهاده شده است. ماشین بردار پشتیبان روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این روش مناسب برای مسائل پیش‌بینی به‌شمار می‌رود [۱۰]. ماشین بردار پشتیبان در اصل یک دسته‌بندی‌کننده دوکلاسه است که کلاس‌ها را توسط نوعی مرز خطی از هم جدا می‌کند. در این روش نزدیک‌ترین نمونه‌ها به مرز تصمیم‌گیری را بردارهای پشتیبان می‌نامند. این بردارها معادله مرز تصمیم‌گیری را مشخص می‌کنند. الگوریتم‌های شبیه‌سازی هوشمند کلاسیک مانند شبکه‌های عصبی مصنوعی، معمولاً قدر مطلق خطا یا مجموع مربعات

نیاز روزافزون به آب سبب شده است که برنامه‌ریزی‌های مدیریتی به‌منظور کنترل مصرف آب در آینده اهمیت بیشتری داشته باشد. با پیش‌بینی جریان رودخانه‌ها علاوه بر مدیریت بهره‌برداری از منابع آب، می‌توان حوادث طبیعی نظیر سیل و خشکسالی را نیز پیش‌بینی و مهار کرد. همچنین می‌توان از نتایج نوعی مدل شبیه‌سازی به‌منظور بررسی صحت داده‌ها و یا اصلاح و تکمیل آنها استفاده کرد. گوویندراجو (۲۰۰۰) معتقد است مدل‌هایی که امروزه در هیدرولوژی استفاده می‌شوند، شامل مدل‌های ریاضی-فیزیکی؛ ژئومورفولوژیکی و تجربی هستند. دسته اول مدل‌ها، براساس خصوصیات فیزیکی سیستم هستند که به‌صورت معادلات دیفرانسیل بیان می‌شوند، اما دسته دوم برمبنای خصوصیات ژئومورفولوژیکی سیستم هیدرولوژیکی مد نظر بیان می‌شوند. مدل‌های تجربی بدون توجه به پارامترها، سعی در ایجاد رابطه‌ای بین داده‌های ورودی و خروجی دارند که این مدل‌ها به مدل‌های جعبه سیاه یا میانگینی نیز معروف‌اند. تا کنون محققان مختلفی در سراسر جهان با روش‌های مختلف اقدام به پیش‌بینی جریان رودخانه‌ها در مقیاس‌های زمانی ساعتی، روزانه و ماهانه کرده‌اند. برنامه‌ریزی ژنتیک^۱ شاخه‌ای از الگوریتم‌های تکاملی است که توانایی مدل‌سازی فرایندهای کاملاً غیرخطی و پویا را دارد. روش برنامه‌ریزی ژنتیک نخستین بار توسط کوزا (۱۹۹۲) توسعه داده شد. این روش جزء روش‌های الگوریتم تکاملی محسوب می‌شود که مبنای آنها براساس نظریه تکامل داروین استوار است. الگوریتم‌های یادشده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی می‌کنند و سپس تابع یادشده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرایند گام‌به‌گام تصحیح ساختار داده‌ها به‌کار می‌گیرند و درنهایت، جواب مناسب را ارائه می‌دهند. روش برنامه‌ریزی ژنتیک جدیدترین شیوه از بین روش‌های الگوریتم تکاملی است که به‌دلیل داشتن دقت کافی، کاربرد بیشتری دارد [۳]. طبق تحقیقات آیتک و همکارانش (۲۰۰۸) روش برنامه‌ریزی ژنتیک روشی مناسب و علمی در پیش‌بینی روابط بارش-رواناب است. وانگ و همکارانش (۲۰۰۹) به مقایسه عملکرد مدل خودهمبسته و چندین مدل هوشمند شامل شبکه عصبی مصنوعی^۲،

3. Adaptive Neuro Fuzzy Inference System (ANFIS)

4. Support Vector Machine (SVM)

5. Neuro-Fuzzy

6. Supervised Learning

7. Statistical Learning Theory

1. Genetic Programming

2. Artificial Neural Network (ANN)

نروفازی در پیش‌بینی سیل رودخانه‌ای در مرکز تایوان بهره بردند. آنها از آمار بارش و دبی به‌عنوان ورودی مدل نروفازی استفاده کردند. نتایج نشان داد روش نروفازی دقت قابل قبولی در پیش‌بینی داشته و استفاده از بارندگی موجب کاهش خطای مدل شده است. هی و همکارانش (۲۰۱۴) در پژوهشی دیگر برای پیش‌بینی جریان رودخانه در منطقه‌ای کوهستانی و نیمه‌خشک واقع در شمال غربی چین از ماشین بردار پشتیبان استفاده کردند و دریافتند که مدل ماشین بردار پشتیبان عملکردی مناسب به‌منظور پیش‌بینی جریان رودخانه در مناطق کوهستانی نیمه‌خشک دارد و نسبت به مدل‌های شبکه عصبی مصنوعی و سیستم استنتاج فازی دقت مطلوبی دارد. با توجه به موارد یادشده می‌توان دریافت که تخمین پارامترهای هیدرولوژیکی همچون پیش‌بینی جریان رودخانه‌ها از دیرباز مورد توجه محققان امر بوده است و بدین‌منظور روش‌های متعددی از جمله مدل‌های تجربی - نیمه‌تجربی، سری‌های زمانی و مدل‌های هوشمند توسعه یافته‌اند که در این میان مدل‌های هوشمند با الهام‌گرفتن از طبیعت قادر به تخمین پارامترهای مختص به پدیده‌های طبیعی با دقت قابل قبول هستند و نیز دقت شایان توجهی نسبت به سایر روش‌ها دارند. بنابراین، در این تحقیق سعی شده است که عملکرد سه مدل هوشمند برنامه‌ریزی ژنتیک (GP)، ANFIS و SVM در پیش‌بینی جریان ماهانه رودخانه نازلوچای با استفاده از داده‌های بارش ماهانه منطقه مطالعه‌شده بررسی و مقایسه شود.

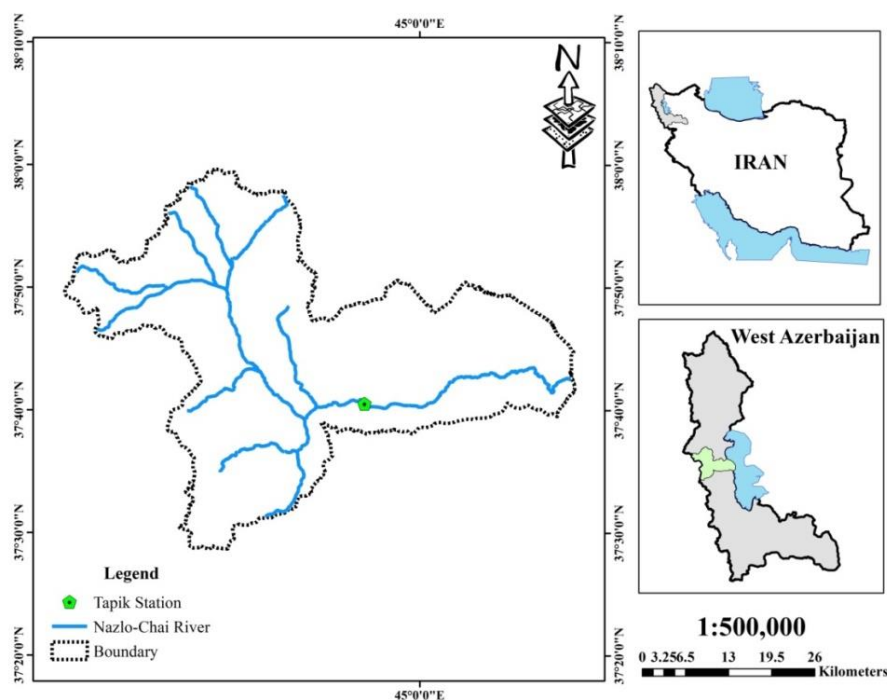
مواد و روش‌ها

منطقه مطالعه‌شده

منطقه بررسی‌شده در این مطالعه حوضه آبریز نازلوچای در استان آذربایجان غربی است که در شکل ۱ قابل مشاهده است. حوضه آبریز نازلوچای در قسمت غربی دریاچه ارومیه واقع شده است. این حوضه با مساحتی برابر ۱۸۸۱/۶۲ کیلومتر مربع و محیطی برابر ۲۹۶/۴۶ کیلومتر در طول جغرافیایی ۱° ۴۹' ۴۴" و عرض جغرافیایی ۱° ۴۰' ۳۷" واقع شده است. طول رودخانه اصلی در این حوضه برابر ۱۱۰/۹۴ کیلومتر، شیب متوسط رودخانه برابر ۳۵ درصد، کمینه ارتفاع حوضه ۱۲۹۴ متر، بیشینه ارتفاع حوضه ۳۵۷۶ متر و میانگین ارتفاع حوضه برابر ۱۸۳۴ متر است.

خطای داده‌های آموزشی را حداقل می‌کنند، ولی مدل‌های SVM اصل حداقل‌سازی خطای ساختاری را به‌کار می‌گیرند [۱۱]. امروزه این مدل‌ها در گستره وسیعی از مسائل هیدرولوژیکی و به‌ویژه پیش‌بینی جریان رودخانه‌ها به‌کار رفته‌اند. در مطالعه‌ای آداموفسکی و پراشر (۲۰۱۲) برای پیش‌بینی جریان روزانه در حوضه کوهستانی سیانجی^۱ واقع در منطقه هیمالیای هندوستان از دو روش رگرسیون ماشین بردار پشتیبان و شبکه‌های موجک^۲ استفاده کردند. مقایسه نتایج پژوهش آنها نشان داد گرچه هر دو روش دقت خوبی در پیش‌بینی جریان روزانه داشتند، اما دقت روش شبکه‌های موجک اندکی بیشتر بود. چویی و همکارانش (۲۰۱۴) از مدل SVM برای پیش‌بینی و داده‌کاوی جریان ورودی به سد نارمادا^۳ واقع در ایالت مادهاپا پرادش^۴ هند استفاده کردند و نشان دادند این روش توانایی بسیار خوبی در مدل‌سازی و پیش‌بینی جریان میان‌مدت ماهانه دارد. امروزه با ترکیب موفقیت‌آمیز بعضی روش‌ها همچون شبکه عصبی، منطق فازی و محاسبات تکاملی، روش‌های جدیدی به نام محاسبات نرم و هوشمند ایجاد شده‌اند و توسعه یافته‌اند. این تکنیک‌های نرم در زمینه‌های مختلف مهندسی به‌کار می‌روند. نروفازی نوعی سیستم هیبریدی مرکب از توانایی تصمیم‌گیری منطق فازی با قابلیت حسابگری شبکه عصبی است و سطح پیچیده و بالایی را برای مدل‌کردن و تخمین پیشنهاد می‌کند. بر این اساس قسمت فازی مختص به گروه‌بندی داده‌های ورودی به مجموعه‌هایی که با درجه عضویت مشخص می‌شوند (که می‌تواند هر عددی بین ۰ و ۱ باشد) و تصمیم‌گیری برای فعالیت بعدی براساس یک سری قوانین و حرکت به مرحله بعدی است. ابتدا سیستم استنتاج عصبی- فازی تطبیقی توسط جانگ (۱۹۹۳) معرفی و سپس به‌عنوان یکی از ابزارهای تقریب‌ساز توابع حقیقی پیوسته (و با هر مقدار دلخواه درجه دقت) در سطح جهانی استفاده شد. این سیستم از نظر عملکرد مشابه سیستم‌های استنتاج فازی است. نویاک و همکارانش (۲۰۰۴) در هند، ثانی‌خانی و کیسی (۲۰۱۲)، طالعی و همکارانش (۲۰۱۰) و گوپیس و همکارانش (۲۰۱۳) روش نروفازی را برای پیش‌بینی جریان مناسب تشخیص دادند. چن و همکارانش (۲۰۰۶) از روش

1. Sianji
2. Wavelet Networks
3. Narmada
4. Madhya Pradesh



شکل ۱. موقعیت منطقه مطالعه شده و ایستگاه هیدرومتری منتخب

جدول ۱. مشخصات آماری داده‌های استفاده شده

نوع ایستگاه	ایستگاه	دوره آماری	تعداد داده‌ها	متوسط	کمینه	بیشینه	انحراف معیار	چولگی
هیدرومتری	تپیک	۲۰۱۴-۱۹۷۱	۵۱۶ ماه	۱۱۱/۷ (m ³ s ⁻¹)	۰/۰۱ (m ³ s ⁻¹)	۱۰۲/۰ (m ³ s ⁻¹)	۱۵/۶	۲/۶
سینوپتیک	تپیک	۲۰۱۴-۱۹۷۱	۵۱۶ ماه	۲۷/۳ (mm)	۰/۰ (mm)	۱۴۸/۶ (mm)	۲۴/۶	۱/۴

تطبیقی بیشتر با استفاده از سیستم فازی سوگنو و به صورت ساختار شبکه‌ای پیش‌رونده استفاده می‌شود که پنج لایه دارد (شکل ۲) [۲۱]. سامانه فازی، منطبق بر قوانین فازی است. هدف از منطق فازی، احتمال کالسیک تحلیل‌های نظریه استخراج نتایج دقیق با استفاده از مجموعه قوانینی است که افراد خبره و متخصص آن را تعریف کرده‌اند. از طرفی، شبکه‌های عصبی قابلیت آموزش و یادگیری دارند و می‌توانند با استفاده از داده‌های مشاهده شده، پارامترهای شبکه را به گونه‌ای معین کنند که به ازای ورودی دلخواه، خروجی مطلوب به دست آید. در عین حال، شبکه‌های عصبی قابلیت استفاده از دانش بشری را ندارند و نمی‌توانند مانند سامانه‌های فازی با استفاده از عبارتهای زبانی نتیجه‌گیری کنند [۲۱].

روش سیستم استنتاج فازی-عصبی تطبیقی (ANFIS)
روش سیستم استنتاج فازی-عصبی تطبیقی، روشی ترکیبی است که در آن بخش فازی رابطه بین متغیرهای ورودی و خروجی را برقرار می‌کند و مشخصه‌های مختص به توابع عضویت بخش فازی نیز به وسیله شبکه عصبی تعیین می‌شود. در سیستم استنتاج فازی-عصبی تطبیقی، ابتدا ساختار مدل با پارامترهای مشخص متناسب با ورودی‌ها، درجه عضویت و قوانین و توابع درجه عضویت خروجی، انتخاب می‌شود. سپس بخشی از داده‌های موجود به عنوان ورودی-خروجی انتخاب می‌شود که قابل استفاده برای آموزش این سیستم باشد. در مرحله آموزش با اصلاح مشخصه‌های درجه عضویت براساس میزان خطای قابل قبول، مشخصه‌های مدل به مقادیر واقعی نزدیک می‌شوند. سیستم استنتاج فازی-عصبی

لایه درجه عضویت تخصیص داده شده به متغیرهای ورودی در مجموعه‌های فازی است که به شکل زیر بیان می‌شود:

$$i = 1, 2 \quad O_{\nu, i} = \mu_{A_i}(X) \quad (3)$$

$$i = 3, 4 \quad O_{\nu, i} = \mu_{B_i-r}(X) \quad (4)$$

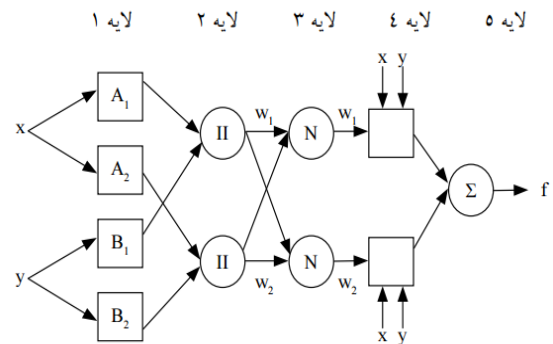
در روابط ۳ و ۴، x و y ورودی‌های غیرفازی گره A_i و B_i و توابع عضویت فازی هستند. همچنین، مقدار عضویت هر ورودی به عنوان خروجی اول با x (μ_{A_i}) و y (μ_{B_i}) مشخص می‌شود. بنابراین، باید پارامترهای توابع عضویت مشخص شوند که به عنوان پارامترهای بخش مقدم قوانین فازی شناخته می‌شوند و از دسته پارامترهای غیرخطی هستند. در ساختار ANFIS معمولاً از فازی‌سازی گوسی^۱ استفاده می‌شود که یکی از انواع فازی‌سازها در سامانه‌های فازی است. به دلیل اینکه سامانه‌های واقعی و کاربردهای مهندسی همواره با متغیرهای عددی سروکار دارند، برای اینکه بتوان از سامانه فازی برای توصیف یا مدل‌سازی هر سامانه واقعی استفاده کرد، باید نگاشتی بین ورودی‌های عددی (غیرفازی) سامانه و یک مجموعه فازی ایجاد کرد. به همین منظور از فازی‌سازها استفاده می‌شود. به بیان دیگر، فازی‌ساز قابلیت آن را دارد تا به شکل یک رابطه، هر ورودی غیرفازی (عددی) را به یک ورودی فازی تبدیل کند. این نوع فازی‌سازها به طور کامل در مرجع (مندانی و عصیلان، ۱۹۷۵) توصیف شده‌اند. لایه دوم شامل گره‌های قوانین است. هر گره در این لایه درجه فعالیت یک قانون را محاسبه می‌کند. در این لایه از عملگر و برای محاسبه درجه مشارکت هر قانون استفاده می‌شود. $O_{2,k}$ نمایانگر خروجی گره k ام در لایه دوم و برابر حاصل ضرب درجه عضویت هر ورودی است.

$$O_{2, k} = \mu_{A_i}(x) * \mu_{B_j}(y) \quad (5)$$

لایه سوم شامل گره‌های نرمال شده است که نسبت درجه مشارکت هر قانون به مجموع درجه مشارکت همه قوانین را محاسبه می‌کند. در نتیجه، این لایه به شکل رابطه ۶ تعریف می‌شود:

$$O_{\nu, i} = \bar{w}_i = \frac{w_i}{\sum_{k=1}^r w_k} \quad (6)$$

لایه چهارم مشتمل بر گره‌های تطبیقی است که با استفاده از پارامترهای نتیجه، خروجی هر گره را محاسبه می‌کند و به شکل رابطه ۷ تعریف می‌شود:



شکل ۲. ساختار شبکه ANFIS

بنابراین، برای دستیابی به قابلیت یادگیری بهتر، دقت تقریب بیشتر و ساختار ساده‌تر، با الهام از قابلیت یادگیری شبکه‌های عصبی و خواص استنباط سامانه‌های فازی در مدل فازی، شبکه‌های فازی-عصبی تطبیقی موسوم به ANFIS در سال ۱۹۹۹ ارائه شد. این شبکه‌ها ضمن داشتن قابلیت یادگیری شبکه‌های عصبی و قدرت استنتاج سامانه‌های فازی، توانایی یافتن هر نوع مدل یا نگاشت غیرخطی را دارند که بتوانند به طور دقیق، ورودی‌ها (مقادیر اولیه) را با خروجی پیش‌بینی شده مرتبط سازند. بنابراین، ANFIS مبتنی بر سامانه‌های فازی است که ساختار آن در شکل ۲ نشان داده شده است. در این ساختار، گره‌های ورودی و خروجی به ترتیب مقادیر پیش‌بینی شده را نشان می‌دهند. برای ساده‌سازی، ورودی و مقادیر یک شبکه، دو ورودی-تک خروجی در نظر گرفته شده است. برای نمونه دو قانون در اینجا آورده شده است:

قانون نخست: اگر x برابر A_1 و y برابر A_1 باشد، آن‌گاه:

$$f_1 = p_1 x + q_1 y + r_1 \quad (1)$$

قانون دوم: اگر x برابر A_2 و y برابر A_2 باشد، آن‌گاه:

$$f_2 = p_2 x + q_2 y + r_2 \quad (2)$$

همان طور که در شکل ۲ مشاهده می‌شود، ANFIS شبکه‌ای پنج‌لایه است و هر لایه گره‌های مختلف دارد و هر گره در یک لایه ثابت یا تطبیقی است. لایه‌های مختلف با گره‌های متناظر آنها به طور خلاصه در اینجا توصیف شده‌اند. لایه نخست، گره‌های ورودی هستند. هر گره از این لایه، مقادیر عضویتی است که به هر یک از متغیرهای ورودی مدل (x و y) نسبت داده می‌شود. مقادیر عضویت بر اساس تعلق ورودی‌ها به هر یک از مجموعه‌های فازی A_1 و B_1 معین می‌شود. به بیان دیگر، خروجی هر گره در این

ماشین بردار پشتیبان

نخستین کاربرد این روش در مسائل آب توسط دیباک و همکارانش در سال ۲۰۰۱ با مدل سازی بارندگی - رواناب ارائه شد [۲۴]. ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه سازی مقید است که از اصل استقرای کمینه سازی خطای ساختاری استفاده می کند و به یک جواب بهینه کلی منجر می شود. در مدل رگرسیون SVM تابعی مرتبط با متغیر وابسته Y برآورد می شود که خود تابعی از چند متغیر مستقل x است. مشابه سایر مسائل رگرسیونی فرض می شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود (رابطه ۱۱) [۲۵].

$$f(x) = W^T \cdot \phi(x) + b \quad (10)$$

$$y = f(x) + noise \quad (11)$$

چنانچه W (بردار ضرایب) و b (ثابت) مشخصه های تابع رگرسیونی و ϕ نیز تابع کرنل باشد، آن گاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این هدف با آموزش مدل SVM توسط مجموعه ای از نمونه ها (مجموعه آموزش) محقق می شود. بنابراین، برای محاسبه w و b لازم است تابع خطا (رابطه ۱۲) در مدل SVM - ε با در نظر گرفتن شرایط مندرج در رابطه ۱۳ بهینه شود.

$$\frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (12)$$

$$W^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*$$

$$y_i - W^T \cdot \phi(x_i) - b \leq \varepsilon + \xi_i \quad (13)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N$$

در روابط یاد شده C عددی صحیح و مثبت است که عامل تعیین جریمه به هنگام رخ دادن خطای آموزش مدل است. ϕ : تابع کرنل^۱، N تعداد نمونه ها و دو مشخصه ξ_i و ξ_i^* متغیر کمبود^۲ هستند که حد بالا و پایین خطای آموزش مختص به مقدار خطای مجاز ε را مشخص می کنند. در مسائل پیش بینی می شود که داده ها، درون بازه ε قرار گیرند. حال اگر داده ای خارج از بازه ε قرار گرفت، یک خطا معادل

$$O_{r,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (7)$$

که w_i خروجی i امین گره از لایه قبلی است و $\{p_i, r_i, q_i\}$ پارامترهای تطبیقی خطی هستند.

لایه پنجم مشتمل بر گره خروجی است که مقدار نهایی خروجی را به شکل مجموع خروجی گره های لایه قبلی بیان می کند:

$$O_{\delta,i} = \bar{w}_i f_i = \frac{\sum_{i=1}^q w_i f_i}{\sum_{i=1}^q w_i} \quad (8)$$

الگوریتم یادگیری شبکه ANFIS نوعی الگوریتم یادگیری هیبریدی مشتمل بر الگوریتم گرادینان نزولی و روش حداقل مربع بازگشتی است. از الگوریتم گرادینان نزولی برای به روزرسانی پارامترهای غیرخطی شبکه و از تخمین زن حداقل مربع بازگشتی برای تنظیم وزن های شبکه استفاده می شود. خطای آموزش شبکه به صورت رابطه ۹ تعریف شده است:

$$E = \sum_{i=1}^N (f_i - \hat{f}_i)^2 \quad (9)$$

که در آن f_i و \hat{f}_i به ترتیب خروجی مطلوب و تخمین زده شده شبکه به ازای i امین ورودی هستند. همچنین، N تعداد کل زوج داده های ورودی - خروجی (داده های آموزشی) شبکه است [۲۲].

برنامه ریزی بیان ژن

روش برنامه ریزی بیان ژن، در سال ۱۹۹۹ توسط فی-یرا ارائه شد [۲۳]. این روش ترکیبی از روش های برنامه ریزی ژنتیک و الگوریتم ژنتیک است که در آن، کروموزوم های خطی و ساده با طول ثابت، مشابه با آنچه در الگوریتم ژنتیک استفاده می شود و ساختارهای شاخه ای با اندازه ها و اشکال متفاوت، مشابه درختان تجزیه در برنامه ریزی ژنتیک، ترکیب می شوند. در این روش پدیده های مختلف با استفاده از مجموعه ای از توابع و ترمینال ها، مدل سازی می شوند. مجموعه توابع معمولاً شامل توابع اصلی حسابی $\{+, -, \times, /\}$ ، توابع مثلثاتی یا هر نوع تابع ریاضی دیگر $\{\sqrt{\quad}\}$ ، $\{x^2, \exp, \log, \sin, \cos, \dots\}$ و یا توابع تعریف شده توسط کاربر است که برای تفسیر مدل مناسب اند. مجموعه ترمینال ها، از مقادیر ثابت و متغیرهای مستقل تشکیل شده اند [۲۳].

1. Kernel Function
2. Slack Variable

بنابراین، با معرفی ۲ ضریب لاگرانژ a_i^* و a_i مسئله بهینه‌سازی با حداکثرسازی عددی تابع درجه دوم زیر (معادله ۱۶) با شرایط معادله ۱۷ حل خواهد شد [۲۵].

$$\sum_{i=1}^N y_i (a_i + a_i^*) - \varepsilon \sum_{i=1}^N (a_i + a_i^*) - 0.5 \sum_{i,j=1}^N (a_i + a_i^*)(a_j + a_j^*) \phi(x_i)^T \phi(x_j) \quad (14)$$

$$\sum_{i=1}^N (a_i + a_i^*) = 0, \quad 0 \leq a_i \leq C, \quad 0 \leq a_i^* \leq C, \quad i = 1, 2, \dots, N \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - Q_i')^2}{S}} \quad (18)$$

۲. آزمون ضریب تبیین (r)

$$(19)$$

۳. ضریب کارایی (نش- ساتکلیف)

$$CE = 1 - \frac{\sum_{i=1}^T (Q_i - Q_i')^2}{\sum_{i=1}^T (Q_i - \bar{Q}_i)^2} \quad (20)$$

که Q_i : داده‌های واقعی یا مشاهده‌ای، Q_i' : داده‌های پیش‌بینی، میانگین داده‌ها، S تعداد داده‌ها و σ_{Q_i} به ترتیب برابر واریانس داده‌های پیش‌بینی و مشاهداتی است [۲۶-۲۸].

نتایج و بحث

در این مطالعه قبل از مدل‌سازی و بررسی دقت مدل‌های مطالعه‌شده، ابتدا برای انتخاب داده‌های بارش که تأثیر زیادی در ایستگاه هیدرومتری تپیک داشته باشد، از روش تیسن برای انتخاب ایستگاه سینوپتیک استفاده شد و ایستگاه‌های سینوپتیک متأثر در هر منطقه مشخص و نتایج در شکل ۳ ارائه شد.

یکی از مراحل مهم در مدل‌سازی، انتخاب ترکیب مناسبی از متغیرهای ورودی است. از این‌رو، ابتدا همبستگی متقابل بین متغیرهای ورودی و خروجی محاسبه شد. پارامترهای ورودی به‌منظور دستیابی به مدل بهینه برای تخمین میزان دبی روزانه انتخاب و در جدول ۲ ارائه شد.

ξ_i و ξ_i^* وجود خواهد داشت. شایان یادآوری است که مدل SVM مشکلات ناشی از کم‌تخمینی و فوق‌برازشی را با کمینه‌کردن هم‌زمان دو ترم $W^T W / 2$ و خطای آموزشی، یعنی $C \sum_{i=1}^N (\xi_i + \xi_i^*)$ در رابطه ۱۴ حل می‌کند.

تابع هدف بالا در رابطه ۱۶ تابع محدب است و بنابراین جواب رابطه ۱۶ یکتا و بهینه خواهد بود. پس از تعریف ضرایب لاگرانژ در رابطه ۱۶ مشخصه‌های w و b در مدل SVM رگرسیونی با استفاده از تئوری کرانش-کوهن-تاکر محاسبه می‌شود که در آن

$$W = \sum_{j=1}^N (a_j + a_j^*) \phi(x_j) \quad (16)$$

است. در نتیجه برای مدل SVM رگرسیونی خواهیم داشت: $W = \sum_{i=1}^N (a_i + a_i^*) \phi(x_i)^T \phi(x) + b$ باید توجه داشت که ترم‌های لاگرانژ $(a_i + a_i^*)$ می‌توانند صفر یا غیرصفر باشند. بنابراین، فقط مجموعه داده‌هایی که ضرایب a_i آنها غیرصفر است در معادله رگرسیون نهایی وارد می‌شوند و این مجموعه داده‌ها به‌عنوان بردارهای پشتیبان شناخته می‌شوند. در نهایت، تابع SVM رگرسیونی را می‌توان به‌صورت زیر بازنویسی کرد:

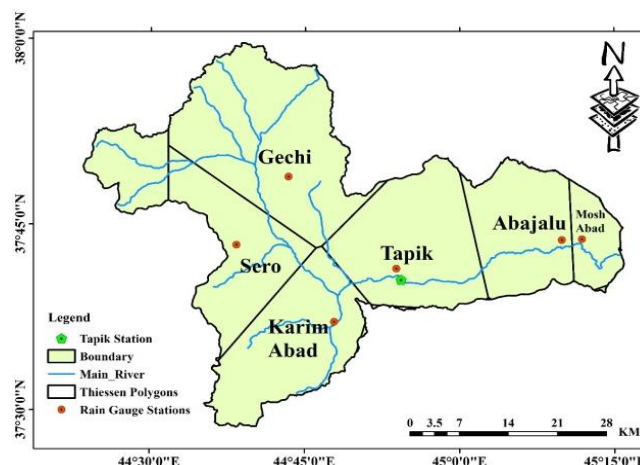
$$f(x) = \sum_{i=1}^N a_i \phi(x_i)^T \phi(x) + b \quad (17)$$

در رابطه ۱۷ محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل SVM رگرسیونی انتخاب یک تابع کرنل به‌صورت $K(x_i, x) = \phi(x_i)^T \phi(x) = \sqrt{b^2 - 4ac}$ است. می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف مدل ε -SVM استفاده کرد [۲۵].

معیارهای ارزیابی

به‌منظور ارزیابی دقت و کارایی مدل‌های بررسی‌شده از ضریب تبیین، جذر میانگین مربعات خطا و نش-ساتکلیف به شرح زیر استفاده شد.

۱. جذر میانگین مربعات خطا (RMSE):



شکل ۳. تیسنبندی ایستگاه‌های هواشناسی حوضه نازلوچای

جدول ۲. ترکیب‌های منتخب پارامترهای ورودی مدل‌های ANFIS، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان

Model	Input	Output
M1	Q_t, Q_{t-1}	$Q(t)$
M2	Q_t, Q_{t-1}, Q_{t-2}	$Q(t)$
M3	$Q_{t-1}, Q_{t-2}, Q_{t-3}$	$Q(t)$
M4	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	$Q(t)$
M5	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}$	$Q(t)$
M6	Q_t, Q_{t-1}, R_{t-1}	$Q(t)$
M7	$Q_t, Q_{t-1}, Q_{t-2}, R_{t-1}, R_{t-2}$	$Q(t)$
M8	$Q_{t-1}, Q_{t-2}, Q_{t-3}, R_{t-1}, R_{t-2}, R_{t-3}$	$Q(t)$
M9	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}$	$Q(t)$
M10	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}$	$Q(t)$

سری زمانی دبی Q و سری زمانی بارش R

در الگوهای ورودی M1 تا M6 خطای مدل‌سازی در مرحله آموزش حدود ۲۲ درصد و در مرحله آزمایش حدود ۳۸ درصد نسبت به متوسط الگوهای ورودی M6 تا M10 کمتر است. همچنین دقت مدل‌سازی در الگوهای ورودی تک‌سری (M1 تا M6) نسبت به الگوی ورودی M6 تا M10 به‌طور متوسط حدود ۱۳ و ۲۲ درصد به‌ترتیب در مرحله آموزش و آزمایش بیشتر است. با مقایسه حالتی که از حافظه سری به‌عنوان ورودی‌های مدل استفاده می‌شود (حالت نخست) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر دبی ماهانه بهره گرفت، کمترین مقدار خطا را ارائه کرد. نتایج بررسی مقادیر کارایی مدل (نش-ساتکلیف) نیز بیانگر آن است که حالت دوم دقت کمتری دارد. میانگین خطا در حالت نخست ۹/۵ و در حالت دوم ۱۱/۵ و نیز میانگین دقت (ضریب همبستگی) ۰/۸۱ و ۰/۷۱ به‌ترتیب برای حالت نخست و دوم به‌دست آمد. الگوی M5 نیز بیشینه ضریب همبستگی

نتایج مدل‌سازی ANFIS

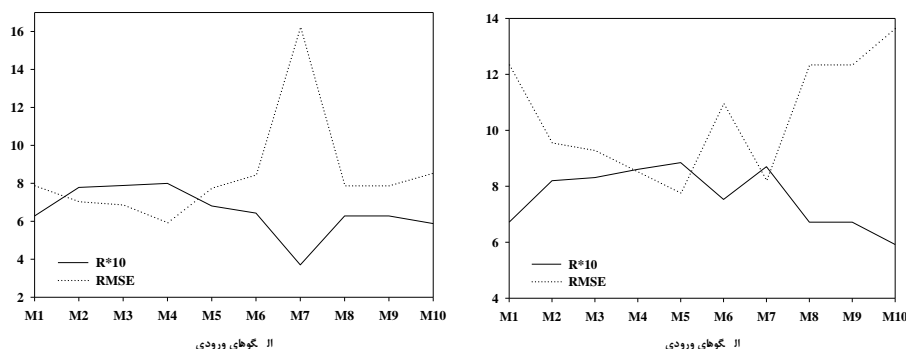
در مطالعه حاضر برای مدل‌سازی دبی ماهانه ایستگاه تپیک رودخانه نازلوچای با استفاده از ANFIS، داده‌های ۳۴ سال (۱۹۷۱-۲۰۰۵) برای آموزش و ۹ سال (۲۰۰۵-۲۰۱۴) به‌عنوان داده‌های صحت‌سنجی انتخاب شدند. به‌طور کلی، ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد نیز برای آزمایش در نظر گرفته شدند. جدول ۳ توابع عضویت استفاده‌شده در مدل ANFIS را برای مدل‌های مختلف نشان می‌دهد. در جدول ۴ نیز نتایج ارزیابی الگوهای مختلف ورودی ANFIS برای مدل‌سازی دبی ماهانه در مراحل آموزش و صحت‌سنجی ارائه شده است. با توجه به این جدول می‌توان نتیجه گرفت که در کل عملکرد ANFIS در پیش‌بینی دبی ماهانه بسیار مناسب است. اگر صرفاً از حافظه سری دبی ماهانه برای پیش‌بینی استفاده شود، براساس جدول ۴ مشاهده می‌شود که دقت مدل‌سازی به بیشترین مقدار خود می‌رسد. به‌طور متوسط

متغیر بود. شکل ۴ نمودار تغییرات مقادیر RMSE و R را در مراحل آموزش و آزمایش نشان می‌دهد. نتایج مدل‌سازی مقادیر دبی ایستگاه مطالعه‌شده در دوره آماری بررسی‌شده در شکل‌های ۵ و ۶ ارائه شد.

و کمترین مقدار خطا را در مرحله آموزش ارائه کرده است. در مرحله آزمایش نیز الگوی M4 بیشینه ضریب همبستگی و ۵ بود. مقدار RMSE برای کل مدل‌های ANFIS ارائه‌شده بین ۷/۷ تا ۱۳/۶ مترمکعب بر ثانیه

جدول ۳. توابع عضویت ورودی و خروجی مدل ANFIS

الگوی ورودی	تابع عضویت خروجی	تعداد توابع عضویت برای هر ورودی	تابع عضویت ورودی
M1	Linear	۴	Tri
M2	Linear	۴	Gauss
M3	Linear	۳	Tri
M4	Linear	۴	Tri
M5	Linear	۴	Gauss
M6	Linear	۴	Tri
M7	Linear	۳	Gauss
M8	Linear	۴	Gauss
M9	Linear	۴	Gauss
M10	Linear	۴	Tri

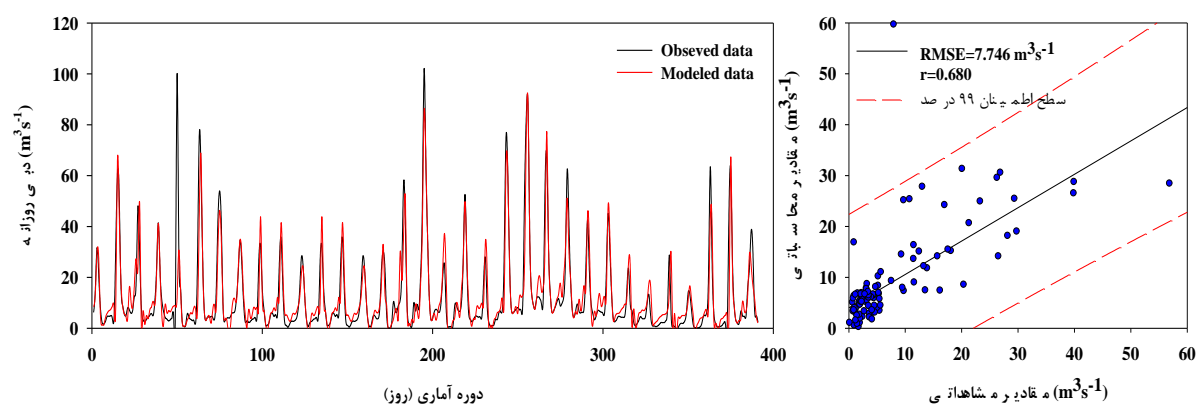


شکل ۴. نتایج بررسی دقت و خطای مدل ANFIS در مدل‌سازی دبی ماهانه ایستگاه تپیک در مرحله آموزش (سمت راست) و مرحله آزمایش (سمت چپ)

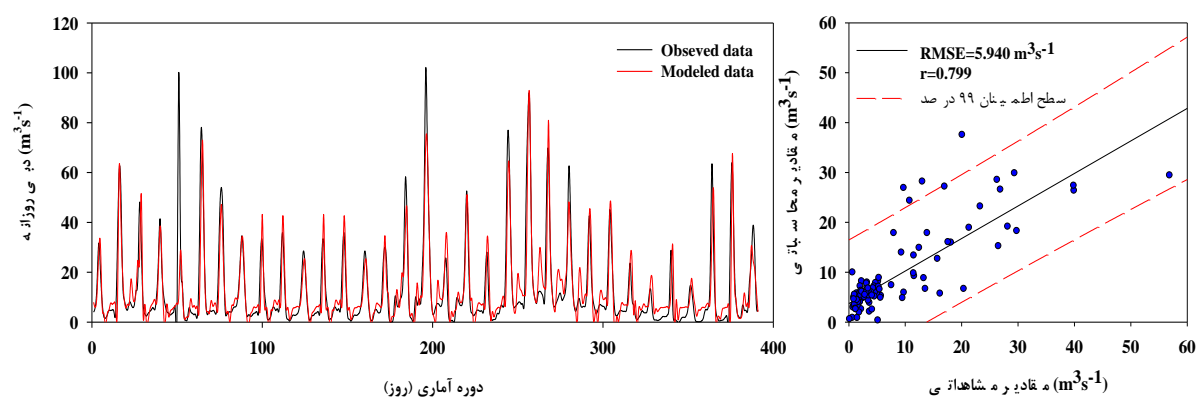
جدول ۴. نتایج آماری الگوهای ورودی مدل ANFIS در پیش‌بینی جریان ماهانه

الگوی ورودی	آموزش			آزمایش		
	RMSE	R	NS	RMSE	R	NS
M1	۱۲/۳۵	۰/۶۷	۰/۴۵۱	۷/۸۹	۰/۶۳	۰/۳۴۶
M2	۹/۵۶	۰/۸۲	۰/۶۷۲	۷/۰۵	۰/۷۸	۰/۴۷۷
M3	۹/۳۹	۰/۸۳	۰/۶۹۱	۶/۸۸	۰/۷۹	۰/۵۰۳
M4	۸/۵۲	۰/۸۶	۰/۷۴۱	۵/۹۴	۰/۸۰	۰/۶۲۹
M5	۷/۷۷	۰/۸۸	۰/۷۸۳	۷/۷۶	۰/۶۸	۰/۵۶۶
M6	۱۰/۹۷	۰/۷۵	۰/۵۶۶	۸/۴۶	۰/۶۴	۰/۲۴۸
M7	۸/۲۱	۰/۸۷	۰/۷۵۷	۱۶/۲۵	۰/۳۷	-۱/۷۷۶
M8	۱۲/۳۵	۰/۶۷	۰/۴۵۱	۷/۸۹	۰/۶۳	۰/۳۴۶
M9	۱۲/۳۵	۰/۶۷	۰/۴۵۱	۷/۸۹	۰/۶۳	۰/۳۴۶
M10	۱۳/۶۵	۰/۵۹	۰/۴۵۱	۸/۵۵	۰/۵۹	۰/۳۴۶

※: واحد RMSE برحسب مترمکعب بر ثانیه است.



شکل ۵. نتایج مدل سازی مقادیر دبی جریان با استفاده از الگوی M5 در دو مرحله Train (سمت چپ) و Test (سمت راست) با استفاده از مدل ANFIS



شکل ۶. نتایج مدل سازی مقادیر دبی جریان با استفاده از الگوی M4 در دو مرحله Train (سمت چپ) و Test (سمت راست) و مدل ANFIS

متغیر γ نیز به صورت آزمون و خطا محاسبه شدند. البته شایان یادآوری است که الگوریتم بهینه سازی جست و جوی شبکه بسیار کند عمل می کند و زمان محاسباتی زیادی را به خود اختصاص می دهد. برای حل این مشکل، طبق توصیه چن و همکارانش (۲۰۰۶) از برنامه اصلاح شده الگوریتم جست و جوی شبکه که به نام الگوریتم جست و جوی شبکه دومرحله ای معروف است به همراه اعتبارسنجی متقاطع^۱ استفاده شد. برای این منظور ابتدا با انتخاب شبکه هایی با ابعاد بزرگ محدوده مشخصه های ε و C به ازای مقدار ثابت مشخصه γ تعیین شد. سپس با مشخص شدن محدوده یاد شده و تقسیم آن به شبکه هایی با ابعاد ریزتر، مقادیر دقیق دو مشخصه ε و C مشخص

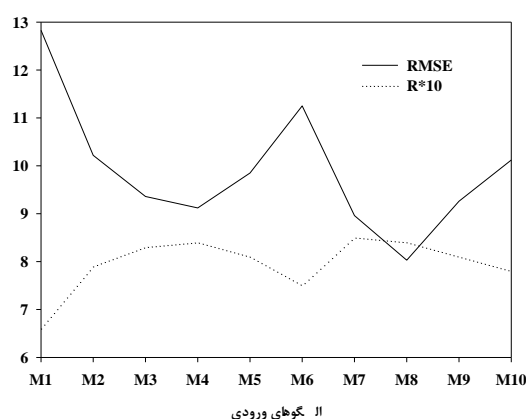
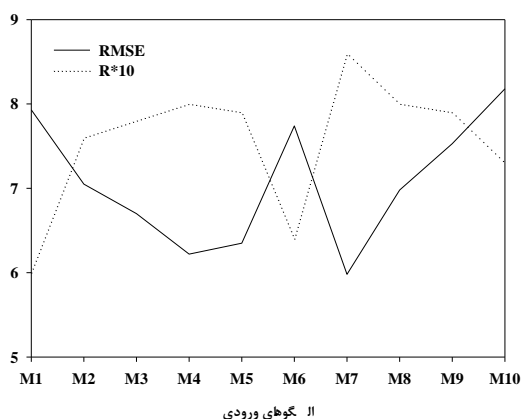
نتایج مدل سازی SVM

در مطالعه حاضر برای مدل سازی دبی ماهانه با استفاده از روش SVM نیز همانند روش ANFIS، داده های ۳۴ سال (۱۹۷۱-۲۰۰۵) برای آموزش و ۹ سال (۲۰۰۵-۲۰۱۴) به عنوان داده های صحت سنجی انتخاب شدند. در مرحله بعد مقادیر بهینه مشخصه های مدل SVM شامل ε و C تعیین می شود. همچنین در این مطالعه تابع کرنل استفاده شده، تابع RBF انتخاب شد چون دقت بیشتری در برآورد دبی ماهانه داشت. در این تابع نیز مشخصه γ باید تعیین شود. بنابراین، در حالت کلی برای پیش بینی دبی ماهانه توسط مدل SVM رگرسیونی، باید مقادیر بهینه سه مشخصه یاد شده به دست آید که بدین منظور دو مشخصه ε و C توسط الگوریتم بهینه سازی جست و جوی شبکه و

1. Cross-Validation

شدند. روند یادشده برای دیگر مقادیر γ نیز تکرار شد و از این راه مدل‌های متفاوتی با تغییر در مقدار γ به دست آمدند. حال می‌توان از بین مدل‌های توسعه داده‌شده، مدل با کمترین خطا را تعیین و مشخصه‌های آن را به‌عنوان مقادیر بهینه ϵ ، C و γ انتخاب کرد. برای هر الگوی ورودی با تغییر ϵ ، C و γ شبکه‌های مختلفی ساخته و آموزش داده شد و در نهایت ساختاری که کمترین خطا و بیشترین دقت را داشت، به‌عنوان مناسب‌ترین الگو انتخاب شد. جدول ۵ شاخص‌های آماری مختص به نتایج به‌دست‌آمده از کاربرد مدل SVM را برای هر الگوی ورودی نشان می‌دهد. با توجه به این جدول می‌توان نتیجه گرفت که در کل عملکرد SVM در پیش‌بینی دبی ماهانه بسیار مناسب است. همان‌طور که در جدول ۵ مشاهده می‌شود، اگر از حافظه خود سری دبی ماهانه برای مدل‌سازی استفاده شود، دقت مدل‌سازی رفته‌رفته تا چهار تأخیر زمانی بهبود می‌یابد و در گام زمانی چهارم بیشینه دقت براساس شاخص‌های NS، RMSE و R به دست می‌آید. با مقایسه بین حالتی که از حافظه سری به‌عنوان ورودی‌های مدل استفاده می‌شود (حالت نخست) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر دبی بهره گرفته می‌شود (حالت دوم) بر خلاف مدل ANFIS، حالت دوم دقت بیشتری دارد. به‌طور متوسط در الگوهای ورودی M6 تا M10 خطای مدل‌سازی در مرحله آموزش حدود ۷ درصد و در مرحله آزمایش حدود ۶ درصد نسبت به

متوسط الگوهای ورودی M1 تا M6 کمتر است. همچنین دقت مدل‌سازی در الگوهای ورودی تک‌سری (M1 تا M6) نسبت به الگوی ورودی M6 تا M10 به‌طور متوسط حدود ۲/۵ و ۲ درصد به‌ترتیب در مرحله آموزش و آزمایش کمتر است. با مقایسه بین حالتی که از حافظه سری به‌عنوان ورودی‌های مدل استفاده می‌شود (حالت نخست) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر دبی ماهانه بهره گرفته می‌شود (حالت دوم) می‌توان نتیجه گرفت که در حالت دوم دقت بیشتری دارد. میانگین خطا در حالت اول ۱۰/۳ و در حالت دوم ۹/۵ و نیز میانگین دقت (ضریب همبستگی) ۰/۷۸ و ۰/۸۰ به‌ترتیب برای حالت اول و دوم به‌دست آمد. الگوهای M7 و M8 نیز بیشینه ضریب همبستگی و الگوی M8 کمترین مقدار خطا در مرحله آموزش ارائه کرده است. براساس شاخص NS الگوهای ۴، ۵، ۷ و ۸ بهترین عملکرد را در بین سایر الگوها دارند. همچنین نتایج نشان داد براساس ضرایب نش - ساتکلیف الگوی ورودی ۱ کمترین میزان کارایی را در بین سایر الگوها ایفا کرده است. در مرحله آزمایش نیز الگوی M7 کمینه مقدار خطا و بیشترین مقدار ضریب همبستگی را ارائه کرد. مقدار RMSE برای کل الگوهای کاربردی در مدل SVM بین ۵/۸ تا ۱۲/۸ مترمکعب بر ثانیه متغیر بود. شکل ۷ نمودار تغییرات مقادیر RMSE و R را در مراحل آموزش و آزمایش نشان می‌دهد.



شکل ۷. نتایج بررسی دقت و خطای مدل SVM در مدل‌سازی دبی ماهانه ایستگاه تپیک در دو مرحله آموزش (سمت راست) و آزمایش (سمت چپ)

جدول ۵. نتایج تحلیل آماری الگوهای مختلف ورودی مدل SVM در پیش‌بینی دبی ماهانه

الگوی ورودی	آزمایش			آموزش		
	NS	R	RMSE	NS	R	RMSE
M1	۰/۳۳۹	۰/۶۰	۷/۹۳	۰/۴۰۸	۰/۶۶	۱۲/۸۳
M2	۰/۴۷۸	۰/۷۶	۷/۰۵	۰/۶۲۰	۰/۷۹	۱۰/۲۲
M3	۰/۵۲۷	۰/۷۸	۶/۷۰	۰/۶۸۲	۰/۸۳	۹/۳۶
M4	۰/۵۹۲	۰/۸۰	۶/۲۲	۰/۷۰۶	۰/۸۴	۹/۱۲
M5	۰/۵۷۶	۰/۷۹	۶/۳۵	۰/۷۱۳	۰/۸۱	۹/۸۵
M6	۰/۳۶۹	۰/۶۴	۷/۷۴	۰/۵۳۹	۰/۷۵	۱۱/۲۵
M7	۰/۶۰۴	۰/۸۶	۵/۹۸	۰/۷۰۸	۰/۸۵	۸/۹۶
M8	۰/۶۰۴	۰/۸۰	۶/۹۸	۰/۷۱۶	۰/۸۴	۸/۰۳
M9	۰/۳۶۷	۰/۷۹	۷/۵۳	۰/۶۲۴	۰/۸۱	۹/۲۶
M10	۰/۰۸۱	۰/۷۳	۸/۱۸	۰/۶۱۵	۰/۷۸	۱۰/۱۲

※: واحد RMSE برحسب مترمکعب بر ثانیه است.

نتایج مدل‌سازی برنامه‌ریزی ژنتیک

استفاده از برنامه‌ریزی بیان ژن به دلیل توان انتخاب متغیرهای مؤثر در مدل و حذف متغیرهایی با تأثیر کمتر و همچنین توانایی ارائه رابطه صریح برای تخمین جریان عبوری از ایستگاه تپیک مد نظر قرار گرفت. در مطالعه حاضر برای مدل‌سازی جریان روزانه رودخانه نالوچای با استفاده از این روش همانند دو روش یادشده، داده‌های ۳۴ سال (۱۹۷۱-۲۰۰۵) برای آموزش و ۹ سال (۲۰۰۵-۲۰۱۴) به‌عنوان داده‌های آزمایش انتخاب شدند. به‌طور کلی، ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد نیز برای آزمایش در نظر گرفته شدند. در برنامه‌ریزی ژنتیک انتخاب جمعیت‌های اولیه که همان الگوهای ورودی هستند، اهمیت زیادی دارند. با توجه به اینکه در این مطالعه توالی دبی ماه‌های قبل در پیش‌بینی دبی ماه‌های حاضر مد نظر بوده است، صرفاً از داده‌های دبی جریان با توالی برگشتی تا پنج ماه قبل به‌عنوان داده‌های آموزشی، به‌صورت ترکیب‌های مختلف ورودی سایر روش‌ها استفاده شده است. نکته مهم دیگر در برنامه‌ریزی ژنتیک، انتخاب عملگرهای مدل برای انجام محاسبات است. در این مطالعه عملگر ترکیب به‌دلیل داشتن دقت زیاد، برای مدل‌سازی جریان ماهانه انتخاب شد. همچنین در این پژوهش برای ساخت درخت تجزیه و تولید یک جمعیت اولیه از راه حل‌هایی استفاده شد که این راه حل‌ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی استفاده‌شده در روابط) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند. چگونگی و دلیل انتخاب این نوع عملگرها با توجه به

مطالعات قربانی و همکارانش (۲۰۱۲) صورت گرفت. در جدول ۶ نتایج ارزیابی الگوهای مختلف ورودی برنامه‌ریزی ژنتیک برای مدل‌سازی جریان ماهانه رودخانه در مراحل آموزش و صحت‌سنجی را نشان می‌دهد. در این مدل نیز به‌طور متوسط در الگوهای ورودی M1 تا M6 (مدل با ورودی تک‌سری) خطای مدل‌سازی در مرحله آموزش نسبت به متوسط الگوهای ورودی M6 تا M10 حدود ۱ درصد و در مرحله آزمایش حدود ۱۴ درصد کمتر است. همچنین دقت مدل‌سازی در الگوهای ورودی تک‌سری (M1 تا M6) نسبت به الگوی ورودی M6 تا M10 به‌طور متوسط حدود ۳ درصد کمتر ولی بیشترین همبستگی بین الگوهای M1 تا M6 با همبستگی ۰/۸ وجود دارد. در مرحله صحت‌سنجی نیز دقت الگوهای ورودی تک‌سری حدود ۶ درصد بیشتر از الگوهای M6 تا M10 است. با مقایسه حالتی که از حافظه سری به عنوان ورودی‌های مدل استفاده می‌شود (حالت نخست) نسبت به حالتی که از متغیرهای هواشناسی در تخمین مقادیر دبی ماهانه بهره گرفته می‌شود (حالت دوم) می‌توان نتیجه گرفت که در مرحله آموزش الگوهای ترکیبی دقت بیشتر و خطای کمتری دارند. در مرحله آموزش میانگین خطا در حالت نخست (الگوی تک‌سری) $10.64\text{m}^3\text{s}^{-1}$ و در حالت دوم (الگوی ترکیبی) $10.54\text{m}^3\text{s}^{-1}$ و در مرحله صحت‌سنجی نیز به ترتیب ۷/۲۵ و ۸/۲۲ مترمکعب بر ثانیه به دست آمد. در مرحله آموزش الگوهای M3 و M10 بیشینه ضریب همبستگی و کمترین مقدار خطا را ارائه کرده‌اند. در مرحله صحت‌سنجی نیز الگوهای M2، M3 و M4 کمینه مقدار خطا

NS را داشته که بیانگر کارایی کمتر آن نسبت به دو مدل یادشده است.

بررسی عملکرد مدل‌های استفاده‌شده

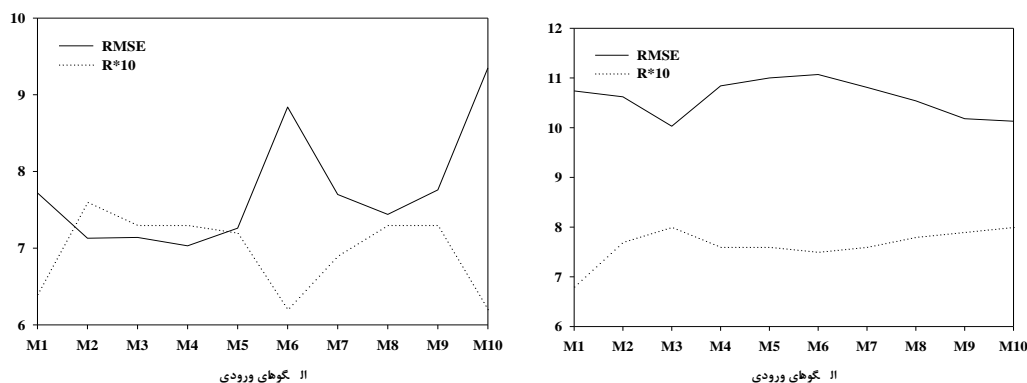
نتایج بررسی و مقایسه عملکرد الگوهای برتر سه مدل GP، ANFIS و SVM در مدل‌سازی دبی جریان رودخانه نازلوچای در محل ایستگاه تپیک براساس کمترین مقدار خطا (RMSE)، بیشترین مقدار کارایی (NS) و بیشترین مقدار تبیین (R) در جدول ۷ ارائه شد.

و بیشترین مقدار ضریب همبستگی را ارائه کردند. براساس مقادیر NS محاسبه‌شده، نتایج بیانگر برتری الگوهای M3، M9 و M10 نسبت به سایر الگوها بود. مقدار RMSE برای کل الگوهای کاربردی در مدل GP بین ۱۰/۰۳ تا ۱۱/۰۷ مترمکعب بر ثانیه برای مرحله آموزش متغیر بود. شکل ۸ نمودار تغییرات مقادیر RMSE و R مدل GP را در دو مرحله آموزش و آزمایش نشان می‌دهد. شایان یادآوری است که مدل GP نسبت به مدل‌های SVM و ANFIS کمترین مقدار

جدول ۶. تحلیل‌های آماری نتایج برنامه‌ریزی ژنتیک برای الگوهای مختلف ورودی جریان ماهانه رودخانه نازلوچای

آموزش			آزمایش			الگوی ورودی
RMSE	R	NS	RMSE	R	NS	
۱۰/۷۴	۰/۶۸	۰/۴۵۵	۷/۷۲	۰/۶۴	۰/۳۷۴	M1
۱۰/۶۲	۰/۷۷	۰/۵۹۵	۷/۱۳	۰/۷۶	۰/۴۶۵	M2
۱۰/۰۳	۰/۸۰	۰/۶۳۹	۷/۱۴	۰/۷۳	۰/۴۵۴	M3
۱۰/۸۴	۰/۷۶	۰/۵۸۰	۷/۰۳	۰/۷۳	۰/۴۸۱	M4
۱۱/۰۰	۰/۷۶	۰/۵۶۷	۷/۲۶	۰/۷۲	۰/۴۴۶	M5
۱۱/۰۷	۰/۷۵	۰/۵۵۹	۸/۸۴	۰/۶۲	۰/۱۷۲	M6
۱۰/۸۱	۰/۷۶	۰/۵۸۰	۷/۷۰	۰/۶۹	۰/۳۷۷	M7
۱۰/۵۴	۰/۷۸	۰/۶۰۰	۷/۴۴	۰/۷۳	۰/۴۱۸	M8
۱۰/۱۸	۰/۷۹	۰/۶۳۶	۷/۷۶	۰/۷۳	۰/۳۶۷	M9
۱۰/۱۳	۰/۸۰	۰/۶۳۰	۹/۳۵	۰/۶۲	۰/۰۸۱	M10

※ واحد RMSE برحسب مترمکعب بر ثانیه است.



شکل ۸. نتایج بررسی دقت و خطای مدل GP در مدل‌سازی دبی ماهانه ایستگاه تپیک در دو مرحله آموزش (سمت راست) و آزمایش (سمت چپ)

جدول ۷. نتایج نهایی بررسی الگوهای مختلف ورودی مدل‌های بررسی‌شده

SVM		GP		ANFIS		مدل
Train	Test	Train	Test	Train	Test	مرحله ارزیابی
M8	M7	M3	M4	M5	M4	الگوی برتر
۳/۰۸	۵/۹۸	۱۰/۰۳	۷/۰۳	۷/۷۷	۵/۹۴	RMSE کمینه
۰/۸۵	۰/۸۶	۰/۸۰	۰/۷۶	۰/۸۸	۰/۸۰	R بیشینه
۰/۷۱۶	۰/۶۰۴	۰/۶۳۹	۰/۴۸۱	۰/۷۸۳	۰/۶۲۹	NS بیشینه

توپوگرافی دشوار است. بنابراین، در این مطالعه تجزیه و تحلیل همبستگی متقابل دبی ماهانه جریان رودخانه برای پیدا کردن تعداد مشاهدات گذشته و ارائه ورودی مؤثر برای مدل‌ها استفاده شد. در هر سه مدل بررسی شده ۸۰ درصد داده‌ها برای مرحله آموزش (۳۴ سال داده آماری در مقیاس ماهانه) و ۲۰ درصد داده‌ها (۹ سال در مقیاس ماهانه) برای بررسی دقت و صحت‌سنجی مدل استفاده شد. در این راستا از الگوهای ورودی ترکیبی (داده‌های دبی و بارش ماهانه با تأخیرهای مختلف) و تک‌سری (داده‌های ماهانه دبی جریان با تأخیرهای مختلف) استفاده شد. نتایج بررسی دقت و میزان خطای مدل‌ها نشان داد الگوی ترکیبی فقط در مدل SVM بهترین نتیجه را می‌دهد و در دو مدل GP و ANFIS الگوهای تک‌سری بهترین نتیجه را ارائه کردند. از بین سه مدل بررسی شده، مدل ANFIS با الگوی ورودی چهار و پنج تأخیر، بهترین نتیجه را ارائه کرد. به‌طور کلی، نتایج نشان داد با به‌کارگیری مدل ANFIS در مدل‌سازی دبی جریان ماهانه رودخانه نازلوچای، خطای مدل نسبت به دو مدل GP و SVM به ترتیب حدود ۲۳ و ۳ درصد (در واحد دبی جریان) کاهش یافته و دقت مدل نیز نسبت به دو مدل GP و SVM به ترتیب حدود ۱۰ و ۴ درصد افزایش یافته است.

منابع

- [1]. Govindaraju RS. Artificial neural network in hydrology. *Journal of hydrologic Engineering*. 2000;5(2): 115-123.
- [2]. Koza JR. *Genetic Programming: on the programming of computers by means of natural selection*. Cambridge, MA: MIT Press. 1992.
- [3]. Alvisi S, Mascellani G, Franchini M, Bardossy A. Water level forecasting through fuzzy logic and artificial neural network approaches. *J Hydrol Earth Sys Sci*. 2005;2:1107-1145.
- [4]. Aytok A, Kisi O. A genetic programming approach to suspended sediment modeling. *Journal of Hydrology*. 2008;351: 288-298.
- [5]. Wang WC, Chau KW, Cheng CT, Qiu L. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*. 2009;374: 294-306.

همان‌طور که در جدول ۷ مشاهده می‌شود، دو مدل GP و ANFIS الگوهای تک‌سری (ورودی داده‌های دبی با تأخیرهای مختلف به‌عنوان ورودی مدل) را به‌عنوان الگوهای برتر در مدل‌سازی داده‌های دبی ماهانه رودخانه نازلوچای در محل ایستگاه تپیک معرفی کردند. همچنین نتایج نشان داد از بین سه مدل بررسی شده، مدل ANFIS با کمترین معیار خطا و بیشترین معیار همبستگی در مرحله آموزش به‌عنوان بهترین مدل شناخته شد. در مرحله صحت‌سنجی نیز دو مدل ANFIS و SVM نتایج تقریباً یکسانی ارائه کردند اما از آنجا که آموزش بهتر توسط مدل ANFIS ارائه شده است، در مرحله صحت‌سنجی به‌عنوان بهترین مدل انتخاب می‌شود. این در حالی است که کارایی مدل ANFIS در هر دو مرحله آموزش و آزمایش براساس شاخص NS بیشتر از سایر مدل‌هاست. به‌طور کلی، نتایج مقایسه مدل‌ها و الگوهای ورودی مختلف نشان داد مدل ANFIS توانایی بهتری در مدل‌سازی دبی جریان رودخانه در محل ایستگاه تپیک دارد. نایاک و همکارانش (۲۰۰۴) نیز در بررسی مدل‌های فازی-عصبی در مدل‌سازی سری‌های زمانی هیدرولوژی نشان دادند مدل ANFIS در مدل‌سازی دبی جریان رودخانه نتایج بهتری ارائه می‌کند. همچنین این مورد در مطالعات طلائی و همکارانش (۲۰۱۰) و گوس و همکارانش (۲۰۱۳) نیز تأیید شده است.

نتیجه‌گیری

در این مطالعه برای مدل‌سازی داده‌های دبی ماهانه ایستگاه تپیک واقع بر رودخانه نازلوچای از رودخانه‌های ورودی به دریاچه ارومیه در شمال غرب ایران، سه مدل ANFIS، SVM و GP بررسی و مقایسه شدند. برای انتخاب داده‌های بارش متناسب با ایستگاه هیدرومتری مطالعه شده از روش چندضلعی تیسن استفاده شد. در هر سه مدل بررسی شده از الگوهای یکسان در ورودی داده‌های اولیه استفاده شد. الگوهای M1 تا M5 داده‌های دبی جریان با تأخیر یک تا پنج و الگوهای M6 تا M10 الگوی ترکیبی با داده‌های بارش و دبی و با تأخیرهای یک تا پنج ماه بررسی شدند. مدل‌های هوش مصنوعی قادر به تجزیه و تحلیل و تعمیم روابط بین پارامترهای مدل هستند، اما استفاده از تمام پارامترها برای توسعه مدل برای رودخانه‌های با ویژگی‌های مختلف آب و هوایی و

- [6]. Guven, A. Linear genetic programming for time-series modeling of daily flow rate. *J Earth Syst. Sci.* 2009;118 (2): 157-173.
- [7]. Ghorbani MA, Kisi O, Aalinezhad, M. A probe into the chaotic nature of daily stream flow time series by correlation dimension and largest Lyapunov methods. *Applied Mathematical Modelling.* 2012;34: 4050-4057.
- [8]. Zahiri A, Azamathulla HM. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Comput & Applic.* 2014;24:413-420.
- [9]. Vapnik VN. *Statistical Learning Theory.* Wiley, New York. 1998
- [10]. Pai PF, Hong WC. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process.* 2007;21: 819-827.
- [11]. Hamel LH. *Knowledge discovery with support vector machines (Vol. 3).* John Wiley & Sons. 2011.
- [12]. Adamowski J, Prasher SO. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *Journal of Water and Land Development.* 2012;17(8):89-97.
- [13]. Choubey V, Mishra S, Pandey SK. Time Series Data Mining in Real Time Surface Runoff Forecasting through Support Vector Machine. *International Journal of Computer Applications.* 2014;98(3): 23-30.
- [14]. Jang JSR. ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on.* 1993; 23(3): 665-685.
- [15]. Nayak PC, Sudheer KP, Rangan DM, Ramasastre KS. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology.* 2004;291(1-2):52-66
- [16]. Sanikhani H, Kisi, O. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. *Water Resources Management.* 2012;26(6): 1715-1729.
- [17]. Talei A, Chua LHC, Wong TSW. Evaluation of rainfall and discharge inputs used by Adaptive Network-based Fuzzy Inference Systems (ANFIS) in rainfall-runoff modeling. *Journal of Hydrology.* 2010;391(3-4): 248-262.
- [18]. Ghose D, Panda P, Swain P. Prediction and optimization of runoff via ANFIS and GA. *Alexandria Engineering Journal.* 2013;52(2): 209-220.
- [19]. Chen SH, Lin YH, Chang LC, Chang FG. The strategy of building a flood forecast model by neuro fuzzy network. *Hydrological Processes.* 2006;20(7): 1525- 1540.
- [20]. He Z, Wen X, Liu H, Du J. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology.* 2014;509:379-386
- [21]. Karakus M, Tutmez B, Fuzzy and multiple regression modeling for evaluation of intact rock strength based on point load, Schmidt hammer and sonic velocity, *Rock Mech. Rock Eng.* 2006;39(1): 45-57.
- [22]. Jang JSR. Fuzzy controllers based on temporal back propagation, *IEEE Trans. Neural Netw.* 1992; 3:714-723.
- [23]. Ferreira C. *Gene Expression Programming: A New Adaptive Algorithm for Solving Problems.* Complex Systems, forthcoming. 2001.
- [24]. Hofmann T, Tsochantaridis I, Altun Y. Learning over structured output spaces via joint kernel functions. *Sixth Kernel Workshop.* 2002
- [25]. Eskandari A, Nouri, R, Meraji H, Kiaghadi A. Developing a Proper Model for Online Estimation of the 5-Day Biochemical Oxygen Demand Based on Artificial Neural Network and Support Vector Machine. *J of Environment Studies.* 2012;38(1):71-82 [Persian].
- [26]. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology.* 1970;10(3): 282-290
- [27]. Swinnow TDV, Campbell MJ. *Statistics at Square One.* London: BMJ Publication. 106 P. 2002.
- [28]. Salas JD, Delleur JW, Yevjevich V, Lane WL. *Applied Modeling of Hydrologic Time Series.* Water resource Publications, P. O. Box 2841, Littleton, Colorado. 80161, U.S.A. 1980; 484 P.