

Nu_D و RF_D ، میانگین جریان رودخانه با دقت بالایی در هر سه مدل محاسبه شد.

کلیدواژه‌ها: آستانه‌گذاری نرم، جریان رودخانه، خرم‌آباد، سیگنال فرکانس بالا، شبیه‌سازی، موجک دابچیز.

مقدمه

پیش‌بینی فرآیندهای هیدرولوژیکی پیچیده و غیرخطی مانند رواناب، موضوعی کلیدی در هیدرولوژی و منابع آب محسوب می‌شود، وانگ و دینگ [۳۳]. با توجه به اهمیت پیش‌بینی جریان رودخانه از نقطه‌نظرهای مختلف، یافتن روشی مناسب در این خصوص حائز اهمیت است، منتصری و قویدل [۲۳]. چنان‌که مدل‌های هیدرولوژی همواره نقش مهمی در مدیریت منابع آب ایفا کرده‌اند، یانگ و همکاران [۳۴].

کاربرد بسیاری از تکنیک‌های تحلیل سری زمانی، مقید به فرض وجود رابطه خطی بین متغیرها هستند. در صورتی که در جهان واقعی، نوسانات گذرا در داده‌ها، نظم و قاعده‌های ساده‌ای را نشان نمی‌دهند؛ بنابراین استفاده از مدل‌های غیرخطی برای تحلیل پدیده‌های پیچیده دنیای واقعی ضرورت دارد، کیشی [۱۷]. در سال‌های اخیر، مدل‌هایی هم‌چون شبکه‌های عصبی مصنوعی (ANN)، ماشین‌های بردار پشتیبان^۵ و K- نزدیک‌ترین همسایه، به‌طور موفقیت‌آمیزی برای مدل‌سازی و پیش‌بینی فرآیندهای پیچیده هیدرولوژیکی مانند رواناب استفاده شده‌اند، کومار و همکاران [۱۸]. از آن‌جاکه سری‌های زمانی غیرخطی غالباً نویزی هستند، به‌منظور کاهش نویز و استخراج مؤثر اطلاعات از این سری‌ها ضرورت دارد که پیش‌پردازش‌هایی بر روی داده‌ها انجام گیرد، گائو و همکاران [۱۱]. این نویزها در دو دسته کلی شامل نویز منوط به نحوه اندازه‌گیری داده‌ها و نویز دینامیکی که مربوط به ماهیت داده‌هاست جای می‌گیرند، دونهو [۹] و حاجی‌زاده و همکاران [۱۲]. از این نویز به‌عنوان نویز سفید یا گوسی نیز یاد می‌شود. البته گاهی نویز به‌صورت غیرسفید در سری زمانی داده‌ها پدید می‌آید که به‌عنوان نویز غیریکنواخت شمرده می‌شود. در نویز غیریکنواخت، نویز به چندین سری زمانی تقسیم می‌شود که هر کدام از آن‌ها دارای واریانس و میانگین منحصر به خود بوده و اختلاف معناداری در پراکندگی نویز این سری‌های زمانی وجود دارد. از

بهبود عملکرد ماشین‌بردار پشتیبان و الگوریتم جنگل تصادفی در پیش‌بینی جریان رودخانه خرم‌آباد با استفاده از نویززدایی غیریکنواخت داده‌ها و الگوریتم سیمپلکس

حمیدمیرهاشمی^۱، داریوش یاراحمدی^۲، سیامک شرفی^۳ و سعید فرزین^۴
تاریخ دریافت: ۱۳۹۷/۰۵/۲۸ تاریخ پذیرش: ۱۳۹۸/۰۱/۲۴

چکیده

در این مطالعه به‌منظور شبیه‌سازی جریان ماهانه رودخانه خرم‌آباد ابتدا سری زمانی این رودخانه طی بازه زمانی ۱۳۳۴-۱۳۹۳ با استفاده از موجک مادر دابچیز سه به سه سطح تجزیه شده است. براین اساس مشخص شد که نویز غیریکنواختی شامل دو دوره زمانی با مرز مهرماه ۱۳۸۷ در این سیگنال وجود دارد. با اتخاذ رویکرد آستانه‌گذاری فراگیر، سطح آستانه برای بازه‌های زمانی ۱۳۳۴-۱۳۸۷ و ۱۳۸۷-۱۳۹۳ به ترتیب برابر ۵/۹۵ و ۲/۲۱ مترمکعب بر ثانیه محاسبه شد. سپس، با طرح آستانه‌گذاری نرم، سیگنال‌های فرکانس بالا جریان رودخانه به طور غیریکنواخت نویززدایی گردید. در ادامه، با استفاده از مدل‌های ماشین‌بردار پشتیبان (نسخه‌های ϵ و Nu) و الگوریتم جنگل تصادفی (RF)، سیگنال اصلی و سیگنال نویززدایی شده جریان رودخانه، شبیه‌سازی و مقایسه شدند. نتایج صحت‌سنجی مدل‌ها نشان داد با نویززدایی غیریکنواخت سیگنال جریان رودخانه، خطای مدل ϵ از ۵/۷ به ۳/۱، مدل Nu از ۵/۸ به ۳/۲ و RF از ۵ به ۲/۹ مترمکعب بر ثانیه کاهش پیدا کرده است. نتایج حاصل از مقایسه دقت محاسباتی مدل‌های مذکور با استفاده از آزمون مرگان - گرنجر - نیوبلد (MGN) به ترتیب به مقدار ۱۵/۷-، ۱۵- و ۱۰/۶۷- برآورد شده است. این مقادیر، نشان از بهبود معنادار عملکرد مدل‌های مورد استفاده در نتیجه نویززدایی غیریکنواخت سیگنال جریان دارد. در نهایت، با کاربرد الگوریتم بهینه‌ساز سیمپلکس در مدل‌های ϵ_D

۱. استادیار آب و هواشناسی گروه جغرافیا دانشگاه لرستان

۲. دانشیار آب و هواشناسی گروه جغرافیا دانشگاه لرستان

۳. استادیار ژئومورفولوژی گروه جغرافیا دانشگاه لرستان

۴. نویسنده مسئول و استادیار، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان. پست الکترونیک: saeed.farzin@semnan.ac.ir

آنجا که شناسایی سیستم، تخمین پارامتر، پیش‌بینی و خوشه‌بندی الگو به‌عنوان مهم‌ترین هدف از تحلیل سری‌های زمانی به شمار می‌روند، لذا وجود نويز در داده‌ها، دستیابی مؤثر به این اهداف را با چالش روبه‌رو می‌کند، گائو و همکاران [۱۱]. بنابراین نويززدایی در سری‌های زمانی، می‌تواند به بهبود کارایی مدل و یا به‌عبارتی به تشخیص بهتر رفتار سری زمانی توسط مدل کمک شایانی کند، حسن‌زاده [۱۳]، خورشیددوست و همکاران [۱۶] لی و همکاران [۱۹] نورانی و همکاران [۲۵].

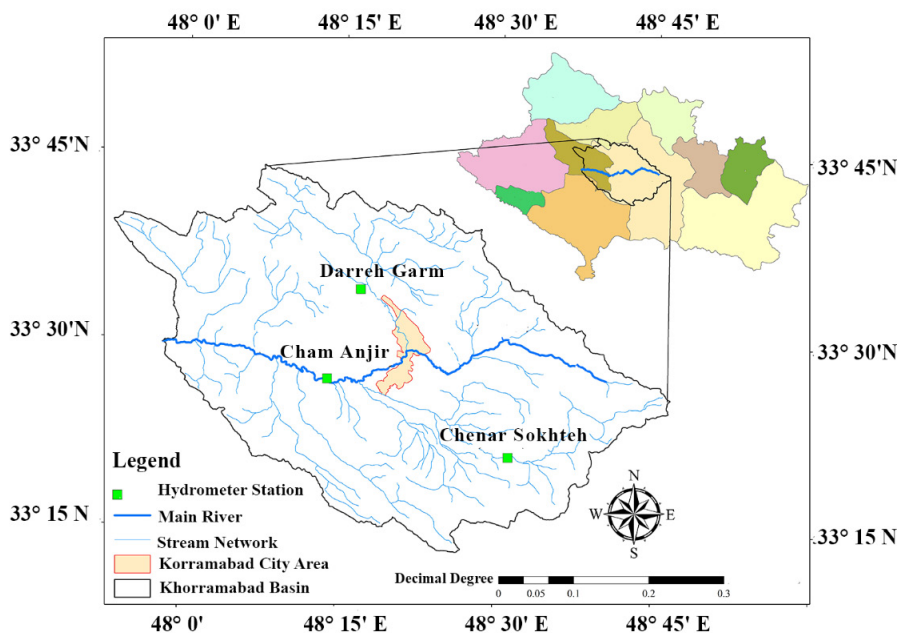
به‌منظور نويززدایی، تبدیل فوریه به‌عنوان یک روش متداول برای تشخیص اطلاعات پنهان به شمار می‌رود پولیکار [۲۷]. تبدیل فوریه اطلاعات مربوط به فرکانس‌های موجود در یک سیگنال را ارائه می‌دهد درحالی‌که از زمان رخداد سیگنال هیچ‌گونه اطلاعی نمی‌دهد. زیرا در تبدیل فوریه، سیگنال از حوزه زمانی به محدوده فرکانسی منتقل می‌شود ولی چون روی شکل سینوسی واقع شده، اطلاعات زمانی بعد از انتقال از دست می‌رود. بنابراین اگر سیگنال نایب‌ا باشد آگاهی داشتن از زمان رخداد یک فرکانس اهمیت پیدا می‌کند، رن و روک [۲۸]. به‌منظور فائق آمدن به محدودیت‌های که در به‌کارگیری تحلیل فوریه وجود دارد، ایده آنالیز موجک در سال ۱۹۸۴ توسط گراسمن معرفی شد، لطف‌الهی‌یقین و کوهدرق [۲۱]. از آنجا که موجک‌ها هم در محدوده زمانی و هم فرکانسی، سیگنال را تجزیه می‌کنند، لذا رویکرد آنالیز موجک برای تجزیه و نويززدایی سری‌های زمانی نایب‌ا بسیار کارآمد است لطف‌الهی‌یقین و همکاران [۲۰].

با توجه به اهمیت موضوع، تاکنون پژوهش‌های چندی انجام پذیرفته است که به اهم آن در دهه اخیر اشاره می‌شود؛ میسرا و همکاران [۲۲] به مقایسه نتایج دو روش شبکه عصبی مصنوعی و ماشین‌بردار پشتیبان در شبیه‌سازی رواناب و بار رسوب رودخانه و امسادهارا در هندوستان در مقیاس زمانی روزانه، هفتگی و ماهانه پرداختند. نتایج این مطالعه نشان داد که ماشین‌بردار پشتیبان در کلیه بخش‌های مدل‌سازی شامل آموزش، آزمون و اعتبارسنجی، رواناب شبیه‌سازی شده دقیق‌تری ارائه داده است. آداموفسکی و پراشر [۱] میزان کارایی دو روش رگرسیون ماشین‌بردار پشتیبان و ترکیب شبکه عصبی مصنوعی - آنالیز موجک در پیش‌بینی جریان پیچیده حوزه کوهستانی سیانچی در منطقه هیمالیا هندوستان را مقایسه کردند. نتایج نشان داد که هر دو مدل قادرند رواناب روزانه حوزه کوهستانی را به‌طور دقیقی پیش‌بینی کنند. چائو و همکاران [۶] به‌منظور پیش‌بینی میان‌مدت و بلندمدت جریان ماهانه منطقه بالا رود رودخانه زرد از ماشین‌بردار پشتیبان، مدل سری زمانی اتورگرسیو و شبکه عصبی مصنوعی شعاعی استفاده کردند. مقایسه نتایج حاصله از این سه روش نشان داد که مدل ماشین‌بردار پشتیبان دارای بهترین اجرا و کم‌ترین خطای شبیه‌سازی بلندمدت و میان‌مدت رواناب در قیاس با دو مدل دیگر است. لی و همکاران [۱۹] با در نظرگیری این مهم که اجتناب از اثرات نويز در داده‌های هیدرولوژیکی مشکل است، با

استفاده از موجک مادر دابچیز و طرح آستانه‌گذاری نرم، سری زمانی داده‌های جریان رودخانه را نويززدایی کردند. سپس با استفاده از سه روش ماشین‌بردار پشتیبان (SVR)، شبکه عصبی مصنوعی پس‌گرد (BPNN) و سیستم تطبیقی نرو - فازی (ANFIS) جریان رودخانه را براساس داده‌های ورودی خام و داده‌های نويززدایی شده پیش‌بینی کردند. نتایج این مطالعه نشان داد که با نويززدایی سری زمانی جریان رودخانه، دقت پیش‌بینی سه مدل یادشده بهبود می‌یابد.

ثانی‌خانی و همکاران [۲۹] با استفاده از روش‌های K-نزدیک‌ترین همسایه، ماشین‌بردار پشتیبان و فازی - عصبی تطبیقی به مدل‌سازی جریان ماهانه رودخانه باراندوزچای پرداختند. نتایج این مطالعه نشان داد که هر سه روش، مقدار دبی بیشینه را کم‌تر از مقدار مشاهداتی برآورد می‌کنند. با این‌حال، ماشین‌بردار پشتیبان عملکرد بهتری نسبت به دو روش دیگر داشته است. احمدی و همکاران [۳] در پیش‌بینی جریان روزانه رودخانه باراندوزچای با استفاده از دو روش برنامه‌ریزی ژنتیک و ماشین‌بردار پشتیبان دریافتند که با افزایش مقادیر دبی، میزان دقت روش‌های یادشده در شبیه‌سازی جریان رودخانه کاهش پیدا می‌کند. از سویی، اگرچه دقت مدل برنامه‌ریزی ژنتیک اندکی بیش از ماشین‌بردار پشتیبان است اما روش ماشین‌بردار پشتیبان به سبب پیچیدگی کم‌تر، به‌عنوان یک روش کاربردی‌تر جهت پیش‌بینی جریان مطرح است. عیسی‌زاده و همکاران [۱۵] عملکرد توابع کرنل خطی، چند جمله‌ای و پایه شعاعی ماشین‌بردار پشتیبان در پیش‌بینی جریان ماهانه رودخانه حوزه خرخره‌چای را ارزیابی کردند. هم‌چنین، کارایی هر یک از توابع یادشده را با مدل‌های سری زمانی مقایسه نمودند. نتایج نشان داد که ابتدا تابع چندجمله‌ای درجه -SVM ۴ و سپس مدل سری زمانی (۶,۲) ARIMA دقیق‌ترین شبیه‌سازی را در قیاس با سایر توابع و مدل‌های سری زمانی داشته‌اند. آهنی و شوریان [۲]، کارایی چهار روش رگرسیون چندگانه، K-نزدیک‌ترین همسایه، شبکه عصبی مصنوعی و سیستم استنتاج عصبی فازی تطبیقی را در پیش‌بینی رواناب ماهانه حوزه آبخیز دریاچه نمک مورد ارزیابی قرار دادند. نتایج این مطالعه نشان داد که استفاده از سیگنال‌های دما و بارش با تأخیرهای زمانی یک و دو ماه، به‌عنوان ورودی، موجب بهبود عملکرد چهار مدل یادشده می‌شود. در این خصوص مدل K-نزدیک‌ترین همسایه با خطای ۲/۶۲ از بهترین عملکرد در میان سه مدل دیگر برخوردار بوده است.

شایان توجه است که داده‌های جریان رودخانه غالباً حاوی مقداری نويزاند. این نويز ممکن است از توزیع نرمال برخوردار نبوده و به‌صورت غیریکنواخت در قالب چند سری زمانی متفاوت توزیع یافته باشد. چنین فرایندی در داده‌ها موجب پیچیدگی بیش‌تر ماهیت پیچیده و غیرخطی سیگنال سری زمانی رودخانه خواهد شد. لذا بدون پیش‌پردازش اولیه، شبیه‌سازی چنین سری‌های زمانی حتی با استفاده از روش‌های هوشمند و غیرخطی نمی‌تواند منتهی به مدل‌های نسبتاً دقیق با قدرت تعمیم‌دهی بالا شود. بنابراین نخستین



شکل ۱- موقعیت جغرافیایی رودخانه خرم‌آباد در حوزه آبخیز خرم‌آباد
Fig 1. Geographical location of Khorramabad river in the Khorramabad Watershed

۹/۶۷ متر مکعب بر ثانیه است که فروردین ماه و شهریور ماه به ترتیب به عنوان پرآب‌ترین کم‌آب‌ترین ماه‌های این رودخانه به شمار می‌روند.

تبدیل موجک و نویزدایی سیگنال سری زمانی

نویزدایی عبارت است از حذف نویز به طوری که تا حد امکان اطلاعات مفید سیگنال حفظ شود. روش کار بدین ترتیب است که ابتدا سیگنال دریافتی را در چندین مرتبه تجزیه کرده و سپس روی مراتب مختلف تجزیه، عمل آستانه‌گذاری انجام می‌شود. چندین روش برای تعیین سطح آستانه ارائه شده که مهم‌ترین آن‌ها عبارتند از: آستانه‌گذاری فراگیر، کمینه کردن تخمین نااریب ریسک اشتاین، کوچک‌ترین حداکثر و آستانه‌گذاری بلوکی. در آستانه‌گذاری فراگیر ضرایب تمام سطوح تجزیه سیگنال با یک سطح آستانه ثابتی که بر اساس رابطه $\sqrt{2h(N)}$ (N: تعداد داده‌ها در سری زمانی است) محاسبه شده و آن‌گاه بر روی ضرایب فرکانس بالا تبدیل موجک، عمل آستانه‌گذاری صورت می‌گیرد. ولی در تخمین نااریب ریسک اشتاین و کوچک‌ترین حداکثر، برای هر سطح تجزیه، سطح آستانه مجزای تعیین می‌شود، دونهو [۹] بعد از تعیین نوع آستانه، مشخص کردن این که آستانه چگونه بر سری زمانی اعمال شود ضرورت دارد. در این خصوص دو طرح به نام آستانه‌گذاری سخت (رابطه ۱) و نرم (رابطه ۲) وجود دارد

$$\eta(x) = \begin{cases} x & |x| \geq T \\ 0 & |x| < T \end{cases} \quad (1)$$

$$\eta(x) = \begin{cases} x - T & x > T \\ 0 & |x| \leq T \\ x + T & x < -T \end{cases} \quad (2)$$

در روابط فوق T و X به ترتیب نشان‌دهنده داده‌ها و آستانه هستند.

هدف این مطالعه، نویزدایی غیریکنواخت سری زمانی داده‌های جریان رودخانه خرم‌آباد به منظور آماده‌سازی این داده‌ها برای فرایند پیش‌بینی است. هدف بعدی این مطالعه، مقایسه عملکرد مدل‌های ماشین‌بردار پشتیبان و جنگل تصادفی^۱ با ورودی سیگنال اصلی و سیگنال نویزدایی شده است؛ تا بر این اساس، روشن شود که آیا نویزدایی داده‌های جریان رودخانه منجر به افزایش دقت پیش‌بینی مدل‌های مذکور خواهد شد.

مواد و روش‌ها

منطقه مورد مطالعه

رودخانه خرم‌آباد در حوزه آبخیز خرم‌آباد با مساحت تقریبی ۱۶۴۰ کیلومتر مربع در مرکز استان لرستان و در شهرستان‌های خرم‌آباد و دوره چگنی جریان دارد. این رودخانه با جمع‌آوری آب سرشاخه‌های کوچک و بزرگ با جهت شرقی - غربی از مرکز شهر خرم‌آباد عبور می‌کند و پس از مسافتی بالغ بر ۴۰ کیلومتر به رودخانه کشکان که از سرشاخه‌های اصلی حوزه آبخیز کرخه است، می‌پیوندد. این حوزه آبخیز از نظر تقسیم‌بندی‌های اقلیمی، جزء مناطق نیمه‌خشک بوده که میانگین بارش سالانه آن براساس داده‌های ایستگاه خرم‌آباد ۴۹۵ میلی‌متر است. در این حوزه چهار ایستگاه هیدرومتری قرار دارد که در این مطالعه از داده‌های ماهانه ایستگاه چم‌انجیر با پوشش سال‌های آبی - زراعی ۱۳۹۳-۱۳۳۴ استفاده شده است. این ایستگاه با موقعیت جغرافیای ۳۳ درجه و ۱۳ دقیقه عرض شمالی و ۴۸ درجه و ۱۴ دقیقه طول شرقی و ارتفاع ۱۱۴۰ متر از سطح دریا در جنوب شهر خرم‌آباد قرار دارد (شکل ۱). میانگین جریان ماهانه رودخانه خرم‌آباد برابر با

1.Random Forest

چنانکه مشخص است در آستانه‌گذاری سخت، ضرایب با اندازه بزرگ‌تر از آستانه، بدون تغییر می‌مانند. ولی در آستانه‌گذار نرم، از اندازه ضرایب بزرگ‌تر از آستانه به مقدار آستانه کاسته می‌شود. بنابراین در این مطالعه به منظور نوپزدایی سری زمانی از تابع موجک مادر دابچیز سه و تعداد سه سطوح تجزیه که تعداد این سطوح براساس رابطه $L = \text{INT}[\log(N)]$ به دست آمده، استفاده شده است. در این رابطه، L درجه تجزیه، INT تابع گردکننده و N تعداد داده‌های سری زمانی است، وانگ و دینگ [۳۳].

ماشین‌های بردار پشتیبان

رویکرد ماشین‌بردار پشتیبان ابتدا به صورت مجموعه‌ای از الگوریتم‌ها توسط وپنیک و چروننکیس [۳۲] بر پایه تئوری یادگیری آماری توسعه داده شد. این روش یک مدل داده‌محور بر پایه یادگیری با نظارت است که در مسائل مربوط به طبقه‌بندی و رگرسیون از آن استفاده می‌شود. به عبارتی، روشی برای طبقه‌بندی دوتایی در فضای ویژگی دلخواه است و بدین ترتیب روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود. این روش یک طبقه‌بندی کننده دو وجهی است که سعی دارد تا در مرز دو طبقه، ابر صفحه‌ای ایجاد کند که فاصله هر طبقه تا ابر صفحه بیشینه باشد. داده‌های نقطه‌ای که به ابر صفحه نزدیک‌ترند برای اندازه‌گیری این فاصله بکار می‌روند. این داده‌های نقطه‌ای بردارهای پشتیبان نام دارند، وپنیک [۳۱]. به طور کلی این روش یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرایی کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌شود، کریستینی و شو - تیلور [۷]. رگرسیون بردار پشتیبان برای توصیف رگرسیون در ماشین‌بردار پشتیبان استفاده می‌شود که در ابتدا توسط سوینکس و واندا [۳۰] در قالب مدل حداقل مربعات ماشین‌بردار پشتیبان معرفی شد. در این خصوص دو تیپ رگرسیون ماشین‌بردار پشتیبان تحت عنوان ϵ -SVM و Nu-SVM وجود دارد که در این مطالعه از این دو روش استفاده شده است.

الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی توسط برایمن به عنوان یک روش یادگیری تجمعی برای مسائل مبتنی بر رگرسیون و خوشه‌بندی براساس توسعه درخت تصمیم، ارائه شده است، برایمن [۵]. یک جنگل تصادفی، مجموعه‌ای از درخت‌های هرس نشده است که هر درخت با الگوریتم جزءبندی بازگشتی حاصل می‌شود، فریدمن و همکاران [۱۰]. به عبارتی، جنگل تصادفی ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه خودسازمانده از داده‌ها شرکت دارند. روش خودسازمانده، روش نمونه‌گیری با جای‌گذاری است یعنی این که هیچ‌کدام از داده‌های انتخاب شده از نمونه‌های ورودی برای تولید زیرمجموعه بعدی حذف نمی‌شوند. بنابراین برخی از داده‌ها ممکن است بیش از یک بار در شاخه‌های آموزشی استفاده شوند و برخی دیگر که در مدل‌سازی تأثیر چندانی ندارند هرگز استفاده نمی‌شوند. برای نمونه خودسازمانده انتخابی،

یک درخت کلاس - بندی با استفاده از الگوریتم جزءبندی بازگشتی رشد می‌کند. در هر گره، عمل تجزیه مبتنی بر یک نمونه تصادفی از تعداد متغیر پیش‌بینی است. الگوریتم تجزیه بازگشتی آن چنان ادامه می‌یابد تا درخت به بزرگترین اندازه خود برسد، بدون آن که درخت هرس شود برایمن [۵].

شالوده کار این الگوریتم در مطالعه حاضر به این ترتیب است که نخست، سری زمانی داده‌ها به دو بخش آموزش و آزمون تقسیم شده است. آن گاه از داده‌های آموزش، نمونه‌گیری مجددی به حجم N گرفته شده و یک سوم از نمونه آموزشی جدید، به عنوان نمونه خارج از کیسه جدا می‌شود. نمونه خارج از کیسه، در هر درخت، نقش نمونه آزمون را برای آن درخت برعهده دارد. تعداد m متغیر با استفاده از معیار $\log_2 M + 1$ از بین M متغیر اصلی انتخاب می‌شوند که در ساخت هر درخت شرکت دارند. وقتی که درخت ساخته شد، داده خارج از کیسه در درخت ساخته شده قرار داده می‌شوند و میزان خطا محاسبه می‌شود. از آن جای که تعداد درخت‌های ساخته شده بر روی نمونه‌های خودسازمانده زیاد است این میزان ضمن کاهش سوگیری مدل، دقت آن را افزایش می‌دهد. به واقع، کلاس‌بندی روی درخت‌های ساخته شده روی نمونه خارج از کیسه‌ها انجام شده است. معیاری که در این قسمت برای میزان دقت در نمونه خارج از کیسه استفاده شده است خطای طبقه‌بندی است، برایمن [۵]. به طور کلی، با افزایش تعداد درخت‌ها، اثر بیش‌برازش که در روش درخت تصمیم رخ می‌دهد از بین می‌رود. بدین صورت که در هر مرحله از رشد درخت، مدل دقیق‌تر شده و جمع مربعات خطا کاهش می‌یابد. هدف در این روش این است که کلاس‌بندی به گونه‌ای انجام شود که مدل دارای کمینه مقدار خطا باشد.

معیارهای ارزیابی مدل

به منظور ارزیابی، مقایسه و گزینش دقیق‌ترین مدل جریان رودخانه، از شاخص نش - ساتکلیف (NCE)، ضریب همبستگی (R)، جذر میانگین مربعات خطا (RMSE) و ضریب CRM استفاده شده است.

آزمون مرگان - گرنجر - نیوبلد (MGN)

هرچند معیارهای یادشده از جمله معیارهای مفید در زمینه بررسی قدرت پیش‌بینی روش‌های مختلف هستند، اما هیچ‌کدام قادر نیستند برتری یک روش را به صورت آماری ارزیابی کنند. بنابراین برای این که از نظر آماری نیز فرضیه برابری صحت پیش‌بینی‌ها در مدل‌های مختلف مورد مقایسه قرار گیرند، از آزمون مرگان - گرنجر - نیوبلد (MGN) استفاده شده است (رابطه ۳). این آزمون قادر است معنادار بودن تفاوت روش‌های مختلف را نیز بررسی کند، دایبلد و ماریانو [۸].

شده است.

$$E = \frac{1}{3}(RF + \varepsilon + Nu) \quad (4)$$

$$VAR = \frac{(RF - E)^2 + (\varepsilon - E)^2 + (Nu - E)^2}{3} \quad (5)$$

$$ErrorBars = E \pm \sqrt{VAR}$$

در روابط ۴ و ۵، E مقدار تجمعی جریان رودخانه حاصل از بکارگیری الگوریتم سیمپلکس بر روی سه مدل جنگل تصادفی (RF) و ماشین‌های بردار پشتیبان (Nu، ε)، VAR واریانس مقدار جریان پیش‌بینی شده توسط سه مدل مورد نظر است.

نتایج

نویززدایی غیریکنواخت

برای نویززدایی سیگنال ماهانه جریان رودخانه خرم‌آباد که شامل ۷۰۸ داده (ماه) است، از تابع موجک مادر دابچیز سه و تابع نویززدایی موجکی با طرح آستانه‌گذاری نرم استفاده شده است. به این ترتیب با انتقال سیگنال اصلی جریان ماهانه رودخانه خرم‌آباد (شکل ۲) به محدوده موجک دابچیز سه و تجزیه این سیگنال به سه زیرسطح مشخص شد که دو فضای متفاوت در این سیگنال وجود دارد. رتبه زمانی این دو فضا عبارت‌اند از یک تا ۶۲۶ امین و ۶۲۶ تا ۷۰۸ امین داده که این رتبه‌ها مبتنی بر بازهای زمانی ۱۳۳۴ - ۱۳۸۷ و ۱۳۸۷ - ۱۳۹۳ هستند. بنابراین رتبه ۶۲۶ امین داده یا همان مهرماه سال زراعی ۱۳۸۷ - ۱۳۸۶ به عنوان مرز بین این دو دوره محسوب می‌شود. یعنی این‌که، نوسان و نویز داده‌های جریان رودخانه قبل از مهرماه ۱۳۸۷ - ۱۳۸۶ (۱۳۳۴ - ۱۳۸۷) با نویز داده‌های بعد از مهرماه ۱۳۸۷ - ۱۳۸۶ (۱۳۸۷ - ۱۳۹۳) بسیار تفاوت دارد. به عبارتی، توزیع نویز سری زمانی جریان ماهانه رودخانه خرم‌آباد، غیریکنواخت است و پارامترهای نویز در طی زمان متحمل تغییرات زیادی شده‌اند. بنابراین، ضرورت دارد که از دو آستانه‌گذاری متفاوت برای نویززدایی این سیگنال استفاده شود.

در این راستا، با اتخاذ روش آستانه‌گذاری فراگیر، سطح آستانه برای دوره اول (۱۳۳۴ - ۱۳۸۷) به مقدار ۵/۹۵۴ و برای دوره دوم (۱۳۸۷ - ۱۳۹۳) به مقدار ۲/۲۱ مترمکعب بر ثانیه محاسبه شد. در مرحله بعد، با طرح آستانه‌گذاری نرم (رابطه ۲)، آستانه‌های یادشده بر روی سه سطح فرکانس بالا جریان رودخانه در طی این دو دوره اعمال شد که نتیجه این فرایند بر روی سیگنال فرکانس بالا سطح یک در شکل ۳ قابل مشاهده است. بنابراین، سه سیگنال فرکانس بالا نویززدایی شده غیریکنواخت با سیگنال فرکانس پایین سطح سه جریان رودخانه ترکیب و بدین ترتیب سیگنال نویززدایی شده جریان ماهانه رودخانه خرم‌آباد به صورت معکوس بازتولید شد (شکل ۴). در ادامه، تأخیرهای زمانی یک، دو و سه ماه این سیگنال

$$\begin{aligned} e_{1,t} &= O - P_1, & e_{2,t} &= O - P_2 \\ S_t &= e_{1,t} + e_{2,t}, & D_t &= e_{1,t} - e_{2,t} \end{aligned} \quad (3)$$

$$MGN = \frac{\hat{\rho}_{sd}}{\sqrt{\frac{1 - \hat{\rho}_{sd}^2}{N - 1}}}$$

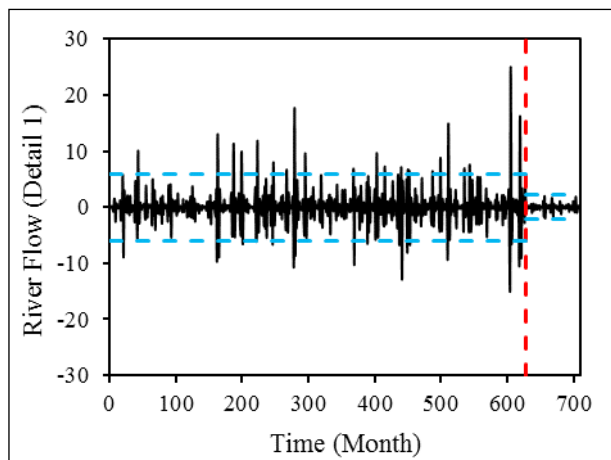
در رابطه ۳، MGN آماره آزمون مرگان - گرنجر - نیوبلد $e_{1,t}$ و $e_{2,t}$ به ترتیب خطای پیش‌بینی مدل اول و دوم، ρ_{sd} ضریب همبستگی بین مجموع خطای پیش‌بینی (S_t) و تفاضل خطای پیش‌بینی (D_t) دو مدل است. همچنین مقدار محاسبه شده MGN از توزیع t_{N-1} پیروی می‌کند.

بهینه‌سازی جریان شبیه‌سازی شده با استفاده از الگوریتم سیمپلکس

الگوریتم سیمپلکس یک تکنیک جستجوی غیر گرادیان مبتنی بر الگوریتم بهینه‌سازی است که برای کمینه‌کردن یک تابع دلخواه در یک تعداد دوره تکرار محدود بکار می‌رود. این الگوریتم برای هر تعداد از متغیرهای پیوسته کار می‌کند و هیچ پیش‌فرضی در مورد تابعی که باید بهینه شود به جزء این‌که پیوسته باشد ندارد. این الگوریتم در ابتدا با ارائه یک راه‌حل و آزمون بهینه‌سازی آن راه‌حل، شروع به کار می‌کند. تمامی مراحل، تا زمانی که یک مجموعه مقادیر مستقل برای دستیابی مدل به متغیر وابسته دلخواه پیدا شوند انجام می‌گیرد اوایل [۲۶].

شایان توجه است که مقایسه عملکرد چندین روش هوشمند و انتخاب بهترین روش همواره مبتنی بر اجرای این مدل‌ها بر روی یک مجموعه داده به عنوان آزمون است که این یک مؤلفه تصادفی است. یعنی این‌که هر یک از مدل‌های مورد استفاده ممکن است بر روی داده‌های آزمون اجرای خوبی داشته باشند ولی بر روی مجموعه داده‌های اعتبارسنجی دیگری از سیگنال مورد استفاده، اجرای رضایت‌بخشی نشان ندهند. براین اساس به نظر می‌رسد که اجرای تجمعی چندین مدل بهتر از اجرای یک مدل منفرد است که در مقایسه با سایر مدل‌ها بهترین اعتبارسنجی را بر روی داده‌های آزمون داشته است (مرحله آزمون مدل). در فرایند مدل‌سازی تجمعی، پیش‌بینی متغیر وابسته مبتنی بر برآورد تجمعی مقدار این متغیر از طریق ترکیب تخمین تمامی مدل‌هاست بیشاپ [۴]. در این مطالعه که از سه مدل جنگل تصادفی و ماشین‌های بردار پشتیبان تیپ یک و دو برای پیش‌بینی جریان ماهانه رودخانه خرم‌آباد استفاده شده، شبیه‌سازی ترکیبی جریان رودخانه توسط این سه مدل و بهینه‌سازی مقدار جریان رودخانه از طریق الگوریتم سیمپلکس انجام گرفته است. به این ترتیب، نتایج حاصل از ترکیب این سه مدل نیز با خروجی حاصل از اجرای منفرد آن‌ها ارزیابی شده است.

مقدار جریان تجمعی به صورت میانگین برآوردهای سه مدل براساس رابطه ۴ و واریانس برآورد این مدل‌ها از طریق رابطه ۵ ارائه



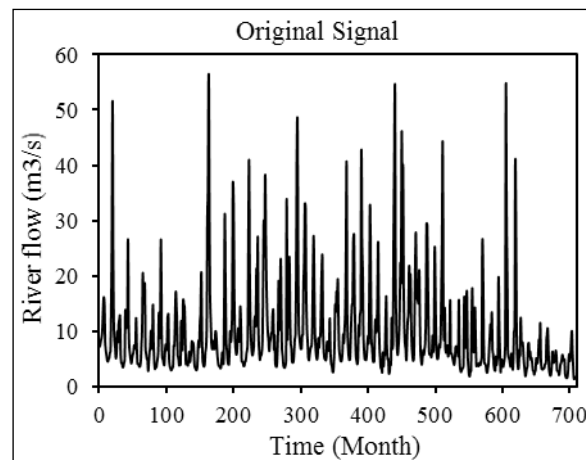
شکل ۳- سیگنال فرکانس بالا سطح ۱ جریان همرا با نویز غیریکنواخت

Fig 3. Detail signal of flow in level 1 associated with Non-Uniform noise

اصلی این رودخانه با استفاده از دو رویکرد ϵ و Nu ماشین‌های بردار پشتیبان انجام شده است. از آنجا که برای اجرای این دو رویکرد از کرنل تابع پایه شعاعی استفاده می‌شود، نخست ضرورت دارد که پارامترهای این کرنل شامل ϵ ، Nu ، C و γ بهینه شوند. با کاربرد روش جستجوی شبکه و اعتبارسنجی متقاطع و در نظرگیری دامنه‌ای از یک تا ۱۰ با گام افزایشی یک برای پارامتر C و دامنه‌ای از ۰/۱ تا ۰/۵ با گام افزایشی ۰/۱ برای ϵ و Nu این پارامترها بهینه شدند. به این ترتیب که مقدار بهینه ϵ و C به ترتیب برابر با ۰/۱ و چهار و مقادیر بهینه Nu و C به ترتیب برابر با ۰/۵ و هفت به دست آمدند. همچنین مقدار بهینه γ در هر دو رویکرد به مقدار ۰/۳۳۳ حاصل شد. نتایج حاصل از پیاده‌سازی این دو مدل ماشین‌بردار پشتیبان در جدول ۱ ارائه شده است. چنان‌که از معیارهای صحت‌سنجی مشخص است، دو مدل ϵ و Nu به ترتیب با دقت ۵۳ و ۵۲ درصد و خطای محاسباتی ۵/۷ و ۵/۸ مترمکعب بر ثانیه در ماه، جریان رودخانه را شبیه‌سازی کرده‌اند. همچنین براساس معیار CRM، هر دو مدل در مجموع، مقدار جریان رودخانه را کمتر از مقدار واقعی تخمین زده‌اند. علاوه بر این، توزیع فراوانی مقادیر خطا نیز نشان از فراوانی بیش‌تر مقادیری دارد که کمتر از مقادیر واقعی شبیه‌سازی شده‌اند (شکل‌های ۵ و ۶).

شبیه‌سازی جریان رودخانه با الگوریتم جنگل تصادفی (RF)

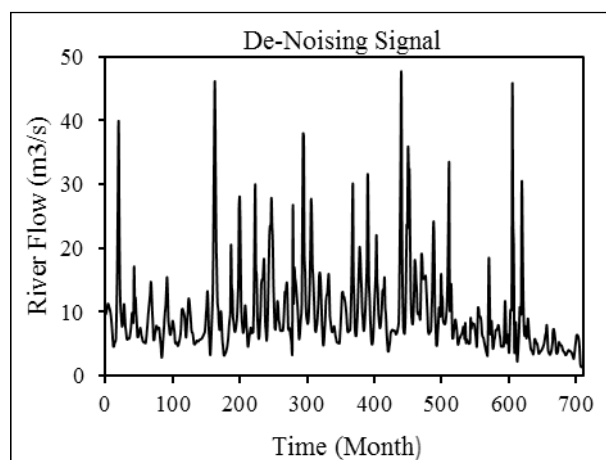
پیاده‌سازی بهینه الگوریتم جنگل‌های تصادفی مستلزم تعیین بهینه تعداد درخت و تعداد متغیر پیش‌بین است که اولی با سعی و خطا و دومی براساس معیار $\log_2 M+1$ قابل‌دستیابی است. با اجرای RF مشخص شد که با تعداد ۱۵۰ درخت، دو متغیر پیش‌بین در هر چرخه از تکرار مدل، الگوریتم جنگل تصادفی بهترین اجرا را در شبیه‌سازی جریان ماهانه رودخانه خرم‌آباد نشان می‌دهد. نتایج حاصل از ارزیابی صحت‌سنجی این مدل در جدول ۱ ارائه شده است. بررسی این معیارها حاکی از آن است که RF توانسته جریان



شکل ۲- سیگنال جریان ماهانه رودخانه خرم‌آباد ۱۳۳۴-۱۳۹۳

Fig 2. Signal of flow monthly for river Khorramabad (1334-1393)

نویززدایی شده به عنوان سیگنال‌های ورودی و تأخیر زمانی یک ماه بعد به عنوان سیگنال هدف مدل‌های پیش‌بینی جریان رودخانه محسوب شده است. همچنین، برای ارزیابی تأثیر نویززدایی غیریکنواخت سیگنال در دقت پیش‌بینی مدل‌های ماشین‌بردار پشتیبان و جنگل تصادفی نسبت به سیگنال نویزی، ابتدا سیگنال اصلی با استفاده از این سه روش مدل شد و آن‌گاه سیگنال نویززدایی شده مدل می‌شود. در مرحله آخر، معناداری تفاوت دقت مدل‌های پیاده‌سازی شده با استفاده از آزمون مرگان - گرنجر - نیولند (MGN)، در سطح اعتماد ۹۵ درصد اطمینان، آزمون شده است.



شکل ۴- سیگنال نویززدایی شده جریان رودخانه خرم‌آباد

Fig 4. Non-uniform De-noising Signal of flow for river Khorramabad

شبیه‌سازی جریان رودخانه با ماشین‌بردار پشتیبان (ϵ ، Nu) با اختصاص ۷۵ و ۲۵ درصد از داده‌های ماهانه جریان رودخانه خرم‌آباد به ترتیب به بخش‌های آموزش و آزمون، شبیه‌سازی سیگنال

رودخانه را با خطای پنج مترمکعب بر ثانیه، در ۵۲ درصد موارد به درستی شبیه سازی کند.

از سویی، این الگوریتم در مجموع، سوگیری متفاوتی نسبت به ماشین بردار پشتیبان نشان می دهد. به این ترتیب که الگوریتم RF برخلاف مدل های ماشین بردار پشتیبان در مجموع، جریان رودخانه را بیش تر از مقادیر مشاهداتی شبیه سازی کرده است. با این حال، ارزیابی نمودار توزیع خطای این مدل نشان می دهد که این مدل از منظر فراوانی، جریان رودخانه را کم تر از مقادیر واقعی شبیه سازی کرده است (شکل ۷). ارزیابی این دو دستاورد به این مفهوم است که مدل RF در برخی موارد با فراوانی کم، نسبت به دو مدل ϵ و Nu متحمل خطای محاسباتی قابل توجهی در شبیه سازی جریان رودخانه شده است. به عبارتی، فرکانس های متفاوتی در سری زمانی جریان این رودخانه وجود دارد که این مدل چنان که از شکل ۸ پیداست قادر به تطبیق مناسب با آن ها نیست. این فرایند منتهی به انعطاف پذیری کم تر الگوریتم RF می شود. مقایسه خروجی جریان شبیه سازی شده مدل RF با خروجی دو مدل ϵ و Nu با استفاده از آزمون MGN حاکی از عدم وجود اختلاف معنادار بین دقت عملکرد این مدل ها است. آماره MGN حاصل از آزمون دقت خروجی RF با خروجی مدل ϵ و Nu به مقدار $-1/1$ به دست آمد (جدول ۱). با توجه به قرارگیری این مقدار در ناحیه بحرانی جدول t نمی توان فرضیه صفر این آزمون را که دال بر عدم تفاوت معنادار بین صحت مقادیر شبیه سازی شده توسط این مدل ها است، رد کرد.

شبیه سازی جریان نویززدایی شده با ماشین بردار پشتیبان (ϵ_D و Nu_D)

در این مرحله، سیگنال نویززدایی شده غیریکنواخت جریان رودخانه خرم آباد (شکل ۴) به عنوان ورودی مدل های ϵ_D و Nu_D ماشین بردار پشتیبان محسوب شده است. پارامترهای کرنل تابع پایه شعاعی با استفاده از روش جستجوی شبکه و اعتبارسنجی متقاطع در همان دامنه های یاد شده در بخش پیشین، برای ϵ و C به ترتیب به مقدار $0/1$ و هفت و Nu و C به ترتیب به مقدار $0/3$ و هفت محاسبه شدند. ارزیابی آمارهای صحت سنجی ارائه شده در جدول ۲ نشان می دهد که دو مدل ϵ_D و Nu_D با خطای سه مترمکعب بر ثانیه، در ۷۸ درصد موارد جریان رودخانه خرم آباد را به درستی

شبیه سازی کرده اند. چنان که سری زمانی جریان محاسباتی دارای ضریب همبستگی $0/90$ ، با سری زمانی جریان مشاهداتی است. شاخص CRM که به عنوان شاخص سنجش سوی گیری مدل در این مطالعه بکار رفته، نشان می دهد که میزان سوگیری هر دو مدل نسبت به دو مدل ϵ و Nu کاهش پیدا کرده است. هم چنین، مقایسه سایر معیارها نشان می دهد با نویززدایی سیگنال جریان رودخانه، میزان خطای برآورد جریان رودخانه به نصف کاهش یافته و ضریب همبستگی بین مقادیر مشاهداتی و محاسباتی و عملکرد مدل های ϵ_D و Nu_D نسبت به مدل های ϵ و Nu نیز دستخوش افزایش شده اند.

به منظور آزمون معناداری چنین فرضی که نویززدایی سیگنال جریان رودخانه با استفاده از آنالیز موجک موجب بهبود عملکرد مدل های ماشین بردار پشتیبان شده، خطای محاسباتی مدل های ϵ_D و Nu_D نسبت به مدل های ϵ و Nu با استفاده از آزمون MGN در سطح اعتماد ۹۵ درصد اطمینان آزمون شده است. آماره این آزمون برای مدل ϵ ، ϵ_D و Nu ، Nu_D به ترتیب به مقدار $-15/7$ و -15 به دست آمد. در نتیجه، با توجه به قرار نگرفتن این مقادیر در ناحیه بحرانی جدول t ، آشکار می شود که نویززدایی غیریکنواخت سیگنال جریان رودخانه موجب افزایش معنادار دقت و کاهش معنادار خطای محاسباتی مدل های ماشین بردار پشتیبان در شبیه سازی جریان رودخانه خرم آباد شده است. از سویی، آماره MGN حاصل از مقایسه خطای محاسباتی ϵ_D و Nu_D با سیگنال ورودی نویززدایی شده به مقدار $2/67$ به دست آمد که این مقدار نشان از دقت بیش تر مدل Nu_D نسبت به مدل ϵ_D دارد.

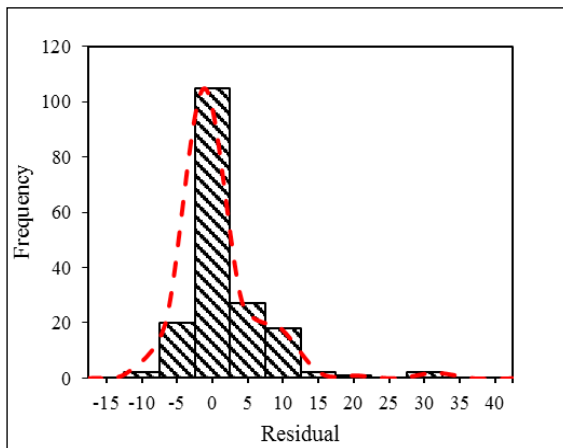
شبیه سازی جریان نویززدایی شده با جنگل تصادفی (RF_D)

مشابه اجرای مدل های ماشین بردار پشتیبان، الگوریتم جنگل تصادفی نیز با ورودی سیگنال نویززدایی غیریکنواخت سری زمانی جریان رودخانه خرم آباد و با اتخاذ 150 درخت و دو متغیر پیش بین در هر مرحله از اجرای مدل، پیاده شده است. معیارهای آماری حاصل از ارزیابی دقت این مدل در جدول ۳ ارائه شده اند. برای اساس مشخص می شود که الگوریتم RF_D با خطای سه مترمکعب بر ثانیه، در ۶۹ درصد موارد جریان رودخانه را به درستی شبیه سازی کرده است. این مقادیر نشان از بهبود عملکرد این الگوریتم مانند مدل های ماشین بردار پشتیبان دارد. هم چنین، ضمن این که مقدار سوگیری در این الگوریتم شبیه سازی، به طور برجسته ای کاهش یافته،

جدول ۱- نتایج عملکرد مدل های ϵ ، Nu و RF در شبیه سازی جریان ماهانه رودخانه خرم آباد

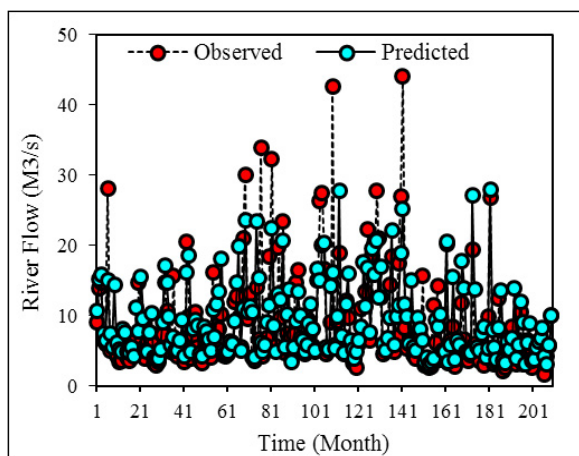
Table 1. Performance results for RF, Nu and ϵ models in simulation Khorammabad River flow monthly

مدل Model	R	RMSE	CRM	NCE%	MGN (RF, ϵ)
ϵ	0.75	5.7	0.025	53	-1.1
Nu	0.74	5.8	0.099	52	MGN (RF, Nu)
RF	0.725	5	-0.048	52	-1.1



شکل ۷- توزیع فراوانی خطای محاسباتی مدل RF

Fig 7. The frequency distribution of computational error



شکل ۸- مقایسه جریان مشاهداتی و جریان محاسباتی مدل RF

Fig 8. Comparison of the observed and computed flow for

RF model

جدول ۲- نتایج عملکرد مدل‌های ϵ_D و Nu_D در شبیه‌سازی جریان ماهانه رودخانه خرم‌آباد

Table 2. Performance results for Nu_D and ϵ_D models in simulation Khorammabad River flow monthly

مدل Model	R	RMSE	CRM	NCE%	MGN (Sig _{De-Noising} , Sig)
ϵ_D	0.9	3.1	0.018	78	-15.7
Nu_D	0.895	3.2	0.039	77	-15

Sig_{De-Noising} و Sig: به ترتیب معرف سیگنال نویززدایی شده و سیگنال اصلی جریان

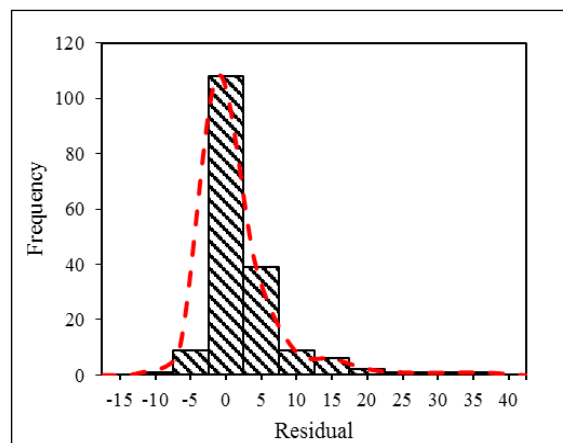
رودخانه خرم‌آباد

جدول ۳- نتایج عملکرد مدل RF_D در شبیه‌سازی جریان ماهانه رودخانه خرم‌آباد

Table 3. Performance results for RF_D model in simulation Khorammabad river flow monthly

مدل Model	R	RMSE	CRM	NCE %	RF _D , ϵ_D	RF _D , RF	RF _D , Nu_D
RF _D	0.83	3	0.018	69	-0.05	-10.67	0.25

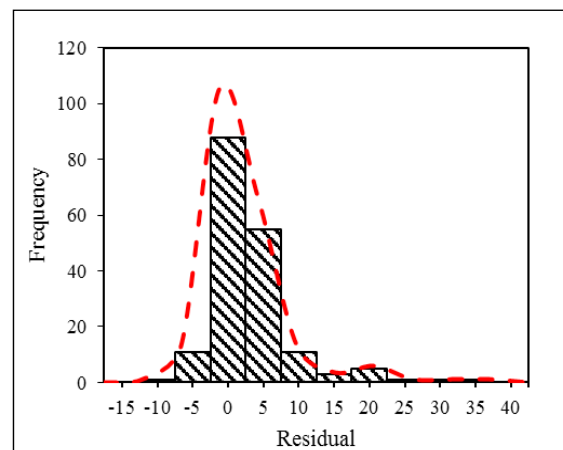
مقادیر جریان رودخانه نیز در مجموع، تنها اندکی بیش‌تر از مقادیر واقعی شبیه‌سازی شده‌اند (جدول ۳). آزمون سنجش کارایی و بهبود عملکرد مدل در نتیجه‌ی کاربرد سیگنال نویززدایی غیریکنواخت در مقایسه با بکارگیری سیگنال اصلی جریان رودخانه نشان از افزایش معنادار دقت الگوریتم جنگل‌های تصادفی در شبیه‌سازی جریان رودخانه خرم‌آباد دارد. چنان‌که آماره آزمون مرگان - گرنجر - نیوبلد در نتیجه آزمون دقت خروجی دو مدل مذکور به مقدار ۱۰/۶۵- محاسبه شد.



شکل ۵- توزیع فراوانی خطای محاسباتی مدل ϵ

Fig 5. The frequency distribution of computational error for

ϵ model



شکل ۶- توزیع فراوانی خطای محاسباتی مدل Nu

Fig 6. The frequency distribution of computational error for

Nu model

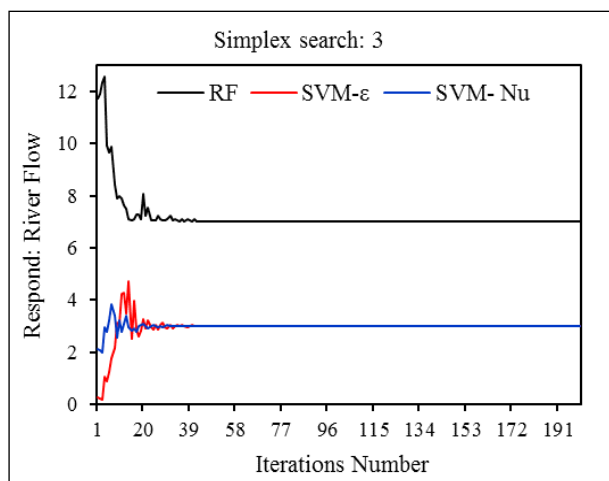
خروجی این عملیات نشان می‌دهد که دو مدل ϵ_D و Nu_D عملکرد بهتری نسبت به الگوریتم RF_D دارند. چنان‌که از شکل‌های ۹ و ۱۰ پیداست، RF_D تنها در تخمین میانگین جریان رودخانه، عملکرد موفق‌تری دارد. اما در تخمین حدود پایین و بالا جریان رودخانه از خطای شدیدی برخوردار است. در صورتی که دو مدل ϵ_D و Nu_D هم میانگین جریان و هم حد پایین و بالا جریان رودخانه را به‌طور دقیقی بهینه و برآورد می‌کنند.

ارزیابی خروجی این مدل‌های ترکیبی نیز نشان می‌دهد که میانگین و حد پایین جریان رودخانه به‌ترتیب به مقدار ۹/۶۷ و سه مترمکعب بر ثانیه به‌طور دقیقی برآورد شده‌اند. ولی حد بالای جریان رودخانه با اختلاف بسیار ناچیزی نسبت به مقدار تعیین‌شده به مقدار ۳۴ مترمکعب بر ثانیه تخمین زده شده است. به‌واقع این مقدار خطای بهینه‌سازی، به سبب وجود واریانس شدید بین دو مدل ϵ_D و Nu_D با الگوریتم RF_D در تخمین حد بالای جریان رودخانه است. در نهایت، برای ارزیابی اثر تجمعی و ترکیبی سه مدل RF_D ، ϵ_D و Nu_D ، ابتدا هر سه مدل باهم ترکیب و آن‌گاه با استفاده از الگوریتم سیمپلکس به جستجوی مقادیر بهینه میانگین جریان، حد بالا و پایین جریان رودخانه اقدام شده است (شکل‌های ۱۱ و ۱۲).

در ادامه، نتایج خروجی دو مدل ماشین‌بردار پشتیبان با نتایج خروجی الگوریتم جنگل تصادفی با استفاده از MGN آزمون شده است. آماره این آزمون به‌ترتیب برابر ۰/۰۵- و ۰/۲۵ به‌دست آمد. باتوجه به قرار گرفتن این مقادیر در ناحیه بحرانی جدول t ، می‌توان استنباط کرد شواهد کافی برای تأیید این که اختلاف معناداری در دقت جریان شبیه‌سازی شده توسط الگوریتم جنگل تصادفی با مدل‌های ماشین‌بردار پشتیبان وجود دارد، در دسترس نیست. بنابراین فرض پوچ آزمون MGN مبنی بر این که اختلاف معناداری بین دقت دو روش ϵ_D و Nu_D با الگوریتم RF_D وجود ندارد، پذیرفته می‌شود.

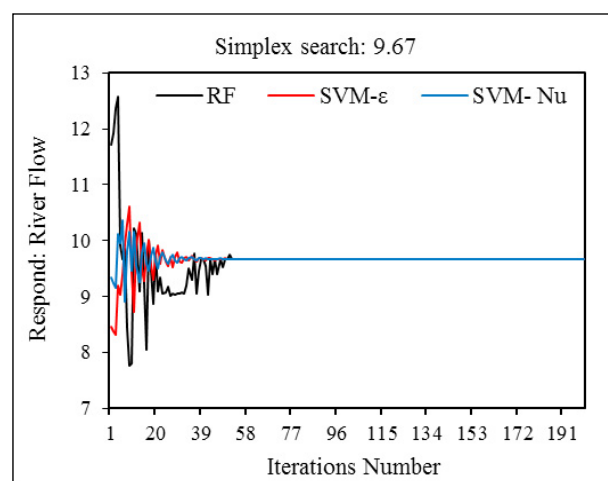
بهینه‌سازی مقدار جریان رودخانه با الگوریتم سیمپلکس

در این مرحله از پژوهش به‌منظور بهینه‌سازی مقدار جریان پیش‌بینی‌شده توسط سه مدل RF_D ، ϵ_D و Nu_D از الگوریتم سیمپلکس استفاده شد. به این ترتیب با اجرای این مدل‌ها و کاربرد الگوریتم سیمپلکس، به جستجوی مقدار بهینه میانگین (۹/۶۷ مترمکعب بر ثانیه)، صدک اول (سه مترمکعب بر ثانیه) و صدک ۹۹ (۳۶ مترمکعب بر ثانیه) جریان رودخانه پرداخته شده است. بر این اساس آشکار می‌شود که کاربرد الگوریتم سیمپلکس بر روی هر کدام از این مدل‌ها، تا چه مقدار قادر به برآورد بهینه میانگین جریان، حدپایین و حد بالای جریان رودخانه است.



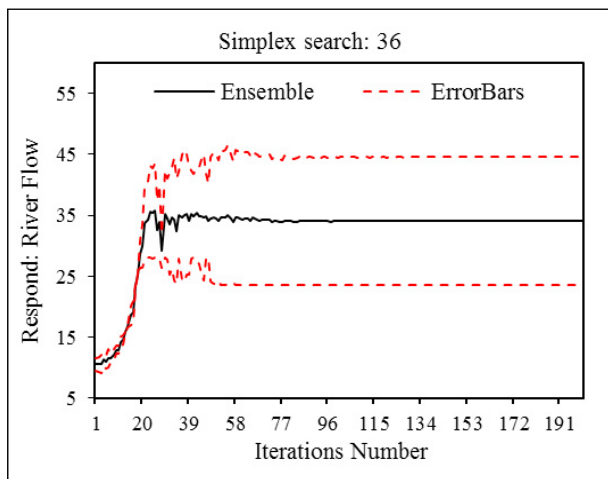
شکل ۱۰- نمودار الگوریتم سیمپلکس در تخمین حد پایین جریان ماهانه

Fig 10. Simplex algorithm diagram for estimating low flow

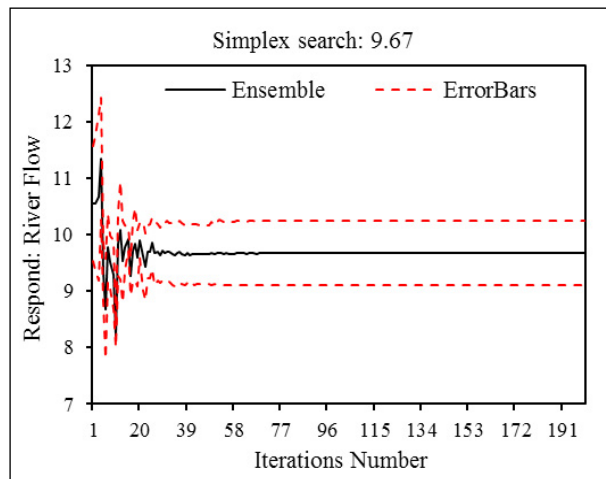


شکل ۹- نمودار الگوریتم سیمپلکس در تخمین میانگین جریان ماهانه

Fig 9. Simplex algorithm diagram for estimating average flow



شکل ۱۲ - نمودار ترکیبی سیمپلکس در تخمین حد بالا جریان ماهانه
Fig 12. Simplex ensemble diagram for estimating high flow



شکل ۱۱ - نمودار ترکیبی سیمپلکس در تخمین میانگین جریان ماهانه
Fig 11. Simplex ensemble diagram for estimating average flow

رودخانه به طور معناداری نسبت به ϵ ، Nu و RF افزایش پیدا کرده است. به این ترتیب که آماره آزمون مرگان - گرنجر - نیولند حاصل از آزمون معناداری اختلاف صحت نتایج جریان شبیه سازی ϵ ، ϵ_D ، Nu ، Nu_D ، RF و RF_D به دست آمد. بنابراین با توجه به قرار نگرفتن این مقادیر در ناحیه بحرانی جدول t می توان نتیجه گرفت که شواهد کافی برای رد فرضیه صفر و پذیرش فرضیه مقابل وجود دارد. به واقع این دستاورد که با نویززدایی سیگنال جریان رودخانه، عملکرد ماشین بردار پشتیبان در فرایند شبیه سازی جریان رودخانه به طور معناداری افزایش پیدا می کند با نتایج لی و همکاران [۱۹]، ابراهیم و احمد [۱۴] نورانی و همکاران [۲۴] هم خوانی دارد. البته لازم به ذکر است که در مطالعه های یادشده، کل سری زمانی را در هر سطح از سیگنال فرکانسی به صورت یکنواخت نویززدایی کرده اند و مشخص نشده که آیا نویز در سری زمانی به صورت یکنواخت یا غیریکنواخت توزیع شده است.

از سویی مشخص شد که اگرچه الگوریتم RF_D در مجموع مقدار جریان رودخانه را بیش تر از مقدار مشاهداتی شبیه سازی کرده است. اما توزیع فراوانی خطاهای این مدل نشان داد که این الگوریتم در مجموع، جریان رودخانه را کم تر از مقادیر مشاهداتی شبیه سازی کرده است. نتیجه این تناقض آن است که الگوریتم جنگل تصادفی در مقابل مقادیر حدی رودخانه نتوانسته انعطاف مناسبی نشان دهد و این الگوریتم تنها از الگوی کلی جریان رودخانه پیروی کرده است. از سویی با وجود آن که خطای خروجی دو مدل Nu_D ، ϵ_D به ترتیب به مقدار $3/1$ و $3/2$ مترمکعب بر ثانیه اندکی بیش تر از خطای الگوریتم RF_D به مقدار $2/9$ مترمکعب بر ثانیه است. اما توزیع زمانی خطای این دو مدل هم راستا با مجموع کلی خطاهای آنهاست. نتیجه آن است که این دو مدل، انعطاف پذیری یکسانی در مقابل مقادیر حدی و غیر حدی نشان داده اند. چنان که با کاربرد الگوریتم بهینه ساز سیمپلکس در سه مدل Nu_D ، ϵ_D و RF_D نیز مشخص شد

بحث و نتیجه گیری

از دیرباز، شبیه سازی و پیش بینی مقدار جریان رودخانه به عنوان یک مسئله مهم برای متخصصان و برنامه ریزان علوم آب محسوب می شود. از سویی، جریان رودخانه از چنان ماهیت پیچیده ای برخوردار است که عملاً شبیه سازی آن حتی با استفاده از روش های هوشمند و غیرخطی، دستخوش خطاهای محاسباتی خواهد شد. در این مطالعه به منظور شبیه سازی جریان ماهانه رودخانه خرم آباد از چندین روش هوشمند بهره گرفته شد. به این ترتیب، نخست با انتقال سیگنال این رودخانه در محیط موجک مشخص شد که نویز غیریکنواختی در سری زمانی آن وجود دارد. این نویز غیریکنواختی بیانگر آن بود که سری زمانی جریان رودخانه علاوه بر این که حاوی نوساناتی با فرکانس بسیار بالا است، فرکانس این نوسان ها در طی زمان نیز تغییر می کند که این فرایند منجر به افزایش پیچیدگی ماهیت سیگنال جریان رودخانه شده است. بنابراین به منظور ارزیابی این مهم که آیا نویززدایی غیریکنواخت جریان رودخانه موجب بهبود عملکرد الگوریتم جنگل تصادفی (RF) و ماشین های بردار پشتیبان (Nu و ϵ) در شبیه سازی جریان ماهانه رودخانه می شود، دو گروه مدل سازی با ورودی سیگنال اصلی (ϵ ، Nu و RF) و سیگنال نویززدایی شده (ϵ_D ، Nu_D و RF_D) انجام گرفت.

نتایج حاصل از ارزیابی دقت سه مدل: ϵ ، Nu و RF نشان داد که میزان همبستگی بین مقادیر مشاهداتی و محاسباتی در هر سه مدل برابر است. اما RF با سوگیری مثبت و خطای محاسباتی $5/8$ مترمکعب بر ثانیه و دو مدل ϵ ، Nu با سوگیری منفی و خطای محاسباتی $5/8$ مترمکعب بر ثانیه جریان رودخانه را شبیه سازی کرده اند.

در ادامه، سیگنال نویززدایی شده به عنوان سیگنال ورودی این سه روش (ϵ_D ، Nu_D و RF_D) در نظر گرفته شد. ارزیابی معیارهای کارائی مدل نشان داد که با نویززدایی غیریکنواخت سیگنال جریان رودخانه، عملکرد سه مدل ϵ_D ، Nu_D و RF_D در شبیه سازی جریان

10. Friedman, J. Hastie, T. and Tibshirani, R. 2001. The elements of statistical learning. Vol. 1r: Springer series in statistics New York, NY, USA:.

11. Gao, J. Sultan, H. Hu, J. and Tung, W.W. 2010. Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison. IEEE signal processing letters. 17(3): 237-240.

12. Hajizadeh, R. Farzin, S. and Hassanzadeh, Y. 2016. Intelligent Models Performance Improvement Based on Wavelet Algorithm and Logarithmic Transformations in Suspended Sediment Water and Soil 30(1): 112-124. (In Persion).

13. Hassanzadeh, Y. Lotfollahi Yaghin, M.A. Shahverdi, S. Farzin, S. and Farzin, N. 2013. De-noising and Prediction of Time Series Based on the Wavelet Algorithm and Chaos Theory (Case Study: SPI Drought Monitoring Index of Tabriz City). Iran Water Resources Research. 8(3): 1-13. (In Persion).

14. Ibrahim, N. and Ahmad, N. 2017. Comparative performance of support vector regressions for accurate streamflow predictions. Malaysian Journal of Fundamental and Applied Sciences. 13(4-1): 325-330.

15. Isazadeh, M. ahmadzadeh, h. and Ghorbani, M.A. 2016. Assessment of Kernel Functions Performance in River Flow Estimation using Support Vector Machine. Journal of Water and Soil Conservation. 23(3): 69-89. (In Persion).

16. Khorshiddoust, A. Jahanbakhsh Asl, S. Abbasi, H. Farzin, S. and Mirhashemi, H. 2017. Prediction of Evaporation Potential through Data De-noising in Tabriz Synoptic Station. Water and Soil Science. 26(4.1): 105-118. (In Persion).

17. Kişi, Ö. 2005. Daily river flow forecasting using artificial neural networks and auto-regressive models. Turkish Journal of Engineering and Environmental Sciences. 29(1): 9-20.

18. Kumar, P.S. Praveen, T. and Prasad, A. 2016. Artificial neural network model for rainfall-runoff-a case study. International Journal of Hybrid Information Technology. 9(3): 263-272.

19. Li, X. Sha, J. Li, Y.M. and Wang, Z.L. 2018. Comparison of hybrid models for daily

که الگوریتم RF_D در تخمین مقدار حد پایین و حد بالای جریان رودخانه از عملکرد ضعیفی نسبت به Nu_D ، ϵ_D برخوردار است. ازسویی، با ترکیب سه مدل مذکور و ارزیابی اثر تجمعی آنها آشکار شد که با وجود واریانس شدید بین مدل RF_D با دو مدل ϵ_D ، Nu_D مقدار حد بالای جریان با انحراف ناچیزی نسبت به مقدار مورد انتظار به وسیله مدل ترکیبی بهینه شده است.

منابع

1. Adamowski, J. and Prasher, S.O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data/Porównanie metod uczenia maszynowego do prognozowania spływu w zlewniach górskich na podstawie ograniczonych danych. Journal of water and land development. 17(1): 89-97.

2. Ahani, A. and Shourian, M. 2017. Prediction of Monthly Streamflow Using Data-driven Models. Iran Water Resources Research. 13(2): 207-214. (In Persion).

3. Ahmadi, F. Radmanesh, F. and Mirabbasi Najaf Abadi, R. 2015. Comparison between Genetic Programming and Support Vector Machine Methods for Daily River Flow Forecasting (Case Study: Barandoozchay River). Water and Soil. 28(6): 1162-1171. (In Persion).

4. Bishop, C. and Bishop, C.M. 1995. Neural networks for pattern recognition. Oxford university press.

5. Breiman, L. 2001. Random forests. Machine learning. 45(1): 5-32.

6. Chu, H. Wei, J. Li, T. and Jia, K. 2016. Application of Support Vector Regression for Mid-and Long-term Runoff Forecasting in "Yellow River Headwater" Region. Procedia Engineering. 154: 1251-1257.

7. Cristianini, N. and Shawe-Taylor, J. 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

8. Diebold, F.X. and Mariano, R.S. 2002. Comparing predictive accuracy. Journal of Business & economic statistics. 20(1): 134-144.

9. Donoho, D.L. 1995. De-noising by soft-thresholding. IEEE transactions on information theory.

27. Polikar, R. 1999. The story of wavelets. Physics and modern topics in mechanical and electrical engineering: 192-197.
28. Ren, W.X. and De Roeck, G. 2002. Structural damage identification using modal data. II: Test verification. Journal of Structural Engineering. 128(1): 96-104.
29. Sanikhani, H. Dinpashoh, y. and Ghorbani, M. 2015. Baranduz-Chay River Flow Modeling Using the K-Nearest Neighbor and Intelligent Methods. Water and Soil Science. 25(1): 219-233. (In Persian).
30. Suykens, J.A. and Vandewalle, J. 1999. Least squares support vector machine classifiers. Neural processing letters. 9(3): 293-300.
31. Vapnik, V. 2013. The nature of statistical learning theory. Springer science & business media.
32. Vapnik, V.N. and Chervonenkis, A.Y. 2015. On the uniform convergence of relative frequencies of events to their probabilities, in Measures of complexity. Springer. p. 11-30.
33. Wang, W. and Ding, J. 2003. Wavelet network model and its application to the prediction of hydrology. Nature and Science. 1(1): 67-71.
34. Young, C.C. Liu, W.C. and Chung, C.E. 2015. Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge. Neural Computing and Applications. 26(7): 1631-1643.
- streamflow prediction in a forested basin. Journal of Hydroinformatics. 20(1): 191-205.
20. Lotfollahi-Yaghin, M.A. Shahverdi, S. and Tarinejad, R. Year. Damage detection in jacket type offshore platforms using wavelet packet transform. in The 5th Civil Engineering Conference in the Asian Region and Australasian Structural Engineering Conference.
21. Lotfollahi Yaghin, M.A. and Koohdaragh, M. 2011. Examining the function of wavelet packet transform (WPT) and continues wavelet transform (CWT) in recognizing the crack specification. KSCE Journal of Civil Engineering. 15(3): 497-506.
22. Misra, D. Oommen, T. Agarwal, A. Mishra, S.K. and Thompson, A.M. 2009. Application and analysis of support vector machine based simulation for runoff and sediment yield. Biosystems engineering. 103(4): 527-535.
23. Montaseri, M. and Ghavidel, S. 2014. River Flow Forecasting by Using Soft computing. Water and Soil. 28(2): 394-405. (In Persian).
24. Nourani, V. Andalib, G. and Sadikoglu, F. 2017. Multi-station streamflow forecasting using wavelet denoising and artificial intelligence models. Procedia Computer Science. 120: 617-624.
25. Nourani, V. Rahimi, A.Y. and Nejad, F. 2013. Conjunction of ANN and threshold based wavelet de-noising approach for forecasting suspended sediment load. International Journal of Management & Information Technology. 3(1): 9-26.
26. O'Neill, R.t. 1971. Algorithm AS 47: function minimization using a simplex procedure. Journal of the Royal Statistical Society. Series C (Applied Statistics). 20(3): 338-345.

Improvement of Support Vector Machine and Random Forest Algorithm in Predicting Khorramabad River Flow Using Non-uniform De-Noising of Data and Simplex Algorithm

H. Mirhashemi¹, D. Yarahmadi², S. Sharafi³ and S. Farzin⁴

Received: 19-08-2018 Accepted: 13-04-2019

Abstract

In this study, in order to simulate the monthly flow of the Khorramabad River, the time series of this river was decomposed into three levels using the wavelet of Daubechies-3, during the period of 1955-2014. Based on this, it was found that there is a Non-uniform noise that includes two periods of time in this signal, with the October 2008 border which required that the signal become non-uniform de-noising. Subsequently, by using two models of support vector machine (ϵ and Nu) and random forest algorithm (RF), the main signal and non-uniform de-noising signal of the river flow were simulated separately. The results validation criteria of the model showed that, with the non-uniform de-noising of the river flow signal, the error of the models dropped ϵ from 5.7 to 3.1, Nu from 5.8 to 3.2 and RF from 5 to 2.9 m³/s. Also, the comparison and testing of the computational error ϵ : ϵD ; Nu: NuD; RF: RFD were obtained by using the MGN test (-15, -15, -107.6), which indicates a significant improvement in the performance of the models used as a result of the non-uniform de-noising signal. In the following, using optimization simplex algorithm in the of three models ϵD , NuD and RFD, the mean flow of the river was very high in all three models.

Keywords: *Soft thresholding, Daubechies wavelet, Simulation, Detail signal.*

1. Assistant Professor in Climatology, Geography Sciences Department, Lorestan University. Khorramabad.

2. Associate Professor in Climatology, Geography Sciences Department, Lorestan University. Khorramabad.

3. Assistant Professor in Geomorphology, Geography Sciences Department, Lorestan University. Khorramabad.

4. Corresponding Author and Assistant Professor, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan. Email: saeed.farzin@semnan.ac.ir