

مقاله علمی-پژوهشی

پیش‌بینی زمانی و مکانی دبی جریان با استفاده از روش‌های تلفیقی هوش مصنوعی و پیش‌پردازش و پس‌پردازش سری زمانی

سید مهدی ثاقبیان^۱

تاریخ دریافت: ۱۳۹۸/۱۲/۱۱ تاریخ پذیرش: ۱۳۹۹/۳/۶

چکیده

پیش‌بینی دبی رودخانه‌ها یکی از موارد مهم در برنامه‌ریزی و مدیریت منابع آبی می‌باشد. در این تحقیق از روش‌های پیش‌پردازش و پس‌پردازش سری زمانی به همراه روش‌های مبتنی بر کرنل ماشین‌بردار پشتیبان (SVM) و رگرسیون فرآیند گاوسی (GPR) جهت تخمین دبی جریان دو رودخانه طبیعی در ایالات متحده با دو ایستگاه هیدرومتری متوالی استفاده شده است. رودخانه اول شامل تقریباً ۲ سال داده بوده و در رودخانه دوم از ۴ سال داده روزانه دبی استفاده شده است. مدل‌های متفاوتی بر اساس مشخصات هیدرولیکی جریان تعریف گردید و کارایی روش‌های تلفیقی پیش‌پردازش و پس‌پردازش در دو حالت درون ایستگاهی و بین ایستگاهی بررسی شد. جهت پیش‌پردازش داده‌ها ابتدا از روش تبدیل موجک گسسته (DWT) استفاده شد. سپس زیر سری‌های با فرکانس بالا انتخاب شده و با روش تجزیه مد تجربی یکپارچه (EEMD) دوباره تجزیه گردیدند. در نهایت زیر سری‌های با انرژی بالا به‌عنوان ورودی مدل‌های مبتنی بر کرنل استفاده شدند. برای پس‌پردازش داده‌ها نیز از مدل میانگین عصبی غیرخطی (NNA) استفاده شد. نتایج حاصل از تحلیل مدل‌های تعریف‌شده، دقت بالای روش‌های تلفیقی به‌کار رفته در تحقیق را در تخمین دبی جریان به خوبی نشان داد. بطوریکه در هر دو ایستگاه، درصد خطا با استفاده از روش‌های تلفیقی پیش‌پردازش و پس‌پردازش نسبت به روش‌های هوشمند مبتنی بر کرنل تقریباً به میزان ۲۰ تا ۲۵ درصد کاهش یافت. مشاهده شد که در حالت بررسی دبی رودخانه بر اساس داده‌های خود ایستگاه مقدار خطای RMSE مدل تقریباً از ۰/۳ به ۰/۲۶ و در حالت استفاده از داده‌های ایستگاه قبلی از مقدار ۰/۴۴ به ۰/۳۳ کاهش یافت. با توجه به قابلیت و دقت بالای روش‌های پیش‌پردازش استفاده شده در این تحقیق، مطالعات مشابه در دیگر رودخانه‌های کشور توصیه می‌گردد.

واژه‌های کلیدی: پس‌پردازش، تجزیه مد تجربی یکپارچه، دبی رودخانه، موجک

مقدمه

ابداع شده است که به طور کلی می‌توانند به دو دسته مدل‌های مفهومی و مدل‌های مبتنی بر داده یا آماری طبقه‌بندی شوند. مدل‌های آماری، کاربردی‌تر از مدل‌های مفهومی می‌باشد. زیرا درک ساختار داخلی آن‌ها، از فرآیندهای فیزیکی مدل‌سازی، مستقل است. با وجود اینکه مدل‌های مختلفی برای پیش‌بینی جریان رودخانه پیشنهاد شده است، اما با توجه به ماهیت غیرخطی و ویژگی‌های متغیر زمانی و مکانی در سیستم گردش آبی، هیچ‌یک از مدل‌های آماری و مفهومی پیشنهاد شده به‌منظور الگوسازی دقیق دبی جریان نتوانسته‌اند به‌عنوان یک مدل برتر و توانا شناخته شوند و عدم قطعیت در پیش‌بینی‌ها از اعتبار برآوردها از طریق مدل‌های مختلف می‌کاهد. از طرفی امروزه محققین، با پیشرفت روش‌های هوشمند، در جستجوی راه‌هایی برای پیش‌بینی پارامترهای مهم هیدرولیکی و هیدرولوژیکی هستند. در دهه‌های اخیر روش‌های نرم برای

با توجه به محدودیت منابع آبی در طبیعت، توزیع زمانی و مکانی غیریکنواخت منابع آبی، افزایش آلودگی‌ها، تخریب منابع طبیعی و همچنین افزایش جمعیت، رشد و توسعه جوامع شهری و فعالیت‌های کشاورزی و صنعتی، ضرورت برنامه‌ریزی و مدیریت صحیح منابع آبی اجتناب‌ناپذیر است. در این میان آگاهی از میزان آورد رودخانه یکی از مسائل مهم و کاربردی در مدیریت منابع آب و از اطلاعات ضروری برای اجرای پروژه‌های آبی است (Kemp et al., 2011). به‌منظور پیش‌بینی جریان رودخانه در طی سالیان گذشته، روش‌های مختلفی

۱- گروه عمران، دانشکده فنی و مهندسی، واحد اهر، دانشگاه آزاد اسلامی، اهر، ایران

(Email: smsaghebian@iau-ahar.ac.ir)

(*)- نویسنده مسئول:

تجربی یکپارچه کامل است. همچنین، ترکیب پیش‌بینی‌های چندین مدل رگرسیونی با استفاده از یک الگوریتم یادگیری معین می‌تواند باعث بهبود نتایج خروجی روش‌های رگرسیونی شود.

بدین ترتیب، با توجه به عدم قطعیت موجود در پیش‌بینی دبی جریان، بهره‌گیری از روش‌های جدید که بتوانند قطعیت مدل را افزایش دهند ضروری به نظر می‌رسد. لذا، هدف از این تحقیق بررسی کارایی روش‌های تلفیقی پیش-پس‌پردازش به‌عنوان روش جدید جهت بهبود دقت پیش‌بینی دبی جریان است که می‌تواند در برنامه‌ریزی و مدیریت بهینه منابع آبی مورد استفاده قرار گیرد. با استفاده از روش‌های تجزیه سیگنال حجم محدودی از نویز سفید وارد سیگنال اصلی شده و با استفاده از جنبه‌های مثبت آماری نویز سفید که توزیع متوازن در حوزه فرکانس دارد، اثر نویز متناوب از روند تجزیه حذف خواهد شد. در این تحقیق، از روش‌های تبدیل موجک گسسته (DWT) و تجزیه مد تجربی یکپارچه (EEMD) و مدل میانگین عصبی غیرخطی (NNA) برای پیش و پس‌پردازش مدل‌ها استفاده شده است. لذا، جهت بررسی کارایی روش‌های به‌کار رفته در تحقیق، مدل‌هایی بر اساس دبی جریان روزهای گذشته برای چهار ایستگاه هیدرومتری واقع در رودخانه‌های هوستونیک و می‌سی‌سی‌پی در ایالات متحده تعریف شده و دو حالت تخمین دبی جریان بر اساس داده‌های خود ایستگاه و داده‌های ایستگاه قبلی مورد بررسی قرار گرفته است.

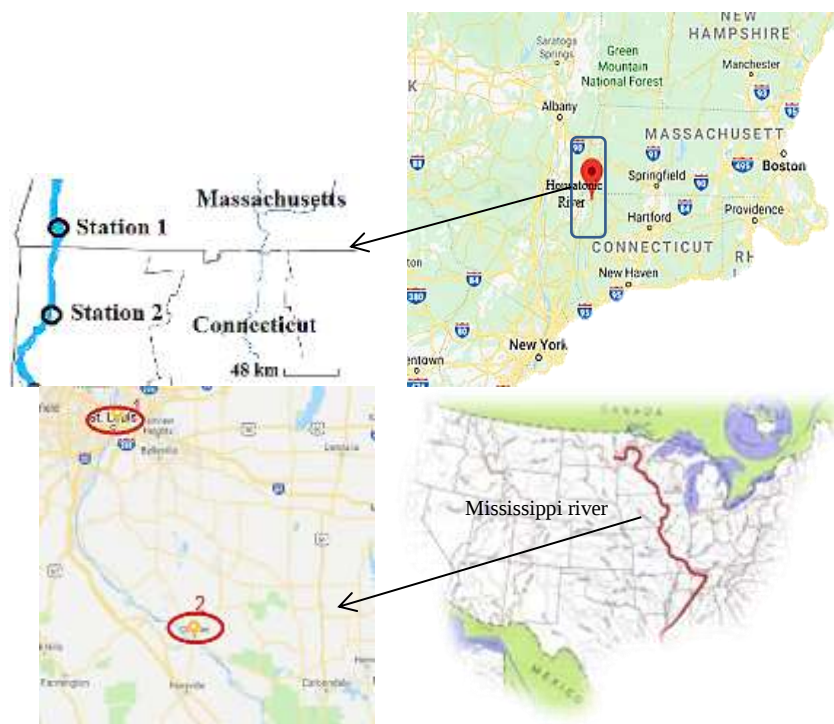
مواد و روش‌ها

منطقه مورد مطالعه

در این تحقیق از داده‌های دبی جریان روزانه‌ی دو حوضه استفاده شده است. حوضه اول شامل دو ایستگاه از رودخانه‌ی هوستونیک (Houstone) در ایالات متحده است که فاصله‌ی بین آن‌ها حدود ۵۰ کیلومتر است. اولین ایستگاه نزدیک برینگتون بزرگ در ایالت ماساچوست و ایستگاه دوم در روستای فالز در ایالت کانکتیکات واقع شده است. مساحت حوضه به ترتیب برای ایستگاه‌ها ۲۸۲ مایل مربع و ۶۳۴ مایل مربع می‌باشد. داده‌های استفاده شده در این ایستگاه‌ها به‌صورت روزانه و حدود دو سال داده می‌باشد. حوضه دوم شامل رودخانه می‌سی‌سی‌پی (Mississippi) با میانگین دبی ۶۱۸۷ مترمکعب بر ثانیه برای محدوده زمانی ۲۰۰۵-۲۰۰۸ است. این رودخانه با داشتن طول ۳۷۳۷ کیلومتر، دومین رودخانه طویل در قاره آمریکا شمالی است که از ایالت‌های مینه‌سوتا، میزوری و آرکانزاس عبور می‌کند. موقعیت ایستگاه‌ها در شکل (۱) نشان شده است. داده‌ها از سایت USGS به نشانی <https://waterdata.usgs.gov> اخذ گردیده است.

شبیه‌سازی رفتار سیستم‌ها در مهندسی آب مورد استفاده قرار گرفته‌اند. این روش‌ها اکثراً الهام گرفته از پدیده‌های موجود در طبیعت هستند و به‌عنوان ابزاری توانمند در حل مسائل پیچیده به شمار می‌آیند (ASCE, 2000). به‌عنوان نمونه سیواپراگسم و لیونگ از مدل ماشین بردار پشتیبان برای پیش‌بینی بارش-رواناب حوضه Tryggevaelde واقع در کشور هلند استفاده کردند (Sivapragasam, & Liong, 2001). اکبری الاشتی و بزرگ حداد (۱۳۹۱) با استفاده از روش‌های مبتنی بر هوش مصنوعی به پیش‌بینی ماهانه‌ی دبی جریان به صورت زمانی پرداختند. نیک‌منش (۱۳۹۴) به پیش‌بینی آبدی متوسط ماهانه رودخانه‌ی کر، با استفاده از مدل تلفیقی شبکه عصبی مصنوعی و تبدیل موجک پرداخت. معروف‌پور و همکاران (۱۳۹۶) از روش‌های هوش مصنوعی برای شبیه‌سازی تبخیر-تعرق استفاده کردند. روشگر و قاسم‌پور (۱۳۹۷) با استفاده از روش‌های هوشمند رگرسیون بردار پشتیبان و سیستم استنتاج عصبی-فازی انطباقی مقاومت جریان را در رودخانه‌های ماریچی بررسی کردند.

بررسی منابع صورت گرفته، نشان از کاربرد روش هوشمند شبکه عصبی مصنوعی در پیش‌بینی سری‌های زمانی از جمله دبی جریان می‌باشد. آداموفسکی و سان به‌منظور تخمین جریان چند رودخانه غیردائمی واقع در حوضه‌های آبریز نیمه‌خشک کشور قبرس استفاده کردند. نتایج نشان داد که روش تلفیقی شبکه عصبی و آنالیز موجک یک روش موثر در تخمین جریان در حوضه مورد مطالعه می‌باشد (Adamowski and Sun, 2010). ناوه و همکاران (۱۳۹۱) الگوی دوخطی را برای پیش‌بینی جریان رودخانه‌های شهرچای و باراندوزچای ارومیه استفاده کردند. نتایج نشان از دقت بالای این الگو در مقایسه با الگوهای خطی بود. خلیلی و همکاران جریان روزانه رودخانه شهرچای ارومیه را با الگوی غیرخطی تلفیقی BL-ARCH پیش‌بینی کردند (Khalili et al., 2013). نتایج نشان داد که الگوی تلفیقی عملکرد بسیار بهتری نسبت به الگوی دوخطی داشته و مقدار خطا را به میزان قابل توجهی کاهش می‌دهد. زمانی و همکاران (۱۳۹۳) به تخمین دبی روزانه رودخانه کارون با روش‌های هوشمند برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی پرداختند و نتایج حاکی از عملکرد مناسب این روش‌ها نسبت به روش‌های خطی بود. این در حالی است که مطالعات چندانی در زمینه کاربرد روش‌های مبتنی بر کرنل مانند رگرسیون فرآیند گاوسی در تخمین دبی جریان انجام نشده است. از طرفی، در سال‌های اخیر استفاده از روش‌های تجزیه سیگنال به‌عنوان روشی نوین و مؤثر در زمینه تحلیل سیگنال‌ها و سری‌زمانی مورد توجه قرار گرفته است (Adamowski et al., 2009; Najah et al., 2011). در این روش سیگنال اصلی به چندین زیرسری تجزیه شده و داده‌های حاصل از تجزیه با معادلات مناسب برازش می‌یابد و مدل به‌دست‌آمده برای پیش‌بینی استفاده می‌شود. از جمله این روش‌ها تبدیل موجک و روش جدید تجزیه مد



شکل ۱- موقعیت ایستگاه‌های استفاده شده در تحقیق

روش‌های مبتنی بر کرنل

ضعف عملکرد روش‌های استخراج ویژگی خطی برای داده‌های پیچیده که به صورت خطی تفکیک‌پذیر نمی‌باشند، سبب پیدایش روش‌های غیرخطی از جمله روش‌های مبتنی بر کرنل برای این منظور شده است. در روش‌های مبتنی بر کرنل ابتدا داده‌ها به فضایی با ابعاد بیشتر نگاشت یافته که در این فضا به طور خطی تفکیک‌پذیر می‌شوند. این روش‌ها بر این ایده استوارند که ضرب داخلی داده‌های نگاشت یافته با یک تابع غیرخطی را می‌توان با تابع کرنل تخمین زد، که به آن حقه کرنل می‌گویند. از جمله روش‌های مبتنی بر کرنل می‌توان به ماشین بردار پشتیبان (SVM) و رگرسیون فرآیند گاوسی (GPR) اشاره کرد. الگوریتم SVM، جزء الگوریتم‌های تشخیص الگو دسته‌بندی می‌شود که برای اولین بار توسط وپنیک در سال ۱۹۹۵ معرفی شد (Vapnik, 1995). نحوه‌ی کارکرد الگوریتم SVM مانند اغلب روش‌های هوشمند به صورت آموزش (Train) و آزمون (Test) می‌باشد. مزیتی که SVM نسبت به سایر الگوریتم‌ها دارد این است که این الگوریتم مبنی بر احتمالات نمی‌باشد و همواره جواب یکه‌ای به همراه دارد و زمان کمتری را برای دستیابی به جواب بهینه به خود اختصاص می‌دهد. از الگوریتم SVM، در هر جایی که نیاز به تشخیص الگو یا دسته‌بندی اشیاء در کلاس‌های خاص باشد می‌توان استفاده کرد. در واقع مبنای کاری SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌شود خطی انتخاب شود که حاشیه

اطمینان بیشتری داشته باشد. هدف اصلی الگوریتم SVM این است که خطی را بیابد که از داده‌های موجود در دو کلاس دارای بیشترین فاصله باشد و یا به عبارت دیگر دارای کمترین ریسک عملیاتی باشد. فرایند گاوسی نیز روشی مفید به منظور تعریف توزیع‌های مقدم برای مدل‌های انعطاف‌پذیر رگرسیون و طبقه‌بندی است که در آن رگرسیون یا توابع احتمال طبقه به اشکال ساده پارامتری محدود نشده باشند. فرایند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد دلخواه محدودی از آن‌ها با توزیع‌های گاوسی ادغام شده‌اند. فرایند گاوسی به طور کامل توسط تابع میانگین $m(x)$ و تابع کوواریانس $k(x, x')$ آن تعیین می‌شود. این فرایند، تعمیمی طبیعی از توزیع گاوسی است که میانگین و کوواریانس آن به ترتیب یک بردار و یک ماتریس است. مدل‌های رگرسیون فرایند گاوسی بر مبنای این فرض هستند که مشاهدات تنظیم باید حامل اطلاعاتی درباره هم‌دیگر باشد. یکی از جذابیت‌های فرایند گاوسی تنوع بسیار توابع کوواریانس آن است که به ایجاد توابعی با درجات مختلف همواری یا انواع مختلف ساختارهای پیوستی منجر می‌شود و به محقق این امکان را می‌دهد که از میان آن‌ها گزینش مناسبی انجام دهد. این مدل‌ها می‌توانند توزیع‌هایی را بین توابعی با تعداد یک یا چندین متغیر ورودی مشخص کنند. هنگامی که چنین تابعی میانگین پاسخ را در یک مدل رگرسیونی با خطاهای گاوسی تعریف می‌کند، می‌توان از محاسبات ماتریس برای استنتاج بهره‌گیری کرد؛ این مسئله برای مجموعه

در آن نوعی توصیف زمان-مقیاس از سیگنال گسسته با استفاده از فیلترهای دیجیتال ارائه می‌گردد. تبدیل موجک، حاصل شباهت‌سنجی (کوریولیشن) بین محتوای فرکانسی (مقیاسی) سیگنال و تابع موجک در مقیاس‌های مختلف است. تجزیه کردن سری‌های زمانی با استفاده از فیلترهای مشخص (موجک و توابع مقیاس‌گذاری شده) دو نوع زیرسری را به وجود می‌آورند؛ بردارهای تقریبی یا باقی‌مانده و بردارهای جزئیات (Chou, 2011). این ضرایب حاصل از در هم پیچیدن سری زمانی اصلی با فیلتر پایین-گذر و فیلتر بالا-گذر می‌باشد. فیلتر پایین-گذر همان تابع مقیاس‌گذاری است و فیلتر بالاگذر همان تابع موجک است. تابع تبدیل موجک قابلیت تجزیه سری زمانی به چندین زیرسری زمانی با مقیاس‌های مختلف را دارد و با بررسی زیرسری‌های زمانی منتجه از سری زمانی کلی، رفتار کوچک مقیاس و بزرگ مقیاس یک فرآیند هیدرولوژیکی را آنالیز می‌کند. روابطی که بیشتر برای تخمین سطح تجزیه مورد استفاده قرار می‌گیرد به صورت زیر است که در آن L سطح تجزیه، N تعداد نقاط داده در سری زمانی و p برابر با زمان محو شونده برای موجک مادر انتخاب شده است.

$$L = \text{int}[\log(N)] \quad , \quad L = \frac{\log(\frac{N}{2^{p-1}})}{\log 2} \quad (3)$$

روش تجزیه مد تجربی EMD

روش تجزیه مد تجربی روشی برای تجزیه سیگنال‌های گوناگون است که این عمل در فرآیندی تحت عنوان غربال کردن صورت می‌گیرد. طی این فرایند سیگنال اصلی به تعدادی مولفه با محتوای بسامدی متفاوت تجزیه می‌شود. روش EMD طبق رابطه (۴) سیگنال اصلی $x(n)$ را به تعدادی مد ذاتی (IMF) تجزیه می‌کند (Amirat et al., 2018).

$$x(n) = \sum_{i=1}^n c_i(x) + r_n(x) \quad (4)$$

که $r_n(x)$ همان مولفه باقی‌مانده، بعد از n تعداد IMFها و $c_i(x)$ است. یک داده ممکن است در یک زمان دارای چندین مد ذاتی باشد. این مدهای نوسانی، تابع‌های مد ذاتی [Intrinsic Mode Functions (IMF)] نام دارند و دارای دو شرایط هستند: ۱) در کل داده‌ها، تعداد نقاط فرین (اکسترمم) و نقاط صفر با هم برابر و یا حداکثر دارای یک واحد اختلاف هستند، ۲) در هر نقطه میانگین پوش برازش داده شده بر نقاط بیشینه محلی و پوش برازش داده شده بر نقاط کمینه محلی باید صفر باشد. به علت وجود تناوب و نویز در سیگنال‌ها، در برخی موارد به علت اختلاط مدها؛ توزیع حوزه زمان-فرکانس دچار انقطاع می‌شود و عملکرد EMD دچار نقص می‌گردد (چون میانگین IMFها حالتی نامشخص به خود می‌گیرد). جهت رفع

داده‌هایی با بیش از هزار نمونه امکان‌پذیر است. طراحی روش‌های مبتنی بر کرنل شامل استفاده از مفهوم تابع کرنل است. در واقع، با یک تبدیل غیرخطی از فضای ورودی به فضای خصیصه با ابعاد بیشتر (حتی نامتناهی) می‌توان مسائل را به صورت خطی تفکیک‌پذیر ساخت. از مهم‌ترین توابع کرنل می‌توان به کرنل خطی، چندجمله‌ای ساده، چندجمله‌ای نرمال‌شده، تابع شعاع محور، تابع کرنل پیرسون و سیگموئید اشاره کرد.

تبدیل موجک

موجک (Wavelet) دسته‌ای از توابع ریاضی است که برای تجزیه سیگنال پیوسته به مؤلفه‌های فرکانسی آن بکار می‌رود که سطح تفکیک هر مؤلفه برابر با مقیاس آن است. تبدیل موجک تجزیه یک تابع بر مبنای توابع موجک است. موجک‌ها (که به عنوان موجک‌های مادر شناخته می‌شوند) نمونه‌های انتقال یافته و مقیاس شده یک تابع (موجک مادر) با طول متناهی و نوسانی شدیداً میرا هستند. تابع موجک، تابعی است که دو ویژگی مهم نوسانی بودن و کوتاه‌مدت بودن را دارد. $\Psi(x)$ تابعی موجک است اگر و فقط اگر تبدیل فوریه آن $\Psi(\omega)$ شرط زیر را ارضا کند (Lau and Weng, 1995).

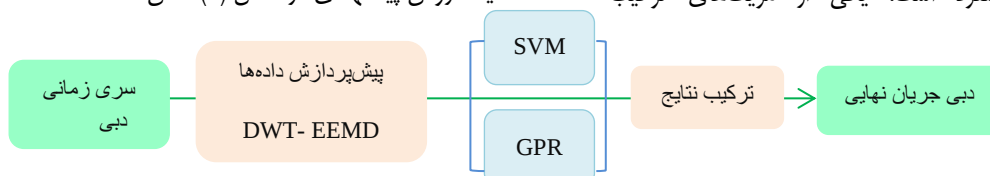
$$\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|}{|\omega|^2} d\omega < +\infty \quad (1)$$

این شرط با عنوان شرط پذیرفتگی برای موجک $\Psi(x)$ شناخته می‌شود $\Psi(x)$ تابع موجک مادر است (رابطه ۲) که ضرایب a و b در آن نقش انتقال و مقیاس سیگنال مورد نظر را دارند.

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \quad (2)$$

توابع موجک دارای انواع بسیاری هستند که مهم‌ترین و پرکاربردترین آن‌ها شامل تابع موجک Haar، Meyer، Mexican hat، Coiflet، Daubechies، Symlet، Morle پیوسته به عنوان روش جایگزین تبدیل فوریه زمان-کوتاه ارائه گردید و هدف آن، فائق آمدن بر مشکلات مربوط به تفکیک‌پذیری در تبدیل فوریه زمان-کوتاه است. در عمل اکثر سری‌های زمانی هیدرولوژیکی به صورت گسسته در زمان‌هایی مشخص اندازه‌گیری می‌شوند. بنابراین کاربرد موجک گسسته می‌تواند تناسب بهتری با این نوع سری زمانی داشته باشد. تبدیل موجک گسسته در عمل همان تبدیل موجک پیوسته است (نسخه گسسته شده آن) که در آن یک سری موجک از تبدیل موجک پیوسته نمونه گرفته است و سیستم‌های کامپیوتری قابلیت محاسبه آن را با هزینه زمانی کمتر دارا است. اصول تبدیل موجک گسسته به روشی تحت عنوان کدینگ زیرباند برمی‌گردد که در سال ۱۹۷۶ سنگ بنای اولیه آن گذارده شد و

پیش‌بینی‌ها این است که وقتی از چندین روش استفاده می‌شود نتایج نسبت به انتخاب خاص روش‌ها بسیار حساس نیستند. به این معنا که استفاده از پیش‌بینی ترکیبی از اتکا به یک روش منفرد، ایمن‌تر و ریسک کمتری دارد. روش‌های مختلفی جهت ترکیب نتایج مدل‌های منفرد وجود دارد که از آن جمله می‌توان به میانگین‌گیری ساده، میانگین‌گیری وزن‌دار و روش غیرخطی اشاره کرد. در مطالعه کنونی، برای بهبود دقت پیش‌بینی سری زمانی دبی جریان، از روش شبکه عصبی غیرخطی (NNEM) به منظور ترکیب خروجی‌های مدل‌های مبتنی بر کرنل استفاده شده است. در مقایسه با روش‌های ترکیبی خطی، در روش شبکه عصبی غیرخطی یک الگوریتم شبکه عصبی مصنوعی (ANN) برای ترکیب کردن غیرمستقیم خروجی‌های مدل‌های GPR و SVM آموزش داده می‌شود. لایه ورودی مدل شبکه عصبی توسط خروجی‌های دو مدل در نظر گرفته شده تغذیه می‌شود، که هر یک در لایه ورودی به یک نورون اختصاص می‌یابد. شماییک روش پیشنهادی در شکل (۲) نشان داده شده است.



شکل ۲- شماییک مدل‌سای در نظر گرفته شده در تحقیق

پیش‌بینی‌شده و N تعداد داده‌هاست. وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت شبکه می‌شود لذا نرمالیزه کردن داده‌ها کمک شایانی به آموزش بهتر و سریع‌تر مدل می‌کند. در تحقیق کنونی داده‌ها مطابق رابطه (۸) نرمالیزه شدند. که در آن $x_{\min}$ و $x_{\max}$ به ترتیب حداقل و حداکثر مقدار داده‌های مشاهده شده و x_n داده نرمال شده می‌باشد. محدوده-ی نرمالیزه کردن داده‌ها در این تحقیق بین ۰/۱ و ۱ انتخاب گردید.

$$x_n = 0.1 + 0.9 \times \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) \quad (8)$$

مدل‌سازی و نتایج و بحث

با توجه به اینکه انتخاب مدل‌های ورودی در سامانه‌های هوشمند می‌تواند بر روی دقت جواب‌های حاصل از تحلیل تأثیر گذارد، لذا سعی گردید پارامترهای مناسب و مؤثر در مدل‌سازی دبی جریان انتخاب گردد. جهت تخمین دبی جریان در ایستگاه اول از داده‌های روزانه مربوط به همان ایستگاه با در نظر گرفتن ترکیبی از داده‌های دبی جریان با تأخیرهای زمانی مختلف و در ایستگاه‌های دوم از داده‌های مربوط به آن ایستگاه و ایستگاه قبل استفاده شد. جدول ۱ مدل‌های تعریف شده در تحقیق را نشان می‌دهد. در این جدول Qw بیانگر دبی جریان است و اندیس t تأخیر زمانی را بر حسب روز نشان

این ایراد، وو و هوانگ روشی متفاوت با عنوان روش تجزیه مد تجربی یکپارچه (EEMD) را پیشنهاد کردند (Wu and Huang, 2004). در روند تجزیه سازی آن، حجم محدودی از نویز سفید وارد سیگنال اصلی می‌شود. با استفاده از جنبه‌های مثبت آماری نویز سفید که توزیع متوازن در حوزه فرکانس دارد، اثر نویز متناوب از روند تجزیه حذف می‌شود.

پس‌پردازش سری زمانی

بسیاری از محققان بر این باورند که ترکیب نتایج خروجی چند مدل منفرد به عنوان یک رویکرد پس‌پردازش می‌تواند ابزار مفیدی برای بهبود دقت کلی پیش‌بینی سری زمانی باشد (Makridakis and Winkler, 1983). هدف از روش‌های یکپارچه‌سازی (Ensembling) ترکیب پیش‌بینی‌های چندین مدل رگرسیونی با استفاده از یک الگوریتم یادگیری معین به منظور بهبود نتایج خروجی روش‌های رگرسیونی منفرد است. یکی از مزیت‌های ترکیب

معیارهای ارزیابی مدل‌ها

به منظور ارزیابی کارایی روش‌های استفاده شده در تحقیق، از سه پارامتر آماری استفاده گردید که عبارتند از ضریب همبستگی بین مقادیر مشاهداتی و محاسباتی (R)، ضریب تبیین (DC) و ریشه میانگین مربعات خطاها ($RMSE$) هر چه مقدار DC برای یک مدل به یک نزدیک‌تر و مقدار $RMSE$ کوچک‌تر باشد به معنی مطلوب بودن آن مدل است. روابط این پارامترهای آماری به صورت زیر می‌باشد:

$$R = \frac{\sum_{i=1}^N (l_o - \bar{l}_o) \times (l_p - \bar{l}_p)}{\sqrt{\sum_{i=1}^N (l_o - \bar{l}_o)^2 \times \sum_{i=1}^N (l_p - \bar{l}_p)^2}} \quad (5)$$

$$DC = 1 - \frac{\sum_{i=1}^N (l_o - l_p)^2}{\sum_{i=1}^N (l_o - \bar{l}_p)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (l_o - l_p)^2}{N}} \quad (7)$$

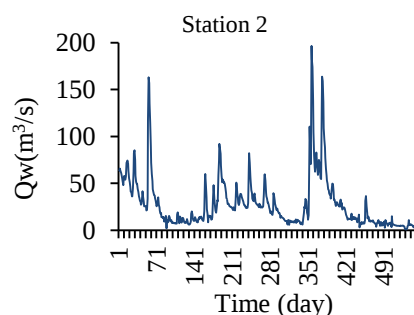
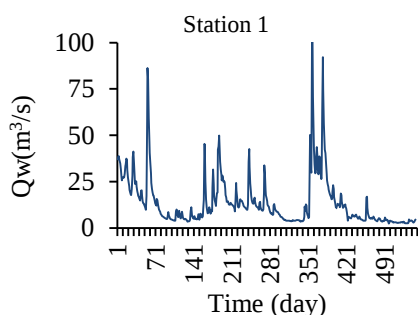
در روابط بالا l_o : مقدار اندازه‌گیری شده، $\bar{l}_o$: متوسط مقدار اندازه‌گیری شده، l_p : مقدار پیش‌بینی شده، $\bar{l}_p$: متوسط مقدار

شکل (۴) برای سری داده‌های آزمون ارائه گردیده است. مطابق با شکل (۴) مشاهده می‌گردد که در بین تمامی توابع کرنل‌ها، تابع کرنل شعاع مداری در پیش‌بینی دبی جریان نسبت به توابع کرنل دیگر دارای خطای کمتری بوده و نتایج دقیق‌تری را ارائه کرده است. بنابراین، در تجزیه و تحلیل تمامی مدل‌ها از کرنل شعاع مداری استفاده شد.

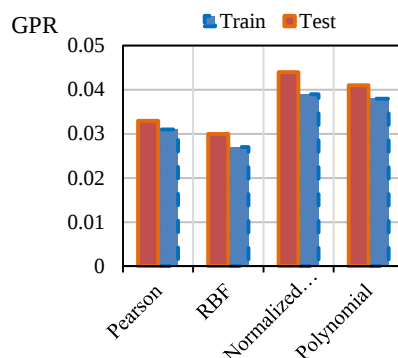
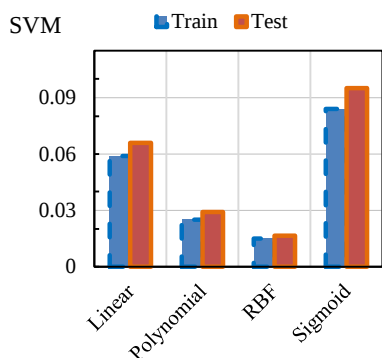
می‌دهد. همچنین در شکل (۳) نمودار مقادیر دبی جریان در دو ایستگاه انتخابی در رودخانه Houstonic نشان داده شده است. جهت انتخاب تابع کرنل مناسب رگرسیون فرایند گاوسی و روش ماشین بردار پشتیبان مدل M3 در ایستگاه اول رودخانه Houstoni با کرنل‌های مختلف مورد ارزیابی قرار گرفت. از معیار ارزیابی RMSE جهت تعیین کارایی هر یک از کرنل‌ها استفاده گردید. نتایج به صورت

جدول ۱- مدل‌های تعریف شده در تحقیق

نام مدل	ساختار مدل	نام مدل	ساختار مدل
مدلسازی بر اساس داده‌ای هر ایستگاه		مدلسازی بر اساس داده‌های ایستگاه قبلی	
Station 1, 2		Station 2	
M1	$Q_w(t) = f[Q_w(t-1)]$	M4	$Q_{w2}(t) = f[Q_{w1}(t)]$
M2	$Q_w(t) = f[Q_w(t-1), Q_w(t-2)]$	M5	$Q_{w2}(t) = f[Q_{w1}(t), Q_{w1}(t-1)]$
M3	$Q_w(t) = f[Q_w(t-1), Q_w(t-2), Q_w(t-3)]$	M6	$Q_{w2}(t) = f[Q_{w1}(t), Q_{w1}(t-1), Q_{w1}(t-2)]$



شکل ۳- نمودار سری زمانی دبی جریان در ایستگاه‌های منتخب رودخانه Houstonic



شکل ۴- پارامترهای آماری مدل SVM و GPR با توابع کرنل متفاوت در ایستگاه اول رودخانه Houstonic

مدل، برای هر چهار ایستگاه در دو رودخانه منتخب مشاهده می‌شود که مدل M3 با پارامترهای ورودی $Q_w(t-1)$, $Q_w(t-2)$, $Q_w(t-3)$ بیشترین کارایی را داراست. این مدل بیشترین ضریب همبستگی و تبیین و کمترین مقدار خطا را داشته و مدل برتر است. مشاهده می‌شود که در حالت استفاده از پارامترهای دبی جریان یک روز قبل

نتایج به‌دست‌آمده برای مدل‌های مبتنی بر کرنل

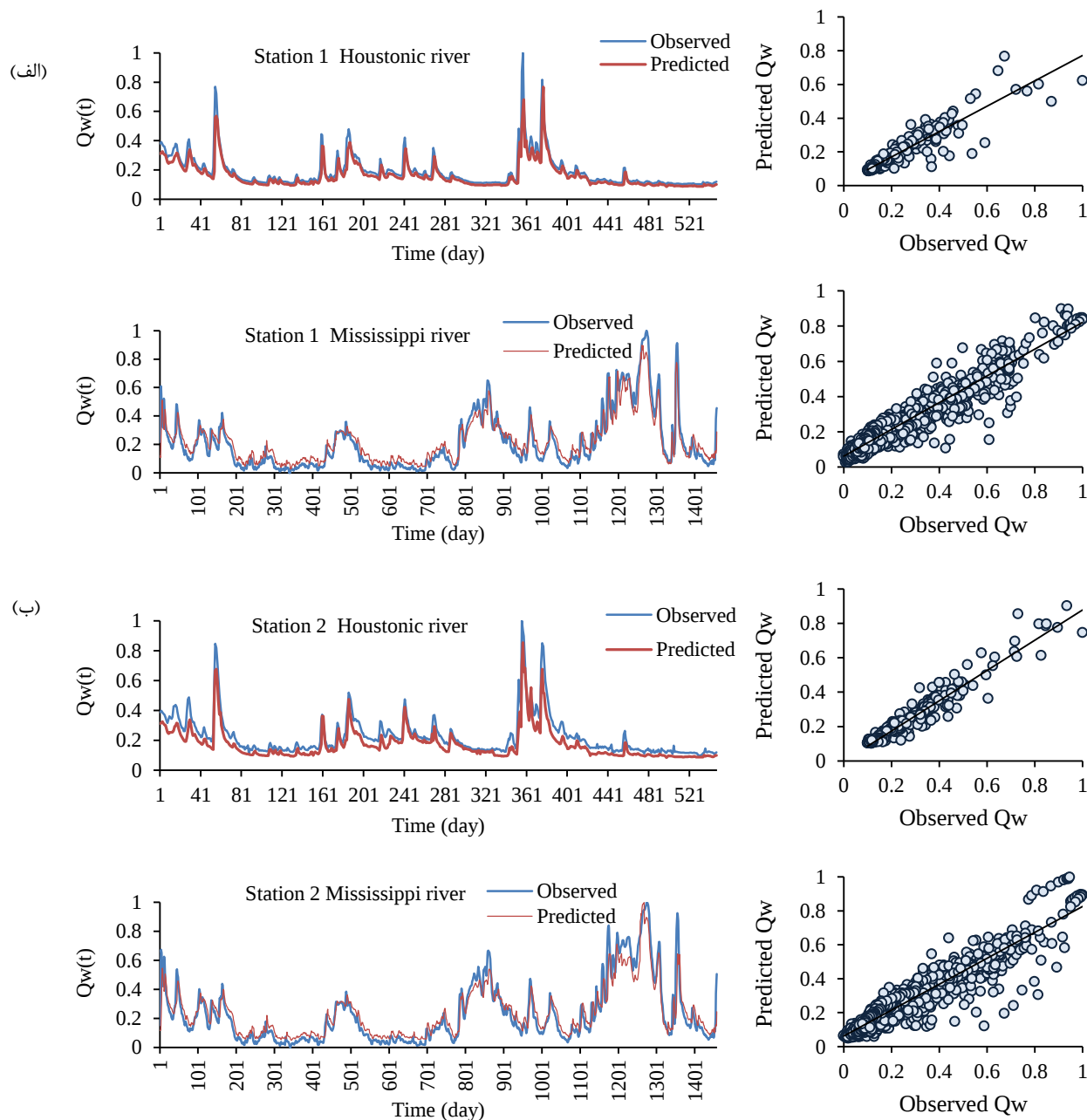
جهت پیش‌بینی دبی جریان بر اساس مقدار دبی جریان مربوط به روزهای قبل، چند مدل تعریف گردید و کارایی روش‌های مبتنی بر کرنل مورد ارزیابی قرار گرفت و نتایج حاصله به صورت جدول ۲ و شکل ۵ ارائه شد. با توجه به نتایج به‌دست آمده از سه معیار ارزیابی

به‌عنوان تنها پارامتر ورودی (مدل M1) نتایج دقیقی ارائه نشده است و مدل‌های ترکیبی دو یا سه پارامتری دقیق‌تر می‌باشند. نتایج فوق با نتایج پژوهش ناوه و همکاران (۱۳۹۱) و خلیلی و همکاران (۲۰۱۳) مطابقت دارد. بین دو روش هوشمند به‌کار رفته روش رگرسیون فرایند گاوسی نتایج دقیق‌تری را ارائه داده است، اما در حالت کلی و با مقایسه نتایج مدل‌ها مشاهده می‌شود که روش‌های مبتنی بر کرنل از دقت مطلوبی جهت مدل‌سازی دبی جریان به‌ویژه در نقاط ماکزیمم و مینیمم برخوردار نمی‌باشد. نمودار مدل برتر برای سری داده‌های آموزش و آزمون برای هر دو ایستگاه با استفاده از روش GPR در شکل ۵ نشان داده شده است. همچنین جهت بررسی ارتباط بین

ایستگاه‌ها، مقدار دبی جریان ایستگاه دوم بر اساس داده‌های ایستگاه اول به‌دست آمد و نتایج در جدول ۲ ارائه شد. در واقع مقدار دبی جریان ایستگاه دوم برای حالتی که هیچ اطلاعاتی در مورد داده‌هایش در دسترس نیست بررسی گردید و از داده‌های ایستگاه اول به‌عنوان ورودی استفاده شد. نتایج پیش‌بینی نشان داد که در صورتی که تنها از داده‌های مربوط به همان روز ایستگاه قبلی (M4) استفاده شود نتایج مطلوبی به‌دست نخواهد آمد و زمانیکه از داده‌های دبی جریان مربوط به یک تا سه روز پیش به‌طور هم‌زمان به‌عنوان ورودی مدل $[Q_{w1}(t), Q_{w1}(t-1), Q_{w1}(t-2)]$ استفاده شود نتیجه‌ی دقیق‌تری به‌دست می‌آید.

جدول ۲- نتایج ارزیابی مدل‌های مبتنی بر کرنل برای ایستگاه‌های منتخب در دو رودخانه

Model	Method	Performance criteria											
		Train			Test			Train			Test		
		R	DC	RMSE	R	DC	RMSE	R	DC	RMSE	R	DC	RMSE
Houstonic river													
		Station 1						Station 2					
M1	GPR	۰.۸۰۵	۰.۶۲۳	۰.۰۴۳	۰.۶۴۸	۰.۵۹۳	۰.۰۴۷	۰.۸۶۹	۰.۶۳۷	۰.۰۵۱	۰.۷۰۰	۰.۶۰۷	۰.۰۵۶
	SVM	۰.۷۸۶	۰.۶۱۰	۰.۰۴۵	۰.۶۳۳	۰.۵۸۰	۰.۰۵۰	۰.۸۴۹	۰.۶۲۴	۰.۰۵۳	۰.۶۸۴	۰.۶۰۱	۰.۰۵۹
M2	GPR	۰.۷۳۷	۰.۶۹۰	۰.۰۳۸	۰.۷۱۸	۰.۶۵۷	۰.۰۴۲	۰.۷۹۶	۰.۷۰۱	۰.۰۴۴	۰.۷۷۵	۰.۶۷۲	۰.۰۴۶
	SVM	۰.۷۲۰	۰.۶۷۶	۰.۰۴۱	۰.۷۰۱	۰.۶۴۳	۰.۰۴۵	۰.۷۷۸	۰.۶۹۹	۰.۰۴۵	۰.۷۵۷	۰.۶۵۸	۰.۰۴۹
M3	GPR	۰.۸۴۵	۰.۷۷۳	۰.۰۲۷	۰.۷۳۱	۰.۷۳۷	۰.۰۳۰	۰.۹۱۳	۰.۷۸۳	۰.۰۳۳	۰.۷۸۹	۰.۷۵۴	۰.۰۳۶
	SVM	۰.۸۲۶	۰.۷۵۷	۰.۰۲۹	۰.۷۱۴	۰.۷۲۲	۰.۰۳۲	۰.۸۹۲	۰.۷۷۴	۰.۰۳۶	۰.۷۷۱	۰.۷۳۹	۰.۰۳۸
		Station 2											
M4	GPR	۰.۷۴۱	۰.۶۳۱	۰.۰۵۰	۰.۷۱۱	۰.۶۱۱	۰.۰۵۴						
	SVM	۰.۷۳۸	۰.۶۲۸	۰.۰۵۳	۰.۷۰۴	۰.۶۰۹	۰.۰۵۶						
M5	GPR	۰.۷۴۸	۰.۶۹۴	۰.۰۴۶	۰.۷۳۵	۰.۶۷۰	۰.۰۴۹						
	SVM	۰.۷۴۱	۰.۶۹۱	۰.۰۴۷	۰.۷۳۳	۰.۶۶۹	۰.۰۵۱						
M6	GPR	۰.۷۶۵	۰.۷۱۶	۰.۰۳۹	۰.۷۵۵	۰.۷۰۸	۰.۰۴۲						
	SVM	۰.۷۳۸	۰.۷۱۱	۰.۰۴۱	۰.۷۳۴	۰.۷۰۵	۰.۰۴۴						
Mississippi river													
		Station 1						Station 2					
M1	GPR	۰.۸۸۶	۰.۶۸۵	۰.۰۴۷	۰.۶۵۸	۰.۶۵۲	۰.۰۵۲	۰.۸۸۲	۰.۷۰۱	۰.۰۵۶	۰.۷۱۱	۰.۶۶۸	۰.۰۶۲
	SVM	۰.۸۶۵	۰.۶۷۱	۰.۰۵۰	۰.۶۴۲	۰.۶۳۸	۰.۰۵۵	۰.۸۶۲	۰.۶۸۶	۰.۰۵۸	۰.۶۹۴	۰.۶۶۱	۰.۰۶۵
M2	GPR	۰.۸۱۱	۰.۷۵۹	۰.۰۴۲	۰.۷۲۹	۰.۷۲۳	۰.۰۴۶	۰.۸۰۸	۰.۷۷۱	۰.۰۴۸	۰.۷۸۷	۰.۷۳۹	۰.۰۵۱
	SVM	۰.۷۹۲	۰.۷۴۴	۰.۰۴۵	۰.۷۱۲	۰.۷۰۷	۰.۰۵۰	۰.۷۹۰	۰.۷۶۹	۰.۰۵۰	۰.۷۶۸	۰.۷۲۴	۰.۰۵۴
M3	GPR	۰.۹۳۰	۰.۸۵۰	۰.۰۳۰	۰.۷۴۲	۰.۸۰۵	۰.۰۳۳	۰.۹۲۷	۰.۸۶۱	۰.۰۳۶	۰.۸۰۱	۰.۸۰۹	۰.۰۴۰
	SVM	۰.۹۰۹	۰.۸۳۳	۰.۰۳۲	۰.۷۲۵	۰.۷۹۴	۰.۰۳۵	۰.۹۰۵	۰.۸۵۱	۰.۰۴۰	۰.۷۸۳	۰.۷۹۹	۰.۰۴۲
		Station 2											
M4	GPR	۰.۸۱۵	۰.۶۹۴	۰.۰۵۵	۰.۷۲۲	۰.۶۷۲	۰.۰۵۹						
	SVM	۰.۸۱۲	۰.۶۹۱	۰.۰۵۸	۰.۷۱۵	۰.۶۷۰	۰.۰۶۲						
M5	GPR	۰.۸۲۳	۰.۷۶۳	۰.۰۵۱	۰.۷۴۶	۰.۷۳۷	۰.۰۵۴						
	SVM	۰.۸۱۵	۰.۷۶۰	۰.۰۵۲	۰.۷۴۴	۰.۷۳۶	۰.۰۵۶						
M6	GPR	۰.۸۴۲	۰.۷۸۸	۰.۰۴۳	۰.۷۶۶	۰.۷۷۹	۰.۰۴۶						
	SVM	۰.۸۱۲	۰.۷۸۲	۰.۰۴۵	۰.۷۴۵	۰.۷۷۶	۰.۰۴۸						



شکل ۵- نمودار مقادیر مشاهداتی و محاسباتی مدل GPR برای ایستگاه‌های دو رودخانه، (الف) مدل سازی با استفاده از داده‌های خود ایستگاه و (ب) بررسی ارتباط ایستگاه‌ها

محاسبه شده و زیر سری‌های با انرژی بالا به عنوان ورودی در مدل-های مبتنی بر کرنل استفاده شدند. اولین قدم برای تجزیه سیگنال با تبدیل موجک، انتخاب موجک مادر است. در این تحقیق از موجک مادر db4 با دو سطح تجزیه استفاده شد، لذا دو زیر سری جزئی و یک زیر سری تقریب حاصل شد. اساس روش EEMD نیز تجزیه سیگنال به IMFهای مختلف و یک سیگنال باقیمانده است که در نهایت مجموعه‌ی این‌ها، همان سیگنال اصلی خواهد بود. روش تشکیل IMFها براساس تفريق تابع پایه از سیگنال اصلی است. این

نتایج به دست آمده برای مدل تلفیقی DWT-EEMD-GPR-SVM

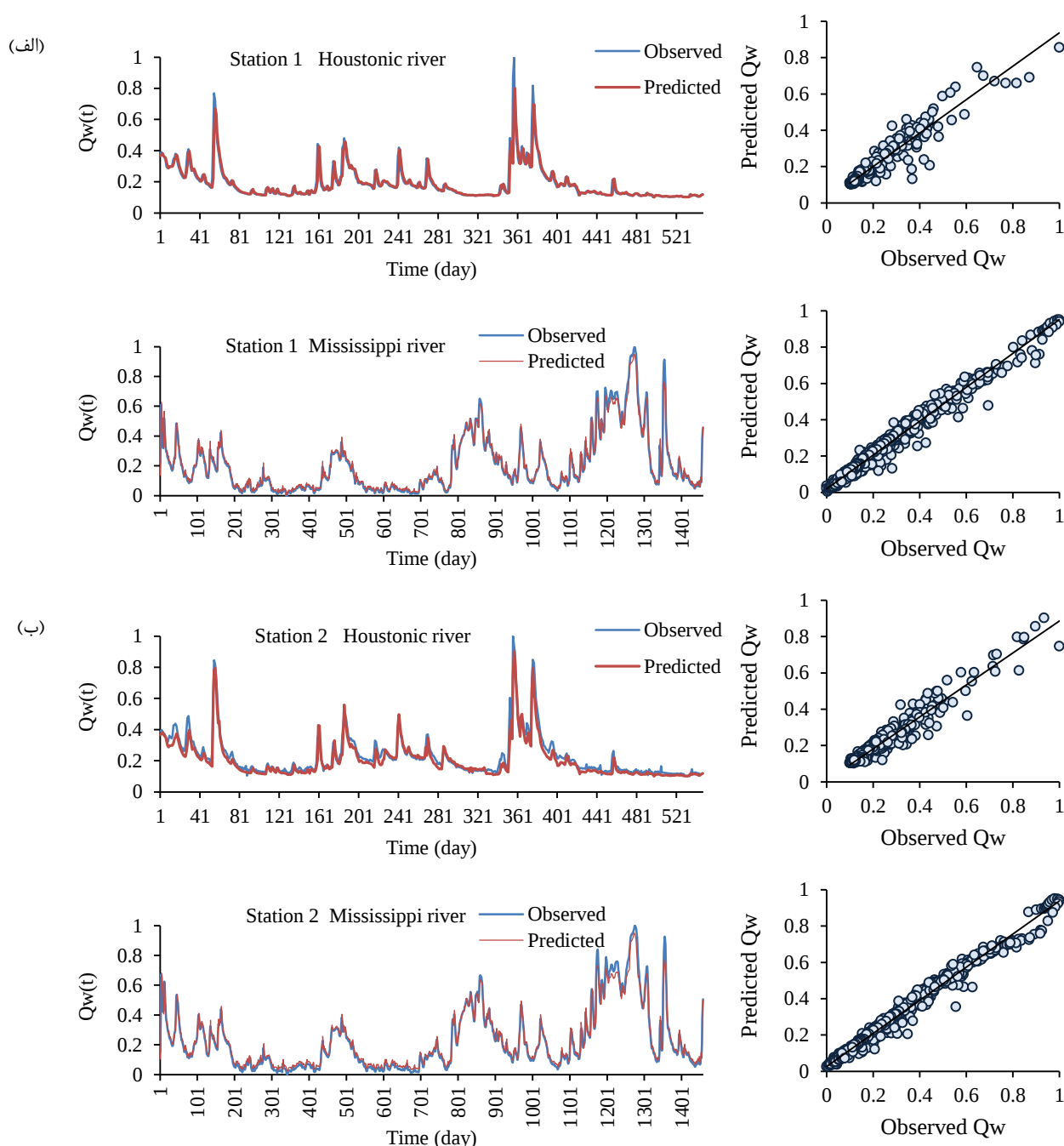
در این قسمت، جهت افزایش دقت پیش‌بینی مدل‌های SVM و GPR از روش‌های پیش‌پردازش سری زمانی شامل تجزیه‌ی یکپارچه مد تجزیه کامل و تبدیل موجک گسسته استفاده شد. بدین منظور، ابتدا سری‌های زمانی توسط روش تبدیل موجک تجزیه گردید. سپس با استفاده از روش EEMD زیر سری جزئی حاصل از تجزیه DWT با فرکانس بالا دوباره تجزیه گردید. سپس انرژی تمامی زیر سری‌ها

روز قبل به‌عنوان تنها پارامتر ورودی نیز دقت نتایج تا حدودی افزایش یافته است. با این حال، مدل‌های ترکیبی دو یا سه پارامتری دقیق‌تر می‌باشند. لازم به ذکر است که در استفاده از مدل‌های مختلف، عملکرد آن‌ها در تخمین مقادیر حداقل و حداکثر سری زمانی باید در نظر گرفته شود. مطابق با نتایج، روش‌های مبتنی بر کرنل تلفیقی از دقت مطلوبی در مدل‌سازی نقاط ماکزیمم و مینییمم دبی جریان برخوردارند. همچنین در حالت تخمین دبی جریان ایستگاه دوم بر اساس داده‌های ایستگاه اول، نتایج پیش‌بینی نشان می‌دهد که مدل M6 با پارامترهای $Qw1(t)$, $Qw1(t-1)$, $Qw1(t-2)$ مدل برتر است. در حالت کلی مشاهده می‌شود که تجزیه سری‌های زمانی دقت مدل‌ها را تقریباً بین ۱۵ تا ۲۰ درصد افزایش داده است.

پروژه تا وقتی ادامه می‌یابد که تقریباً سیگنال باقیمانده ثابت شود. نتایج حاصل از تحلیل مدل‌ها در جدول ۳ و شکل ۶ ارائه شده است. با مقایسه نتایج جداول ۲ و ۳ ملاحظه می‌شود که تجزیه سری‌های زمانی با روش DWT و تجزیه بیشتر آن‌ها با روش EEMD تا حدود زیادی باعث بهبود نتایج شده است که مطابق با تحقیق آداموفسکی و سان (۲۰۱۰) می‌باشد که نشان دادند استفاده از روش‌های تلفیقی و تجزیه سری‌های زمانی تا حدودی خطای پیش‌بینی را کاهش می‌دهد. با توجه به نتایج به‌دست آمده از سه معیار ارزیابی مشاهده می‌شود که در این حالت نیز در ایستگاه اول مدل M3 با پارامترهای ورودی $Qw(t-1)$, $Qw(t-2)$, $Qw(t-3)$ بیشترین کارایی را دارا است. ملاحظه می‌شود که در حالت استفاده از پارامترهای دبی جریان یک

جدول ۳- نتایج ارزیابی مدل‌های مبتنی بر کرنل با پیش‌پردازش داده‌ها برای ایستگاه‌های منتخب در دو رودخانه

Model	DWT-CEEMD-	Performance criteria											
		Train			Test			Train			Test		
		R	DC	RMSE	R	DC	RMSE	R	DC	RMSE	R	DC	RMSE
Houstonic river													
		Station 1						Station 2					
M1	GPR	۰.۸۳۳	۰.۷۸۰	۰.۰۳۲	۰.۸۱۱	۰.۷۴۲	۰.۰۳۶	۰.۸۴۰	۰.۷۷۱	۰.۰۳۶	۰.۸۳۳	۰.۷۳۹	۰.۰۳۷
	SVM	۰.۸۱۴	۰.۷۶۴	۰.۰۳۵	۰.۷۹۲	۰.۷۲۷	۰.۰۳۸	۰.۸۲۱	۰.۷۶۹	۰.۰۳۷	۰.۸۱۴	۰.۷۲۴	۰.۰۴۰
M2	GPR	۰.۹۳۰	۰.۸۵۰	۰.۰۲۳	۰.۸۱۹	۰.۸۲۵	۰.۰۲۶	۰.۹۶۳	۰.۸۶۱	۰.۰۲۷	۰.۸۴۸	۰.۸۲۹	۰.۰۲۹
	SVM	۰.۹۰۹	۰.۸۳۳	۰.۰۲۵	۰.۸۰۰	۰.۸۰۹	۰.۰۲۷	۰.۹۴۱	۰.۸۵۱	۰.۰۳۰	۰.۸۲۹	۰.۸۱۳	۰.۰۳۱
M3	GPR	۰.۹۶۴	۰.۸۵۷	۰.۰۱۸	۰.۹۱۴	۰.۸۴۱	۰.۰۲۴	۰.۹۷۱	۰.۸۷۹	۰.۰۲۵	۰.۹۶۴	۰.۸۵۵	۰.۰۲۸
	SVM	۰.۹۴۱	۰.۸۴۹	۰.۰۲۱	۰.۸۹۳	۰.۸۲۴	۰.۰۲۶	۰.۹۶۳	۰.۸۷۲	۰.۰۲۶	۰.۹۴۳	۰.۸۳۷	۰.۰۲۹
		Station 2											
M4	GPR	۰.۸۰۱	۰.۷۲۳	۰.۰۴۱	۰.۷۶۱	۰.۷۱۸	۰.۰۴۴						
	SVM	۰.۷۹۹	۰.۷۱۵	۰.۰۴۵	۰.۷۳۴	۰.۷۰۱	۰.۰۴۶						
M5	GPR	۰.۸۱۹	۰.۷۸۸	۰.۰۲۹	۰.۸۱۵	۰.۷۵۶	۰.۰۳۴						
	SVM	۰.۸۰۱	۰.۷۷۱	۰.۰۳۲	۰.۷۹۶	۰.۷۴۰	۰.۰۳۷						
M6	GPR	۰.۹۲۱	۰.۸۶۴	۰.۰۲۷	۰.۹۰۴	۰.۸۲۸	۰.۰۳۱						
	SVM	۰.۸۹۹	۰.۸۵۶	۰.۰۲۸	۰.۸۸۳	۰.۸۱۱	۰.۰۳۲						
Mississippi river													
		Station 1						Station 2					
M1	GPR	۰.۹۱۶	۰.۸۵۸	۰.۰۳۵	۰.۸۲۳	۰.۸۱۶	۰.۰۴۰	۰.۸۵۳	۰.۸۴۸	۰.۰۴۰	۰.۸۴۵	۰.۸۱۳	۰.۰۴۱
	SVM	۰.۸۹۵	۰.۸۴۰	۰.۰۳۹	۰.۸۰۴	۰.۸۰۰	۰.۰۴۲	۰.۸۳۳	۰.۸۴۶	۰.۰۴۱	۰.۸۲۶	۰.۷۹۶	۰.۰۴۴
M2	GPR	۰.۹۵۵	۰.۹۳۵	۰.۰۲۵	۰.۸۳۱	۰.۹۰۸	۰.۰۲۹	۰.۹۷۷	۰.۹۴۷	۰.۰۳۰	۰.۸۶۱	۰.۹۱۲	۰.۰۳۲
	SVM	۰.۹۲۱	۰.۹۱۶	۰.۰۲۸	۰.۸۱۲	۰.۸۹۰	۰.۰۳۰	۰.۹۵۵	۰.۹۳۶	۰.۰۳۳	۰.۸۴۱	۰.۸۹۴	۰.۰۳۴
M3	GPR	۰.۹۴۴	۰.۹۴۳	۰.۰۲۰	۰.۹۲۸	۰.۹۱۸	۰.۰۲۶	۰.۹۸۶	۰.۹۶۷	۰.۰۲۸	۰.۹۷۸	۰.۹۲۱	۰.۰۳۱
	SVM	۰.۹۷۴	۰.۹۳۴	۰.۰۲۳	۰.۹۰۶	۰.۹۰۶	۰.۰۲۹	۰.۹۷۷	۰.۹۵۹	۰.۰۲۹	۰.۹۵۷	۰.۹۱۱	۰.۰۳۲
		Station 2											
M4	GPR	۰.۸۸۱	۰.۷۹۵	۰.۰۴۵	۰.۷۷۲	۰.۷۹۰	۰.۰۴۸						
	SVM	۰.۸۷۹	۰.۷۸۷	۰.۰۵۰	۰.۷۴۵	۰.۷۷۱	۰.۰۵۱						
M5	GPR	۰.۹۰۱	۰.۸۶۷	۰.۰۳۲	۰.۸۲۷	۰.۸۳۲	۰.۰۳۷						
	SVM	۰.۸۸۱	۰.۸۴۸	۰.۰۳۵	۰.۸۰۸	۰.۸۱۴	۰.۰۴۱						
M6	GPR	۰.۹۸۸	۰.۹۱۸	۰.۰۳۰	۰.۹۱۸	۰.۸۹۵	۰.۰۳۳						
	SVM	۰.۹۸۹	۰.۹۰۵	۰.۰۳۱	۰.۸۹۶	۰.۸۹۲	۰.۰۳۵						



شکل ۶- (الف) نمودار مقادیر مشاهداتی و محاسباتی مدل تلفیقی GPR برای ایستگاه اول با استفاده از داده‌های خود ایستگاه و (ب) بررسی ارتباط ایستگاه‌ها

نتایج به‌دست‌آمده در حالت پس‌پردازش داده‌ها

جهت بررسی کارایی روش‌های پس‌پردازش، خروجی حاصل از روش‌های تلفیقی با هم ترکیب شده و نتایج پس از تحلیل به‌صورت جدول ۴ و شکل ۷ ارائه شد. جهت پس‌پردازش نتایج حاصله از روش شبکه عصبی غیرخطی استفاده شد. مطابق با این جدول ملاحظه شد که ترکیب نتایج مدل‌ها دقت پیش‌بینی را افزایش داده است. مشاهده

گردید که روش غیرخطی استفاده‌شده به ترتیب حدود ۴ تا ۷٪ در داده‌های آموزش و ۶ تا ۱۰٪ در داده‌های آزمون کارایی مدل را افزایش داده‌اند. در شکل ۸ نمودار مقایسه روش‌های به‌کار رفته در تحقیق با استفاده از معیار RMSE برای رودخانه Houstonian آورده شده است. ملاحظه می‌شود که در صورت استفاده از هر دو روش‌های پیش و پس‌پردازش نتایج دقیق‌تری در تخمین دبی جریان حاصل

می‌شود. در حالات کلی، در صورت استفاده از هر دو روش‌های تلفیقی پیش‌پردازش و پس‌پردازش دقت مدل‌ها تقریباً به میزان ۲۰ تا ۲۵ درصد نسبت به روش‌های منفرد مبتنی بر کرنل افزایش می‌یابد.

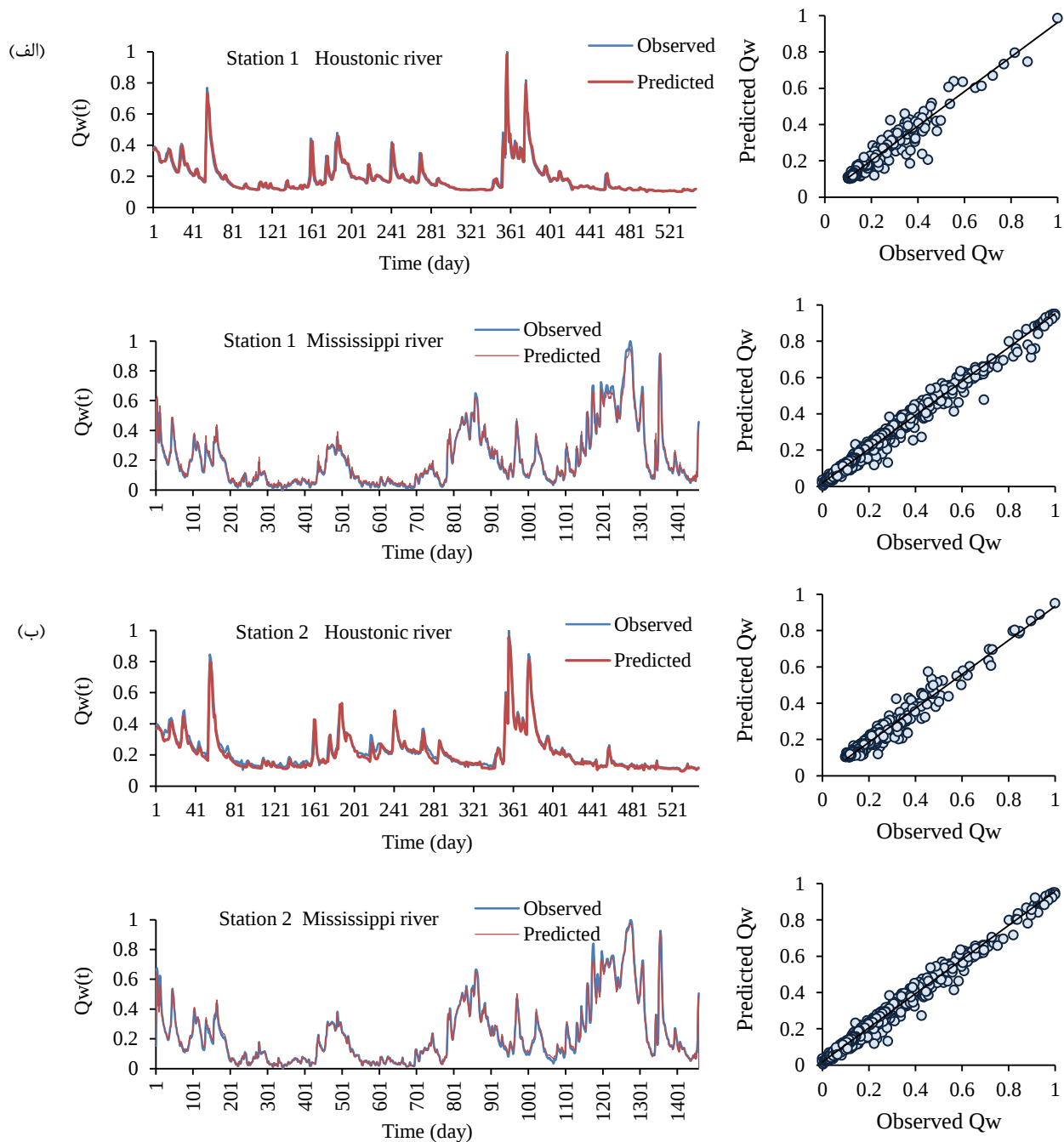
جدول ۴- نتایج ارزیابی مدل‌های تلفیقی با پیش و پس پردازش داده‌ها برای دو ایستگاه

Model	Performance criteria											
	Train			Test			Train			Test		
	R	DC	RMSE	R	DC	RMSE	R	DC	RMSE	R	DC	RMSE
Houstonic river												
Station 1						Station 2						
M1	۰.۸۸۷	۰.۸۱۹	۰.۰۲۷	۰.۸۶۴	۰.۷۸۰	۰.۰۲۹	۰.۸۷۴	۰.۸۱۳	۰.۰۳۱	۰.۸۷۰	۰.۸۰۲	۰.۰۳۳
M2	۰.۹۴۴	۰.۸۸۰	۰.۰۲۵	۰.۹۲۷	۰.۸۵۳	۰.۰۲۷	۰.۹۵۸	۰.۸۹۶	۰.۰۲۴	۰.۹۵۹	۰.۸۸۱	۰.۰۲۸
M3	۰.۹۶۶	۰.۸۹۹	۰.۰۱۶	۰.۹۵۰	۰.۸۹۳	۰.۰۲۱	۰.۹۷۴	۰.۹۰۵	۰.۰۲۰	۰.۹۷۱	۰.۸۹۹	۰.۰۲۳
Station 2												
M4	۰.۸۸۸	۰.۸۳۳	۰.۰۲۹	۰.۸۷۵	۰.۸۰۸	۰.۰۳۲						
M5	۰.۸۹۲	۰.۸۵۴	۰.۰۲۷	۰.۸۸۷	۰.۸۱۵	۰.۰۳۰						
M6	۰.۹۷۴	۰.۸۸۷	۰.۰۲۵	۰.۹۴۲	۰.۸۶۳	۰.۰۲۹						
Mississippi river												
Station 1						Station 2						
M1	۰.۹۷۶	۰.۹۱۱	۰.۰۳۰	۰.۹۰۵	۰.۸۹۴	۰.۰۳۲	۰.۹۴۹	۰.۹۲۴	۰.۰۲۴	۰.۹۱۸	۰.۹۱۷	۰.۰۲۸
M2	۰.۹۸۱	۰.۹۲۴	۰.۰۲۰	۰.۹۴۱	۰.۹۲۴	۰.۰۲۵	۰.۹۷۲	۰.۹۴۱	۰.۰۲۱	۰.۹۷۳	۰.۹۲۵	۰.۰۲۳
M3	۰.۹۸۴	۰.۹۴۸	۰.۰۱۷	۰.۹۶۴	۰.۹۴۴	۰.۰۱۸	۰.۹۸۹	۰.۹۵۰	۰.۰۱۶	۰.۹۸۶	۰.۹۴۴	۰.۰۱۹
Station 2												
M4	۰.۹۳۷	۰.۸۹۹	۰.۰۲۹	۰.۸۸۸	۰.۸۸۸	۰.۰۳۱						
M5	۰.۹۸۱	۰.۹۰۸	۰.۰۲۶	۰.۹۴۳	۰.۸۹۱	۰.۰۳۰						
M6	۰.۹۸۱	۰.۹۱۵	۰.۰۲۴	۰.۹۵۶	۰.۹۰۶	۰.۰۲۷						

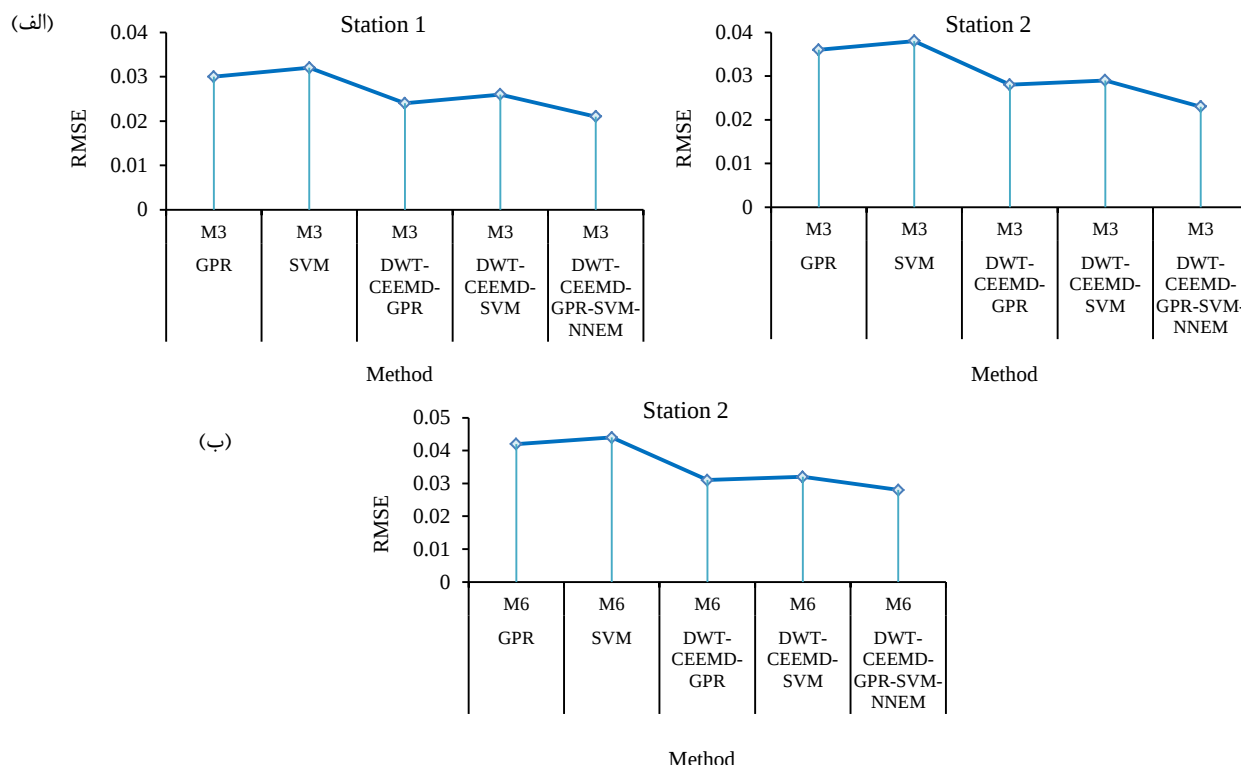
نتیجه‌گیری

به‌دست آمده جهت پیش‌بینی دبی روزانه براساس داده‌های خود ایستگاه مطابق با نتایج زمانی و همکاران (۱۳۹۳) بوده و مشخص گردید که در صورتی که تنها از داده‌های مربوط به روز قبل استفاده شود نتایج مطلوبی به‌دست نیامده و زمانی که از داده‌های دبی جریان مربوط به یک تا سه روز پیش به‌طور هم‌زمان به‌عنوان ورودی مدل استفاده شود نتیجه‌ی دقیق‌تری به‌دست خواهد آمد. مشاهده شد که تجزیه سری‌های زمانی دقت مدل‌ها را تقریباً بین ۱۵ تا ۲۰ درصد افزایش داده و سبب کاهش خطای مدل می‌شود. این نتیجه با نتایج تحقیق آداموفسکی و سان (۲۰۱۰) تطابق مناسبی دارد. در تحقیق نامبردگان در مرحله آزمون مقدار خطا (RMSE) از ۰/۱۳۱۵ به ۰/۱۲۵۸ کاهش یافت. مشاهده شد که در صورت استفاده از هر دو روش‌های پیش و پس‌پردازش نتایج دقیق‌تری در تخمین دبی جریان حاصل شده و روش پس‌پردازش غیرخطی استفاده‌شده حدود ۶ تا ۱۰٪ کارایی مدل‌های تلفیقی را برای داده‌های آموزش افزایش داده است. استفاده هم‌زمان از روش‌های پیش-پس‌پردازش نیز خطای مدل را تقریباً به میزان ۲۰ تا ۲۵ درصد کاهش داد. مطابق با نتایج، روش‌های مبتنی بر کرنل تلفیقی دقت مطلوبی را در مدل‌سازی نقاط ماکزیمم و مینیمم دبی جریان نشان دادند.

پیش‌بینی دبی جریان در رودخانه‌ها یکی از مهم‌ترین موضوعات مهندسی آب می‌باشد. عدم قطعیت و پیچیدگی موجود در سری زمانی دبی جریان از اعتبار بسیاری از مدل‌های پیش‌بینی می‌کاهد. لذا، در این تحقیق به پیش‌بینی دبی جریان با استفاده از روش‌های تلفیقی پیش-پس‌پردازش به‌عنوان روش جدید جهت بهبود دقت پیش‌بینی مدل‌ها پرداخته شد. مدل‌های ورودی متفاوتی با استفاده از داده‌های مربوط به دبی جریان و با در نظر گرفتن دو نوع مدل‌سازی با استفاده از داده‌های مربوط به خود ایستگاه و داده‌های مربوط به ایستگاه قبل تعریف گردید و مورد بررسی قرار گرفت. نتایج حاصله نشان داد که تجزیه سری‌های زمانی با هر روش DWT و تجزیه بیشتر زیرسری‌های با فرکانس بالا با روش EEMD تا حدود زیادی باعث بهبود نتایج می‌شود. مطابق با نتایج حاصله، برای مدل‌های تعریف‌شده براساس داده‌های خود ایستگاه، مدل با پارامترهای $Qw(t-1)$, $Qw(t-3)$, $Qw(t-2)$ و در حالت بررسی ارتباط بین ایستگاه‌ها مدل M6 با پارامترهای ورودی $Qw1(t)$, $Qw1(t-1)$, $Qw1(t-2)$ با داشتن بیشتر R و DC و کمترین مقادیر خطا مدل برتر به‌دست آمدند. نتیجه



شکل ۷- (الف) نمودار مقادیر مشاهداتی و محاسباتی مدل تلفیقی پیش-پس پردازش برای ایستگاه اول با استفاده از داده‌های خود ایستگاه و (ب) بررسی ارتباط ایستگاه‌ها



شکل ۸- نمودار مقایسه خطای حاصل از روش‌های به کار رفته در تحقیق، (الف) مدلسازی با استفاده از داده‌های هر ایستگاه و (ب) بررسی ارتباط ایستگاه‌ها در رودخانه Houstonic

خاک. ۲۶(۵): ۱۲۹۹-۱۳۰۷.

منابع

- نیک‌منش، م. ۱۳۹۴. پیش‌بینی آبدی متوسط ماهانه با استفاده از مدل تلفیقی شبکه عصبی مصنوعی و تبدیلات موجک (مطالعه موردی: رودخانه کر- ایستگاه پل خان). نشریه پژوهش‌های آب و خاک. ۲۲(۳): ۲۳۱-۲۳۹.
- Adamowski, K., Prokoph, A., and Adamowski, J. 2009. Development of a new method of wavelet aided trend detection and estimation. *Hydrology Process*, 23(18): 2686-2696.
- Adamowski, J., and Sun, K. 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390: 85-91.
- Amirat, Y., Benbouzid, M., Wang, T., Bacha, K., and Feld, G. 2018. EEMD-based notch filter for induction machine bearing faults detection. *Applied Acoustics*, 133: 202-209.
- ASCE, Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. *Artificial Neural Networks in hydrology. I: Preliminary concepts*. Hydrological Engineering, ASCE, 5(2): 115-123.
- Chou, C. 2011. Complexity analysis of rainfall and
- اکبری الاشقی، ح.، و بزرگ حداد، ا. ۱۳۹۱. مقایسه‌ی شبیه‌ها و روش‌های مختلف پیش‌بینی ماهانه‌ی جریان مبتنی بر هوش مصنوعی. مهندسی منابع آب. ۵(۱۳): ۳۷-۵۵.
- روشنگر، ک.، و قاسم‌پور، ر. ۱۳۹۷. بررسی عملکرد روش‌های کلاسیک و هوش مصنوعی در تخمین ضریب زبری در پیچان‌رودها. نشریه آبیاری و زهکشی ایران. ۱۲(۴): ۸۱۱-۸۲۲.
- زمانی، ر.، احمدی، ف.، و رادمنش، ف. ۱۳۹۳. مقایسه روش‌های برنامه‌ریزی بیان ژن، سری زمانی غیرخطی، خطی و شبکه عصبی مصنوعی در تخمین دبی روزانه (مطالعه موردی: رودخانه کارون). آب و خاک (علوم و صنایع کشاورزی). ۶: ۱۱۷۲-۱۱۸۲.
- معروف‌پور، س.، معروف‌پور، ع.، و ثانی‌خانی، ه. ۱۳۹۶. شبیه‌سازی تبخیر-تعرق مرجع ماهانه بدون داده‌های اقلیمی با استفاده از محاسبات نرم در غرب و شمال‌غرب ایران. مهندسی منابع آب. ۱(۳۴): ۳۷-۵۰.
- ناوه، ح.، خلیلی، خ.، عالمی، م.، و بهمنش، ج. ۱۳۹۱. پیش‌بینی جریان رودخانه توسط مدل سری زمانی غیرخطی Bilinear (مطالعه موردی: رودخانه‌های باراندوز چای و شهر چای). علوم آب و

- Science, 29(9): 987-996.
- Najah, A., El-Shafie, A., Karim, O.A., Jaafar, O., and El-Shafie, AH. 2011. An application of different artificial intelligences techniques for water quality prediction. *International Journal of Physical Sciences*, 6: 5298-5308.
- Siviapragasam, C., and Liong, S. 2001. Rainfall and runoff forecasting with SSA-SVM approach. *Hydroinformation*, 3(5): 141-152.
- Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. Data Mining and Knowledge Discovery, pp. 1-47.
- Wu, Z., and Huang, N. 2004. A study of the characteristics of white noise using the empirical mode decomposition method, *Proc RS Lond* 460A: 1597-1611.
- runoff time series based on sample entropy in different temporal scales. *Stochastic Environmental Research and Risk Assessment*, 6: 1401-1408.
- Khalili, K., Ahmadi, F., Abghari, H., and Besharat, S. 2013. Stationary and nonlinearity analysis of river flow at different time scales (Case study: Barandouz-Chai River). 9th International River Engineering Conference, Ahwaz, Iran.
- Kemp, P., Sear, D., Collins, A., Naden, P., and Jones, I. 2011. The impacts of fine sediment on riverine fish. *Hydrological Processes*, 25: 1800-1821.
- Lau, KM., and Weng, HY. 1995. Climate signal detection using wavelet transform, How to make time series sing. *Bulletin of the American Meteorological Society*, 76: 2391-2402.
- Makridakis, S., and Winkler, RL. 1983. Average of forecasts: some empirical results. *Management*

Temporal and Spatial Flow Discharge Prediction using Integrated Artificial Intelligence and Pre and Post-processing Time Series Methods

S.M. Saghebani^{1*}

Recived: Mar.01, 2020

Accepted: May.26, 2020

Abstract

Forecasting of river discharge is an important aspect of efficient water resources planning and management. In this study, time series pre and post-processing methods along with support vector machine (SVM) and Gaussian process regression (GPR) kernel based approaches were used to estimate flow discharge of two natural river in the United States with two consecutive hydrometric stations. The first river contained about 2 years of data and in the second river 4 years of daily discharge data was used. Different models were defined based on hydraulic characteristics and the capability of integrated pre and post-processing methods in two states of inter-station and between-stations was investigated. For data pre-processing, the Discrete Wavelet Transform (DWT) method was first used. Then, the high-frequency sub-series were selected and re-decomposed using the Ensemble Empirical Mode Decomposition (EEMD). Finally, sub-series with higher energy were imposed as inputs for kernel-based models. Non-linear neural average (NNA) model was also used for data post-processing. The obtained results from the defined models showed the high accuracy of the integrated methods used in the research in estimating flow discharge. At both stations, the error percentage was reduced by approximately 20 to 25% using the integrated pre-post-processing methods compared to the intelligent kernel based models. It was observed that in the case of river flow prediction based on the station's own data, the RMSE error value of the model decreased from approximately 0.3 to 0.26 and in the case of using the previous station data decreased from 0.44 to 0.33. Due to the high capability and accuracy of the pre-processing methods used in this study, similar studies are recommended in other rivers of the country.

Keywords: Experimental Mode Decomposition, Flow discharge, Post-processing, Wavelet

1- Department of Civil Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran
(*- Corresponding Author Email: smsaghebani@iau-ahar.ac.ir)