

مقاله پژوهشی

تخمین بار معلق رسوبی با استفاده از روش‌های هوشمند تلفیقی با در نظر گرفتن

عدم قطعیت مدل

سید مهدی ثاقبیان*

تاریخ دریافت: ۱۳۹۹/۱۲/۰۵

تاریخ پذیرش: ۱۴۰۰/۰۴/۲۸

چکیده

پیش‌بینی دبی رسوبی با دقت بالاتر، از مهم‌ترین مولفه‌های فرآیندهای هیدرولوژیکی در مدیریت منابع آب می‌باشد. به دلیل پیچیدگی پدیده انتقال رسوب و وجود چندین پارامتر موثر در تخمین آن، تعیین روابط حاکم مشکل می‌باشد. این تحقیق به منظور ارائه مدل بهینه برآورد بار رسوب معلق در دو ایستگاه هیدرومتری متوالی یک رودخانه طبیعی انجام گرفته است. در این راستا، جهت کاهش مقدار خطا در پیش‌بینی دبی رسوبی از روش‌های پیش‌پردازش سری زمانی به همراه روش‌های هوشمند مبتنی بر کرنل ماشین بردار پشتیبان (SVM) و رگرسیون فرآیند گاوسی (GPR) استفاده شده است. دو سناریو بررسی دبی رسوب معلق برای حالت تک ایستگاهی و ارتباط بین ایستگاهی در نظر رفته شد و مدل‌های متفاوتی بر اساس مشخصات هیدرولیکی و ذرات رسوبی تعریف گردید و مورد ارزیابی قرار گرفت. در استفاده از روش‌های پیش‌پردازش، ابتدا روش تبدیل موجک (WT) به کار رفت، سپس زیر سری‌های با فرکانس بالای بدست آمده از روش WT با روش تجزیه مد تجزیه (EMD) دوباره تجزیه گردیدند. در نهایت زیر سری‌های تأثیرگذار به عنوان ورودی مدل‌های مبتنی بر کرنل استفاده شدند. نتایج حاصل از تحلیل مدل‌های تعریف شده، دقت بالای روش‌های تلفیقی به کار رفته در تحقیق را در تخمین رسوب معلق به خوبی نشان داد. نتایج نشان داد که در حالت اول، مقدار معیار خطا برای مدل برتر به ترتیب از ۰,۰۳۵ و ۰,۰۳۷ برای روش‌های GPR و SVM به ۰,۲۸ و ۰,۲۹ برای مدل‌های تلفیقی کاهش یافت. به طور کلی، مدل‌های ترکیبی دقت مدل‌سازی را بین ۲۰ تا ۲۵ درصد افزایش دادند. جهت ارزیابی قابلیت اطمینان مدل برتر، از تحلیل عدم قطعیت مونت کارلو استفاده شد و نتایج نشان داد که مدل GPR دارای درجه عدم اطمینان مطلوبی در مدل‌سازی است.

واژه‌های کلیدی: پیش‌پردازش، تجزیه مد تجزیه، رگرسیون فرآیند گاوسی، رسوب معلق

مقدمه

طیف وسیعی از مسائل از قبیل طراحی مخازن سدها، انتقال رسوب و آلودگی در رودخانه‌ها، دریاچه‌ها و مصب‌ها، طراحی کانال‌های پایدار، ارزیابی زیست‌محیطی پروژه‌ها و مدیریت آبخیز به کار گرفته شده است و می‌تواند موجب کنترل بهتر عواقب زیان‌بار این پدیده و استفاده بهینه از تأسیسات آبی (مانند سد، کانال، ...) و جلوگیری از صرف هزینه‌های اضافی شود. مواد رسوبی در رودخانه‌ها به دو صورت کلی بار معلق و بار بستر می‌باشد که محاسبه مقدار هر یک به روش خاصی انجام می‌پذیرد. از جمله این روش‌ها مدل‌های تجربی و استفاده از تکنیک رگرسیون است که در این روش یک یا چند منحنی بر داده‌های رسوب برآزش داده می‌شود (۱۱). ساده‌ترین شکل معادلات

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 مواد رسوبی که توسط جریان حمل یا ترسیب می‌گردد، یکی از مسائل مهم و کاربردی در مدیریت منابع آب و از اطلاعات ضروری برای اجرای پروژه‌های آبی است (۸). تخمین صحیح رسوبات حمل شده به وسیله رودخانه در

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

*) - نویسنده مسئول:

DOI: 10.22067/JSW.2021.68665.1021

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

ایران

پارامترهای مختلف در پیش‌بینی آن، بهره‌گیری از روش‌های جدید که بتوانند قطعیت مدل را افزایش دهند ضروری به نظر می‌رسد. لذا، هدف از این تحقیق بررسی کارایی روش‌های تلفیقی پیش‌پردازش به‌عنوان روش جدید جهت بهبود دقت پیش‌بینی دبی رسوبی 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 (WT) و تجزیه مد تجربی (EMD) برای پیش‌پردازش مدل‌ها استفاده شده است. لذا، جهت بررسی کارایی روش‌های به‌کار رفته در تحقیق، مدل‌هایی بر اساس پارامترهای هیدرولیکی و دبی رسوبی برای دو ایستگاه هیدرومتری واقع در رودخانه هوستونیک تعریف شده و دو حالت تخمین بار رسوبی بر اساس داده‌های خود ایستگاه و ایستگاه قبلی مورد بررسی قرار گرفته است. همچنین میزان قابل اطمینان بودن نتایج مدل‌های به‌کار رفته با استفاده از روش عدم قطعیت مونت کارلو مورد بررسی قرار گرفت.

مواد و روش‌ها

منطقه مورد مطالعه

داده‌های مورد استفاده در این تحقیق دبی جریان و بار معلق روزانه‌ی دو ایستگاه هیدرومتری متوالی از رودخانه‌ی هوستونیک (Houstonic) واقع در ایالات متحده آمریکا در بازه زمانی تقریباً دو ساله می‌باشد. اولین ایستگاه نزدیک برینگتون بزرگ در ایالت ماساچوست و ایستگاه دوم در روستای فالز در ایالت کانکتیکات واقع شده و فاصله‌ی بین آن‌ها حدود ۵۰ کیلومتر است. مساحت حوضه به ترتیب برای ایستگاه‌ها ۲۸۲ مایل مربع و ۶۳۴ مایل مربع می‌باشد و مسیر جریان از ایستگاه اول به سمت ایستگاه دوم است. داده‌های مربوط به بار رسوب روزانه برای منطقه مطالعاتی در سایت USGS دسترس است که اطلاعات کافی در مورد دبی رسوب و دبی جریان را به دست می‌دهد. داده‌های مشاهده شده‌ی دو ایستگاه در حدود تقریباً دو سال از ۱ آوریل ۲۰۰۰ تا ۳۰ سپتامبر ۲۰۰۱ می‌باشد. این داده‌ها به‌صورت روزانه و به‌طور پیوسته برداشته شده و حدود ۵۴۹ داده می‌باشند. موقعیت ایستگاه‌ها در شکل نشان شده است.

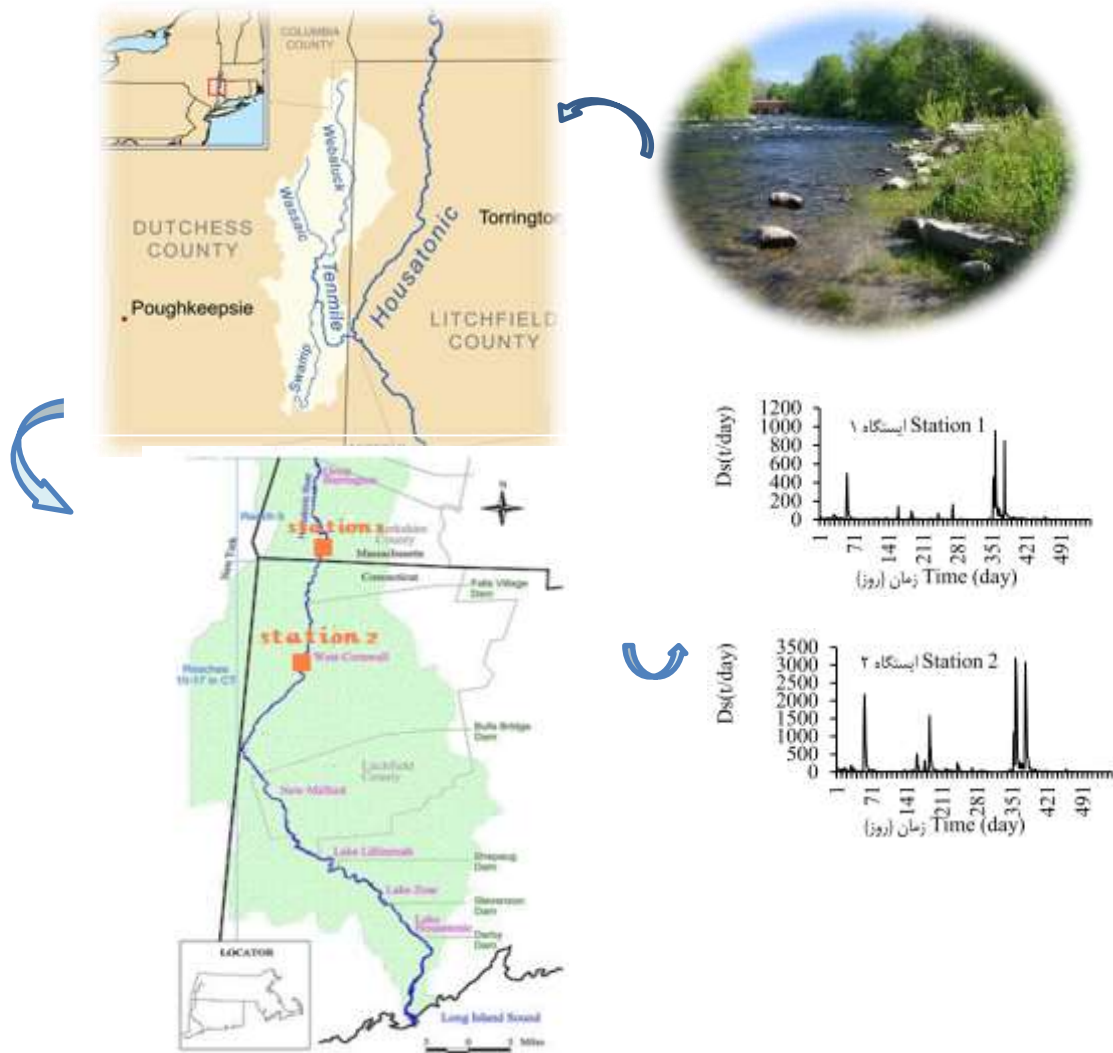
روش‌های هوشمند

ضعف عملکرد روش‌های استخراج ویژگی خطی برای داده‌های پیچیده که به‌صورت خطی تفکیک‌پذیر نمی‌باشند، سبب پیدایش روش‌های غیرخطی از جمله روش‌های مبتنی بر کرنل برای این منظور شده است.

رگرسیون که مستلزم تغییر شکل لگاریتمی داده‌های دبی رسوبی و دبی جریان است به شکل $Qs=aQ^b$ بوده که در آن Qs دبی رسوب و Q دبی جریان آب و a و b پارامترهای رگرسیونی می‌باشند (۱۳). با وجود اینکه مدل‌های مختلفی برای پیش‌بینی رسوب معلق 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 رسوب 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 شبیه‌سازی رفتار سیستم‌ها در مهندسی آب مورد استفاده قرار گرفته‌اند. این روش‌ها اکثراً الهام گرفته از پدیده‌های موجود در طبیعت هستند و به‌عنوان ابزارهای توانمند در حل مسائل پیچیده به‌شمار می‌آیند (۳). به‌عنوان نمونه سیواپراگسم و لیونگ (۱۵) از مدل ماشین بردار پشتیبان برای پیش‌بینی بارش- رواناب حوضه Tryggevaelde واقع در کشور هلند استفاده کردند. دو همکاران (۱۸) از مدل رگرسیون فرآیند گاوسی برای پیش‌بینی دبی رودخانه استفاده کردند. معروف‌پور و همکاران (۱۰) از روش‌های هوش مصنوعی برای شبیه‌سازی تبخیر-تعرق استفاده کردند. روشنگر و قاسم‌پور (۱۴) با استفاده از روش‌های هوشمند رگرسیون بردار پشتیبان و سیستم استنتاج عصبی- فازی انطباقی مقاومت جریان را در رودخانه‌های ماریچی بررسی کردند. حیدری و همکاران (۷) از روش هوشمند مبتنی بر کرنل برای مدل‌سازی مشخصات پرش هیدرولیکی استفاده کردند. عاشقی و همکاران (۴) از روش‌های هوش مصنوعی برای شبیه‌سازی بار بستر استفاده کردند.

بررسی منابع صورت گرفته، نشان از کاربرد روش هوشمند شبکه عصبی مصنوعی در پیش‌بینی سری‌های زمانی از جمله رسوب می‌باشد. این در حالی است که مطالعات چندانی در زمینه کاربرد روش‌های مبتنی بر کرنل مانند رگرسیون فرآیند گاوسی در تخمین رسوب معلق 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 (۱). 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 تجربی یکپارچه کامل است (۵). 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 پیچیده بودن پدیده رسوب و تأثیر



شکل ۱- موقعیت ایستگاه‌های رودخانه هوستونیک

Figure 1- The location of the Houstonic river stations

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

در روش‌های مبتنی بر کرنل ابتدا داده‌ها به فضایی با ابعاد بیشتر نگاشت یافته که در این فضا به‌طور 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 فرایند گاوسی (GPR) اشاره کرد. الگوریتم SVM، جزء الگوریتم‌های تشخیص الگو دسته‌بندی می‌شود که برای اولین بار توسط وینیک در سال ۱۹۹۵ معرفی شد (۱۶) 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

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 Mexican hat، Meyer، Haar، Coiflet، Daubechies، Symlet، Morle پیوسته به‌عنوان 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 به‌صورت 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 در آن نوعی توصیف زمان-مقیاس از سیگنال گسسته با استفاده از فیلترهای دیجیتال ارائه می‌گردد. تبدیل موجک، حاصل شباهت‌سنجی (کورولیشن) 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 می‌آورد؛ بردارهای تقریبی یا باقی‌مانده و بردارهای جزئیات (۹) این ضرایب حاصل از در هم پیچیدن سری زمانی اصلی با فیلتر پایین‌گذر 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 به‌صورت زیر است که در آن L سطح تجزیه، N تعداد نقاط داده در سری زمانی و p برابر با زمان محو‌شونده برای موجک مادر انتخاب شده است.

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

روش تجزیه مد تجربی EMD^۱

روش تجزیه مد تجربی روشی برای تجزیه سیگنال‌های گوناگون است که این عمل در فرآیندی تحت عنوان غربال کردن صورت

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 به‌طور 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 گزینه 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 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 به‌صورت 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 به‌عنوان موجک‌های مادر شناخته می‌شوند) نمونه‌های انتقال‌یافته 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 (۹).

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

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

در روابط بالا I_0 : مقدار اندازه‌گیری‌شده، I_p : مقدار پیش‌بینی‌شده، I_0 : متوسط مقدار اندازه‌گیری‌شده، I_p : مقدار پیش‌بینی‌شده، I_p : متوسط مقدار 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 مشاهده‌شده و x_n داده نرمال شده می‌باشد. محدوده‌ی نرمالیزه کردن داده‌ها در این تحقیق بین $0/1$ و 1 انتخاب گردید.

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

عدم قطعیت مونت کارلو

به صورت کلی، شبیه‌سازی مونت کارلو به هر تکنیکی اطلاق می‌شود که از طریق نمونه‌سازی آماری، پاسخ‌های تقریبی برای مسائل کمی فراهم می‌کند. شبیه‌سازی مونت کارلو بیشتر برای توصیف روشی جهت انتشار عدم قطعیت‌های موجود در ورودی مدل به عدم قطعیت‌ها در خروجی مدل، به کار می‌رود. بنابراین مونت کارلو، نوعی شبیه‌سازی است که به طور صریح و کمی، عدم قطعیت را نمایش می‌دهد. شبیه‌سازی مونت کارلو متکی به فرآیند نمایش صریح عدم قطعیت با تعیین ورودی‌ها به عنوان توزیع‌های احتمال است. اگر ورودی‌های توصیف‌کننده یک سیستم، غیرقطعی باشند، آنگاه پیش‌بینی عملکرد پیش رو به طور قطع غیرقطعی است. این بدان معنی است که نتیجه هر گونه تحلیل مبتنی بر ورودی‌های نمایش داده شده با توزیع‌های احتمال، خود یک توزیع احتمال است. در محاسبه عدم قطعیت به روش شبیه‌سازی مونت کارلو بر خلاف روش‌های تحلیلی، اساس کار شبیه‌سازی پدیده مورد نظر به تعداد ۱۰۰۰ بار می‌باشد. بدین منظور هر یک از پارامترهای ورودی (داده‌های تصادفی) به تعداد ۱۰۰۰ بار توسط روش‌های موجود تولید داده، شبیه‌سازی شده و با قرار دادن آن‌ها در مدل مورد نظر ۱۰۰۰ تکرار خروجی حاصل می‌گردد. سپس ضریب تغییرات پارامتر خروجی محاسبه و به‌عنوان معیار عدم قطعیت خروجی در نظر گرفته می‌شود. برای تولید هر پارامتر ابتدا بهترین توزیع آماری به هر یک از پارامترهای ورودی برآزش داده می‌شود. به هر بار شبیه‌سازی، تحقق سیستم گفته می‌شود. برای هر تحقق، تمام پارامترهای غیرقطعی نمونه‌برداری خواهند شد که یعنی یک مقدار تصادفی از توزیع اختصاصی مربوط به هر پارامتر، انتخاب می‌شود (۱۷). سپس، این سیستم با پارامترهای ورودی معین، در طول زمان، شبیه‌سازی خواهد شد. این شبیه‌سازی به گونه‌ای انجام می‌شود که در نهایت بتوانیم کارایی سیستم را محاسبه کنیم. این امر منتج به ایجاد تعداد زیادی

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^i) تجزیه می‌کند (۶).

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

که $r_n(x)$ همان مولفه باقی‌مانده، بعد از n تعداد IMF's و $c_i(x)$ است. یک داده ممکن است در یک زمان دارای چندین مد ذاتی باشد. این مدهای نوسانی، تابع‌های مد ذاتی (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 نقض می‌گردد (چون میانگین IMF‌ها حالتی نامشخص به خود می‌گیرد). جهت رفع این ایراد، وو و هوانگ (۲) روشی متفاوت با عنوان روش تجزیه مد تجربی یکپارچه $(EEMD^2)$ را پیشنهاد کردند. در روند تجزیه‌سازی 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.

معیارهای ارزیابی مدل‌ها

به‌منظور ارزیابی کارایی روش‌های استفاده‌شده در تحقیق، از سه پارامتر آماری استفاده گردید که عبارت‌اند 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 به‌صورت زیر می‌باشد:

$$R = \frac{\sum_{i=1}^N (I_0 - \bar{I}_0) \times (I_p - \bar{I}_p)}{\sqrt{\sum_{i=1}^N (I_0 - \bar{I}_0)^2 \times (I_p - \bar{I}_p)^2}} \quad (5)$$

$$DC = 1 - \frac{\sum_{i=1}^N (I_0 - I_p)^2}{\sum_{i=1}^N (I_0 - \bar{I}_p)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (I_0 - I_p)^2}{N}} \quad (7)$$

ایستگاه قبل استفاده شد. جدول ۱ مدل‌های تعریف شده در تحقیق را نشان می‌دهد. در این جدول D_s و D_w به ترتیب بیانگر دبی جریان و دبی بار معلق رسوبی می‌باشند و اندیس t تأخیر زمانی را بر حسب روز نشان می‌دهد. همچنین در شکل ۲ نمودار مقادیر دبی جریان و رسوب معلق در دو ایستگاه انتخابی نشان داده شده است. لازم به ذکر است که ۷۵ درصد از ابتدای داده‌ها جهت آموزش مدل‌ها و ۲۵ درصد باقیمانده جهت آزمون مدل‌ها مورد استفاده قرار گرفت.

جهت انتخاب تابع کرنل مناسب رگرسیون فرایند گاوسی و روش ماشین بردار پشتیبان مدل $M5$ با کرنل‌های مختلف مورد ارزیابی قرار گرفت. از معیار ارزیابی $RMSE$ جهت تعیین کارایی هر یک از کرنل‌ها استفاده گردید. نتایج به صورت شکل ۳ برای سری داده‌های آزمون ارائه گردیده است. مطابق با شکل ۳ مشاهده می‌گردد که در بین تمامی توابع کرنل‌ها، تابع کرنل شعاع مداری در پیش‌بینی دبی بار معلق 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.

نتیجه مستقل و جداگانه می‌شود، که هر کدام بیانگر یک «آینده» احتمالی برای سیستم هستند. به عبارت دیگر، یعنی یک مسیر احتمالی خواهیم داشت که سیستم، با گذشت زمان آن را دنبال خواهد کرد. نتایج تحقق‌های مستقل سیستم به شکل توزیع‌های احتمالی خروجی‌های ممکن خواهد بود. در نتیجه، خروجی‌ها به صورت مقادیر تک نیستند، بلکه توزیع احتمال هستند.

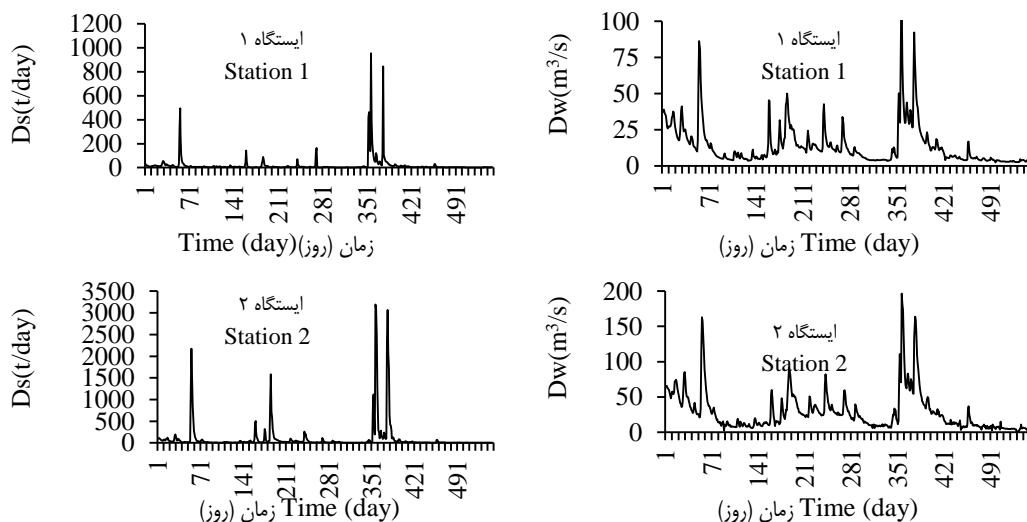
مدل‌سازی و نتایج و بحث

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 بار رسوب معلق انتخاب گردد. جهت تخمین بار معلق رسوبی در ایستگاه اول از داده‌های روزانه مربوط به همان ایستگاه و با در نظر گرفتن ترکیبی از داده‌های دبی جریان و دبی بار معلق رسوبی با تأخیرهای زمانی مختلف و در ایستگاه‌های دوم از داده‌های مربوط به آن ایستگاه و

جدول ۱- مدل‌های تعریف شده در تحقیق

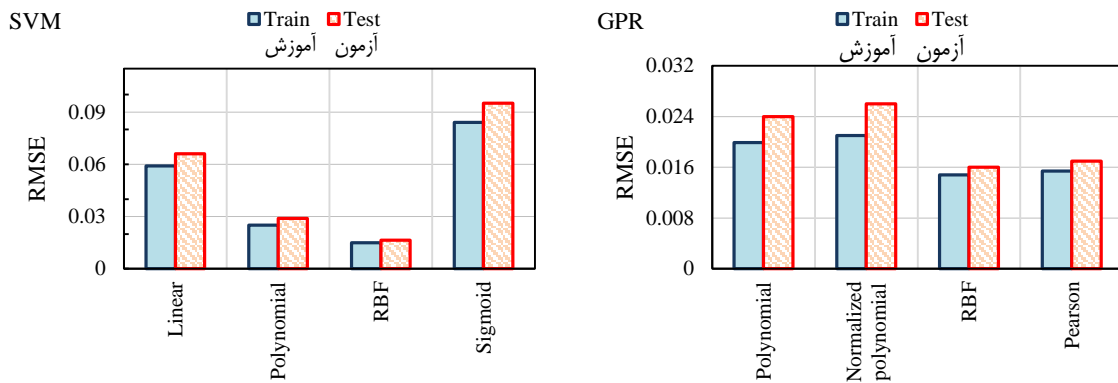
Table 1- Developed models in the study

مدل	ساختار مدل	مدل	ساختار مدل
Model	Model structur	Model	Model structur
M1	$D_s(t)=f [Dw_t]$	M6	$Ds_2(t)=f [(Dw_{t-1})1]$
M2	$D_s(t)=f [Dw_t, Dw_{t-1}]$	M7	$Ds_2(t)=f [(Dw_{t-1})1, (Ds_{t-1})1]$
M3	$D_s(t)=f [Ds_{t-1}]$	M8	$Ds_2(t)=f [(Dw_{t-1})1, (Ds_{t-1})1, (Ds_{t-2})1]$
M4	$D_s(t)=f [Dw_t, Ds_{t-1}]$		
M5	$D_s(t)=f [Dw_t, Dw_{t-1}, Ds_{t-1}]$		



شکل ۲- نمودار سری زمانی دبی جریان و دبی رسوب معلق ایستگاه‌های منتخب

Figure 2- The plots of the flow discharge and suspended sediment discharge for the selected stations



شکل ۳- پارامترهای آماری مدل SVM و GPR با توابع کرنل متفاوت
 Figure 3- The statistical parameters of the GPR and SVM models with different kernel types

خطا را داشته و مدل برتر است. مشاهده می‌شود که در حالت استفاده از پارامترهای دبی جریان و دبی رسوبی به‌عنوان تنها پارامتر ورودی (مدل‌های M1 و M3) نتایج دقیقی ارائه نشده است و مدل‌های ترکیبی دو یا سه پارامتری دقیق‌تر می‌باشند. بین دو روش هوشمند به کار 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 بار معلق به‌ویژه در نقاط ماکزیمم و مینیمم برخوردار نمی‌باشد.

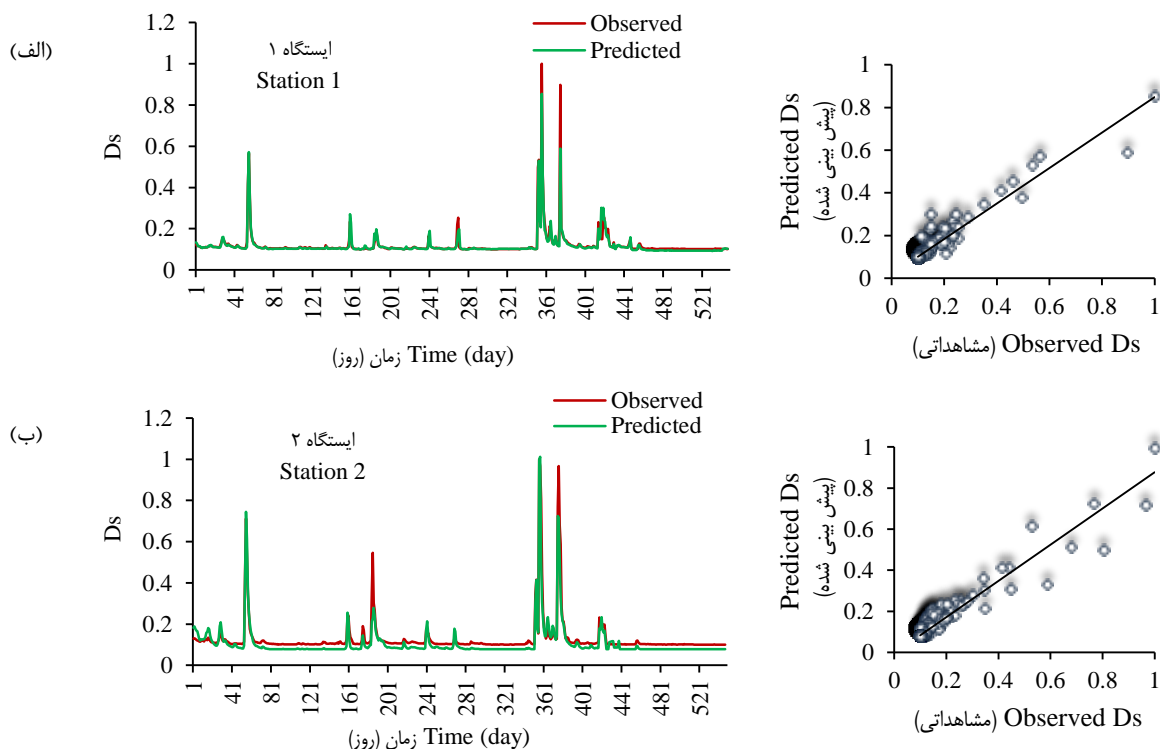
نتایج به‌دست‌آمده برای مدل‌های مبتنی بر کرنل

جهت پیش‌بینی رسوب معلق بر اساس دبی جریان و بار معلق رسوبی مربوط به روز قبل، چند مدل تعریف گردید و کارایی روش‌های مبتنی بر کرنل مورد ارزیابی قرار گرفت و نتایج حاصله به‌صورت جدول ۲ و شکل ۴ ارائه شد. با توجه به نتایج به‌دست آمده از سه معیار ارزیابی مدل، برای هر دو ایستگاه منتخب مشاهده می‌شود که مدل M5 با پارامترهای ورودی Dw_t , Dw_{t-1} , Ds_{t-1} بیشترین کارایی را داراست. این مدل بیشترین ضریب همبستگی و تبیین و کمترین مقدار

جدول ۲- نتایج ارزیابی مدل‌های مبتنی بر کرنل برای دو ایستگاه

Table 2- Results of evaluating kernel-based models for two stations

Model روش	Method مدل	Performance criteria معیارهای ارزیابی											
		Train			Test			Train			Test		
		R	DC	RMSE	R	DC	RMSE	R	DC	RMSE	R	DC	
		Station 1						Station 2					
M1	GPR	0.755	0.700	0.038	0.736	0.667	0.042	0.816	0.712	0.044	0.794	0.682	
	SVM	0.738	0.686	0.041	0.719	0.653	0.045	0.797	0.709	0.045	0.776	0.668	
M2	GPR	0.866	0.785	0.027	0.749	0.748	0.030	0.936	0.795	0.033	0.809	0.765	
	SVM	0.847	0.768	0.029	0.732	0.733	0.032	0.914	0.786	0.036	0.790	0.750	
M3	GPR	0.825	0.632	0.043	0.664	0.602	0.047	0.891	0.647	0.052	0.718	0.616	
	SVM	0.806	0.619	0.045	0.649	0.589	0.051	0.870	0.633	0.054	0.701	0.610	
M4	GPR	0.842	0.773	0.033	0.812	0.731	0.036	0.909	0.792	0.034	0.876	0.748	
	SVM	0.822	0.757	0.034	0.793	0.720	0.038	0.888	0.785	0.036	0.857	0.736	
M5	GPR	0.932	0.811	0.021	0.852	0.771	0.028	1.007	0.814	0.031	0.919	0.789	
	SVM	0.910	0.794	0.025	0.832	0.755	0.031	0.983	0.805	0.032	0.899	0.772	
		Station 2											
M6	GPR	0.762	0.716	0.034	0.731	0.708	0.041						
	SVM	0.745	0.701	0.038	0.741	0.694	0.043						
M7	GPR	0.841	0.744	0.035	0.800	0.738	0.039						
	SVM	0.844	0.743	0.036	0.802	0.733	0.040						
M8	GPR	0.886	0.783	0.032	0.842	0.777	0.036						
	SVM	0.865	0.777	0.034	0.822	0.760	0.038						



شکل ۴- نمودار مقادیر مشاهداتی و محاسباتی مدل GPR برای دو ایستگاه، (الف) مدلسازی با استفاده از داده‌های خود ایستگاه و (ب) بررسی ارتباط ایستگاه‌ها

Figure 4- The plots of the observed and predicted values of GPR model for two stations, (a): modeling based on the station's own data and (b): investigating the stations relationship

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 محاسبه شده و زیر سری‌های با انرژی بالا به‌عنوان ورودی در مدل‌های 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 سه سطح تجزیه استفاده شد، لذا دو سه 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 مجموعه‌ای این‌ها، همان سیگنال اصلی خواهد بود. روش تشکیل IMFها براساس تفریق تابع پایه از سیگنال اصلی است. این پروسه تا وقتی ادامه می‌یابد که تقریباً سیگنال باقیمانده ثابت شود. در شکل (۵-الف) زیر سری‌های حاصل از تجزیه بار رسوب معلق ایستگاه اول نشان داده شده است. مطابق با این شکل IMFهای دو و سه از زیرسری جزئی اول و IMFهای سه و چهار از زیرسری جزئی دوم، زیرسری‌های غالب حاصل از تجزیه با روش EEMD هستند. نتایج حاصل از تحلیل مدل‌ها در جدول ۳ و شکل (۵-ب و ج) ارائه شده است. با مقایسه نتایج جدول ۲ و ۳ ملاحظه می‌شود که تجزیه سری‌های زمانی با روش DWT و تجزیه بیشتر آن‌ها با روش

نمودار مدل برتر برای سری داده‌های آموزش و آزمون برای هر دو ایستگاه با استفاده از روش GPR در شکل ۴ نشان داده شده است. همچنین جهت بررسی ارتباط بین ایستگاه‌ها، مقدار دبی رسوبی ایستگاه دوم بر اساس داده‌های ایستگاه اول به‌دست آمد و نتایج در انتهای جدول ۲ ارائه شد. در واقع مقدار بار رسوبی ایستگاه دوم برای حالتی که هیچ اطلاعاتی در مورد داده‌هایش در دسترس نیست بررسی گردید و از داده‌های ایستگاه اول تحت عنوان مدل‌های M6، M7 و M8 به‌عنوان ورودی استفاده شد. نتایج پیش‌بینی نشان داد که در صورتی که از داده‌های دبی جریان و بار رسوبی به‌طور هم‌زمان به‌عنوان ورودی مدل استفاده شود نتیجه‌ی بهتری نسبت به حالتی که تنها از دبی رسوبی به‌عنوان ورودی استفاده شود، به‌دست می‌آید. در این حالت مدل با پارامترهای ورودی DS_{t-1} , DW_{t-1} , DS_{t-2} عملکرد بهتری نسبت به دو مدل دیگر دارد.

نتایج به‌دست‌آمده 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 تجربی کامل و تبدیل موجک گسسته استفاده شد. بدین منظور،

بررسی عدم قطعیت مدل

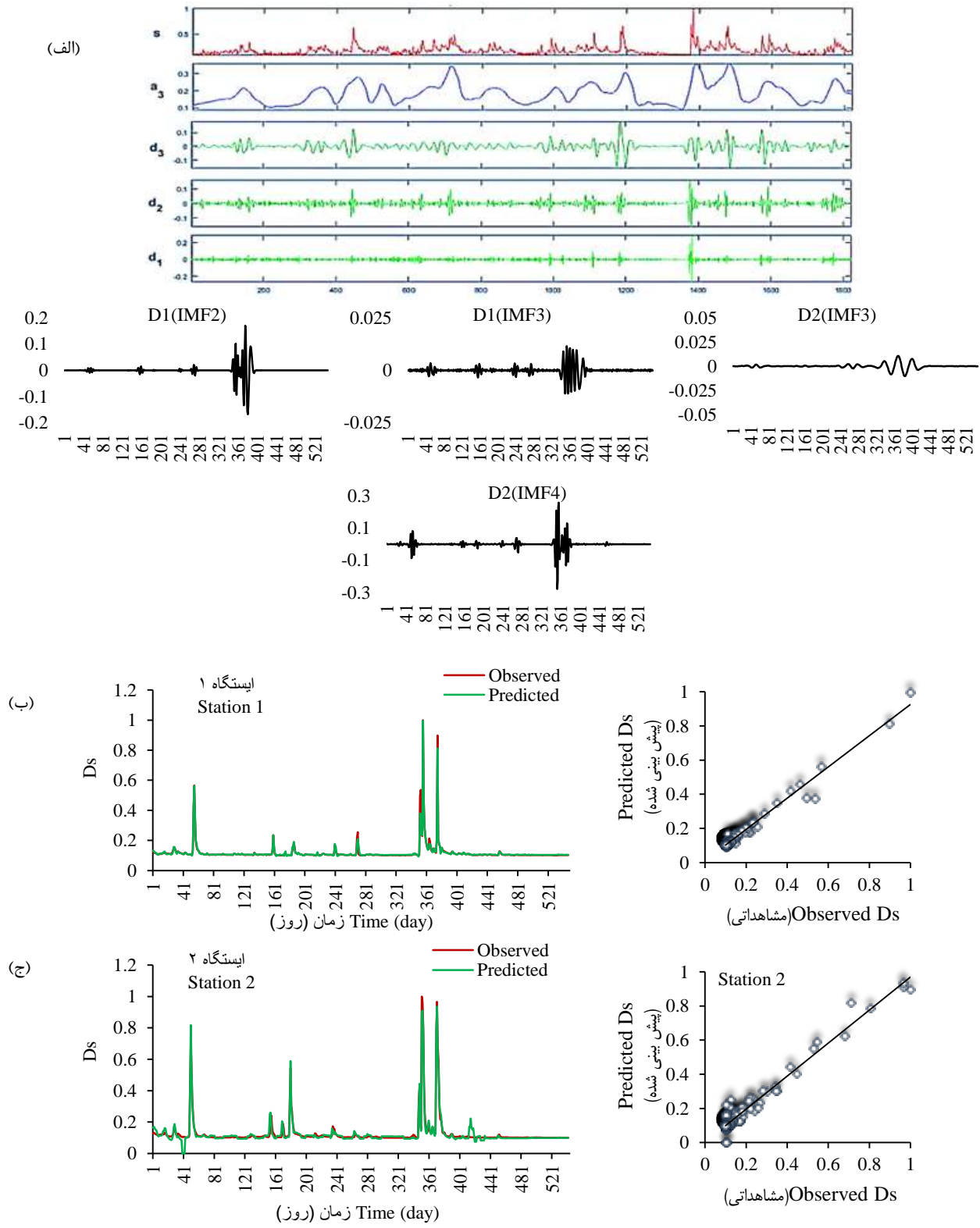
برای بررسی عدم قطعیت، از دو فاکتور استفاده می‌شود که اولی درصد خروجی‌های مورد مطالعه است که در محدوده 95% PPU قرار داشته باشد و مورد بعدی میانگین فاصله بین باندهای بالا (X_U) و پایین (X_L) عدم قطعیت است (۱۲). در این راستا، مدل در نظر گرفته شده باید بارها اجرا شود (۱۰۰۰ بار در این مطالعه)، و احتمال توزیع تجمعی مدل‌ها محاسبه شود. باندهای بالا و پایین به ترتیب ۲٫۵٪ و ۹۷٫۵٪ احتمال توزیع تجمعی در نظر گرفته می‌شوند. در سطح اطمینان مناسب باید دو نکته مهم در نظر گرفته شود. اولاً، باند 95% PPU شامل بیشترین داده‌های مشاهداتی باشد و دوماً میانگین فاصله بین باندهای بالا و پایین (d-Factor) از انحراف استاندارد داده‌های مشاهداتی کمتر باشد. می‌توان متوسط عرض باند فاصله اطمینان را به شرح زیر محاسبه کرد که در آن σ_x و \bar{d}_x به ترتیب انحراف معیار داده‌ها و عرض متوسط باند اطمینان و K تعداد داده‌های مشاهداتی است.

EEMD تا حدود زیادی باعث بهبود نتایج شده است. با توجه به نتایج به دست آمده از سه معیار ارزیابی مشاهده می‌شود که در این حالت نیز در ایستگاه اول مدل M5 با پارامترهای ورودی DW_t, DW_{t-1}, DS_{t-1} بیشترین کارایی را دارا است. ملاحظه می‌شود که در حالت استفاده از پارامترهای دبی جریان و دبی رسوبی به عنوان تنها پارامتر ورودی نیز دقت نتایج تا حدودی افزایش یافته است. باین‌حال، 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 مینیمم بار رسوبی معلق برخوردارند. همچنین در حالت تخمین دبی رسوبی ایستگاه دوم بر اساس داده‌های ایستگاه اول، نتایج پیش‌بینی نشان می‌دهد که مدل بر اساس هر دو داده‌های دبی جریان و دبی رسوبی مدل برتر بوده و با تجزیه داده‌ها کارایی مدل تا حدود زیادی افزایش یافته است. در حالت کلی، در صورت استفاده از روش‌های تلفیقی پیش‌پردازش دقت مدل‌ها تقریباً به میزان ۲۰ تا ۲۵ درصد نسبت به روش‌های منفرد هوشمند افزایش می‌یابد.

جدول ۳- نتایج ارزیابی مدل‌های مبتنی بر کرنل با پیش‌پردازش داده‌ها برای دو ایستگاه

Table 3- Results of evaluating kernel-based models with data preprocessing for two stations

Model مدل	DWT- EEMD-	Performance criteria معیارهای ارزیابی											
		Train			Test			Train			Test		
		R	DC	RM SE	R	DC	RMSE	R	DC	RMSE	R	DC	RMSE
		Station 1 ایستگاه ۱						Station 2 ایستگاه ۲					
M1	GPR	0.854	0.792	0.032	0.831	0.753	0.036	0.861	0.783	0.036	0.854	0.750	0.037
	SVM	0.834	0.775	0.035	0.812	0.738	0.038	0.842	0.781	0.037	0.834	0.735	0.040
M2	GPR	0.953	0.863	0.023	0.839	0.837	0.026	0.987	0.874	0.027	0.869	0.841	0.029
	SVM	0.932	0.845	0.025	0.820	0.821	0.027	0.965	0.864	0.030	0.850	0.825	0.031
M3	GPR	0.908	0.705	0.037	0.744	0.684	0.040	0.940	0.712	0.042	0.772	0.678	0.045
	SVM	0.887	0.698	0.038	0.727	0.681	0.043	0.918	0.696	0.043	0.753	0.671	0.048
M4	GPR	0.926	0.851	0.028	0.909	0.824	0.031	0.959	0.871	0.028	0.942	0.823	0.032
	SVM	0.904	0.833	0.029	0.889	0.806	0.032	0.937	0.863	0.030	0.921	0.810	0.033
M5	GPR	0.988	0.870	0.018	0.937	0.854	0.024	0.995	0.892	0.025	0.988	0.868	0.028
	SVM	0.965	0.862	0.021	0.915	0.836	0.026	0.987	0.885	0.026	0.967	0.850	0.029
		Station 2 ایستگاه ۲											
M6	GPR	0.860	0.823	0.029	0.856	0.786	0.033						
	SVM	0.842	0.804	0.031	0.837	0.770	0.036						
M7	GPR	0.919	0.856	0.024	0.904	0.839	0.028						
	SVM	0.920	0.855	0.026	0.905	0.822	0.031						
M8	GPR	0.968	0.901	0.021	0.951	0.861	0.024						
	SVM	0.944	0.893	0.023	0.929	0.844	0.027						



شکل ۵- (الف) زیرسری‌های حاصل از تجزیه بار رسوب معلق ایستگاه اول، (ب) نمودار مقادیر مشاهداتی و محاسباتی مدل تلفیقی GPR برای دو ایستگاه با استفاده از داده‌های خود ایستگاه و (ج) بررسی ارتباط ایستگاه‌ها

Figure 5- (a): The suspended sediment load decomposed subseries for the station 1, (b): plots of the observed and predicted values of integrated GPR model for two stations using the station's own data, and (c): investigating the stations relationship

جدول ۴- نتایج تحلیل عدم اطمینان مدل‌های GRP و WT-EEMD-GPR
Table 4- Uncertainty analysis results for the GPR and WT-EEMD-GPR methods.

روش Method	معیارهای ارزیابی Performance criteria					
	Station 1		Station 2		Station 2-1	
	95PPU	d-Factor	95PPU	d-Factor	95PPU	d-Factor
GPR	72.1%	0.341	74.8%	0.242	71.14%	0.258
WT-EEMD-GPR	83.2%	0.093	85.1%	0.108	83.4%	0.110

روش جدید جهت بهبود دقت پیش‌بینی بار رسوب معلق پرداخته شد. مدل‌های ورودی متفاوتی با استفاده از داده‌های مربوط به دبی جریان و دبی رسوب 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 WT و تجزیه بیشتر زیرسری‌های با فرکانس بالا با روش EEMD تا حدود زیادی باعث بهبود نتایج می‌شود. مطابق با نتایج حاصله، برای مدل‌های تعریف شده براساس داده‌های خود ایستگاه، مدل M5 با پارامترهای ورودی D_{wt} , D_{wt-1} و در حالت بررسی ارتباط بین ایستگاه‌ها مدل M8 با پارامترهای ورودی D_{st-2} , D_{wt-1} , D_{st-1} با داشتن بیشتر R و DC و کمترین مقادیر خطا مدل برتر به دست آمدند. مشاهده شد که در صورت استفاده از هر دو روش‌های تلفیقی نتایج دقیق‌تری در تخمین بار رسوب معلق حاصل شده و دقت مدل‌ها حدود ۲۰ تا ۲۵٪ افزایش می‌یابد. همچنین نتایج عدم قطعیت نشان داد که روش تلفیقی WT-EEMD-GPR دارای درجه اطمینان مناسبی در مدل‌سازی بار معلق رسوبی است.

$$d - \text{factore} = \frac{\overline{dx}}{\sigma x} \quad (9)$$

$$\text{Bracketed by 95PPU} = \frac{1}{k} \text{Cont}(j | X_L < X_{reg} < X_U) \quad (10)$$

نتایج بدست آمده برای تجزیه و تحلیل عدم قطعیت در جدول ۴ نشان داده شده است. براساس مقادیر بدست آمده برای شاخص‌های d-Factor و 95% PPU، مشاهده می‌شود که بیشتر مقادیر مشاهده شده و پیش‌بینی شده در محدوده باند 95% PPU قرار دارند. همچنین، مقدار شاخص d-Factor برای مجموعه داده‌های آموزش و آزمون کمتر از انحراف استاندارد داده‌های مشاهداتی است. بنابراین، براساس نتایج، می‌توان بیان نمود که مدل‌سازی بار رسوب معلق با استفاده از مدل تلفیقی GPR دارای درجه عدم قطعیت مجاز می‌باشد.

نتیجه‌گیری

پیش‌بینی بار معلق رسوبی به دلیل اثرات آن روی شاخص‌های کیفی آب، تقلیل گنجایش مخازن و تغییر در مورفولوژی رودخانه‌ها از اهمیت ویژه‌ای برخوردار است. لذا، در این تحقیق به پیش‌بینی بار رسوبی معلق با استفاده از روش‌های تلفیقی پیش‌پردازش به‌عنوان

منابع

- 1- 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.
- 2- 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.
- 3- 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.
- 4- Ashoghi R., Hosseini S.A., and Sanei M. 2021. Intelligent hybridized modeling approach to predict the bedload sediments in gravel-bed rivers. *Modeling Earth Systems and Environment* 8: 1-10.
- 5- Bahmani R., and Ouarda T.B. 2021. Groundwater level modeling with hybrid artificial intelligence techniques. *Journal of Hydrology* 595: 125659.
- 6- Chou C. 2011. Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales. *Stochastic Environmental Research and Risk Assessment* 6: 1401–1408.
- 7- Heydari M., Shabanlou S., and Sanahmadi B. 2020. Self-adaptive extreme learning machine-based prediction of roller length of hydraulic jump on rough bed. *ISH Journal of Hydraulic Engineering* 45: 1-11.
- 8- 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.
- 9- 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.
- 10- Maroufpour S., Maroufpour E., and Sanikhani H. 2018. Monthly Evapotranspiration Modeling in the Absence of

- Climatic Data Using the Soft Computing Methods in West and Northwest of Iran. *Water Resource Engineering* 10(34): 37-50. (In Persian with English abstract)
- 11- Najah A., El-Shafie A., Karim OA., 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.
 - 12- Noori R., Deng Z., Kiaghadi A., and Kachoosangi F.T. 2015 How reliable are ann, anfis, and svm techniques for predicting longitudinal dispersion coefficient in natural rivers?. *Hydraulic Engineering* 142(1): 04015039.
 - 13- Nourani V., Alizadeh F., and Roushangar K. 2016. Evaluation of a two-stage SVM and spatial statistics methods for modeling monthly river suspended sediment load. *Water Resources Management* 30(1): 393-407.
 - 14- Roushangar K., and Ghasempour R. 2019. Investigation of the Performance of Classical and Artificial Intelligence Approaches in Prediction of Roughness Coefficient in Meanders. *Iranian Journal of Irrigation and Drainage* 12(4): 811-822. (In Persian with English abstract)
 - 15- Siviapragasam C., and Liang S. 2001. Rainfall and runoff forecasting with SSA-SVM approach. *Hydroinformation* 3(5): 141-152.
 - 16- Vapnik V. 1995. The Nature of Statistical Learning Theory. *Data Mining and Knowledge Discovery*, pp. 1–47.
 - 17- 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.
 - 18- Zhou Y., Lu X., Huang Y., and Zhu Y. 2007. Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the upper Yangtze catchment. *China Geomorphology* 84: 111-125.



Estimation of Suspended Sediment Load Using Integrated Intelligent Methods with Considering Model Uncertainty

S.M. Saghebian^{1*}

Received: 23-02-2021

Accepted: 19-07-2021

Introduction: Sediment transportation and accurate estimation of its rate is a significant issue for river engineers and researchers. So far, various and complex relationships have been proposed to predict the amount of suspended sediment transport rate, such as velocity and critical shear stress based equations. However, the complex nature of sediment transport and lack of validated models make it difficult to model the suspended sediment concentration and suspended sediment discharge carried by rivers. Although the developed models led to promising results in sediment transport prediction, due to the importance of sediment transport and its impact on hydraulic structures it is necessary to use other methods with higher efficiency. On the other hand, in recent years, the Meta model approaches have been applied in investigating the hydraulic and hydrologic complex phenomena. Hybrid models involving signal decomposition have also been shown to be effective in improving the prediction accuracy of time series prediction methods, as indicated in. Complementary Ensemble Empirical Mode Decomposition analysis is one of the widely used signal decomposition methods for hydrological time series prediction. Decomposition of time series reduces the difficulty of forecasting, thereby improving forecasting accuracy.

In this study, due to the complexity of the sediment and erosion phenomenon and the effect of different parameters in estimating, time series pre-processing methods along with support vector machine (SVM) and Gaussian process regression (GPR) kernel based approaches were used to estimate suspended sediment load of a natural river at two consecutive hydrometric stations. For this purpose, different models were defined based on hydraulic and sediment particles characteristics. Moreover, the capability of integrated pre-processing and post-processing methods in two states of inter-station and between-stations was investigated. First, the Wavelet Transform (WT) method was used for data pre-processing then, the high-frequency sub-series were selected and re-decomposed using the Empirical Mode Decomposition (EMD). Finally, the most effective sub-series were imposed as inputs for kernel-based models. In addition, to assess the reliability of the superior model, Monte Carlo uncertainty analysis was used. The results showed that the GPR model had a desirable degree of uncertainty in modeling.

Materials and Methods: In this study, data of two stations of Housatonic River was used. The distance between stations was approximately 50 km. The first station is located near Great Brighton, Massachusetts, and the second station is in Connecticut. The basin area for the stations is 282 and 634 square miles, respectively. The flow path is from the first station to the second station. SVM and GPR models are based on the assumption that adjacent observations should convey information about each other. Gaussian processes are a way of specifying a prior directly over function space. This is a natural generalization of the Gaussian distribution whose mean and covariance are a vector and matrix, respectively. Due to prior knowledge about the data and functional dependencies, no validation process is required for generalization, and GP regression models are able to understand the predictive distribution corresponding to the test input. Wavelet Transform (WT) uses a flexible window function (mother wavelet) in signal processing. The flexible window function can be changed over time according to the signal shape and compactness. After using WT, the signal will decompose into two approximations (large-scale or low-frequency component) and detailed (small-scale component) components. EEMD was proposed to solve the mode mixing issue of empirical mode decomposition (EMD) which specifies the true IMF as the mean of an ensemble of trials. Each trial consists of the decomposition results of the signal plus a white noise of finite amplitude. EMD can be used to decompose any complex signal into finite intrinsic mode functions and a residue, resulting in subtasks with simpler frequency components and stronger correlations that are easier to analyze and forecast. Another important feature of empirical model of decomposition is that it can be used for noise reduction of noisy time series, which can be effective in improving the accuracy of model predictions. In the uncertainty analysis method, two elements are used to test the robustness and to analyze the models uncertainty. The first one is the percentage of the studied outputs which are in the range of 95PPU and the next one is the average distance between the upper (X_U) and lower (X_L) uncertainty bands. In this regard, the considered model should be run many

1- Department of Civil Engineering, Ahar Branch, Islamic Azad University Ahar, Iran

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

DOI: 10.22067/JSW.2021.68665.1021

times (1000 times in this study), and the empirical cumulative distribution probability of the models be calculated. The upper and lower bands are considered 2.5% and 97.5% probabilities of the cumulative distribution, respectively.

Results and Discussion: In order to evaluate and review the performance of the tested models and determine the accuracy of the selected models, three performance criteria named Correlation Coefficient (CC), Determination Coefficient (DC), and Root Mean Square Errors (RSME) were used. The obtained results indicated that the accuracy of the applied integrated models was higher than the single SVM and GPR models. The use of integrated methods decreased the error criteria between 20 to 25 %. The obtained results for the uncertainty analysis showed that in suspended sediment load modeling the observed and predicted values were within the 95 PPU band in most of the cases. Moreover, it was found that the amount of d-Factors for train and test datasets were smaller than the standard deviation of the observed data. Therefore, based on the results, it could be induced that the suspended sediment modeling via integrated WT-EEMD-GPR model led to an allowable degree of uncertainty.

Conclusion: Comparison of the developed models' accuracy revealed that integrated GPR and SVM models had higher performance compared with single GPR and SVM models in predicting the suspended sediment discharge. The use of these two methods approximately decreased the error criteria between 20 to 25 %. According to the results, for the models that were developed based on the station data, the model with the input parameters of Dw_t , Dw_{t-1} , and Ds_{t-1} and in the case of investigating the relationship between the stations, the model with the input parameters of Ds_{t-2} , Dw_{t-1} , and Ds_{t-1} were superior models. Also, based on the uncertainty analysis, the integrated GPR model had an allowable degree of uncertainty in suspended sediment modeling. However, it should be noted that the used methods are data sensitive models. Therefore, further studies using data ranges out of this study and field data should be carried out to determine the merits of the models to estimate suspended sediment load in the real conditions of flow.

Keywords: Experimental mode decomposition, Gaussian process regression, Pre-processing, Suspended sediment