

برآورد رسوب معلق با استفاده از مدل‌های ناپارامتریک M5 و رگرسیون تطبیقی چند متغیره اسپلاین (MARS)

(مطالعه موردی: رودخانه‌های تیره-ماربره لرستان)

بابک شاهی نژاد^{۱*}، سیده حدیث شاهرخی^۲، حسین یوسفی^۳

تاریخ ارسال: ۱۳۹۹/۰۷/۱۶

تاریخ پذیرش: ۱۳۹۹/۱۲/۲۳

مقاله پژوهشی

چکیده

در پژوهش حاضر از الگوریتم درختی M5 و رگرسیون تطبیقی چند متغیره اسپلاین MARS به‌عنوان روش‌های نوین جهت برآورد بار معلق رسوب در مقایسه با روش منحنی سنج رسوب استفاده شده است. اطلاعات مورد استفاده در این پژوهش شامل دبی جریان آب و دبی رسوب مربوط به چهار ایستگاه هیدرومتری تیره دورود و تیره مروک بر روی رودخانه تیره و همچنین ماربره دورود و ماربره دره تخت بر روی رودخانه ماربره در استان لرستان است. جهت ساخت و صحت سنجی مدل‌ها، دبی جریان با یک، دو و سه روز تأخیر و دبی جریان همان روز به همراه باران به‌عنوان پارامترهای ورودی و دبی بار معلق رسوب به‌عنوان پارامتر خروجی در نظر گرفته شد. جهت بررسی کارایی مدل‌ها و مقایسه نتایج آن‌ها با روش‌های متداول موجود از آنالیزهای آماری استفاده شد. در ایستگاه ماربره دورود، مقدار $RMSE$ و R^2 مربوط به مدل M5 به ترتیب ۰/۴۷ و ۰/۷۱ و برای مدل MARS به ترتیب ۰/۴۶ و ۰/۷۲ بوده درحالی‌که در روش منحنی سنج رسوب مقدار ۰/۵۶ و ۰/۶۴ است. عملکرد مدل‌های ارائه‌شده، نشان‌دهنده بهبود دقت و توانایی آن‌ها در تخمین بار معلق رسوب می‌باشد. نتایج حاصل نشان داد که معادلات ارائه‌شده توسط مدل‌های درختی M5 و MARS دقت بیشتری نسبت به منحنی سنج رسوب دارند. بر اساس نتایج به‌دست‌آمده مشاهده شد که دو روش M5 و MARS پاسخ‌های نزدیک به هم ارائه داده‌اند، اما در نهایت با توجه به ساختار ساده و مفهومی مدل M5 این روش به‌عنوان روش مناسب‌تر جهت برآورد بار معلق در محدوده مورد مطالعه انتخاب گردید. علاوه بر آن بررسی روابط به‌دست‌آمده از دو مدل M5 و MARS نشان داد که از میان پارامترهای ورودی، دبی جریان روز قبل و همان روز جهت برآورد بار معلق مورد استفاده قرار گرفته‌اند و مقادیر پیش‌بینی بیش از هر عاملی متأثر از این دو عامل بوده‌اند.

واژه‌های کلیدی: تیره-ماربره، رسوب معلق، مدل درختی M5، منحنی سنج رسوب، MARS

*۱- استادیار گروه مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان، خرم‌آباد، ایران، پست الکترونیکی: Shahinejad.b@lu.ac.ir

۲- دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان، خرم‌آباد، ایران، ۰۹۱۶۳۶۷۶۰۹۰ پست الکترونیکی: sh.hadis815@gmail.com

۳- دانشیار، گروه انرژی‌های نو و محیط‌زیست، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران، ایران، ۰۹۱۲۴۱۵۱۱۷۴ پست الکترونیکی: hosseinyousefi@ut.ac.ir

مقدمه

در سال‌های اخیر استفاده از الگوریتم‌های هوشمند در شاخه‌های مختلف مهندسی از جمله مهندسی رودخانه رو به افزایش است. الگوریتم‌های فراکاوشی توانایی تخمین پدیده‌ها و فرآیندهای پیچیده طبیعی را دارا بوده و اغلب دارای دقت بیشتری نسبت به روابط تجربی هستند. اعتماد شهیدی و محجوبی از مدل درختی M5 جهت شبیه‌سازی ارتفاع امواج رودخانه استفاده کردند. آن‌ها علاوه بر مدل درختی از شبکه عصبی نیز جهت مقایسه کارایی مدل درختی استفاده کردند و نشان دادند که مدل درختی با دقت بیشتری می‌تواند ارتفاع امواج را شبیه‌سازی کند (اعتماد شهیدی و محجوبی، ۲۰۰۹). در مطالعه‌ای که توسط آداموسکی و همکاران از مدل MARS جهت پیش‌بینی سیلاب در زیرحوضه‌های هیمالیا که اطلاعات محدودی از آن‌ها در دسترس بود، استفاده کردند. این پژوهشگران جهت مقایسه کارایی مدل MARS از شبکه عصبی نیز استفاده کردند که در نهایت دقت مدل تطبیقی جهت تخمین سیلاب در حوضه‌های مذکور در مقایسه با شبکه عصبی بیشتر بود (Adamowski et al, 2012). ظهیری (۱۳۹۴) از مدل درختی M5 و CART جهت تخمین عمق آب شستگی اطراف پایه پل استفاده کرد. نتایج این پژوهش، دقت بالاتر مدل M5 را در مقایسه با مدل درختی CART و سایر روش‌های متداول نشان داد. شاهرخی و همکاران (۱۳۹۵) در مطالعه‌ای، کاربرد الگوریتم درختی M5 را برای برآورد رسوب معلق رودخانه جراحی واقع در استان خوزستان را بررسی کردند و در نهایت این مدل را با توجه به دقت بالا و ساختار ساده‌ی آن، مناسب معرفی کردند.

بیلماز و همکاران در مطالعه‌ای به تخمین بار معلق رسوب رودخانه کوره در ترکیه با استفاده از مدل‌های MARS، الگوریتم کلونی زنبور عسل مصنوعی و الگوریتم بهینه‌سازی مبتنی بر آموزش و یادگیری پرداختند و نتایج حاصل از این پژوهش نشان داد که مدل MARS با دقت بیشتری بار معلق رسوب را شبیه‌سازی می‌کند (Yilmaz et al, 2018). صادق صفر (۲۰۲۰) پیش‌بینی بار معلق رسوب با استفاده از مدل‌های رگرسیون تطبیقی چند متغیره اسپلاین و جنگلی تصادفی، به همراه یک مدل

امروزه مدیریت کمی و کیفی منابع آب به سبب افزایش نیازمندی‌ها و مصارف جوامع انسانی، بیش‌ازپیش از اهمیت ویژه برخوردار است (حزباوی و همکاران، ۱۳۹۱). مدل‌سازی و پیش‌بینی رسوب رودخانه جهت مدیریت منابع آب و مهندسی محیط زیست، بسیار مهم و ضروری است (رجایی و جعفری، ۲۰۲۰). پدیده فرسایش خاک و انتقال رسوب از جمله فرآیندهای مهم هیدرودینامیک هستند که به‌عنوان یکی از بزرگ‌ترین چالش‌های بهره‌برداری از منابع آب‌های سطحی در جهان شناخته می‌شوند (دهقانی و همکاران، ۱۳۸۹). تخمین مقدار بار رسوبی یا میزان انتقال آن برای بسیاری از کاربری‌های پروژه‌های منابع آب از قبیل طراحی سدها و حوضچه‌های ذخیره آب سطحی، انتقال رسوب و آلودگی در رودخانه‌ها و دریاچه‌ها و نیز طراحی کانال‌ها و نگهداری آن‌ها حیاتی است (Yang et al, 2009). از دیگر دلایل اهمیت تخمین رسوب معلق رودخانه‌ها می‌توان به تأثیر آن بر روی کیفیت سامانه‌های آبی اشاره کرد (ابراهیمی محمدی و همکاران، ۱۳۹۱). جهت برآورد بار معلق رودخانه‌ها اصولاً از دو روش تجربی و رگرسیون آماری استفاده می‌شود. روابط مورداستفاده در روش تجربی مبتنی بر مفاهیم فیزیکی بوده و معمولاً به داده‌های متنوعی نیاز دارند. تعدد روابط تجربی ارائه‌شده، لزوم اصلاح این روابط و ارائه ضرایب واسنجی جدید، لزوم دستیابی به اطلاعات گسترده و عدم وجود روش تحلیلی یا تجربی مناسب که بر اساس آن بتوان به تخمین درستی از میزان رسوب حمل شده توسط جریان دست یافت از مهم‌ترین محدودیت‌های روش‌های تجربی برآورد بار رسوبی است. در روش رگرسیون آماری از پردازش بین مقادیر دبی جریان و رسوب معلق متناظر آن‌ها به صورت رابطه توانی $Q_s = aQ_w^b$ استفاده می‌شود. در این رابطه Q_s میزان دبی بار معلق و Q_w دبی جریان می‌باشد. ضرایب a و b ضرایب معادله هستند که با توجه به اطلاعات مورداستفاده، از طریق رگرسیون‌های آماری محاسبه می‌شوند (Achite and Ouillon, 2007).

مدل‌ها از دقت بیش‌تری برخوردار بوده و قابل‌اعتمادتر است (Zahiri and Nezratian, 2020).

در پژوهش حاضر جهت مدل‌سازی میزان بار معلق رودخانه‌های تیره-ماربره لرستان از مدل‌های ناپارامتریک شامل مدل درختی M5 و مدل رگرسیون تطبیقی چند متغیره اسپلاین MARS استفاده شد که امروزه از قوی‌ترین الگوریتم‌های داده‌کاوی محسوب می‌شوند. جهت ساخت و صحت‌سنجی مدل‌های درختی از اطلاعات میدانی مربوط به بار معلق رسوب ایستگاه‌های تیره مروک، تیره درود و نیز ماربره دره تخت و ماربره درود استفاده شد. پارامترهای ورودی به مدل شامل دبی جریان و خروجی آن نیز به‌صورت بار معلق رسوب در نظر گرفته شد. پس از ساخت و صحت‌سنجی مدل M5 و مدل رگرسیون تطبیقی چند متغیره اسپلاین MARS آنالیزهای آماری بر روی مدل‌های درختی صورت گرفت.

مواد و روش‌ها

منطقه مورد مطالعه

استان لرستان در غرب ایران بین $51^{\circ} 46'$ تا $50^{\circ} 3'$ طول شرقی و $32^{\circ} 37'$ تا $34^{\circ} 22'$ عرض شمالی واقع شده است و وسعت آن حدود ۲۸۵۵ کیلومترمربع است. رودخانه سزار یکی از رودخانه‌های دائمی و پرآب استان لرستان است. این رودخانه که سرشاخه اصلی رود دز در خوزستان است از مناطق کوهستانی الیگودرز، ازنا، دورود، و بروجرد سرچشمه می‌گیرد. رودخانه سزار از به هم پیوستن دو رودخانه ماربره (که از به هم پیوستن آب‌های الیگودرز، دورود و ازنا تشکیل می‌شود) و رودخانه تیره (که آب‌های گله رود، سراب سفید، کرتول و چند رود دیگر را دریافت می‌کند) در شهر دورود تشکیل می‌شود که از این محل تا ایستگاه تنگ پنج که با رودخانه بختیاری تلاقی می‌کند، سزار نامیده می‌شود. این رودخانه در پایین‌دست (در استان خوزستان) رودخانه دز نامیده می‌شود و سد بزرگ دز روی این رودخانه ساخته شده است. رودخانه سزار که گاهی به آن «آب دز» نیز می‌گویند از شاخه‌های اصلی رودخانه دز است که به‌عنوان یکی از بزرگ‌ترین رودخانه‌های زیرحوضه کارون بزرگ در استان لرستان واقع است. جهت ساخت و

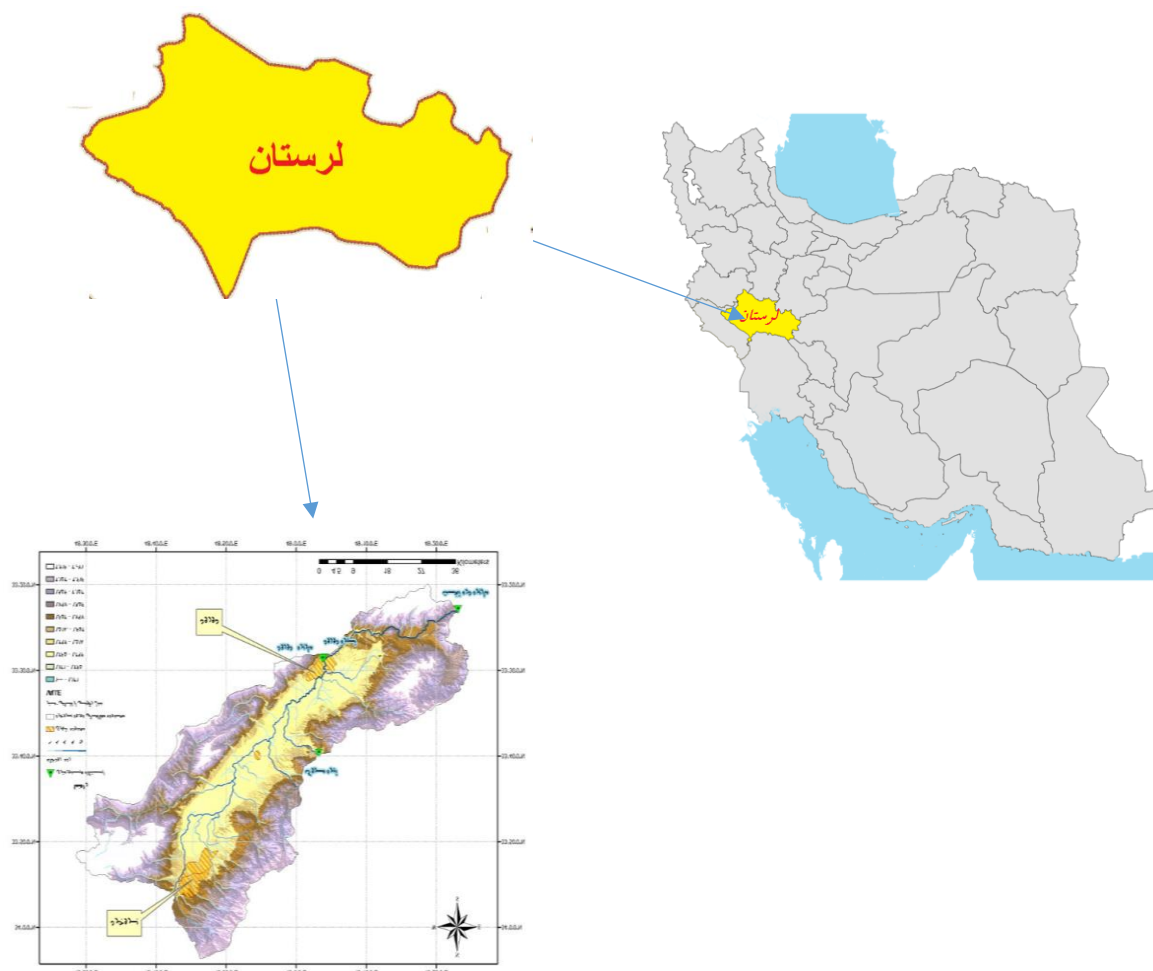
تجربی را مورد مطالعه قرار داد. بر اساس نتایج به‌دست‌آمده مشاهده گردید که مدل رگرسیون تطبیقی چند متغیره اسپلاین از دقت مناسبی در پیش‌بینی بار معلق برخوردار بوده است (Sadegh Safar, 2020).

بهمنی و همکاران (۲۰۲۰) به شبیه‌سازی سطح آب‌های زیرزمینی با استفاده مدل‌های بیان ژن، M5 و مدل ترکیبی پرداختند که نتایج مطالعه آن‌ها بیانگر آن بود که مدل مستقل بیان ژن، M5 با دقت بیشتری سطح آب‌های زیرزمینی را شبیه‌سازی می‌کند (Bahmani et al., 2020). هددام (۲۰۲۱) با استفاده مدل‌های رگرسیون تطبیقی چند متغیره اسپلاین، M5 و شبکه عصبی پرسپترون چندلایه تخمین رطوبت خاک را مورد مطالعه قرار داد (Heddham, 2021). حقی آبی پیش‌بینی ضریب پراکندگی طولی با استفاده از مدل MARS و شبکه عصبی را مورد مطالعه قرار داد. بر اساس نتایج این پژوهش، مدل MARS می‌تواند تخمین دقیق‌تری از ضریب پراکندگی طولی را نسبت به شبکه عصبی و روش‌های تجربی ارائه دهد (Haghiabi, 2016). از مسائل مهم در مدیریت منابع آب، تبخیر است. قائمی و همکاران (۲۰۱۹) با استفاده از مدل‌های MARS و M5 تبخیر ماهانه را پیش‌بینی کردند. نتایج مطالعه مذکور، نشان‌دهنده برتری این دو مدل نسبت به سایر روش‌ها بود (Ghaemi et al, 2019). تابش خورشیدی از نیازهای اساسی برای تولید برق در هر نیروگاه خورشیدی محسوب می‌گردد. راجیت و همکاران (۲۰۱۹) با استفاده از مدل‌های MARS، M5 و CART تابش خورشید را پیش‌بینی کردند. نتایج نشان‌دهنده دقت بیشتر MARS و M5 در این شبیه‌سازی بود (Rachit et al, 2019).

گارسیا و همکاران (۲۰۱۹) نیز با استفاده از مدل MARS و M5 به شبیه‌سازی یک نوع جلبک که عامل اصلی تخریب دریاچه و اکوسیستم دریاچه است پرداختند (García-Nieto et al., 2019). ظهیری و نازاریان به ارزیابی نتایج حاصل از برآورد ضریب اختلاط عرضی با استفاده از روش‌های MARS، M5، GA و PSO پرداختند. نتایج نشان داد مدل‌های MARS و M5 نسبت به سایر

شامل دبی جریان آب و دبی رسوب مربوط به چهار ایستگاه آب‌سنجی تیره مروک و تیره درود بر روی رودخانه تیره و نیز ماربره دره تخت و ماربره درود بر روی رودخانه ماربره در استان لرستان است. در شکل ۱ موقعیت ایستگاه‌های آب‌سنجی و حوضه آبریز آن نشان داده شده است.

صحت سنجی مدل‌های ناپارامتریک از داده‌ها و اطلاعات ایستگاه‌های هیدرومتری مورد تأیید شرکت آب منطقه‌ی لرستان استفاده شد. این اطلاعات به صورت میدانی برداشت شده و دبی جریان و نمونه‌برداری رسوب از نقاط مختلف رودخانه را شامل می‌شود. اطلاعات مورد استفاده



شکل (۱): موقعیت حوضه آبریز تیره و ماربره در سطح منطقه

الگوریتم M5

دوتایی و تنها بر اساس یک متغیر ایجاد می‌کند، بدین گونه که بر اساس شرطی که در هر گره تعریف می‌شود، اطلاعات در آن گره به دو قسمت تقسیم می‌شود (جباری و صمدی، ۱۳۹۲). در روش M5 فضای مساله به زیردامنه‌هایی

الگوریتم M5 اولین بار توسط کوئینلن در سال ۱۹۹۲ معرفی شد و پس از آن توسط ونگ و ویتن در سال ۱۹۹۷ توسعه و بهبود یافت. روش M5 شاخه‌های خود را به صورت



در این حالت، مدل دچار بیش برآزش شده که از طریق هرس کردن می‌توان این مشکل را رفع کرد. عمل هرس کردن باعث کاهش خطای مورد انتظار جهت داده‌های غیر آموزشی می‌شود. پس از هرس کردن، مدل‌های خطی مورد استفاده در برگ‌های مجاور درخت هرس شده به شدت دچار ناپیوستگی می‌گردند که این امر سبب از بین رفتن پیوستگی سیستم می‌گردد. بر همین اساس الگوریتم M5 از مکانیزی به نام هموارسازی جهت رفع ناپیوستگی ایجاد شده در مدل‌های خطی استفاده می‌کند.

در این فرآیند مقدار تخمین زده شده در هر برگ تصحیح می‌شود. چنانچه نمونه مورد نظر در شاخه S_i از زیر درخت S باشد، n_i تعداد نمونه‌های آموزشی در S_i ، $PV(S_i)$ مقدار محاسبه شده در S_i و $M(S)$ مقدار محاسبه شده توسط مدل در S باشد، آنگاه مقدار اصلاح شده $(Si(PV))$ از رابطه ۲ محاسبه خواهد شد.

$$PV = \frac{n_i \times pv(si) + k \times M(s)}{n_i + k} \quad (2)$$

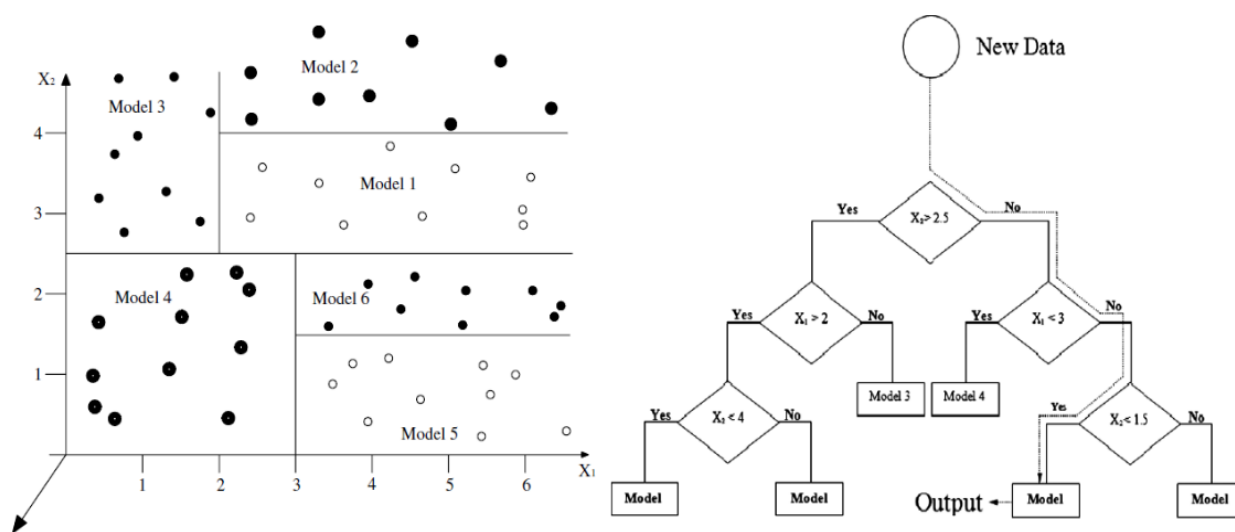
در رابطه فوق k ثابت هموارسازی بوده که به صورت پیش فرض برابر ۱۵ در نظر گرفته می‌شود. هموارسازی به‌ویژه در مواقعی که مدل‌های خطی در برگ‌های مجاور، مقادیر کاملاً متفاوتی ارائه می‌دهند و یا مدل‌هایی که بر اساس داده‌های آموزشی محدود ساخته می‌شوند، می‌تواند به میزان زیادی مؤثر باشد. تحقیقات انجام شده توسط (Wang and Witten, 1997) نشان داد که هموارسازی می‌تواند کارایی الگوریتم M5 را به میزان زیادی افزایش دهد.

تقسیم شده و برای هر زیر دامنه یک مدل رگرسیون خطی چندمتغیره برآزش داده می‌شود.

این الگوریتم، جداسازی‌های ممکن را در فضای چندمتغیره انجام داده و به‌طور خودکار مدل‌هایی را برای هرکدام از دامنه‌ها ایجاد می‌کند (Quinlan, 1992). در این الگوریتم از پارامتر انحراف معیار، مقادیر متغیر هدف به‌عنوان معیار اندازه‌گیری خطا در هر گره، جهت ایجاد شاخه در آن گره استفاده می‌شود. بدین صورت که صفتی که موجب کاهش بیشتر انحراف معیار در هر گره می‌شود به‌عنوان صفت مورد نظر جهت ایجاد شاخه انتخاب می‌شود. کاهش انحراف استاندارد (SDR) که به‌عنوان تابع خطا در الگوریتم M5 به کار می‌رود به‌صورت رابطه ۱ است:

$$SDR = sd(T) - \sum \left| \frac{Ti}{T} \right| \times sd(T) \quad (1)$$

در رابطه بالا T شامل نمونه‌هایی است که به گره مورد نظر رسیده‌اند و Ti شامل نمونه‌های است که از تقسیم گره مورد نظر بر اساس صفت انتخاب شده به دست آمده‌اند. Sd نیز معرف انحراف معیار می‌باشد. الگوریتم M5 تمامی حالت‌های مختلف جهت ایجاد شاخه بر اساس صفت خاص را بررسی کرده و در نهایت حالتی را انتخاب می‌کند که بتواند بیشتر از حالت‌های دیگر تابع خطای فوق را افزایش دهد. پس از تکمیل الگوریتم درختی برای نمونه‌های موجود در هر گره داخلی، یک مدل رگرسیون خطی چند متغیره برآزش داده می‌شود. شکل ۲ نحوه تقسیم فضای مسئله به اجزای کوچک‌تر و کاربرد مدل‌های خطی را برای هرکدام از زیر دامنه‌ها با توجه به ساختار درختی نشان می‌دهد (Etemad-Shahidi, and Taghipour, 2012). در طی ساخت مدل درختی توسط الگوریتم M5 فرآیند جداسازی در گره‌های تقسیم ممکن است بارها تکرار شده و در نتیجه درخت بسیار بزرگی ایجاد شود.



شکل (۲): تقسیم فضای مسئله و ارائه مدل‌های خطی به ازای هر کدام از فضاها توسط الگوریتم M5 (Etemad-Shahidi, and Taghipour, 2012)

دو تابع پایه، مجاور، لزوماً در یک گره متقاطع هستند که به موجب آن پیوستگی توابع پایه حفظ شود. توابع پایه به صورت مرحله‌ای به هر متغیر ورودی اعمال شده و مکان گره‌ها یعنی جایی که مقدار تابع تغییر می‌کند (یا شیب خطوط تغییر می‌کند)، تعیین می‌شوند. تعداد گره‌ها بر اساس یک فرآیند سعی و خطا حاصل می‌شوند. فرم عمومی مدل MARS به شکل زیر است:

$$Y=f(x)=\beta_0+\sum_{m=1}^M\beta_m h_m(x) \quad (4)$$

در رابطه فوق Y مقدار پیش‌بینی شده (متغیر هدف) بوده که توسط تابع $f(x)$ که به صورت ترکیبی از یک مقدار ثابت اولیه β_0 و مجموع M عبارت که هر کدام از یک ضریب β_m و یک تابع پایه $h_m(x)$ تشکیل شده‌اند، تعریف می‌شود. در اولین گام مدل MARS مقدار تابع هدف را با یک مقدار ثابت تخمین می‌زند و سپس در مسیری روبه‌جلو با جستجو در میان زوج‌ها (متغیر، گره) بهترین برازش را ایجاد می‌نماید. فرآیند جستجو تا زمانی که تمامی توابع پایه ممکن، به مدل اضافه شوند، ادامه می‌یابد. در این مرحله مدلی بسیار پیچیده، و بیش برازش شده با تعداد گره‌های زیاد به دست می‌آید. در گام بعد، با یک فرآیند هرس‌کننده

مدل MARS

مدل رگرسیون تطبیقی چند متغیره اسپیلاین (MARS) شکلی از الگوریتم‌های رگرسیونی بوده که توسط فریدمن در سال ۱۹۹۱ جهت پیش‌بینی خروجی‌های عددی پیوسته معرفی شد (Adamowski, 2012).

این تکنیک به وسیله تقسیم فضای مسئله به بازه‌هایی از متغیرهای پیش‌بینی کننده (ورودی) و برازش یک اسپیلاین (تابع پایه) در هر بازه، مدل‌های رگرسیونی انعطاف‌پذیری را برای پیش‌بینی متغیر هدف ایجاد می‌نماید (Zhang and Goh, 2016). تابع پایه نشان‌دهنده اطلاعاتی در بردارنده یک یا چند متغیر مستقل است. یک تابع پایه در یک بازه معین تعریف شده و نقاط ابتدایی و انتهایی آن گره نامیده می‌شود. گره، مفهوم کلیدی در این روش است و بیانگر نقطه‌ای است که رفتار تابع در آن نقطه تغییر می‌کند. توابع پایه ارتباط بین متغیرهای ورودی و متغیر هدف را بیان می‌کنند (Abraham et al., 2001) و توسط یکی از روابط زیر بیان می‌شود:

$$h_m(x)=\max(0, X-c) \text{ or } h_m(x)=\max(0, c-X) \quad (3)$$

که در رابطه بالا c یک مقدار آستانه^۱ است.

¹ Threshold

ترتیب از برنامه‌های WEKA و STATISTICA استفاده شد. جهت مقایسه نتایج مدل‌های مختلف و بررسی دقت آن‌ها از تحلیل‌های آماری متفاوتی استفاده شده است که از آن جمله می‌توان به مجموع مربعات خطا (SSE)^۲، خطای متوسط مطلق (MAE)^۳، جذر میانگین مربعات خطا (RMSE)^۴، ضریب تعیین (R^2)، ضریب کارایی نش-ساتکلیف (NSE) و DR^۵ اشاره کرد. معادلات مورد استفاده جهت تحلیل‌های آماری در روابط ۷ تا ۱۰ ارائه شده است.

$$SSE = \sum \frac{(d_s^{measured} - d_s^{computed})^2}{\sum (d_s^{measured})^2} \quad (7)$$

$$MAE = \frac{1}{N} \sum |d_s^{measured} - d_s^{computed}| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum (d_s^{measured} - d_s^{computed})^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum (d_s^{measured} - d_s^{computed})^2}{\sum (d_s^{measured} - d_s^{measured})^2} \quad (10)$$

$$NSE = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (d_s^{computed} - d_s^{measured})^2}{\frac{1}{N} \sum_{i=1}^N (d_s^{computed} - d_s^{computed})^2} \quad (11)$$

در روابط فوق $d_s^{measured}$ مقادیر مشاهداتی، $d_s^{computed}$ مقادیر پیش‌بینی شده، n تعداد داده‌ها و $d_s^{measured}$ میانگین مقادیر مشاهداتی هستند. مقدار ضریب R^2 بین صفر و یک است. هرچه این مقدار به عدد یک نزدیک‌تر باشد، قدرت پیش‌بینی مدل بالاتر و دقت بیشتر است و هر چه به صفر نزدیک‌تر باشد، قدرت و دقت پیش‌بینی پایین‌تر است. عدد صفر نشان می‌دهد که مدل هیچ‌یک از تغییرپذیری داده‌های پاسخ در اطراف میانگین آن را تبیین نمی‌کند و عدد یک نشان می‌دهد که مدل همه تغییرپذیری داده‌های پاسخ در اطراف میانگین آن را تبیین می‌کند. ضریب کارایی

رو به عقب بر اساس مدل ایجادشده، توابع پایه‌ای که اهمیت کمتری دارند شناسایی شده و حذف می‌شوند. این فرآیند تا بررسی کامل همه توابع پایه ادامه می‌یابد سرانجام بهترین مدل بر مبنای کمینه بودن معیاری به نام اعتبار سنجی تقاطعی تعمیم‌یافته^۱ (GCV) انتخاب می‌گردد. اگر فرض شود GCV_K مقدار GCV برای K امین مدل در مرحله حذفی باشد. آنگاه این کمیت به صورت رابطه زیر تعریف می‌شود:

$$GCV_K = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - f_K(x_i))^2}{\left(\frac{1 - C(K)}{n}\right)^2} \quad (5)$$

در رابطه فوق، f_K مدل برآورد شده در گام K ام مرحله حذفی پسرو بوده و $C(K)$ از رابطه ۶ به دست می‌آید. در این معادله M_K برابر است با تعداد جملات مدل در مرحله K ام، m برابر است با تعداد گره‌های توابع اسپلاین موجود در مدل و λ ضریبی بین عدد ۲ و ۴ خواهد بود.

$$C(K) = M_K + \lambda m \quad (6)$$

در پژوهش حاضر با توجه به اینکه مدل M5 تنها قادر به شبیه‌سازی مدل‌های خطی بوده و این امر با اساس معادلات متداول بار معلق رسوب در تناقض است، کلیه داده‌ها در ابتدا به صورت لگاریتم طبیعی به مدل M5 معرفی شدند و در پایان مدل‌های خطی به دست آمده از حالت لگاریتمی به توانی تبدیل شدند. در مدل MARS هم در ابتدا از لگاریتم طبیعی اعداد استفاده شد و پس از ساخت مدل، خروجی‌ها از حالت لگاریتمی به حالت توانی تبدیل شدند. با این کار از برآورد مقادیر منفی، برای کلیه پارامترها جلوگیری می‌شود. همچنین داده‌های ورودی در مراحل آموزش و صحت سنجی برای هر دو مدل به صورت یکسان معرفی شدند. در این تحقیق از کل اطلاعات در دسترس ۸۰ درصد جهت آموزش مدل‌ها و ۲۰ درصد باقیمانده جهت صحت سنجی مورد استفاده قرار گرفت. جهت ساخت مدل درختی دبی جریان با یک، دو و سه روز تأخیر و دبی جریان همان روز به همراه باران به عنوان پارامترهای ورودی به مدل‌ها معرفی شدند. برای مدل‌سازی M5 و MARS به

⁴ Root Mean Square Error

⁵ Determination Coefficient

¹ Generalised Cross Validation

² Sum of the Square Error

³ Mean-Absolute Error

این صورت ($DR < 0$)، بار معلق رسوب کمتر از مقدار واقعی محاسبه شده است.

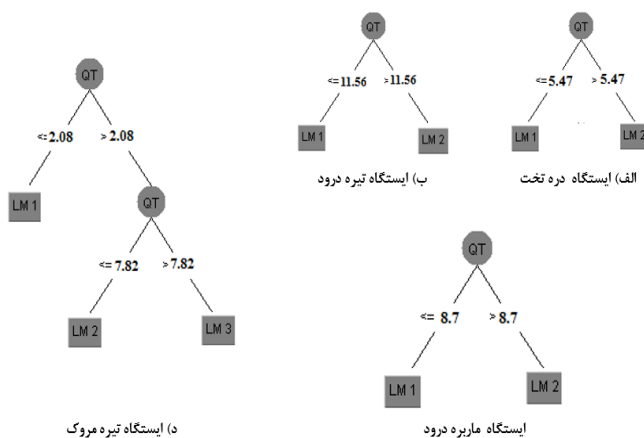
نتایج و بحث

در پژوهش حاضر پس از معرفی داده‌های مشاهده‌ای به صورت لگاریتم طبیعی به مدل M5، ساختار درختی توسط این مدل برای ایستگاه‌های مورد مطالعه ایجاد شد که این ساختار در شکل ۳ ارائه شده است. همان‌طور که قبلاً اشاره شد مدل M5 تنها قادر به ساخت مدل‌های خطی است و به همین دلیل و با توجه به مکانیزم انتقال رسوب که دارای رابطه توانی است، همه‌ی مقادیر در این پژوهش به صورت لگاریتمی به مدل معرفی شدند. مقادیر ارائه شده در شکل ۳ مقادیر واقعی بوده که از حالت لگاریتمی خارج شده‌اند.

نش- ساتکلیف (NSE) یک آمار نرمال شده است که میزان نسبی واریانس باقی مانده در مقایسه با واریانس داده‌های اندازه‌گیری شده را مشخص می‌کند. گستره ضریب کارایی نش ساتکلیف (NSE) بین $-\infty < NSE < 1$ است که عدد یک، مقدار بهینه است؛ به عبارتی دیگر هر چه این مقدار به سمت یک میل کند، جواب بهینه‌تری خواهد داشت. مقادیر بین صفر و یک معمولاً به عنوان سطح قابل قبول عملکرد، مشاهده می‌شوند.

مقادیر $NSE \leq 0$ نشان می‌دهد که میانگین مقدار مشاهداتی پیش‌بینی بهتری نسبت به مقدار محاسباتی داشته است.

همچنین یکی دیگر از مهم‌ترین شاخص‌های مورد استفاده، DR است. چنانچه DR برابر صفر باشد، نشان- دهنده تطابق کامل نتایج مدل و مقادیر واقعی است. در صورتی که ($DR > 0$) باشد مدل مورد استفاده، بار معلق رسوب را بیشتر از مقدار واقعی برآورد کرده است. در غیر



شکل (۳): ساختار درختی مدل M5 جهت برآورد بار معلق رسوب در ایستگاه‌های مختلف

و رسوب در این ایستگاه مربوط دانست که نسبت به بقیه ایستگاه‌ها دارای بازه تغییرات بیشتری است. معادلات ارائه شده توسط مدل درختی که در شکل ۳ با عنوان LM_i نشان داده شده است، معادلات خطی بوده که توسط الگوریتم درختی برای هر کدام از حالت‌ها توسعه داده شده است. این معادلات در جدول ۱ ارائه شده است.

شکل ۳ نشان می‌دهد که در تمامی ایستگاه‌ها، دبی جریان در همان روز به عنوان متغیر اصلی در ایجاد شاخه‌ها مورد استفاده قرار گرفته است. علاوه بر این در تمامی ایستگاه‌ها، مدل درختی جهت برآورد بار معلق تنها از دو شاخه استفاده کرده است و تنها در ایستگاه تیره مروک از چهار شاخه جهت محاسبه بار معلق استفاده شده است. مهم‌ترین دلیل این امر را می‌توان به پراکندگی مقادیر دبی

جدول (۱): معادلات مربوط به مدل M5

معادلات	محدوده کاربرد	ایستگاه
$LM 1: Q_S = 3.22Q_T^{1.2}$	$Q_T \leq 11.65$	تیره دورود
$LM 2: Q_S = 0.6Q_T^{2.15}$	$Q_T > 11.65$	
$LM 1: Q_S = 1.7Q_T^{1.7}$	$Q_T \leq 2.08$	تیره مروک
$LM 2: Q_S = 5.003Q_T^{0.85}$	$2.08 < Q_T \leq 7.82$	
$LM 3: Q_S = 1.46Q_T^{2.065}$	$Q_T > 7.82$	ماربره دورود
$LM 1: Q_S = 1.57Q_{T-1}^{0.53}Q_T^{1.13}$	$Q_T \leq 8.7$	
$LM 2: Q_S = 0.3Q_{T-1}^{0.43}Q_T^{2.16}$	$Q_T > 8.7$	ماربره دره تخت
$LM 1: Q_S = 2.013Q_{T-1}^{1.04}Q_T^{0.7}$	$Q_T \leq 5.47$	
$LM 2: Q_S = 0.91Q_{T-1}^{2.03}Q_T^{0.02}$	$Q_T > 5.47$	

کاهش داده شده و یا این پارامتر از معادلات ارائه شده حذف گردیده است.

مدل رگرسیون تطبیقی چند متغیره اسپلین MARS در مرحله اول آماده سازی مدل برای چهار ایستگاه ۲۱ تابع پایه در نظر گرفته شده سپس در مرحله دوم (هرس کردن) به ترتیب ۱۹، ۱۹، ۱۷ و ۱۶ تابع پایه برای ایستگاه های تیره درود، تیره مروک، ماربره درود و ماربره دره تخت هرس شده و در نهایت مدل بهینه رگرسیون تطبیقی چند متغیره به ترتیب برای ایستگاه های مورد استفاده در این تحقیق با ۲، ۲، ۴ و ۵ تابع اصلی به دست آمد. مقادیر پارامتر هرس (GCV) برای ایستگاه های تیره درود، تیره مروک، ماربره درود و ماربره دره تخت به ترتیب برابر با ۱/۰۲، ۱/۴۳، ۱/۳۷ و ۰/۹۷ می باشد. مدل های بهینه ارائه شده توسط مدل رگرسیون تطبیقی چند متغیره در جدول ۲ ارائه شده اند. همچنین در جدول ۳ توابع پایه به دست آمده از مدل رگرسیون تطبیقی چند متغیره برای چهار ایستگاه مورد استفاده در این تحقیق شرح داده شده است.

در این روابط Q_S دبی رسوب و Q_T و Q_{T-1} دبی جریان همان روز و یک روز قبل است. در تمامی معادلات ارائه شده توسط مدل درختی، میان دبی رسوب و دبی جریان مربوط به همان روز و روز قبل رابطه مستقیم وجود داشته که با فیزیک مسئله انتقال رسوب مطابقت دارد. در ایستگاه تیره مروک که بیشترین شاخه ها را در میان بقیه ایستگاه ها دارد، سه معادله ارائه شده است که تنها شامل دبی همان روز است که این نشان دهنده تأثیر بیشتر این متغیر نسبت به دبی یک روز قبل بر روی میزان غلظت رسوب معلق است. ولی در ایستگاه ماربره از دو پارامتر Q_T و Q_{T-1} استفاده شده است و مشاهده می گردد که توان مربوط به دبی همان روز به طور محسوسی نسبت به دبی روز قبل بیشتر است. با توجه به ساختار درختی ارائه شده این ایستگاه، معادله به دست آمده مربوط به دبی های بالای ۶۰ مترمکعب بر ثانیه بوده که به عنوان دبی سیلابی در این ایستگاه به حساب می آیند. همین روند در مورد بقیه ایستگاه ها نیز صادق است. به گونه ای که در دبی های بالا، توان مربوط به دبی روز قبل نسبت به دبی همان روز

می‌نماید. بر اساس تعداد تکرارها و ضریب هر پارامتر ورودی در مدل MARS، می‌توان اهمیت آن پارامتر را تعیین کرد (Haghiabi, 2016; Samadi et al, 2016; Zhang and Goh, 2016)

مدل MARS به‌وسیله تقسیم فضای مسئله به بازه‌هایی از متغیرهای پیش‌بینی کننده (ورودی) و برازش یک اسپلاین (تابع پایه) در هر بازه، مدل‌های رگرسیونی انعطاف‌پذیری را برای پیش‌بینی متغیر هدف ایجاد

جدول (۲): مدل‌های بهینه ارائه شده توسط مدل MARS

معادلات	ایستگاه
$Q_S = \exp(3.49 + \sum_{i=1}^2 \beta_m h_m(x))$	تیره دورود
$Q_S = \exp(1.49 + \sum_{i=1}^2 \beta_m h_m(x))$	تیره مروک
$Q_S = \exp(2.57 + \sum_{i=1}^4 \beta_m h_m(x))$	ماربره دورود
$Q_S = \exp(2.04 + \sum_{i=1}^5 \beta_m h_m(x))$	ماربره دره تخت

جدول (۳): توابع پایه و ضرایب مربوط به مدل ماریس

تابع پایه	معادله	ضریب (β_i)	ایستگاه
$h_1(x)$	MAX (0; QT -1.95)	2.25	تیره دورود
$h_2(x)$	MAX (0; 1.95-QT)	- 1.16	
$h_1(x)$	MAX (0; QT -0.93)	2.47	تیره مروک
$h_2(x)$	MAX (0; 0.93- QT)	- 1.04	
$h_1(x)$	MAX (0; QT-1.78)	2.472	ماربره دورود
$h_2(x)$	MAX (0; 1.8-QT)	- 1.023	
$h_3(x)$	MAX (0; QT-1-1.36)	- 0.95	
$h_4(x)$	MAX (0; QT-1 -0.83)	1.27	

ادامه جدول (۳): توابع پایه و ضرایب مربوط به مدل مارس

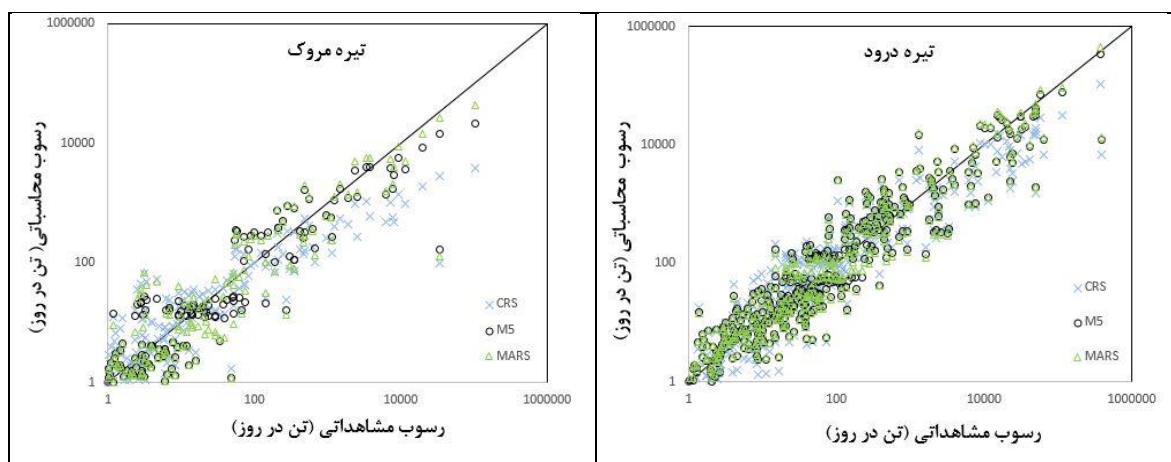
$h_1(x)$	MAX (0; QT-1.3)	2.07	
$h_2(x)$	MAX (0; 1.3-QT)	- 0.85	
$h_3(x)$	MAX (0; QT-1 -3.09)	3.2	
$h_4(x)$	MAX (0; QT-1 -2.18)	- 1.5	ماربره دره تخت
$h_5(x)$	MAX (0; QT-1 -0.74)	0.49	

پراکندگی داده‌ها در اطراف خط ۴۵ درجه بیشتر باشد، نشان‌دهنده دقت بالاتر مدل خواهد بود. در کلیه گراف‌ها نتایج مدل‌های M5 و MARS در مقایسه با منحنی سنجه انطباق بهتری با مقادیر مشاهداتی داشته و پراکندگی بهتری در اطراف خط ۴۵ درجه نشان می‌دهند.

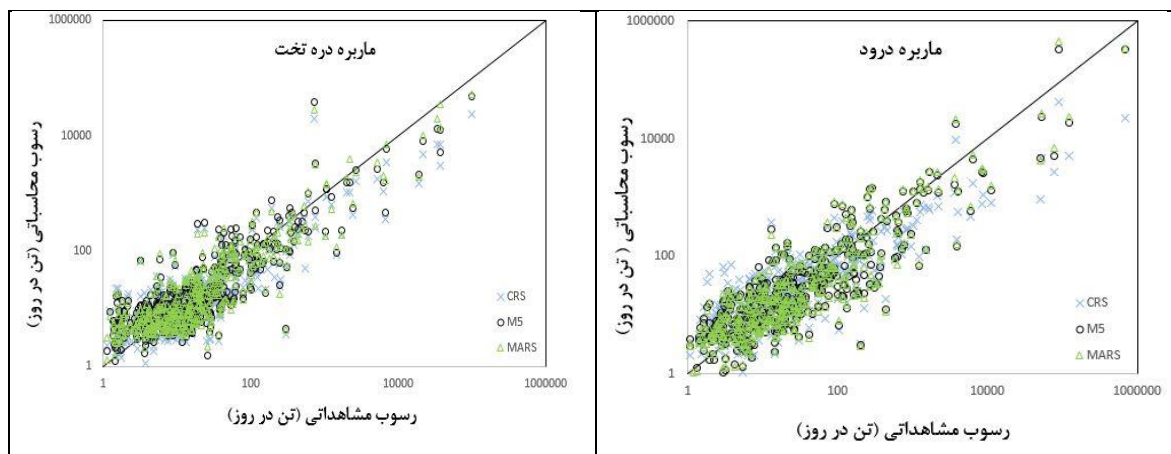
علاوه بر این روش منحنی سنجه که متداول‌ترین روش محاسبه بار معلق به حساب می‌آید در دبی‌های سیلابی که حداکثر رسوب را به همراه دارند، مقدار رسوب معلق را نسبت به الگوریتم‌های درختی کمتر از مقدار واقعی برآورد نموده است. این امر با توجه به اهمیت بار معلق رسوب در دبی‌های سیلابی و در نظر گرفتن ضریب اطمینان برای پروژه‌های مهندسی رودخانه می‌تواند به‌عنوان یکی از نقاط ضعف روش منحنی سنجه رسوب به حساب آید.

همان‌طور که در توابع پایه به‌دست‌آمده از مدل MARS در جدول ۳ مشاهده می‌شود در تمامی ایستگاه‌ها دبی جریان در همان روز به‌عنوان متغیر اصلی در تابع پایه مورد استفاده قرار گرفته است که بیانگر اهمیت این پارامتر در تخمین رسوب است. نتایج مربوط به مقادیر بار معلق محاسبه‌شده توسط مدل‌های درختی M5، MARS و منحنی سنجه رسوب در ایستگاه‌های هیدرومتری مطالعه شده در این پژوهش در شکل‌های ۴ تا ۵ نشان داده شده است.

در تمام گراف‌های ارائه شده، محور افقی نشان‌دهنده مقادیر رسوب معلق مشاهده‌شده و محور قائم نیز معرف مقادیر محاسبه‌شده توسط الگوریتم‌های درختی و منحنی سنجه رسوب است. بر اساس گراف‌های ارائه شده هر چه



شکل (۴): مقایسه پراکندگی نتایج مدل‌های درختی و منحنی سنجه رسوب در ایستگاه‌های تیره درود و تیره مروک



شکل (۵): مقایسه پراکندگی نتایج مدل‌های درختی و منحنی سنج رسوب در ایستگاه‌های ماربره درود و ماربره دره تخت

در جدول ۴ ضریب تبیین برای مدل‌های داده‌کاوی نسبت به منحنی سنج رسوب به ۱ نزدیک‌تر است. در مورد جذر میانگین مربعات خطا و ضریب نش ساتکلیف نیز همین روند مشاهده می‌شود. همچنین در مورد دقت مدل‌ها با توجه به مقادیر DR در تمامی ایستگاه‌ها دو مدل M5 و MARS نسبت به منحنی سنج رسوب دارای دقت بالاتری می‌باشند. $DR < -1$ نشان‌دهنده مقادیری است که توسط روش موردنظر، بیش از ۱۰ برابر کمتر از مقادیر مشاهداتی محاسبه شده است و در $DR > 1$ بار معلق رسوب بیش از ۱۰ برابر بیشتر از مقادیر مشاهداتی برآورد شده است.

جهت بررسی دقت مدل‌ها در این پژوهش از شاخص‌های آماری متنوعی استفاده شد که نتایج آنالیزهای صورت گرفته در جدول ۴ ارائه شده است. در جدول ۴ درصد DR مربوط به مقادیر پیش‌بینی شده ارائه شده است. محدوده DR بین -0.3 و 0.3 به‌عنوان بازه مناسب برای مقادیر پیش‌بینی شده دقت هر مدل در نظر گرفته شده است.

نتایج جدول ۴ نشان می‌دهد که در مجموع مدل‌های داده‌کاوی بار رسوب معلق را با خطای کمتر و ضریب تبیین بالاتری در مقایسه با منحنی سنج رسوب برآورد کرده‌اند.

جدول (۴): بررسی کارایی مدل‌های هوشمند و تجربی در تخمین بار معلق رسوب

NSE	$RMSE$	R^2	دقت	$DR > 0.3$	$0 < DR < 0.3$	$-0.3 < DR < 0$	$DR < -0.3$	مدل	ایستگاه
۰/۷۸	۰/۴۲	۰/۸۷	۵۴	۲۳	۲۷	۲۷	۲۰	مدل M5	تیره درود
۰/۷۷	۰/۴۲	۰/۸۷	۵۴	۲۳	۲۸	۲۶	۲۰	مدل MARS	
۰/۶۹	۰/۴۶	۰/۸۳	۴۷	۲۵	۲۴	۲۳	۲۲	منحنی سنج رسوب	
۰/۷۹	۰/۵۱	۰/۸۵	۴۵	۲۶	۲۰	۲۵	۲۵	مدل M5	تیره مروک
۰/۷۸	۰/۵۲	۰/۸۵	۴۵	۲۴	۲۰	۲۵	۲۵	مدل MARS	
۰/۶۷	۰/۵۹	۰/۷۸	۳۸	۲۵	۲۰	۱۸	۲۵	منحنی سنج رسوب	ماربره درود
۰/۵۷	۰/۴۷	۰/۷۱	۴۹	۲۵	۲۳	۲۶	۲۲	مدل M5	
۰/۵۹	۰/۴۶	۰/۷۲	۴۷	۲۴	۲۱	۲۶	۲۲	مدل MARS	

ادامه جدول (۴): بررسی کارایی مدل‌های هوشمند و تجربی در تخمین بار معلق رسوب

۰/۴۷	۰/۵۶	۰/۶۴	۴۴	۲۵	۲۳	۲۱	۲۴	منحنی سنجه رسوب
۰/۵۹	۰/۴۴	۰/۶۵	۵۵	۲۰	۲۴	۳۱	۲۱	مدل M5
۰/۵۶	۰/۴۳	۰/۶۵	۵۷	۱۹	۲۶	۳۱	۲۰	مدل MARS ماربره
۰/۴۸	۰/۴۹	۰/۵۷	۴۹	۲۱	۲۵	۲۴	۲۶	منحنی سنجه رسوب دره تخت

نتیجه‌گیری

بیشتری در تخمین مقدار رسوب معلق است که این نکته می‌تواند به‌عنوان یک ضعف اساسی برای این روش محسوب شود. در نهایت، مقایسه کارایی دو مدل درختی نشان داد که نتایج دو مدل نزدیک به هم بوده اما روابط پیشنهادی توسط مدل MARS به‌مراتب پیچیده‌تر از روابط مدل M5 است. آداموسکی و همکاران (۲۰۱۲) نیز از مدل MARS و شبکه عصبی جهت پیش‌بینی سیلاب استفاده کردند که در نهایت دقت مدل تطبیقی از شبکه عصبی بیشتر بود. نتایج مطالعه‌ی ظهیری (۱۳۹۴) نیز دقت بالاتر مدل M5 را در مقایسه با مدل درختی CART و سایر روش‌های متداول نشان داد. شاهرخی و همکاران (۱۳۹۵) نیز مدل M5 را مدلی مناسب و ساده و دارای دقت بالا معرفی کردند. نتایج مطالعه‌ی ییلماز و همکاران (۲۰۱۸) نیز نشان داد که مدل MARS نیز با دقت بالایی، بار معلق رسوب را شبیه‌سازی می‌کند. حقی‌آبی (۲۰۱۶) نیز مدل MARS را برتر از شبکه عصبی معرفی کرد. نتایج مطالعه ظهیری و نزارتجان (۲۰۲۰) نیز نشان داد که مدل‌های M5 و MARS نسبت به سایر مدل‌ها از دقت بیشتری برخوردار بوده و قابل‌اعتمادتر است. در نهایت با توجه به دقت بالا و سادگی معادلات ارائه شده توسط مدل M5 و قابل‌درک بودن آن جهت برآورد بار معلق رسوب، این مدل به‌عنوان روش برتر در برآورد بار معلق رودخانه پیشنهاد می‌گردد.

در پژوهش حاضر با به‌کارگیری الگوریتم‌های M5 و MARS با استفاده از پارامترهای ورودی شامل دبی جریان با یک، دو و سه روز تأخیر و همان روز به همراه بارش اقدام به برآورد بار معلق رسوب گردید. همچنین از روش منحنی سنجه رسوب که روش متداول محاسبه بار معلق در رودخانه‌ها به‌حساب می‌آید برای مقایسه با نتایج حاصل از مدل‌های درختی و در نهایت تعیین بهترین روش استفاده شد. نتایج بدست آمده نشان داد که از میان پارامترهای ورودی فقط دبی جریان روز قبل و همان روز جهت برآورد بار معلق مورد استفاده قرار گرفته‌اند و مقادیر پیش‌بینی بیش از هر چیزی متأثر از این دو عامل بوده‌اند. همچنین بر اساس آنالیزهای آماری صورت گرفته مشاهده گردید که نتایج دو مدل درختی نسبت به منحنی سنجه رسوب از دقت بهتر و بالاتری برخوردار هستند. ایستگاه ماربره درود بیشترین تفاوت را میان نتایج مدل‌های درختی و منحنی سنجه نشان داد به‌گونه‌ای که جذر میانگین مربعات خطا برای مدل‌های M5 و MARS به ترتیب به میزان ۹ و ۱۰ درصد بهبود یافته است. در ایستگاه تیره درود نیز کمترین تفاوت میان مدل‌های درختی و منحنی سنجه مشاهده شد که این اختلاف به میزان ۴ درصد برآورد شد. علاوه بر آن بررسی مقادیر پیش‌بینی شده نشان داد که منحنی سنجه رسوب بیشترین مقادیر با $DR < -1$ را داراست که این نشان‌دهنده برآورد بسیار پایین مقدار رسوب معلق است. همین نکته در مورد $DR > 1$ نیز وجود دارد، بدین معنی که منحنی سنجه رسوب درصد بیشتری از مقادیر بار معلق را با اختلاف زیاد بیشتر از مقادیر واقعی برآورد می‌کند. بر این اساس، روش منحنی سنجه نسبت به دو روش دیگر دارای درصد برآوردهای با دقت پایین

منابع

- ابراهیمی محمدی، ش.، س. ح. ر. صادقی و ک. چپی. ۱۳۹۱. تحلیل آورد رواناب، رسوب معلق مغذی ورودی‌های مختلف به دریاچه زریوار در پایه زمانی رگبار و آب پایه. نشریه حفاظت منابع آب و خاک، ۲: ۷۴-۶۱.
- جباری، ا.، م. صمدی. ۱۳۹۲. کاربرد الگوریتم M5 در پیش‌بینی عمق آبستگي در پایین دست سريزها. نشریه وزارت نیرو.
- حزبواي، ز.، س. ح. ر. صادقی و ح. ا. یونسی. ۱۳۹۱. تحلیل و ارزیابی تأثیرپذیری مؤلفه‌های رواناب از کاربرد سطوح مختلف پلی‌آکريل آميد. نشریه حفاظت منابع آب و خاک، نشریه حفاظت منابع آب و خاک. ۲(۲): ۱-۱۴.
- دهقانی، ا.، م. ملک‌محمدی و ا. هزارجریبی. ۱۳۸۹. تخمین بار معلق رودخانه بهشت آباد با استفاده از شبکه عصبی مصنوعی. مجله پژوهش‌های حفاظت آب و خاک. ۱(۱۷): ۱۵۹-۱۶۸.
- ظهیری، ج. ۱۳۹۴. کاربرد مدل‌های ناپارامتریک CART و M5 در محاسبه عمق آبستگي اطراف پایه‌های پل. مجله کرمان، ۵(۲۰): ۳۵-۵۰.
- شاهرخی، س. ح.، ج. ظهیری و ا. جعفری. ۱۳۹۵. کاربرد الگوریتم درختی M5 در برآورد رسوب معلق رودخانه‌ها. نشریه علمی پژوهشی مهندسی آبیاری و آب ایران. ۶(۴): ۱۶-۲۸.

Abraham, A., D. Steinberg, N. S. Philip. 2001. Rainfall forecasting using soft computing models and multivariate adaptive regression splines IEEE SMC Transactions, Special Issue on Fusion of Soft Computing and Hard.

Achite, M., S. Ouillon. 2007. Suspended sediment transport in a semiarid watershed, Wadi Abd, Algeria (1973-1995). Journal of Hydrology, 343: 187-202.

Adamowski, J., H. F. Chan, S. O. Prasher, V.N. Sharda. 2012. Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. Journal of Hydroinformatics 14(3): 731-744.

Bahmani, B., B. M. Taha, J. Ouarda. 2020. Groundwater level modeling with hybrid artificial intelligence techniques. Journal of Hydrology, <https://doi.org/10.1016/j.jhydrol.2020.125659>.

Etemad-Shahidi, A., N. Ghaemi, 2011. Model tree approach for prediction of pile groups scour due to waves. Journal of Ocean Engineering, 38, 1522-1527.

Etemad-Shahidi, A., J. Mahjoobi. 2009. Comparison between M5' model tree and neural networks for prediction of significant wave height. Journal of Ocean Engineering, 36(15-16): 1175-1181.

Etemad-Shahidi, A., M. Taghipour. 2012. Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. Journal of Hydraulic Engineering, 138(6): 542-554.

García-Nieto, P.J., E. García-Gonzalo, J.R. Alonso Fernández, C. Díaz Muñoz. 2019. Modeling algal atypical proliferation using the hybrid DE-MARS-based approach and M5 model tree in LaBarca reservoir: A case study in northern Spain. Journal homepage.

Ghaemi, A., M. Rezaie-Balfa, J. Adamowski, O. Kisić, J. Quilty. 2019. On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. Journal homepage. <https://doi.org/10.1016/j.agrformet.2019.107647>.

Haghiabi, A. H. 2016. Prediction of longitudinal dispersion coefficient using multivariate adaptive regression splines. J Earth Syst Sci 125:985-995.

Heddum, S. 2021. New formulation for predicting soil moisture content using only soil temperature as predictor: multivariate adaptive regression splines versus random forest, multilayer perceptron neural



network, M5Tree, and multiple linear regression. *Journal Water Engineering Modeling and Mathematic Tools*, Pages 45-62, <https://doi.org/10.1016/B978-0-12-820644-7.00027-X> Get rights and content.

Quinlan, J. R. 1992. Learning with continuous classes. *Proc., 5th Australian Joint Conf. on Artificial Intelligence*, World Scientific, Singapore, 343-348.

Rachit, S., A.N. Tiwari, V.K. Giri. 2019. Solar radiation forecasting using MARS, CART, M5, and random forestmodel: A case study for India. *Journal homepage*. <https://doi.org/10.1016/j.heliyon.2019>.

Rajae, T., h. Jafari. 2020. Two decades on the artificial intelligence models 2advancement for modeling river sediment concentration: 3State-of-the-art. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2020.125011>.

Sadegh Safar, M. J. 2020. Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. *Journal Pre-proofs*, S0022-1694(20)30852-0, <https://doi.org/10.1016/j.jhydrol.2020.125392>.

Samadi, M., E. Jabbari, H. M. Azamathulla, M. Mojallal. 2015. Estimation of scour depth below free overfall spillways using multivariate adaptive regression splines and artificial neural networks. *Eng ApplComp Fluid* 9:291-300. <https://doi.org/10.1080/19942060.2015.1011826>.

Wang, Y., I. H. Witten, 1997. Induction of Model Trees for predicting Continuous Classes. *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague.

Yang, C.T., R. Marsooli, M. Taghi Aalami. 2009. Evaluation of total load sediment transport formulas using ann. *International Journal of Sediment Research*, 24, 274-286.

Yilmaz, B., E. Aras, S. Nacar, M. Kankal. 2018. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Sci Total Environ* 639:826-840. <https://doi.org/10.1016/j.scitotenv.2018.05.153>.

Zahiri1, Z., H. Nezaratian. 2020. Estimation of transverse mixing coefficient in streams using M5, MARS, GA, and PSO approaches. Springer-Verlag GmbH Germany, part of Springer Nature 2020.

Zhang, W., A.T.C. Goh. 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front* 7:45-52. <https://doi.org/10.1016/j.gsf.2014.10.003>.

Zheng, G., W. Zhang, H. Zhou, P. Yang. 2020. Multivariate adaptive regression splines model for prediction of the liquefaction-induced settlement of shallow foundations. *Journal homepage* <https://doi.org/10.1016/j.soildyn.2020.106097>.



Estimation of suspended sediment using non-parametric M5 models and spline multivariate adaptive regression (MARS) (Case study: Tireh-Marbareh Rivers of Lorestan)

Babak Shahinejad^{*}, Seyedeh Hadis Shahrokhi^۲, Hossein Yousefi^۳

Abstract

In the present study, the M5 tree algorithm and MARS multivariate comparative regression have been used as new methods for estimating the suspended sediment load in comparison with the sediment measurement curve method. The information used in this study includes water flow and sediment discharge related to four hydrometric stations of Doroud and Tireh Marvak hydrometric stations on Tireh River, as well as Marbareh Doroud and Marbareh Darreh Takht on Marbareh River in Lorestan province. For fabrication and validation of the models, the flow rate with one, two and three days delay and the flow rate of the same day with rain as input parameters and the suspended sediment load flow rate were considered as output parameters. Statistical analyzes were used to evaluate the efficiency of the models and compare their results with conventional methods. In Marbareh Doroud station, the values of RMSE and R2 for the M5 model were 0.47 and 0.71, respectively, and for the MARS model were 0.46 and 0.72, respectively, while in the measurement curve method they were 0.56 and 0.64. The performance of the proposed models indicates an improvement in their accuracy and ability to estimate the suspended sediment load. The results showed that the equations presented by the M5 and MARS tree models are more accurate than the measurement curve. Based on the results, it was observed that the two methods M5 and MARS have provided close answers to each other, but finally, due to the simple and conceptual structure of the M5 model, this method is a more appropriate method for estimating the suspended load in the case range. The study was selected. In addition, the study of the relationships obtained from the two models M5 and MARS showed that among the input parameters, the flow rate of the previous day and the same day were used to estimate the suspended load and the prediction values were affected more than anything. These two factors have been.

Keywords: MARS, Measurement curve, Suspended sediment, Tireh-Marbareh, Tree model M5

¹ Assistant Professor, Department of Water Engineering, Faculty of Agriculture and Natural Resources, Lorestan University, Khorramabad, Iran, Phone: 09163676090 Email: Shahinejad.b@lu.ac.ir

² PhD Student in Hydraulic Structures, Faculty of Agriculture and Natural Resources, Lorestan University, Khorramabad, Iran, Phone: 09163676090 Email: sh.hadis815@gmail.com

³ Associate Professor, Department of New Energy and Environment, Faculty of Modern Science and

Technology, University of Tehran, Tehran, Iran, Phone: 09124151174 E-mail: hosseinyousefi@ut.ac.ir

Extended Abstract

Research Paper

Estimation of suspended sediment using non-parametric M5 models and spline multivariate adaptive regression (MARS) (Case study: Tireh-Marbareh Rivers of Lorestan)**Babak Shahinejad^{1*}, Seyedeh Hadis Shahrokhi², Hossein Yousefi³**

1. Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran Corresponding Author: Shahinejad.b@lu.ac.ir

2 PhD student in Hydraulic Structures, Lorestan University, Khorramabad, Iran.,

3. Associate Professor, Department of New Energy and Environment, University of Tehran, Iran.



10.22125/IWE.2021.279055.1493

Received:

07. October.2020

Accepted:

30. March.2021

Available online:

10. January.2022

Keywords: MARS, Measurement curve, Suspended sediment, Tireh-Marbareh, Tree model M5

Abstract

In the present study, the M5 tree algorithm and MARS multivariate comparative regression have been used as new methods for estimating the suspended sediment load in comparison with the sediment measurement curve method. The information used in this study includes water flow and sediment discharge related to four hydrometric stations of Doroud and Tireh Marvak hydrometric stations on Tireh river, as well as Marbareh Doroud and Marbareh Darreh Takht on Marbareh river in Lorestan province. For fabrication and validation of the models, the flow rate with one, two- and three-days delay and the flow rate of the same day with rain as input parameters and the suspended sediment load flow rate were considered as output parameters. Statistical analyzes were used to evaluate the efficiency of the models and compare their results with conventional methods. In Marbareh Doroud station, the values of RMSE and R2 for the M5 model were 0.47 and 0.71, respectively, and for the MARS model were 0.46 and 0.72, respectively, while in the measurement curve method they were 0.56 and 0.64. The performance of the proposed models indicates an improvement in their accuracy and ability to estimate the suspended sediment load. The results showed that the equations presented by the M5 and MARS tree models are more accurate than the measurement curve. Based on the results, it was observed that the two methods M5 and MARS have provided close answers to each other, but finally, due to the simple and conceptual structure of the M5 model, this method is a more appropriate method for estimating the suspended load in the case range. The study was selected. In addition, the study of the relationships obtained from the two models M5 and MARS showed that among the input parameters, the flow rate of the previous day and the same day were used to estimate the suspended load and the prediction values were affected more than anything. These two factors have been.

1. Introduction

Modeling and forecasting river sediment is very important and necessary for water resources management and environmental engineering (Rajaei and Jafari, 2020). Etemad Shahidi and Mahjoubi used the M5 tree model to simulate the height of river waves. In addition to the tree model, they used a neural network to compare the efficiency of the tree model and showed that the tree model can more accurately simulate the height of the waves (Etemad Shahidi and Mahjoubi, 2009) in the present study to model Non-parametric models including M5 tree model and MARS multivariate comparative regression model were used to construct the suspended load of Lorestan Tireh-Marboreh rivers, which are considered as the most powerful data mining algorithms today. In order to construct and validate the tree models, field information related to the suspended sediment load of Tireh Marvak, Tireh Droud stations, as well as Darreh Takht and Marbreh Dorod stations were used. Input parameters to the model including flow rate and output were also considered as suspended sediment load. After fabrication and validation of M5 model and MARS multivariate comparative regression model, statistical analyzes were performed on tree models.

2. Materials and Methods

Cesar River is one of the permanent rivers of Lorestan province. Cesar River is one of the main tributaries of Dez River, which is located as one of the largest rivers in the Karun Bozorg Basin in Lorestan Province. Data and information of hydrometric stations of Lorestan Regional Water Company were used to construct and validate non-parametric models. This information is collected in the field and includes flow rate and sediment sampling from different parts of the river. The information used includes water flow and sediment flow related to four hydrometric stations. The M5 algorithm was first introduced by Quinlan in 1992 and later developed and improved by Wong and Witten in 1997. This algorithm performs possible separations in multivariate space and automatically creates models for each domain. Research conducted by (Wang and Witten, 1997) showed that smoothing can greatly increase the efficiency of the M5 algorithm. The Spillin multivariate adaptive regression model (MARS) is a form of regression algorithms introduced by Friedman in 1991 to predict continuous numerical outputs (Adamowski, 2012). This technique creates flexible regression models for predicting the target variable by dividing the problem space into intervals of predictor variables (input) and fitting a spline (base function) in each interval (Zhang and Goh, 2016). To evaluate the accuracy of the models in this study, various statistical indicators were used.

3. Results

Comparing the performance of the two tree models showed that the results of the two models were close to each other, but the relationships proposed by the MARS model were more complex than the relationships of the M5 model. Adamowski et al. (2012) also used MARS model and neural network to predict flood, which was more accurate than adaptive model than neural network. The results of Zahiri and Nizartian (2020) study also showed that M5 and MARS models are more accurate and more reliable than other models. Finally, due to the high accuracy and simplicity of the equations presented by the M5 model and its comprehensibility for estimating the suspended sediment load, this model is proposed as a superior method in estimating the suspended river load.

4. Discussion and Conclusion

In the present study, after introducing observational data as a natural logarithm to the M5 model, a tree structure was created by this model for the studied stations. Model M5 is only able to build linear models and due to the sediment transport mechanism which has a power relationship, all values in this study were introduced to the model in a logarithmic way. In all stations, the flow rate on the same day is used as the main variable in creating branches. In addition, in all stations, the tree model used only two branches to estimate the suspended load, and only in the dark station of Marvak, four branches were used to calculate the suspended load. The most important reason for this is the dispersion of discharge and sediment in this station, which has a greater range of changes than other stations. In the basic functions obtained from the MARS model, it is observed that in all stations, the flow rate on the same day is used as the main variable in the basic function, which indicates the importance of this

parameter in estimating sediment. According to the graphs provided, the higher the data scatter around the 45 ° line, the higher the accuracy of the model. In all graphs, the results of M5 and MARS models are in better agreement with the observed values compared to the measurement curve and show better scatter around the 45 ° line. The measurement curve method, which is the most common method of calculating suspended load in flood discharges with maximum sediment, has estimated the amount of suspended sediment less than the actual value compared to tree algorithms. This can be considered as one of the weaknesses of the sediment measurement curve method due to the importance of suspended sediment load in flood discharges and considering the reliability coefficient for river engineering projects.

.5. Six important references

1. Adamowski, J., H. F. Chan, S. O. Prasher, V.N. Sharda. 2012. Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *Journal of Hydroinformatics* 14(3): 731-744.
2. Etemad-Shahidi, A., J. Mahjoobi. 2009. Comparison between M5' model tree and neural networks for prediction of significant wave height. *Journal of Ocean Engineering*, 36(15-16): 1175-1181.
3. Rajaei, T., h. Jafari. 2020. Two decades on the artificial intelligence models 2advancement for modeling river sediment concentration: 3State-of-the-art. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2020.125011>.
4. Wang, Y., I. H. Witten, 1997. Induction of Model Trees for predicting Continuous Classes. *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague.
5. Zahiri1, Z., H. Nezaratian. 2020. Estimation of transverse mixing coefficient in streams using M5, MARS, GA, and PSO approaches. Springer-Verlag GmbH Germany, part of Springer Nature 2020.
6. Zhang, W., A.T.C. Goh. 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front* 7:45–52. <https://doi.org/10.1016/j.gsf.2014.10.003>

Conflict of Interest

Authors declared no conflict of interest.