



ISSN 2251-7480

ارزیابی عملکرد روش‌های مدل درختی M5 و رگرسیون بردار پشتیبان در مدل‌سازی رسوب معلق رودخانه

محمدتقی ستاری^۱، علی رضازاده جودی^{۲*}، فروغ صفدری^۳ و فراز قهرمانزاده^۴

۱) استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

۲*) باشگاه پژوهشگران جوان و نخبگان، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران

*نویسنده مسئول مکاتبات: alijoudi66@gmail.com

۳) دانش آموخته کارشناسی ارشد منابع آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

۴) دانش آموخته کارشناسی ارشد عمران - آب، دانشگاه آزاد اسلامی واحد اهر، اهر، ایران

تاریخ دریافت: ۱۳۹۴/۱۲/۱۶ تاریخ پذیرش: ۱۳۹۵/۰۳/۰۸

چکیده

همواره پدیده انتقال رسوب، بسیاری از سازه‌های رودخانه‌ای و سازه‌های عمرانی را تحت تأثیر قرار داده و عدم اطلاع از میزان دقیق آن خسارات بسیاری را موجب می‌شود. از این جهت برآورد صحیح بار رسوبی در رودخانه‌ها از نقطه نظر رسوب، فرسایش و کنترل سیلاب بسیار حایز اهمیت است. در این تحقیق، از دو روش نوین داده‌کاوی شامل مدل درختی M5 و رگرسیون بردار پشتیبان برای برآورد بار رسوبی معلق رودخانه اهرچای در مقایسه با روش کلاسیک منحنی‌سنجه رسوب استفاده گردید. جهت ارزیابی عملکرد روش‌های استفاده شده از سه آماره شامل ضریب همبستگی، جذر میانگین مربعات خطا و میانگین خطای مطلق استفاده گردید. با انجام تحلیل حساسیت مدل‌ها به متغیر ورودی مشخص گردید، متغیر دبی جریان در ماه جاری دارای بیش‌ترین تأثیر بر روی میزان بار رسوبی معلق می‌باشد. در حالت کلی نتایج بدست آمده بیانگر دقت بسیار بالای روش‌های داده‌کاوی نسبت به منحنی‌سنجه رسوب می‌باشد. اگرچه هر دو روش داده‌کاوی بررسی شده دقت بیش‌تر و خطای کم‌تری نسبت به روش متداول منحنی‌سنجه رسوب داشته‌اند، اما با توجه به روابط خطی ساده و قابل فهم ارائه شده توسط مدل درختی M5، کاربرد این روش کارآمد در موارد مشابه توصیه می‌گردد.

کلید واژه‌ها: اهرچای؛ تخمین بار رسوبی؛ داده‌کاوی؛ مدل درختی M5؛ رگرسیون ماشین بردار پشتیبان

مقدمه

رسوب گذاری در کف کانال و بسیاری مسایل و مشکلات دیگر را در بردارد. طراحی بهینه و عملکرد مناسب سازه‌های آبی، نظیر مخازن، سدها و کانال‌ها، نیازمند تخمین دقیق از بار رسوبی رودخانه می‌باشد. با توجه به اینکه رژیم جریان و به تبع آن رژیم رسوب در حوضه‌های آبریز ثابت نیست، لذا همچنانکه با پیش‌بینی جریان می‌توان نسبت به مدیریت بهتر تقاضا در ماه‌های آتی اقدام نمود، پیش‌بینی دبی رسوب نیز می‌تواند کمک شایانی در پیش‌بینی رسوب تجمع یافته پشت سازه‌های آبی و به ویژه سدها و تعیین حجم مرده مخازن در ماه‌های آتی نموده و

در ایران اطلاع دقیقی از میزان فرسایش، انتقال رسوب و رسوب‌گذاری وجود نداشته و در بسیاری از موارد، بین اندازه‌گیری‌ها و برآوردهای انجام شده نیز اختلاف زیادی مشاهده می‌شود. این در حالیست که رودخانه‌ها همواره با پدیده فرسایش و انتقال رسوب مواجه می‌باشند. انتقال رسوب و رسوب‌گذاری، پی‌آمدهایی چون، کاهش ظرفیت انتقال جریان‌های سیلابی، کاهش عمر مفید سدها و ظرفیت ذخیره مخازن، خوردگی تاسیسات سازه‌های رودخانه‌ای و وارد شدن خسارات به ابنیه آبی و مزارع،

منحنی‌های سنجه مختلف در حوضه آبریز گرگان‌رود را با یکدیگر مقایسه کردند. این پژوهش نشان داد که شبکه‌های پرسپترون چندلایه با ورودی‌های دبی همان روز و دبی یک و دو روز قبل، نسبت به سایر روش‌ها از دقت بالاتری در تخمین میزان بار رسوب معلق برخوردار است. رجیبی و همکاران (۱۳۹۴) بار معلق رودخانه گیوی‌چای را با استفاده از مدل‌های تیرید تدریجی عصبی^۳، انفیس و شبکه عصبی مصنوعی تخمین زدند. نتایج بیانگر توانایی بالاتر روش تیرید تدریجی عصبی بود. *Aytek* و *Kishi* (2008) در مطالعه‌ای برنامه‌ریزی ژنتیک را به عنوان روشی جدید برای فرموله کردن رابطه جریان و بار رسوب معلق روزانه در رودخانه تنگو واقع در مونتانا به کار گرفته و با منحنی سنجه و رگرسیون چند متغیره خطی مقایسه کردند. نتایج نشان داد که برنامه‌ریزی ژنتیک در مقایسه با دو روش دیگر عملکرد بهتری دارد. *Nourani* (2009) بار رسوبات معلقی که در دهانه رودخانه تلخه‌رود تخلیه می‌شود را، با استفاده از شبکه عصبی مصنوعی مدلسازی نمود. در مدل شبکه عصبی مصنوعی سری زمانی داده‌های جریان رودخانه در زمان حال و گذشته به عنوان ورودی و بار رسوبی به عنوان خروجی مدل در نظر گرفته شد. نتایج بدست آمده از روش شبکه عصبی مصنوعی با نتایج دو روش رگرسیون خطی و روش منحنی سنجه مورد مقایسه قرار گرفت. براساس یافته‌ها روش شبکه عصبی مصنوعی بعنوان کارآمدترین و بهترین روش پیشنهادی انتخاب و مورد استفاده قرار گرفت. میزان بار رسوبی معلق روزانه با استفاده از مدل ترکیبی موجک و عصبی-فازی تخمین زده شد و توانایی قابل قبول این روش در مقابل منحنی سنجه مشاهده گردید (*Rajaei et al.*, 2010). در مطالعه‌ای بار معلق رسوب رودخانه ال^۴ واقع در کالیفرنیا با استفاده از داده‌های بارش روزانه، جریان و غلظت رسوب و توسط روش‌های برنامه‌ریزی بیان ژن^۵، شبکه عصبی و سیستم

با اتخاذ تمهیدات به موقع، مدیریت تخلیه به‌هنگام رسوب را تاحدی تسهیل نماید. با توجه به اهمیت موضوع همواره تحقیقات زیادی در مورد تخمین و مدل‌سازی بار رسوبی رودخانه‌ها صورت پذیرفته و روش‌ها و روابط گوناگونی معرفی گردیده است. بار معلق رسوبی رودخانه‌ها معمولاً از طریق اندازه‌گیری مستقیم یا معادلات تجربی تعیین می‌شود. گرچه اندازه‌گیری مستقیم از درجه اعتبار زیادی برخوردار است، ولی بسیار هزینه‌بر بوده و نمی‌تواند برای سطح وسیعی از رودخانه مورد استفاده قرار گیرد. با این حال، اکثر معادلات برآورد کننده بار معلق رسوبی، نیازمند داده‌های دبی و رسوب حاصل از اندازه‌گیری‌ها می‌باشند (*Ozturk et al.*, 2001). یکی از روش‌های متداول محاسبه بار معلق رودخانه استفاده از منحنی سنجه رسوب^۱ است. این رابطه به دلیل داشتن اریب^۲ در بیشتر مواقع نتوانسته غلظت رسوب در دبی‌های مختلف جریان را به خوبی نشان دهد، اریب در واقع باعث شده تا مقدار باقی‌مانده‌ها (اختلاف بین مقادیر مشاهده‌ای و محاسباتی) توزیع نرمالی نداشته باشد (*Kao et al.*, 2005). در این میان می‌توان از روش‌های نوین داده‌کاوی که مبتنی بر روابط حاکم در میان داده‌ها بوده و بدنبال کشف الگوهای پنهان بین داده‌ها هستند و از قدرت انعطاف‌پذیری بالایی نیز برخوردار می‌باشند، استفاده نمود.

حسین زاده دلیر و همکاران (۱۳۸۸) رسوب معلق رودخانه اهرچای را توسط روش شبکه عصبی مصنوعی پیش‌بینی کردند. نتایج نشان داد داده‌های سطح آب نسبت به دبی جریان از دقت کمتری در پیش‌بینی بار معلق برخوردار می‌باشد. داننده مهر و همکاران (۱۳۸۹) بار معلق رودخانه ليقوان چای را با استفاده از برنامه‌ریزی ژنتیک پیش‌بینی کردند. نتایج نشان داد که برنامه‌ریزی ژنتیک در مقایسه با روش شبکه عصبی مصنوعی در پیش‌بینی بار معلق عملکرد بهتری دارد. دهقانی و وفاخواه (۱۳۹۲) در مطالعه‌ای نتایج حاصل از دو نوع شبکه عصبی مصنوعی و

^۳- Neural differential evolution

^۴- Eel

^۵- Gene Expression Programming

^۱- Suspended Rating Curve

^۲- Bias

طبق بررسی مطالعات پیشین مطالعه‌ای در زمینه کاربرد مدل درختی M5 و رگرسیون بردار پشتیبان در ایران و بخصوص حوضه آبریز اهرچای صورت پذیرفته است. از این‌روی، هدف از تحقیق حاضر امکان‌سنجی کاربرد مدل درختی M5 و روش رگرسیون بردار پشتیبان جهت پیش‌بینی بار معلق رودخانه اهرچای در ایستگاه اورنگ در مقایسه با روش متداول منحنی سنجه رسوب و همچنین شناسایی تاثیرگذارترین متغیر در این زمینه می‌باشد.

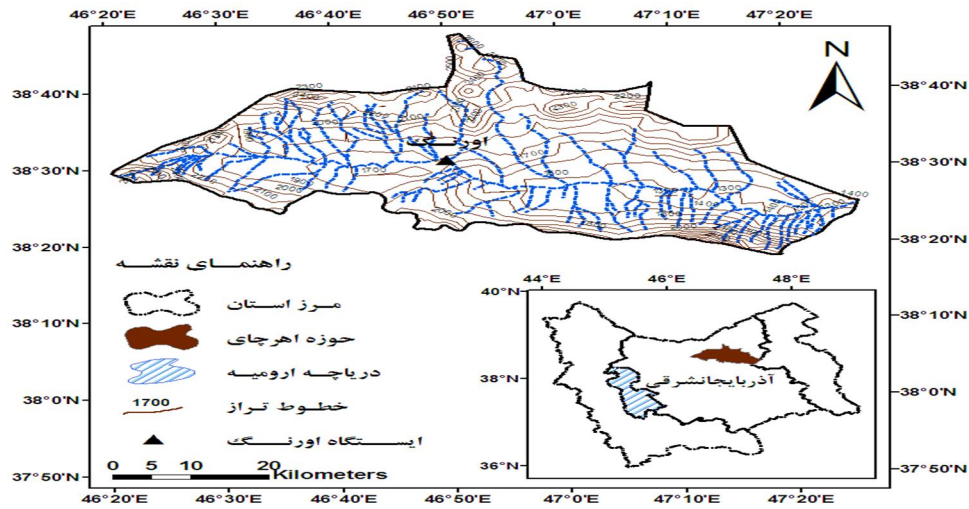
مواد و روش‌ها

منطقه مورد مطالعه و داده‌های مورد استفاده

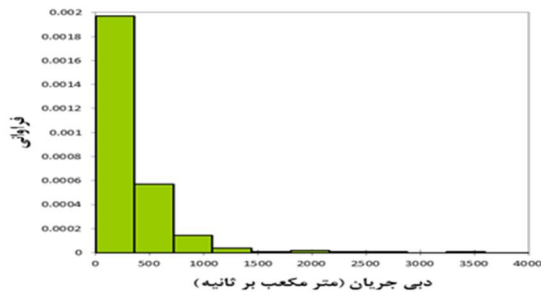
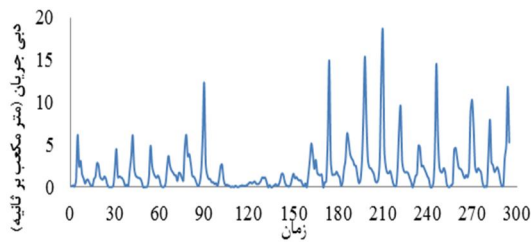
حوضه آبریز اهرچای در شمال شرقی شهرستان تبریز با مساحتی بالغ بر ۲۴۲۶/۵ کیلومتر مربع یکی از مهمترین حوضه‌های آبریز منطقه می‌باشد. این حوضه بین مختصات جغرافیایی "۲۴'۱۹'۴۶" تا "۴۴' ۵۶" عرض شمالی در شمال شرقی زیر حوضه ارس گسترش یافته است. شکل ۱ موقعیت ایستگاه اورنگ واقع در حوضه آبریز اهرچای را نشان می‌دهد. در این تحقیق برای برآورد بار رسوبی معلق رودخانه اهرچای از داده‌های ماهانه دبی رسوب (تن در روز)، بارش (میلی‌متر) و دبی جریان (متر مکعب بر ثانیه) مربوط به ایستگاه اورنگ به علت وقوع شدن آن قبل از ورودی سد ستارخان استفاده گردید. محدوده تغییرات و مشخصات آماری پارامترهای دبی جریان، دبی رسوب و بارش ماهانه در جدول (۱) ارائه شده‌اند. طول آماری داده‌های استفاده شده در این تحقیق مربوط به سال‌های ۱۳۶۴ تا ۱۳۹۳، برابر با ۳۲۲ مجموعه داده می‌باشد. از این تعداد داده ۶۶ درصد آن (ابتدای مجموعه داده)، برابر با ۲۱۳ داده برای قسمت آموزش مدل و ۳۴ درصد آن (بین ماه مرداد ۱۳۸۶ تا ماه شهریور ۱۳۹۳)، برابر با ۱۰۹ مجموعه داده برای قسمت آزمون مدل در نظر گرفته شد. سری زمانی و فراوانی بارش، رسوب و جریان ماهانه رودخانه اهرچای در شکل‌های (۲) تا (۴) ارائه شده است.

استنتاجی فازی-عصبی تطبیقی پیش‌بینی گردید. نتایج نشان داد روش برنامه‌ریزی بیان ژن در پیش‌بینی بار رسوب معلق روزانه عملکرد بهتری نسبت به دو روش دیگر دارد (Kisi and Shiri, 2012). Kakaei Lafdani و همکاران (2013) با استفاده از دو مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان به پیش‌بینی بار معلق رسوب روزانه رودخانه پرداختند و نتایج را با مدل‌های رگرسیونی مقایسه نمودند. نتایج نشان داد که مدل‌های رگرسیونی عملکرد ضعیف‌تری نسبت به دو مدل دیگر در پیش‌بینی رسوب دارند. Ghorbani و همکاران (2013) بار رسوبی را توسط روش‌های رگرسیون چند متغیره، شبکه عصبی مصنوعی و مدل عصبی-فازی تخمین زدند. نتایج حاکی از برتری شبکه عصبی مصنوعی در این زمینه بود. Kia و همکاران (2013) کارایی شبکه عصبی مصنوعی در مقایسه با منحنی سنجه رسوب را در تخمین و مدل‌سازی بار رسوبی بررسی نمودند. نتایج نشان‌دهنده برتری شبکه عصبی مصنوعی با استفاده از داده‌های دبی در ماه جاری بود. (Heng and Suetsugi, 2013) کارایی شبکه عصبی مصنوعی در تخمین بار رسوبی حوضه آبخیز رودخانه تونله ساپ در کامبوج را مورد ارزیابی قرار دادند. نتایج بیانگر عملکرد قابل قبول شبکه عصبی مصنوعی در این زمینه می‌باشد. (Kumar Goyal, 2014) عملکرد مدل درختی M5 و رگرسیون موجکی را در مقایسه با شبکه عصبی مصنوعی برای پیش‌بینی آورد رسوبی در حوضه آبریز ناگوا واقع در هند، مورد بررسی قرار داد. نتایج نشان‌دهنده برتری مدل درختی M5 و رگرسیون موجکی می‌باشد.

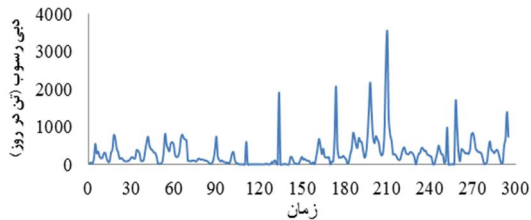
طبق بررسی مطالعات پیشین صورت گرفته در زمینه تخمین و مدل‌سازی بار رسوبی رودخانه‌ها، اکثر مطالعات مربوط به کاربرد روش‌های شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن در این زمینه بوده و مطالعات بسیار اندکی در زمینه کاربرد روش‌هایی مانند مدل درختی M5 و رگرسیون بردار پشتیبان صورت پذیرفته است. همچنین



شکل ۱. موقعیت جغرافیایی حوضه آبخیز اهرچای و ایستگاه اورنگ.

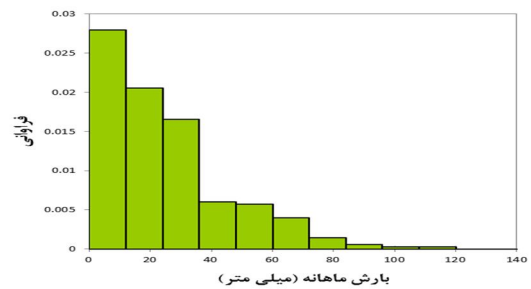
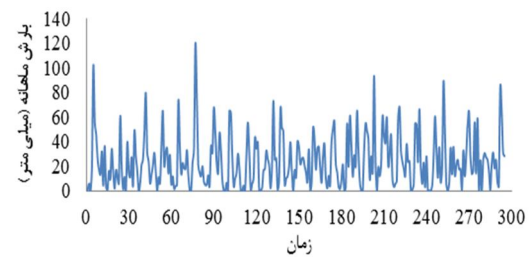


شکل ۳. نمودار سری زمانی و فراوانی دبی جریان



جدول ۱. محدوده تغییرات و مشخصات آماری داده های ماهانه رودخانه اهرچای (۹۳-۱۳۶۴)

پارامتر	دبی جریان (مترمکعب در ثانیه)	دبی رسوب (تن در (میلی متر))	دبی بارش (میلی متر)
حدافل	۰/۰۰۴	۰/۰۰۲	۰
حداکثر	۲۹/۹۰۶	۸۶۱۲/۲۴۸	۱۱۹/۶
میانگین	۳/۵۱۹	۴۹۵/۲۸۲	۲۳/۹۴۳
انحراف معیار	۴/۰۹	۱۱۲۷/۱۱	۲۰/۹۷۳

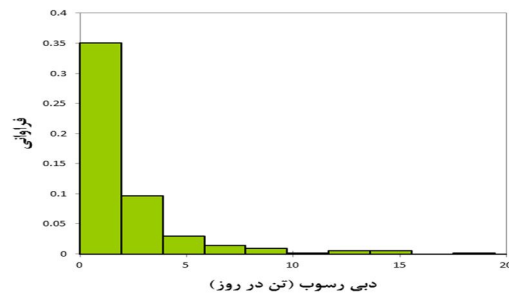


شکل ۲. نمودار سری زمانی و فراوانی بارش ماهانه

رسوبی حمل شده توسط آب اندازه‌گیری شده و منحنی مربوطه رسم می‌گردد (دهقانی و وفاخواه، ۱۳۹۲).

مدل درختی M5^۲

مدل درختی M5 یکی از روش‌های داده‌کاوی می‌باشد که اخیراً بسیار مورد توجه قرار گرفته است و در مدل‌سازی مسایل مختلف به کار می‌رود. ایده کلی این مدل برگرفته از درختان رگرسیونی است با این تفاوت که به جای مقادیر ثابت و برچسب‌های طبقه‌بندی، در برگ‌های خود دارای توابع خطی می‌باشد. مزیت عمده مدل درختی M5 نسبت به درختان رگرسیونی این است که مدل درختی M5 بسیار کوچک‌تر از درختان رگرسیون بوده و توابع رگرسیون به صورت طبیعی شامل بسیاری از متغیرها نمی‌شوند. یک درخت تصمیم معمولاً از چهار بخش ریشه^۳، شاخه^۴، گره‌ها^۵ و برگ‌ها^۶ تشکیل شده است. هر گره مربوط به یک خصوصیت معین است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند، این بازه‌ها مقادیر بخش‌های مختلف مجموعه مقادیر معلوم را برای خصوصیت‌ها به دست می‌دهند. عمل انشعاب توسط یکی از متغیرهای پیش‌بینی کننده انجام می‌پذیرد. بازه‌های انشعاب طوری انتخاب می‌شوند که مجموع مجذور انحراف از میانگین داده‌های هر گره را حداقل کنند. معیار انشعاب بیانگر میزان خطا در آن گره می‌باشد و مدل حداقل خطای مورد انتظار را به عنوان نتیجه آزمایش هر صفت در آن گره محاسبه می‌کند. خطای مدل عموماً با اندازه‌گیری دقت پیش‌بینی مقادیر هدف موارد دیده نشده سنجش می‌شود. فرآیند انشعاب در هر گره بارها تکرار می‌شود تا به گره پایانی (برگ) برسد که در برگ، مجموع مجذور انحراف از میانگین داده‌ها حدوداً به صفر می‌رسد. با این کار درخت بزرگی توسعه پیدا خواهد کرد (Alberg et al., 2012). کار با این درخت بزرگ که شاخه‌ها و گره‌های زیادی دارد



شکل ۴: نمودار سری زمانی و فراوانی دبی رسوب

همچنانکه از نمودارهای فراوانی در شکل‌های (۲) تا (۴) ملاحظه می‌شود، مقادیر بارش ماهانه، دبی جریان و دبی رسوب رودخانه اهرچای از توزیع نرمال تبعیت نمی‌کنند. لازم به ذکر است که در روش‌های داده‌کاوی لزومی به تبعیت پارامترهای بکار رفته در مدل‌سازی از توزیع نرمال نبوده و از این‌روی مشکلی در فرآیند مدل‌سازی ایجاد نمی‌کند. به عنوان یک قانون کلی در تکنیک‌های آماری فرض اولیه بر این اساس است که توزیع داده‌ها مشخص است که بیشتر موارد فرض بر این است که توزیع نرمال است و در نهایت درستی یا نادرستی نتایج نهایی به درست بودن فرض اولیه وابسته است. در مقابل روش‌های داده‌کاوی از هیچ فرض در مورد داده‌ها استفاده نمی‌کنند (مهریزی حائری، ۱۳۸۲).

روش منحنی سنجه رسوب یک خطی یا روش (USBR)

به دلیل در دسترس نبودن آمار دقیق میزان فرسایش و رسوب در حوضه‌های آبریز در بیشتر مواقع از منحنی سنجه رسوب، تهیه شده با داده‌های دبی و غلظت رسوب یا دبی رسوب معلق، استفاده می‌شود. شکل کلی منحنی سنجه رسوب به صورت رابطه (۱) می‌باشد:

$$Q_s = a Q_w^b \quad (1)$$

که در این رابطه، Q_w برابر دبی آبراهه (m^3/s)، Q_s برابر دبی وزنی مواد معلق (تن بر روز) و a و b ضرایب ثابت هستند. منحنی‌های سنجه رسوب براساس اندازه‌گیری‌های انجام شده در رودخانه‌ها به دست می‌آیند. به این صورت که متناظر با هر دبی خاص در رودخانه، غلظت مواد

^۲- M5 Model tree

^۳- Root

^۴- Beach

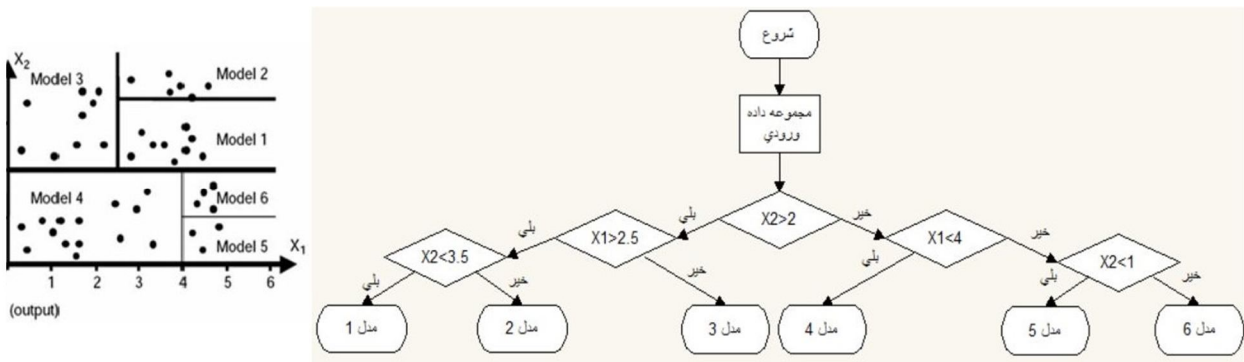
^۵- Nodes

^۶- Leafs

^۱- United States Bureau of Reclamation

از شکل‌گیری درخت حداکثر (۲) هرس بعد از شکل‌گیری درخت حداکثر.

سخت می‌باشد، بنابراین برای رسیدن به یک درخت بهینه و کارآمد باید شاخه‌های اضافی درخت هرس شود. دو روش برای هرس کردن درخت وجود دارد: (۱) هرس قبل



شکل ۵. نمایی شمانیک از نحوه تقسیم فضای ورودی توسط مدل درختی *M5* و تولید مدل.

انحراف می‌باشد (Vapnik, 1995). در برخی مواقع، به ساختارهای پیچیده و غیر خطی برای جداسازی داده‌ها نیاز است. در این صورت، ماشین بردار پشتیبان، داده‌های اصلی را با به کارگیری مجموعه‌ای از توابع ریاضی که کرنل نام دارند در فضای جدیدی نگاشت و بازآرایی می‌کند که به این کار تبدیل (نگاشت) گفته می‌شود (شهرابی و حجازی، ۱۳۹۱). اگر داده‌های آموزش به صورت k سری به صورت $(x_1, y_1), \dots, (x_k, y_k)$ باشند، تابع خطی‌سازی مربوطه به صورت رابطه (۲) قابل بیان است:

$$f(x) = (w, x_1) + b \quad (2)$$

$$\text{with } (w, x) \in R^N, b \in R$$

که در این رابطه، x بردار ورودی، w وزن بردار و b میزان اختلال می‌باشد. تابع $f(x)$ همواره می‌بایست به گونه‌ای تعیین گردد که به طور همزمان میزان کمترین انحراف (ϵ) تعیین شده و همچنین تابع، مقدار مناسب w را اختیار نماید. ایجاد چنین شرایطی با حل معادله بهینه‌سازی زیر قابل دستیابی است:

$$\begin{aligned} & \text{Minimise } \frac{1}{2} \|W\| \text{ Subject to} \\ & \begin{cases} y_i - (w, x_i) - b \leq \epsilon \\ (w, x_i) + b - y_i \leq \epsilon \end{cases} \end{aligned} \quad (3)$$

در روش اول فرآیند هرس اجازه نمی‌دهد شاخه‌های اضافی تولید شوند، ولی در روش دوم ابتدا درخت حداکثر تشکیل می‌شود، سپس فرآیند هرس انجام می‌گیرد. برای اطلاعات بیشتر به (Quinlan, 1992) مراجعه شود. در شکل (۵) شمانیکی از نحوه تقسیم فضای ورودی داده‌ها توسط مدل درختی *M5* ارائه شده است. در این تحقیق جهت مدل‌سازی روش *M5* از نرم‌افزار WEKA^۱ که در دانشگاه وایکاتو نیوزیلند توسعه داده شده، استفاده گردیده است.

رگرسیون ماشین بردار پشتیبان

ماشین‌های بردار پشتیبان^۲ شامل دو دسته می‌باشند: طبقه‌بندی کننده بردار پشتیبان و رگرسیون بردار پشتیبان. ماشین‌های بردار پشتیبان بر پایه مفهوم صفحات تصمیم هستند که مرز تصمیم را تعریف می‌کنند. یک صفحه تصمیم، داده‌های با برجسب‌های مختلف را از هم تفکیک می‌کند (شهرابی و حجازی، ۱۳۹۱). در یک الگوریتم خطی‌سازی به کمک ماشین بردار پشتیبان با فرض مقادیر ورودی x_i و مقادیر خروجی y_i هدف یافتن تابعی است که کمترین انحراف (ϵ) را از y_i ها داشته باشد، ϵ میزان

^۱- Waikato Environment for Knowledge Analysis

^۲- Support vector machines

آزمون شاپیرو-ویلک^۲

آزمون شاپیرو-ویلک، یک آزمون نرمال بودن است که توسط ساموئل سنفورد شاپیرو و مارتین ویلک، در سال ۱۹۶۵ ارائه شده است. آزمون شاپیرو-ویلک، برای بررسی اینکه یک نمونه از مجموعه داده x_1, \dots, x_n دارای توزیع طبیعی نرمال است یا نه از اصل فرض صفر استفاده می‌کند. آماره این آزمون از رابطه (۴) محاسبه می‌گردد.

$$W = \frac{(\sum_{i=1}^n a_x x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

که در رابطه (۴) $x_{(i)}$ i امین آماره ترتیبی است، مثلاً n امین عدد کوچک در نمونه، \bar{x} میانگین نمونه و ثابت‌های a_i از رابطه (۵) به دست می‌آیند.

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}} \quad (5)$$

که در رابطه (۵) مقدار m از رابطه (۶) قابل محاسبه است.

$$m = (m_1, \dots, m_n)^T \quad (6)$$

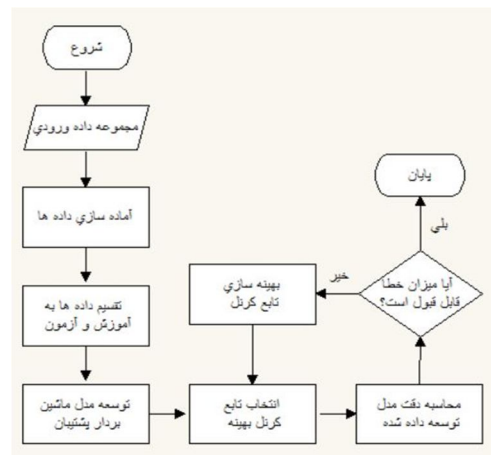
m_1, \dots, m_n امیدهای ریاضی متغیرهای تصادفی مستقل با توزیع یکسان هستند که از توزیع طبیعی استاندارد انتخاب شده‌اند و V ماتریس کوواریانس این آماره‌های ترتیبی است. اگر W کمتر از سطح اطمینان (۰/۰۵) باشد، فرض صفر با احتمال زیادی رد می‌شود (Shapiro & Wilk, 1965).

آزمون اندرسون دارلینگ^۳

آزمون اندرسون دارلینگ یک آزمون آماری قدرتمند جهت بررسی تبعیت یک نمونه مشخص از یک مجموعه داده از توزیع نرمال می‌باشد. این آماره از طریق رابطه (۷) قابل محاسبه است.

که در این رابطه، w, b, x_i و y_i همان پارامترهای معادله ۱ بوده و ϵ نیز میزان انحراف اعمال شده می‌باشد. دو مشکل عمده در این روش وجود دارد: اول این که آموزش در آن به کندی صورت می‌گیرد و دوم اینکه الگوریتم آموزش در آن پیچیده بوده و اغلب حل آن مشکل است (شهرابی و حجازی، ۱۳۹۱). بهینه‌سازی حداقل متوالی^۱، الگوریتم جدیدی است که برای آموزش این مدل به کار می‌رود. این الگوریتم، به جای استفاده از تکرارهای عددی برای حل مسائل درجه دو از گام‌های تحلیلی برای حل این گونه مسائل استفاده می‌کند.

این الگوریتم به ویژه برای داده‌های پراکنده چه از نوع باینری و چه از نوع غیرباینری مفید است. برای حل مسائل درجه دوم در ماشین بردار پشتیبان حداقل دو ضریب لاگرانژ وجود دارد، زیرا ضرایب لاگرانژ باید از یک نامساوی غیر خطی تبعیت کنند. این الگوریتم در هر گام دو ضریب لاگرانژ را به طور مشترک جهت بهینه‌سازی انتخاب می‌کند و مقادیر بهینه را برای این ضرایب پیدا کرده ماشین بردار پشتیبان را برای منعکس نمودن مقادیر جدید بهینه به‌روز رسانی می‌کند (Plat, 2000). در این تحقیق برای اولین بار از الگوریتم بهینه‌سازی حداقل متوالی، جهت آموزش رگرسیون بردار پشتیبان استفاده گردید. در شکل ۶ نمودار گردش ماشین بردار پشتیبان ارائه شده است.



شکل ۶. نمودار گردش ماشین بردار پشتیبان.

^۲- Shapiro-Wilk test

^۳- Anderson-Darling test

^۱- Sequential minimal optimization

میانگین خطای مطلق مورد ارزیابی قرار گرفت. روابط محاسبه آماره‌های فوق در روابط (۱۰) تا (۱۲) ارائه گردیده است.

$$cc = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=0}^n |x_i - y_i| \quad (12)$$

نتایج و بحث

با توجه به اهمیت استفاده از داده‌های صحیح آماری، ابتدا کلیه داده‌های موجود از نظر همگن بودن توسط آزمون همگنی نرمال استاندارد^۱ که یکی از روش‌های متداول برای ارزیابی همگنی داده‌ها می‌باشد، مورد بررسی قرار گرفت. نتایج مربوط به آزمون همگنی نرمال استاندارد در جدول (۲) ارائه شده است.

جدول ۲. نتایج آزمون همگنی برای داده‌های استفاده شده.

متغیر	P-Value	ریسک رد کردن فرض صفر (%)	سطح اطمینان (a)
بارش	۰/۶۰۲	۶۰/۲	۰/۰۵
دبی جریان	۰/۵۷۵	۵۷/۴۹	۰/۰۵
دبی رسوب	۰/۲۸۹	۲۸/۹۲	۰/۰۵

در این آزمون فرض صفر، همگنی داده‌ها و فرض یک، ناهمگن بودن داده‌ها می‌باشد. چنان‌که مقدار P -Value از مقدار درجه اطمینان بزرگ‌تر باشد، فرض صفر صحیح است، در غیر این صورت فرض یک قابل قبول می‌باشد. نتایج جدول ۲ نشان‌دهنده این مطلب است که داده‌های ماهانه بارش، دبی جریان و دبی رسوب در ایستگاه و طول آماری مورد مطالعه همگن بوده و می‌توانند با اطمینان به کار برده شوند. یکی از مباحثی که در مطالعات هیدرولوژیکی مطرح می‌باشد مسئله کفایت یا عدم کفایت طول آماری داده‌ها برای انجام مطالعات و مدل‌سازی می‌باشد. روش

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) (\ln \Phi(Y_i)) \quad (7)$$

که در رابطه ۷ n تعداد داده‌ها، Φ ضریب تابع توزیع تجمعی داده‌ها و Y_i مقادیر استاندارد شده داده‌ها با استفاده از رابطه ۸ می‌باشد.

$$Y_i = \frac{X_i - \mu}{\sigma} \quad (8)$$

که در رابطه (۸) X داده‌های مشاهداتی، μ میانگین داده‌های مشاهداتی و σ واریانس داده‌های مشاهداتی می‌باشند. در این آزمون نیز اگر آماره به دست آمده از مقدار مشخص سطح اطمینان (۰/۰۵) کمتر باشد، فرض صفر با احتمال زیادی رد می‌شود (Stephens, 1979).

آزمون لیلیه فورس^۱

این آزمون بر اساس آزمون کولموگروف-اسمیرنوف^۲ برای تعیین نرمال بودن یک مجموعه داده توسط هوبرت لیلیه فورس^۳ ارائه شده است. در این آزمون فرض صفر نرمال بودن داده‌ها می‌باشد. اگر مقدار به دست آمده توسط این آزمون بیشتر از سطح اطمینان (۰/۰۵) باشد، بدین معنی است که فرض صفر درست بوده و در صورت کوچکتر بودن این آماره از سطح اطمینان، داده‌ها با احتمال بسیار زیادی از توزیع نرمال تبعیت نمی‌کنند. این آماره برای مجموعه داده x_1, \dots, x_n از رابطه (۹) محاسبه می‌گردد.

$$T = \max |N^*(X) - F_N(X)| \quad (9)$$

که در رابطه (۹) $N^*(X)$ برابر با تابع توزیع تجمعی نرمال مجوعه داده و $F_N(X)$ تابع توزیع تجمعی مجوعه داده می‌باشد (Graziose, 2013).

معیارهای ارزیابی

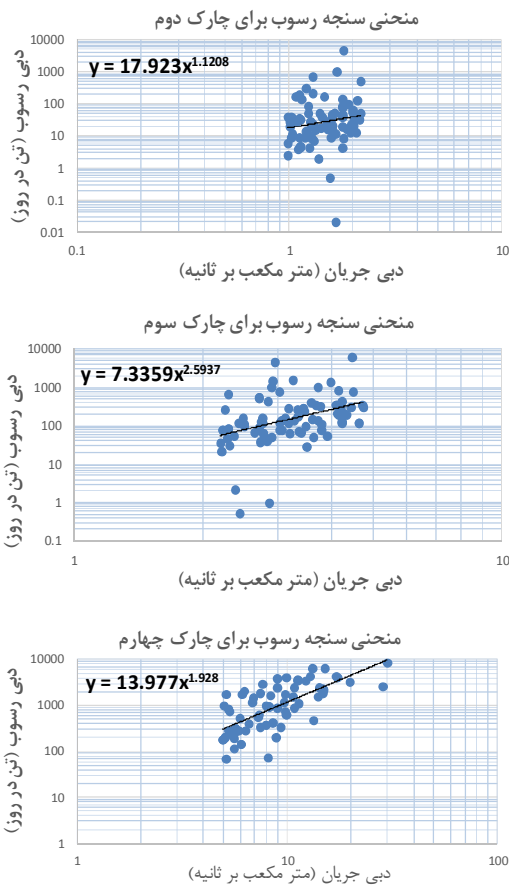
در این تحقیق عملکرد مدل درختی قوانین $M5$ ، رگرسیون بردار پشتیبان و منحنی سنجه رسوب بر پایه سه آماره ضریب همبستگی، جذر میانگین مربعات خطا و

¹ - Lilliefors test

² - Kolmogorov-Smirnov test

³ - Hubert Lilliefors

⁴ - Standard Normal Homogeneity Test



شکل ۷. منحنی‌های سنج‌ رسوب به دست آمده برای رودخانه اهر جای.

معادلات به دست آمده از منحنی سنج رسوب به صورت روابط (۷) تا (۱۰) می‌باشند. برای مقایسه کارایی روش منحنی سنج رسوب با سایر روش‌های مورد بررسی، داده‌های مربوط به بخش آزمون مدل‌ها (۱۰۹ داده در بخش انتهایی مجموعه داده‌ها) مورد استفاده قرار گرفتند و در روابط (۱۳) تا (۱۶) بسته به مقدار دبی جای‌گذاری گردیدند تا مقادیر رسوب مربوطه برآورد گردد. پس از محاسبه رسوب توسط منحنی‌های به دست آمده، مقادیر محاسباتی شده با مقادیر مشاهده‌ای توسط شاخص‌های آماری مطرح شده مورد مقایسه قرار گرفتند.

$$Q_s = 13/206 Q_w^{1/0.248} \quad (13)$$

$$Q_s = 17/923 Q_w^{1/1.208} \quad (14)$$

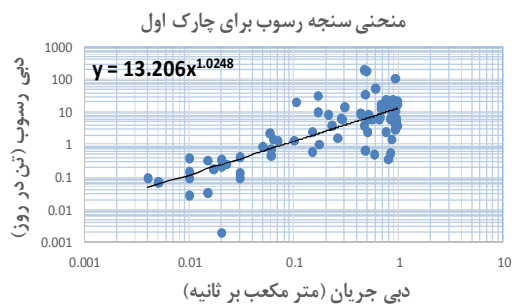
$$Q_s = 7/3359 Q_w^{2/0.937} \quad (15)$$

متداول در این زمینه محاسبه ضریب هرست^۱ است (صفوی، ۱۳۹۰). اگر این ضریب بیشتر از مقدار ۰/۵ باشد، طول آماری داده‌ها برای انجام مطالعات کافی و اگر کمتر باشد، بدین معنی است که قبل از انجام مطالعات باید طول آماری داده‌ها تطویل گردد. محاسبه و بررسی ضریب هرست نشان داد طول آماری داده‌ها با مقدار ضریب هرست برابر با ۰/۵۷۴ برای انجام مطالعات کافی و مناسب می‌باشد.

با توجه به اینکه مقایسه عملکرد روش‌های مدل درختی M5 و رگرسیون بردار پشتیبان با روش متداول منحنی سنج رسوب یکی از اهداف اصلی این مطالعه می‌باشد، ابتدا منحنی سنج رسوب رسم گردید. برای رسم منحنی سنج رسوب داده‌ها به چهار چارک توسط نرم‌افزار XLSTAT تقسیم‌بندی شده و سپس برای هر یک از چارک‌ها منحنی منحصر‌بفردی رسم گردید. کمینه و بیشینه مقدار هر یک از چارک‌ها در جدول (۳) ارائه شده است. همچنین در شکل (۷) منحنی‌های سنج رسوب به دست آمده از داده‌های موجود در بازه زمانی ۲۹ ساله مورد بررسی، ارائه گردیده است.

جدول ۳. بازه قرارگیری داده‌ها در هر چارک.

چارک	چارک دوم	چارک سوم	چارک چهارم	اول
کمینه	۰/۹۸۲	۲/۱۹	۴/۹۱	۰/۰۰۴
بیشینه	۲/۱۸	۴/۷۳	۲۹/۹۰۶	۰/۹۷۰
تعداد داده	۸۱	۸۵	۷۶	۸۰



^۱ - hurst coefficient

و همچنین اندیس t بیانگر ماه جاری و اندیس $t-n$ بیانگر n ماه قبل است.

همچنان که از نتایج ارائه شده در جدول (۴) قابل مشاهده می‌باشد، هر دو روش مدل درختی $M5$ و رگرسیون بردار پشتیبان نتایج بسیار قابل قبولی ارائه کرده‌اند، ولیکن بهترین نتیجه مربوط به ترکیب پارامترهای سناریو شماره پنج می‌باشد که نشان‌دهنده تاثیر بیشتر این متغیرها در میزان بار رسوبی معلق رودخانه اهرچای می‌باشد. در سناریوهای شماره شش و هفت به منظور پیش‌بینی و تخمین میزان بار رسوبی معلق اهرچای از داده‌های ماه جاری استفاده نگردید که این امر موجب کاهش دقت مدل در هر دو روش داده‌کاوی بررسی شده گردید. این امر نشان‌دهنده حساسیت بالای مدل‌های مذکور نسبت به متغیر دبی جریان در ماه جاری می‌باشد. در سناریو پنج، مدل درختی $M5$ برای مدل‌سازی بار رسوبی معلق در هنگام استفاده از ترکیب متغیر ورودی بهینه، با توسعه سه رابطه خطی ساده و قابل فهم اگر-آنگاه بار رسوبی معلق را مدل‌سازی نمود که این روابط در جدول (۵) ارائه گردیده‌اند.

$$Q_s = 13/977 Q_w^{1/928} \quad (16)$$

در این روابط Q_s برابر با دبی وزنی مواد رسوبی (تن در روز) و Q_w برابر با دبی جریان می‌باشد. برای مدل‌سازی و تخمین بار رسوب معلق توسط مدل درختی $M5$ و رگرسیون بردار پشتیبان، با توجه به مقادیر خودهمبستگی و همبستگی متقابل پارامترها در گام‌های زمانی مختلف، تعداد ۷ سناریو متفاوت شامل ترکیب‌های گوناگون از متغیرها در نظر گرفته شد. همچنین برای مدل‌سازی میزان بار رسوب معلق ماهانه توسط روش رگرسیون بردار پشتیبان، پس از آزمون توابع کرنل مختلف، مشخص شد که تابع چند جمله‌ای ساده از نوع درجه دو، بهترین عملکرد را در مدل‌سازی میزان بار رسوب معلق دارد. ترکیب‌های گوناگون بررسی شده در این تحقیق و نتایج مدل‌های $M5$ و رگرسیون بردار پشتیبان برای این سناریوها در مرحله آزمون (۲۰٪ داده‌ها برابر ۱۱۲ داده) در جدول شماره (۴) ارائه شده است. در این جدول پارامتر Q نشان‌دهنده دبی جریان (متر مکعب بر ثانیه)، Q_s نشان‌دهنده دبی رسوب (تن در روز) و P نشان‌دهنده بارش (میلی متر)

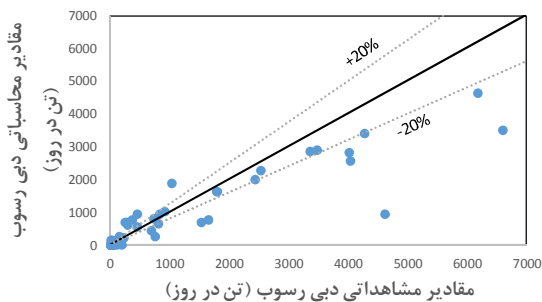
جدول ۴. نتایج صحت سنجی برای سناریوهای مختلف با روشهای مدل درختی $M5$ و رگرسیون بردار پشتیبان

سناریو	پارامترهای ورودی	رگرسیون بردار پشتیبان			مدل درختی $M5$		
		RMSE (تن در روز)	MAE (تن در روز)	CC (تن در روز)	RMSE (تن در روز)	MAE (تن در روز)	CC (تن در روز)
۱	$Q_b, Q_s(t-1), P_{t-1}$	۹۰/۱۶۳	۳۰/۱۶۳	۰/۸۳	۷۸۱/۴۱	۲۹۵/۱۶	۰/۸۹
۲	$Q_b, Q_s(t-1), P_b, P_{t-1}$	۸۹۲/۴۵	۲۹۸/۴۵	۰/۸۴	۷۶۵/۲۲	۲۵۰/۹۵	۰/۹۰
۳	$P_b, Q_b, Q_s(t-1)$	۸۰۹/۲۱	۳۰۹/۲۱	۰/۸۶	۷۷۵/۶۴	۲۵۱/۶۵	۰/۸۹
۴	$Q_b, Q_s(t-1)$	۸۰۰/۱۴	۳۰۰/۱۴	۰/۸۷	۷۸۲/۷۱	۲۵۲/۲۴	۰/۸۹
۵	$P_b, Q_b, Q_s(t-1), Q_{t-b}, P_{(t-1)}$	۷۶۷/۵۸	۲۸۳/۲۰	۰/۸۹	۵۵۸/۶۹	۱۹۷/۲۸	۰/۹۳
۶	$Q_s(t-1), Q_{t-1}, P_{(t-1)}, Q_{t-2}, P_{(t-2)}, Q_s(t-2)$	۹۵۰/۴۲	۴۷۱/۶۹	۰/۶۹	۸۴۹/۴۷	۴۶۹/۱۴	۰/۷۲
۷	$Q_{t-1}, Q_s(t-1), P_{t-1}$	۱۰۵۳/۳۳	۵۲۰/۱۸	۰/۶۵	۹۶۳/۶۳	۴۷۵/۸۲	۰/۷۰

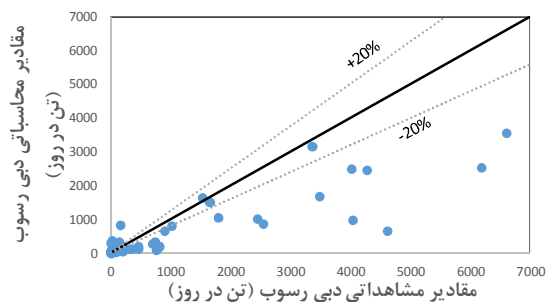
جدول ۵. قوانین اگر- آنگاه ارائه شده توسط مدل درختی $M5$ برای برترین ترکیب متغیر ورودی (سناریوی ۵).

قانون	اگر	آنگاه
۱	$Q_t > 2.656$	$Q_{St} = (427.2198 * Q_t) + (0.7328 * Q_{St-1}) - (298.2964 * Q_{t-1}) - (35.6134 * P_t) + 345.258$
۲	$Q_t \leq 1.12$	$Q_{St} = (18.2029 * Q_t) + (4.7638 * Q_{t-1}) - (0.95 * Q_{St-1}) + 4.769$
۳	$1.12 < Q_t < 2.656$	$Q_{St} = (0.104 * Q_{St-1}) - (211.1885 * Q_t) + 342.58994$

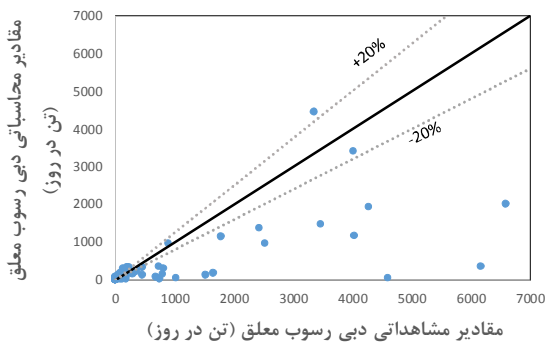
مورد استفاده در این پژوهش است. به طوری که تغییرات زیادی در مقادیر رسوب در طول دوره آماری مورد مطالعه وجود دارد که مقدار بالای انحراف معیار برای داده‌های رسوب (1127/11) تاییدکننده این مطلب است. برای درک بهتر نتایج مربوطه نمودارهای پراکنش داده‌های محاسباتی به دست آمده از روش‌های مختلف نسبت به داده‌های مشاهداتی موجود، در اشکال (8) تا (10) نشان داده شده است.



شکل 8. نمودار پراکنش مقادیر محاسباتی نسبت به مشاهداتی به دست آمده از مدل درختی M5 برای سناریوی برتر.



شکل 9. نمودار پراکنش مقادیر محاسباتی نسبت به مشاهداتی به دست آمده از رگرسیون بردار پشتیبان برای سناریوی برتر.



بر اساس روابط ارائه شده در جدول (5) مشاهده می‌شود که اگر دبی ماه جاری جریان بیشتر از 1/656 متر مکعب در ثانیه باشد، دبی رسوب مطابق جدول (5) از قانون یک و اگر دبی ماه جاری حداکثر برابر با 1/12 مترمکعب در ثانیه باشد، دبی رسوب در ماه جاری مطابق قانون دو به دست خواهد آمد. همچنین اگر دبی در ماه جاری بین مقادیر 1/12 و 1/656 متر مکعب بر ثانیه باشد، دبی رسوب در ماه جاری مطابق قانون سه به دست خواهد آمد.

روابط خطی ارائه شده در جدول (5) نشان می‌دهد کمترین ضریب مربوط به دبی رسوب در دوره قبل (t-1) و بعد از آن مربوط به بارش دوره زمانی جاری (t) می‌باشد. عبارت دیگر، در مدل توسعه داده شده به ترتیب دبی جریان در دوره های زمانی فعلی (t) و قبلی (t-1)، بارش دوره زمانی جاری (t) و دبی رسوب در دوره قبلی (t-1) بیشترین تاثیر و اهمیت را داشته‌اند. در جدول (6) نتایج بهینه به دست آمده از روش‌های داده‌کاوی شامل مدل درختی M5 و رگرسیون بردار پشتیبان برای بهترین ترکیب متغیر ورودی (سناریو پنج) و همچنین منحنی سنج رسوب در بازه آزمون داده‌ها (بین ماه مرداد 1386 تا ماه شهریور 1393) ارائه شده است.

جدول 6. نتایج بهینه به دست آمده از روش‌های مختلف.

روش مورد استفاده	RMSE	MAE	CC
	(تن در روز)	(تن در روز)	(تن در روز)
منحنی سنج رسوب	۹۹۰/۸۲	۳۴۲/۶۱	۰/۶۹
رگرسیون بردار پشتیبان	۷۶۷/۵۸	۲۸۳/۲۰	۰/۸۹
مدل درختی M5	۵۵۸/۶۹	۱۹۷/۲۸	۰/۹۳

همچنانکه مشاهده می‌شود هر دو روش داده‌کاوی بررسی شده به مراتب نتایج دقیق‌تر و با خطای پایین‌تری نسبت به منحنی سنج رسوب ارائه کرده‌اند که نشان‌دهنده کاربردی بودن این روش‌ها در این زمینه می‌باشد. بالا بودن نسبی مقادیر به دست آمده برای آماره‌های ریشه میانگین مربعات خطا و میانگین خطای مطلق به دلیل نوع داده‌های

شکل ۱۰. نمودار پراکنش مقادیر محاسباتی نسبت به مشاهداتی به دست آمده از منحنی سنجه رسوب برای داده‌های قسمت آزمون.

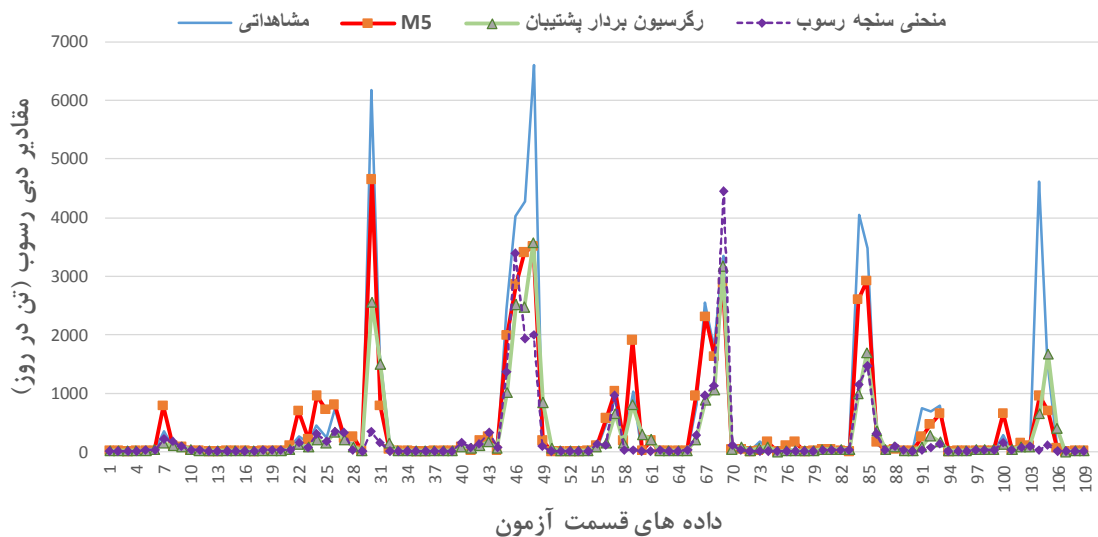
جهت مقایسه بصری نتایج داده‌های محاسباتی بدست آمده رسوب معلق از روش‌های مختلف نسبت به داده‌های مشاهداتی موجود، نمودار سری زمانی داده‌های محاسباتی دبی رسوب نسبت به داده‌های مشاهداتی در بازه آزمون مدل‌ها (بین ماه مرداد ۱۳۸۶ تا ماه شهریور ۱۳۹۳) در شکل (۱۱) ارائه شده است.

با بررسی شکل‌های (۸) تا (۱۰) مشاهده می‌گردد که مقادیر به دست آمده از روش‌های نوین داده‌کاوی پراکنش و همخوانی بسیار مناسبی نسبت به داده‌های مشاهداتی داشته و نقاط بیشتری در بین خطوط خطای $\pm 20\%$ درصد و 20% درصد قرار گرفته‌اند. در صورتی که نتایج به دست آمده از روش متداول منحنی سنجه رسوب نشان‌دهنده تمرکز اکثر نقاط در خارج از این خطوط خطا می‌باشد. نتیجه دیگری که از بررسی دقیق‌تر شکل‌های (۸) تا (۱۱)

قابل استنباط می‌باشد این است که روش کلاسیک منحنی سنجه رسوب در تخمین میزان رسوب رودخانه اهر چای در مقادیر بالای دبی جریان و دبی رسوب، دقت بسیار پایینی داشته و استفاده از این روش برای محاسبه رسوب معلق در اهرچای مناسب نمی‌باشد. برای بررسی صحت این نتیجه، دقت مدل‌ها در محاسبه مقادیر محاسباتی دبی رسوب در مقادیر حدی دبی (دبی‌های مربوط به چارک چهارم داده‌ها در بازه آزمون) مورد واکاوی قرار گرفت که نتایج مربوطه در جدول (۷) ارائه شده است.

جدول ۷. دقت روش‌های مختلف در مدل‌سازی دبی رسوب در مقادیر اوج در قسمت آزمون.

روش مورد استفاده	MAE (ton/day)	RMSE (ton/day)	CC
منحنی سنجه رسوب	۱۵۷۶/۳۳	۲۰۲۱/۶۷	۰/۴۵
رگرسیون بردار پشتیبان	۱۴۳۹/۲۰	۱۷۲۴/۷۴	۰/۸۲
مدل درختی M5	۷۸۶/۸۶	۱۱۵۲/۲۳	۰/۹۲



شکل ۱۱. نمودار سری زمانی مقادیر محاسباتی و مشاهداتی رسوب معلق در بازه داده‌های آزمون.

در صورتی که روش مدل درختی M5 و رگرسیون بردار پشتیبان عملکرد بهتری در این زمینه داشته‌اند. این مسأله می‌تواند در اثر اریب موجود در این روش باشد. اریب در واقع باعث می‌شود تا مقدار باقی‌مانده‌ها (اختلاف بین

با بررسی جدول (۷) مشاهده می‌شود که روش منحنی سنجه رسوب مقادیر دبی رسوب مربوط به دبی‌های سیلابی که اکثر حجم رسوب در این مواقع بوقوع می‌پیوندد را با خطای بسیار زیادی تخمین زده است،

در تخمین میزان رسوب کل در بازه آزمون مدل‌ها در شکل (۱۳) نشان داده شده است.



شکل ۱۳. مقایسه مقدار تجمعی دبی رسوب مشاهداتی و محاسباتی.

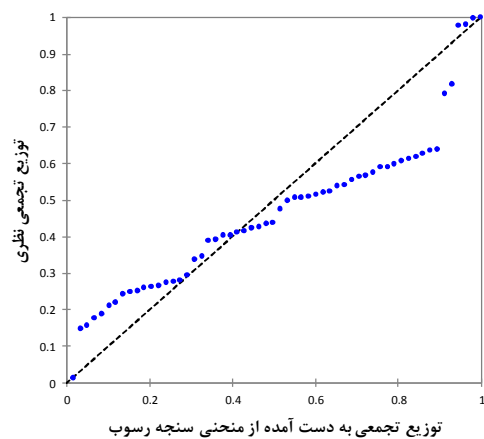
همچنان که مشاهده می‌گردد روش منحنی سنجه رسوب مقادیر رسوب را بسیار کمتر از میزان واقعی آن تخمین زده است، در صورتی که روش‌های نوین داده‌کاوی استفاده شده عملکرد مناسبی در این زمینه از خود نشان داده‌اند. از آنجا که پیش‌بینی کمتر از حد معمول میزان رسوب کل مشکلات عدیده‌ای در بهره‌برداری و استفاده بهینه از منابع آبی موجود ایجاد می‌کند، استفاده از روش‌های داده‌کاوی که توانایی بالایی در این زمینه از خود نشان داده‌اند، برای تخمین میزان رسوب معلق توصیه می‌شود.

نتیجه‌گیری کلی

مدل‌سازی بار رسوب معلق در رودخانه‌ها به دلیل تاثیر آن بر مدیریت و بهره‌برداری سازه‌های هیدرولیکی، محاسبه حجم مرده سدها و همچنین تغییرات مورفولوژی رودخانه از اهمیت زیادی برخوردار است. در مطالعه حاضر، عملکرد دو روش مدل درختی M5 و روش رگرسیون ماشین بردار پشتیبان در برآورد بار رسوبی معلق رودخانه اهرچای مورد ارزیابی قرار گرفت و نتایج به‌دست آمده با نتایج منحنی سنجه رسوب مقایسه گردید. تحلیل حساسیت مدل نسبت به متغیرهای ورودی نشان‌دهنده این مطلب است که میزان رسوب معلق

مقادیر مشاهده‌ای و محاسباتی) توزیع نرمالی نداشته باشند. آزمون نرمالیته برای باقی مانده‌ها در روش منحنی سنجه رسوب با انجام آزمون‌هایی توسط نرم افزار آماری XLSTAT مورد بررسی قرار گرفته و نتایج مربوطه در جدول (۸) و شکل (۱۲) ارائه گردیده است. جدول ۸ نتایج آزمون‌های تعیین نرمال بودن مقادیر باقی مانده به دست آمده از روش منحنی سنجه رسوب.

سطح اطمینان (a)	ریسک رد کردن فرض صفر	آماره p-value	آزمون آماری نرمال بودن
۰/۰۵	۰/۰۱	۰/۰۰۰۱	آزمون شاپیرو ویلک
۰/۰۵	۰/۰۱	۰/۰۰۰۱	آزمون اندرسون دارلینگ
۰/۰۵	۰/۰۱	۰/۰۰۰۱	آزمون لیلیه فورس



شکل ۱۲. نمودار احتمال نرمال به‌دست آمده از مقادیر باقی‌مانده در روش منحنی سنجه رسوب.

در این آزمون‌ها فرض صفر برابر با نرمال بودن داده‌ها و فرض مقابل برابر با غیرنرمال بودن داده‌ها در نظر گرفته شده است. چنانکه مشاهده می‌گردد مقادیر باقی مانده به‌دست آمده از روش منحنی سنجه رسوب در تمامی آزمون‌های مورد بررسی نرمال نبوده و به همین علت نتایج به دست آمده از این روش نسبت به روش‌های داده‌کاوی اشاره شده از دقت کمتری برخوردار است. از آنجا که میزان رسوب کل در یک بازه زمانی مشخص در مطالعات منابع آب و طراحی سدها اهمیت زیادی دارد، نحوه عملکرد روش‌های داده‌کاوی و منحنی سنجه رسوب

بیشترین تاثیرپذیری را از متغیر دبی جریان در ماه جاری داشته و اطلاع دقیق از دبی جریان می‌تواند در تخمین و برآورد میزان رسوب معلق بسیار موثر باشد. مقایسه نتایج پژوهش حاضر با سایر مطالعات انجام یافته در این زمینه نشان‌دهنده دقت قابل قبول و در اکثر موارد بیشتر روش‌های به کار رفته در این پژوهش می‌باشد. چنانکه در پژوهش مشابهی، خزائی پول و طالبی (۱۳۹۲) رسوبات معلق رودخانه قطورچای را با استفاده از شبکه عصبی مصنوعی و منحنی سنج رسوب پیش‌بینی نمودند و هنگام مدل‌سازی رسوبات معلق این رودخانه توسط شبکه عصبی مصنوعی، ضریب همبستگی را برابر با ۰/۸۷۳ و مقدار ریشه میانگین مربعات خطا را برابر ۶۲۲ (تن در روز) بدست آوردند. حسین‌زاده دلیر و همکاران (۱۳۸۸) در مطالعه‌ای رسوب را با روش شبکه عصبی در رودخانه اهرچای پیش‌بینی کردند. نتایج نشان داد که، حداکثر ضریب همبستگی صحت سنجی ۰/۸۰ در ایستگاه اورنگ و کمترین آن ۰/۶۹ در ایستگاه اشدلق می‌باشد. همچنین (Battacharya & Solomatine, 2006) مدل

شبکه عصبی پرسپترون چند لایه و مدل درختی *M5* را برای پیش‌بینی رسوب در حوضه آبخیز هارپور در بندر روتردام هلند به کار بردند. نتایج نشان داد که، شبکه عصبی مصنوعی با ضریب همبستگی ۰/۸۲ نسبت به مدل درختی با ضریب همبستگی ۰/۶۶ عملکرد بهتری در تخمین بار رسوب معلق داشته است. در مطالعه حاضر نیز نتایج حاکی از آن است که روش مدل درختی *M5* با ضریب همبستگی ۰/۹۳ و روش رگرسیون بردار پشتیبان با ضریب همبستگی ۰/۸۹ عملکرد بهتری نسبت به روش متداول منحنی سنج رسوب در تخمین بار رسوب رودخانه اهرچای داشته و می‌تواند بعنوان روش جایگزین در مدل‌سازی رسوب معلق مورد استفاده قرار گیرند. با توجه به این‌که مدل درختی *M5* با ارائه قوانین ساده خطی اگر-آنگاه بار رسوب معلق رودخانه اهرچای را با دقت نسبتاً خوبی مدل‌سازی می‌نماید، می‌توان نتیجه گرفت این روش نسبت به روش رگرسیون بردار پشتیبان کاربردی‌تر است.

فهرست منابع

- حسین زاده دلیر، ع.، فرسادی‌زاده، د. و قربانی، م. ع. ۱۳۸۸. تاثیر نوسانات سطح آب و دبی جریان بر میزان بار معلق با استفاده از شبکه‌های عصبی مصنوعی (مطالعه موردی: رودخانه اهرچای). دانش آب و خاک. ۱۹(۱): ۲۳-۳۳.
- خزائی پول، ا. و طالبی، ع. ۱۳۹۲. بررسی امکان پیش‌بینی رسوبات معلق با استفاده از ترکیب منحنی سنج رسوب و شبکه‌ی عصبی مصنوعی (مطالعه موردی: رودخانه قطورچای، پل یزدکان). پژوهش‌های فرسایش محیطی. ۹: ۷۳-۸۲.
- داننده مهر، ع.، علیایی، ا. و قربانی، م. ع. ۱۳۸۹. پیش‌بینی بار معلق رودخانه‌ها بر مبنای دبی جریان با استفاده از برنامه ریزی ژنتیک. پژوهش‌های آبخیزداری (پژوهش و سازندگی). ۸۸ (۲): ۴۴-۵۴.
- دهقانی، ن. و وفاخواه، م. ۱۳۹۲. مقایسه روشهای تخمین رسوب معلق روزانه با استفاده از روش‌های منحنی سنج رسوب و شبکه عصبی (مطالعه موردی: ایستگاه قزاقلی، استان گلستان). پژوهش‌های حفاظت آب و خاک. ۲۰(۲): ۱-۱۰.
- رجبی، م.، فیض‌اله پور، م. و روستایی، ش. ۱۳۹۴. استفاده از مدل تبرید تدریجی عصبی (*NDE*) در تخمین بار معلق رسوبی و مقایسه آن با مدل *ANFIS* و *RBF* (مطالعه موردی: رودخانه گیوی‌چای). جغرافیا و توسعه، ۳۹(۲): ۱-۱۶.
- مه‌ریزی حائری، ع. ا. ۱۳۸۲. داده‌کاوی: مفاهیم، روش‌ها و کاربردها. پایان‌نامه کارشناسی ارشد آمار اقتصادی و اجتماعی، دانشکده اقتصاد، دانشگاه علامه طباطبائی.
- شهرابی، ج. و حجازی، ط. ح. ۱۳۹۱. داده‌کاوی. انتشارات دانشگاه امیرکبیر. تهران. ۱۳۱ صفحه.
- صفوی، ح. ۱۳۹۰. هیدرولوژی مهندسی، انتشارات ارکان دانش، چاپ سوم، اصفهان. ۶۹۹ صفحه.
- Alberg, D., Last, M. and Kandel, A. 2012. Knowledge Discovery In Data Streams With Regression Tree Methods. Wiley Interdisciplinary Reviews Data Mining Knowledge Discovery, 2: 69-78.

- Aytek, A. and Kisi, O. 2008. A Genetic Programming Approach to Suspended Sediment Modelling. *Hydrology*, 351: 288-298.
- Battacharya, B. and Solomatine, D. P. 2006. Machine Learning in Sedimentation Modeling, *Neural Network*, 19, pp: 208-214.
- Ghorbani, M. A., Hosseini, S. H., Fazelifard, M. H. and Abbasi, H. 2013. Sediment Load Estimation by MLR, ANN, NF and Sediment Rating Curve (SRC) In Rio Chama River. *Journal of Civil Engineering and Urbanism*, 3(4): 136-141.
- Graziose, J. 2013. Is My Data Normal? Using Technology to Test For Normality. 25th anniversary International Conference on Technology in Collegiate Mathematics, 169-175.
- Heng, S. and Suetsugi, T. 2013. Using artificial neural network to estimate sediment load in ungauged catchments of the Tonle Sap River Basin, Cambodia. *Journal of Water Resource and Protection*, 5: 111-123.
- Kakaei Lafdani, E., Moghaddam Nia, A. and Ahmadi, A. 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Hydrology*, 478: 50-62.
- Kao, Sh., Lee, T. and Milliman, J. D. 2005. Calculating Highly Fluctuated Suspended Sediment Fluxes From Mountainous Rivers In Taiwan. *Terrestrial, Atmospheric and Oceanic Sciences*, 16(3): 653-675.
- Kia, E., Emadi, A. R. and Fazlola, R. 2013. Investigation and Evaluation of Artificial Neural Networks in Babolroud River Suspended Load Estimation. *Journal of Civil Engineering and Urbanism*, 3(4): 183-190.
- Kisi, O. and Shiri, J. 2012. River Suspended Sediment Estimation by Climate Variables Implication: Comparative Study among Soft Computing Techniques. *Computer and Geosciences*, 43: 73-82.
- Kumar Goyal, M. 2014. Modeling of Sediment Yield Prediction Using M5 Model Tree Algorithm and Wavelet Regression. *Journal of Water Resources Management*, 28: 1991-2003.
- Nourani, V. 2009. Using Artificial Neural Network (ANNs) For Sediment Load Forecasting of Talkherood River Mouth. *Urban and Environmental Engineering*, 3(1): 1- 6.
- Ozturk, F., Apaydin, H. and Walling, D. E. 2001. Suspended Sediment Loads through Flood Events for Streams of Sakarya Basin. *Journal of Engineering Environment*, 25, 643-650.
- Platt, J. 2000. Fast Training Support Vector Machine Using Sequential Minimal Optimization. <http://www.research.microsoft.com/~jplatt>. 41-65.
- Quinlan, J. R. 1992. Learning with Continuous Classes, Singapore. In proceedings AI, 92 (Adams & Sterling, Eds). World Scientific, pp: 343-348.
- Rajaei, T., Mirbagheri, S. A., Nourani, V., and Alikhani, A. 2010 Prediction Of Daily Suspended Sediment Load Using Wavelet And Neuro Fuzzy Combined Model. *Environment Sciences, Tech*, 7 (1): 93-110.
- Shapiro, S. S.; Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52 (3-4): 591-611.
- Stephens, M.A. 1979. Test of fit for the logistic distribution based on the empirical distribution function, *Biometrika*, 66(3), 591-5.
- Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. Springer, New York. 314p.



Performance evaluation of M5 tree model and support vector regression methods in suspended sediment load modeling

Mohammad Taghi Sattari¹, Ali Rezazadeh Joudi^{2*}, Forough Safdari³ and Faraz Ghahramanian⁴

1) Assistant Professor, Department of Water Engineering, Agriculture Faculty, University of Tabriz, Tabriz, Iran

2) Young Researchers and Elite Club, Maragheh Branch, Islamic Azad University, Maragheh, Iran

* Corresponding author: alijoudi66@gmail.com

3) M.Sc. Graduated of Water Resources, Department of Agriculture, University of Tabriz, Tabriz, Iran

4) M.Sc. Graduated of Water-Civil Engineering, Department of Civil Engineering, Islamic Azad University, Ahar, Iran

Received: 06-03-2016

Accepted: 28-05-2016

Abstract

Sediment transport has always affected the river and civil structures and the lack of knowledge about its exact amount causes high damages. Therefore, it is very important to properly estimate the sediment load in rivers in terms of sediment, erosion and flood control. This study used two new data mining methods including M5 model tree and support vector regression comparing with the classic method of sediment rating curve to estimate the suspended sediment load in Aharchay River. To assess the performance of the used methods, three criteria including the correlation coefficient, root mean square error and mean absolute error were used. Analyzing the sensitivity of models to the input variables, it was found that the variable of flow discharge in the current month had the greatest effect on the amount of suspended sediment load. The results showed the high accuracy of new data mining methods in comparison with the sediment rating curve. Although, both considered data mining methods had more accuracy and less error compared to the conventional sediment rating curve, given the simple understandable linear relationships provided by the M5 model tree, this method is recommended for similar cases.

Keywords: data mining, M5 tree model, sediment load estimation, support vector regression and Aharchay