Archive of SID

کاربرد الگوریتم درختی M5 در برآورد رسوب معلق رودخانه‌ها

سیده حدیث شاهرخی[۱]، جواد ظهیری[۲] و احمد جعفری[۳]

تاریخ دریافت: ۱۳۹۵/۰۱/۳۰
تاریخ پذیرش: ۱۳۹۵/۰۵/۱۶

چکیده

برآورد میزان بار رسوب در بسیاری از پروژه‌های آبی از قبیل طراحی سدها و حوضچه‌های ذخیره آب سطحی، انتقال رسوب و آلودگی در رودخانه‌ها و دریاچه‌ها و نیز طراحی کانال‌ها و نگهداری آن‌ها اهمیت زیادی دارد. روش‌های متداول برآورد بار رسوب عمدتا دقت پایینی داشته و نتایج روش‌های مختلف تفاوت زیادی با یکدیگر دارد. بر همین اساس در این تحقیق پس از تهیه داده‌های دبی رسوب مربوط به ایستگاه‌های هیدرومتری جوکنک، بهبهان، شادگان، مشراگه و چم نظام، از الگوریتم درختی M5 که یکی از مدل‌های نوین داده کاوی به حساب می‌آید، جهت محاسبه بار معلق استفاده گردید. الگوریتم درختی M5 فضای مساله را به شاخه‌های متعدد تقسیم کرده و معادلاتی را جهت تخمین بار معلق رسوب در شاخه‌های مختلف ارائه می‌دهد. دبی جریان یک روز قبل و دبی همان روز به عنوان پارامترهای ورودی و دبی بار معلق رسوب نیز به عنوان پارامتر خروجی جهت ساخت و صحت‌سنجی مدل در نظر گرفته شد. آنالیزهای مختلف آماری بر روی معادلات ارائه شده توسط الگوریتم M5 نشان دهنده دقت بهتر مدل درختی در مقایسه با منحنی سنجه می‌باشد. در ایستگاه‌های مورد بررسی مقدار RMSE مدل درختی با توجه به داده‌های آموزش و صحت‌سنجی ۰/۵۹ محاسبه گردید در حالی‌که در مورد منحنی سنجه این مقدار ۰/۷۴ به دست آمد. بر اساس نتایج این تحقیق مدل M5 علاوه بر اینکه می‌تواند میزان رسوب معلق را با دقت بالایی برآورد کند دارای ساختار ساده و قابل فهم بوده و از نتایج آن می‌توان جهت مسائل کاربردی سود جست.

واژه‌های کلیدی: ایستگاه‌های هیدرومتری ، مدل‌های داده کاوی، منحنی سنجه رسوب.

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی آب دانشگاه کشاورزی و منابع طبیعی رامین خوزستان_ خرم آباد، میدان کیو، خیابان انقلاب، ۰۹۱۶۰۵۰۶۱۸۱، sh.hadis815@gmail.com

[۲] استادیار، گروه مهندسی آب دانشگاه کشاورزی و منابع طبیعی رامین خوزستان، ۰۹۱۶۶۵۳۱۸۹۶، Zahiri_Javad@yahoo.com (مسئول مکاتبه)

[۳] استادیار، گروه مهندسی آب دانشگاه کشاورزی و منابع طبیعی رامین خوزستان، ۰۹۱۳۱۲۶۱۳۵۰، Jafary_Ahmad@yahoo.com

www.SID.ir

Archive of SID

مقدمه

پدیده فرسایش خاک و انتقال رسوب از جمله فرآیندهای مهم هیدرودینامیک هستند که به عنوان 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 and et al., 2009). از دیگر دلایل اهمیت تخمین رسوب معلق رودخانه‌ها می‌توان به تاثیر آن بر روی کیفیت سامانه‌های آبی اشاره کرد (ابراهیمی محمدی و همکاران، ۱۳۹۱). جهت برآورد بار معلق رودخانه‌ها اصولا از دو روش تجربی و رگرسیون آماری استفاده می‌شود. روابط مورد استفاده در روش تجربی مبتنی بر مفاهیم فیزیکی بوده و معمولا به داده‌های متنوعی نیاز دارند. تعدد روابط تجربی ارائه شده، 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 (معتمد وزیری، ۱۳۸۸). در روش رگرسیون آماری از پردازش بین مقادیر دبی جریان و رسوب معلق متناظر آن‌ها به صورت رابطه توانی $Q_s = a\,Q_w^b$ استفاده می‌گردد. در این رابطه Q_s میزان دبی بار معلق و Q_w دبی جریان می‌باشد. ضرایب a و b ضرایب معادله بوده که با توجه به اطلاعات مورد استفاده، از طریق رگرسیون‌های آماری محاسبه می‌شوند (Achite and Ouillon, 2007). اکبرزاده و همکاران (۱۳۹۳) جهت برآورد رسوب رودخانه اترک از روش‌های هیدرولوژیکی استفاده کردند. نتایج این تحقیق نشان داد که هر کدام از روش‌های حد وسط دسته‌ها، مدل خطی، مدل ترسیمی و مدل ترکیبی در ایستگاه‌های مشخصی بهترین تخمین را داشته‌اند. در طی سال‌های اخیر سعی شده است که روش‌های داده کاوی از قبیل شبکه‌های عصبی مصنوعی و منطق فازی جایگزین روش منحنی سنجه که همواره با خطای نسبتا زیادی همراه است، شوند. از جمله تحقیقات جدید صورت گرفته در زمینه استفاده از مدل‌های داده‌کاوی در برآورد بار معلق می‌توان به Zounemat-Kermani et al., 2016 اشاره کرد. این محققین از شبکه‌های عصبی مصنوعی و مدل‌های رگرسیون بردار پشتیبان جهت مدل‌سازی رسوب معلق روزانه در ایستگاه‌های هیدرومتری آرکانزاس، دلور و آیداهو در ایالات متحده استفاده کردند. مقایسه نتایج مدل‌های مورد استفاده با روش‌های معمول از قبیل روش چند رگرسیونی و منحنی سنجه نشان داد که مدل‌های داده کاوی بهتر از روش‌های معمول می‌توانند رسوب روزانه را مدل نمایند. روش‌های داده کاوی از قبیل شبکه‌های عصبی و منطق فازی هر چند دقت بهتری در مقایسه با روش‌های متداول داشته ولی از آنجا که نتایج آن‌ها به صورت یک مدل آموزش دیده می‌باشد و روابطی جهت محاسبه بار معلق ارائه نمی‌دهند، نمی‌توان به سادگی از آن‌ها استفاده کرده و به‌کار بردن آن‌ها در مدل‌های دیگر نیز امکان‌پذیر نیست. بر همین اساس در سال‌های اخیر نسل جدید از مدل‌های داده کاوی از قبیل الگوریتم‌های درختی مورد توجه قرار گرفته است که می‌توان از آن‌ها جهت استخراج قوانین حاکم بر یک پدیده با استفاده از داده‌های موجود استفاده کرد (ظهیری، ۱۳۹۴). این مدل‌ها با استفاده از فرآیند تکراری تفکیک در گره‌ها، داده‌ها را به کلاس‌های مجزا تقسیم کرده و توسعه پیدا می‌کنند. از جمله مهمترین تحقیقات صورت گرفته بر روی مدل‌های مذکور می‌توان به Bhattacharya and Solomatine(2006) اشاره کرد که از الگوریتم M5 جهت شبیه‌سازی رسوبات در بندر رتردام استفاده کردند. بر اساس نتایج به‌دست آمده در این تحقیق می‌توان از مدل‌های ذکر شده جهت تصمیمات کاربردی سود جست. Etemad-Shahidi and Mahjoobi. (2009) از مدل درختی M5 جهت

www.SID.ir

Archive of SID

شبیه‌سازی ارتفاع امواج رودخانه استفاده کردند. آن‌ها علاوه بر مدل درختی از شبکه عصبی نیز جهت مقایسه کارآیی مدل درختی استفاده کردند و نشان دادند که مدل درختی با دقت بیشتری می‌تواند ارتفاع امواج را شبیه‌سازی کند. Mahjoobi et al. (2010) نیز از مدل های درختی CART و M5 جهت تخمین عمق آبشستگی اطراف پایه‌های پل استفاده کردند. Etemad-Shahidi and Ghaemi. (2011) جهت بررسی کارآیی الگوریتم M5 در محاسبه عمق آبشستگی اطراف مجموعه پایه‌های پل از اطلاعات آزمایشگاهی و میدانی استفاده کردند. نتیجه این تحقیق دقت بالاتر مدل‌های درختی در مقایسه با معادلات رایج بوده است. Kumar et al., 2016 جهت برآورد بار معلق رودخانه کوپیلی از روش‌های مختلف هیدرولوژیکی و داده کاوی استفاده کردند. مقایسه نتایج روش‌های مختلف نشان داد که روش‌های بردار پشتیبان و شبکه عصبی مصنوعی نسبت به روش‌های هیدرولوژیکی و الگوریتم‌های درختی از دقت بالاتری برخوردارند. ظهیری (۱۳۹۴) از مدل درختی M5 و CART جهت تخمین عمق آبشستگی اطراف پایه پل استفاده کرد. نتیجه این تحقیق دقت بالاتر M5 در مقایسه با مدل درختی CART و سایر روش‌های متداول می‌باشد. در این تحقیق جهت مدل‌سازی میزان بار معلق رودخانه‌ها از الگوریتم درختی M5 استفاده گردید. پارامترهای ورودی به مدل درختی شامل دبی جریان و خروجی آن نیز به صورت بار معلق رسوب در نظر گرفته شد. پس از ساخت و صحت‌سنجی مدل M5، آنالیزهای آماری بر روی مدل درختی صورت گرفت.

مواد و روش‌ها

اطلاعات مورد استفاده جهت ساخت و صحت سنجی الگوریتم درختی M5 از منابع سازمان آب و برق استان خوزستان استخراج شده است. این اطلاعات به صورت میدانی برداشت شده و دبی جریان و نمونه برداری رسوب از نقاط مختلف رودخانه را شامل می‌شود. اطلاعات مورد استفاده شامل دبی جریان آب و دبی رسوب مربوط به پنج ایستگاه هیدرومتری مشراگه و شادگان بر روی رودخانه مارون، جوکنک بر روی رودخانه الله و چم نظام و بهبهان بر روی رودخانه جراحی در استان خوزستان می‌باشد. موقعیت ایستگاه‌های مورد استفاده و مشخصات اطلاعات مربوط به این ایستگاه‌ها در جدول ۱ ارائه شده است. همچنین موقعیت حوضه آبریز جراحی، مارون و الله در سطح منطقه در شکل ۱ نشان داده شده است.

مدل درختی M5

الگوریتم M5 اولین بار توسط Quinlan (1992) ابداع شد و پس از آن توسط Wang and Witten (1997) توسعه و بهبود یافت. روش M5 شاخه‌های خود را به صورت دوتایی و تنها براساس 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 زیر دامنه‌هایی تقسیم شده و برای هر زیر دامنه یک مدل رگرسیون خطی چند متغیره برازش داده می‌شود. این الگوریتم جداسازی‌های ممکن را در فضای چند متغیره انجام داده و به طور خودکار مدل‌هایی را برای هر کدام از دامنه‌ها ایجاد می‌کند (Quinlan., 1992). در این الگوریتم از پارامتر انحراف معیار مقادیر متغیر هدف به عنوان 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 به عنوان صفت مورد نظر جهت ایجاد شاخه انتخاب می‌شود. کاهش انحراف استاندارد (SDR) که به عنوان تابع خطا در الگوریتم M5 به کار می‌رود به صورت زیر است:

$$SDR = sd(T) - \sum \left| \frac{Ti}{T} \right| \times sd(T) \quad (۱)$$

www.SID.ir

Archive of SID

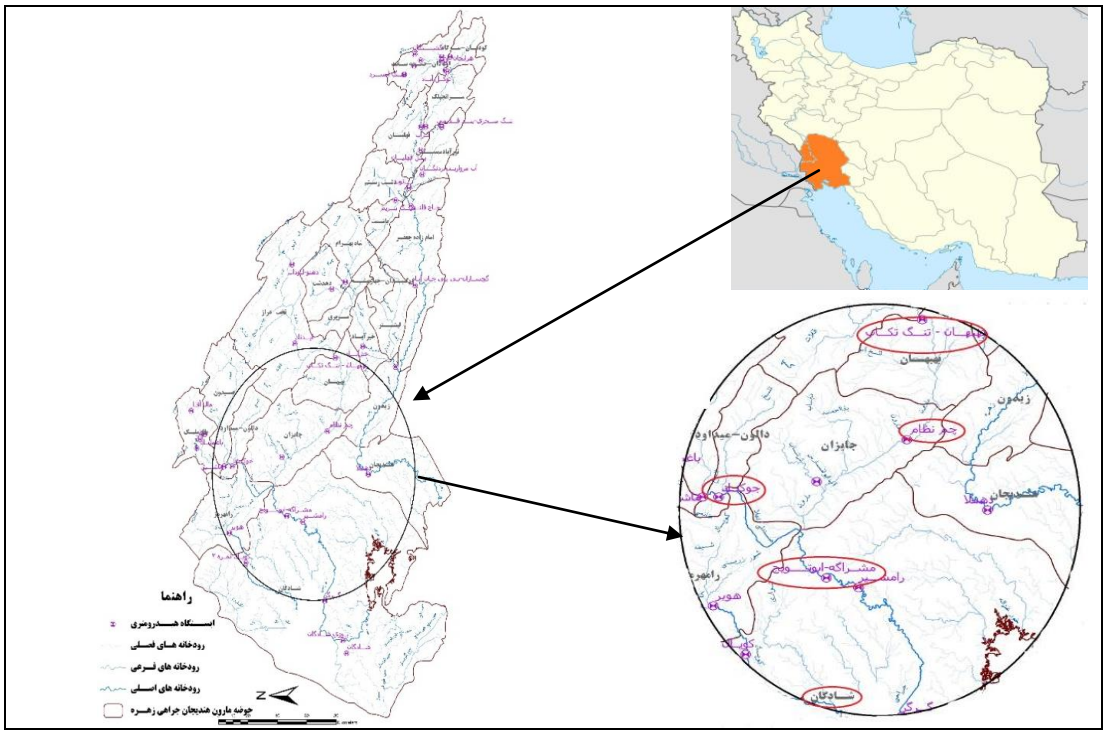
در رابطه بالا T شامل نمونه‌هایی است که به گره مورد نظر رسیده‌اند و Ti شامل نمونه‌های است که از تقسیم گره مورد نظر بر اساس صفت انتخاب شده به دست آمده‌اند. Sd نیز معرف انحراف معیار می‌باشد. الگوریتم M5 تمامی حالت‌های مختلف جهت ایجاد شاخه بر اساس صفت خاص را بررسی کرده و در نهایت حالتی را انتخاب می‌کند که 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 برازش داده می‌شود. شکل (۲) نحوه تقسیم فضای مسأله به اجزای کوچک‌تر و کاربرد مدل‌های خطی را برای هر کدام از زیر دامنه‌ها با توجه به ساختار درختی نشان می‌دهد (Etemad Shahidi and Taghipour, 2012). در طی ساخت مدل درختی توسط الگوریتم M5 فرآیند جدا سازی در گره‌های تقسیم ممکن است بارها تکرار شده و در نتیجه درخت بسیار بزرگی ایجاد شود. در این حالت مدل دچار بیش برازش 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.

جدول (۱): مشخصات ایستگاه‌های هیدرومتری مورد استفاده جهت ساخت و صحت‌سنجی مدل درختی M5

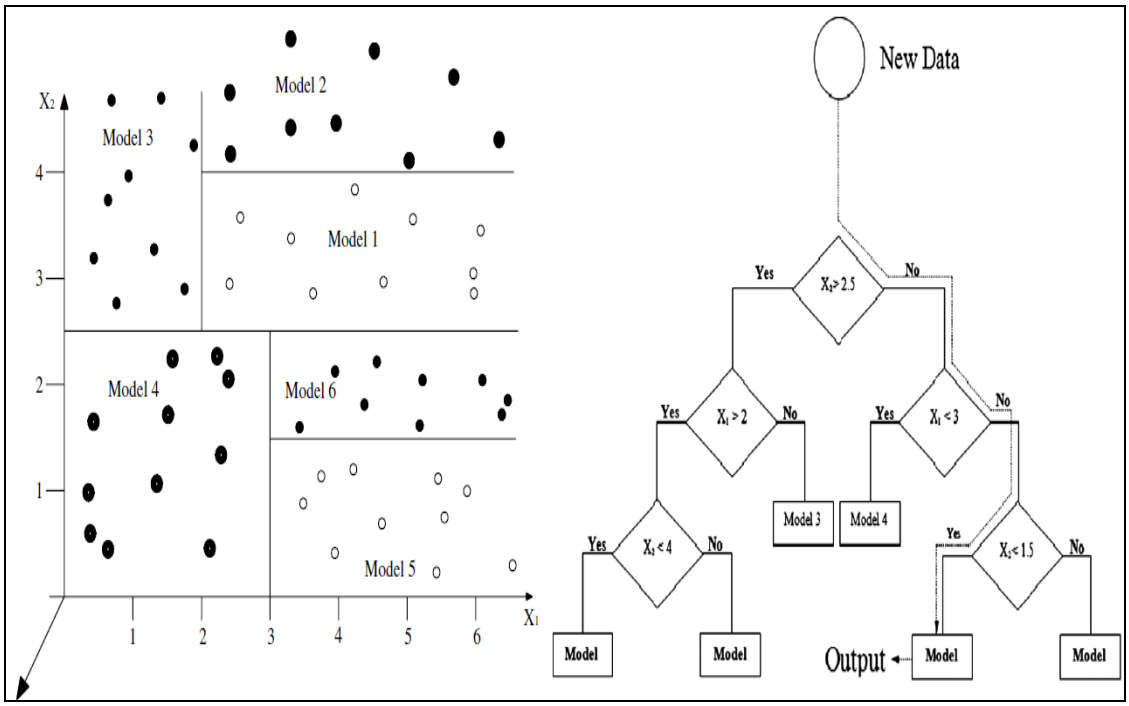
ردیف	ایستگاه	سال	موقعیت		
			عرض جغرافیایی	طول جغرافیایی	ارتفاع
۱	مشراگه	۱۳۹۰-۱۳۶۶	۳۱۰۰۲۲	۴۹۲۶۲۴	۲۲
۲	شادگان	۱۳۹۱-۱۳۶۰	۳۰۳۹۱۸	۴۸۴۳۲۴	۶
۳	چم نظام	۱۳۹۱-۱۳۷۰	۳۱۱۴۲۴	۴۹۳۴۵۷	۱۷۸
۴	جوکنک	۱۳۹۱-۱۳۷۰	۳۱۲۰۲۸	۴۹۴۳۲۵	۳۰۶
۵	بهبهان	۱۳۹۱-۱۳۷۰	۳۰۴۰۲۹	۵۰۲۰۳۲	۱۷۸

www.SID.ir

Archive of SID



شکل(۱): موقعیت حوضه آبریز جراحی، مارون و الله در سطح منطقه



شکل(۲): تقسیم فضای مسأله و ارائه مدل‌های خطی به ازای هر کدام از فضاها توسط الگوریتم M5

www.SID.ir

Archive of SID

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 مکانیزمی به نام هموارسازی[1] جهت رفع ناپیوستگی ایجاده شده در مدل‌های خطی استفاده می‌کند. در این فرآیند مقدار تخمین زده شده در هر برگ تصحیح می‌شود. چنانچه نمونه مورد نظر در شاخه Si از زیر درخت S باشد، ni تعداد نمونه‌های آموزشی در Si، PV(Si) مقدار محاسبه شده در Si و M(S) مقدار محاسبه شده توسط مدل در S باشد، آنگاه مقدار اصلاح شده (PV(Si)) از رابطه زیر محاسبه خواهد شد.

$$PV = \frac{ni \times pv(si) + k \times M(s)}{ni + k} \quad (۱۲)$$

در رابطه فوق k ثابت هموارسازی بوده که به صورت پیش‌فرض برابر ۱۵ در نظر گرفته می‌شود. هموارسازی بویژه در مواقعی که مدل‌های خطی در برگ‌های مجاور، مقادیر کاملا 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 داده است که هموارسازی می‌تواند کارآیی الگوریتم M5 را به میزان زیادی افزایش دهد. در این تحقیق جهت ساخت و آموزش مدل درختی M5 از ۸۰ درصد اطلاعات به صورت تصادفی استفاده گردید و ۲۰ درصد باقیمانده نیز جهت صحت سنجی مدل درختی مورد استفاده قرار گرفت. مدلM5 تنها قادر به شبیه‌سازی مدل‌های خطی بوده که این امر با اساس معادلات متداول بار معلق در تناقض است. بر همین اساس کلیه داده‌ها به صورت لگاریتم طبیعی به مدل معرفی شدند تا پس از خارج کردن روابط از حالت لگاریتمی به روابط توانی تبدیل شوند. جهت ساخت مدل درختی دبی جریان مربوط به ۵، ۴، ۳، ۲ و یک روز قبل و دبی جریان همان روز به عنوان پارامترهای ورودی به مدل معرفی شد و نتایج مدل مورد بررسی قرار گرفت. نتایج نشان داد که تنها دبی جریان مربوط به یک روز قبل و دبی همان روز بر روی میزان دبی رسوب معلق تاثیر معنادار داشته و به همین دلیل روابط ارائه شده در این تحقیق بر اساس تنها دو پارامتر ورودی می‌باشند. 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)[2]، خطای متوسط مطلق (MAE)[3]، جذر میانگین مربعات خطا (RMSE)[4] و ضریب تعیین (R^2)[5] اشاره کرد. معادلات مورد استفاده جهت تحلیل‌های آماری در زیر ارائه شده است.

$$SSE = \sum \frac{\left(ds^{measuerd} - ds^{computed}\right)^2}{\sum \left(ds^{measuerd}\right)^2} \quad (۳)$$

$$MAE = \frac{1}{N} \sum \left| ds^{measuerd} - ds^{compused} \right| \quad (۴)$$

$$RMSE = \sqrt{\frac{1}{N} \sum \left(d_s^{measured} - d_s^{computed}\right)^2} \quad (۵)$$

$$R^2 = 1 - \frac{\sum \left(d_s^{measured} - d_s^{computed}\right)^2}{\sum \left(d_s^{measured} - \overline{d_s^{measured}}\right)^2} \quad (۶)$$

نتایج و بحث

پس از ساخت مدل درختی M5 معادلات به دست آمده از حالت لگاریتمی خارج شده و به صورت معادلات ۷ تا ۱۸ تبدیل شده‌اند. معادلات ۷ تا ۱۰ مربوط به ایستگاه چم نظام، معادلات ۱۱ و ۱۲ مربوط به ایستگاه بهبهان، معادلات ۱۳ و ۱۴ مربوط به جوکنک، معادلات ۱۵ و ۱۶ مربوط به ایستگاه شادگان

[1] Smoothing process

[2] Sum of the Square Error
[3] Mean-Absolute Error
[4] Root Mean Square Error
[5] Determination Coefficient

www.SID.ir

Archive of SID

و معادلات ۱۷ و ۱۸ مربوط به ایستگاه مشراگه است. در این روابط Q_s دبی رسوب، Q_t دبی جریان همان روز و Q_{t-1} دبی جریان یک روز قبل می‌باشد. شکل ۳ ساختار درختی مدل M5 را در ایستگاه‌های مشراگه، شادگان، چم نظام، جوکنک و بهبهان نشان می‌دهد. همانگونه که اشاره شد مدل M5 تنها قادر به ساخت مدل‌های خطی بوده به همین دلیل و با توجه به مکانیزم انتقال رسوب که دارای رابطه توانی است، در این تحقیق کلیه مقادیر به صورت لگاریتمی به مدل معرفی گردید. مقادیر ارائه شده در شکل ۳ مقادیر واقعی بوده که از حالت لگاریتمی خارج شده‌اند. در شکل ۳ مشاهده می‌شود که در تمامی ایستگاه‌ها بجز ایستگاه بهبهان، 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 چم نظام از چهار شاخه جهت محاسبه بار معلق استفاده شده است. مهم‌ترین دلیل را می‌توان به مقادیر دبی و رسوب در این ایستگاه مربوط دانست که نسبت به بقیه ایستگاه‌ها دارای بازه تغییرات بیشتری است.

Cham Nezam

$$LM1: Q_S = 5.81 Q_{t-1}^{0.38} Q_t^{0.44} \quad (۷)$$

$$LM2: Q_S = 487.84 Q_{t-1}^{1.16} Q_t^{0.82} \quad (۸)$$

$$LM3: Q_S = 1.07 Q_{t-1}^{0.68} Q_t^{0.6} \quad (۹)$$

$$LM4: Q_S = 0.004 Q_{t-1}^{0.35} Q_t^{2.98} \quad (۱۰)$$

Behbahan

$$LM1: Q_S = 5.002 Q_{t-1}^{0.83} Q_t^{0.074} \quad (۱۱)$$

$$LM2: Q_S = 0.88 Q_t^{1.45} \quad (۱۲)$$

Jow Kanak

$$LM1: Q_S = 4.9 Q_{t-1}^{0.95} Q_t^{1.37} \quad (۱۳)$$

$$LM2: Q_S = 0.57 Q_t^{2.17} \quad (۱۴)$$

Shadegan

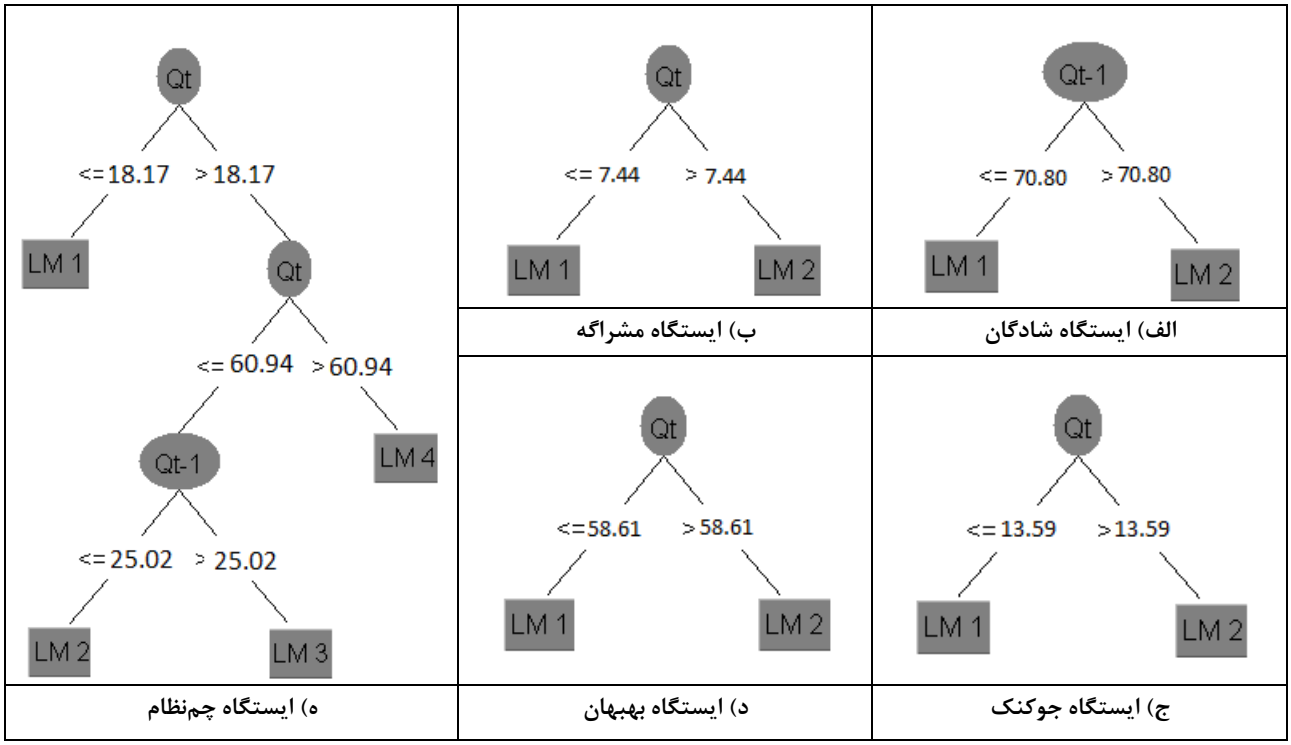
$$LM1: Q_S = 16.11 Q_t^{0.6} \quad (۱۵)$$

$$LM2: Q_S = 1.5 Q_t^{2.29} \quad (۱۶)$$

Moshrageh

$$LM1: Q_S = 2.27 Q_{t-1}^{1.4} \quad (۱۷)$$

$$LM2: Q_S = 0.15 Q_t^{2.18} \quad (۱۸)$$

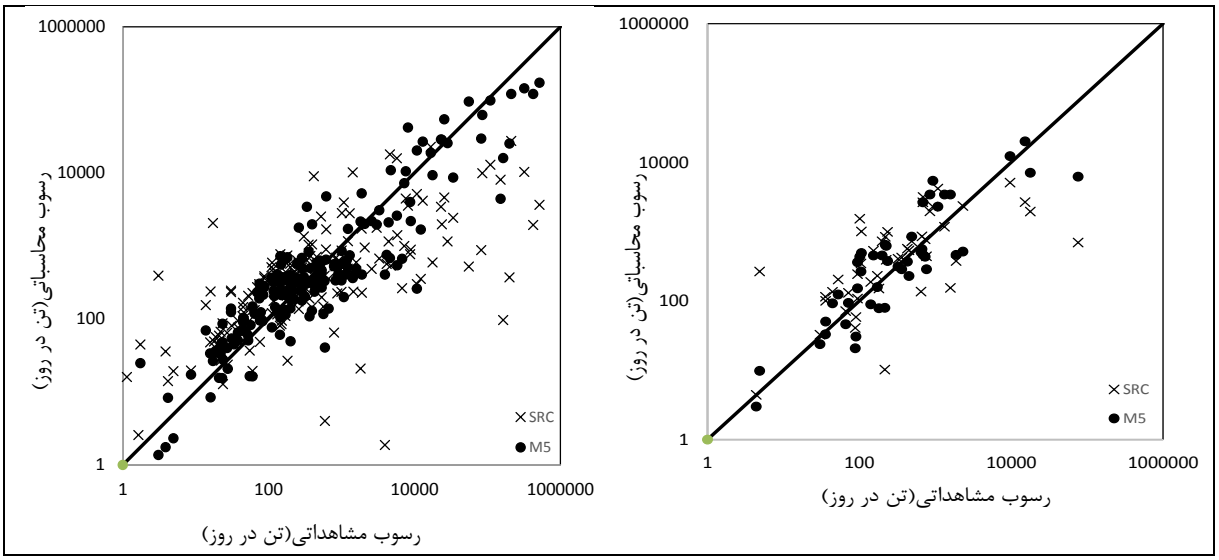


شکل(۳): ساختار درختی M5 جهت برآورد بار معلق رسوب در ایستگاه‌های مختلف

www.SID.ir

Archive of SID

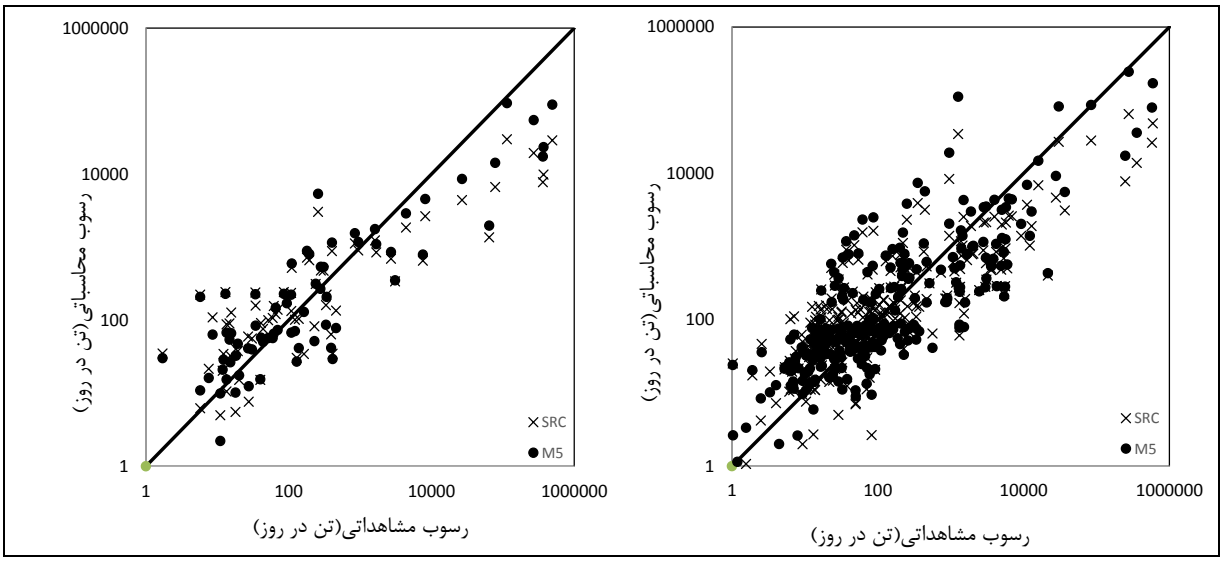
در تمامی معادلات ارائه شده توسط مدل درختی (معادلات ۷ تا ۱۸)، میان دبی رسوب و دبی جریان مربوط به همان روز و روز قبل رابطه مستقیم وجود داشته که با فیزیک مساله مطابقت دارد. در حقیقت معادلات ارائه شده توسط مدل M5 به مانند روش منحنی سنجه بوده که دو تفاوت اصلی را با منحنی سنجه دارند. تفاوت اول مربوط به تقسیم فضای مساله به دو یا چند زیردامنه بوده و ارائه معادله به ازای هر زیردامنه می‌باشد، در حالیکه در منحنی سنجه کل فضای مساله به عنوان یک دامنه در نظر گرفته می‌شود. از طرفی در منحنی سنجه تنها پارامتر ورودی دبی همان روز بوده ولی در معادلات ارائه شده توسط مدل درختی، دبی جریان روز قبل هم بر روی میزان رسوب معلق تاثیرگذار می‌باشد. نتایج مربوط به مقادیر بار معلق محاسبه شده توسط مدل درختی M5 و منحنی سنجه در ایستگاه‌های هیدرومتری مطالعه شده در این تحقیق در اشکال ۴ تا ۸ نشان داده شده است. در کلیه گراف‌های ارائه شده محور افقی نشان‌دهنده مقادیر رسوب معلق مشاهده شده و محور قائم معرف مقادیر محاسبه شده توسط الگوریتم درختی و منحنی سنجه می‌باشد. بر اساس گراف‌های ارائه شده هر چه پراکندگی داده‌ها در اطراف خط ۴۵ درجه بیشتر باشد، نشان‌دهنده دقت بالاتر مدل خواهد بود. در تمامی شکل های ارائه شده نتایج مدل M5 در مقایسه با منحنی سنجه انطباق بهتری با مقادیر مشاهداتی داشته و پراکندگی بهتری در اطراف خط ۴۵ درجه نشان می‌دهند. در اغلب موارد مدل درختی M5 میزان رسوب معلق را بیشتر از منحنی سنجه برآورد کرده است که این امر به ویژه در ایستگاه بهبهان به وضوح مشاهده می‌شود.



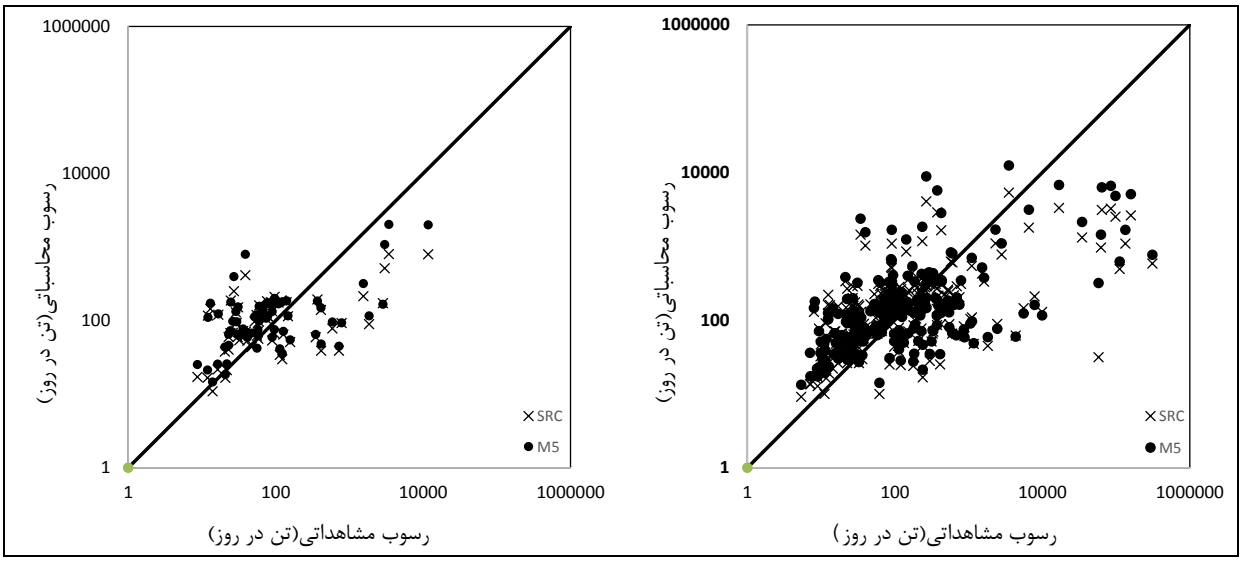
شکل(۴): مقادیر پیش بینی شده در مقابل مقادیر مشاهداتی در مرحله آزمون و صحت سنجی ایستگاه مشراگه

www.SID.ir

Archive of SID



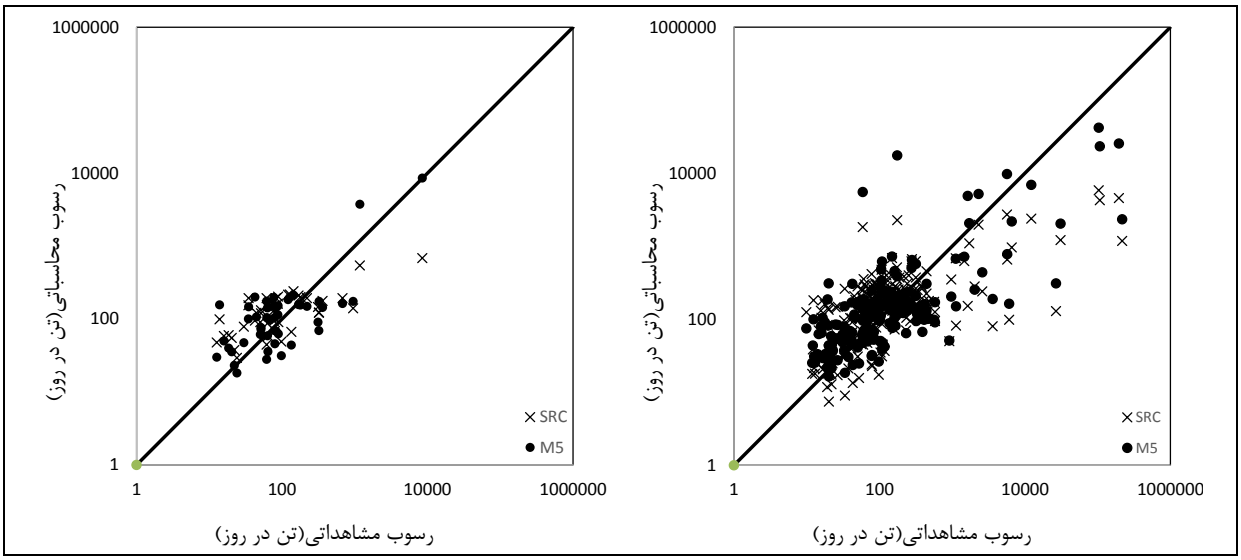
شکل(۵): مقادیر پیش بینی شده در مقابل مقادیر مشاهداتی در مرحله آزمون و صحت سنجی ایستگاه جوکنک



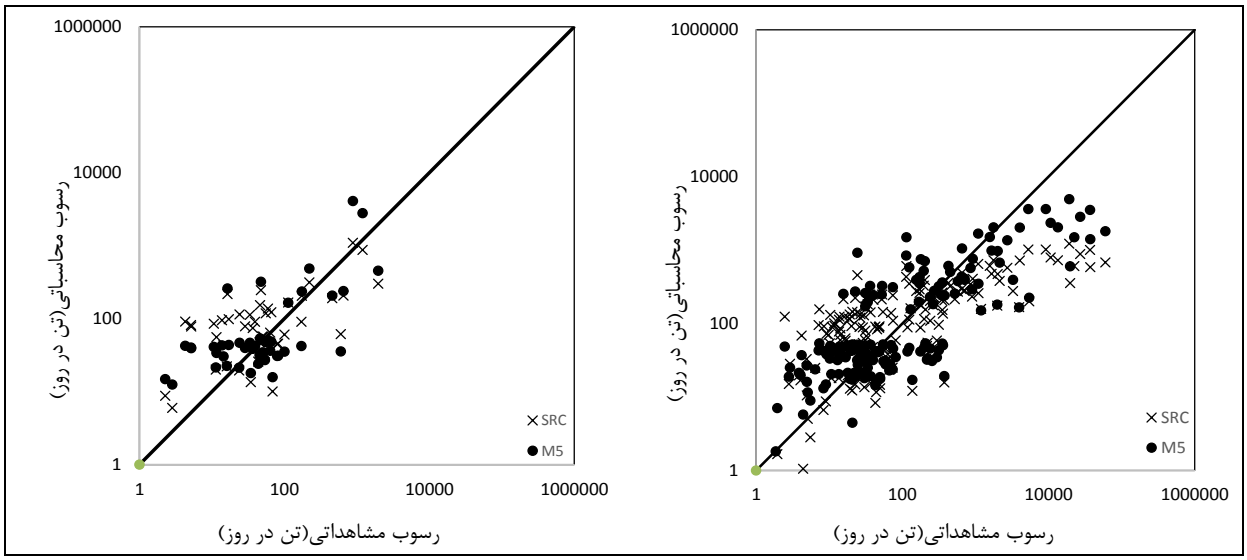
شکل(۶): مقادیر پیش بینی شده در مقابل مقادیر مشاهداتی در مرحله آزمون و صحت سنجی ایستگاه بهبهان

www.SID.ir

Archive of SID



شکل(۷): مقادیر پیش بینی شده در مقابل مقادیر مشاهداتی در مرحله آزمون و صحت سنجی ایستگاه چم نظام



شکل(۸): مقادیر پیش بینی شده در مقابل مقادیر مشاهداتی در مرحله آزمون و صحت سنجی ایستگاه شادگان

جهت بررسی دقیق‌تر نتایج حاصل از مدل درختی و مقایسه آن با منحنی سنجه از تحلیل‌های آماری استفاده شده است. نتایج تحلیل‌های آماری صورت گرفته در جدول ۲ نشان داده شده است. تحلیل‌های صورت گرفته نشان دهنده دقت بالاتر مدل M5 در مقایسه با منحنی سنجه رسوب که در تمام مطالعات فرسایش و رسوب مورد استفاده قرار می‌گیرد، می‌باشد. در جدول ۲ همانگونه که مشاهده می شود، مقدار RMSE در مدل M5 نسبت به منحنی سنجه مقادیر پایین‌تری را نشان می‌دهد. تغییرات این پارامتر زیاد بوده به گونه‌ای که در مورد ایستگاه مشراگه مقدار RMSE در مراحل آموزش و صحت‌سنجی به ترتیب به میزان ۴۵ و ۱۸ درصد و در مورد ایستگاه چم نظام ۱۴ و ۶ درصد نسبت به منحنی سنجه کاهش داشته است. در مورد سایر پارامترهای آماری محاسبه شده نیز همین روند مشاهده می‌گردد. از آن‌جا که معادلات ارائه شده توسط مدل M5 شباهت زیادی به معادلات ارائه شده توسط منحنی سنجه دارد، در بعضی از

www.SID.ir

Archive of SID

ایستگاه‌ها از قبیل ایستگاه بهبهان شباهت زیادی میان نتایج دو روش مشاهده می شود. این امر نشان دهنده قابلیت جایگزینی مدل M5 برای منحنی سنجه می‌باشد. بر اساس نتایج این جدول چنانچه مقدار RMSE جهت کلیه ایستگاه‌ها در نظر گرفته شود، مقدار آن در مدل درختی M5 برابر ۰/۵۹ و در منحنی سنجه رسوب برابر ۰/۷۴ خواهد بود که به میزان ۲۰٪ کاهش داشته است.

جدول (۲) نتایج خروجی از مدل درختی و منحنی سنجه رسوب در ایستگاه‌های مختلف

مدل	M5								منحنی سنجه							
	TRAIN				TEST				TRAIN				TEST			
ایستگاه	R^2	RMSE	MAE	SSE	R^2	RMSE	MAE	SSE	R^2	RMSE	MAE	SSE	R^2	RMSE	MAE	SSE
مشراگه	۰/۸۳	۰/۴۴	۰/۳۴	۲/۴۸	۰/۶۶	۰/۵۲	۰/۳۸	۳/۹۰	۰/۴۵	۰/۸۱	۰/۵۷	۸/۳۴	۰/۴۱	۰/۶۴	۰/۴۷	۶
جوکنک	۰/۸۰	۰/۶۲	۰/۴۸	۵/۳۱	۰/۶۵	۰/۷۱	۰/۵۴	۸/۴۵	۰/۷۴	۰/۷۲	۰/۵۷	۷/۰۶	۰/۶۰	۰/۷۷	۰/۵۹	۹/۷۶
بهبهان	۰/۳۳	۰/۷۷	۰/۵۵	۱۰/۸۷	۰/۳۰	۰/۶۲	۰/۴۷	۸/۶۸	۰/۲۸	۰/۸۰	۰/۵۷	۱۱/۶۰	۰/۲۵	۰/۶۴	۰/۵۰	۹/۳۹
چم نظام	۰/۵۳	۰/۵۹	۰/۴۱	۶/۵۶	۰/۴۴	۰/۷۴	۰/۴۷	۱۱/۳۲	۰/۳۶	۰/۶۹	۰/۴۸	۸/۸۹	۰/۱۴	۰/۷۹	۰/۵۱	۱۲/۵۰
شادگان	۰/۵۶	۰/۵۱	۰/۴۱	۸/۰۷	۰/۴۵	۰/۵۶	۰/۴۴	۱۰/۱۱	۰/۳۸	۰/۸۰	۰/۶۳	۱۲/۸۶	۰/۳۷	۰/۵۷	۰/۴۶	۱۰/۱۲

نتیجه گیری

هدف از این تحقیق ارائه روشی بر مبنای الگوریتم‌های درختی جهت محاسبه بار معلق رسوب در رودخانه می‌باشد. مدل درختی M5 برخلاف سایر روش‌های داده‌کاوی موجود از قبیل شبکه‌های عصبی مصنوعی و منطق فازی این مزیت را دارا می‌باشد که خروجی آن به صورت روابط رگرسیونی بوده و می‌توان در سایر مدل‌ها از آن استفاده کرد. علاوه بر این می‌توان مدل‌های ارائه شده توسط این الگوریتم را با مدل‌های تجربی موجود مقایسه کرد. در این تحقیق از الگوریتم درختی M5 جهت شبیه سازی بار معلق در ایستگاه‌های مشراگه، جوکنک، بهبهان، چم نظام و شادگان استفاده گردید. در این تحقیق دو پارامتر دبی جریان همان روز و دبی یک روز قبل به عنوان پارامترهای ورودی و دبی رسوب معلق به عنوان پارامتر خروجی به مدل معرفی گردید. نتایج بدست آمده از آنالیزهای مختلف آماری نشان داد که مقادیر RMSE مدل درختی M5 نسبت به منحنی سنجه مقادیر کمتری را نشان می‌دهد. این امر به ویژه در مورد ایستگاه مشراگه که مقدار RMSE آن در مدل درختی M5 در دو مرحله آموزش و صحت‌سنجی به ترتیب ۰/۴۴ و ۰/۵۲ و در مورد منحنی سنجه ۰/۸۱ و ۰/۶۴ به دست آمد به چشم می خورد. علاوه بر دقت بالاتر مدل درختی، در اغلب اوقات جهت برآورد بار رسوب با استفاده از منحنی سنجه می‌بایستی از دو یا سه رابطه استفاده نمود که برازش این روابط به صورت چشمی انجام شده و همراه با خطای زیادی است. این در حالی است که مدل M5 با دارا بودن ساختار درختی فضای مساله را به چند زیرشاخه تقسیم کرده و به ازای هر زیرشاخه یک مدل ارائه می‌دهد.

www.SID.ir

Archive of SID

منابع

اکبرزاده، م. ر.، ع. قزل سوفلو و م. حاجی بیگلو. ۱۳۹۳. ارزیابی و برآورد رسوب رودخانه اترک در محل ایستگاه‌های هیدرومتری دارای آمار با استفاده از روش‌های هیدورلوژیکی. مجله علوم و مهندسی آبیاری، دوره ۳۷، شماره ۴، ص ۱۳۲-۱۲۱.

جباری، ا و م. صمدی. ۱۳۹۲. کاربرد الگوریتم M5 در پیش بینی عمق آبشستگی در پایین‌دست سرریزها. طرح تحقیقات کاربردی، شرکت سهامی مدیریت منابع آب ایران.

دهقانی، ا.، م. ملک محمدی و ا. هزارجریبی. ۱۳۸۹. تخمین بار معلق رودخانه بهشت آباد با استفاده از شبکه عصبی مصنوعی. مجله پژوهش‌های حفاظت آب و خاک، شماره ۱۷، ص ۱۶۸-۱۵۹.

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، شماره ۲، ص ۶۱-۷۴.

ظهیری، ج. ۱۳۹۴. کاربرد مدل‌های ناپارامتریک CART و M5 در محاسبه عمق آبشستگی اطراف پایه‌های پل. فصلنامه علمی پژوهشی مهندسی آبیاری و آب ، شماره ۲۰، ص ۵۰-۳۵.

معتمدی‌وزیری، ب.، ح. احمدی، م. مهدوی و ف. شریفی. ۱۳۸۸. مقایسه روش‌های رگرسیون آماری و فازی جهت برآورد رسوب معلق رودخانه تلوار. فصلنامه جغرافیایی سرزمین علمی پژوهشی، شماره ۲۱، ص ۴۲-۳۱.

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

Bhattacharya, B. and D. P. Solomatine. 2006. Machine Learning in Sedimentation Modeling. Neural Networks Journal, 19: 208–214.

Etemad-Shahidi, A. and 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. and M. Taghipour. 2012. Predicting Longitudinal Dispersion Coefficient in Natural Streams Using M5′ Model Tree. Journal of Hydraulic Engineering, 138(6): 542-554.

Etemad-Shahidi, A. and N. Ghaemi. 2011. Model Tree Approach for Prediction of Pile Groups Scour due to Waves. Ocean Engineering, 38: 1522–1527.

Kumar, D., A. Pandey, N. Sharma and W. Flugel. 2016. Daily suspended sediment simulation using machine learning approach. Catena, 138: 77-90.

Mahjoobi, J., A. Sabzianpoor and E. Jabbari. 2010. Application of Meta-Heuristic Models for Local Scour Evaluation. Ninth International Conference on Computing Anticipatory Systems, AIP Conference Proceedings, 1303:389-397.

Quinlan, J. R.1992. Learning with Continuous Classes. Proceedings of AI'92, World Scientific, 343–348.

Wang, Y. and 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 and M. Taghi Aalami. 2009. Evaluation of total load sediment transport formulas using ann. International Journal of Sediment Research, 24: 274-286.

Zounemat-Kermani, M., O. Kişi, J. Adamowski and A. Ramezani-Charmahineh. 2016. Evaluation of data driven models for river suspended sediment concentration modeling, Journal of Hydrology, 535: 457-472.

www.SID.ir

Archive of SID

Application of M5 Model for Rivers Suspended Sediment Evaluation

Seyedeh Hadis Shahrokhi[1], Javad Zahiri[2*], Ahmad Jafary[3]

Abstract

Estimation of the suspended sediment load plays a very important role in water projects such as dam and surface water storage design, pollution and sediment transport in rivers and lakes, as well as design and maintenance the channels. The accuracy of conventional methods of sediment estimation, usually low and there are a big difference between their results. Therefore, in this study after preparing sediment and flow discharge data of hydrometric stations; Jow Kanak, Behbahan, Shadegan, Moshrageh and Cham Nezam, the M5 algorithm, has been used to determine the suspended sediment load. M5 tree algorithm is one of the newest data-driven models which divides the problem space into multiple branches and proposes equations for each branches. Flow discharge of each day and the day before have been used as an input variables and sediment discharge has been considered as an output variable to build and validate the tree model. Statistical analysis on tree model showed better compliance of tree model results with observed data in compare with sediment rating curve method. RMSE for training and validation phases at studied stations were 0.59 and 0.74 for M5 and rating curve methods, respectively. The results of this study showed that suspended sediment load could be accurately predicted using M5 tree algorithm. The results of this model, in addition to have an understandable and simple structure, could be used in practical issues as well.

Keywords: Data-driven models, Hydrometric stations, Sediment rating curve.

[1] M. Sc. Water structures, Khuzestan Ramin Agriculture and Natural Resources University, Sh.Hadis815@gmail.com
[2] Assistant Professor, Department of Water Engineering, Khuzestan Ramin Agriculture and Natural Resources University, Zahiri_Javad@yahoo.com
[3] Assistant Professor, Department of Water Engineering, Khuzestan Ramin Agriculture and Natural Resources University, Jafary_Ahmad@yahoo.com .* Corresponding author: Zahiri_Javad@yahoo.com, Tel: +98 916 653 1896

www.SID.ir