بررسی کارایی مدل درختان تصمیم‌گیری در برآورد رسوبات معلق رودخانه‌ای (مطالعه موردی: حوضه سد ایلام)

علی طالبی* و زینب اکبری[۱]

(تاریخ دریافت: ۱۳۸۹/۹/۱۴ ؛ تاریخ پذیرش: ۱۳۹۱/۲/۱۰)

چکیده

تخمین درست حجم رسوبات حمل شده توسط رودخانه‌ها در پروژه‌های آبی از اهمیت بسیاری برخوردار است. در حقیقت حصول روش‌هایی برای محاسبه دبی رسوبات مهم‌ترین هدف تحقیقات مربوط به فرآیند رسوب شده است. از جمله این روش‌ها می‌توان به روش‌های یادگیری ماشین از قبیل مدل درختان تصمیم‌گیری که مبتنی بر اصول یادگیری می‌باشند، اشاره کرد. روش درخت تصمیم یک روش سلسله مراتبی یا چند مرحله‌ای است که در آن به‌صورت بازگشتی مجموعه داده‌ها به روش دودویی به تقسیمات فرعی و کوچک‌تر تقسیم‌بندی می‌شود تا زمانی‌که تقسیمات فرعی نهایی نتوانند بیشتر از آن تجزیه شوند. درختان تصمیم استقرایی مجموعه‌ای از داده‌های معلوم را می‌گیرد و یک درخت تصمیم را از آن استنتاج می‌کند. سپس درخت می‌تواند به‌صورت مجموعه قوانینی برای پیش‌بینی ویژگی‌های معلوم استفاده شود. در این تحقیق کارایی این تکنیک در پیش‌بینی میزان آورد رسوبات رودخانه گل گل حوضه سد ایلام مورد بررسی قرار گرفته است. جهت ارزیابی دقت و صحت نتایج این مدل که در محیط برنامه‌نویسی MATLAB اجرا شده است معیارهای آماری r2, RMSE, BIAS, R و MAE مورد استفاده قرار گرفته است. نتایج نشان داده که به‌طورکلی و براساس تمامی معیارهای آماری ذکر شده، مدل درخت تصمیم‌گیری در مقایسه با روش مرسوم منحنی سنجه رسوب تطابق بسیار بیشتری با مقادیر اندازه‌گیری شده داشته و مهم‌ترین عامل ایجاد درخت که هم‌بستگی بالایی با مقادیر رسوب داشته است، دبی متناظر با رسوب و سپس بارش‌های روزانه بوده است.

واژه‌های کلیدی: بار معلق، درخت تصمیم‌گیری، حوضه سد ایلام

۱. گروه آبخیزداری، دانشکده منابع طبیعی، دانشگاه یزد

* : مسئول مکاتبات، پست الکترونیکی: talebisf@yazduni.ac.i r

مقدمه

هر ساله هزاران تن خاک از سطح حوضه‌های آبخیز توسط آب شسته شده و از محلی به محل دیگر انتقال می‌یابد. به‌طوری‌که در سال بیش از ۲۰ میلیارد تن رسوب توسط رودخانه‌های جهان انتقال می‌یابد و در آب‌های ساکن ته‌نشین می‌گردد. در ایران نیز سالانه بالغ بر ۱۰۰ میلیون مترمکعب از گنجایش سدهای مخزنی به خاطر رسوب‌گذاری کاسته می‌شود(۱). کشور ایران با دارا بودن رودخانه‌های متعدد پتانسیل بالایی جهت ایجاد سد در نقاط مختلف دارد که در موارد متعددی بهره‌وری مخزن سد به واسطه انباشتگی رسوبات کاهش یافته و یا به‌طورکلی غیرقابل استفاده شده است. از طرفی همه ساله سیل در نقاط مختلف کشور باعث ایجاد خسارت‌های فراوانی می‌گردد، که علت پدید آمدن سیل در موارد قابل توجهی، کاهش یافتن ظرفیت حمل آب توسط مقطع رودخانه به‌دلیل انباشتگی رسوبات می‌باشد (۳). لذا بررسی پدیده رسوب و برآورد رسوب حمل شده توسط رودخانه اهمیت خاصی خواهد داشت.

در خصوص مسأله پیش‌بینی آورد رسوب، عموماً با استفاده از مدل‌های فیزیکی هیدرولیکی روابط تجربی متعددی به‌دست آمده است. این روابط که براساس قوانین دینامیک و هیدرولیک استوارند، مطابق با خصوصیات هیدرولیک جریان، همانند دبی، سرعت، سطح مقطع و عمق جریان، دانه‌بندی و قطر ذرات رسوب، وزن مخصوص، لزجت و دمای آب به‌دست آمده‌اند، در حالی‌که برخی از این اطلاعات (مانند دانه‌بندی مواد بستر و یا دست کم اندازه D_{50} مواد بستر) در بسیاری از ایستگاه‌های هیدرومتری کشور برداشت نمی‌شوند و این امر استفاده از این روابط را دچار مشکل می‌کند. لذا استفاده و گسترش شیوه‌های نوین تخمین رسوبی که علاوه بر دقت کافی دارای سهولت کاربرد باشند در این میان نقش مهمی خواهد داشت. با توجه به اهمیت این موضوع، در سال‌های گذشته روش‌ها و مدل‌های مختلفی توسط محققان و متخصصان امر جهت ارزیابی، پیش‌بینی و تخمین رسوبات ارائه شده است. از جمله این روش‌ها می‌توان به روش‌های یادگیری ماشین از قبیل شبکه‌های عصبی مصنوعی و روش نوین درختان تصمیم‌گیری (مدل‌های درختی) که مبتنی بر اصول یادگیری می‌باشد، اشاره کرد. درختان تصمیم ابزاری است که قابلیت پاسخ‌گویی به مسائل پیچیده و غیرخطی را دارد و از آنجایی که در علوم دیگر از جمله برق و الکترونیک که اساساً شامل مسائل غیرخطی هستند به خوبی از عهده حل مسائل برآمده است لذا کم کم پای این ابزار جدید به علوم مهندسی آب و منابع طبیعی که جزء پدیده‌های غیرخطی و پیچیده هستند نیز باز شده است. سابقه استفاده از روش درختان تصمیم‌گیری در بررسی رسوبات تنها در یک تحقیق توسط بهاتاچاریا و همکاران (۴) گزارش شده است به‌طوری‌که آنها مدل درختان تصمیم را با چندین معادله تجربی برآورد رسوبات بارکف برای داده‌های آزمایشگاهی استفاده کردند و نتیجه گرفتند که مدل درختان تصمیم خطای بسیار کمتری نسبت به روش‌های تجربی دارد.

اما در زمینه‌های مختلف دیگر افرادی همچون محجوبی و صمغی(۲) و محجوبی و شهیدی (۹)، در ارتباط با پیش‌بینی ارتفاع موج شاخص، هیتوشی و همکاران(۸) در زمینه آنالیز حساسیت زمین لغزش در منطقه وسیع کوهستان‌های آکایشی ژاپن، روسجان و میکاس (۱۰)، نقش متقابل وضعیت‌های فصلی و هیدرولوژیکی آبشویی نیترات، هم‌چنین گیسن و همکاران (۶) در فرسایش آبی سطحی و زیرزمینی، زمازک و همکاران (۱۲) برای پیش‌بینی زمین لرزه و هندرسون (۷) در پیش‌بینی خصوصیات خاک گستره استرالیا، با استفاده از این روش تحقیقاتی را به انجام رسانده‌اند. نتایج این تحقیقات به‌طور عمده مدل درخت‌های تصمیم‌گیری را به‌عنوان ابزاری دقیق و قابل اعتماد معرفی کرده است. بنابراین استفاده از این روش به‌عنوان روشی نو و ابزاری قدرتمند و قابل اعتماد می‌تواند یاریگر مسئولان و مهندسان امور آب برای برنامه‌ریزی و استفاده بهینه از منابع آبی کشور باشد. هدف از تحقیق حاضر ارزیابی کارایی مدل درخت‌های تصمیم‌گیری رگرسیونی در مدل‌سازی و پیش‌بینی رسوبات معلق ایستگاه سرجوی رودخانه

گل گل حوضه سد ایلام می‌باشد. نتایج حاصل با روش مرسوم منحنی سنجه رسوب نیز مورد مقایسه قرار گرفت و تجزیه و تحلیل‌های لازم انجام پذیرفت.

مواد و روش‌ها

منطقه مورد مطالعه

حوضه آبخیز سد ایلام در جنوب شرقی شهرستان ایلام قرار داشته و از لحاظ تقسیمات سیاسی جزء محدوده شهرستان‌های ایلام و مهران می‌باشد. حوضه مورد مطالعه دارای مختصات ″۳۰ ′۲۰ °۴۶ تا ″۲۳ ′۲۹ °۴۶ تا ″۲۳ ′۲۹ °۴۶ طول شرقی و ″۳۳ ′۲۳ °۳۳ تا ″۳۱ ′۳۸ °۳۳ عرض شمالی است. شکل ۱ موقعیت جغرافیایی منطقه مورد مطالعه را نشان می‌دهد. آمار۳۳ ساله ایستگاه سینوپتیک، متوسط بارش سالانه را ۵۹۵/۵ میلی‌متر نشان می‌دهد. این حوضه در قسمت شمال غربی رشته کوه‌های زاگرس قرار گرفته است و با توجه به نقشه تقسیمات ساختمان ایران جزء واحد زاگرس چین‌خورده می‌باشد.

به‌طورکلی سنگ‌های موجود در سطح حوضه مورد بررسی، متعلق به دوران‌های دوم و سوم می‌باشد. قدیمی‌ترین سازند در حوضه، سازند سروک با سن آلبین و جدیدترین سازند نیز گچساران به سن میوسن آغازی می‌باشد. عموماً دو نوع سنگ در سطح حوضه وجود دارد سنگ‌های آهکی و سنگ‌های مارنی که به‌صورت متناوب با سنگ‌های شیلی و ماسه‌ای هم‌پوشانی دارند.

سنگ‌های مارنی و شیلی در سطح حوضه در مقابل عوامل فرسایش بسیار ضعیف عمل کرده و سالانه حجم زیادی از رسوب در حوضه تولید می‌کنند که در پشت سد ایلام جمع شده و عامل مؤثری در کاهش عمر مفید سد می‌باشند.

مواد و روش‌ها

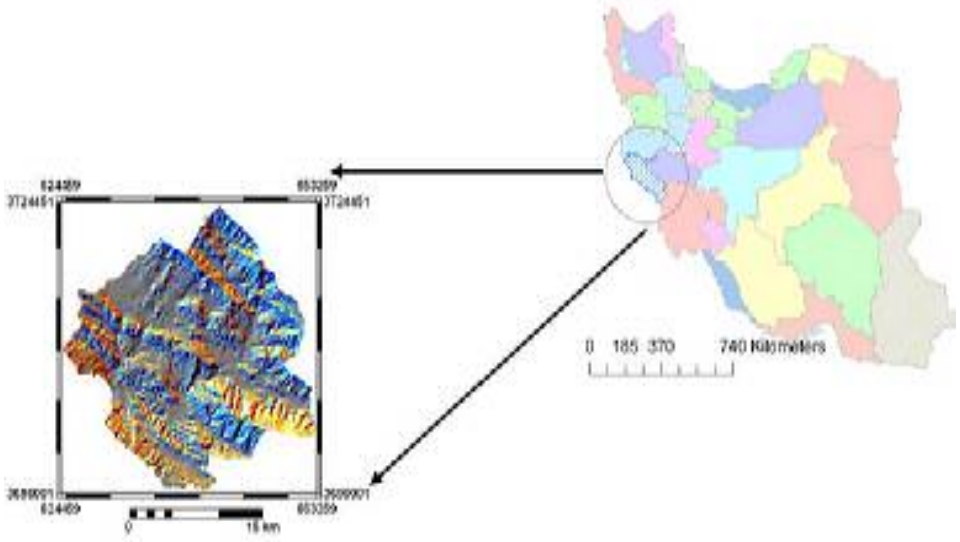
به منظور ساخت و ارزیابی مدل درختان تصمیم‌گیری در برآورد رسوبات معلق از آمار ایستگاه رسوب‌سنجی سرجوی واقع در رودخانه گل گل حوضه سد ایلام استفاده شد. این داده‌ها شامل دبی رسوب و دبی آب متناظر، بارش و دبی روزانه مربوط به دوره آماری از سال آبی ۶۸-۱۳۶۷ تا سال ۸۷-۱۳۸۶ می‌باشد. پس از پردازش داده‌ها، دو سوم آنها جهت آموزش و از یک سوم باقی‌مانده جهت آزمون کارایی مدل استفاده گردید. اطلاعات بیشتر در مورد مقادیر حداقل، حداکثر و میانگین پارامترهای مختلف آموزشی و آزمایشی در جداول ۱ و ۲ نشان داده شده است.

درخت تصمیم‌گیری

روش درخت تصمیم یک روش سلسله مراتبی یا چند مرحله‌ای است که در آن به‌صورت بازگشتی مجموعه داده‌ها به روش دودویی به تقسیمات فرعی و کوچک‌تر تقسیم‌بندی می‌شود تا زمانی‌که تقسیمات فرعی نهایی نتوانند بیشتر از آن تجزیه شوند. درختان تصمیم استقرایی مجموعه‌ای از داده‌های معلوم را می‌گیرد و یک درخت تصمیم را از آن استنتاج می‌کند. سپس درخت می‌تواند به‌صورت مجموعه قوانینی برای پیش‌بینی نتیجه ویژگی‌های معلوم استفاده شود. درختان تصمیم‌گیری به دو نوع طبقه‌بندی و رگرسیونی تقسیم می‌شوند به این صورت که اگر متغیر پاسخ مقداری گسسته داشته باشد به آن طبقه‌بندی و زمانی‌که درخت مقادیر پیوسته را پیش‌بینی کند نوع رگرسیونی را به آن نسبت می‌دهند. الگوریتم‌های مختلفی جهت ساخت درخت‌های تصمیم طبقه‌بندی و رگرسیونی توسط افراد مختلف معرفی شده است که در این تحقیق از الگوریتم (CART درختان رگرسیونی و طبقه‌بندی) و در محیط برنامه‌نویسی MATLAB استفاده شده است. الگوریتمCART درختان تصمیم دودویی را ایجاد می‌کند ساخت آن بر سه اصل استوار است (۵):

۱. مجموعه‌ای از سوالات به شکل $x \le d$? که در آن x یک متغیر و d یک مقدار ثابت است وجواب هر سوال بله/خیر است.
۲. بهترین معیار شکاف جهت انتخاب بهترین متغیر برای شکاف.
۳. ایجاد آمار خلاصه برای گره انتهایی.

تجزیه در درختان رگرسیونی براساس مجموع مربع خطا



شکل ۱. موقعیت منطقه مورد مطالعه

جدول ۱. مقادیر حداقل، حداکثر و میانگین پارامترهای مختلف(داده‌های آموزش)

پارامتر	حداقل	حداکثر	میانگین	واحد
دبی رسوب(QS)	۰/۰۹۳	۳۷۸۷۴۷/۲	۴۷۸۰/۹۶	Ton/day
دبی آب (متناظر با رسوب) (QW)	۰/۱۶	۹۸/۹۱	۳/۸۷	m3/s
دبی روزانه(Q)	۰/۱۶	۴۳/۱۹	۳/۶۱	m3/s
بارش روزانه(P)	۰	۵۲	۳/۰۱۹	mm

جدول ۲. مقادیر حداقل، حداکثر و میانگین پارامترهای مختلف(داده‌های آزمون)

پارامتر	حداقل	حداکثر	میانگین	واحد
دبی رسوب (Q_S)	۰/۰۶۸	۲۰۳۴۷۵/۹	۴۹۲۴	Ton/day
دبی آب (متناظر با رسوب) (Q_W)	۰/۱۴۶	۱۰۱/۵۶	۳/۸۴	m^3/s
دبی روزانه (Q)	۰/۱۴۶	۷۸/۳	۲/۹۳	m^3/s
بارش روزانه (P)	۰	۵۳	۵/۲	mm

(Sum of the Squared of Error, SSE) که از رابطه (۱) محاسبه می‌شود صورت می‌گیرد. مقدار SSE برای یک زیرمجموعه N مشاهده‌ای محاسبه می‌شود که در آن Oi هر مقدار تکی(فردی) از متغیر پاسخ است و $\bar{O}$ مقدار میانگین زیر مجموعه است.

$$SSE = \sum_{i=1}^{n} (o_i - \bar{o})^2 \quad [1]$$

نتیجه تقسیم باید به‌طور ایده‌آل منجر به مجموعه‌هایی شود که مقدار پاسخ به میانگین مجموعه نزدیک باشد. کمترین مقدار SSE برای مجموعه نشان‌دهنده نزدیک بودن مقادیر مجموعه به میانگین آن است. برای هر شکاف بالقوه مقدار SSE برای هر گره حاصل محاسبه می‌شود. امتیاز هر شکاف به‌وسیله مجموع مقادیر SSE محاسبه شده برای هر گره نتیجه به‌دست می‌آید.

بنابراین همه شکاف‌ها برای همه متغیرها محاسبه می‌شود سپس شکاف با کمترین امتیاز انتخاب می‌شود.

درختان ایجاد شده ممکن است نتایج بسیار پیچیده‌ای را ناشی شوند و شامل صدها سطح شوند، بنابراین قبل از استفاده از آنها برای طبقه‌بندی داده جدید باید بهینه‌سازی شوند. بهینه‌سازی درخت به معنای انتخاب اندازه صحیح درخت، برش گره‌های بی اهمیت و حتی زیر درختان است. در عمل دو الگوریتم هرس ممکن است که مورد استفاده قرار گیرد: بهینه‌سازی به‌وسیله تعداد نقاط در هر گره و اعتبارسنجی که در این تحقیق از روش اعتبارسنجی استفاده گردیده است (۱۱).

در این پژوهش علاوه بر درختان تصمیم‌گیری از روش مرسوم منحنی سنجه رسوب نیز جهت برآورد رسوبات معلق برای ایستگاه مذکور و دوره زمانی مربوطه استفاده شده و نتایج دو روش مورد ارزیابی و مقایسه قرار گرفته است. برای ارزیابی نتایج روش‌های مورد استفاده از معیارهای آماری‌ضریب هم‌بستگی(R^2)، جذر میانگین مربعات خطا (RMSE)، ضریب نش (r^2)، انحراف(Bias)، و خطای مطلق(MAE) طبق روابط ارائه شده زیر استفاده شده است.

$$R^2 = \frac{\sum (o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum (o_i - \bar{o})^2 \sum (p_i - \bar{p})^2}} \quad [2]$$

$$RMSE = \left(\frac{\sum_{i=1}^{n} (p_i - o_i)^2}{n} \right)^{1/2} \quad [3]$$

$$r^2 = 1 - \left(\frac{\sum_{i=1}^{n} (o_i - p_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2} \right) \quad [4]$$

$$Bias = \bar{p} - \bar{o} \quad [5]$$

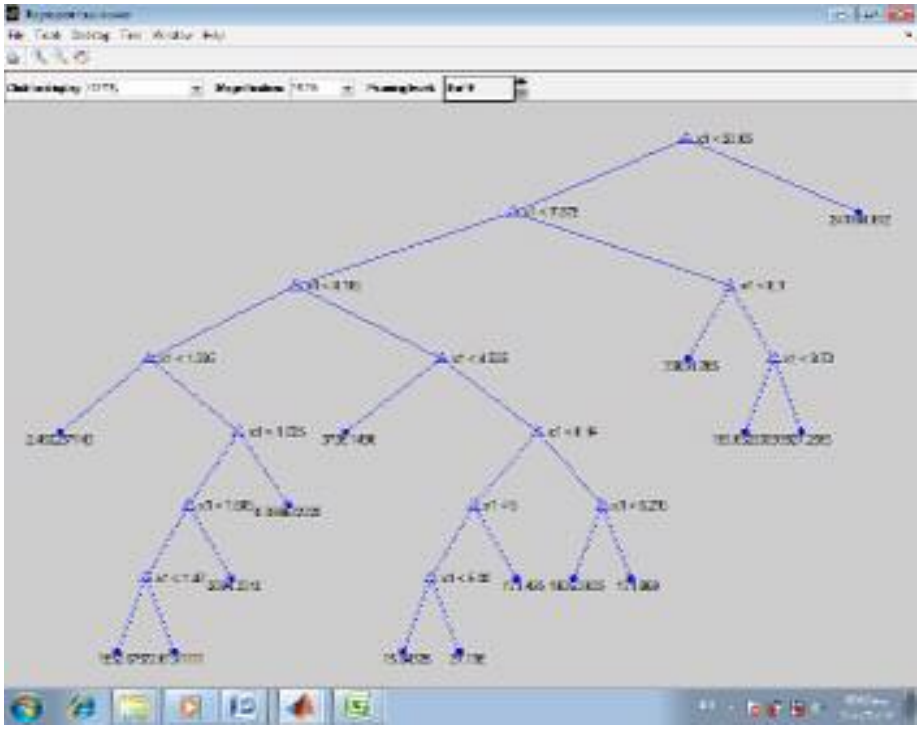
$$SSE = \frac{\sum |o_i - p_i|}{n} \quad [6]$$

در روابط بالا، Oi مقادیر مشاهداتی، Pi مقادیر برآورد شده، $\bar{o}$ میانگین مقادیر مشاهداتی، $\bar{p}$ میانگین مقادیر برآوردی و n تعداد مشاهدات می‌باشد.

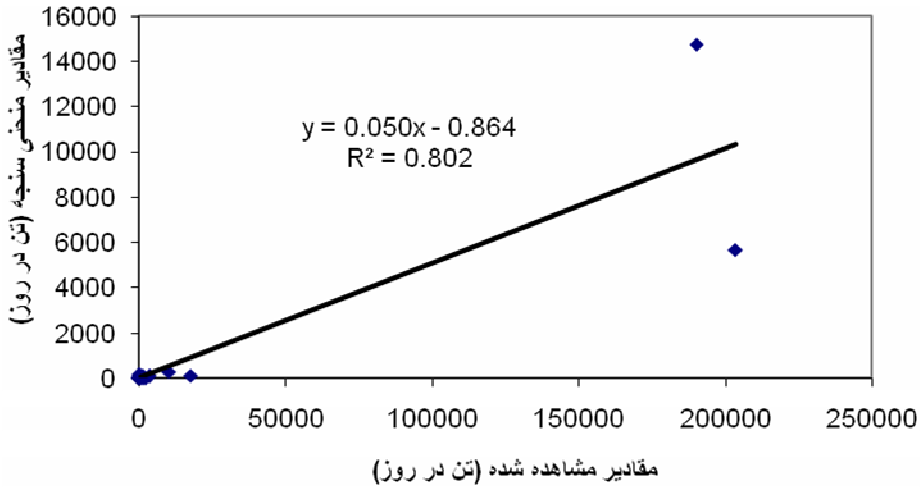
نتایج و بحث

از آنجا که در روش منحنی سنجه رسوب تنها از پارامتر دبی متناظر جهت برآورد رسوبات استفاده می‌گردد لذا به منظور بالا بردن دقت مقایسه، در مرحله اول، جهت ایجاد درخت تصمیم از دو پارامتر دبی آب و دبی رسوب متناظر استفاده گردیده است. سپس بر روی درخت ساخته شده که درخت حداکثر (شکل۲) نامیده می‌شود عملیات هرس جهت جلوگیری از پدیده بیش برازش (Overfitting) صورت گرفته است. در این مرحله درخت ایجاد شده جهت حذف شاخه‌ها و زیرشاخه‌های اضافه‌ای که منجر به ظهور خطا در مرحله آزمون می‌شود مورد بررسی قرار گرفته است. نتایج نشان داده که درخت ساخته شده نیاز به حذف هیچ شاخه و زیرشاخه‌ای نداشته و درخت اولیه کمترین خطا را در برآورد نشان داده است. در نهایت با استفاده از درخت نهایی (شکل۲) میزان رسوب برای هر ورودی دبی متناظر برآورد شده است. میزان هم‌بستگی بین مقادیر برآورد شده توسط درخت تصمیم‌گیری و مقادیر مشاهداتی که در شکل۳ نشان داده شده است برابر با ۹۸/. می‌باشد. در ادامه با استفاده از منحنی سنجه رسوب (شکل۴) که با مجموعه داده‌های آموزشی ترسیم شده است، مقادیر رسوب برآورد گردیده است. همان‌طورکه در شکل ۵ نیز نشان داده شده است ضریب تشخیص مقادیر مشاهداتی و مقادیر برآورد شده توسط منحنی سنجه برابر با ۰/۸می‌باشد. جهت ارزیابی بیشتر دقت دو روش از معیارهای آماری R، r^2, RMSE, BIAS, MAE استفاده گردید که نتایج در جدول ۳ نشان داده شده است.

همان‌طور که در جدول۳ نشان داده شده است، مقادیر به‌دست آمده برای تمامی پارامترهای خطا، مدل درخت تصمیم‌گیری را برتر از روش منحنی سنجه رسوب نشان می‌دهد. هم‌چنین مقایسه مقادیر مشاهداتی و مقادیر برآورد شده توسط هر دو روش که در شکل ۶ مشاهده می‌شود، نشان می‌دهد که روش منحنی سنجه بار معلق حوضه را به خصوص در مواقع سیلابی که در برنامه‌ریزی‌ها بسیار حائز اهمیت بوده بسیار کمتر از مقدار مشاهداتی برآورد کرده درحالی‌که مدل



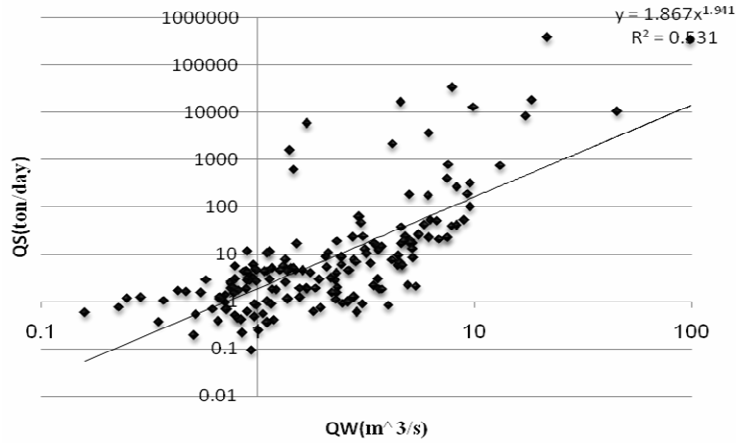
شکل۲. درخت تصمیم ایجاد شده با استفاده از مجموعه داده‌های آموزش(دبی و رسوب متناظر)



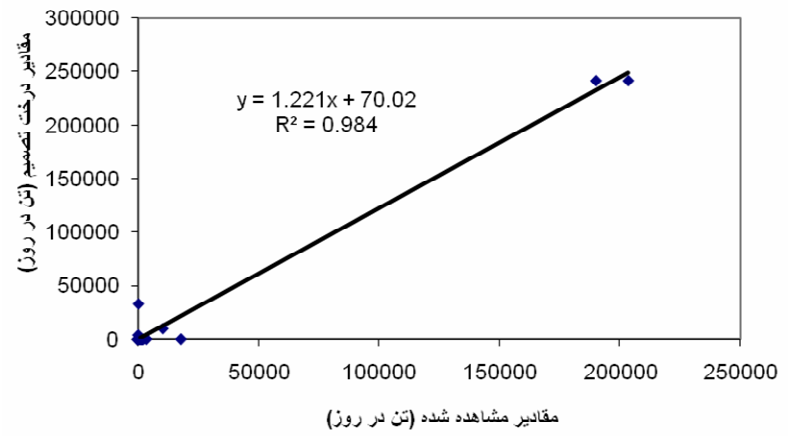
شکل۳. میزان هم‌بستگی بین مقادیر برآورد شده توسط درخت تصمیم‌گیری و مقادیر مشاهداتی

درخت تصمیم مقدار رسوبات را بهتر از روش منحنی سنجه و نزدیک‌تر به مقادیر مشاهداتی پیش‌بینی کرده که این موضوع در دبی‌های اوج نمود بیشتری داشته است.

در مرحله بعد جهت بهبود نتایج حاصل از درختان تصمیم، داده‌های بارش و دبی روزانه نیز به‌عنوان متغیر ورودی وارد مدل شده است. به این صورت که با ثابت بودن دبی متناظر(Qw) در هر بار اجرای مدل، ابتدا بارش همان روز(P) وارد مدل شده و دبی به‌ترتیب از دبی همان روز(Q) تا



شکل۴. منحنی سنجه رسوب حوضه مورد مطالعه

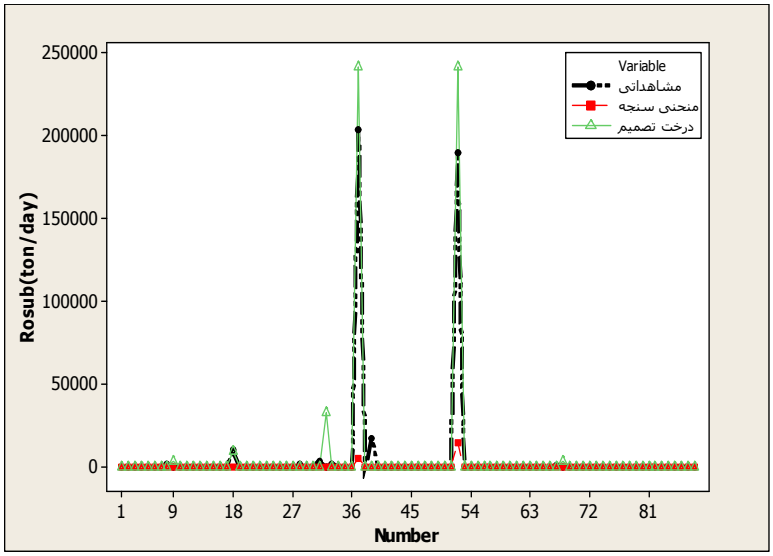


شکل۵. میزان هم‌بستگی بین مقادیر برآوردی توسط منحنی سنجه و مقادیر مشاهداتی

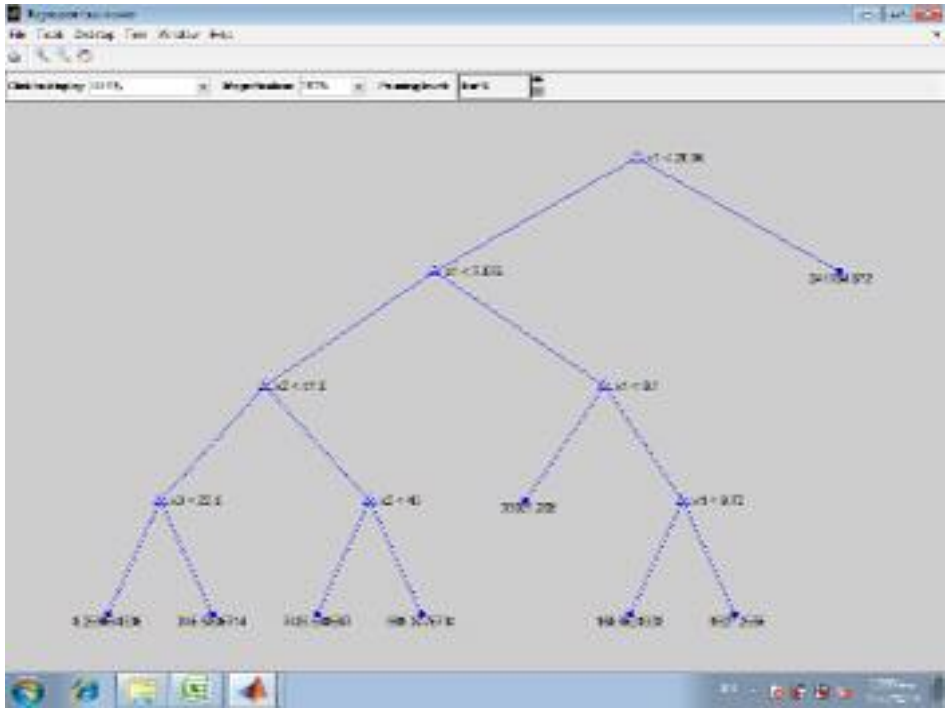
جدول۳. مقادیر معیارهای مختلف خطا در برآورد بار معلق برای داده‌های آزمون

پارامتر	R	r^2	RMSE	MAE	BIAS
منحنی سنجه رسوب	۰/۹	۰/۰۷	۴۳۸۵۸/۵	۴۶۷۸/۷	-۴۶۷۵/۳
درخت تصمیم	۰/۹۹	۰/۹۳	۱۰۸۹۰/۴	۱۸۰۷/۹۳	-۱۱۶۰/۹

مجموع چهار روز قبل(Q4) تغییر داده شده است. سپس ورودی بارش را به مجموع بارش همان روز(P) و روز قبل(P1) تغییر داده و دبی مطابق حالت قبل یعنی از دبی همان روز تا مجموع چهار روز(Q+Q1+...Q4) قبل تغییر داده شده است. این روند تا مجموع بارش پنج روز قبل(P+P1+...P5)، که جمعاً شامل ۳۰ حالت مختلف که در جدول۴ آمده است ادامه یافته و معیارهای آماری برای ارزیابی دقت در تمام حالات محاسبه گردیده است. از بین تمامی ترکیبات وارد شده بهترین ترکیب که کمترین خطا را تولید کرده به‌عنوان بهترین ورودی‌ها در نظر گرفته شده است. به این ترتیب پس از تحلیل‌های صورت گرفته بهترین ترکیب شامل دبی متناظر، مجموع بارش چهار روز قبل و مجموع دبی همان روز و روز قبل تعیین گردیده است. مطابق مراحل ساخت مدل درختی عملیات هرس روی درخت ایجاد شده با بهترین ترکیب ورودی (شکل۷)



شکل۶. سری زمانی مقادیر مشاهداتی، مقادیر برآوردی توسط منحنی سنجه و درخت تصمیم



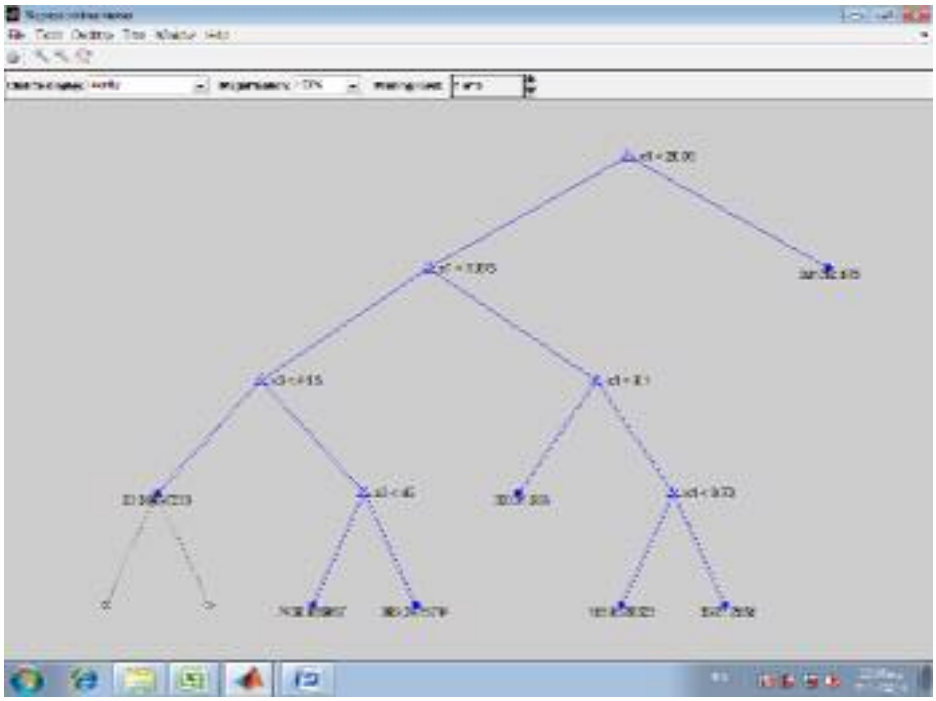
شکل۷. درخت تصمیم ایجاد شده با استفاده از بهترین ترکیب ورودی

صورت گرفته و پارامترهای خطا برای درخت هرس شده (شکل۸) نیز محاسبه شده است. نتایج نشان داده که هرس درخت در اولین سطح نسبت به سطوح دیگر خطای کمتری را در برآورد نشان می دهد. اما همان‌طورکه در جدول ۴ نیز مشاهده می‌شود درخت هرس شده در برخی از پارامترها (MAE) نا مناسب‌تر از درخت اولیه بوده ولی به‌طورکلی نتایج بهتری را ایجاد کرده است. سری زمانی مقادیر برآورد شده توسط بهترین ترکیب و مقادیر مشاهداتی در شکل۹ نشان داده شده است.

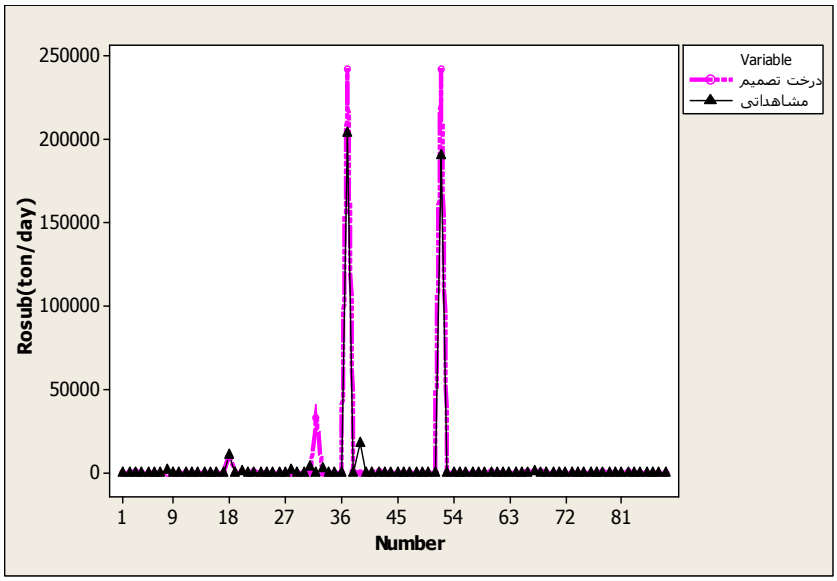
از مقایسه نتایج پارامترهای آماری برای درخت تصمیمی که

جدول۴. نتایج معیارهای آماری مربوط به مدل درخت تصمیم در ترکیبات ورودی مختلف

	ورودی	R	r^2	RMSE	MAE	BIAS
۱	$Q_{W,\ Q,\ P}$	۱	۰/۹۳	۱۳۶۴۲	۱۹۸۵	۱۴۵۴
۲	$Q_{W,\ (\ Q+Q1),\ P}$	۱	۰/۹۳	۱۳۶۲۲	۱۹۸۶	۱۴۵۲
۳	$Q_{W,\ (\ Q+Q1+Q2),\ P}$	۱	۰/۹۳	۱۳۶۳۶	۱۹۸۷	۱۴۵۳
۴	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ P}$	۱	۰/۹۳	۱۳۶۳۸	۱۹۸۵	۱۴۵۳
۵	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ P}$	۱	۰/۹۳	۱۴۲۰۹	۲۰۲۰	۱۵۱۴
۶	$Q_{W,\ Q,\ (P+P1)}$	۰/۹۹	۰/۹۳	۱۴۷۱۳	۲۳۳۴	۱۵۶۸
۷	$Q_{W,\ (\ Q+Q1),\ (P+P1)}$	۰/۹۹	۰/۹۳	۱۴۷۱۳	۲۳۳۴	۱۵۶۸
۸	$Q_{W,\ (\ Q+Q1+Q2),\ (P+P1)}$	۰/۹۹	۰/۹۳	۱۴۷۱۳	۲۳۳۴	۱۵۶۸
۹	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ (P+P1)}$	۰/۹۹	۰/۹۳	۱۴۷۱۳	۲۳۳۴	۱۵۶۸
۱۰	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ (P+P1)}$	۰/۹۹	۰/۹۳	۱۴۷۱۳	۲۳۳۴	۱۵۶۸
۱۱	$Q_{W,\ Q,\ (P+P1+P2)}$	۱	۰/۹۳	۱۲۸۵۳	۱۹۳۱	۱۳۷۰
۱۲	$Q_{W,(\ Q+Q1),\ (P+P1+P2)}$	۱	۰/۹۴	۱۱۴۵۸	۱۷۸۶	۱۲۲۱
۱۳	$Q_{W,\ (\ Q+Q1+Q2),\ (P+P1+P2)}$	۱	۰/۹۴	۱۲۹۶۹	۱۹۴۶	۱۳۸۲
۱۴	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ (P+P1+P2)}$	۱	۰/۹۴	۱۲۹۶۹	۱۹۴۶	۱۳۸۲
۱۵	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ (P+P1+P2)}$	۱	۰/۹۴	۱۳۰۴۲	۱۹۴۷	۱۳۹۰
۱۶	$Q_{W,\ Q,\ (P+P1+P2+P3)}$	۰/۹۹	۰/۹۳	۱۳۲۵۹	۲۰۱۶	۱۴۱۳
۱۷	$Q_{W,\ (\ Q+Q1),\ (P+P1+P2+P3)}$	۰/۹۹	۰/۹۳	۱۳۲۵۹	۲۰۱۶	۱۴۱۳
۱۸	$Q_{W,\ (\ Q+Q1+Q2),\ (P+P1+P2+P3)}$	۰/۹۹	۰/۹۳	۱۳۲۵۹	۲۰۱۶	۱۴۱۳
۱۹	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ (P+P1+P2+P3)}$	۰/۹۹	۰/۹۳	۱۳۲۵۹	۲۰۱۶	۱۴۱۳
۲۰	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ (P+P1+P2+P3)}$	۰/۹۹	۰/۹۳	۱۳۲۵۹	۲۰۱۶	۱۴۱۳
۲۱	$Q_{W,\ Q,\ (P+\ldots+P4)}$	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
۲۲	$Q_{W,(\ Q+Q1),\ (P+\ldots+P4)}$	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
۲۳	$Q_{W,\ (\ Q+Q1+Q2),\ (P+\ldots+P4)}$	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
۲۴	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ (P+\ldots+P4)}$	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
۲۵	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ (P+\ldots+P4)}$	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
۲۶	$Q_{W,\ Q,\ (P+\ldots+P5)}$	۰/۹۹	۰/۹۳	۱۱۶۲۵	۱۸۳۴	۱۲۳۹
۲۷	$Q_{W,(\ Q+Q1),\ (P+\ldots+P5)}$	۰/۹۹	۰/۹۳	۱۱۶۲۵	۱۸۴۳	۱۲۳۹
۲۸	$Q_{W,\ (\ Q+Q1+Q2),\ (P+\ldots+P5)}$	۰/۹۹	۰/۹۳	۱۱۶۲۵	۱۸۴۳	۱۲۳۹
۲۹	$Q_{W,\ (\ Q+Q1+\ldots+Q3),\ (P+\ldots+P5)}$	۰/۹۹	۰/۹۳	۱۱۶۲۵	۱۸۴۳	۱۲۳۹
۳۰	$Q_{W,\ (\ Q+Q1+\ldots+Q4),\ (P+\ldots+P5)}$	۰/۹۹	۰/۹۳	۱۱۶۲۵	۱۸۴۳	۱۲۳۹



شکل۸ .درخت تصمیم نهایی (هرس شده) ایجاد شده با استفاده از بهترین ترکیب ورودی



شکل۹. سری زمانی مقادیر برآورد شده توسط بهترین ترکیب ورودی و مقادیر مشاهداتی

در ساخت آن فقط از دبی متناظر استفاده شده (جدول۳) با جدول ۵، که دبی و بارش روزانه نیز وارد مدل شده است می‌توان مشاهده کرد که ورود پارامترهای بارش و دبی روزانه به مدل منجر به افزایش دقت در برآورد شده است. جدول ۶ درصد افزایش دقت پارامترهای آماری را درحالتی‌که دبی و بارش روزانه وارد مدل شده است نسبت به حالتی که تنها از دبی متناظر استفاده گردیده است را نشان می‌دهد. اما همان‌طورکه از روی شکل۸ (در شکل۸، x1 دبی متناظر و x3 بارش‌های روزانه می‌باشند) نیز نمایان است دبی‌های روزانه تأثیر چشم‌گیری در ایجاد درخت نداشته و مهم‌ترین عامل ایجاد درخت که هم‌بستگی بالایی با مقادیر رسوب داشته است، دبی متناظر با رسوب و سپس بارش‌های روزانه بوده است.

همان‌طورکه در بررسی منابع ذکر شده تنها مرجع موجود در مورد استفاده از درختان تصمیم‌گیری در مطالعات رسوب

جدول۵. مقادیر معیارهای مختلف خطا در برآورد بار معلق برای بهترین ترکیب ورودی

پارامتر	R	r^2	RMSE	MAE	BIAS
درخت تصمیم اولیه	۰/۹۹	۰/۹۳	۱۰۷۲۱	۱۷۵۹	۱۱۴۲
درخت تصمیم هرس شده	۰/۹۹	۰/۹۳	۱۰۵۹۹/۳	۱۷۶۱/۵	۱۱۲۹/۸۹

جدول۶. میزان افزایش دقت پارامترهای خطا با ورود داده‌های بارش و دبی روزانه

پارامتر	R	r^2	RMSE	MAE	BIAS
٪ افزایش دقت	۰	۰	۲/۹	۰/۴۶	۰/۳۱

تحقیقی است که توسط بهاتاچاریا و همکاران (۴) انجام شده است، آنها در مطالعه خود از دو روش یادگیری ماشینی (Machine learning)، شبکه عصبی مصنوعی و مدل‌های درختی جهت مدل‌سازی انتقال بار کل و بار کف استفاده کرده‌اند و مدل‌های انتقال بار کف را با مدل‌های بگنولد، اینشتین، پارکر و همکاران و وان ریجن، و مدل انتقال بار کل را با مدل‌های وایت وایکر، بگنولد، انگلوند و هانسون و وان ریجن از نظر RMSE، میانگین درصد خطا و نسبت مقادیر انتقال اندازه‌گیری شده و محاسبه شده مقایسه کرده‌اند. طبق نتایجی که آنها گرفته‌اند مدل‌های یادگیری ماشینی دقت بیشتری نسبت به دیگر روش‌ها داشته و از بین مدل‌های یادگیری ماشینی، مدل درختی کمترین خطا را نشان داد که با نتایج این تحقیق که حاکی از برتری مدل درختان تصمیم‌گیری رگرسیونی نسبت به روش مرسوم منحنی سنجه رسوب است مطابقت دارد.

نتیجه‌گیری

برای انجام این تحقیق که در راستای اهمیت موضوع و با هدف ارزیابی کارایی مدل درختان تصمیم‌گیری در برآورد رسوبات معلق رودخانه گل گل حوضه سد ایلام انجام شده از آمار۲۰ ساله ایستگاه هیدرومتری و رسوب‌سنجی سرجوی استفاده گردیده است. داده‌های ورودی به مدل شامل دبی و رسوب متناظر، دبی‌های روزانه و بارش‌های روزانه بوده است. در حالت‌های مختلف و با تغییر داده‌های ورودی، مدل مورد ارزیابی قرار گرفته و در نهایت نتایج با روش مرسوم منحنی سنجه مقایسه گردیده است. نتایج نشان داده که براساس تمامی ضرایب آماری مورد استفاده جهت ارزیابی دقت برآورد انجام شده (ضرایب R ،r^2 ,RMSE BIAS و MAE)، مدل درخت تصمیم‌گیری به‌ویژه در مواقع سیلابی مقدار رسوبات را با دقت بسیار بیشتری نسبت به منحنی سنجه برآورد کرده است. به‌طوری‌که مقدار R و RMSE محاسبه شده در درخت تصمیم‌گیری به‌ترتیب برابر با ۰/۹۹و ۱۰۸۹۰/۴ بوده در حالی‌که منحنی سنجه به‌ترتیب مقادیر ۰/۹ و ۴۳۸۵۸/۵ را نشان داده است. با ورود داده‌های بارش و دبی روزانه به مدل دقت مدل افزایش یافته به گونه‌ای که RMSE به میزان ۲/۹ و MAE به میزان ۰/۴۶ درصد کاهش نشان داده است. هم‌چنین نتایج بیانگر این مطلب است که بعد از دبی متناظر با رسوب، بارندگی‌های روزانه بیشترین تأثیر را بر میزان رسوبات داشته است.

پیشنهادات

- ✓ جهت واسنجی دقیق‌تر مدل درختان تصمیم‌گیری توصیه می‌شود در چند ایستگاه با شرایط محیطی مختلف این مدل اجرا شده تا در صورت اثبات دقت این مدل در برآورد رسوبات معلق بتوان این روش را به‌عنوان یک روش پایه در مطالعات هیدرولوژی معرفی کرد .
- ✓ پیشنهاد می‌گردد از سایر الگوریتم‌های ایجاد درخت تصمیم‌گیری برای برآورد رسوبات معلق استفاده شود و نتایج حاصل با یکدیگر مورد مقایسه قرار گیرد.

منابع مورد استفاده

۱. کاظمی، ی.۱۳۸۷. برآورد نسبت بار بستر به بار معلق در رودخانه‌های البرز مرکزی (مطالعه موردی رودخانه‌های طالقان و جاجرود). پایان‌نامه کارشناسی‌ارشد، دانشگاه تهران.

۲. محجوبی،ج وا، صمغی.۱۳۸۸. پیش‌بینی پارامترهای ناشی از امواج باد با استفاده از روش درختان تصمیم رگرسیونی و شبکه‌های عصبی مصنوعی. نشریه مهندسی دریا ۵(۹): ۷۱-۶۵.

۳. نظم آرا ،ح. ۱۳۸۵. پیش‌بینی بار معلق رودخانه توسط شبکه عصبی مصنوعی (مطالعه موردی رودخانه آجی چای). پایان‌نامه کارشناسی‌ارشد، دانشگاه تبریز.

4. Bhattacharya, B., R.K. Price and D. P. Solomatine. 2007. Machine learning approach to modeling sediment transport. J. Hydraul. Eng. 133(4): 440-450.
5. Breiman, L., J. Friedman, R. Olshen and C. Stone. 1984. Classification and Regression Trees. Chapman & Hall/CRC Press, Boca Raton, FL,.
6. Geissen, V. and C. Kampichler. 2007. Superficial and subterranean soil erosion in Tabasco, tropical mexico: development of a decision tree modeling approach .Geoderma 139: 277-287.
7. Henderson, B.L., E.N. Bui, C.J. Moran and D.A.P. Simon. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124: 383–398.
8. Hitoshi, S., D. Nakayama and H. Matsuyama. 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan.
9. Mahjoobi, J. and E. Shahidi. 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. Appl. Ocean Res. 30: 172-177.
10. Rusjan, S. and M. Micos. 2008. Assessment of hydrolpgical and seasonal controls over the nitrate flushing from a forested watershed using a data mining technique. Hydrol.Earth Sys. Sci. 12: 645-656.
11. Timofeev, R. 2004. Classification and regression trees (CART), Theory and applications. CASE - Center of Applied Statistics and Economics Humboldt University, Berlin.
12. Zmazek, B., L. Todorovski, S. Dmeroski, J. Vaupoti and I. Kobal. 2003. Application of decision trees to the analysis of soil radon data for earthquake prediction. Appl. Radiation and Isotopes 58: 697–706.

J. Sci. & Technol. Agric. & Natur. Resour., Water and Soil Sci., Vol. 17, No. 63, Spring 2013, Isf. Univ. Technol., Isf., Iran.

Investigation of Ability of Decision Trees Model to Estimate River Suspended Sediment (Case Study: Ilam Dam Basin)

A. Talebi* and Z. Akbari [1]

(Received : Dec. 5-2010 ; Accepted : Apr. 29-2012)

Abstract

The real estimation of the volume of sediments carried by rivers in water projects is very important. In fact, achieving the most important ways to calculate sediment discharge has been considered as the objective of the most research projects. Among these methods, the machine learning methods such as decision trees model (that are based on the principles of learning) can be presented. Decision tree method is a hierarchical multi step method which is a recursive data collection technique to binary and smaller sub-divisions until the final analysis cannot be divided. Decision trees consider a priori known set of data and derive a decision tree from it. Then, tree can be used as the set of laws to predict unknown features. In this research, the efficiency of this technique for predicting the suspended sediments in Ilam dam basin has been investigated. To evaluate the accuracy of the methods (written by MATLAB software), statistical criteria such as R, BIAS, RMSE, r^2 and MAE were computed. The results showed that based on all the statistical criteria, decision tree in comparison with the sediment rating curve had most consistency with the observed data. Meanwhile, the most important factors for creating tree in the model (that had high correlation with sediment data) are the corresponding discharge and daily rainfall.

Keywords: Suspended sediment, Decision trees model, Ilam dam basin.

1. Dept. of Watershed Manage., College of Natur. Resour., Yazd Univ., Yazd, Iran.
*: Corresponding Author, Email: talebisf@yazduni.ac.i r