



دانشگاه گنجینه‌های آب و خاک

نشریه پژوهش‌های حفاظت آب و خاک

جلد بیست و چهارم، شماره چهارم، ۱۳۹۶

<http://jwsc.gau.ac.ir>

مدل‌سازی تصادفی بار رسوب با استفاده از جنگل تصادفی و رگرسیون چندک

مهسا جوادی علی‌نژاد^۱، سیدمرتضی سیدیان^۲، حامد روحانی^۲ و ابوالحسن فتح‌آبادی^۲

^۱ دانش‌آموخته کارشناسی ارشد گروه آبخیزداری، دانشگاه گنبد کاووس، ^۲ استادیار گروه آبخیزداری، دانشگاه گنبد کاووس

تاریخ دریافت: ۹۵/۱۱/۱؛ تاریخ پذیرش: ۹۶/۷/۲۱

چکیده

سابقه و هدف: ارزیابی بار معلق رسوبی رودخانه‌ها بسیار مهم است. کیفیت آب رودخانه‌ها و محیط‌زیست تحت تأثیر بار رسوب می‌باشد. همچنین طراحی سازه‌های هیدرولیکی و سایر تأسیسات آبرسانی، مدیریت حوزه آبخیز و اجرای برنامه‌های حفاظت خاک و مشکلات عمده دیگر ناشی از آورد رسوب رودخانه‌ها به تخمین صحیح بار رسوب وابسته است. از آنجایی که برآورد مستقیم بار رسوبی بسیار دشوار و وقت‌گیر است، بنابراین این امر سبب شد پژوهشگران به برآورد غیرمستقیم بار رسوبی که به روش‌های گوناگون امکان‌پذیر است روی آورند. یکی از راه‌های آسان برآورد غیرمستقیم بار رسوبی، منحنی سنج رسوب است. این روش تنها می‌تواند معرف یک مقدار رسوب در یک دبی معین باشد و به علت عوامل مختلفی در طبیعت ممکن است چندین مقدار بار رسوبی برای یک دبی مشخص وجود داشته باشد. بر این اساس در پژوهش حاضر از روش‌های رگرسیون چندک و جنگل تصادفی که بتوانند مقدار بار رسوب را برای یک مقدار دبی معین در احتمالات مختلف تخمین بزنند استفاده شد. با به‌کارگیری این دو روش می‌توان احتمال وقوع بار رسوب در رویدادهای استثنایی و سیلاب‌های عظیم را تحلیل کرد.

مواد و روش‌ها: در این پژوهش از مدل‌های منحنی سنج رسوب، رگرسیون چندک و جنگل تصادفی به منظور برآورد بار رسوب چهار ایستگاه جنگلده، نوده، ارازکوسه و قزاقلی واقع در رودخانه گرگانرود در استان گلستان استفاده گردید. به این منظور داده‌های دبی-رسوب متناظر چهار ایستگاه مورد مطالعه به دو بخش ۷۵٪ برای آموزش و ۲۵٪ برای آزمون تفکیک شدند. در روش منحنی سنج، مقدار رسوب با استفاده از معادله توانی برازش داده شده بین دبی و رسوب متناظر، حاصل گردید. الگوریتم‌های رگرسیون چندک و جنگل تصادفی با استفاده از نرم‌افزار آماری R اجرا گردیدند. مقدار بهینه پارامترهای متغیر این دو روش با استفاده از آزمون و خطا تعیین شد. با اجرای مدل، مقدار رسوب مربوط به یک دبی در سطوح احتمال مختلف (۱٪ تا ۹۹٪) محاسبه شد.

یافته‌ها: با به‌کارگیری این دو روش، بار رسوبی در چندک‌های ۲/۵، ۵۰ و ۹۷/۵٪ تعیین و دامنه عدم قطعیت در هر ایستگاه مشخص شد. روش جنگل تصادفی در ایستگاه‌های جنگلده و نوده با مقدار RMSE به ترتیب برابر ۹۶ و ۲۱۰ تن بر روز و رگرسیون چندک در ایستگاه‌های ارازکوسه و قزاقلی با مقدار RMSE به ترتیب ۶۴۵۳ و ۲۴۸۸۶ تن بر روز به‌عنوان بهترین روش برآورد بار رسوبی انتخاب شدند. مقدار معیار ارزیابی RMSE رسوب برآورد شده توسط

* مسئول مکاتبه: s.m.seyedian@gmail.com

منحنی سنج رسوب کلاسیک در ایستگاه‌های جنگلده، نوده، ارازکوسه و قراقلی به ترتیب برابر ۱۹۹، ۲۸۸، ۷۵۰۵ و ۲۵۸۱۱ تن بر روز به دست آمد.

نتیجه‌گیری: نتایج نشان داد منحنی سنج رسوب کلاسیک علاوه بر این که قادر به برآورد بار رسوبی در دامنه عدم قطعیت‌های مختلف برای یک مقدار دبی معین نیست، بار رسوبی را نیز با مقدار خطای بیش‌تری برآورد می‌کند. با استفاده از روش‌های رگرسیون چندک و جنگل تصادفی برای یک دبی معین مقدار رسوب در احتمالات مختلف قابل پیش‌بینی است و این امر کمک زیادی به برنامه‌ریزی صحیح و جامع برای ساخت سازه‌های آبی می‌کند و از این طریق، خطرات تخریب این تأسیسات را که ناشی از سیلاب‌های عظیم می‌باشد کاهش می‌دهد.

واژه‌های کلیدی: منحنی سنج رسوب، رگرسیون چندک، جنگل تصادفی، رودخانه گرگانرود

مقدمه

امروزه منابع آبی و فرسایش و رسوب یکی از چالش‌های عمده زیست‌محیطی در زمینه مدیریت حوزه به‌شمار می‌رود، به طوری که حدود ۸۵٪ تخریب زمین و محیط‌زیست در سطح جهان در ارتباط با فرسایش خاک است. مقداری از خاک فرسایش یافته در بالادست آبراهه ته‌نشین می‌شود و بقیه رسوبی که در واحد زمان از یک حوزه خارج می‌گردد را بار رسوبی می‌نامند. رسوبات معلق در ابتدا مواد بار بستر هستند که با جریان آب رودخانه حرکت می‌کنند و به صورت معلق برای زمان قابل‌توجهی بدون تماس با بستر رودخانه به مسیر خود در آبراهه ادامه می‌دهند. وجود بار معلق در جریان رودخانه موضوع مهمی است زیرا به طور مستقیم بر اکوسیستم، تنوع زیستی، ژئومورفولوژی رودخانه و همچنین جنبه‌های مختلف مهندسی به خصوص طراحی و بهره‌برداری از سازه‌های هیدرولیکی تأثیرگذار است. برآورد رسوبات حمل شده و انتقال آن توسط رودخانه‌ها یک مسأله مهم و اساسی در برنامه‌ریزی منابع آب است. میزان انتقال رسوب پارامتری است که در طراحی مخازن و سازه‌های هیدرولیکی، کیفیت آب و تأثیر آن بر محیط‌زیست و مشکلات ناشی از ته‌نشین شدن رسوبات در کانال‌های کوچک (شبکه‌های انتقال آب) و آبراهه‌های بزرگ (جهت کشتیرانی) نیاز به مدیریت

دارد (۳۵). بنابراین برآورد مقدار دقیق رسوبات معلق موجود در رودخانه‌ها حیاتی و شاخص مناسبی برای ارزیابی شرایط آینده مدیریت منابع آب است (۲۲). تاکنون، مطالعات متعددی برای برآورد بار رسوب رودخانه‌ها صورت گرفته است. با این وجود، با توجه به مکانیزم بسیار پیچیده و ماهیت غیرخطی فرایند انتقال رسوب در رودخانه و رفتار غیرخطی پارامترهای مؤثر بر مقدار رسوب معلق، این روش‌ها قادر به برآورد مقدار دقیق رسوبات نیستند (۱۶). بنابراین بعضی از پژوهشگران برای بهبود نتایج از ضرایب اصلاحی استفاده کردند (۲۳ و ۲۷). از طرف دیگر در بسیاری از رودخانه‌ها نمونه‌برداری از رسوب به ندرت و به صورت دوره‌ای انجام می‌شود. در سال‌های اخیر مطالعات قابل‌توجهی برای ایجاد رابطه تجربی بین بار رسوب و دبی انجام شده است. این روش‌ها شامل شبکه عصبی مصنوعی (آفن و همکاران، ۲۰۱۵؛ حیات‌زاده و همکاران، ۲۰۱۷) برنامه‌ریزی ژنتیک (کیتسیکودس و همکاران، ۲۰۱۴)، منطق فازی (لوهانی و همکاران، ۲۰۰۷)، شبکه موجک عصبی (پارتال و سبجیزاگلو، ۲۰۰۸) و ماشین‌های بردار پشتیبان (فتح‌آبادی و همکاران، ۲۰۱۶؛ کاکائی‌لفدانی و همکاران، ۲۰۱۳) و یا ترکیبی از این روش‌ها (شیخعلی‌پور و همکاران، ۲۰۱۵) می‌باشند می‌باشند (۱، ۹، ۱۲، ۱۴، ۱۶، ۱۹، ۲۴ و ۳۰). نتایج پژوهش‌های مختلف نشان می‌دهد پیچیدگی

مصطفی زاده و همکاران (۲۰۱۵) نشان دادند منحنی سنجه رسوب از خط به حلقه‌هایی تبدیل می‌شود که شکل آن با توجه به شرایط منطقه و نوع بارش متفاوت است (۲۱). به علت وجود این پدیده (پدیده پسماند^۱)، برای یک دبی معین چندین مقدار رسوب وجود دارد و روش‌های مختلف هوش مصنوعی قادر به پیش‌بینی دقیق مقادیر رسوب مختلف برای یک دبی مشخص نیستند (۲۹). بنابراین برآورد بار رسوب با روش‌هایی مانند جنگل تصادفی (RF^۲) و رگرسیون چندک (QR^۳) که می‌تواند برای یک دبی، بار رسوب مختلفی را با توجه به احتمال وقوع پیش‌بینی کنند دارای اهمیت است. فلاح و همکاران (۲۰۱۵) به منظور تعیین قابلیت داده‌های سنجنده ASTER از روش جنگل تصادفی برای تهیه نقشه تیپ استفاده کردند و نتایج قابل قبول بود (۷). فرهادی و شاهسونی (۲۰۱۶) برای خوشه‌بندی داده‌ها از روش عدم تشابه جنگل تصادفی (RFD^۴) استفاده کردند و نشان دادند این روش معیار مناسبی برای سنجش عدم تشابه داده‌ها است (۸). سالاری جزی (۲۰۱۷) نشان داد رگرسیون چندک متغیرهای سیلاب را با دقت بیشتری نسبت به رگرسیون خطی تحلیل می‌کند (۲۶). با استفاده از رگرسیون چندک فرنک و همکاران (۲۰۰۸) غلظت رسوب را با استفاده از روش‌های خطی و مدل‌های غیرخطی جنگل تصادفی (RF) و جنگل رگرسیون چندک (QRF^۵) پیش‌بینی کردند. نتایج نشان داد که همه روش‌های ارائه شده با مدل‌های خطی، عملکرد پایین‌تری نسبت به RF و QRF دارند. همچنین جنگل رگرسیون چندک علاوه بر این که پیش‌بینی و برآورد رسوبات را با دقت زیادی انجام می‌دهد قادر است دامنه عدم قطعیت خروجی مدل را برآورد کنند که یکی از مزیت‌های این روش

ذاتی در فرآیند انتقال رسوب قطعاً بین دبی و رسوب متناظر با آن پراکندگی ایجاد می‌کند.

منحنی سنجه بار رسوب معلق، رابطه‌ای لگاریتمی بین دبی و غلظت رسوب برآزش می‌دهد (۶ و ۳۲). منحنی سنجه بار رسوب را کم‌تر از مقدار واقعی پیش‌بینی می‌کند (۳۶ و ۳۹). اغلب این روابط غیرخطی هستند و تعدادی پژوهش‌ها به صورت منطقه‌ای برای تعریف منحنی سنجه روش‌های رگرسیون را وزن داده‌اند (۱۳، ۱۵، ۲۴ و ۳۴). به‌طور کلی به علت پویایی و پیچیدگی تخمین بار رسوب با استفاده از منحنی سنجه، پراکندگی نقاط در این منحنی زیاد و بنابراین عدم قطعیت آن نیز زیاد است (۳۸). والینگ و وب (۱۹۸۸) نشان دادند این پراکندگی محدودیت بیشتری بر عدم قطعیت بار رسوبی ایجاد می‌کند (۴۰). ترس- والبرگ و لین (۲۰۰۳) عدم قطعیت مقدار بار رسوب را محاسبه کردند و نتایج نشان داد مقادیر کم عدم قطعیت با دقت کمی پیش‌بینی شده است (۳۴).

ارائه اطلاعات عدم قطعیت به‌منظور تخمین دقیق بار رسوبی در برنامه‌ریزی و مدیریت منابع آب دارای اهمیت است. مطالعات مختلفی به این منظور صورت گرفته است: مکین و آل- نصری (۱۹۸۸) برای هر دبی یک بازه اطمینان از غلظت رسوب معلق تعیین کردند و عدم قطعیت مقدار رسوب را بررسی کردند (۲۰). کلارک (a1۹۹۰، b1۹۹۰) برای پیش‌بینی غلظت رسوب معلق مربوط به یک دبی از ویژگی‌های آماری مانند اریبی و واریانس با فرض توزیع لوگ نرمال دو متغیره استفاده کرد (۳ و ۴). رستمجی و ویلکینسن (۲۰۰۸) با استفاده از تکنیک‌های نمونه‌گیری مجدد بوت استرپ برای برآورد بار معلق به‌منظور کاهش عدم قطعیت در شکل منحنی سنجه رسوب^۱ استفاده کردند (۲۵). ویجیاک و بند- میچل (۲۰۱۳) کارایی دو روش بوت استرپ و استنتاج بی‌زی را برای ارزیابی عدم قطعیت بار رسوب مقایسه کردند (۳۷).

- 2- Hysteresis
- 3- Random Forests
- 4- Quantile Regression
- 5- Random Forest Dissimilarity
- 6- Quantile Regression Forests

- 1- Sediment Rating Curve

۳۳۳ کیلومتر و در بخش جنوب‌شرقی دریای خزر قرار دارد. این حوزه از شمال و غرب به حوزه رودخانه اترک، از جنوب به حوزه‌های آبریز شاهرود و دامغان و از جنوب‌شرقی به حوزه قره‌سو و حوزه رودخانه نکا محدود می‌شود. در این حوزه ارتفاعات البرز بخش‌های جنوبی و شرقی آن را تشکیل داده که با توجه به پهنای کم چین‌خوردگی‌های البرز در این ناحیه رودخانه‌های واقع در این ارتفاعات فاقد حوزه‌های آبخیز وسیع می‌باشند. حوزه‌های واقع در این ارتفاعات با شیب تندی به دشت وسیع گرگان منتهی می‌شوند. در این مطالعه چهار ایستگاه هیدرومتری قزاقلی، ارازکوسه، نوده و جنگله به‌عنوان ایستگاه مینا بر روی رودخانه گرگانرود برای مدل‌سازی رسوب انتخاب گردید. مساحت تحت پوشش ایستگاه قزاقلی ۷۰۷۲ کیلومترمربع است. بالادست ایستگاه قزاقلی نوع اقلیم منطقه بر اساس تقسیم‌بندی دومارتن نیمه‌خشک تا خیلی مرطوب بوده که بیش‌ترین مقدار بارش در ماه‌های فروردین و آذر و کم‌ترین مقدار در ماه‌های تیر و مرداد می‌باشد. عمده کاربری اراضی در حوزه مذکور کاربری‌های مرتع-بوته‌زار و مخلوط جنگل‌های سوزنی برگ-پهن‌برگ به‌ترتیب با ۳۵/۰۸ و ۲۷/۰۱ درصد می‌باشد (۲۸). مقدار ارتفاع حداقل و حداکثر ایستگاه‌های ارازکوسه، نوده و جنگله به‌ترتیب (۲۰ و ۲۸۷۸)، (۵۰۶ و ۲۰۶۱) و (۱۶۱ و ۱۸۹۲) متر می‌باشد. شیب متوسط آن‌ها به‌ترتیب ۴/۵۸، ۴/۲۷ و ۳۰/۸ درصد است. عمده کاربری اراضی در حوزه ارازکوسه کشاورزی-جنگل و کوهستانی-دشتی، در حوزه نوده کشاورزی و جنگل و در حوزه جنگله کشاورزی-مسکونی و جنگل می‌باشد. موقعیت قرارگیری این چهار ایستگاه در شکل ۱ مشخص شده است. در جدول ۱ پارامترهای آماری چهار ایستگاه جنگله، نوده، ارازکوسه و قزاقلی به تفکیک مراحل واسنجی و صحت‌سنجی آورده شده است.

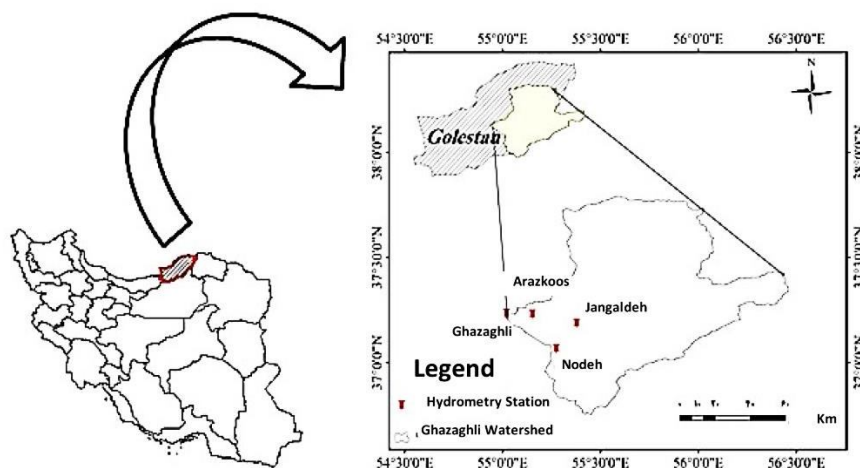
نسبت به روش‌های دیگر است (۱۰). زنگ‌شیا و جوچن (۲۰۱۵) برای برآورد مقادیر احتمالی بار رسوب روزانه و سالانه ایستگاه لئونگ واقع در جنوب تایوان به‌جای ایجاد یک منحنی سنجه واحد از رویکرد جدید رگرسیون چندک استفاده کردند (۳۵). فرنک (۲۰۰۹) به‌منظور اندازه‌گیری و مدل‌سازی بار رسوب در حوزه‌های با اندازه متوسط از منحنی سنجه، مدل‌های خطی تعمیم‌یافته (GLM) و رگرسیون ناپارامتری و برای بررسی عدم قطعیت از مدل‌های جنگل تصادفی و جنگل رگرسیون چندک استفاده کرد. نتایج نشان داد روش‌های منحنی سنجه رسوب معمولی و مدل‌های خطی تعمیم‌یافته با وجود در نظر گرفتن متغیرهایی مانند شدت بارش، پیش‌بینی مناسبی انجام ندادند اما دو روش RF و QRF توانستند با دقت زیادی رسوب را تخمین بزنند. همچنین RF و QRF عدم قطعیت مقدار رسوب برآورد شده را نیز محاسبه می‌کنند (۱۱).

در این پژوهش دو روش ناپارامتری و غیرخطی QR و RF برای تخمین بار رسوبی در سطوح احتمالات مختلف در چهار ایستگاه هیدرومتری با مساحت‌های مختلف و کاربری غالب متفاوت، در شرق حوزه آبخیز گرگانرود ارزیابی شد. همچنین کارایی سه روش منحنی سنجه رسوب، رگرسیون چندک و جنگل تصادفی مقایسه شد.

مواد و روش‌ها

منطقه مورد مطالعه: حوزه آبخیز گرگانرود یکی از سه حوزه اصلی در استان گلستان است که در جنوب‌شرقی این استان در حد فاصل عرض‌های شمالی ۳۴°، ۳۴' و ۴۷'، ۳۷° و طول‌های جغرافیایی ۰۲'، ۵۴° و ۱۶'، ۵۶° شرقی جای دارد و با راستای تقریبی غربی-شرقی وارد دشت گرگان می‌شود. این حوزه با مساحت ۱۳۱۷۰ کیلومترمربع و طول رودخانه اصلی

1- Generalized Linear Models



شکل ۱- موقعیت جغرافیایی حوزه آبخیز گرگانرود در استان گلستان.

Figure 1. Geographical location watershed Gorganrood in Golestan province.

جدول ۱- پارامترهای آماری ایستگاه‌های جنگلده، نوده، ارازکوسه و قزاقلی.

Table 1. Statistically parameters Jangal-deh, Nodeh, Arazkoose and Ghazagzli stations.

ضریب چولگی Skewness	ضریب تغییرات Coefficient of Variation	انحراف معیار Standard deviation	حداکثر max	میانگین average	حداقل min	تعداد داده The number of data	مرحله Stage	نوع داده Data type	سال آماری Statistical year	نام ایستگاه Station name
3.4	1.5	2.1	13	1.4	0.02	115	آموزش (Train)	دبی (m ³ /s) (Discharge)	1971-2013	جنگلده Jangal-deh
7.3	4.6	273.5	2437	59.5	0.03		رسوب (Ton/day) (Sediment)			
1.7	1.2	2.5	11	2.06	0.13	39	آزمون (Test)	دبی (m ³ /s) (Discharge)	1971-2013	نوده Nodeh
5.8	3.4	229.2	1424	66.7	0.14		رسوب (Ton/day) (Sediment)			
2.7	0.75	1.7	15	2.3	0.04	370	آموزش (Train)	دبی (m ³ /s) (Discharge)	1971-2013	نوده Nodeh
6.1	2.8	336.5	3348	121.4	0.07		رسوب (Ton/day) (Sediment)			
2.5	0.84	2.0	12	2.4	0.01	124	آزمون (Test)	دبی (m ³ /s) (Discharge)	1971-2013	نوده Nodeh
4.4	2.3	256.3	1629	112.02	0.04		رسوب (Ton/day) (Sediment)			
3.7	1.7	15.6	143	9.5	0.02	532	آموزش (Train)	دبی (m ³ /s) (Discharge)	1968-2014	ارازکوسه Arazkoose
9.2	4.4	18038.0	257421	4065.7	0.09		رسوب (Ton/day) (Sediment)			
3.0	1.5	13.9	82	9.5	0.03	177	آزمون (Test)	دبی (m ³ /s) (Discharge)	1971-2013	قزاقلی Ghazagzli
5.0	3.1	10436.3	74152	3409.3	0.17		رسوب (Ton/day) (Sediment)			
3.3	1.22	23.32	265	19.2	0.26	2425	آموزش (Train)	دبی (m ³ /s) (Discharge)	1971-2013	قزاقلی Ghazagzli
9.3	3.82	34614.57	705767	9058.2	0.44		رسوب (Ton/day) (Sediment)			
3.2	1.26	25.83	219	20.5	0.25	808	آزمون (Test)	دبی (m ³ /s) (Discharge)	1971-2013	قزاقلی Ghazagzli
8.1	3.65	39274.07	580129	107	0.87		رسوب (Ton/day) (Sediment)			

(۳)

$$\min \sum_{i: Q_{s,i} \geq a_q Q_i^{b_q}} q |Q_{s,i} - a_q Q_i^{b_q}| + \sum_{i: Q_{s,i} < a_q Q_i^{b_q}} (1 - q) |Q_{s,i} - a_q Q_i^{b_q}|$$

راه‌حل تحلیلی برای تعیین ضریب رگرسیون چندک a_q و b_q وجود ندارد. کنکر و دی‌اری (۱۹۸۷) با استفاده از برنامه‌ریزی خطی الگوریتمی را برای تخمین پارامترهای رگرسیون چندک ارائه دادند (۱۸). مدل رگرسیون چندک بار رسوبی در چندک‌های مختلف برای یک دبی معین را برآورد می‌کند. کل تابع توزیع بار رسوب برای یک دبی معین نیز با استفاده از رگرسیون چندک قابل بررسی است (۳۵). محاسبات مدل رگرسیون چندک در این مطالعه در محیط نرم‌افزار آماری R انجام شد، به‌منظور انتخاب مناسب‌ترین خط برازش‌یافته توسط روش رگرسیون چندک در سه احتمال (۲/۵، ۵۰، ۹۷/۵ درصد) دو متغیر a_q و b_q با آزمون و خطا تعیین می‌گردد.

جنگل تصادفی: جنگل تصادفی یک نوع مدرن از روش‌های درخت-پایه است که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی می‌باشد همچنین یکی از روش‌های ناپارامتریک مناسب برای مدل‌سازی داده‌های پیوسته و گسسته روش درخت تصمیم است. از مشکلات این روش نوسانات بالای نتایج هر درخت است. به‌منظور کاهش این نوسانات و کاهش واریانس برآورد، روش جنگل تصادفی پیشنهاد شده است. جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن n داده مشاهده‌ای مستقل ساخته می‌شود:

$$(Y_i, X_i), \quad i=1, \dots, n \quad (۴)$$

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به‌طور تصادفی

منحنی سنجه رسوب: منحنی سنجه رسوب نشان‌دهنده رابطه تجربی بین بار رسوب (یا غلظت) اندازه‌گیری شده و دبی رودخانه است. معمولاً از یک رابطه توانی برای برازش استفاده می‌شود:

$$Q_s = aQ^b \quad (۱)$$

که در آن، Q_s نشان‌دهنده بار یا غلظت رسوب معلق بر حسب تن بر روز یا میلی‌گرم بر لیتر، Q دبی رودخانه بر حسب مترمکعب در ثانیه، a و b ضرایب ثابت هستند که با استفاده از داده‌های اندازه‌گیری شده تعیین می‌شوند. در این مطالعه از روش تندترین شیب با در نظر گرفتن معیار حداقل مربعات برای محاسبه پارامترهای ثابت منحنی سنجه استفاده شد (۲۲).

رگرسیون چندک: رگرسیون چندک توسط کنکر و باست (۱۹۷۸) توسعه داده شد (۱۷). از آنجایی که رگرسیون چندک می‌تواند متغیرهایی از کل دامنه توزیع داده‌ها بگیرد، اخیراً در طیف گسترده‌ای از برنامه‌های کاربردی مانند اقتصاد و امور مالی (الاجید و پنایجیوتایدس، ۲۰۱۲) زیست‌شناسی و محیط‌زیست (کازلی و همکاران، ۲۰۱۳) و علوم زیست‌محیطی (شیثائو و هوایانگ، ۲۰۱۵) استفاده شده است (۲)، (۵) و (۳۱).

رگرسیون چندک بر اساس رابطه توانی منحنی سنجه رسوب برای مقادیر مختلف چندک بیان شده است به‌طوری‌که:

$$Q_{s,q} = a_q Q^{b_q} \quad (۲)$$

که در آن، q مقدار چندک بین ۰ تا ۱، $Q_{s,q}$ بار رسوب در چندک q برای یک دبی معین Q و a_q و b_q ضرایب رگرسیون چندک می‌باشند که برای هر چندک تعیین می‌شود. به‌عنوان مثال:

روش کار: در این پژوهش از مدل‌های منحنی سنجه رسوب، رگرسیون چندک و جنگل تصادفی به منظور برآورد بار رسوب چهار ایستگاه جنگلده، نوده، ارازکوسه و قزاقلی استفاده شد. به این منظور داده‌های دبی- رسوب متناظر چهار ایستگاه مورد مطالعه به دو بخش ۷۵٪ برای آموزش و ۲۵٪ برای تست تفکیک شدند. در روش منحنی سنجه، مقدار رسوب با استفاده از معادله توانی برازش داده شده بین دبی و رسوب متناظر، حاصل گردید. به منظور استفاده از الگوریتم رگرسیون چندک و جنگل تصادفی، داده‌های آموزش و تست به نرم‌افزار آماری R و بسته رگرسیون چندک و جنگل تصادفی معرفی شدند. با اجرای مدل، مقدار رسوب در سطوح احتمال مختلف (۱٪ تا ۹۹٪) محاسبه شد. به منظور انتخاب مناسب‌ترین خط برازش یافته توسط RF و QR در سه احتمال (۲/۵، ۵۰ و ۹۷/۵) درصد، به ترتیب متغیر اندازه گره برای مدل جنگل تصادفی و دو متغیر a_q و b_q برای مدل رگرسیون چندک با آزمون و خطا تعیین شدند و بهترین روش با استفاده از مقدار RMSE کم‌تر و تعداد داده‌های مشاهداتی قرار گرفته در دامنه عدم قطعیت (بیش‌تر از رسوب برآورد شده در صدک ۲/۵ و کم‌تر از رسوب برآورد شده صدک ۹۷/۵) انتخاب شدند. برای رسم تابع جرم احتمال (PMF^1) از نمودار تابع توزیع تجمعی (CDF^3) استفاده شد. ابتدا در یک دبی معین، بار رسوب در سطوح احتمال ۰/۰۱ تا ۰/۹۹ توسط دو روش رگرسیون چندک و جنگل تصادفی برآورد می‌گردد. سپس خطی برازش داده می‌شود که این ۹۹ نقطه را به یکدیگر وصل کند که خط برازش داده شده منحنی CDF است. بر اساس منحنی CDF، احتمال وقوع مقادیر مختلف بار رسوبی در هر دبی تعیین می‌گردد.

تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت استرپ به تعداد زیاد (به عنوان مثال ۲۰۰۰ مرتبه) نمونه‌های n تایی از مجموعه داده‌های مشاهداتی اولیه، نمونه‌برداری همراه با جای‌گذاری می‌شوند. در طی فرآیند نمونه‌گیری حدود یک سوم از داده‌ها نمونه‌گیری نمی‌شوند و به عنوان نمونه خارج از کیسه (از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد ناریب خطا استفاده می‌شود) در نظر گرفته می‌شوند. سپس بر روی هر نمونه بوت استرپ یک درخت گسترش داده می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به صورت تصادفی m متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت $\frac{m}{M}$ برابر با یک سوم است و برای کلاسه‌بندی برابر با $m = \sqrt{M}$ پیشنهاد شده است. پس از ساخت تمام درخت داده‌های تست به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود.

روش درخت رگرسیون جنگل تصادفی به‌ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد یک روش پیش‌بینی کارآمد است (۳۳). در این پژوهش محاسبات مدل جنگل تصادفی در محیط نرم‌افزار R و بسته جنگل تصادفی انجام شد. به منظور انتخاب مناسب‌ترین خط برازش یافته توسط این روش در سه احتمال (۲/۵، ۵۰ و ۹۷/۵ درصد)، متغیر اندازه گره^۱ (که نشان‌دهنده تعداد برگ‌ها در هر شاخه است) با آزمون و خطا تعیین شد.

2- Probability Mass Function
3- Cumulative Distribution Function

1- Nodesize

دخالت‌های انسانی و نیز به سبب وجود پدیده پسماند در طبیعت در میزان فرسایش خاک و تولید رسوب نوساناتی پدید خواهد آمد. این امر موجب می‌شود که چندین مقدار بار رسوبی برای یک دبی مشخص وجود داشته باشد. به همین دلیل در یک دبی معین میزان تولید رسوب در زمان‌های مختلف متفاوت خواهد بود و این تفاوت در رویدادهای طبیعی نیز قابل مشاهده است. به‌عنوان مثال در ایستگاه جنگلده میزان رسوب متناظر با دبی $0.17 \text{ m}^3/\text{s}$ در ۲۰ آبان ۱۳۸۶ برابر 0.38 ، در ۶ مهر ۱۳۸۹ برابر 0.23 و در ۹ شهریور ۱۳۹۱ معادل $1/26$ تن بر روز است. این در حالی است که روش SRC رسوب برآورد شده در این دبی را در هر سه زمان 0.54 تن بر روز برآورد کرده است. در ایستگاه نوده در دبی $1/51 \text{ m}^3/\text{s}$ مقدار بار رسوبی حاصل از منحنی سنج رسوب $23/71$ تن بر روز محاسبه شد در حالی که در ۲۶ تیر ۱۳۵۵ برابر $119/59$ ، در ۱۳ مرداد ۱۳۵۵ برابر 0.26 و در ۶ آبان ۱۳۸۲، $31/18$ تن بر روز ثبت شده است. تفاوت حداکثر و حداقل رسوب ثبت شده برای دبی $1/51 \text{ m}^3/\text{s}$ در ایستگاه نوده، تقریباً 459 تن بر روز می‌باشد که این مقدار تفاوت در بار رسوبی با روش SRC قابل برآورد نیست.

در ایستگاه ارازکوسه برای دبی $11 \text{ m}^3/\text{s}$ مقدار رسوب در ۳ بهمن ۱۳۴۸، 11785 تن بر روز، در ۲۰ اسفند ۱۳۵۰، 687 تن بر روز و در ۲۲ فروردین ۱۳۶۸، 1119 تن بر روز مشاهده شده و مقدار رسوب با روش SRC 1143 تن بر روز محاسبه شد. همچنین در ایستگاه قزاقلی مقادیر رسوبات ثبت شده برای دبی $11/30 \text{ m}^3/\text{s}$ در ۲۷ آذر ۱۳۵۸، ۱۲ خرداد ۱۳۶۱، ۸ اردیبهشت ۱۳۶۳ و ۹ دی ۱۳۶۳ به ترتیب برابر 3520 ، 6501 ، 412 و 202 تن بر روز می‌باشد و منحنی سنج رسوب 759 تن بر روز را تخمین می‌زند. در ایستگاه ارازکوسه و قزاقلی حداکثر رسوب

معیارهای ارزیابی: به‌منظور انتخاب مناسب‌ترین روش پیش‌بینی بار رسوب در هر ایستگاه برای مدل‌های جنگل تصادفی و رگرسیون چندک از شاخص ریشه دوم میانگین مربع خطا ($RMSE^1$) و تعداد داده‌های رسوب برآورد شده که در دامنه عدم قطعیت $2/5\%$ و $97/5\%$ قرار دارد، استفاده می‌شود. برای مقایسه سه روش منحنی سنج رسوب، رگرسیون چندک و جنگل تصادفی نیز شاخص $RMSE$ محاسبه شد. بر این اساس هر اندازه مقدار $RMSE$ به صفر نزدیک‌تر باشد و مجموع نقاط درون محدوده عدم قطعیت $2/5\%$ و $97/5\%$ بیش‌تر باشد مدل با دقت بیش‌تری بار رسوب را برآورد کرده است. شاخص $RMSE$ بر اساس رابطه ۵ قابل محاسبه است:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{sp} - Q_{so})^2}{n}} \quad (5)$$

که در آن، n تعداد داده‌ها، Q_{sp} مقادیر رسوب برآورد شده از مدل رگرسیون چندک و جنگل تصادفی در سطح اطمینان 50% و نیز مقدار رسوب محاسبه شده از روش منحنی سنج رسوب، Q_{so} مقادیر رسوب مشاهده شده است.

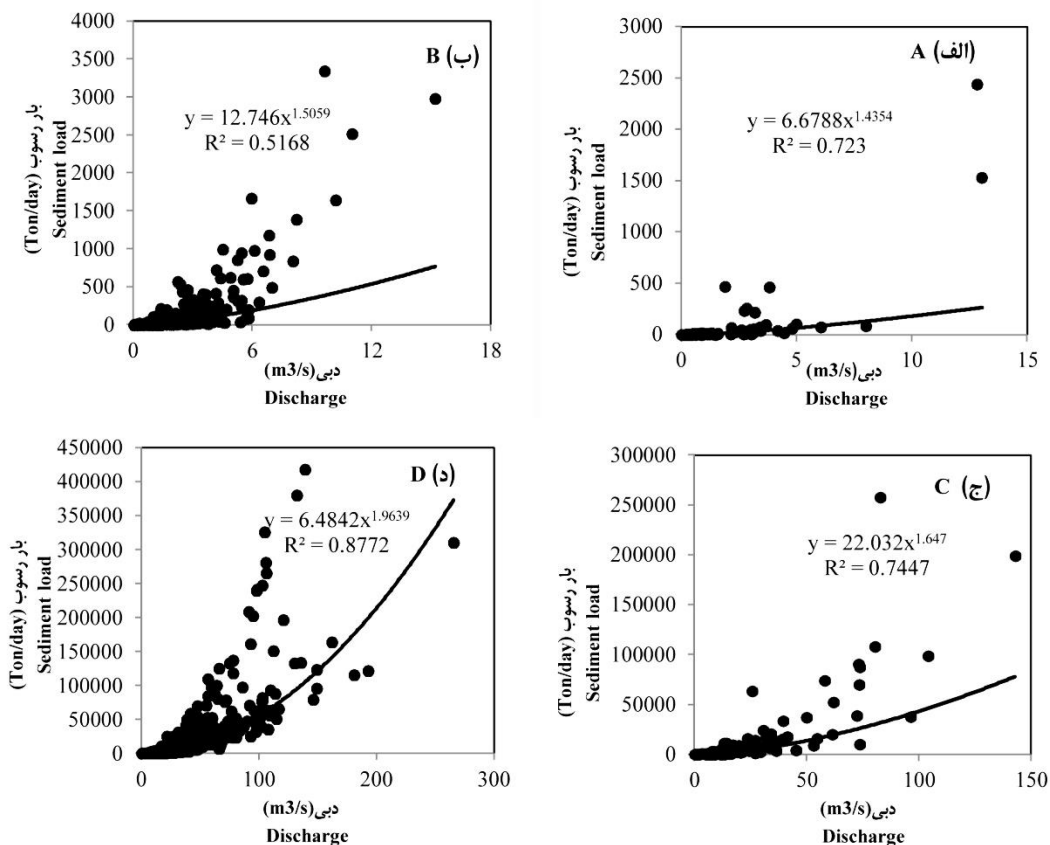
نتایج و بحث

در شکل ۲ منحنی سنج ایستگاه‌های جنگلده، نوده، ارازکوسه و قزاقلی نشان داده شده است. به‌علت وقوع واکنش‌های فیزیکی و شیمیایی گوناگون در سطح حوزه آبخیز و تغییر در شرایط ژئوشیمیایی و بیولوژیکی بین منابع حوزه (خاک، پوشش گیاهی، زمین‌شناسی، کاربری اراضی و ...) و به‌علت تغییر در متغیرهای هواشناسی (دما، بارش، ذوب برف و ...) و

1- Root Mean Square Error

یک خط (خطی یا غیرخطی) به داده‌ها برازش می‌دهد نمی‌توان هم‌زمان مقدار چند رسوب برای یک دبی را برآورد کرد در این حالت راهکار عملی برآورد دامنه عدم قطعیتی است که بتواند این مقادیر رسوب را در برگرد.

به ترتیب ۱۷ و ۳۲ برابر مقدار حداقل رسوب ثبت شده می‌باشد؛ در حالی که در تمامی تاریخ‌های ذکر شده مقدار رسوب با روش منحنی سنجه یکسان محاسبه شد. در حالتی که برای یک دبی مشخص چند مقدار رسوب وجود دارد با استفاده از هر روشی که



شکل ۲- نمودار پراکنندگی دبی- رسوب در ایستگاه‌های (الف) جنگلده و (ب) نوده برای دوره ۱۳۵۰-۱۳۹۲، (ج) ارازکوسه برای دوره ۱۳۹۳-۱۳۴۷ و (د) قزاقلی برای دوره ۱۳۵۰-۱۳۹۲.

Figure 2. Scatter plot of discharge-sediment load in stations (A) Jangalده and (B) Nodeh for 1971-2013 period, (C) Arazkose for 1968-2014 period and (D) Ghazaghlī for of 1971-2013 period.

ایستگاه‌های جنگلده و نوده، جنگل تصادفی و بهترین روش برای ایستگاه‌های ارازکوسه و قزاقلی، روش رگرسیون چندک می‌باشد. برای مثال در شکل ۳، مقدار رسوب برآورد شده در دامنه عدم قطعیت ۲/۵٪، ۵۰٪ و ۹۷/۵٪ با روش رگرسیون چندک برای ایستگاه ارازکوسه به همراه منحنی سنجه رسوب کلاسیک

در جدول ۲ مقادیر متغیرهای a_q و b_q و اندازه گره برای چهار ایستگاه جنگلده، نوده، ارازکوسه و قزاقلی آورده شده است. با استفاده از پارامترهای بهینه شده مدل‌های رگرسیون چندک و جنگل تصادفی، مقدار رسوب محاسبه شد. با توجه به جدول ۲ مناسب‌ترین روش به‌منظور تخمین بار رسوب برای

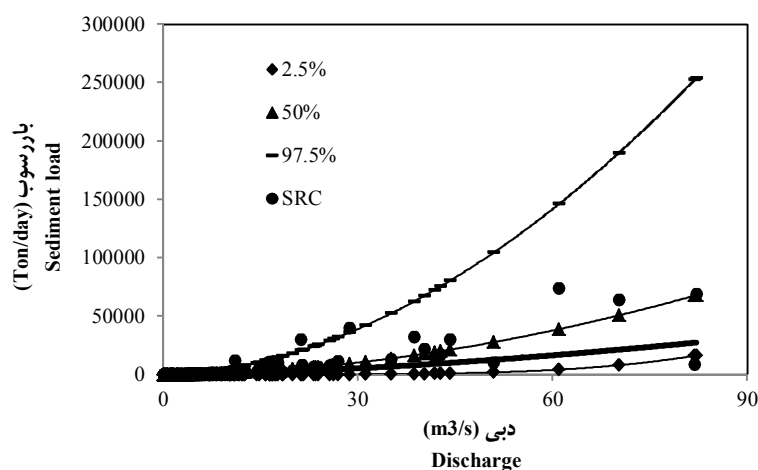
۱۶۲ نقطه (۹۲ درصد داده‌های موجود) درون دامنه عدم قطعیت ۲/۵ و ۹۷/۵٪ قرار گرفته‌اند.

نمایش داده شده است. مناسب‌ترین متغیر a_q و b_q برای ایستگاه ارازکوسه به ترتیب برابر ۳ و ۴/۸ با مقدار RMSE برابر ۶۴۵۳ تن بر روز و در مجموع

جدول ۲- بهترین مقادیر متغیرهای مدل‌های رگرسیون چندک و جنگل تصادفی در ایستگاه‌های جنگلده، نوده، ارازکوسه و قزاقلی.

Table 2. The best values of Quantile Regression and Random Forest variables in Jangalده, Nodeh, Arazkooose and Ghazaghli stations.

نام ایستگاه Station name	مدل Model	a_q	b_q	اندازه گره Nodesize	RMSE (Ton/day)	۲.۵% < X < ۹۷.۵%	مناسب‌ترین مدل The best model
جنگلده Jangalده	رگرسیون چندک Quantile Regression	1.5	2	-	167	6	جنگل تصادفی Random Forest
	جنگل تصادفی Random Forest	-	-	47	96	23	جنگل تصادفی Random Forest
نوده Nodeh	رگرسیون چندک Quantile Regression	2	4.5	-	141	11	جنگل تصادفی Random Forest
	جنگل تصادفی Random Forest	-	-	7	210	80	جنگل تصادفی Random Forest
ارازکوسه Arazkooose	رگرسیون چندک Quantile Regression	3	4.8	-	6453	162	رگرسیون چندک Quantile Regression
	جنگل تصادفی Random Forest	-	-	18	9225	110	رگرسیون چندک Quantile Regression
قزاقلی Ghazaghli	رگرسیون چندک Quantile Regression	2.2	4.5	-	24886	788	رگرسیون چندک Quantile Regression
	جنگل تصادفی Random Forest	-	-	10	23121	373	رگرسیون چندک Quantile Regression



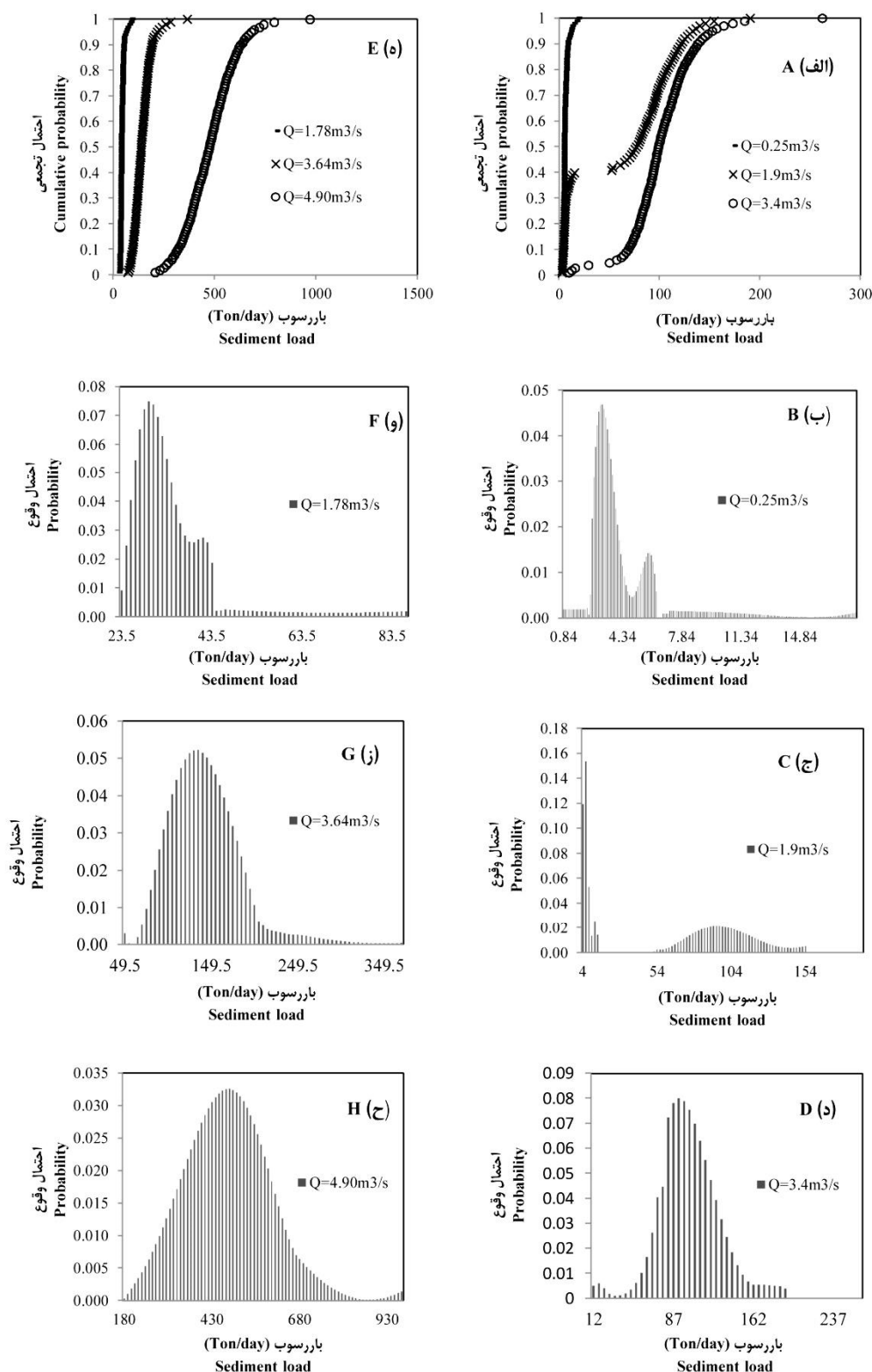
شکل ۳- منحنی‌های سنج رسوب در دامنه عدم قطعیت ۲/۵٪، ۵۰٪ و ۹۷/۵٪ با روش رگرسیون چندک برای ایستگاه ارازکوسه.

Figure 3. Sediment rating curves in the range of uncertainty 2.5%, 50% and 97.5% quantile regression for Arazkooose station.

توجه به (شکل ۵ الف)) نمودار احتمال وقوع تجمعی مربوط به ایستگاه ارازکوسه برای دبی m^3/s ۲۱/۲ هنگامی که مقدار رسوب در احتمال تجمعی ۰/۸۹ به ۹۴۲۸/۱ تن بر روز می‌رسد، احتمال وقوع رسوبات بیش‌تر از این مقدار کاهش نشان می‌دهد و شیب خط بسیار ملایم می‌شود در صورتی که در محدوده احتمال ۰/۲۵-۰/۰۱ شیب خط زیاد است (۲۰۱۵) نیز در پژوهش‌های خود نشان دادند با افزایش مقدار رسوب، شیب خط ملایم می‌شود که بیانگر احتمال وقوع کم‌تر است (۳۵). از طریق تابع جرم احتمال می‌توان احتمال وقوع بار رسوب را برای یک دبی پیش‌بینی کرد. برای مثال با توجه به (شکل ۵ ب)) برای دبی m^3/s ۱/۵ به ترتیب در ۳ آذر ۱۳۵۱ و ۱۱ مهر ۱۳۹۲ بار رسوب ۱۴/۸ و ۱۱۹/۴ تن بر روز، برای دبی m^3/s ۴/۹ نیز بار رسوب ۳۱۳/۲ تن بر روز در ۱۲ اردیبهشت ۱۳۵۰ و ۹۰/۱۱ تن بر روز در ۲۹ دی ۱۳۶۱ ثبت شده است. مقدار رسوب برای ایستگاه ارازکوسه با روش رگرسیون چندک (شکل ۵ ب)) نشان می‌دهد که برای دبی m^3/s ۱/۵ احتمال وقوع بار رسوبی ۱۴/۸ و ۱۱۹/۴ تن بر روز به ترتیب حدود ۰/۰۳۴ و ۰/۰۰۸ است. همچنین در دبی m^3/s ۴/۹ احتمال وقوع بار رسوب ۹۰/۱۱ و ۳۱۳/۲ تن بر روز به ترتیب برابر ۰/۰۳۴ و ۰/۰۲۳ می‌باشد (شکل ۵ ج)). (شکل ۵ ب)، (ج) و (د)) نشان می‌دهد در ایستگاه ارازکوسه بیش‌ترین آورد رسوب برای دبی‌های ۱/۵، ۴/۹ و ۲۱/۲ مترمکعب بر ثانیه به ترتیب برابر ۲/۵، ۴۹/۵ و ۱۹۱۰ تن بر روز است. تابع جرم احتمال ایستگاه قزاقلی (شکل ۵ و)، (ز) و (ح)) نشان می‌دهد برای دبی‌های ۱۴، ۳۲/۳ و ۱۰۹ مترمکعب بر ثانیه حداکثر بار رسوبی که احتمال وقوع آن وجود دارد به ترتیب برابر ۵۸۰، ۲۷۳۰ و ۳۱۷۳۰ تن بر روز است.

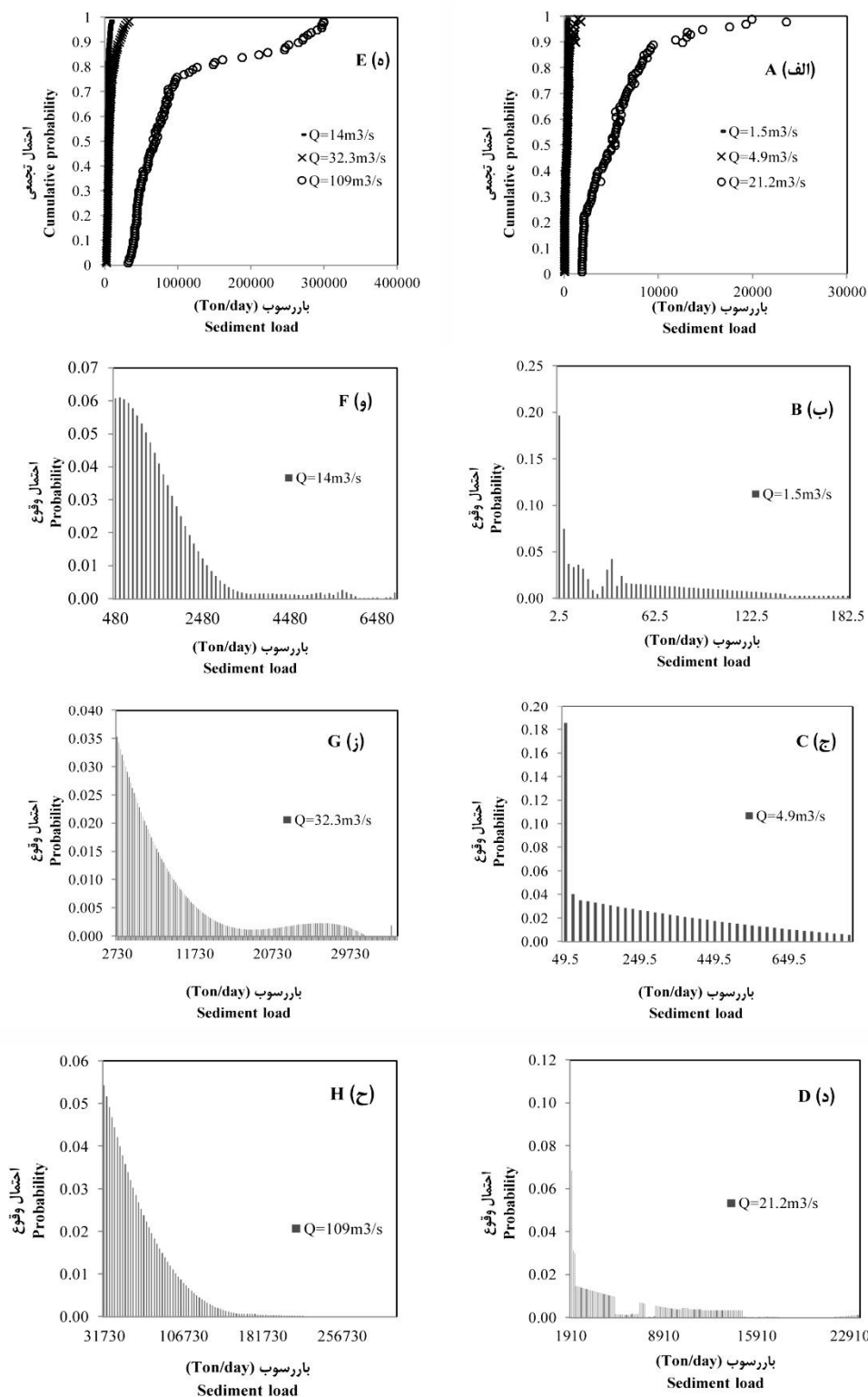
شکل ۴ تابع توزیع تجمعی و تابع جرم احتمال در دو ایستگاه جنگلده و نوده که با استفاده از روش جنگل تصادفی به دست آمده را نشان می‌دهد. تابع توزیع تجمعی مقادیر بار رسوب در سطح احتمالات مختلف را نشان می‌دهد. تابع جرم احتمال از تابع توزیع تجمعی بر اساس متوسط رسوب در احتمالات مختلف تعیین می‌شود. به عنوان مثال در ایستگاه جنگلده (شکل ۴ ب)) بار رسوب برای دبی m^3/s ۰/۲۵ در محدوده‌ای بین ۰/۸۴ و ۱۷/۸۴ تن بر روز متغیر است و به سمت راست دارای چولگی است. با احتمال ۰/۹۷ بار رسوب بین ۰ و ۱۱/۹۴ تن بر روز متغیر است (شکل ۴ الف)) و در دبی m^3/s ۰/۲۵ بیش‌ترین بار رسوب ۳/۰۴ تن بر روز با احتمال ۰/۰۴۶ است (شکل ۴ ب)). برای یک دبی ثابت با توجه به شرایط در روزها و یا فصل‌های مختلف، بار رسوبی متفاوت است که با استفاده از تابع جرم احتمال قابل تحلیل و شناسایی می‌باشد. به عنوان مثال در ایستگاه جنگلده بار رسوب برای دبی m^3/s ۰/۲۵ در ۸ شهریور ۱۳۸۳ و برای دبی m^3/s ۱/۹ در ۱۷ فروردین ۱۳۷۸ به ترتیب برابر ۱/۷۱ و ۱۵/۶ تن بر روز بوده است. با توجه به (شکل ۴ ب) و (ج)) مشخص می‌شود احتمال وقوع بار رسوب ۱/۷۱ و ۱۵/۶ تن بر روز به ترتیب برابر با ۰/۰۰۲ و ۰/۰۰۰۷ است. مقایسه تابع جرم احتمال ((و)، (ز) و (ح)) نشان می‌دهد برای دبی‌های ۱/۷۸، ۳/۶۴، ۴/۹۰ m^3/s بیش‌ترین احتمال وقوع برای رسوبات به ترتیب معادل ۲۹/۵، ۱۳۴/۵ و ۴۸۰ تن بر روز می‌باشد. فرنک و همکاران (۲۰۰۸) نشان دادند دو روش جنگل تصادفی و رگرسیون چندک قادر به تحلیل پدیده پسماند می‌باشد (۱۰).

شکل ۵ تابع توزیع تجمعی و تابع جرم احتمال در دو ایستگاه ارازکوسه و قزاقلی تخمین زده شده با استفاده از روش رگرسیون چندک را نشان می‌دهد. با



شکل ۴- تابع توزیع تجمعی (CDF) در ایستگاه‌های: الف) جنگلده و ه) نوده؛ تابع جرم احتمال (PMF) در ایستگاه جنگلده برای دبی‌های: ب) 0.25 ، ج) 1.9 و د) 3.4 m^3/s و تابع جرم احتمال در ایستگاه نوده برای دبی‌های: و) 1.78 ، ز) 3.64 و ح) 4.90 m^3/s .

Figure 4. The Cumulative Distribution Function (CDF) in stations: A) Jangaldehy and E) Nodeh; the Probability Mass Function (PMF) Jangaldehy station for discharges: B) 0.25, C) 1.9 and D) 3.4 m^3/s and the Probability Mass Function Nodeh at the station for discharges: F) 1.78, G) 3.64 and H) 4.90 m^3/s .



شکل ۵- تابع توزیع تجمعی در ایستگاه‌های: الف) ارازکوسه و ه) قزاقلی و تابع جرم احتمال در ایستگاه ارازکوسه برای دبی‌های: ب) ۱/۵، ج) ۴/۹ و د) ۲۱/۲ و تابع جرم احتمال در ایستگاه قزاقلی برای دبی‌های: و) ۱۴، ز) ۳۲/۳ و ح) ۱۰۹ m³/s.

Figure 5. The Cumulative Distribution Function in stations: A) Arazkoose and E) Ghazaghli and the Probability Mass Function Arazkoose station for discharges: B) 1.5, C) 4.9 and D) 21.2 m³/s and the Probability Mass Function Ghazaghli at the station for discharges: F) 14, G) 32.3 and H) 109 m³/s.

(الف)). در (شکل ۷ ب)) با افزایش دبی مقدار برد و محدوده چارک‌های وسط افزایش می‌یابد که بیانگر عدم قطعیت بیش‌تر برای دبی‌های زیاد است. هنگامی که دبی از ۳ به $15 \text{ m}^3/\text{s}$ افزایش می‌یابد محدوده میان چارکی^۱ (فاصله بین چارک سوم و اول) از ۱۵۹ به ۲۱۸۲ تن بر روز می‌رسد. همچنین در (شکل ۷ د)) با افزایش دبی از ۱۱۰ به $165 \text{ m}^3/\text{s}$ محدوده میان چارکی از ۵۸۵۰۶ به 103090 تن بر روز افزایش می‌یابد، هنگامی که دبی $220 \text{ m}^3/\text{s}$ می‌شود این محدوده به 184548 تن بر روز می‌رسد.

گستره وسیع‌تر بار رسوبی برای دبی‌های بالاتر نیز امکان وقوع حوادث سهمگین را نشان می‌دهد. حداکثر بار رسوبی برای دبی $220 \text{ m}^3/\text{s}$ در ایستگاه قزاقلی برابر 1306000 تن بر روز است. اگرچه با افزایش مقدار دبی یا رسوب احتمال وقوع، کم‌تر می‌شود اما این رخداد یک تهدید بالقوه برای آسیب رساندن به بستر، کناره‌های رودخانه‌ها، کانال‌ها و سایر سازه‌های آبی محسوب می‌شود. در نظر گرفتن این خطرها در برنامه‌ریزی و مدیریت منابع آب بسیار ضروری است اما این اطلاعات با روش منحنی سنجه رسوب کلاسیک حاصل نمی‌شود، بنابراین در اینجا لزوم استفاده از مدل‌های رگرسیون چندک و جنگل تصادفی که بار رسوبی را در سطوح احتمالات گوناگون محاسبه می‌کنند مشخص می‌شود. در جدول ۳ مقادیر رسوبات برآورد شده با روش‌های منحنی سنجه رسوب کلاسیک، رگرسیون چندک و جنگل تصادفی به همراه مقادیر مشاهده‌ای آورده شده است. به‌منظور افزایش دقت مقایسه از متوسط آورد رسوب در هر بازه دبی استفاده گردید. با توجه به جدول ۳ مناسب‌ترین رسوب برآورد شده در بازه $2-0$ ،

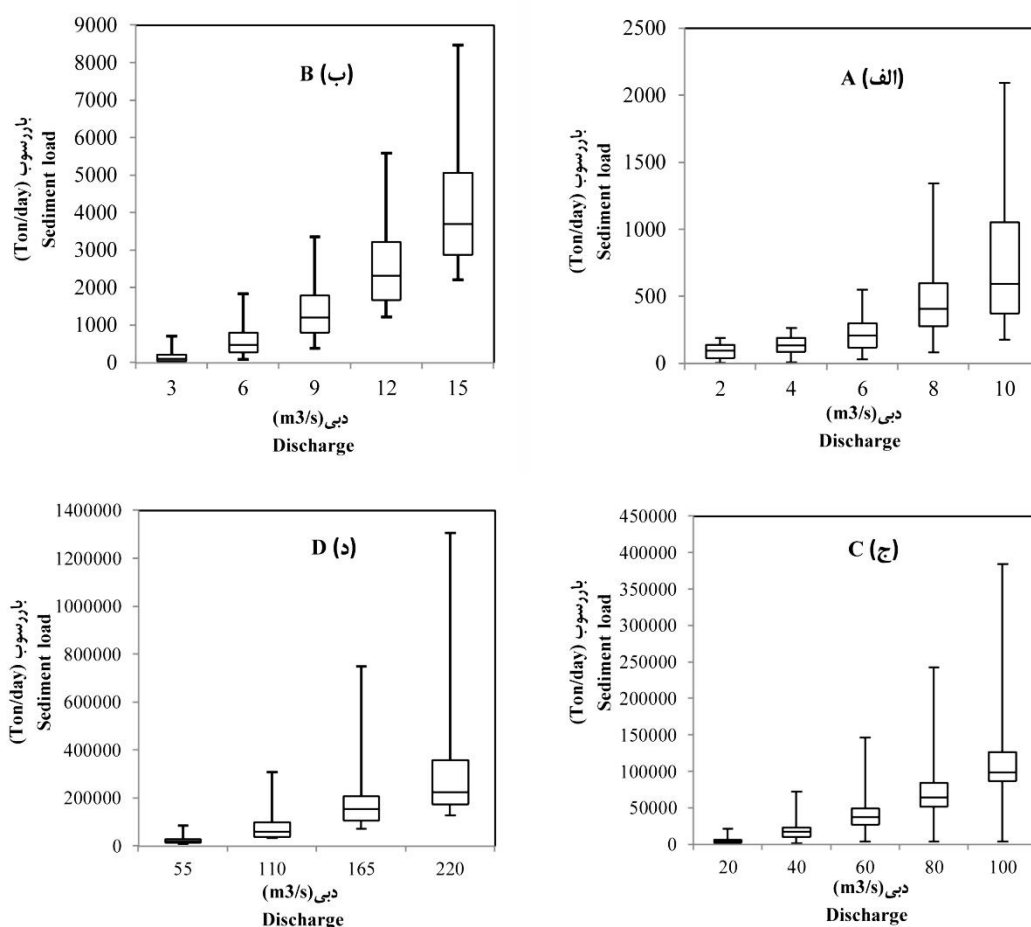
کم‌ترین احتمال آورد رسوب برای دبی $1/5 \text{ m}^3/\text{s}$ در ایستگاه ارازکوسه (شکل ۵ ب)) $0/003$ است که برای رسوبات بیش‌تر از $143/5$ تن بر روز رخ می‌دهد و کم‌ترین احتمال وقوع برای دبی $4/9 \text{ m}^3/\text{s}$ برابر $0/006$ است که در بار رسوب 8010 تن بر روز حاصل شده است (شکل ۶ ج)). کم‌ترین مقدار آورد رسوب در دبی $21/2 \text{ m}^3/\text{s}$ (شکل ۶ د)) مربوط به بار رسوب بیش‌تر از 14810 تن بر روز است.

باکس پلات چارک اول (اولین صدک)، میانه (چارک دوم)، چارک سوم و حداکثر (صدک صدم) تغییرات بار رسوب در هر دبی را نشان می‌دهد که برای مقایسه توزیع احتمال مقدار رسوب در دبی‌های مختلف در برنامه‌ریزی‌ها مناسب است. (شکل - ۶) باکس پلات بار رسوبی برای دبی‌های مختلف در ایستگاه‌های جنگلده، نوده، ارازکوسه و قزاقلی را نشان می‌دهد. در تمام شکل‌ها چولگی به چپ وجود دارد ولی در دبی‌های کم‌تر، مقدار چولگی کم می‌باشد و مشابه حالت نرمال است. مثلاً در دبی‌های ۲، ۴ و ۶ مترمکعب بر ثانیه برای ایستگاه جنگلده (شکل ۷ الف))، دبی‌های ۳، ۶ و ۹ مترمکعب بر ثانیه ایستگاه نوده (شکل ۷ ب)) و دبی‌های ۲۰، ۴۰ و ۶۰ مترمکعب بر ثانیه ایستگاه ارازکوسه (شکل ۷ ج)) چولگی به چپ به مقدار کمی وجود دارد و میانه تقریباً در وسط باکس قرار گرفته است و توزیع داده‌ها تقریباً متقارن است اما با افزایش دبی، مقدار چولگی نیز افزایش یافته است. در دبی‌های ۸ و ۱۰ مترمکعب بر ثانیه، خط میانه فاصله کم‌تری با چارک اول دارد و زمانی که دبی از ۴ به ۱۰ مترمکعب بر ثانیه افزایش می‌یابد، فاصله چارک سوم با میانه نیز از ۵۴ به ۴۵۸ تن بر روز می‌رسد که نشان می‌دهد این فاصله با افزایش دبی حدوداً ۸ برابر شده است (شکل ۷

1- Interquartile

مشاهده‌ای با رسوب مدل‌سازی شده به ترتیب در بازه‌های ۰-۳ و ۰-۲۰ مترمکعب بر ثانیه مربوط به شبیه‌سازی رسوب با روش‌های جنگل تصادفی و رگرسیون چندک است. در ایستگاه قزاقلی برآورد منحنی سنجه با روش رگرسیون چندک تفاوت زیادی ندارد و برتری خاصی بین این دو روش مشاهده نشد.

در ایستگاه جنگلده و با مدل جنگل تصادفی بوده است که میانگین رسوب حاصل شده از روش جنگل تصادفی کم‌ترین اختلاف را با میانگین مشاهده‌ای داشته است و نامطلوب‌ترین رسوب مدل‌سازی شده در این ایستگاه، مربوط به بازه $8-10 \text{ m}^3/\text{s}$ با روش SRC با اختلاف ۱۲۲۹ تن بر روز می‌باشد. در دو ایستگاه نوده و ارازکوسه، کم‌ترین اختلاف مقادیر



شکل ۶- باکس پلات بار رسوب روزانه برای دبی‌های مختلف در ایستگاه‌های: الف) جنگلده، ب) نوده، ج) ارازکوسه و د) قزاقلی.

Figure 6. Box plot of daily suspended sediment load given various discharge in stations : A) Jangal-deh, B) Nodeh, C) Arazkoose and D) Ghazaghlī.

جدول ۳- مقادیر رسوب برآورد شده (Ton/day) با سه مدل SRC، QR و RF در ایستگاه‌های جنگلده، نوده، ارازکوسه و قزاقلی.

Table 3. Sediment load (Ton/day) estimated using three models QR, SRC and RF in Jangaldehy, Nodeh, Arazkoose and Ghazaghli Stations.

میانگین رسوب مدل SRC	میانگین رسوب مدل RF	میانگین رسوب مدل QR	میانگین رسوب مشاهده‌ای	دبی (m ³ /s)	نام ایستگاه
Average sediment model SRC	Average sediment model RF	Average sediment model QR	Average sediment observational	Discharge	Station name
5	9	-	8	0-2	جنگلده Jangaldehy
29	88	-	37	2-4	
65	90	-	79	4-6	
104	102	-	171	6-8	
195	1983	-	1424	8-10	
26	33	-	35	0-3	نوده Nodeh
100	218	-	159	3-6	
238	668	-	547	6-9	
438	2170	-	1461	9-12	
558	2211	-	1591	12-15	
500	-	691	721	0-20	ارازکوسه Arazkoose
5014	-	8526	13958	20-40	
11236	-	21216	18804	40-60	
21696	-	44836	69216	60-80	
31253	-	67849	38890	80-100	
1868	-	1861	2748	0-55	قزاقلی Ghazaghli
33114	-	34670	61080	55-110	
90152	-	96496	107354	110-165	
220708	-	241382	173692	165-220	

نتیجه‌گیری

ارزیابی حجم رسوبات معلق در طیف وسیعی از مباحث، مانند طراحی بهینه و عملکرد مناسب سازه‌های آبی: مخزن، سد و کانال؛ ارزیابی زیست‌محیطی پروژه‌ها: آلودگی رودخانه‌ها، دریاچه‌ها و منابع آب نیازمند تخمین دقیق بار رسوب رودخانه است. زمان‌بر و پرهزینه بودن آزمایش‌ها مربوط به نمونه‌برداری رسوبات معلق سبب شده پژوهشگران از روش‌های غیرمستقیم برآورد بار رسوبی استفاده کنند. یکی از

به‌طورکلی در ایستگاه‌های مختلف و در دبی‌های متفاوت، روش‌های جنگل تصادفی و رگرسیون چندک نتایج بهتر و یا یکسان با منحنی سنجه ارائه کرده‌اند. اگر برآورد بار رسوبی با هدف مدیریت و به‌منظور جلوگیری یا کاهش خسارت‌های حاصل از پدیده‌های سیلابی باشد، دو روش رگرسیون چندک و جنگل تصادفی به‌دلیل شبیه‌سازی مناسب‌تر بار رسوبی و برآورد در احتمالات مختلف کارایی بیشتری در پیش‌بینی بار رسوب دارند.

می‌کنند و عدم قطعیت مقدار بار رسوب شبیه‌سازی شده را پیش‌بینی می‌کنند. این نتایج با پژوهش‌های فرنک و همکاران (۲۰۰۸)، فرنک (۲۰۰۹) و زنگ‌شیا و جوچن (۲۰۱۵) مطابقت داشته است (۱۰، ۱۱ و ۳۵). یکی از مهم‌ترین مزایای استفاده از دو روش جنگل تصادفی و رگرسیون چندک تخمین احتمال وقوع حوادث سهمگین سیلابی است که با بار رسوبی فراوانی همراه است. استفاده از این دو روش، این فرصت را در اختیار برنامه‌ریزان و مدیران منابع آب قرار می‌دهد تا با ارزیابی خطر وقوع چنین رویدادهایی که سبب آسیب به سازه‌های هیدرولیکی می‌شوند و متعاقباً خسارات مالی و جانی فراوانی دارند اقدامات و برنامه‌ریزی‌های لازم برای پیشگیری یا مقابله با این پدیده‌ها صورت گیرد. اگرچه استفاده از روش‌های جنگل تصادفی و رگرسیون چندک قابلیت زیادی در شبیه‌سازی و تعیین عدم قطعیت بار رسوب رودخانه‌ها دارند اما استفاده از این روش‌ها به سادگی استفاده از منحنی سنج رسوب کلاسیک نیست و کاربر باید ضرایب مربوط به این روش‌ها را بهینه کند که ممکن است فرآیندی وقت‌گیر باشد.

راه‌های ساده و آسان برآورد غیرمستقیم بار رسوبی، منحنی سنج رسوب است که بین دبی و رسوب متناظر رابطه تجربی برقرار می‌کند. از آنجایی که این روش تنها می‌تواند معرف یک مقدار رسوب در یک دبی معین باشد و به علت وجود پدیده پسماند در طبیعت و شرایط فصلی، برای یک دبی مشخص چندین مقدار بار رسوبی وجود دارد و نیز با پیشرفت روزافزون کاربرد مدل‌های کامپیوتری در سال‌های اخیر، استفاده از مدل‌های شبیه‌سازی برای پیش‌بینی بار رسوبی رودخانه‌ها گسترش یافته است. از این‌رو، در پژوهش حاضر از روش‌هایی که بتوانند مقدار بار رسوب را برای یک دبی معین در احتمالات گوناگون تخمین بزنند استفاده شد. بدین‌منظور از دو روش رگرسیون چندک و جنگل تصادفی برای مدل‌سازی بار رسوب چهار ایستگاه (جنگلده، نوده، ارازکوسه و قراقلی) واقع بر رودخانه گرگانرود در استان گلستان بهره گرفته شد. نتایج نشان داد که روش‌های رگرسیون چندک و جنگل تصادفی فقط بار رسوب را در دبی‌های مختلف پیش‌بینی نمی‌کنند بلکه مقدار احتمال آورد رسوب در یک دبی ثابت را تعیین

منابع

1. Afan, H.A., El-Shafie, A., Yaseen, Z.M., Hameed, M.M., Wan Mohtar, W.H.M., and Hussain, A. 2015. ANN based sediment prediction model utilizing different input scenarios. *Water Resour. Manage.* 29: 4. 1231-1245.
2. Alagidede, P., and Panagiotidis, T. 2012. Stock returns and inflation evidence from quantile regressions. *Economics Letters.* 117: 1. 283-286.
3. Clarke, R.T. 1990a. Statistical characteristics of some estimators of sediment and nutrient loadings. *Water Resources Research.* 26: 9. 2229-2233.
4. Clarke, R.T. 1990b. Bias and variance of some estimators of suspended sediment load. *Hydrological Sciences Journal-des Sciences Hydrologiques.* 35: 3. 253-261.
5. Cozzoli, F., Bouma, T.J., Ysebaert, T., and Herman, P.M.J. 2013. Application of non-linear quantile regression to macrozoobenthic species distribution modelling: comparing two contrasting basins. *Marine Ecology Process Series.* 475: 119-133.
6. Crawford, C.G. 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. *J. Hydrol. Elsevier Science publishers B.V., Amsterdam.* 129: 331-348.
7. Fallah, A., Kalbi, S., Shatai Juybari, S.H., and Karami, A. 2015. Determine the classification and regression tree methods landsat and ASTER random forest mapping forest type. *Forest and wood chips, Iran. J. Natur. Resour.* 67: 4. 573-584. (In Persian)

8. Farhadi, Z., and Shahsavani, D. 2016. Gene expression data clustering with random forest dissimilarity. *Razi J. Med. Sci.* 22: 136-109-118. (In Persian)
9. Fath Abadi, A., Seyedian, S.M., and Rouhani, H. 2016. Suspended sediment load using a combination of support vector machines and genetic algorithms. First national conference on soil conservation and watershed management. 8-10 February. (In Persian)
10. Francke, T., Lopez-Tarazon, J.A., and Schroder, B. 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. *Hydrological Process.* 22: 25. 4892-4904.
11. Francke, T.K.O. 2009. Measurement and modeling of water and sediment fluxes in meso-scale dryland catchmentes. PhD Thesis. Germany: University of Potsdam, 137p.
12. Hayatzadeh, M., Ekhtesasi, M.R., Malekinezhad, H., Fathzadeh, A., and Azimzadeh, H.R. 2017. Determination of the Optimal Method to Estimate the Suspended Sediment Load in Arid Regions (Case study: Fakhrabad basin in Mehriz-Yazd). *J. Water Soil Sci. (Science and Technology Agriculture and Natural Resources)*. 21: 1. 113-125. (In Persian)
13. Hicks, D.M., Gomez, B., and Trustrum, N. 2000. Erosion thresholds and suspended sediment yields, Waipoa River Basin, New Zealand. *Water Resources Research.* 36: 1129-1142.
14. Kitsikoudis, V., Sidiropoulos, E., and Hrissanthou, V. 2014. Machine learning utilization for bed load transport in gravel-bed Rivers. *Water Resources Management.* 28: 11. 3727-3743.
15. Krishnaswamy, J., Richter, D., Halpin, P., and Hofmockel, M. 2001. Spatial patterns of suspended sediment yields in a humid tropical watershed in Costa Rica. *Hydrological Processes.* 15: 12. 2237-2257.
16. Kakaei Lafdani, E., Moghaddam Nia, A., and Ahmadi, A. 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *J. Hydrol.* 478: 50-62.
17. Koenker, R., and Basset, G. 1978. Regression quantiles. *Econometrica.* 46: 1. 33-50.
18. Koenker, R., and D'Orey, V. 1987. Computing regression quantiles. *Apply Statistics.* 36: 3. 383-393.
19. Lohani, A.K., Goel, N.K., and Bhatia, K.K.S. 2007. Deriving stage-discharge-sediment concentration relationship using fuzzy logic Hydrological sciences. *Journal-des Science Hydrologiques.* 52: 4. 793-807.
20. McBean, E.A., and Al-Nassri, S. 1988. Uncertainty in suspended sediment transport curves. *J. Hydr. Engin.* 114: 1. 63-74.
21. Mostafazadeh, R., Sadeghi, S.H.R., and Sadoddin, A. 2015. Analysis of storm-wise sedimentgraphs and rating loops in Galazchai Watershed, West-Azarbaijan. *J. Water Soil Cons.* 21: 5. 175-191. (In Persian)
22. Mustafa, M.R. 2016. Modelling daily suspended sediments of a hyper-concentrated river in Malaysia. *Arpn J. Engin. Appl. Sci.* 11: 4. 2141-2145.
23. Najafinejad, A., Mardian, M., Varvani, J., and Sheikh, V.B. 2011. Performance Evaluation of Correction Factors in Optimization of Sediment Rating Curve (Case study: Kamal Saleh Dam Watershed, Markazi province, Iran). *J. Water Soil Cons.* 18: 2. 105-122. (In Persian)
24. Partal, T., and Cigizoglu, H.K. 2008. Estimation and forecasting of daily suspended sediment data using wavelet-neural network. *J. Hydrol.* 358: 317-331.
25. Rustomji, P., and Wilkinson, S.N. 2008. Applying bootstrap resampling to quantify uncertainty in fluvial suspended sediment loads estimated using rating curves. *Water Resources Research.* 44: 10. 1029/2007WR006088.
26. Salarijazi, M. 2017. Quantiles trend estimation of variables of annual maximum floods. *J. Water Soil Cons.* 24: 1. 25-46. (In Persian)
27. Salarijazi, M., Abdolhosseini, M., Ghorbani, Kh., and Eslami, S. 2016. Evaluation of quasi-maximum likelihood and smearing estimator to improve sediment rating curve estimation. *Inter. J. Hydrol. Sci. Technol.* 6: 4. 173-181.

28. Salmani, H., Mohseni Saravi, M., Rouhani, H., and Salajeghe, A. 2013. Evaluation of land use change and its impact on the hydrological process in the Ghazaghli watershed Golestan province. *J. Water. Manage.* 3: 6. 43-60. (In Persian)
29. Seyedian, S.M., and Rouhani, H. 2015. Assessing ANFIS accuracy in estimation of suspended sediments. *Gradevinar.* 67: 12. 1225-1236.
30. Sheikhal Pour, Z., Hassan Pour, F., and Azimi, V. 2015. Comparison of artificial intelligence methods in estimation of suspended sediment load (Case study: Sistan River). *J. Water Soil Cons.* 22: 2. 41-60. (In Persian)
31. Shiau, J.T., and Huang, W.H. 2015. Detecting distributional changes of annual rainfall indices in Taiwan using quantile regression. *J. Hydro. - Environ. Res.* 9: 3. 368-380.
32. Singh, K.P., and Durgunog-lu, A. 1989. Developing accurate and reliable stream sediment yields. *Sediment and the Environment*, edited by R.F. Hadley and E.D. Ongley, IAHS Publ. 184: 193-199.
33. Svetnik, V., Liaw, A., Tong, C., Culberson, J., Sheridan, R.P., and Feuston, B.P. 2003. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Com. Sci.* 43: 1947-1958.
34. Tarras-Wahlberg, N.H., and Lane, S.N. 2003. Suspended sediment yield and metal contamination in a river catchment affected by El Niño events and gold mining activities: The Puyango River basin, southern Ecuador. *Hydrological Processes.* 17: 15. 3101-3123.
35. Tzong Shiau, J., and Ju Chen, T. 2015. Quantile regression based probabilistic estimation scheme for daily and annual suspended sediment loads. *Water Resources Management.* 29: 8. 2805-2818.
36. Varvani, J., and Khalighi Sigarudi, Sh. 2007. Check the error sediment rating curves for estimating sediment load of flood events in Gharechay River. *The new finding Agriculture.* 1: 3. 201-214. (In Persian)
37. Vigiak, O., and Bende-Michl, U. 2013. Estimating bootstrap and Bayesian prediction intervals for constituent load rating curves. *Water Resources Research.* 49: 8565-8578.
38. Walling, D.E. 1977b. Assessing the accuracy of suspended sediment rating curves for a small basin. *Water Resources Research.* 13: 3. 531-538.
39. Walling, D.E., and Webb, B.W. 1981. The reliability of suspend sediment load data. *Erosion and Sediment Transport Measurement. Proceeding of Florence Congress.* IAHS Pub. 133: 177-194.
40. Walling, D., and Webb, B.W. 1988. The reliability of rating curve estimates of suspended sediment yield some further comments. *Sediment Budgets.* Edited by M.P. Bordas and D.E. Walling. IAHS Publ. 174: 337-350.



Stochastic modeling of sediment yield using random forest and quantile regression

M. Javadi Alinejad¹, *S.M. Seyedian², H. Rouhani² and A.H. Fathabadi²

¹M.Sc. Graduate, Dept. of Watershed Management, University of Gonbad Kavous,

²Assistant Prof., Dept. of Watershed Management, University of Gonbad Kavous

Received: 01/20/2017; Accepted: 10/13/2017

Abstract

Background and Objectives: Assessment of suspended sediment load is very important. Water quality and environmental is under impression of sediment load. As well as the design of hydraulic structures and other water supply facilities, watershed management, soil conservation programs and another major problem caused by sediment yield is dependent on the accurate estimation of sediment load. As a direct estimation of sediment load is very difficult and time consuming, so this led the researchers to estimate sediment load as indirect that it is possible to resort to various methods. One easy way to indirectly estimate the sediment load is sediment rating curve. It can only represent invariable amount of sediment in flow and due to various factors in nature may be there is several sediment load for a known flow rate. On the basis of this study quantile regression and random forest methods was used that can estimate sediment load for a given flow rate in the various probability. The use of these two methods can be analyzed sediment load in great flood and special events.

Materials and Methods: In this study, sediment rating curve models, quantile regression and random forest was used to estimate sediment load in four stations Gorganrood River Jangaldehy, Nodeh, Arazkoose and Ghazaghli in Golestan province. For this purpose, flow and sediment data was collected at four studied stations and separated into two parts, 75% for training and 25% for testing. The rating curve was obtained using fitted power equation between discharge and sediment load. Quantile regression and random forest algorithms were implemented using R statistical software. The optimal values of the variable parameters of the two methods were determined using trial and error method. By running the model, the amounts of sediment associated with specified flow were calculated in different probability level (1% to 99%).

Results: Using these two methods, sediment load was estimated in quantiles 2.5, 50 and 97.5%, respectively and range of uncertainty was determined in each station. In Jangaldehy and Nodeh stations random forest were selected as best method with RMSE criterion 96 and 210 tons per day and quantile regression were selected as best method with RMSE criterion 6453 and 24886 tons per day in Arazkoose and Ghazaghli stations. Classic rating curve method estimate sediment load in Jangaldehy, Nodeh, Arazkoose and Ghazaghli stations with RMSE 199, 288, 7505 and 25811 tons per day respectively.

Conclusion: The results showed that classic sediment rating curve not only unable to estimate the sediment load in the range of uncertainties in specified flow rate but also estimates sediment load with more error. Quantile regression and random forest methods can be estimate sediment load in various probabilities for a specified flow and this has contributed greatly to accurate and comprehensive planning for the construction of hydraulic structures and in this way, the dangers of the destruction of the facility reduction due to the great flood.

Keywords: Sediment rating curve, Quantile regression, Random forest, Gorganrood River

* Corresponding Author; Email: s.m.seyedian@gmail.com