

## ارزیابی روش‌های رگرسیونی نوین جهت تخمین بار رسوبی معلق در رودخانه سیستان

سارا محمدی<sup>۱</sup>، فرزاد حسن‌پور<sup>۲</sup>، سلمان شریف‌آذری<sup>۳</sup>، فرید فروغی<sup>۴</sup>

تاریخ ارسال: ۱۳۹۸/۱۱/۱۰

تاریخ پذیرش: ۱۳۹۹/۰۷/۰۷

مقاله پژوهشی

### چکیده

برآورد صحیح حجم رسوب حمل شده توسط رودخانه از اهمیت زیادی در پروژه‌هایی آبی برخوردار است. این مطالعه بر روی ۱۶۸۲ زوج داده برداشت شده که در بازه زمانی ۱۳۷۴-۱۳۹۰ صورت گرفت از روابط تجربی انتقال رسوب، مدل‌های رگرسیونی بردار پشتیبان و نزدیکترین همسایه توسعه داده شده به منظور پیش‌بینی بار معلق رسوب رودخانه سیستان استفاده گردید. در این راستا علاوه بر متغیرهای دمای حداقل، دمای حداکثر، دبی و بار معلق برداشت شده، دبی کلاسه نیز به عنوان یک متغیر ورودی تاثیرگذار در مدل‌سازی رسوب معین مورد بررسی قرار گرفت. برای ترکیب‌های متفاوت از ورودی‌ها در هر کدام از روش‌های رگرسیون بردار پشتیبان و رگرسیون ناپارامتری بهترین ساختار مدل رگرسیونی توجه به معیارهای ارزیابی تعیین گردید. نتایج این تحقیق نشان داد که در بین روابط تجربی روش توفالتی با مقدار جذر میانگین مربعات ۶۶۵۵۸ تن در روز و ضریب تبیین ۰/۷۰۵ بهتر از سایر روش‌ها عمل کرده است. همچنین مدل رگرسیون بردار پشتیبان با ترکیب ورودی‌های دبی، دمای حداقل، دمای حداکثر و دبی کلاسه با مقدار جذر میانگین مربعات خطا ۲۸۰۹/۳ تن در روز و ضریب تبیین ۰/۹۶ بهترین نتیجه را در برآورد بار معلق رسوب رودخانه دارا بود. بر اساس نتایج این تحقیق روش‌های رگرسیونی با تفاوت قابل ملاحظه‌ای بهتر از روابط تجربی، بار معلق رسوب رودخانه سیستان را برآورد کردند.

**واژه‌های کلیدی:** انتقال رسوب، رگرسیون بردار پشتیبان، رگرسیون نزدیک‌ترین همسایه، روش توفالتی، مدل‌های تجربی.

<sup>۱</sup>دانشجو دکتری مهندسی عمران- سازه های هیدرولیکی، گروه مهندسی عمران، دانشگاه باهنر کرمان، کرمان، ایران. ۰۹۱۷۸۲۴۶۱۹۹. saramohamadi494@yahoo.com

<sup>۲</sup>دانشیار، گروه مهندسی آب، دانشگاه زابل، زابل، ایران. ۰۹۱۵۱۴۴۴۴۷۴. hassanpourir@uoz.ac.ir (نویسنده مسئول)

<sup>۳</sup>مریی، گروه مهندسی آب، دانشگاه زابل، زابل، ایران. ۰۹۱۴۷۲۹۸۷۹۴. s.sharifazari@gmail.com

<sup>۴</sup>بخش مرتع و آبخیزداری، دانشکده کشاورزی و منابع طبیعی داراب، دانشگاه شیراز، داراب، ایران. ۰۹۱۷۷۲۳۱۱۸۱. frooghi14@yahoo.com



## مقدمه

توابع عضویت در بردارد ( Kisi, 2005; Kisi et al., 2009; Kisi et al., 2006). عملکرد دو مدل شبکه عصبی چند لایه پیش‌خور و شبکه پایه شعاعی نسبت به رگرسیون چند متغیره در برآورد بار معلق روزانه حوضه جانیتا در پنسیلوانیای آمریکا مورد بررسی قرار گرفت. در این بررسی ورودی مدل‌های شبکه عصبی و رگرسیون چند متغیره، میزان دبی و بارندگی در گام‌های زمانی قبل بوده است. نتایج پژوهش انجام شده نشان داد که روش شبکه عصبی عملکرد بهتری نسبت به رگرسیون چند متغیره داشته است (Alp and Cigizoglu, 2007). بررسی مقایسه‌ای مدل‌های مختلف هوش مصنوعی نظیر شبکه عصبی مصنوعی، سیستم استنتاج فازی و برنامه‌ریزی بیان ژن جهت برآورد بار رسوبی رودخانه سیستان نشان داد سیستم تطبیقی عصبی-فازی کارایی بالایی جهت برآورد بار رسوبی رودخانه مذکور دارد (Sheikhalipour and Hassanpour, 2013). علاوه بر روش‌های هوشمند ذکر شده روش‌های رگرسیونی نوینی نیز در راستای برآورد رسوب معلق اخیراً مورد استفاده قرار گرفته است. در تحقیقی به منظور مدیریت منابع آب با کمک روش بردارهای پشتیبان مشخص شد به علت وجود شرایط مرسر در توابع کرنل مسائل بهینه‌سازی مربوطه همگرا بوده، بنابراین مشکل نقطه کمینه‌ی محلی وجود ندارد. نتایج بدست آمده از مدل‌ها نشان دادند که این روش عملکرد مناسبی در حل مسائل مکانی و لحظه‌ای در مدیریت منابع آب دارد (Asefa et al., 2006). مدل‌سازی رابطه‌ی دبی - رسوب با استفاده از ماشین بردار پشتیبان و مقایسه‌ی آن با مدل‌های شبکه عصبی و منحنی سنج رسوب بیان‌گر این مهم بود که مدل ماشین بردار پشتیبان کارایی بهتری نسبت به مدل‌های دیگر به کار رفته دارد (Kisi, 2012). مقایسه‌ی بین شبکه عصبی و رگرسیون ماشین‌های بردار پشتیبان جهت تخمین بار معلق کارایی بالای هر دو روش را در پی داشت (Kakaei lafdani et al., 2013). بررسی سوابق مطالعاتی در این زمینه گویای این واقعیت است که روش بردار پشتیبان خیلی سریعتر از شبکه‌های عصبی مصنوعی متداول آموزش می‌بیند (Cimen, 2008; Cristianini and Shawe, 2000; Jie )

برآورد صحیح حجم رسوب حمل شده توسط رودخانه از اهمیت زیادی در پروژه‌هایی آبی برخوردار است. از آنجا که نرخ انتقال رسوبات یا بزرگی آن‌ها می‌تواند با متغیر-هایی مثل دبی آب، سرعت جریان، شیب خط انرژی، تنش‌برشی، انرژی جریان، زبری نسبی، عدد فرود و ... تعیین شود، روش‌های تئوری انتقال رسوبات بر مبنای یک سری فرضیات می‌باشد. استفاده از این متغیرها بستگی به منطقه مورد مطالعه دارد، بگونه‌ای که گاهی اوقات فقط دو یا سه تا از این متغیرها استفاده می‌شود. در حقیقت در برخی موارد یک سری از معادلات در یک محل خاص بی-اعتبار است. در دهه‌های اخیر، محققان روش‌های متفاوتی را برای این امر در مقالات مختلف ارائه نموده و بکار گرفته‌اند که این روش‌ها عمدتاً شامل تکنیک‌های رگرسیونی، مدل‌های مفهومی و روش‌های پیچیده‌تر بر پایه‌ی هوش مصنوعی و اخیراً مدل‌های رگرسیون بردار پشتیبان است (فراحی و همکاران، ۱۳۹۰).

در دهه‌های اخیر تحقیقات ارزشمندی در راستای به کارگیری مدل‌های هوش مصنوعی در پیش‌بینی و شبیه‌سازی متغیرهای هیدرواقليمی انجام شده است (دهقانی و همکاران، ۱۳۹۰؛ رجایی و ابراهیمی، ۱۳۹۳). کارایی شبکه عصبی مصنوعی در مقایسه با منحنی سنج رسوب به منظور برآورد بار معلق رودخانه مورد ارزیابی واقع شده است که نتایج نشان داد که شبکه عصبی مصنوعی با دقت بسیار مناسب و با اطمینان بیشتر می‌تواند نسبت به روش منحنی سنج به و بدون دسته بندی داده‌ها برای تخمین بار معلق رسوب مورد استفاده قرار گیرد (دستورانی و همکاران، ۱۳۹۱؛ دهقانی و همکاران، ۱۳۸۸؛ دهقانی و وفاخواه، ۱۳۹۱؛ Shoushtari and Kashefipour, 2007). تخمین رسوبات بار معلق روزانه با استفاده از منطق فازی و مقایسه‌ی نتایج به دست آمده را با روش منحنی سنج نشان داد که عملکرد مدل عصبی - فازی نسبت به منحنی سنج بهتر بوده است. همچنین به کارگیری تابع عضویت مثلثی نتایج بهتری نسبت به سایر

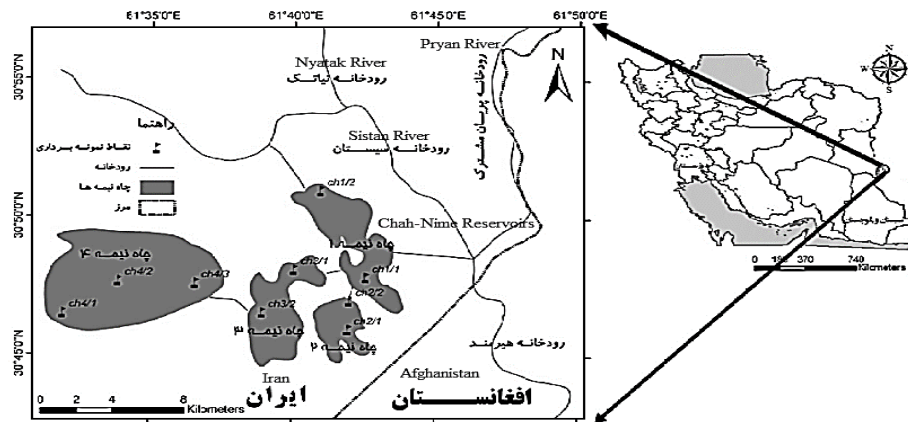
مدل‌های داده مبنا مانند روش‌های رگرسیونی نوین به عنوان راهکارهای جایگزین می‌تواند در نظر گرفته شود. لذا در تحقیق حاضر سعی می‌شود با به کارگیری مدل‌های ذکر شده و مقایسه‌ی نتایج آن‌ها با معادلات تجربی انتقال رسوب و مطالعات پیشین مدلی مناسب به منظور برآورد بار رسوبی رودخانه‌ی سیستان معرفی گردد. همچنین بررسی دقت و کارایی مدل توسعه داده شده‌ی رگرسیون نزدیک‌ترین همسایه برای تخمین مقادیر خارج از محدوده‌ی نوین در برآورد بار رسوبی معلق رودخانه سیستان از اهداف اصلی این تحقیق است.

## مواد و روش‌ها

### منطقه تحقیق و داده‌ها

این پژوهش با استفاده از آمار و اطلاعات مربوط به رودخانه سیستان واقع در جنوب شرق کشور به عنوان یکی از دو شاخه اصلی منشعب از رودخانه هیرمند، انجام شد. رودخانه مذکور وظیفه آبیاری ۷۰ درصد از اراضی کشاورزی دشت سیستان را به عهده دارد. این رودخانه به دلیل قرارگرفتن در پایاب حوضه آبریز وسیع هیرمند، مستعد ته‌نشینی رسوبات ریزدانه می‌باشد و انباشت سالانه مقادیر متناهی رسوب در طول مسیر رودخانه باعث افزایش تراز بستر و در نتیجه کاهش ظرفیت عبور جریان گردیده به نحوی که حتی در دبی‌های متوسط و کم نیز فرار آب از کناره رودخانه به دشت‌های مجاور به چشم می‌خورد (حسن‌پور و طباطبائی، ۱۳۸۸). شکل یک موقعیت جغرافیایی رودخانه‌ی سیستان را نشان می‌دهد.

(and Yu, 2011; Torabi and dehghani, 2018.)  
نتایج آنالیز مدل‌سازی رواناب و تخمین میزان رسوب حمل شده با استفاده از مدل‌های ANN، LS-SVR، REPTREE و M5 بیان‌کننده دقیق‌تر بودن روش ANN و LS-SVR نسبت به سایر مدل‌ها مورد استفاده بوده است (Bharti et al., 2017). در روش رگرسیون بردار پشتیبان مبتنی بر الگوریتم ژنتیک (GA) جهت تعیین پارامترهای بهینه، مدل‌های GA-SVR با هسته عملکرد شعاعی نمایی (ERBF) دقت بالاتری نسبت به سایر مدل‌های کاربردی GA-SVR دارند (Roshangar and koshi, 2015). روش‌های رگرسیونی بردار پشتیبان در حالت طرح ترکیبی مبتنی بر پیوند موجک (wavelet conjunction) در مقایسه با حالت طرح عادی عملکرد بهتری را به همراه دارد (Hazarika et al., 2020). همچنین در ارزیابی که بین روش‌های رگرسیون بردار پشتیبان و مدل درختی (M5) جهت مدل‌سازی رسوب معلق رودخانه انجام گرفت کارایی بالای هر دو روش بیان شد (ستاری و همکاران، ۱۳۹۵). از دیگر روش‌های رگرسیونی که در سال‌های اخیر به طور گسترده در پیش بینی و برنامه ریزی منابع آب کاربرد داشته است روش ناپارامتری نزدیکترین همسایه است (Lall and Sharma, 1996; Prairie et al., 2006. Vinod, 2008). روش نزدیکترین همسایه همانند روش رگرسیون بردار پشتیبان دارای کارایی بالایی جهت برآورد بار رسوبی رودخانه می‌باشد (joudi and sattari, 2017). مزیت بارز برآوردهای ناپارامتری این است که منجر به حداقل فروض به منظور فرآیند تولید داده‌ها می‌شود (Vinod, 2014). با توجه به غیر خطی بودن رابطه‌ی رسوب و سایر پارامترهای موثر بر آن، استفاده از مدل‌هایی که بتوانند فرآیندهای غیر خطی را با اطمینان بالا شبیه سازی کنند یک راهکار مطمئن در برآورد رسوب می‌باشد. به دلیل زمان‌بر بودن آموزش مدل‌هایی نظیر شبکه‌های عصبی مصنوعی و سیستم استنتاج فازی استفاده از دیگر



شکل (۱) موقعیت جغرافیایی رودخانه سیستان

به مدل‌های مورد بحث استفاده شد. علاوه بر متغیرهای ذکر شده جهت افزایش دقت برآورد رسوب روزانه دبی کلاسه رودخانه ( $Q_c$ )، متوسط دمای روزانه ( $T_{mean}$ ) و اختلاف دمای بیشینه و کمینه روزانه ( $T_d$ ) نیز به عنوان دیگر متغیرهای ورودی در نظر گرفته شد. پس از عملیات استاندارد سازی و حذف داده‌های پرت تعداد اطلاعات ثبت شده و قابل اعتماد برابر با ۱۶۸۲ زوج داده بود که از این تعداد ۷۰ درصد به منظور آموزش مدل‌های ذکر شده و ۳۰ درصد باقی مانده برای آزمایش (تست) آن‌ها به کار رفت. مشخصات آماری داده‌های دبی و دبی رسوب در برای هر دو دسته‌ی آموزش و تست در جدول شماره ۱ ذکر شده است.

به منظور تخمین بار رسوبی معلق رودخانه سیستان از معادلات تجربی موجود در مدل HEC-RAS، نرم افزار Libsvm جهت انجام عملیات یادگیری و استفاده از ماشین پشتیبان استفاده شد. همچنین جهت برازش مدل نزدیک‌ترین همسایه نیز مدل توسعه داده شده توسط شریف آذری و عراقی نژاد با کمی تغییرات در راستای دسترسی به عملکرد بهتر مورد استفاده قرار گرفت.

از داده‌های ثبت شده‌ی دبی جریان ( $Q$ )، دبی رسوب ( $Q_s$ )، دمای حداکثر ( $T_{max}$ ) و دمای حداقل ( $T_{min}$ ) حوضه‌ی آبریز رودخانه در بازه‌ی زمانی ۱۳۷۴ تا ۱۳۹۰ به صورت روزانه برداشت شده‌اند، به عنوان اطلاعات ورودی

جدول (۱) ویژگی‌های آماری داده‌های مشاهداتی

پارامتر	میانگین	انحراف از معیار	کمینه	بیشینه
دبی ( $m^3/s$ )	91.78	167.58	0.2	1105.5
دبی رسوب (ton/day)	44763.16	123405.5	1.33	1566836
دبی ( $m^3/s$ )	91.04	176.38	0.4	1178.45
دبی رسوب (ton/day)	41544.18	120394	1.62	1121680

## رگرسیون بردار پشتیبان

$$L_{\varepsilon}(\hat{y}_i) = |y - f(x)|_{\varepsilon} = \begin{cases} 0 & \text{for } |y - f(x)| < \varepsilon \\ |y - f(x)| & \text{for } |y - f(x)| \geq \varepsilon \end{cases} \quad (3)$$

این رابطه بیانگر این است که خطاهای کوچکتر از  $\varepsilon$  در نظر گرفته نمی‌شود. وپنیک (۱۹۹۵) مسئله SVR را به صورت مسئله بهینه‌سازی زیر بیان کرد:

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi'_i) \\ & \text{subject to } \begin{cases} y_i - (\langle w, x_i \rangle + b) \leq \varepsilon + \xi_i \\ (\langle w, x_i \rangle + b) - y_i \leq \varepsilon + \xi'_i \\ \xi_i, \xi'_i \geq 0 \text{ for all } i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (4)$$

در این رابطه،  $\xi_i$  و  $\xi'_i$  متغیرهای سرپاره هستند که حدود بالا و پایین خطای آموزش را نشان می‌دهند و  $C$  خطای تولید بوده ( $C > 0$ ) و به منظور موازنه بین  $\varepsilon$  و  $f$  لحاظ شده است. برای اینکه مدل بتواند داده‌های با پیچیدگی بالا را پیش‌بینی کند، داده‌ها به وسیله تابع  $\Phi$  به فضای با ابعاد بالا منتقل می‌شوند. برای حل مسئله با ابعاد خیلی بالا از قضیه دوگانگی لاگرانژ برای تبدیل مسئله کمینه‌سازی مورد نظر به فرم دوگانی آن استفاده می‌گردد که در آن به جای تابع پیچیده  $\Phi$  که به فضایی با ابعاد بالا می‌رود، از تابع ساده‌تری به نام تابع کرنل استفاده نمود. توابع کرنل مختلفی نظیر خطی<sup>۴</sup>، چندجمله‌ای<sup>۵</sup>، تابع پایه شعاعی<sup>۶</sup> و سیگموئید برای مدل کردن رگرسیون داده‌ها وجود دارد. بنابراین در مسائل غیرخطی کفایت از کرنل مقادیر ورودی به جای خود تابع استفاده گردد. مقدار پارامترهای  $C$  و  $\varepsilon$  و نیز نوع تابع کرنل می‌تواند در میزان خطای مدل تاثیر بسزایی داشته باشد. در این پژوهش دو نوع رگرسیون بردار پشتیبان (epsilon-SVR و nu-SVR)، با توابع کرنل خطی، چند جمله‌ای، تابع شعاعی (RBF) و سیگموئید مورد بررسی قرار گرفت.

رگرسیون بردار پشتیبان (SVR<sup>۱</sup>) که تعمیم یافته‌ی ماشین بردار پشتیبان (SVM<sup>۲</sup>) است، توسط وپنیک<sup>۳</sup> معرفی گردید (Vapnik, 1995). در این روش بر خلاف روش SVM که داده‌های خروجی در فضای باینری بودند، خروجی‌ها در فضای پیوسته بوده و هر نوع خروجی می‌تواند داشته باشند. به عبارت

دیگر اگر مجموعه داده‌های آموزشی به صورت  $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$  باشد، که بردار ورودی  $x_i$  و بردار خروجی  $y_i$  بوده، هدف SVR یافتن تابع رگرسیون  $f(x)$  به صورت زیر است.

$$\hat{y} = f(x) = \langle w, x \rangle + b \quad (1)$$

در این رابطه،  $w$  و  $b$  به ترتیب وزن‌ها و عرض از مبدأ تابع رگرسیون و  $\langle \rangle$  علامت ضرب داخلی می‌باشد. هستند. هرچه مقدار  $\|w\|$  کمتر باشد، پیچیدگی مدل نیز کمتر خواهد بود. بر اساس اصل استقرای خطای ساختاری پارامترهای  $w$  و  $b$  با کمینه کردن عبارت زیر که حاصل-جمع  $\|w\|$  و میزان ریسک عملیاتی است، به دست می‌آید.

$$R = \frac{1}{2} w^T w + C \sum_{i=1}^n L_{\varepsilon}(\hat{y}_i) \quad (2)$$

در این رابطه،  $\varepsilon$  = میزان انحراف اعمال شده است. وپنیک (۱۹۹۵) برای ساخت شکل رگرسیونی SVM یا همان SVR تابع خطای جدیدی به نام  $L_{\varepsilon}$  را معرفی کرد که این تابع مقدار خطاهایی که در یک  $\varepsilon$ -insensitive (فاصله معین) از مقادیر واقعی باشند را نادیده می‌گیرد.

<sup>۱</sup> Linear

<sup>۲</sup> Polynomial

<sup>۳</sup> Radial Basis Function

<sup>۱</sup> Support Vector Regression

<sup>۲</sup> Support Vector Machine

<sup>۳</sup> Vapnik



$$GVC = \frac{\sum_{i=1}^n e_i^2 / n}{(1 - 1 / \sum_{j=1}^k 1/j)^2} \quad (۶)$$

مقادیر وزن‌ها از صفر تا یک با تغییرات ۰/۱ در نظر گرفته شد. تعداد همسایه‌ها برای  $n$  بزرگتر از ۵۰ و کمتر از ۲۰۰ از هفت تا ۱۴ در نظر گرفته شد (رجائی و ابراهیمی، ۱۳۹۳) و برای مقادیر بیشتر از ۲۰۰ نیز از هفت تا  $n^{0.5}$  در نظر گرفته شد. مقادیر کم  $k$  باعث می‌شود که مقدار پیش‌بینی شده شبیه یکی از همسایه‌ها و مقادیر زیاد آن سبب می‌شود که مقدار پیش‌بینی شده نزدیک به مقدار میانگین داده‌های وابسته مشاهداتی شود. همچنین جهت رفع ناتوانی الگوریتم در بازتولید اعداد خارج از محدوده‌ی تاریخی از روش به کار رفته توسط شریف آذری و عراقی‌نژاد (۱۳۹۲) استفاده شد. در روش مذکور تخمین مقدار متغیر وابسته در سه مرحله صورت می‌گیرد. ابتدا با استفاده از مدل بازنمونه‌گیری KNN داده‌های مشاهداتی شبیه‌سازی می‌شوند. سپس از تفاضل داده‌های واقعی و داده‌های شبیه‌سازی شده، سری باقی‌مانده‌ها به دست می‌آید. در گام آخر مقدار تخمینی متغیر وابسته از حاصل جمع مقدار به دست آمده با استفاده از رگرسیون ناپارامتری KNN و یکی از مقادیر باقی‌مانده‌ها که به صورت احتمالاتی از سری باقی‌مانده‌ها انتخاب می‌شود، محاسبه می‌گردد. جزئیات بیشتر به شریف آذری و عراقی‌نژاد (۱۳۹۲) مراجعه شود.

#### بررسی کارایی روش‌های رگرسیونی نوین با استفاده معادلات تجربی

به منظور بررسی کارایی مدل‌های رگرسیونی به کار رفته در این پژوهش علاوه بر مقایسه خروجی آن‌ها با داده‌های واقعی نتایج به دست آمده با نتایج معادلات تجربی مشهور در زمینه تخمین رسوب معلق مقایسه شد. برای شبیه‌سازی رسوب روزانه از معادلات تجربی مدل HEC-RAS استفاده شد. معادلات بکار گرفته شده در این مدل ریاضی عبارتند از: معادله یانگ، توفالتی، انگلند - هسنس، ایگرز - وایت، ویلکوک، مایر - پیتر و مولر .

#### رگرسیون ناپارامتری K- نزدیکترین همسایه

یکی از روش‌های آماری که می‌توان برای دسته‌بندی، شناخت الگوها، پیش‌بینی و مدل‌سازی داده‌ها استفاده کرد، روش ناپارامتری برآورد توابع است. یکی از شناخته شده ترین مدل‌های ناپارامتری روش  $K$ -NN<sup>۱</sup> می‌باشد که به دلیل سهولت ذاتی و کاربردها و قابلیت‌های آن مورد استفاده گسترده قرار گرفته است. در این روش تشابه الگوی داده‌های مورد نظر با الگوی کلی داده‌ها مورد مقایسه و ارزیابی قرار می‌گیرد. در این روش به منظور تعیین بهترین همسایگی‌ها نیاز به یک تابع فاصله‌سنجی است (شریف‌آذری و عراقی‌نژاد، ۱۳۹۲). تابع فاصله‌سنجی در این مطالعه تابع فاصله‌سنجی اقلیدسی انتخاب گردید، که براساس فاصله متریک دو نقطه در فضای اقلیدسی تعریف می‌گردد. برای هر نقطه  $x$ ، فاصله اقلیدسی تا همه نقاط مرجع  $v = 1, 2, \dots, n$  به شرح زیر بدست می‌آید. اگر  $m$  متغیر پیش‌بینی‌کننده و  $n$  تعداد مشاهدات در نظر گرفته شود، فاصله اقلیدسی به شرح زیر بدست می‌آید.

$$D_{rs} = \sqrt{\sum_{j=1}^m W_j (X_{jr} - X_{js})^2} \quad (۵)$$

در این رابطه،  $X_{jr}$  متغیر پیش‌بینی‌کننده  $r$ ام نقطه  $r$  در سری نقاط  $i = 1, 2, \dots, n$ ،  $X_{js}$  متغیر پیش‌بینی‌کننده مشاهده شده  $r$ ام که متغیر وابسته  $Y_i$  توسط این مقدار تخمین زده می‌شود و مقادیر  $W_j$  وزن‌هایی است که برای پیش‌بینی‌کننده‌ها در نظر گرفته می‌شود. در این الگوریتم دو نوع پارامتر مهم وجود دارد که دقت پیش‌بینی را کنترل می‌کنند: اول تعداد همسایه‌ها ( $k$ ) و دوم وزن هر یک از متغیرهای مستقل ( $W$ ). جهت تعیین مقدار بهینه وزن‌ها و تعداد مناسب همسایه‌ها از روش اعتبارسنجی تعمیم یافته<sup>۲</sup> استفاده شد (رابطه (Error! Reference source not found.)).

<sup>۱</sup> K Nearest Neighbor

<sup>۲</sup> Generalized Cross Validation (GCV)

برای رسیدن به بهترین جواب با بیشترین دقت که دارای بالاترین ضریب همبستگی و کمترین خطای آموزش و آزمون است به صورت بهینه و با استفاده از سعی و خطا تعیین محدوده مشخص، تعیین شد. پارامترهای رگرسیون ناپارامتری نزدیک‌ترین همسایه نیز با استفاده از روش ذکر شده به صورت بهینه تعیین شد.

### نتایج و بحث

نتایج به دست آمده برای ترکیبات ورودی‌های مختلف در مدل‌های مورد استفاده در جداول ۳ و ۲ ارائه شده است. جهت سنجیدن حساسیت مدل‌های مورد استفاده به تعداد و نوع متغیرهای ورودی و تعیین بهینه‌ترین ترکیب، هفت مدل متفاوت از ترکیب ورودی‌ها به منظور مشخص نمودن متغیرهای تأثیرگذارتر معرفی شد. نتایج بیانگر عملکرد بهتر روش رگرسیون بردار پشتیبان در مقایسه با روش نزدیک‌ترین همسایه می‌باشد که می‌تواند به دلیل به کار گیری اصل کمینه سازی خطای ساختاری در فرآیند مدل‌سازی باشد. در فرآیند بهینه سازی روش مذکور برای هر یک از مدل‌های ارائه شده مشخص شد که تابع کرنل RBF و روش nu-SVR بهترین عملکرد را در بالابردن دقت برآورد روش دارد که با نتایج حاصله از مطالعات شهبازی و کاکایی لافدانی همخوانی دارد (Kakaei, lafdani Et al., 2013؛ نیک‌بخت شهبازی و همکاران، ۱۳۹۱). به طوری که بهترین نتیجه به دست آمده در مورد مدل هفت نشان داد که هسته تابع پایه شعاعی با پارامترهای C برابر با ۱/۸، E برابر با ۰/۰۰۵ و گامای هشت در روش nu-SVR بهترین عملکرد را دارد. در مدل هفتم علاوه بر متغیرهای ذکر شده متغیر دبی کلاسه رودخانه (Qc)، به منظور رسیدن به بهترین جواب یا بیشترین دقت به عنوان یک پارامتر مستقل ورودی در نظر گرفته شد و همان‌طور که مشاهده می‌شود در هر دو روش رگرسیونی به کار گرفته شده استفاده از دبی کلاسه به عنوان ورودی مدل تأثیر بسزایی در افزایش دقت برآورد دبی رسوب دارد. در روش رگرسیون ناپارامتری با افزایش تعداد متغیرهای مستقل دقت مدل نیز بهتر شد. همچنین نتایج مرحله‌ی بهینه سازی پارامترهای این روش نشان داد

هرکدام از این معادلات به پارامترهای متفاوتی بستگی دارند، بنابراین تخمین مقدار رسوب در آن‌ها متفاوت است. در این راستا داده‌های روزانه ایستگاه هیدرومتری کهک در طی سال‌های مذکور، خصوصیات هیدرولیکی جریان و رژیم رسوبی آن، مجموعه‌ای از داده‌های صحرائی به عنوان ورودی‌های مدل، جهت دستیابی به بهترین معادله انتقال رسوب در رودخانه سیستان مورد استفاده قرار گرفت.

### معیارهای ارزیابی

به منظور ارزیابی عملکرد مدل‌های استفاده شده چند شاخص ارزیابی به کار گرفته شد. شاخص‌های آماری به کار برده شده جهت بررسی مدل‌ها عبارتند از:

جزر میانگین مربعات خطا (RMSE)<sup>۱</sup>

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (7)$$

ضریب تبیین<sup>۲</sup> (R<sup>2</sup>)

$$R^2 = \frac{\left[ \sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P}) \right]^2}{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2} \quad (8)$$

در این روابط، n تعداد داده‌ها، O<sub>i</sub> مقادیر مشاهداتی و P<sub>i</sub> مقادیر پیش‌بینی شده می‌باشد. حداقل مقدار برای RMSE و حداقل و حداکثر مقدار برای R<sup>2</sup> به ترتیب برابر صفر و یک می‌باشد.

به منظور دست یافتن به بهترین عملکرد روش‌های به کار رفته در این تحقیق هفت ترکیب متفاوت (مدل مختلف) از ورودی‌ها در نظر گرفته شد. برای هر کدام از مدل‌ها و هر کدام از روش‌های تجربی، بردار پشتیبان و رگرسیون ناپارامتری بهترین مدل از هر نوع با توجه به معیارهای ارزیابی ذکر شده و نتایج به دست آمده انتخاب شد. پارامترهای موجود در روش رگرسیون بردار پشتیبان

<sup>2</sup> Root Mean Square Error

<sup>3</sup> Determination coefficient



مدل به دست آمده تعداد همسایه‌ها در روند بهینه سازی برابر با ۳۶ بدست آمد. همچنین مقادیر به دست آمده برای معیارهای  $R^2$  و RMSE برای مدل‌های رگرسیونی در ترکیب‌های ورودی ۵ تا ۷ بیانگر عملکرد بهتر آن‌ها نسبت به روابط تجربی می‌باشد.

که در اکثر مدل‌ها وزن متغیر دبی بیشتر از وزن دما بوده به طوری که در بهترین مدل نیز وزن متغیرهای مستقل دبی و دبی کلاسه به ترتیب  $0/3$  و  $0/7$  و وزن دیگر متغیرها (دمای بیشینه و کمینه) برابر با صفر بود. تعداد نزدیک‌ترین همسایه‌ها نیز در مورد همه‌ی مدل‌های به کار رفته در محدوده‌ی ۳۲ تا ۳۶ قرار داشت و برای بهترین

جدول (۲) مقایسه نتایج به دست آمده از روش رگرسیون بردار پشتیبان (SVR) در ترکیبات مختلف ورودی‌ها

Model	ترکیبات ورودی	RMSE		$R^2$	
		Train	Test	Train	Test
مدل ۱	Q	10933.3	11520.9	0.65	0.57
مدل ۲	Q, $T_{mean}$	10007.9	10063.3	0.71	0.67
مدل ۳	Q, $T_d$	11211.6	10992.4	0.65	0.61
مدل ۴	Q, $T_{min}$ , $T_{max}$	7818.8	9358	0.78	0.70
مدل ۵	Q, $T_{min}$ , $T_{max}$ , $Q_{S(t-1)}$	5730.9	6527.4	0.86	0.86
مدل ۶	Q, $T_{mean}$ , $Q_{S(t-1)}$	9906.7	9206.4	0.70	0.71
مدل ۷	Q, $T_{min}$ , $T_{max}$ , $Q_c$	3575.1	3523.5	0.96	0.96

جدول (۳) مقایسه نتایج به دست آمده از روش رگرسیون نزدیکترین همسایه (K-NN) در ترکیبات مختلف ورودی‌ها

Model	ترکیبات ورودی	RMSE		$R^2$	
		Train	Test	Train	Test
مدل ۱	Q	10957.6	12076.8	0.63	0.56
مدل ۲	Q, $T_{mean}$	10033.3	11668.5	0.69	0.59
مدل ۳	Q, $T_d$	11077.2	11813.8	0.62	0.60
مدل ۴	Q, $T_{min}$ , $T_{max}$	10033.7	11185.2	0.69	0.64
مدل ۵	Q, $T_{min}$ , $T_{max}$ , $Q_{S(t-1)}$	8386.6	9667.9	0.79	0.7
مدل ۶	Q, $T_{mean}$ , $Q_{S(t-1)}$	8525.5	9563.3	0.78	0.72
مدل ۷	Q, $T_{min}$ , $T_{max}$ , $Q_c$	4766.7	4165.5	0.93	0.95

مختلف و مقدار دبی رابطه‌ای رگرسیونی از نوع توانی به دست آمد. سپس رابطه‌ی به دست آمده برای تخمین رسوب از روی مقادیر دبی مشاهداتی به کار گرفته شد. نتایج به دست آمده در جدول شماره ۴ ذکر شده است.

به منظور مقایسه‌ی روش‌های تجربی و رگرسیونی، ابتدا برای هر یک از روابط تجربی موجود در نرم افزار HEC-RAS، رابطه‌ای توانی بین مقدار دبی (Q) بر حسب متر مکعب بر ثانیه و دبی رسوب ( $Q_s$ ) بر حسب تن در روز به دست آمد. برای این منظور بین مقادیر دبی رسوب تخمینی توسط روابط موجود در دبی‌های

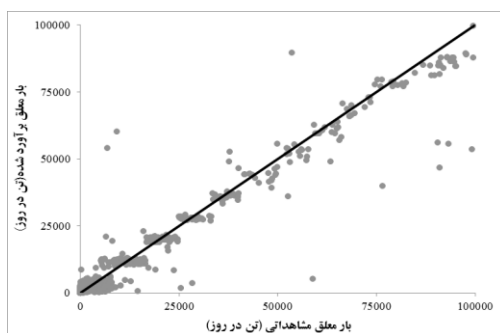


جدول (۴) نتایج بدست آمده از تخمین رسوب معلق با استفاده از روش‌های تجربی

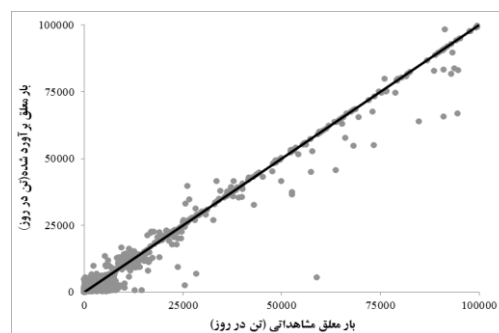
نام روش	رابطه توانی دبی رسوب (ton/day) - دبی (m <sup>3</sup> /s)	RMSE (ton/day)	R <sup>2</sup>
توفالتی	$Q_S = 191.08Q^{1.1888}$	66557.8	0.705
انگلند - هانسن	$Q_S = 432.71Q^{1.0791}$	70623.6	0.706
لارسن - کوپلند	$Q_S = 361.04Q^{1.0272}$	71932.3	0.706
یانگ	$Q_S = 560.08Q^{0.9172}$	81525.8	0.7
مایر - پیتر و مولر	$Q_S = 29.896Q^{0.8593}$	126082.8	0.695
ویل - کاک	$Q_S = 2.6853Q^{0.9778}$	127683.5	0.704
ایگر وایت	$Q_S = 795.48Q^{1.2641}$	816474.4	0.701

برای دو روش SVR و KNN به ترتیب برابر با ۳۵۲۳/۵ و ۴۱۶۵/۵ می‌باشد، که تفاوت زیادی با مقادیر به دست آمده از روش‌های تجربی دارد. همچنین مقادیر بالای ضریب تبیین در نتایج به دست آمده از این روش‌ها بیانگر تطابق زیاد مقادیر برآورد شده و مقادیر مشاهداتی بار رسوبی توسط آن‌ها در مقایسه با روش‌های تجربی می‌باشد. شکل‌های دو تا چهار نشان دهنده‌ی نمودارهای مقایسه‌ای عملکرد هر سه روش ذکر شده می‌باشند.

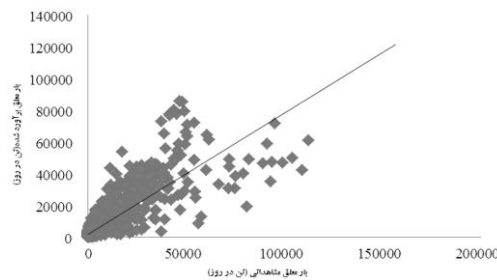
بررسی نتایج بدست آمده مبین عملکرد ضعیف‌تر آن‌ها در مقایسه با روش‌های رگرسیونی است بگونه‌ای که بهترین معادله تجربی رابطه توفالتی با ضریب تبیین ۰/۷۰۵ و مقدار جذر میانگین مربعات خطا ۶۶۵۵۷/۸ تن در روز را نتیجه داده است و این خود بیان کننده همبستگی پایین مقادیر پیش‌بینی شده و مشاهده شده است. نتایج ارائه شده در جدول شماره ۳ و ۲ نشان می‌دهد که مقدار RMSE در مدل شماره ۷



شکل (۳) نمودار نقطه‌ای بین مقادیر مشاهداتی و برآورد شده توسط KNN در مدل ۷



شکل (۲) نمودار نقطه‌ای بین مقادیر مشاهداتی و برآورد شده توسط SVR در مدل ۷



شکل (۴) نمودار نقطه‌ای بین مقادیر مشاهداتی و برآوردی توسط روش توفالتی

### نتیجه‌گیری

بر اساس آنچه که ذکر شد بدلیل کاربرد وسیع مقادیر رسوب جریان، محققین تلاش‌های زیادی را برای برآورد حجم رسوب معلق براساس اطلاعات آزمایشگاهی و میدانی بر مبنای اصول فیزیکی انجام داده‌اند ولی روابط ارائه شده ضمن پیچیدگی از پارامترهای مختلفی که غالباً اندازه‌گیری نمی‌شوند، تاثیر می‌پذیرند. لذا استفاده از مدل‌های داده مینا می‌تواند تا حدود زیادی مشکلات موجود را رفع نماید. در این بین روش‌هایی نظیر رگرسیون بردار پشتیبان و رگرسیون نزدیک‌ترین همسایه به سبب توانایی در مدل‌سازی روابط غیر خطی می‌توانند گزینه‌های مناسبی جهت برآورد رسوب معلق باشند. نتایج به دست آمده از این تحقیق نشان داد که مدل‌های رگرسیونی به کار رفته عملکرد مناسبی در برآورد رسوب رودخانه سیستان دارند. بهترین معادله تجربی بار معلق رسوبی را با جذر میانگین مربعات ۶۶۵۵۸ و ضریب تبیین ۰/۷۰۵ برآورد کرد در حالی که معیارهای ذکر شده در روش بردار پشتیبان به ترتیب ۲۸۰۹/۳ و ۰/۹۶ محاسبه شدند. همچنین روش به کار رفته در توسعه رگرسیون ناپارامتری به منظور بازتولید اعداد خارج از محدوده تاریخی کارایی مناسبی داشته به طوری که در بیشتر مدل‌های شماره یک تا هفت قادر به برآورد مقادیری خارج از محدوده‌ی ذکر شده بود

بررسی کمی نمودارهای فوق نشان می‌دهد که هر دو روش رگرسیونی به کار رفته در این تحقیق توانایی بالایی در برآورد رسوب معلق دارند. همان‌طور که مشاهده می‌شود رسوب برآوردی و مشاهداتی همبستگی بالایی با هم داشته و پراکنش آن‌ها در اطراف خط یک به یک مناسب است. فاصله‌ی کم نقاط از خط یک به یک در شکل‌های شماره ۲ و ۳ نسبت به شکل شماره ۴ نشان دهنده‌ی دقت بهتر روش‌های رگرسیونی نسبت به روش‌های تجربی است. با این وجود همان‌طور که مشاهده می‌شود تراکم نقاط زیر خط یک به یک بیشتر از تراکم آن‌ها بالای خط می‌باشد و این بیانگر این واقعیت است که این مدل‌ها تا حدودی مقدار رسوب معلق را کم‌تر از مقدار واقعی برآورد می‌کنند. مقایسه نتایج به دست آمده از این روش‌ها در مطالعات پیشین (Kakaei lafdani Et al., 2013; Kisi, 2012) نشان می‌دهد که این امر می‌تواند به دلیل بالا بودن حجم رسوب معلق در رودخانه‌ی سیستان نسبت به رودخانه‌های دیگر باشد. یکی دیگر از توانایی هر دو روش ذکر شده پیش‌بینی یا تخمین مقادیر خارج از محدوده‌ی مشاهداتی می‌باشد که نزدیک به مقادیر مشاهداتی متناظر هستند.



## منابع

- جودی، ع. و م. ت. ستاری. ۱۳۹۶. مقایسه کارایی روش‌های رگرسیون بردار پشتیبان و نزدیکترین همسایگی در برآورد بار رسوبی معلق در رودخانه، مطالعه موردی: رودخانه ليقوان چای. مجله منابع طبیعی ایران، دوره ۷۰، شماره ۲، ص ۳۴۵-۳۵۸.
- حسن‌پور، ف. و س. م. طباطبایی. ۱۳۸۸. لایروبی موضعی روشی کم‌هزینه برای حفظ حیات رودخانه سیستان. اولین همایش ملی اصلاح الگوی مصرف با محوریت منابع طبیعی، کشاورزی و دامپزشکی. زابل، دانشگاه زابل.
- دستورانی، م. ت. خ. عظیمی فشمی، ع. طالبی و م. ر. اختصاصی. ۱۳۹۱. برآورد رسوبات معلق با استفاده از شبکه عصبی مصنوعی، مطالعه موردی: حوزه آبخیز چامیشان استان کرمان. پژوهشنامه مدیریت حوزه آبخیز، دوره ۳، شماره ۶، ص ۶۱-۷۴.
- دهقانی، ا. م. ا. زنگانه، ا. مساعدی و ن. کوهستانی. ۱۳۸۸. مقایسه تخمین بار معلق به دو روش منحنی سنج رسوب و شبکه عصبی مصنوعی، مطالعه موردی: رودخانه دوغ استان گلستان. نشریه علوم کشاورزی و منابع طبیعی، دوره ۱۶، شماره ۱، ص ۲۷۶-۲۶۶.
- دهقانی، ا. م. قبائی سوق، ح. شریفان و ن. دهقانی. ۱۳۹۰. واسنجی معادلات تجربی تبخیر و تعرق روزانه و مقایسه با نتایج شبکه‌های عصبی مصنوعی در شرایط مختلف کمبود داده‌های هواشناسی. مجله مدیریت آب و آبیاری، دوره ۱، شماره ۱، ص ۴۱-۵۴.
- دهقانی، ن. و م. وفاخواه. ۱۳۹۲. مقایسه روش‌های تخمین رسوب معلق روزانه با استفاده از روش‌های منحنی سنج رسوب و شبکه عصبی، مطالعه موردی: ایستگاه قزاقلی، استان گلستان. مجله پژوهش‌های حفاظت آب و خاک، دوره ۲۰، شماره ۲، ص ۲۳۰-۲۲۱.
- رجایی، ط. و ه. ابراهیمی. ۱۳۹۳. مدل‌سازی نوسان‌های ماهانه آب زیرزمینی به وسیله تبدیل موجک و شبکه عصبی پویا. مجله مدیریت آب و آبیاری، دوره ۴، شماره ۱، ص ۷۳-۸۷.
- ستاری، م. ت. ع. ر. جودی، ف. صفدری، و ف. قهرمان زاده. ۱۳۹۵. ارزیابی عملکرد روش‌های مدل درختی و رگرسیون بردار پشتیبان در مدل‌سازی رسوب معلق رودخانه. نشریه حفاظت منابع آب و خاک، دوره ششم، شماره اول، ص ۱۰۹-۱۲۳.
- شریف آذری، س. و ش. عراقی‌نژاد. ۱۳۹۲. توسعه مدل ناپارامتری شبیه ساز داده‌های ماهانه هیدرولوژیکی. مجله مدیریت آب و آبیاری، دوره ۳، شماره ۱، ص ۸۳-۹۵.
- فراهی، گ. س. ر. خدائشناس و ا. علی‌زاده. ۱۳۹۰. برآورد رسوب حوزه‌های آبخیز شمال استان خراسان با استفاده از مدل رگرسیون فازی. مجله علوم و مهندسی آبخیزداری، سال پنجم، شماره ۱۵، ص ۱۱-۲۴.
- نیک‌بخت شهبازی، ع. ب. زهرایی و م. ناصری. ۱۳۹۱. پیش‌بینی فصلی خشکسالی هواشناسی با استفاده از ماشین‌های بردار پشتیبان. مجله آب و فاضلاب، دوره ۲۳، شماره ۲، ص ۷۳-۸۵.

Alp, M. and H.K. Cigizoglu. 2007. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environmental Modelling & Software*, 22(1): 2-13.

Asefa, T., M. Kemblowski, M. Mckee and A. Khalil. 2006. Multi time scale stream flow predictions: the support vector machines approach. *Hydrology*, 318(1-4): 7-16.



- Bharti, B., A. Pandey, S. K. Tripathi and D. Kumar. 2017. Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. *Hydrologie research*, 211(2-8):1489-1506.
- Cimen M. 2008. Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal*, 53 (3):656–666.
- Cristianini, N. and T. Shawe. 2000. *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge, 189 p.
- Hazarika, B.B., D. Gupta and M. Belin. 2020. Modeling suspended sediment load in river using extreme learning machine and twin support vector regression with wavelet conjunction. *Environmental Earth Sciences*, 79(2):234-248.
- Joudi, A. R. and M.T. Sattari. 2017. Comparison of the Efficiency of Support Vector Regression and K-Nearest Neighbor Methods in suspended sediment load Estimation in river (Case Study: Lighvan Chay River). *Journal of RANNGE AND Watershed MANAGEMENT*, 122-138p.
- Kakaei lafdani, E., A. Moghaddam nia and A. Ahmadi, 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Hydrology Research* 478: 50–62.
- Kisi, O. 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. *Hydrology*, 456-457: 110-120.
- Kisi, O. 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrological Sciences Journal*, 50(4): 683-696.
- Kisi, O., M.E. Karahan and Sen, Z. 2006. River suspended sediment modeling using fuzzy logic approach. *Hydrological Processes*, 20(20): 4351-4362.
- Kisi, O.T. Haktanir, M. Ardiclioglu, O. Ozturk, E. Yalcin and S. Uludag. 2009. Adaptive neuro-fuzzy computing technique for suspended sediment estimation. *Advances in Engineering Software*, 40(6): 438-444.
- Lall, U. and A. Sharma. 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3): 679-694.
- Prairie, J. R., B. Rajagopalan, T.J. Fulp and E.A. Zagona. 2006. Modified K-NN model for stochastic streamflow simulation. *Hydrologic Engineering*, 11(4):371-378.
- Roshangar, K. and A. Koshi. 2015. Evaluation of GA-SVR method for modeling bed load transport in gravel-bed Rivers. *Journal of Hydrology*, 527:1142-4452.
- Sheikhalipour, Z. and F. Hassanpour. 2013. Estimation of Suspended Sediment Load Using Genetic Expression Programming. *Civil Engineering and Urbanism*, 3(5):292-299.
- Shoushtari, Sh. and Kashefipour, M. 2007. Estimating of suspended sediment load using artificial neural networks case study Ahvaz station. In: *Proceeding of 7<sup>th</sup> International Conference on River Engineering*, Shahid Chamran University, Ahwaz, Iran, 343-356.
- Torabi, H. and R. Dehghani, 2018. Comparison and evaluation of intelligent models for river suspended sediment estimation (case study: Kakareza River, Iran). *Environmental Resources Research*, 852:139-15
- Vapnik, V. N. 1995. *The nature of statistical learning theory*. Springer-Verlag, New York, 314p.
- Vinod, H. D. 2014. Matrix Algebra Topics in Statistics and Economics Using R, in *Handbook of Statistics: Computational Statistics with R*, Vol. 34, eds. Rao, M. B. and Rao, C. R., New York: North Holland, Elsevier Science, pp. 143-176.



## Assessment of Modern Regression Methods to Suspended Sediment Load Estimation in the Sistan River

Sara Mohammadi<sup>1</sup>, Farzad Hassanpour<sup>2</sup>, Salman Sharifazari<sup>3</sup>, Farid Foroughi<sup>4</sup>

### Abstract:

Correct estimation of the river sediment volume is important for many water resources projects. In this study, the empirical equations of sediment transport, support vector regression (SVR), and developed k-nearest neighbor regression (KNN) were used in order to estimate suspended sediment load in the *Sistan* River. In this regard, in addition to the maximum temperature, minimum temperature, and discharge and Suspended Load in the period 1374 to 1390, it was paired on 1682, discharge, the classified discharge was used as effecting variable to suspended sediment load modeling. For each of the input combinations in support vector regression and development k-nearest neighbor regression the best structure of the regression model determined based on the performance criteria. The Result showed that the *Toffaleti* method is the best method between empirical equations with  $R^2$  equal to 0.705 and RMSE equal to 66558 Ton/day. Also, the support vector regression model with discharge, minimum temperature, maximum temperature and classified discharge as the set of input data is the best model in estimation of suspended sediment load with  $R^2$  equal to 0.96 and RMSE equal to 2809.3 Ton/day. The results indicate that the regression method estimate suspended sediment load much better than empirical equations in the *Sistan* River.

**Key Words:** Empirical Models, K-Nearest Neighbor Regression, Sediment Transport, Support Vector Regression, Toffalti Method.

---

<sup>1</sup> M. Sc. of Hydro Structures Engineering, University of Zabol; saramohamadi494@yahoo.com

<sup>2</sup> Associate Professor, University of Zabol; hassanpourir@uoz.ac.ir

<sup>3</sup> Instructor, University of Zabol; s.sharifazari@gmail.com

<sup>4</sup> Instructor, Darab School of Agriculture, Shiraz University; frooghi14@yahoo.com

## Extended Abstract

## Research Paper

## Assessment of Modern Regression Methods to Suspended Sediment Load Estimation in the Sistan River

Sara Mohammadi<sup>1</sup> Farzad Hassan pour<sup>2</sup>, Salman Sharifazari<sup>3</sup> Farid Foroughi<sup>4</sup>

<sup>1</sup> PhD Student in Civil Engineering-Hydraulic Structures, Department of Civil Engineering, Bahonar University of Kerman, Kerman, Iran; saramohamadi494@yahoo.com

<sup>2</sup> Associate Professor of Department of Water Engineering, Faculty of Water and Soil Sciences, University of Zabol Iran; hassanpourir@uoz.ac.ir (Corresponding Author)

<sup>3</sup> Instructor, University of Zabol; Iran; s.sharifazari@gmail.com,

<sup>4</sup> Darab, University of Agriculture, Shiraz University; Iran; frooghi14@yahoo.com



10.22125/IWE.2022.323757.1590

Received:

30. January.2020

Accepted:

28. September.2020

Available online:

22. January.2022

**Keywords:**

**Empirical Models, K-Nearest Neighbor Regression, Sediment Transport, Support Vector Regression, Toffalti Method**

### Abstract

Correct estimation of the river sediment volume is important for many water resources projects.

In this study, the empirical equations of sediment transport, support vector regression (SVR), and developed k-nearest neighbor regression (KNN) were used in order to estimate suspended sediment load in the Sistan River. In this regard, in addition to the maximum temperature, minimum temperature, and discharge and Suspended Load In the period 1374 to 1390, it was paired on 1682, discharge, the classified discharge was used as effecting variable to suspended sediment load modeling. For each of the input combinations in support vector regression and development k-nearest neighbor regression the best structure of the regression model determined based on the performance criteria. The Result showed that the Toffaleti method is the best method between empirical equations with R2 equal to 0.705 and RMSE equal to 66558 Ton/day. Also, the support vector regression model with discharge, minimum temperature, maximum temperature and classified discharge as the set of input data is the best model in estimation of suspended sediment load with R2 equal to 0.96 and RMSE equal to 2809.3 Ton/day. The results indicate that the regression method estimate suspended sediment load much better than empirical equations in the Sistan River.

### 1. Introduction

Researchers have made great efforts to accurately estimate the volume of suspended sediment, and based on laboratory and field information, generally based on physical principles, have developed relationships that have led to the advancement of this science, but in most existing methods. In addition to the existence of various parameters that are not measured in most rivers, it is also necessary to solve complex mathematical equations. On the other hand, due to the lack of complete and accurate information on the parameters affecting the sedimentation process and structure. Completely non-linear, a comprehensive model cannot be introduced to it model (Sheikh Alipour, 2013). Therefore, due to the nonlinearity of the sediment relationship and other it parameters affecting, the use of models that can simulate nonlinear processes with high reliability is a reliable solution in sediment estimation. Examination of study records in this field indicates the fact that the support vector method is trained much faster than conventional artificial neural networks. Therefore, due to the time-consuming training of models such as artificial neural networks and fuzzy inference system, the use of other data-based models such as modern regression methods can be considered as alternative solutions (Cimen, 2008;

Cristianini and Shawe, 2000; Jie and Yu, 2011; Torabi and deghani, 2018). The study area is the Sistan plain, which is irregularly quadrangular in shape and one of its vertices is formed by the collision of two branches of the Helmand River. One branch is the Sistan River, which flows west into the Sistan Plain, and the other is called Paryan, which flows along the southeastern side of the Sistan Plain along the border between Iran and Afghanistan. The eastern side of Sistan is approximately 60 km long and its southern side is 50 km long. Its western and northern sides are limited to Helmand and Puzak hamuns, respectively.

## 2. Materials and Methods

Sistan river is prone to sedimentation of fine-grained sediments due to its location in the end of Helmand large catchment area and the annual accumulation of large amounts of sediments along the river route increases the bed level and thus reduces the flow capacity so that even in Dubai Medium and low water escape from the river bank to the adjacent plains can be seen. This river is one of the two main tributaries of the Helmand River, located in the southeast of the country and in the Sistan plain, which is responsible for irrigating 70% of the agricultural lands of the Sistan plain (Hassanpour, 2000). The low slope of the Sistan River bed makes it prone to sedimentation and on the other hand, the negative effects of the construction of drainage and galaxy dams have encouraged this sedimentation and increased the amount of sedimentation (Hassan Pouro Tabatabai, 2009).

In this study, the empirical equations of sediment transport, support vector regression (SVR), and developed k-nearest neighbor regression (KNN) were used in order to estimate suspended sediment load in the Sistan River. From the recorded data of flow rate (Q), sediment flow rate (QS), maximum temperature (Tmax) and minimum temperature (Tmin) of the river catchment in the period 1374 to 1390, were collected daily as input to the models in question. Used. In addition to the mentioned variables, in order to increase the accuracy of estimating the daily sediment of the river class (QC), the average daily temperature (Tmean) and the difference between the maximum and minimum daily temperature (Td) were also considered as other input variables. After standardization and deletion of outliers, the number of recorded and reliable data was equal to 1682 data pairs, of which 70% was used to teach the mentioned models and the remaining 30% was used to test them,,

## 3. Discussion and Conclusion

For each of the input combinations in support vector regression and development k-nearest neighbor regression the best structure of the regression model determined based on the performance criteria. The Result showed that the Toffaleti method is the best method between empirical equations with R2 equal to 0.705 and RMSE equal to 66558 Ton/day. Also, the support vector regression model with discharge, minimum temperature, maximum temperature and classified discharge as the set of input data is the best model in estimation of suspended sediment load with R2 equal to 0.96 and RMSE equal to 2809.3 Ton/day. The results indicate that the regression method estimate suspended sediment load much better than empirical equations in the Sistan River.

## 4. Six important references

1. Hazarika, B.B., D. Gupta and M. Belin. 2020. Modeling suspended sediment load in a river using extreme learning machine and twin support vector regression with wavelet conjunction. *Environmental Earth Sciences*, 79(2):234-248.
2. Joudi, A. R. and M.T. Sattari. 2017. Comparison of the Efficiency of Support Vector Regression and K-Nearest Neighbor Methods in suspended sediment load Estimation in river (Case Study: Lighvan Chay River). *Journal of RANNGE AND Watershed MANAGEMENT*, 122-138p.
3. Kakaei lafdani, E., A. Moghaddam nia and A. Ahmadi, 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Hydrology Research* 478: 50–62.
4. Kisi, O. 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. *Hydrology*, 456-457: 110-120.
5. Sheikhalipour, Z. and F. Hassanpour. 2013. Estimation of Suspended Sediment Load Using Genetic Expression Programming. *Civil Engineering and Urbanism*, 3(5):292-299.

## Conflict of Interest

Authors declared no conflict of interest.