

Uncertainty Analysis of SVM Model Parameters for Estimating Suspended and Bed Sediment Load at Sierra Station in Karaj by Monte-Carlo Simulation Method

ALIREZA KEIHANI¹, ALI MOHAMMAD AKHONDALI^{1*}, HOSSEIN FATHIAN²

1. Department of Hydrology and Water Resources, Faculty of Water Engineering and Environment, Shahid Chamran University, Ahvaz, Iran.

2. Department of Water Resources Engineering, Islamic Azad University, Ahvaz Branch, Ahvaz, Iran.

(Received: Aug. 15, 2020- Revised: Nov. 22, 2020- Accepted: Nov. 30, 2020)

ABSTRACT

Estimation of sediment transported by the streamflow is important for planning and storing water resources of dam reservoirs and river bed changes, watershed management, coastal protection and the environment. Sediment transport in the river is an inherently uncertain and complex phenomenon. Incomplete knowledge of processes and data create uncertainty in estimating sediment transport. Parameters uncertainty is one of the main sources of uncertainty in estimating the suspended and bed sediment load. In this paper, the Monte Carlo (MC) simulation method is used to estimate the uncertainty of suspended and bed sediment load due to uncertainty in the parameters of the support vector machine (SVM) model in the Karaj Dam Basin. The partial mutual information (PMI) algorithm was used to select the efficient input variables in the SVM model to estimate the suspended and bed sediment load. The results of using PMI algorithm show that the only efficient variable in estimating the suspended and bed sediment loads is the current stream discharge. The results show that the uncertainty in estimating the suspended sediment load with SVM model for training, test and total data is equal to 12.8%, 17% and 13.5%, respectively. Also, the uncertainty in estimating the bed sediment load with SVM model for training, test and total data is equal to 23.5%, 36.8% and 27.2%, respectively. Therefore, the uncertainty in estimating the bed sediment load with SVM model is more than the one in estimating the suspended sediment load. Therefore, the use of optimization methods can be useful for accurate estimation of parameter values and reducing uncertainty in estimating the suspended and bed sediment load.

Keywords: Parameter Uncertainty, SVM Model, Suspended and Bed Sediment Load, PMI Algorithm, Monte-Carlo.

* Corresponding Author's Email: aliakh@scu.ac.ir

تحلیل عدم قطعیت پارامترهای مدل SVM برای برآورد بار رسوبات معلق و بستر در ایستگاه سیرا کرج با روش شبیه سازی مونت کارلو

علیرضا کیهانی^۱، علی محمد آخوندعلی^{۲*}، حسین فتحیان^۳

۱. گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران، اهواز، ایران.

۲. گروه مهندسی منابع آب، دانشگاه آزاد اسلامی، واحد اهواز، اهواز، ایران.

(تاریخ دریافت: ۱۳۹۹/۵/۲۵ - تاریخ بازنگری: ۱۳۹۹/۹/۲ - تاریخ تصویب: ۱۳۹۹/۹/۱۰)

چکیده

برآورد میزان رسوب حمل شده توسط جریان برای برنامه ریزی و ذخیره منابع آب مخازن سدها و تغییرات بستر رودخانه‌ها، مدیریت آبخیز، حفاظت سواحل و محیط زیست حائز اهمیت است. انتقال رسوب در رودخانه یک پدیده ذاتا غیرقطعی و پیچیده می‌باشد. دانش ناکامل در مورد فرآیندها و داده‌ها، عدم قطعیت در برآورد انتقال رسوب را ایجاد می‌کند. عدم قطعیت پارامترها، از جمله منابع اصلی عدم قطعیت در برآورد بار رسوبات معلق و بستر است. در این مقاله از روش شبیه سازی مونت کارلو برای برآورد عدم قطعیت بار رسوبات معلق و بستر به علت عدم قطعیت در پارامترهای مدل ماشین بردار پشتیبان (SVM) در حوضه سد کرج استفاده شده است. برای انتخاب متغیرهای ورودی موثر در مدل SVM برای برآورد بار رسوبات معلق و بستر، از الگوریتم PMI استفاده شد. نتایج به کارگیری الگوریتم PMI نشان می‌دهد که تنها متغیر موثر در برآورد بار رسوبات معلق و بستر، دبی جریان در زمان حال است. نتایج نشان می‌دهد که عدم قطعیت در برآورد بار رسوب معلق با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها به ترتیب برابر با ۱۲/۸، ۱۷ و ۱۳/۵ درصد است. همچنین عدم قطعیت در برآورد بار رسوب بستر با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها به ترتیب برابر با ۲۳/۵، ۳۶/۸ و ۲۷/۲ درصد است. بنابراین عدم قطعیت در برآورد بار رسوب بستر با مدل SVM بیشتر از عدم قطعیت در برآورد بار رسوب معلق است. به کارگیری روش‌های بهینه سازی می‌تواند برای برآورد دقیق مقادیر پارامترها و کاهش عدم قطعیت در برآورد بار رسوبات معلق و بستر مفید باشد.

واژه‌های کلیدی: عدم قطعیت پارامترها، مدل SVM، بار رسوب معلق و بستر، الگوریتم PMI، مونت-کارلو.

مقدمه

رسوبات حمل شده از سطح حوضه‌ها توسط رواناب‌ها وارد سرشاخه‌ها و آبراهه‌های اصلی شده و در نهایت در مخازن سدها ته نشین گردیده و باعث کاهش عمر مفید آنها می‌شود. برآورد میزان رسوب حمل شده توسط جریان می‌تواند شاخصی برای نشان دادن فرسایش خاک حوضه باشد و در پروژه‌های آبی جهت برنامه ریزی و ذخیره منابع آب مخازن سدها و تغییرات بستر رودخانه‌ها، مدیریت آبخیز، حفاظت سواحل و محیط زیست و غیره حائز اهمیت است. رابطه رواناب و بار رسوبی با توسعه، بهره برداری و مدیریت منابع آب، بهره برداری ایمن از پروژه‌های مربوط با آب و محیط زیست مرتبط است (Miao et al., 2011; Peng et al., 2010). به علت تغییرات محیط ناشی از تغییر اقلیم و فعالیت های انسانی مشخصات آماری سری‌های هیدرولوژیکی در حوضه‌های مختلف ممکن است تغییر کند که تحت عنوان غیرایستایی^۱

نامیده می‌شود (Jiang et al., 2015; Gilroy and McCuen, 2012; Vogel et al., 2011)

رسوبات معلق متغیر مهم محیط زیستی و هیدرولوژیکی می‌باشد که به فرسایش خاک، شرایط اکولوژیکی حوضه، شرایط آبراهه‌ها، اقدامات هیدرولیکی و فراوانی وقایع حدی باران وابسته است (Bonacci and Oskorus, 2010). این واقعیت به خوبی شناخته شده است که اکثر بار رسوب معلق در چند واقعه بارندگی شدید منتقل می‌شوند (Rodríguez-Blanco, 2010; Tena et al., 2011). بنابراین منطقی است که وقایع غلظت رسوب معلق متناظر (تعریف شده بر مبنای دبی پیک) در تحلیل فراوانی سیل در نظر گرفته شود. درک تغییرات درازمدت دبی رسوب معلق و عوامل موثر بر آن در حوضه‌های آبخیز برای مطالعه تخریب خاک، کیفیت آب، دوره‌های بیوشیمیایی زمین و پایداری کشاورزی بسیار مهم است (Walling, 2006; Syvitski, and Milliman; 2007, Walling, 2009; Pelletier, 2012; Vanmaercke et al.,

(2014).

فزاینده‌ای برای ارزیابی پدیده‌های هیدرولوژیکی و هیدرولیکی استفاده شده‌اند (Adnan et al., 2019). Zounemat-Kermani (2016) et al. توانایی مدل‌های ANN و SVM برای پیش‌بینی و برآورد غلظت‌های رسوب معلق روزانه با استفاده از داده‌های هشت ساله از یک ایستگاه در ایالات متحده آمریکا را ارزیابی کردند. عملکرد مدل‌های ANN بر اساس ریشه میانگین مربعات خطا، ضریب همبستگی پیرسون و معیار کارایی ناش-ساتکلیف ارزیابی شدند. مدل‌های ANN و SVM نتایج بهتری نسبت به مدل‌های رگرسیون خط چندگانه و منحنی سنجه رسوب نشان دادند. (Wieprecht et al., 2013) بار بستر و بار مواد بستر کل در رودخانه رییم^۷ با سیستم استنتاج تطبیقی عصبی-فازی^۸ (ANFIS) پیش-بینی کردند. پارامترهای مورد استفاده برای توسعه مدل عمدتاً پارامترهایی بودند که بر انتقال بار رسوب در مقیاس بزرگ تأثیرگذار هستند. نتایج نشان داد که رویکرد مدل‌سازی عصبی-فازی تطبیقی مبتنی بر داده‌ها می‌تواند یک روش جایگزین قدرتمند برای تخمین بار بستر و بار مواد جامد معلق حوضه سد علویان (et al., 2018) در استان آذربایجان شرقی با مدل SWAT^۹ پیش‌بینی کردند. جهت واسنجی و تحلیل عدم قطعیت مدل SWAT از دو روش SUFI-2^{۱۰} و GLUE^{۱۱} استفاده کردند. نتایج حاصل از کاربرد دو روش مذکور نشان داد روش SUFI-2 دارای الگوریتم موثرتری جهت واسنجی و تعیین عدم قطعیت مدل‌ها می‌باشد. بر این اساس با اطمینان بیشتری می‌توان از مدل SWAT واسنجی شده با الگوریتم SUFI-2 در مدیریت منابع آب، کمی‌سازی سناریوهای تغییر اقلیم و تغییر کاربری و تخمین تأثیر اجرای بهترین راهکارهای مدیریتی در داخل حوضه استفاده کرد. (Dams et al., 2015) نیز اثر سناریوی انتخاب شده برای تغییر اقلیم، ساختار مدل بارش-رواناب و تابع هدف مورد استفاده در واسنجی پارامترهای مدل را بر عدم قطعیت دبی‌های ماهانه پیش‌بینی شده در حوضه‌ای در شمال بلژیک بررسی کردند. نتایج نشان داد نوع مدل بارش-رواناب سهم نسبتاً قابل توجهی از عدم قطعیت پیش‌بینی دبی میانگین و بیشینه را به خود اختصاص می‌دهد و این عدم قطعیت، قابل مقایسه با عدم قطعیت حاصل از تغییر اقلیم است. (Lee & Kang, 2016) از ANN برای پیش‌بینی دبی روزانه جریان در حوضه‌ای در کره جنوبی استفاده کردند و با استفاده از روش دامنه پیش‌بینی، عدم قطعیت خروجی مدل را با

در دهه‌های اخیر بارهای رسوبی در بسیاری از رودخانه‌ها در سرتاسر جهان، تغییرات قابل توجهی بعلت تغییرات اقلیمی و فعالیت‌های انسانی تجربه کرده‌اند. انتقال رسوب در رودخانه یک پدیده ذاتاً غیرقطعی و پیچیده می‌باشد. دانش ناکامل در مورد فرآیندها و داده‌ها، عدم قطعیت در برآورد انتقال رسوب را ایجاد می‌کند. منابع عدم قطعیت عواملی هستند که منجر به افزایش عدم قطعیت در برآورد و پیش‌بینی یک سیستم مدل‌سازی شده می‌شوند. تحلیل عدم قطعیت تلاش می‌کند یک ساختار اصولی برای شناسایی عدم قطعیت مربوط به خروجی مدل فراهم کند. خطاهای تقریب‌های به کار رفته در اندازه‌گیری داده‌های ورودی، مقادیر پارامترها، ساختار مدل و الگوریتم روابط مابین پارامترها و متغیرها در مدل، همگی منابع عدم قطعیت هستند (Liu and Gupta, 2007).

Shen et al., 2012 روش برآورد عدم قطعیت درست‌نمایی عمومی (GLUE) را مدل SWAT^۱ ترکیب کردند تا عدم قطعیت پارامترها مربوط به شبیه‌سازی جریان آبراهه و رسوب در حوضه رودخانه دانیگ^۲ در TGRA^۳ در چین به کمیت در بیارند. نتایج نشان داد که شبیه‌سازی رسوب عدم قطعیت بیشتری نسبت به جریان آبراهه دارد و عدم قطعیت در شرایط بارندگی زیاد (از ماه مه تا سپتامبر) حتی بیشتر از فصل خشک است. منابع اصلی عدم قطعیت جریان آبراهه از فرآیندهای در سطح حوضه آبریز حاصل می‌شود در حالی که فرآیندهای در کانال تأثیر زیادی بر شبیه‌سازی رسوب دارد. (Rymaszewicz et al., 2018) برای شبیه‌سازی تغییرات مکانی و زمانی بار رسوبات معلق سالانه در حوضه‌های کوچک کشاورزی، از مدل‌های مبتنی بر رگرسیون متغیرهای هیدرولوژیکی استفاده کردند. تغییرات مکانی عواملی مانند خصوصیات خاک (درصد خاک‌های زهکشی یافته بصورت ضعیف و درصد خاک‌های توری)، کاربری اراضی (درصد علفزار و درصد زمین‌های قابل کشت)، شیب کانال و خصوصیات شبکه زهکشی (تراکم زهکشی) به‌عنوان پارامترهای مدل‌ها در نظر گرفته شدند. روش معرفی شده، بیانگر پتانسیل قابل قبول مدل‌های مبتنی بر رگرسیون جهت شبیه‌سازی بار رسوبات معلق در حوضه‌های کوچک می‌باشد.

الگوریتم‌های شبکه‌های عصبی مصنوعی^۴ (ANN)، ماشین بردار پشتیبان^۵ (SVM) و ماشین یادگیری شدید^۶ (ELM) به‌طور

7 Rhime

8 Adaptive Neuro-Fuzzy Inference System

9 Soil and Water Assessment Tool

10 Sequential Uncertainty Fitting Algorithm

11 Generalized Likelihood Uncertainty Estimation

1 Soil and Water Assessment Tool

2 Daning

3 Three Gorges Reservoir Region

4 Artificial Neural Networks

5 Support Vector Machin

6 Extreme Learning Machine

تهران، با استفاده از کمیت‌های PPU^{۹۵}، p-factor و d-factor مورد بررسی قرار دادند. نتایج نشان داد که SVM با p-factor نهایی معادل با ۸۲ درصد در مرحلهٔ آزمون، قابل‌اعتمادترین مدل برای پیش‌بینی دبی ماهانه در حوضهٔ مورد بررسی است. Fathian et al., (2020) عدم‌قطعیت پارامترهای مدل بارش-رواناب HEC-HMS در برآورد حداکثر سیلاب محتمل^۸ (PMF) در حوضه سد بختیاری با روش مونت کارلو تحلیل کردند. نتایج نشان داد که عدم‌قطعیت در دبی اوج و حجم هیدروگراف PMF بعلت عدم قطعیت تمام پارامترها به ترتیب برابر با ۱۷/۱۳ و ۶/۷۹ درصد است. Nourani et al., (2019) یک روش موجک بر اساس داده کاوی به نام wavelet-M5 (WM5) برای پیش‌بینی بار رسوب معلق معرفی کردند. آنها با مقایسه این مدل با مدل‌های موجک wavelet-ANN (WANN) و درخت M5، دریافتند که مدل WM5 پیش‌بینی بهتری دارد.

Sharafati et al., (2020) مدل‌های نوین ماشین یادگیری گروهی^۹ (EML) توسعه‌یافته را برای پیش‌بینی بار معلق رسوب روزانه در رودخانه می‌سی‌سی‌پی بکار بردند. این مدل‌ها شامل رگرسیون افزایش گرادیان^{۱۰} (GBR)، رگرسیون AdaBoost (ABR) و رگرسیون جنگل تصادفی (RFR) بودند. بر اساس معیارهای عملکرد و مشاهداتی، مدل RFR اندکی برتری در پیش‌بینی بار معلق رسوب روزانه نسبت به دو مدل دیگر نشان داد. تحلیل عدم‌قطعیت نشان داد که ترکیب متغیرهای ورودی تأثیر بیشتری بر پیش‌بینی به‌دست‌آمده نسبت به انتخاب ساختار مدل نشان می‌دهند.

اخیرا بر اساس روش آماری، الگوریتم‌هایی جهت انتخاب متغیرهای ورودی موثر در مدل‌های بر مبنای پردازش داده‌ها ارائه شده‌اند. از جمله الگوریتم غیرخطی جهت تعیین متغیرهای ورودی موثر در مدل‌های بر مبنای پردازش داده‌ها، الگوریتم اطلاعات متقابل جزئی^{۱۱} (PMI) می‌باشد. الگوریتم PMI از شدت روابط بین ورودی‌ها و خروجی‌ها و بین متغیرهای اضافی جهت انتخاب متغیرهای ورودی موثر استفاده می‌کند. در مطالعات قبلی از الگوریتم PMI برای تعیین متغیرهای موثر ورودی برای پیش‌بینی دو هفته آینده شوری رودخانه (Bowden et al., 2005) ، پیش‌بینی کیفیت آب در سیستم توزیع آب (May et al., 2008)، پیش‌بینی تبخیر و تعرق مرجع (Fang et al., 2018)، مدل‌سازی

تغییر دادن وزن‌های اولیه و تعداد عناصر پردازشگر در لایهٔ میانی و همچنین تغییر دادن داده‌های مورد استفاده در آموزش برآورد کردند. در این تحقیق از چهار نوع تابع انتقال برای ANN استفاده شد و نتایج نشان داد مدلی که از تابع انتقال Logistic استفاده می‌کند دارای کمترین عدم‌قطعیت است. در این مدل، درصدی از داده‌های مشاهده‌شده که در دامنهٔ پیش‌بینی قرار داشتند برابر با ۹۹ درصد و میانگین عرض دامنه برابر با ۱/۱۴ مترمکعب در ثانیه به‌دست آمد. Isazadeh et al. (2017) عدم‌قطعیت دو مدل ANN و SVM را در برآورد پارامترهای کیفی آبخوان دشت گیلان بررسی کردند. آنها با تغییر در پارامترهای دو مدل، به ازای هر مقدار مشاهده‌شده از پارامترهای کیفی آبخوان، دامنه‌ای از برآوردها به‌دست آوردند و عدم‌قطعیت هر مدل با محاسبهٔ معیارهای p-factor (درصدی از داده‌های مشاهده‌شده که در دامنهٔ برآوردها قرار می‌گیرند) و d-factor (میانگین عرض باند برآوردها) ارزیابی کردند. نتایج نشان داد که عدم‌قطعیت SVM نسبت به ANN کمتر می‌باشد به‌طوری که به ازای p-factor یکسان (۹۹ درصد)، SVM دارای d-factor کمتری است.

Karami cheme and Mazaheri (2018) یک روش جدید برای تعیین موقعیت‌هایی است که ضریب پراکندگی تأثیر معنی‌داری بر حمل و نقل مواد آلاینده رودخانه دارد، ارائه کردند که مبتنی بر روش شبیه‌سازی مونت کارلو است. نتایج نشان داد که الگوی زمانی منبع آلودگی، عامل اصلی در تأثیر ضریب پراکندگی در حمل و نقل مواد آلاینده است. Fathabadi et al. (2018) روش‌های ناپارامتریک LEC^۱، رگرسیون چندک^۲ و جنگل تصادفی^۳ در برآورد عدم‌قطعیت مدل یکپارچه HBV^۴ در حوضه چهل‌چای استان گلستان بررسی کردند. نتایج برآورد عدم‌قطعیت نشان داد بزرگترین و کوچکترین مقدار دامنه عدم‌قطعیت به ترتیب توسط روش‌های LEC در حالتی که توسط ماشین بردار رگرسیون آموزش دیده باشد و روش جنگل تصادفی، بدست می‌آید.

Salehpoor et al., (2019) عدم‌قطعیت خروجی پنج مدل مبتنی بر هوش مصنوعی شامل مدل‌هایی از نوع ماژولار^۵ (MNN)، تحلیل مولفه‌های اصلی^۶ (PCA)، شبکه عصبی مصنوعی برگشتی با تأخیر زمانی^۷ (TLRN)، ANFIS و SVM در پیش‌بینی دبی ماهانهٔ حوضهٔ آبریز حبله‌رود در شرق استان

8 Probable Maximum Flood
9 Ensemble Machine Learning
10 Gradient boost regression
11 Partial Mutual Information

1 Local Errors and Clustering
2 Quantile Regression
3 Random Forest
4 Hydrologiska Byråns Vattenbalansavdelning
5 Modular Neural Network
6 Principal Component Analysis
7 Time Lag Recurrent Network

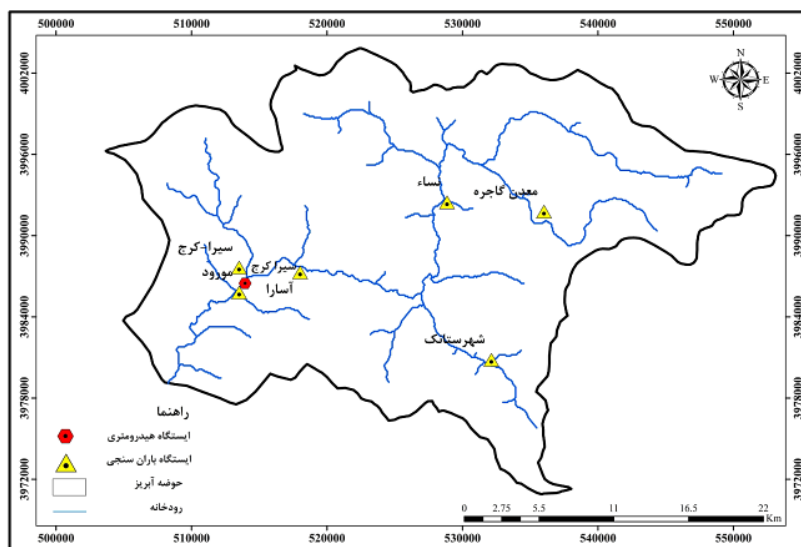
دارد که مساحت آن ۷۱۶ کیلومتر مربع در محل ایستگاه هیدرومتری سیرا کرج است. اقلیم منطقه بر اساس روش دومارتن از نوع مرطوب و بر اساس روش آمبرژه از نوع اقلیم ارتفاعات است و طی یک ارزیابی کلی در گروه سرد قرار می‌گیرد. رژیم بارندگی منطقه از نوع مدیترانه‌ای است. خاک‌های منطقه اغلب از نوع لیتوسل و رگوسل است و عمق خاک در بیش از نیمی از اراضی، کم‌عمق سنگریزه‌دار است. نیم‌رخ خاک‌ها تحول کم تا ناچیز دارد و دارای مواد مادری توف، سازند کرج، ماسه سنگ و کنگلومرا است. بافت خاک در منطقه مورد مطالعه به طور عمده لومی شنی می‌باشد. شکل (۱) نقشه حوضه کرج و موقعیت ایستگاه‌های باران‌سنجی و هیدرومتری را نشان می‌دهد. تنها ایستگاه هیدرومتری در حوضه کرج که بار رسوبات بستر و معلق اندازه‌گیری می‌شود، ایستگاه سیرا کرج است. به منظور برآورد بار رسوبات معلق و بستر در ایستگاه سیرا کرج، از سری زمانی مقادیر بارش در ایستگاه‌های باران‌سنجی و دبی جریان، بار بستر، معلق در ایستگاه سیرا کرج استفاده گردید. دوره زمانی مشترک سه ایستگاه از سال آبی ۸۹-۱۳۸۸ تا سال ۹۸-۱۳۹۷ می‌باشد. جدول (۱) مشخصات آماری متغیرهای مورد استفاده را نشان می‌دهد.

تک واقعه بارش-رواناب (Chang et al. 2017) شبیه‌سازی پیوسته بارش-رواناب (Shafeizadeh et al., 2019) استفاده شده است. ولی تاکنون، الگوریتم PMI برای انتخاب متغیرهای ورودی مؤثر در مدل‌های داده محور برای برآورد بار رسوبات معلق و بستر در رودخانه مورد استفاده قرار نگرفته است. هدف از این تحقیق تحلیل عدم قطعیت در برآورد بار رسوبات معلق و بستر در رودخانه با مدل SVM بعلت عدم قطعیت در پارامترهای مدل با روش شبیه‌سازی مونت کارلو می‌باشد. همچنین متغیرهای ورودی مؤثر بر برآورد بار رسوبات معلق و بستر، با الگوریتم PMI تعیین می‌شوند.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

حوضه آبریز سد کرج قسمتی از دامنه جنوبی ارتفاعات البرز در شمال غرب استان تهران بین طول شرقی ۵-۵۱ تا ۳۰-۵۱ و عرض شمالی ۵۲-۳۵ تا ۱۱-۳۶ قرار گرفته است. این حوضه در تقسیم‌بندی حوضه‌های آبریز کشور جزو حوضه دریاچه نمک قرار



شکل ۱- نقشه حوضه کرج و موقعیت ایستگاه‌های باران‌سنجی و هیدرومتری

جدول ۱- معیارهای آماری متغیرهای مورد استفاده

معیار آماری	بارش متوسط	دبی جریان	بار رسوب معلق	بار رسوب بستر (تن در روز)
حداکثر	۳۱۹/۴	۲۸۲/۷	۲۳۶۲۲۹/۴	۸۹۴۹/۳
حداقل	۰/۰	۲/۹	۰/۷۱	۰/۰
انحراف معیار	۳۵/۲	۲۸/۴	۳۰۲۳۰/۲	۸۷۹/۷
میانگین	۱۱/۶	۱۷/۵	۴۲۴۱/۱	۱۵۷/۴
ضریب تغییرات	۳/۰	۱/۶	۷/۱	۵/۶
دامنه تغییرات	۳۱۹/۴	۲۷۹/۸	۳۳۶۳۲۸/۶	۸۹۴۹/۴

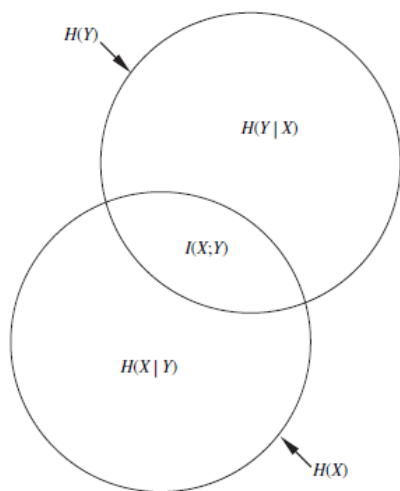
الگوریتم PMI، از جمله الگوریتم غیرخطی برای انتخاب

انتخاب متغیرهای ورودی مؤثر در مدل SVM با الگوریتم PMI

احتمال با تقریب عددی انتگرال در رابطه (۱) خواهیم داشت (May et al., 2008):

$$I(X; Y) \approx \frac{1}{n} \sum_{i=1}^n \log \left[\frac{f(x_i, y_i)}{f(x_i) f(y_i)} \right] \quad (\text{رابطه ۲})$$

که در آن f نمایانگر چگالی برآورد شده بر مبنای نمونه ای از n مشاهده از (x, y) است. از مبناهای مختلفی برای لگاریتم در رابطه (۲) می توان استفاده کرد ولی معمولاً از ۲ یا e استفاده می شود. اگر مبنای لگاریتم مشخص نشده باشد، لگاریتم طبیعی در نظر گرفته می شود. با فرض رابطه (۲) می توان گفت که برآورد دقیق و موثر MI (اطلاعات متقابل) بستگی زیادی به روش بکار گرفته شده در برآورد توابع چگالی احتمال حاشیه ای و توام دارد.



شکل ۲- دیاگرام وین^۶، نشان دهنده رابطه بین اطلاعات متقابل (MI) و آنتروپی برای خروجی Y و متغیر ورودی منفرد X (May et al., 2008)

در مجموع سه معیار توقف الگوریتم PMI وجود دارد که عبارتند از: ۱- مقادیر بحرانی جدول بندی شده^۷، ۲- معیار مبنای اطلاعات اکائیکه^۸ (AIC) (Akaike, 1974)، ۳- معیار آزمون همپل^۹. جداول مقادیر بحرانی ضریب همبستگی (R) به سهولت در دسترس هستند، که بر پایه فرمول تحلیلی برای توزیع خطای یک تخمین برای اندازه نمونه فرض شده استوار است. در مورد ضریب همبستگی خطی (R)، توزیع تخمین نمونه از یک توزیع t پیروی می کند. جداول مقادیر بحرانی ضریب همبستگی (R) بر پایه توزیع t ساخته شده اند (David, 1966)، که مقدار

متغیرهای ورودی موثر در مدل های بر مبنای پردازش داده ها است. الگوریتم انتخاب ورودی بر مبنای PMI (PMIS)، در ابتدا توسط (Sharma 2000) جهت شناسایی متغیرهای ورودی موثر در مدل های هیدرولوژیکی توسعه داده شد. الگوریتم PMI هر تکرار را با در نظر گرفتن یک ورودی (C) و یک خروجی (Y) و با پیدا کردن C_s (فرض می شود که C_s متفاوت از C است) که مقدار PMI را با توجه به متغیر خروجی ماکزیمم می کند (با توجه به ورودی هایی که از قبل انتخاب شده اند)، انجام می دهد. مفهوم آماری که PMI برای C_s برآورد می کند بر مبنای محدوده های اطمینان^۱ است که از توزیع تشکیل شده توسط یک حلقه ی خودراه انداز^۲، تعیین شده است. اگر ورودی معنی دار باشد، C_s به S (مجموعه متغیرهای ورودی انتخاب شده) اضافه شده و انتخاب کردن ادامه می یابد، تا جایی که هیچ ورودی معنی داری باقی نماند، سپس متعاقباً الگوریتم متوقف می شود.

برآورد اطلاعات متقابل جزئی (PMI)

با توجه به یک متغیر خروجی تصادفی Y ، مقداری عدم قطعیت در مورد یک مشاهده y که عضوی از Y می باشد، وجود دارد که می توان آن را با توجه به آنتروپی شانون^۳ H ، (Shannon, 1948) تعریف کرد. ولی با فرض یک متغیر ورودی تصادفی X که Y وابسته به آن است، مشاهدات متقابل (x, y) این عدم قطعیت را کاهش می دهد، زیرا اطلاع از x ، اجازه می دهد تا مقدار y استنتاج شود و برعکس. بنا به تعریف اطلاعات متقابل $I(X; Y)$ ، کاهش در عدم قطعیت متغیر Y به دلیل مشاهده X است (Cover and Thomas, 1991). این مسئله بصورت یک قسمت مشترک بین دو دایره در شکل (۲) نمایش داده شده است. این قسمت مشترک جایی است که عدم قطعیت کاهش یافته حول X و Y به ترتیب توسط آنتروپی شرطی $H(Y|X)$ و $H(X|Y)$ مشخص شده است. اطلاعات متقابل (MI) را می توان با رابطه زیر بصورت مستقیم محاسبه کرد (May et al., 2008):

$$I(X; Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (\text{رابطه ۱})$$

که در آن $p(x)$ و $p(y)$ به ترتیب توابع چگالی احتمال حاشیه ای^۴ (pdfs) X و Y هستند و $p(x, y)$ تابع چگالی توام^۵ است. به هر حال بصورت عملی، فرم صحیح توابع چگالی احتمال در رابطه (۱) مجهول است. از این رو، از برآورد چگالی های احتمال بجای آن استفاده می شود. با جایگذاری برآوردهای چگالی

۶ Venn diagram

۷ Tabulated critical values

۸ Akaike information criterion

۹ Hampel test criterion

۱ Confidence bounds

۲ Bootstrap loop

۳ Shannon entropy

۴ Marginal probability density functions

۵ Joint probability density function

برای آزمون مستقل بودن بر مبنای MI به کار گرفته شود. برای هر اندازه نمونه، ابتدا یک سری $\varepsilon_y \sim N(0,1)$ ساخته می‌شود و سپس توابع چگالی احتمالی حاشیه‌های f_{ε_y} محاسبه می‌شود. مجموع یکصد هزار تکرار از سری $\varepsilon_x \sim N(0,1)$ ، مستقل از $\varepsilon_y \sim N(0,1)$ تولید می‌شود. برای هر نمونه از ε_x ، توابع چگالی احتمالی حاشیه‌های f_{ε_x} و $f_{\varepsilon_x \varepsilon_y}$ تخمین زده و متعاقباً $\hat{I}(\varepsilon_x \varepsilon_y)$ ارزیابی می‌شود. نتایج مقادیر بحرانی I در جدول (۲) برای سطوح اعتماد مختلف درج شده است. دو ضابطه جایگزین متوقف‌کننده فرمول‌بندی شده که به وسیله آن در هر تکرار، $I'_{CSY.S}$ با مقادیر بحرانی مربوطه $Ib^{(95)}$ و $Ib^{(99)}$ که از جدول (۲) بدست آمده مقایسه می‌شود، که از این دو معیار بجای محاسبات مستقیم از سیستم خودراه انداز، جهت تعیین اینکه کدام متغیر باید انتخاب شود یا اینکه الگوریتم متوقف شود، استفاده می‌شود. حذف محاسباتی حلقه سیستم خودراه انداز باعث می‌شود تا انتخاب متغیرهای ورودی بسیار سریعتر انجام شود (May et al., 2008).

بحرانی برای R برای تعداد نمونه‌ها و یک سطح اعتماد معین را تهیه می‌کنند. با این حال برخلاف ضریب همبستگی خطی، یک تعریف تحلیلی معادل برای I را نمی‌توان برطبق رابطه (۲) نتیجه گیری کرد (Goebel et al., 2005). از این رو محققان باید از خودراه انداز کردن، برای محاسبه $f(\hat{I})$ کمک بگیرند (Granger et al., 2004; Sharma, 2000).

به جای استفاده از مقادیر تحلیلی، یک روش برای ساختن جداول مقادیر بحرانی محاسبه شده برای I، استفاده از شبیه‌سازی مونت کارلو که توسط May et al. (2006) شرح داده شده، می‌باشد. شبیه‌سازی مونت کارلو برای تعیین عملی توزیع تخمینگر MI که در گام اول در توسعه یک معیار متوقف‌کننده بر مبنای مقادیر بحرانی تقریبی، به کار گرفته شد. در هر شبیه سازی، MI برای یک سری داده‌ها در مقایسه با داده‌های نویز-سفید^۱ گوسین محاسبه می‌شود (با اندازه نمونه n، که مقداری بین ۵۰ تا ۵۰۰۰ دارد) که این محاسبه برای به دست آوردن داده‌هایی با مقادیر بحرانی استفاده می‌شود که این داده‌ها می‌تواند

جدول ۲- مقادیر بحرانی تخمینگر اطلاعات متقابل بر مبنای برآورد چگالی هسته^۲ (KDE) (May et al., 2006)

$I^{(99)}$	$I^{(95)}$	$I^{(90)}$	\bar{I}	n	$I^{(99)}$	$I^{(95)}$	$I^{(90)}$	\bar{I}	n
۰/۱۰۶۲	۰/۰۹۲۸	۰/۰۸۶۲	۰/۰۶۶۱	۲۸۰	۰/۲۷۰۵	۰/۲۲۲۴	۰/۱۹۹۰	۰/۱۳۲۳	۵۰
۰/۱۰۲۲	۰/۰۸۹۶	۰/۰۸۳۴	۰/۰۶۴۲	۳۰۰	۰/۲۴۵۲	۰/۲۰۳۱	۰/۱۸۲۵	۰/۱۲۳۶	۶۰
۰/۰۸۷۶	۰/۰۷۷۵	۰/۰۷۲۴	۰/۰۵۶۷	۴۰۰	۰/۲۲۵۴	۰/۱۸۷۹	۰/۱۶۹۴	۰/۱۱۶۶	۷۰
۰/۰۷۷۵	۰/۰۶۸۹	۰/۰۶۴۶	۰/۰۵۱۳	۵۰۰	۰/۲۰۹۱	۰/۱۷۵۶	۰/۱۵۹۲	۰/۱۱۰۶	۸۰
۰/۰۷۰۲	۰/۰۶۲۷	۰/۰۵۸۹	۰/۰۴۷۳	۶۰۰	۰/۱۸۵۸	۰/۱۵۷۲	۰/۱۴۲۹	۰/۱۰۱۳	۱۰۰
۰/۰۶۴۴	۰/۰۵۷۸	۰/۰۵۴۴	۰/۰۴۴۱	۷۰۰	۰/۱۶۸۸	۰/۱۴۳۴	۰/۱۳۰۹	۰/۰۹۴۳	۱۲۰
۰/۰۵۹۷	۰/۰۵۳۹	۰/۰۵۰۹	۰/۰۴۱۵	۸۰۰	۰/۱۵۴۶	۰/۱۳۲۱	۰/۱۲۱۱	۰/۰۸۸۳	۱۴۰
۰/۰۵۶۳	۰/۰۵۰۷	۰/۰۴۷۹	۰/۰۳۹۳	۹۰۰	۰/۱۴۴۴	۰/۱۲۳۷	۰/۱۱۳۸	۰/۰۸۳۹	۱۶۰
۰/۰۵۳۱	۰/۰۴۸۱	۰/۰۴۵۵	۰/۰۳۷۵	۱۰۰۰	۰/۱۳۵۶	۰/۱۱۶۶	۰/۱۰۷۲	۰/۰۷۹۸	۱۸۰
۰/۰۳۶۱	۰/۰۳۳۳	۰/۰۳۱۸	۰/۰۲۷۰	۲۰۰۰	۰/۱۲۷۶	۰/۱۱۰۳	۰/۱۰۱۹	۰/۰۷۶۳	۲۰۰
۰/۰۲۸۹	۰/۰۲۶۸	۰/۰۲۵۷	۰/۰۲۲۲	۳۰۰۰	۰/۱۲۱۵	۰/۱۰۵۵	۰/۰۹۷۵	۰/۰۷۳۵	۲۲۰
۰/۰۲۴۷	۰/۰۲۳۰	۰/۰۲۲۱	۰/۰۱۹۲	۴۰۰۰	۰/۱۱۵۸	۰/۱۰۰۵	۰/۰۹۳۲	۰/۰۷۰۷	۲۴۰
۰/۰۲۱۸	۰/۰۲۰۴	۰/۰۱۹۶	۰/۰۱۷۲	۵۰۰۰	۰/۱۱۰۸	۰/۰۹۶۵	۰/۰۸۹۴	۰/۰۶۸۲	۲۶۰

$$AIC = n \log_e \left(\frac{1}{n} \sum_{i=1}^n u_i^2 \right) + 2p \quad (\text{رابطه ۳})$$

که در آن n تعداد مشاهدات، u_i باقیمانده‌های n و p تعداد پارامترهای مدل است. در حالت رگرسیون خطی، تعداد پارامترها برابر با K + 1 است که K برابر با تعداد متغیرها می‌باشد. جهت انتخاب متغیرهای ورودی موثر با استفاده از معیار AIC، متغیرهایی انتخاب می‌شوند که مقدار آنها کاهش یافته است. به منظور برآورد بار رسوبات معلق و بستر در ایستگاه

معیار بر مبنای اطلاعات اکائیکه (AIC)

معیار اطلاعاتی اکائیکه (AIC) (Akaike, 1974)، بعنوان سنجش ارتباط بین دقت فیلتر رگرسیونی و اندازه مجموعه ورودی S، جهت فرمول‌بندی کردن این معیار توقف پذیرفته شده است. مقیاس‌هایی مانند AIC معمولاً به عنوان یک مبنای مقایسه در انتخاب مدل‌ها، مورد استفاده قرار می‌گیرند. رابطه AIC بصورت زیر است:

۱ White-noise

۲ Kernel density estimation

و بستر در ایستگاه هیدرومتری سیرا کرج را نشان می‌دهد. با توجه به جدول (۳)، بار رسوبات معلق و بستر در ایستگاه هیدرومتری سیرا کرج در زمان حال، متغیرهای خروجی هستند. به منظور تعیین متغیرهای موثر بر بار رسوبات معلق و بستر از الگوریتم PMI استفاده شد. این الگوریتم با محاسبه مقدار PMI، متغیرهای ورودی موثر را مشخص می‌سازد. لازم به ذکر است که این الگوریتم به صورت پیش‌فرض در نظر می‌گیرد که متغیرهای ورودی دارای چولگی صفر هستند و از یک توزیع نرمال پیروی می‌کنند. در همین راستا برای استفاده از این الگوریتم بایستی ابتدا داده‌ها را نرمال کرد. جهت نرمال‌سازی داده‌های بارش متوسط در سطح حوضه و بار بستر در ایستگاه سیرا کرج، تمام مقادیر به توان یک عدد کوچک (۰/۰۰۱) رسانده شد. جهت نرمال‌سازی داده‌های دبی جریان و بار رسوب معلق در ایستگاه سیرا کرج، از تبدیل لگاریتمی استفاده گردید و داده‌ها بصورت تبدیل شده به الگوریتم PMI معرفی شدند.

هیدرومتری سیرا کرج، از متغیرهای موثر بر بار رسوبات معلق و بستر بعنوان ورودی در مدل SVM استفاده گردید. بنابراین ابتدا مجموعه پتانسیل متغیرهای ورودی با در نظر گرفتن تاخیر زمانی تا ۱۰ روز قبل برای هر یک از متغیرها شامل بارش متوسط روزانه در سطح حوضه و دبی جریان روزانه بدست آمد. بعد از آماده‌سازی داده‌ها و انتخاب دوره زمانی مشترک بین متغیرها و اعمال تاخیر زمانی جهت ساخت مجموعه پتانسیل متغیرهای ورودی، کل الگوهای آموزشی برابر با ۱۲۵ الگو بدست آمد. از این تعداد کل الگوهای آموزشی، ۷۰٪ الگوها (۸۸ الگو) به منظور آموزش (واسنجی) و ۳۰٪ الگوها (۳۷ الگو) جهت آزمون (اعتبارسنجی) مدل SVM استفاده گردید. بنابراین به منظور آموزش مدل SVM داده‌های از تاریخ ۱۳۸۸/۷/۲۷ تا ۱۳۹۵/۶/۲۲ و جهت آزمون مدل SVM داده‌های از تاریخ ۱۳۹۵/۷/۲۸ تا ۱۳۹۸/۶/۴ در نظر گرفته شد. جدول (۳) مجموعه پتانسیل متغیرهای ورودی و خروجی حاصل از ساخت الگوهای آموزشی برای برآورد بار رسوبات معلق

جدول ۳- مجموعه پتانسیل متغیرهای ورودی و خروجی حاصل از ساخت الگوهای آموزشی در مدل SVM

نام متغیر	معرفی	نوع متغیر
Rave(t-10)	بارش متوسط در ۱۰ روز قبل	ورودی
Rave(t-9)	بارش متوسط در ۹ روز قبل	ورودی
Rave(t-8)	بارش متوسط در ۸ روز قبل	ورودی
Rave(t-7)	بارش متوسط در ۷ روز قبل	ورودی
Rave(t-6)	بارش متوسط در ۶ روز قبل	ورودی
Rave(t-5)	بارش متوسط در ۵ روز قبل	ورودی
Rave(t-4)	بارش متوسط در ۴ روز قبل	ورودی
Rave(t-3)	بارش متوسط در ۳ روز قبل	ورودی
Rave(t-2)	بارش متوسط در ۲ روز قبل	ورودی
Rave(t-1)	بارش متوسط در ۱ روز قبل	ورودی
Rave(t)	بارش متوسط در زمان حال	ورودی
Qw(t-10)	دبی جریان در ۱۰ روز قبل	ورودی
Qw(t-9)	دبی جریان در ۹ روز قبل	ورودی
Qw(t-8)	دبی جریان در ۸ روز قبل	ورودی
Qw(t-7)	دبی جریان در ۷ روز قبل	ورودی
Qw(t-6)	دبی جریان در ۶ روز قبل	ورودی
Qw(t-5)	دبی جریان در ۵ روز قبل	ورودی
Qw(t-4)	دبی جریان در ۴ روز قبل	ورودی
Qw(t-3)	دبی جریان در ۳ روز قبل	ورودی
Qw(t-2)	دبی جریان در ۲ روز قبل	ورودی
Qw(t-1)	دبی جریان در ۱ روز قبل	ورودی
Qw(t)	دبی جریان در زمان حال	ورودی
Qss(t)	بار رسوب معلق در زمان حال	خروجی
Qbs(t)	بار رسوب بستر در زمان حال	خروجی

که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. ماشین بردار پشتیبان یک طبقه بندی کننده دودویی غیرآماری است که

ماشین بردار پشتیبان، یکی از روش‌های یادگیری با نظارت^۱ است

ماشین‌های بردار پشتیبان با حالت‌های مختلف از سطوح تصمیم-گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید کرده و بدین منظور نیاز به تعریف معادله خط می‌باشد. معادله خط در فضای دوبعدی با استفاده از رابطه (۶)، معادله صفحه از رابطه (۷) و معادله صفحه نمایش از رابطه (۸) محاسبه می‌گردد (Scholkopf, 2001).

$$w_1x_1 + w_2x_2 + b = 0 \quad (\text{رابطه ۶})$$

$$w_1x_1 + w_2x_2 + w_3x_3 + b = 0 \quad (\text{رابطه ۷})$$

$$w^T x + b = 0 \quad (\text{رابطه ۸})$$

در مسائلی که داده‌ها به‌طور خطی تفکیک‌پذیر نباشند، می‌توان با نگاشت داده‌ها به یک فضای ویژگی، آنها را به‌صورت خطی جداپذیر نمود. در حالت کلی ابعاد این فضا بینهایت بوده، بنابراین انجام محاسبات در فضای ویژگی می‌تواند زمان‌بر باشد. برای غلبه بر این مشکل از توابع کرنل استفاده می‌شود، در نتیجه معادله صفحه تفکیک‌کننده برای حالت غیرخطی با دخالت تابع کرنل $K(X)$ به‌صورت رابطه زیر می‌باشد (Scholkopf, 2001).

$$w^T K(X) + b = 0 \quad (\text{رابطه ۹})$$

که در آن تابع $K(X)$ با انجام نگاشت، داده‌ها را از فضای غیرخطی به یک فضای خطی منتقل می‌نماید. می‌توان این تابع را به صورت $K(X_i, X_j)$ نیز تعریف نموده و به وسیله آن فضاهای پیچیده‌تری را مورد بررسی قرار داد. به این فرآیند ترفند کرنل^۱ گفته می‌شود. در مدل SVM انتخاب تابع کرنل بسیار مهم بوده و در مسایل مختلف با توجه به ماهیت مسئله می‌توان موارد متفاوتی را در نظر گرفت. بنابراین نمی‌توان تابعی را به‌طور قطعی به عنوان تابع مناسب برای SVM معرفی کرد. انواع مختلف تابع در جدول (۴) ارائه شده است. تابع کرنل خطی، حالت خاصی از تابع کرنل چندجمله‌ای بوده و در مسایل ساده‌تر کارایی خوبی از خود نشان می‌دهد ولی در مسایل پیچیده تابع کرنل چندجمله‌ای می‌تواند بسیار مفیدتر باشد. در تابع کرنل چندجمله‌ای پارامتر d باید از طرق آزمون و خطا یا از طریق الگوریتم‌های بهینه‌سازی محاسبه شود. تابع کرنل گوسی یا شعاعی (RBF) نوع دیگری از توابع کرنل بوده و معمولاً در مسائلی به کار برده می‌شود که هیچگونه شناختی از نوع و ماهیت داده‌ها در دسترس نباشد.

جدول ۴- توابع کرنل مورد استفاده در ماشین‌های بردار پشتیبان

نوع تابع	تابع کرنل
خطی	$K(X_i, X_j) = X_i^T \times X_j$
چند جمله ای	$K(X_i, X_j) = (\gamma X_i^T \times X_j + C)^d$
RBF	$K(X_i, X_j) = -\exp(-\gamma X_i - X_j ^2)$

عدم قطعیت پارامترها با روش مونت-کارلو

در سالهای اخیر بسیار مورد توجه قرار گرفته است. خصوصیات SVM عبارتند از: (۱) ایجاد حدکثر تعمیم در طراحی طبقه‌بندی کننده‌ها، (۲) دارا بودن توانایی یافتن جواب بهینه کلی تابع، (۳) تهیه خودکار ساختار و سازوکار بهینه در حل مسائل طبقه‌بندی، (۴) مدل‌سازی توابع غیرخطی با استفاده از کرنل‌های غیرخطی و شیوه حاصلضرب داخلی در فضاهای هیلبرت. ماشین بردار پشتیبان الگوریتمی است که مدل‌های خطی منحصر به فرد را یافته و با استفاده از آنها حداکثر حاشیه ابر صفحه را حاصل می‌کند. حداکثر شدن حاشیه ابر صفحه، حداکثر شدن تفکیک بین طبقات را حاصل نموده و موجبات بهبود دقت در فرآیند مدل‌سازی را فراهم می‌نماید. به نزدیکترین نقاط آموزشی موجود در حاشیه ابر صفحه، بردارهای پشتیبان اطلاق گردیده و از آنها برای مشخص کردن مرز بین طبقات استفاده می‌شود (Shin et al., 2005).

تفاوت اساسی این روش با طبقه‌بندی‌کننده‌های مرسوم آماری در قابلیت پردازش و طبقه‌بندی داده‌های ابرطیفی بوده و دیگر کاهش تعداد باندها در فرآیندهای مختلف مدل‌سازی نیاز نمی‌باشد. همچنین در این روش با استفاده از یک الگوریتم بهینه-سازی تعداد نمونه‌های مورد نیاز برای تشکیل مرزهای کلاس‌ها به دست آمده و یک مرز تصمیم‌گیری خطی بهینه برای جداسازی کلاس‌ها محاسبه می‌شود. اگر داده‌ها به‌صورت خطی و جدا از هم باشند، SVM با استفاده از ماشین‌های خطی یک سطح بهینه را با کمترین خطا و حداکثر فاصله میان صفحه و نزدیکترین نقاط آموزشی (بردارهای پشتیبان) تفکیک نموده و آموزش می‌دهد (Shin et al., 2005).

اگر نقاط آموزشی را به صورت $[x_i, y_i]$ و بردار ورودی حائز شرط $x_i \in R^n$ باشد، در این صورت ارزش هر طبقه به صورت $y_i \in \{-1, 1\} \quad i = 1, \dots, n$ آنگاه قواعد تصمیم‌گیری که تعریف می‌شود توسط یک صفحه بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند، به‌صورت رابطه (۴) قابل بیان می‌باشد:

$$Y = \text{sgn}(\sum_{i=1}^N y_i a_i (X \times X_i) + b) \quad (\text{رابطه ۴})$$

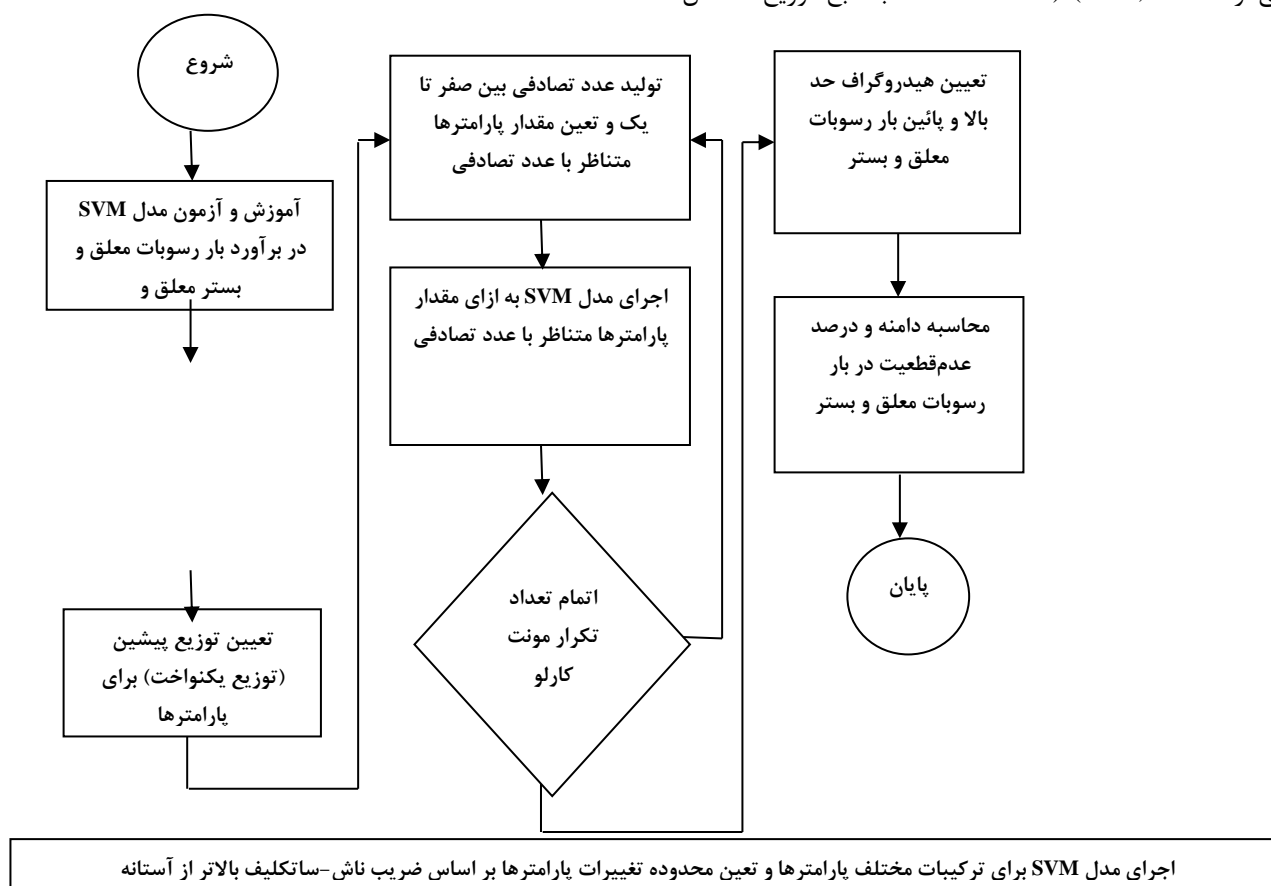
که در آن Y خروجی، y_i ارزش طبقه نمونه X_i و a_i و b پارامترهای تعیین کننده ابر صفحه می‌باشند. اگر امکان تفکیک داده‌ها بصورت خطی فراهم نباشد، در این صورت رابطه (۴) به شرح زیر تغییر می‌یابد:

$$Y = \text{sgn}(\sum_{i=1}^N y_i a_i K(X \times X_i) + b) \quad (\text{رابطه ۵})$$

در این رابطه $K(X \times X_i)$ تابع کرنلی است که برای ایجاد

برای متغیر ورودی، یکی از مسائل موجود در روش مونت کارلو می‌باشد که به دلیل محدودیت اطلاعات و به دنبال آن در دسترس نبودن سری زمانی پارامترها، با توجه به وجود تنها یک عدد برای تمام پارامترها در تمام زیرحوضه و آبراهه‌ها، تابع توزیع یکنواخت به عنوان تابع توزیع احتمال پارامتر مذکور در نظر گرفته شد. شکل (۳) فلوچارت تحلیل عدم قطعیت پارامترهای مدل SVM در برآورد بار رسوبات معلق و بستر در ایستگاه هیدرومتری سیرا کرج به روش شبیه‌سازی مونت کارلو نشان می‌دهد.

اساس کار روش شبیه‌سازی مونت کارلو، تولید ترکیبات تصادفی با توجه به حالات ممکن از عدم قطعیت‌هایی است که در یک مطالعه رخ می‌دهد. بدین صورت که هر متغیر با استفاده از تابع توزیع احتمال آن، شبیه‌سازی و سپس متناظر با آن‌ها خروجی هر مدل محاسبه می‌گردد. این عمل بارها تکرار و سپس با استفاده از داده‌های تولید شده به تجزیه و تحلیل میزان و نحوه تغییرات نتایج مدل پرداخته می‌شود. میزان عدم قطعیت پارامتر خروجی نیز از طریق محاسبه شاخص‌های آماری یا تعیین تابع توزیع آن بیان می‌گردد (Eckhardt et al., 2003). انتخاب تابع توزیع احتمال



اجرای مدل SVM برای ترکیبات مختلف پارامترها و تعیین محدوده تغییرات پارامترها بر اساس ضریب ناش-ساتکلیف بالاتر از آستانه

شکل ۳- فلوچارت تحلیل عدم قطعیت بار رسوبات معلق و بستر به روش مونت کارلو با بکارگیری مدل SVM

تعیین محدوده تغییرات پارامترهای واسنجی، ترکیبات مختلف پارامترها در نظر گرفته و با اعمال ترکیب پارامترها در مدل SVM، مدل اجرا شده و ضریب ناش-ساتکلیف محاسبه گردید. آستانه قابل قبول ضریب ناش-ساتکلیف برای ترکیب پارامترها، ۰/۷ در نظر گرفته شد (Bhavsar and Patel, 20120). جدول (۵) محدوده تغییرات پارامترهای واسنجی مدل SVM در برآورد بار رسوبات معلق و بستر در روش مونت کارلو را نشان می‌دهد. تعداد تکرار مونت کارلو در فلوچارت برای تعیین حد بالا و پائین هیدروگراف بار رسوبات معلق و بستر برابر با ۵۰۰۰ در نظر گرفته شد.

با توجه به اینکه مدل SVM با تابع کرنل پایه شعاعی، بهترین مدل در برآورد بار رسوبات معلق است، بنابراین پارامترهای گاما و واریانس به عنوان پارامترهای واسنجی در نظر گرفته شدند. همچنین با توجه به اینکه بهترین مدل SVM در برآورد بار رسوبات بستر، مدل با تابع کرنل چند جمله‌ای است، بنابراین پارامتر گاما به عنوان پارامترهای واسنجی در نظر گرفته شدند. سپس حدود پائین و بالا تغییرات پارامترهای واسنجی به الگوریتم مونت کارلو معرفی شدند. محدوده تغییرات پارامترهای واسنجی با به کارگیری دوره زمانی واسنجی مدل SVM تعیین گردید. برای

جدول ۵- محدوده تغییرات پارامترهای واسنجی مدل SVM در برآورد بار رسوبات معلق و بستر در روش مونت کارلو

متغیر	تابع کرنل	پارامتر	حد پائین	حد بالا
بار رسوب معلق	پایه شعاعی	σ^2	۰/۲	۱
		γ	۱/۵	۸/۵
بار رسوب بستر	چند جمله‌ای	γ	۹۸/۹	۱۱۳/۵

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2} \quad (\text{رابطه ۱۱})$$

$$(\text{رابطه ۱۲})$$

$$R = (Cov(O_i, S_i) / \sqrt{Cov(O_i, O_i) \cdot Cov(S_i, S_i)})^2$$

نتایج و بحث

جدول (۶) نتایج الگوریتم PMI برای ۱۰ متغیر موثر به ترتیب اولویت برای انتخاب متغیرهای موثر در برآورد بار رسوب معلق را نشان می‌دهد. علاوه بر این نتایج الگوریتم PMI برای ۱۰ متغیر موثر به ترتیب اولویت برای انتخاب متغیرهای موثر در برآورد بار رسوب بستر در جدول (۷) نشان داده شده است. در جداول (۶) و (۷)، Iteration تعداد تکرار الگوریتم PMI، Variable نام متغیر، MC-I*(95) مقدار اطلاعات متقابل جزئی (PMI) برای هر متغیر، -MC-I*(99) مقدار بحرانی MI و AIC(k) و AIC(p) مقدار معیار اطلاعات آکائیکه برای هر متغیر است. برای تعیین متغیرهای ورودی موثر با توجه به مقادیر ضابطه AIC، تا متغیرهایی موثر هستند که مقدار AIC آنها منفی و بصورت نزولی باشند. بنابراین با توجه به مقدار AIC(k) و AIC(p) در جداول (۶) و (۷) تنها متغیر موثر در برآورد بار رسوبات معلق و بستر در ایستگاه هیدرومتری سیرا کرج، دبی جریان در زمان حال است.

معیارهای آماری ارزیابی دقت مدل

برای مقایسه بار رسوب معلق و بستر اندازه‌گیری شده با بار رسوب معلق و بستر برآورد شده در مراحل واسنجی و صحت‌سنجی مدل SVM، شاخص‌های نیکوئی برازش مختلفی را می‌توان استفاده نمود. این شاخص‌ها به دو گروه معیارهای گرافیکی و پارامترهای آماری سنجش خطا قابل تقسیم هستند. از دسته معیارهای گرافیکی می‌توان به ترسیم نمودار تغییرات بار رسوب معلق و بستر اندازه‌گیری شده و برآورد شده در کنار هم اشاره نمود. از شاخص‌های آماری سنجش خطا می‌توان به ضریب ناش-ساتکلیف^۱، ریشه میانگین مربعات خطا و ضریب همبستگی اشاره کرد که به ترتیب در روابط (۱۰) تا (۱۲) آمده است. ضریب ناش-ساتکلیف بیانگر کارایی مدل بوده که می‌تواند مقادیری از منفی بی نهایت تا یک را به خود اختصاص دهد که عدد یک نشان‌دهنده برازش کامل و انطباق کامل بین مقادیر اندازه‌گیری شده و برآورد شده می‌باشد. در روابط (۱۰) تا (۱۲)، n برابر با تعداد داده‌ها، O_i و S_i مقادیر بار رسوب معلق و بستر اندازه‌گیری شده و برآورد شده در گام زمانی نام، \bar{O} میانگین بار رسوب معلق و بستر اندازه‌گیری شده می‌باشد (Nash and Sutcliffe, 1970; Abrahart et al., 2004).

$$NS = 1 - \left[\frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \right] \quad (\text{رابطه ۱۰})$$

جدول ۶- نتایج الگوریتم PMI برای انتخاب متغیرهای موثر در برآورد بار رسوب معلق

تکرار	متغیر	I(x;y)	MC-I*(95)	MC-I*(99)	AIC(k)	AIC(p)
0	logQw(t)	۰/۵۲	۰/۱۳۹	۰/۱۶۴	-۱۰۴/۲	-۹۸/۱
1	logQw(t-2)	۰/۱۰۴	۰/۱۳۹	۰/۱۶۴	-۱۰۰	-۸۸/۷
2	logQw(t-10)	۰/۱۰۷	۰/۱۳۹	۰/۱۶۴	-۹۲/۲	-۷۳/۷
3	logQw(t-3)	۰/۰۹۸	۰/۱۳۹	۰/۱۶۴	-۸۵/۸	-۶۶/۵
4	logQw(t-9)	۰/۱۰۸	۰/۱۳۹	۰/۱۶۴	-۷۸/۳	-۵۸/۴
5	logQw(t-4)	۰/۰۹	۰/۱۳۹	۰/۱۶۴	-۷۴/۵	-۵۴/۵
6	logQw(t-7)	۰/۱۰۲	۰/۱۳۹	۰/۱۶۴	-۷۰/۱	-۵۰/۸
7	logQw(t-8)	۰/۰۸۴	۰/۱۳۹	۰/۱۶۴	-۶۵/۵	-۴۷/۴
8	logQw(t-5)	۰/۰۷۶	۰/۱۳۹	۰/۱۶۴	-۶۲/۲	-۴۴/۶
9	logQw(t-1)	۰/۰۷	۰/۱۳۹	۰/۱۶۴	-۵۶/۷	-۳۶/۸

جدول ۷- نتایج الگوریتم PMI برای انتخاب متغیرهای موثر در برآورد بار رسوب بستر

تکرار	متغیر	I(x;y)	MC-I*(95)	MC-I*(99)	AIC(k)	AIC(p)
0	logQw(t)	۰/۱۳۱	۰/۱۳۹	۰/۱۶۴	-۴۹/۱	-۴۳/۲
1	logQw(t-4)	۰/۱۹	۰/۱۳۹	۰/۱۶۴	-۴۳/۹	-۳۱/۱
2	logQw(t-3)	۰/۰۸۴	۰/۱۳۹	۰/۱۶۴	-۴۰/۵	-۲۶/۵
3	logQw(t-9)	۰/۱۱۱	۰/۱۳۹	۰/۱۶۴	-۳۶/۸	-۱۸/۷
4	logQw(t-7)	۰/۰۷۹	۰/۱۳۹	۰/۱۶۴	-۳۴/۳	-۱۶/۴
5	logQw(t-6)	۰/۰۷	۰/۱۳۹	۰/۱۶۴	-۳۱/۸	-۱۴/۹
6	logQw(t-2)	۰/۰۵۵	۰/۱۳۹	۰/۱۶۴	-۳۰/۱	-۱۱/۹
7	logQw(t-1)	۰/۰۵۱	۰/۱۳۹	۰/۱۶۴	-۲۷/۸	-۷/۵
8	logQw(t-5)	۰/۰۶	۰/۱۳۹	۰/۱۶۴	-۲۵/۲	-۵/۴
9	logQw(t-10)	۰/۰۴۶	۰/۱۳۹	۰/۱۶۴	-۲۲/۹	-۳/۳

تعیین بهترین ساختار مدل SVM

به منظور آموزش و آزمون مدل SVM و تعیین ساختار بهینه مدل SVM در برآورد بار رسوبات معلق و بستر در ایستگاه هیدرومتری سیرا کرج، توابع مختلف کرنل و مقادیر مختلف پارامترهای تابع کرنل در نظر گرفته شد. مقادیر شاخص‌های آماری و مقادیر بهینه پارامترها مربوط به بهترین توابع کرنل مختلف در مدل SVM برای برآورد بار رسوبات معلق به ازای دوره زمانی آموزش و آزمون در جدول (۸) نشان داده شده است. با توجه به جدول (۸)، تابع کرنل پایه شعاعی در دوره زمانی آزمون دارای ضریب ناش-ساتکلایف بیشتری نسبت به دو تابع چندجمله‌ای و خطی داشته است. عبارتی تابع کرنل پایه شعاعی کارایی بهتری نسبت به تابع چندجمله‌ای و خطی در برآورد بار معلق دارد. همچنین جدول (۹) مقادیر شاخص‌های آماری و مقادیر بهینه پارامترها مربوط به بهترین توابع کرنل مختلف در مدل SVM در برآورد بار رسوبات بستر به ازای دوره زمانی آموزش و آزمون نشان می‌دهد. با توجه به جدول (۹)، تابع کرنل چندجمله‌ای در دوره زمانی آزمون دارای ضریب ناش-ساتکلایف بیشتری نسبت به دو تابع کرنل پایه شعاعی و خطی داشته است. عبارتی تابع کرنل چندجمله‌ای کارایی بهتری نسبت به تابع پایه شعاعی و خطی در برآورد بار بستر دارد. اشکال (۴) و (۵) مقایسه نمودار تغییرات بار رسوب معلق مشاهداتی و شبیه‌سازی شده با مدل SVM با تابع کرنل پایه

شعاعی به ترتیب برای دوره های زمانی آموزش و آزمون مدل را نشان می‌دهند. با توجه به اشکال (۴) و (۵) می‌توان دریافت که مدل SVM از لحاظ برآورد بار رسوب معلق بخوبی آموزش دیده است. مقایسه تغییرات بار رسوب معلق مشاهداتی و برآورد شده با مدل SVM در اشکال (۴) و (۵) نشان می‌دهد که مدل SVM مقادیر اوج بار رسوب معلق را با دقت بیشتری نسبت به مقادیر پائین بار رسوب معلق برآورد کرده است.

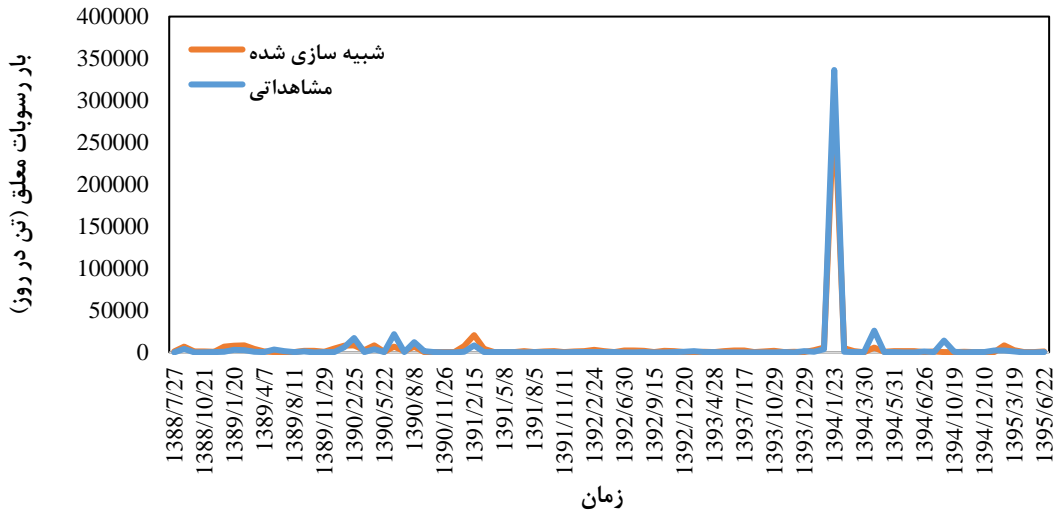
با توجه به جدول (۸) مقدار ضریب ناش-ساتکلایف بین بار رسوب معلق مشاهداتی و شبیه‌سازی شده به ازای دوره زمانی آزمون با بهترین مدل SVM (تابع کرنل پایه شعاعی) برابر با ۰/۸۷ می‌باشد که بیانگر دقت و راندمان خوب مدل در برآورد بار رسوب معلق است. مقایسه نمودار تغییرات بار رسوب بستر مشاهداتی و شبیه‌سازی شده با بهترین مدل SVM (مدل با تابع کرنل چندجمله‌ای) برای دوره‌های زمانی آموزش و آزمون مدل به ترتیب در اشکال (۶) و (۷) نشان داده شده است. با توجه به اشکال (۶) و (۷) می‌توان دریافت که مدل SVM از لحاظ برآورد بار رسوب بستر بخوبی واسنجی شده است. با توجه به جدول (۹) مقدار ضریب ناش-ساتکلایف بین بار رسوب بستر مشاهداتی و شبیه‌سازی شده به ازای دوره زمانی آزمون با بهترین مدل SVM برابر با ۰/۸۴ می‌باشد که بیانگر دقت و راندمان خوب مدل در برآورد بار رسوب بستر است.

جدول ۸- مقادیر شاخص‌های آماری مربوط به توابع کرنل مختلف در مدل SVM برای برآورد بار رسوب معلق به ازای دوره‌های زمانی آموزش و آزمون مدل

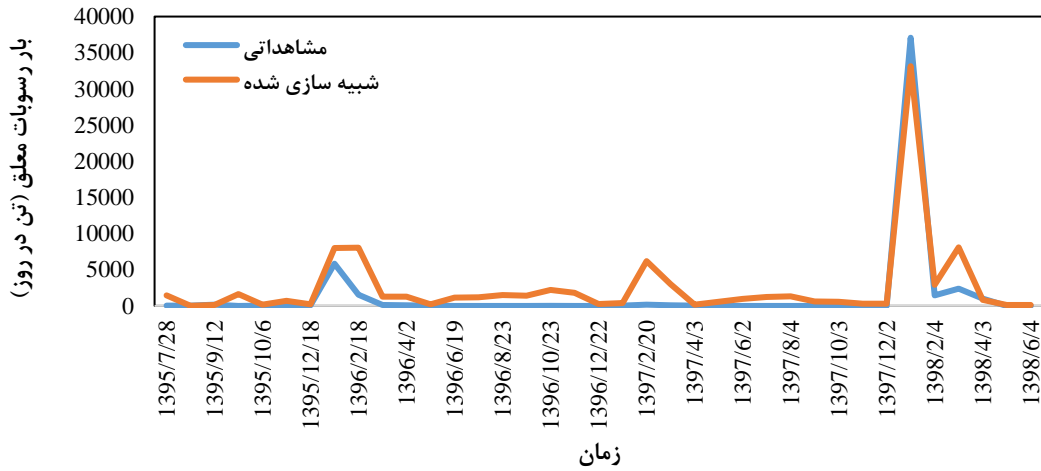
تابع کرنل	آموزش		آزمون					مقادیر بهینه پارامترها	
	RMSE (ton/day)	NS	R	RMSE (ton/day)	NS	R	γ	σ^2	c
چند جمله‌ای	۴۵۷۹/۲	۰/۹۸	۰/۹۹۲	۲۳۳۰/۵	۰/۸۵	۰/۹۶	۱۰۰	-	۴
پایه شعاعی	۵۸۰۴/۳	۰/۹۷	۰/۹۹۳	۲۱۴۱/۲	۰/۸۷	۰/۹۵۵	۵	۰/۶	-
خطی	۱۴۵۶۲/۱	۰/۸۳	۰/۹۴۷	۱۳۲۹۶/۴	-۳/۸۴	۰/۸۵۸	۵	-	-

جدول ۹- مقادیر شاخص‌های آماری مربوط به توابع کرنل مختلف در مدل SVM برای برآورد بار رسوب بستر به ازای دوره‌های زمانی آموزش و آزمون مدل

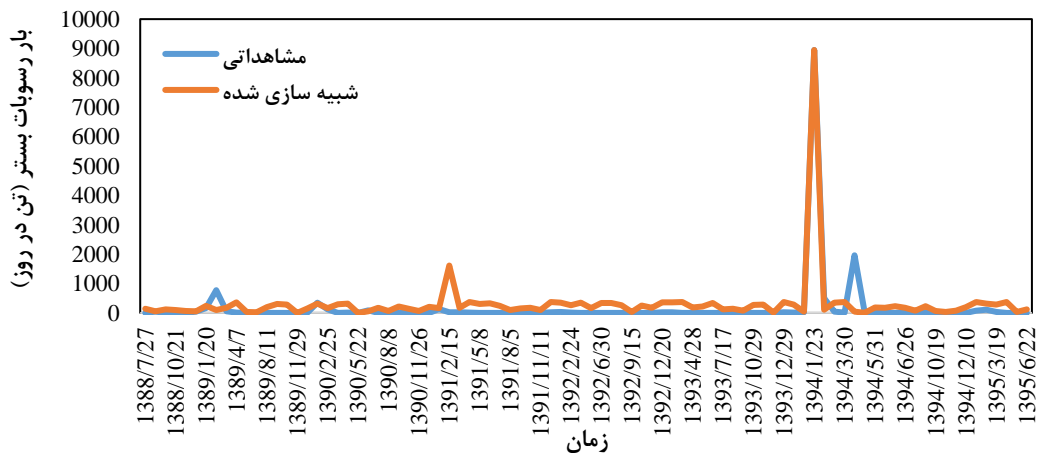
تابع کرنل	آموزش			آزمون			مقادیر بهینه پارامترها			
	RMSE (ton/day)	NS	R	RMSE (ton/day)	NS	R	γ	σ^2	d	c
چند جمله‌ای	۳۴۷/۵	۰/۸۷	۰/۹۴۷	۲۱۸/۴	۰/۸۴	۰/۹۶۹	۱۱۰	-	۴	-۱
پایه شعاعی	۴۰۳/۸	۰/۸۲	۰/۹۶۵	۳۹۰/۱	۰/۴۹	۰/۸۹۱	۱/۲	۰/۵	-	-
خطی	۴۱۸/۲	۰/۸۱	۰/۹۳۳	۳۵۷/۵	۰/۵۷	۰/۸۴۸	۵	-	-	-



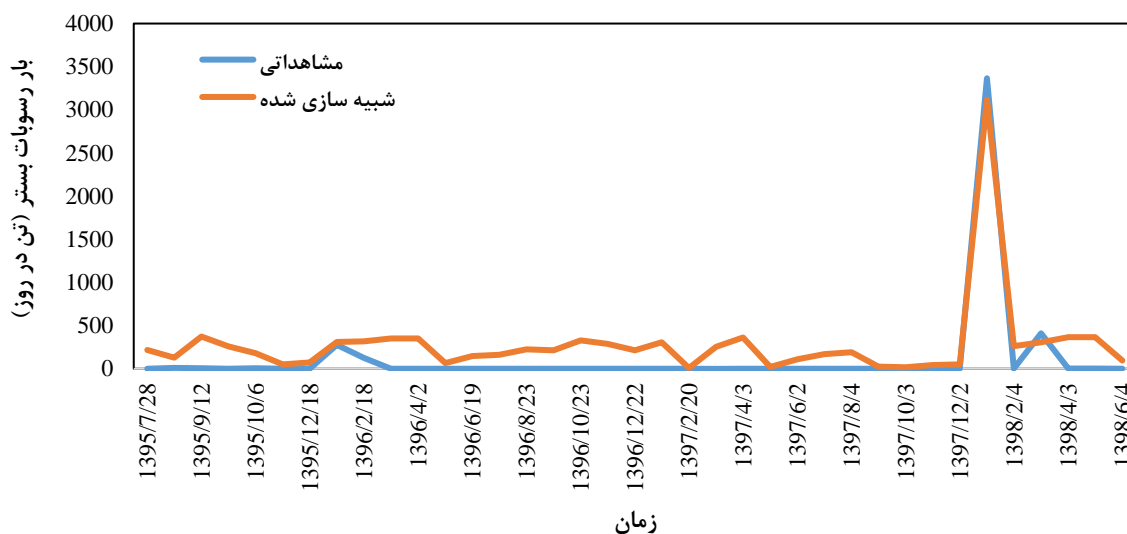
شکل ۳- مقایسه بار رسوب معلق مشاهده‌ای و شبیه‌سازی شده با بهترین مدل SVM برای دوره زمانی آموزش



شکل ۴- مقایسه بار رسوب معلق مشاهده‌ای و شبیه‌سازی شده با بهترین مدل SVM برای دوره زمانی آزمون



شکل ۵- مقایسه بار رسوب بستر مشاهده‌ای و شبیه‌سازی شده با بهترین مدل SVM برای دوره زمانی آموزش

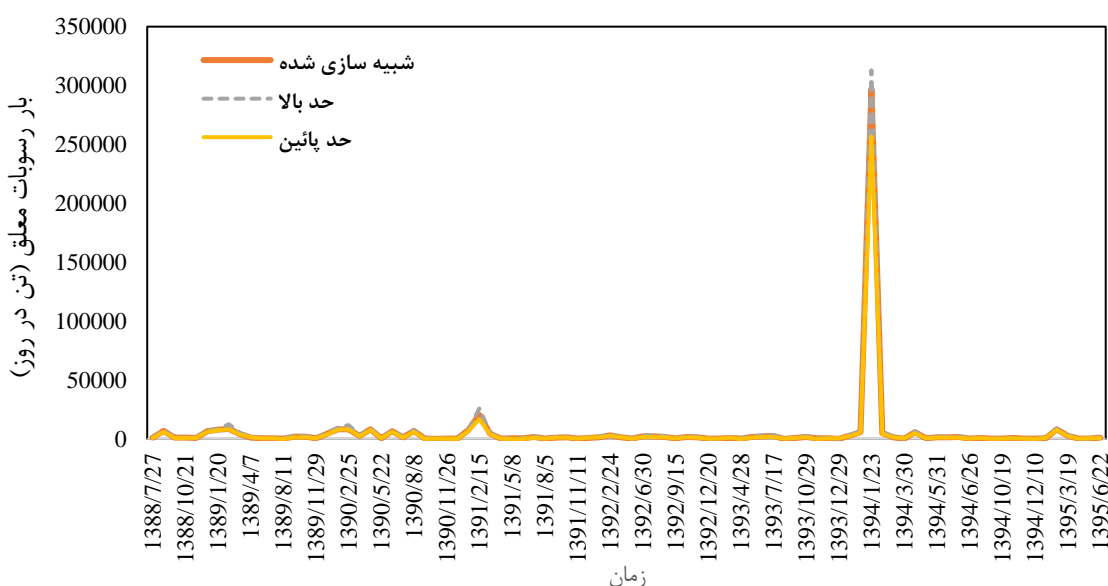


شکل ۷- مقایسه بار رسوب بستر مشاهداتی و شبیه‌سازی شده با بهترین مدل SVM برای دوره زمانی آزمون

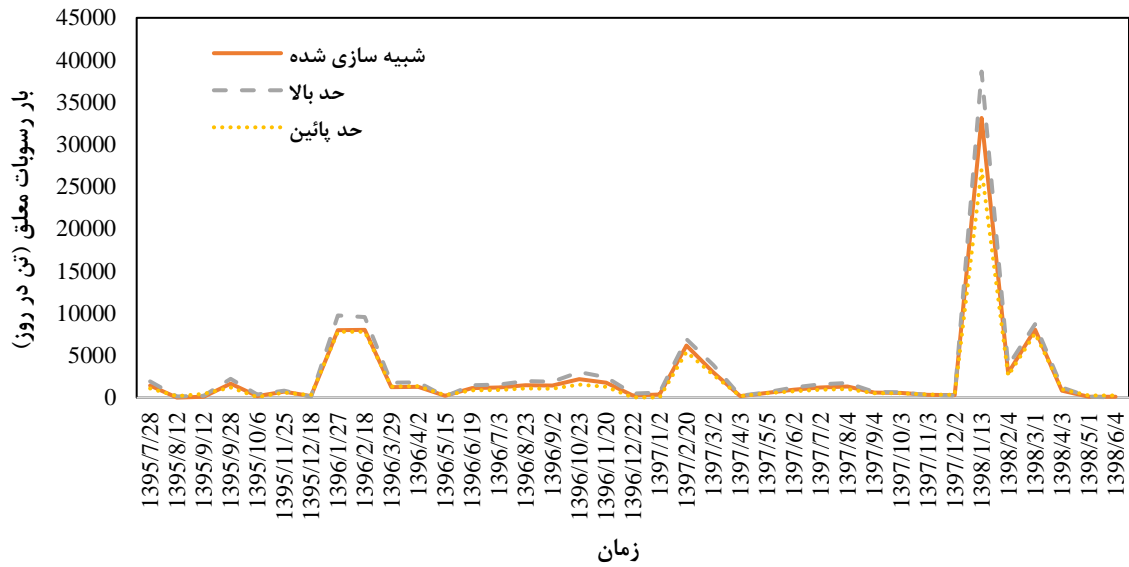
معلق و بستر با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها بعلت عدم قطعیت پارامترها را نشان می‌دهد. با توجه به جدول (۹)، عدم قطعیت در برآورد بار رسوب معلق با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها به ترتیب برابر با ۱۲/۸، ۱۷ و ۱۳/۵ درصد است. همچنین عدم قطعیت در برآورد بار رسوب بستر با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها به ترتیب برابر با ۲۳/۵، ۳۶/۸ و ۲۷/۲ درصد است. بنابراین عدم قطعیت در برآورد بار رسوب بستر با مدل SVM بیشتر از عدم قطعیت در برآورد بار رسوب معلق است. همچنین عدم قطعیت در برآورد بار رسوب معلق و بستر با مدل SVM برای داده‌های آزمون بیشتر از داده‌های آموزش است.

تحلیل عدم قطعیت پارامترها در برآورد بار رسوبات معلق و بستر به منظور تحلیل عدم قطعیت در برآورد بار رسوبات و بستر به علت عدم قطعیت در پارامترهای مدل SVM، از روش شبیه‌سازی مونت کارلو استفاده گردید. بنابراین مقادیر پارامترها در محدوده مورد نظر تغییر داده شد و هیدروگراف رسوب بار معلق و بستر به ازای هر تکرار تعیین گردید. اشکال (۸) و (۹) هیدروگراف حد پائین و بالا بار رسوب معلق و شبیه‌سازی شده با مدل SVM به ترتیب برای داده‌های آموزش و آزمون را نشان می‌دهد. همچنین هیدروگراف حد پائین و بالا بار رسوب بستر و شبیه‌سازی شده با مدل SVM برای داده‌های آموزش و آزمون به ترتیب در اشکال (۱۰) و (۱۱) نشان داده شده است.

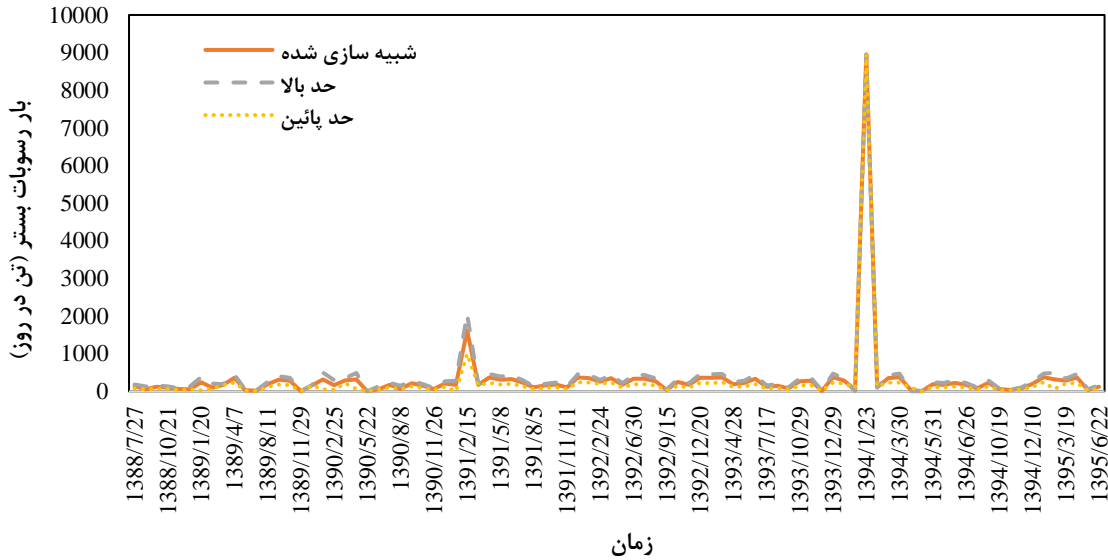
جدول (۱۰) دامنه و درصد عدم قطعیت در برآورد بار رسوب



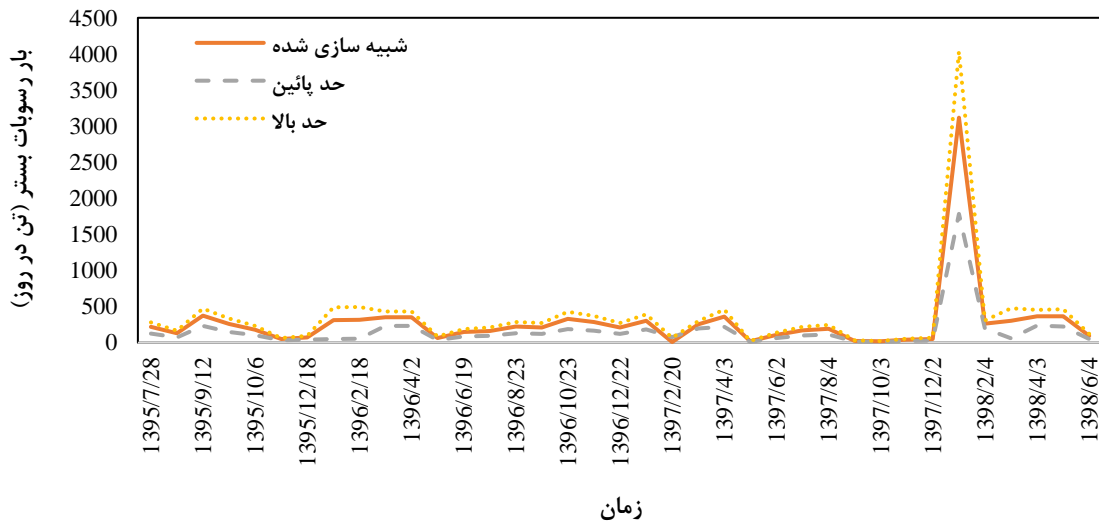
شکل ۸- هیدروگراف حد پائین و بالا بار رسوب معلق و شبیه‌سازی شده با مدل SVM برای داده‌های آموزش



شکل ۹ - هیدروگراف حد پائین و بالا بار رسوب معلق و شبه‌سازی شده با مدل SVM برای داده‌های آزمون



شکل ۱۰ - هیدروگراف حد پائین و بالا بار رسوب بستر و شبه‌سازی شده با مدل SVM برای داده‌های آموزش



شکل ۱۱ - هیدروگراف حد پائین و بالا بار رسوب بستر و شبه‌سازی شده با مدل SVM برای داده‌های آزمون

جدول ۱۰- دامنه و درصد عدم قطعیت در برآورد بار رسوب معلق و بستر با مدل SVM برای داده‌های آموزش، آزمون و کل داده‌ها

متغیر	داده	متوسط بار رسوب حد پائین (تن در روز)	متوسط بار رسوب حد بالا (تن در روز)	متوسط بار رسوب با مدل SVM (تن در روز)	دامنه عدم قطعیت (تن در روز)	عدم قطعیت (%)
بار رسوب معلق	آموزش	۴۸۵۴/۱	۶۲۷۹/۱	۵۵۷۴/۲	۱۴۲۵	۱۲/۸
	آزمون	۲۱۹۷/۱	۳۰۵۲/۳	۲۵۱۹/۴	۸۵۵/۲	۱۷
	کل داده‌ها	۴۰۶۷/۶	۵۳۲۴	۴۶۶۹/۹	۱۲۵۶/۳	۱۳/۵
بار رسوب بستر	آموزش	۲۲۶	۳۷۱/۶	۳۰۹/۳	۱۴۵/۵	۲۳/۵
	آزمون	۱۵۸/۹	۳۶۴/۳	۲۷۹	۲۰۵/۴	۲۶/۸
	کل داده‌ها	۲۰۶/۲	۳۶۹/۴	۳۰۰/۳	۱۶۳/۲	۲۷/۲

نتیجه‌گیری

در این تحقیق بار رسوبات معلق و بستر با مدل SVM که متغیرهای موثر ورودی مدل با الگوریتم PMI انتخاب شده بودند، برآورد گردید. سپس دامنه و درصد عدم قطعیت در بار رسوبات معلق و بستر بعلاوه عدم قطعیت در پارامترهای مدل SVM، با روش شبیه‌سازی مونت-کارلو تعیین گردید. نتایج این تحقیق نشان می‌دهد که بکارگیری الگوریتم PMI جهت تعیین متغیرهای ورودی موثر در مدل SVM موجب افزایش دقت مدل در برآورد بار رسوبات معلق و بستر می‌گردد. علاوه بر این استفاده از الگوریتم PMI موجب کاهش قابل توجه در زمان موردنیاز برای تعیین متغیرهای ورودی موثر و در نتیجه توسعه مدل می‌گردد. (2008) May et al. با مقایسه نتایج مدل ANN توسعه یافته با الگوریتم PMI با نتایج مدل ANN بدون بکارگیری الگوریتم PMI، دریافتند که مدل توسعه یافته با الگوریتم PMI از دقت پیش‌بینی بیشتری همراه با صرفه‌جویی قابل توجه در توسعه مدل برخوردار است. (2020) Sharafati et al. بار معلق رسوب روزانه را با مدل‌های EML پیش‌بینی کردند و با تحلیل عدم قطعیت دریافتند داد که ترکیب متغیرهای ورودی تاثیر بیشتری بر پیش‌بینی به‌دست آمده نسبت به انتخاب ساختار مدل دارد. نتایج حاصل از بکارگیری الگوریتم PMI نشان می‌دهد که مقادیر دبی جریان در همان روز در ایستگاه هیدرومتری، متغیر ورودی موثر بر انتقال بار رسوبات معلق و بستر می‌باشند.

نتایج این تحقیق نشان می‌دهد که عدم قطعیت در برآورد بار رسوب بستر با مدل SVM بعلاوه عدم قطعیت پارامترهای مدل، بیشتر از عدم قطعیت در برآورد بار رسوب معلق است. همچنین عدم قطعیت در برآورد بار رسوب معلق و بستر با مدل SVM برای داده‌های آزمون بیشتر از داده‌های آموزش است. بنابراین در تعیین پارامترهای مدل برای برآورد بار بستر باید دقت بیشتری کرد. (2012) Shen et al. عدم قطعیت پارامترها مربوط به شبیه‌سازی جریان آبراهه و رسوب با روش GLUE در ترکیب با مدل SWAT برآورد کردند. نتایج تحقیق آنها نشان داد که شبیه‌سازی رسوب عدم قطعیت بیشتری نسبت به جریان آبراهه دارد و عدم قطعیت در شرایط بارندگی زیاد (از ماه مه تا سپتامبر) بیشتر از فصل خشک است. منابع اصلی عدم قطعیت جریان آبراهه از فرآیندهای در سطح حوضه آبریز حاصل می‌شود در حالی که فرآیندهای در کانال تأثیر زیادی بر شبیه‌سازی رسوب دارد. با توجه به اینکه عدم قطعیت در برآورد بار رسوبات معلق و بستر بعلاوه عدم قطعیت در پارامترهای مدل قابل ملاحظه می‌باشد، بنابراین به‌کارگیری روش‌های بهینه‌سازی می‌تواند برای برآورد دقیق مقادیر پارامترها و کاهش عدم قطعیت در برآورد بار رسوبات معلق و بستر مفید باشد. در ادامه این تحقیق پیشنهاد می‌گردد که عدم قطعیت پارامترهای مدل SVM در انتقال بار رسوبات معلق و بستر با الگوریتم DREAM روش مونت کارلو زنجیره مارکوف تحلیل شود و با نتایج روش مونت کارلو مقایسه گردد.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

REFERENCES

- Abrahart, R., Kneale, P.E. and See, L.M. (2004). Neural networks for hydrological modeling. CRC Press, 316p
- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716-723
- Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M. and Kisi, O. (2019). Prediction of Suspended Sediment Load Using Data-Driven Models. Water, 11(10), p.2060.
- Bhavsar, P. N. and Patel, J. N. (2020). Event-based rainfall-run-off modeling and uncertainty analysis for lower Tapi Basin, India. *ISH Journal of Hydraulic Engineering*, 26(3), 353-362.
- Bonacci, O. and Oskoruš, D. (2010). The changes in the lower Drava River water level, discharge and suspended sediment regime. *Environmental Earth Sciences*, 59(8), 1661-1670.
- Bowden, G.J., Maier, H.R. and Dandy, G.C. (2005). Input determination for neural network models in water resources applications. Part 2. Case study: Forecasting salinity in a river. *Journal of*

- Hydrology, 301(1-4), 93-107
- Chang, T. K., Talei, A., Alaghmand, S. and Ooi, M. P. L. (2017). Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques. *Journal of Hydrology*, 545, 100-108.
- Cover, T.M. and Thomas, J.A. (1991). Elements of information theory. John Wiley & Sons, Inc., New York, 776p
- Dams, J., Nossent, J., Senbeta, T. B., Willems, P. and Batelaan, O. (2015). Multi-model approach to assess the impact of climate change on runoff. *Journal of Hydrology*, 529, 1601-1616.
- David, F.N. (1966). Tables of the correlation coefficient. In: Pearson ES, Hartley HO (Eds.) *Biometrika tables for statisticians*, third ed., vol. 1. Cambridge University Press, Cambridge.
- Eckhardt, K., Breuer, L., & Frede, H. G. (2003). Parameter uncertainty and the significance of simulated land use change effects. *Journal of Hydrology*, 273(1-4), 164-176.
- Fang, W., Huang, S., Huang, Q., Huang, G., Meng, E., & Luan, J. (2018). Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information. *Journal of Hydrology*, 561, 764-779.
- FathAbadi, A., Ruohani, H., Seyedian, S. M. (2018). The efficiency of nonparametric methods based on residual analyzes and parametric method to estimate hydrological model uncertainty. *Iran Water and Soil Research Journal*, 49(2), 281-292. (In Farsi)
- Fathian, H., AkhondAli, A.M., Sharifi, M.R. (2020). Parameters Uncertainty Analysis in Estimation of Probable Maximum Flood in Bakhtiary Dam Basin by Monte Carlo Method. *Iran Water and Soil Research Journal*, 51(4), 855-871. (In Farsi)
- Gilroy, K. L. and McCuen, R. H. (2012). A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of hydrology*, 414, 40-48.
- Goebel, B., Dawy, Z., Hagenauer, J. and Mueller, J. C. (2005). An approximation to the distribution of finite sample size mutual information estimates. In *IEEE International Conference on Communications*, 2, 1102-1106
- Isazadeh, M., Biazar, S. M. and Ashrafzadeh, A. (2017). Support vector machines and feed-forward neural networks for spatial modeling of groundwater qualitative parameters. *Environmental Earth Sciences*, 76(17), 1-14.
- Jiang, C., Xiong, L., Xu, C. Y. and Guo, S. (2015). Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. *Hydrological Processes*, 29(6), 1521-1534.
- Karami cheme, E. and Mazaheri, M. (2018). Determine of the importance of longitude dispersion coefficient on solute transport in rivers using the Monte Carlo simulation. *Iran Water and Soil Research Journal*, 50(4), 763-776. (In Farsi)
- Lee, D. H. and Kang, D. S. (2016). The application of the artificial neural network ensemble model for simulating streamflow. *Procedia Engineering*, 154, 1217-1224.
- Liu, Y. and Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43, 1-18.
- Miao, C., Ni, J., Borthwick, A. G. and Yang, L. (2011). A preliminary estimate of human and natural contributions to the changes in water discharge and sediment load in the Yellow River. *Global and Planetary Change*, 76(3-4), 196-205.
- May, R. J., Dandy, G. C., Maier, H. R. and Fernando, T. G. (2006). Critical values of a kernel density-based mutual information estimator. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 4898-4903.
- May, R. J., Maier, H. R., Dandy, G. C. and Fernando, T. G. (2008). Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, 23(10-11), 1312-1326.
- Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models; part I: A discussion of principles. *Journal of Hydrology*, 10, 282-290.
- Nourani, V., Molajou, A., Tajbakhsh, A.D. and Najafi, H. (2019). A wavelet based data mining technique for suspended sediment load modeling. *Water Resources Management*, 33(5), 1769-1784.
- Peng, J., Chen, S. and Dong, P. (2010). Temporal variation of sediment load in the Yellow River basin, China, and its impacts on the lower reaches and the river delta. *Catena*, 83(2-3), 135-147.
- Pelletier, J. D. (2012). A spatially distributed model for the long-term suspended sediment discharge and delivery ratio of drainage basins. *Journal of Geophysical Research: Earth Surface*, 117(F2).
- Rodríguez-Blanco, M. L., Taboada-Castro, M. M., Palleiro, L. and Taboada-Castro, M. T. (2010). Temporal changes in suspended sediment transport in an Atlantic catchment, NW Spain. *Geomorphology*, 123(1-2), 181-188.
- Rymszewicz, A., Bruen, M., O'Sullivan, J. J., Turner, J. N., Lawler, D. M., Harrington, J. R., Conroy, E. and Kelly-Quinn, M. (2018). Modelling spatial and temporal variations of annual suspended sediment yields from small agricultural catchments. *Science of The Total Environment*, 619, 672-684.
- Salehpour, j, Ashraf Zadeh, A. and Mosavi S.A. (2019). Investigating the uncertainty of data-based models in forecasting monthly flow of the Hablehroud River. *Iran Water and Soil Research Journal*, (In Farsi)
- Scholkopf, B. (2001). The kernel trick for distances. *Advances in neural information processing systems*, 301-307.
- Shafeizadeh, M., Fathian, H., Nikbakht Shahbazi, A. (2019). Continuous rainfall-runoff simulation by artificial neural networks based on efficient input

- variables selection using partial mutual information (PMI) algorithm. *Iran Water Resources Research*, 15(2), 144-161. (In Farsi)
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423
- Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part 1: A strategy for system predictor identification. *Journal of Hydrology*, 239, 232-239
- Shen, Z.Y., Chen, L., Chen, T. and Di Baldassarre, G. (2012). Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. *Hydrology & Earth System Sciences*, 16(1), 121-132.
- Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert systems with applications*, 28(1), 127-135.
- Sharafati, A., Haji Seyed Asadollah, S.B., Motta, D. and Yaseen, Z.M. (2020). Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal*, 65(12), 2022-2042.
- Syvitski, J. P., & Milliman, J. D. (2007). Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. *The Journal of Geology*, 115(1), 1-19.
- Tena, A., Batalla, R. J., Vericat, D. and López-Tarazón, J. A. (2011). Suspended sediment dynamics in a large regulated river over a 10-year period (the lower Ebro, NE Iberian Peninsula). *Geomorphology*, 125(1), 73-84.
- Vanmaercke, M., Poesen, J., Broeckx, J. and Nysen, J. (2014). Sediment yield in Africa. *Earth-Science Reviews*, 136, 350-368.
- Vogel, R. M., Yaindl, C. and Walter, M. (2011). Nonstationarity: flood magnification and recurrence reduction factors in the United States 1. *JAWRA Journal of the American Water Resources Association*, 47(3), 464-474.
- Walling, D. E. (2006). Human impact on land-ocean sediment transfer by the world's rivers. *Geomorphology*, 79(3-4), 192-216.
- Walling, D. E. (2009). The impact of global change on erosion and sediment transport by rivers: current progress and future challenges. *Unesco*.
- Wieprecht, S., Tolossa, H. G. and Yang, C. T. (2013). A neuro-fuzzy-based modelling approach for sediment transport computation. *Hydrological sciences journal*, 58(3), 587-599.
- Zounemat-Kermani, M., Kişi, Ö., Adamowski, J. and Ramezani-Charmahineh, A. (2016). Evaluation of data driven models for river suspended sediment concentration modeling. *Journal of Hydrology*, 535, 457-472.