



Research Paper

Performance Analysis of Hydrological and Data-Based Models in Estimation of Suspended Sediment Rate

Sahar Javidan¹, Mohammad Taghi Sattari^{2*}, Paria Karimzadeh¹ and Ahmad Mehrabi³

¹M.Sc. Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

²Assoc. Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

³PhD in Water Engineering, Expert in the Office of Water Resources Studies, Qazvin Regional Water Company, Qazvin, Iran

Article information

Received: September 19, 2021

Revised: October 24, 2021

Accepted: October 31, 2021

Keywords:

Gene Expression Programming
Linear Regression
Sediment Rating Curves
Smearing Method

Abstract

Data-driven models are proposed as an alternative to hydrological methods in sediment estimation calculations. This study aimed to compare the performance and accuracy of hydrological and data-based methods in estimating the amount of suspended sediment. For this purpose, discharge and sediment data were collected over 20 yr (2001-2020), and then the amount of suspended sediment at Bagh Kalayeh hydrometric station on Alamut River in Qazvin province, Iran was estimated. In this study hydrological methods including Smearing, FAO, and Sediment Rating Curves versus data-driven methods including Gene Expression Programming, Instance-Based Learning with parameter K and Linear Regression methods were used. The model performances were compared using two statistical methods of RRMSE and NS. The results showed that two techniques such as IBK model with evaluation criteria of ($R = 0.94$, $RRMSE = 0.29$ and $NS = 0.24$) and the GEP model with ($R = 0.85$, $RRMSE = 0.59$ and $NS = 0.65$) estimated suspended sediment in more accurate way than other studies methods. Thus, the superiority of data-driven methods in estimating the amount of suspended sediment in the study area was proved. Therefore, the use of data-based techniques as a competitor and alternative to hydrological methods to estimate the amount of suspended sediment in areas similar to the study area is recommended.

*Corresponding author:
mtsattar@tabrizu.ac.ir



© Authors, Published by **Environment and Water Engineering** journal. This is an open-access article distributed under the CC BY (license <http://creativecommons.org/licenses/by/4.0/>).



Introduction

In countries such as Iran, which are faced with reduced precipitation and a lack of vegetation, watershed management is one of the main challenges for water resources and watershed managers. Sediment transfer from upstream

catchments and sedimentation downstream of basins or reservoirs and supply facilities can have adverse consequences. The creation of sedimentary islands along rivers, especially during heavy rains upstream of the basin, will reduce the transmission capacity of river floods

and incur human and financial losses. Freshwater resources such as rivers are polluted due to sedimentation and this issue negatively affects human life and health. Due to the different physical, meteorological and hydrological conditions of basins, various hydraulic and hydrological methods have been proposed by water engineers to estimate the suspended sediment load.

Generally, in hydrological methods, the sediment flow rate is considered a function of flow rate, regardless of other possible factors. Sediment rating curves are one of the most common methods for estimating suspended sediment loads in rivers. The nonlinear behavior of hydrological variables in the catchment creates skewed relations in linear relationships and reduces the accuracy of classical methods such as sediment rating curves.

However, data-based methods can handle the nonlinear behavior of meteorological and hydrological variables in basins without sufficient attention to the physical conditions of the watersheds. Based on the literature review, in recent years a set of data-based methods including intelligent methods and machine learning has been used as a powerful tool in modeling most hydrological variables such as sediment. One of these successful intelligent methods in estimating the suspended sediment load is the gene expression programming (GEP) model, which is one of the evolutionary methods. This method has a high ability compared to regression methods and other intelligent methods. According to the literature review, it was observed that new data-based methods have good performance in estimating the amount of suspended sediment in different basins with different hydrological and meteorological characteristics.

This study aimed to estimate the amount of suspended sediment in the Bagh Kelayeh hydrometric station in Qazvin province with a semi-arid climate. Hydrological methods including FAO, and Smearing, and data-based methods including linear regression, Instance-Based Learning with parameter K, and gene expression programming have been used. Finally, the performance of these two categories of methods is compared.

Material and Methods

Qazvin province is located in the northwestern of Iran and its area is about 15820 km². Moreover, the average annual rainfall in the last 20 years is

302.51 mm and the average annual temperature is 15.21 C. According to De Martonne's climate classification approach, this region has a semi-arid climate and based on the Ambrose climate classification method, has a cold dry climate.

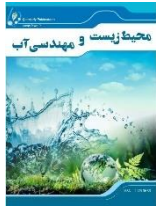
Bagh Kelayeh is a village in Rudbar Alamut district of Qazvin city in Qazvin province and the average rainfall for 20 yr at this station is 423.06 mm. In the present study, a total of 417 observed data related to discharge (m³/s) and sediment (tons/day) parameters in the Bagh Kelayeh hydrometric station located in Qazvin province in a 20-yr statistical period (2001-2020) have been used. The Grubbs statistical test was used for outlier detections, the Standard Normal Homogeneity Test (SNHT) was used to determine the homogeneity of the data, and the Mann-Kendal test method in the Xlstat software environment is used to determine the trend of data. In this study, hydrological methods include sediment rating curve (USBR), FAO and modifier (Ls), and data-based methods including linear regression (LR), gene expression programming (GEP), and Instance-Based k (IBK) are used to estimate the amount of suspended sediment. During the modeling process, 70% of the data were allocated to training and the remaining 30% to the test section. The Weka data mining software was used for applying linear regression/IBK and Gene Xpro Tools software was used for gene expression programming.

To compare the accuracy of the sediment rate estimation models with the observed sediment rates for selecting the appropriate model, the evaluation metrics of correlation coefficient (R), relative root means square error (RRMSE), and Nash-Sutcliffe (NS) have been used. In the sediment rating curve method, a mathematical (usually nonlinear) relationship is established between the suspended sediment load of the river and its discharge. The best fit is passed based on the least-squares method.

Results

The results of the standard normal homogeneity test and Mann-Kendall test showed that the data were homogeneous without any trend, and the data were random.

The IBK method had the highest accuracy in estimating the amount of suspended sediment in Bagh Kelayeh hydrometric station due to its higher correlation and lower relative root mean square error. Other data-based methods also had acceptable accuracy compared to hydrological



ISSN: 2476-3683

محیط زیست و مهندسی آب

Homepage: www.jewe.ir



مقاله پژوهشی

تحلیل عملکرد روش‌های هیدرولوژیکی و داده-مبنا در برآورد میزان رسوب معلق

سحر جاویدان^۱، محمدتقی ستاری^{۲*}، پریا کریمزاده^۱ و احمد مهربانی^۴

^۱ دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران
^۲ دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران
^۳ دکترای مهندسی آب، کارشناس دفتر مطالعات منابع آب، شرکت آب منطقه‌ای استان قزوین، قزوین، ایران

اطلاعات مقاله

تاریخ دریافت: [۱۴۰۰/۰۶/۲۸]

تاریخ بازنگری: [۱۴۰۰/۰۸/۰۲]

تاریخ پذیرش: [۱۴۰۰/۰۸/۰۹]

واژه‌های کلیدی:

بیان ژن
 رگرسیون خطی
 روش اصلاح گر
 منحنی سنجه رسوب

*نویسنده مسئول:

mtsattar@tabrizu.ac.ir



چکیده

مدل‌های داده مبنا به‌عنوان یک جایگزین برای روش‌های هیدرولوژیکی در محاسبات مربوط به تخمین رسوب مطرح هستند. هدف پژوهش حاضر مقایسه عملکرد و دقت روش‌های هیدرولوژیکی و داده-مبنا در برآورد میزان رسوب معلق بود. بدین منظور داده‌های دبی و رسوب در بازه زمانی ۲۰ یژ (۱۳۸۰-۱۳۹۹) جمع‌آوری و سپس میزان رسوب معلق ایستگاه هیدرومتری باغ کلابه بر روی رودخانه الموت در استان قزوین برآورد شد. در این پژوهش از روش‌های هیدرولوژیکی شامل منحنی سنجه رسوب، فائو و روش اصلاح گر و روش‌های داده-مبنای برنامه‌ریزی بیان ژن، یادگیری بر پایه نمونه K و رگرسیون خطی استفاده شد. عملکرد روش‌های مذکور با معیارهای R، RRMSE و NS مورد ارزیابی قرار گرفت. نتایج نشان داد که به ترتیب روش یادگیری بر پایه نمونه K با معیارهای ارزیابی $R=0/94$ ، $RRMSE=0/29$ و روش برنامه‌ریزی بیان ژن با $R=0/85$ ، $RRMSE=0/59$ و $NS=0/65$ رسوب معلق را با دقت بیشتری نسبت به سایر روش‌های مورد مطالعه برآورد کرده است. بدین ترتیب برتری روش‌های داده-مبنا در برآورد میزان رسوب معلق در منطقه مورد مطالعه به اثبات رسید. از این‌رو استفاده از روش‌های داده-مبنا به‌عنوان رقیب و جایگزین روش‌های هیدرولوژیکی برای تخمین میزان رسوب معلق در مناطقی شبیه با منطقه مورد مطالعه توصیه می‌شود.

۱- مقدمه

بالادست حوزه آبخیز باعث کاهش ظرفیت انتقال جریان‌های سیلابی رودخانه و وقوع خسارات جانی و مالی خواهد شد. با توجه به شرایط فیزیکی، آب و هواشناختی و هیدرولوژیکی جهت برآورد بار رسوب معلق، روش‌های متعدد هیدرولیکی و هیدرولوژیکی از سوی مهندسی آب ارائه گردیده است. به‌طور کلی در روش‌های هیدرولوژیکی دبی رسوب، بدون توجه به سایر عوامل ممکن تنها به‌صورت تابعی از دبی

در کشورهایی چون ایران که با کاهش بارش‌های جوی و کمبود پوشش گیاهی مواجه هستند، مدیریت حوزه‌های آبخیز از چالش‌های اساسی مدیران صنعت آب می‌باشد. انتقال رسوب از بالادست حوضه‌ها و رسوب‌گذاری در پایین‌دست حوضه و یا تأسیسات ذخیره و تأمین، می‌تواند پیامدهای منفی داشته باشد. ایجاد جزایر رسوبی در مسیر رودخانه‌ها به‌ویژه در زمان وقوع بارش‌های شدید در

پوشش گیاهی و اقلیمی به ترتیب با داشتن کمترین میزان خطا، از بهترین مدل‌های برآورد بار رسوب معلق در این حوضه می‌باشند. این در حالی است که مدل سالانه که در آن هیچ‌گونه کلاسه‌بندی برای داده‌های دبی جریان و دبی رسوب متناظر آن صورت نگرفته به‌عنوان ناکارآمدترین مدل معرفی گردید. نتایج پژوهش Rezazadeh and Sattari (2017) نشان داد که روش رگرسیون بردار پشتیبان میزان بار رسوبی معلق رودخانه ليقوان چای را با ارائه مقادیر ضریب همبستگی برابر با ۰/۹۵۹ و ریشه میانگین مربعات خطا برابر با ۴۳/۵۴۷ (تن در روز) با دقت بیشتری نسبت به روش k- نزدیک‌ترین همسایگی پیش‌بینی می‌کند. Khosravi et al. (2018) دریافتند که مدل‌های ترکیبی عملکرد بهتری نسبت به مدل‌های جداگانه دارند. بهترین قابلیت پیش‌بینی مربوط به دسته‌بندی مدل درختی (Bagging M5P) و ضعیف‌ترین آن مربوط به مدل REPT^۴ بود. نتایج به‌دست‌آمده از تحقیق Roushangar and Shahnazi (2019) نشان داد که شبکه عصبی مصنوعی با دارا بودن ضریب همبستگی و معیار ناش- ساتکیف به ترتیب برابر با ۰/۹۵۲ و ۰/۹۰۳ برای داده‌های صحت سنجی از عملکرد بهتری برخوردار می‌باشد. Shadkani et al. (2020) نشان دادند که مدل پرسپترون چند لایه نزول گرادیان تصادفی^۴ (MLP-SGD) برای برآورد بار رسوب معلق از دقت بالاتری برخوردار می‌باشد. Zounemat- Kermani et al. (2020) برای پیش‌بینی رسوب معلق رودخانه، مدل GA-ANFIS^۵ و GA-SVR^۶ را به‌عنوان مدل‌های پیش‌بینی کننده برتر در مقایسه با ANFIS، SVR، رگرسیون خطی چندگانه و مدل‌های سنتی معرفی کردند. Keihani et al. (2021) با تحلیل عدم قطعیت پارامترهای مدل ماشین بردار پشتیبان در برآورد بار رسوبات معلق به این نتیجه رسیدند که عدم قطعیت در برآورد بار رسوب معلق با این مدل برای داده‌های آموزش، آزمون و کل داده‌ها به ترتیب برابر با ۱۲/۸، ۱۷ و ۱۳/۵٪ است. Saghebian (2021) در مطالعه‌ای بار معلق رسوبی را با استفاده از روش‌های هوشمند تلفیقی با در نظر گرفتن عدم قطعیت تخمین زده و دریافت که مدل‌های ترکیبی دقت مدل‌سازی را بین ۲۰ تا ۲۵٪ افزایش می‌دهد. ایشان مدل رگرسیون فرآیند گاوسی را به‌عنوان مدلی که

جریان در نظر گرفته می‌شود. منحنی‌های سنجه رسوب یکی از معمول‌ترین روش‌های برآورد بار رسوب معلق رودخانه‌ها هستند. رفتار غیرخطی متغیرهای هیدرولوژیکی در حوضه آبریز باعث به وجود آمدن اریب در روابط خطی شده و از دقت روش‌های کلاسیکی مانند منحنی سنجه رسوب می‌کاهد (Iadanza and Napolitano 2006). این در حالی است که روش‌های مبتنی بر داده، قابلیت در نظر گرفتن رفتار غیرخطی متغیرهای هواشناسی و هیدرولوژیکی حوضه‌ها را بدون توجه کافی بر شرایط فیزیکی حاکم بر مسئله دارا هستند. همچنان‌که در ادامه بررسی منابع خواهیم دید در سال‌های اخیر مجموعه‌ای از روش‌های داده-مبنای از روش‌های هوشمند و یادگیری ماشینی به‌عنوان ابزاری توانمند در مدل‌سازی اغلب متغیرهای هیدرولوژیکی و به‌ویژه رسوب به‌کار گرفته می‌شود. یکی از این روش‌های هوشمند موفق در برآورد بار رسوب معلق، روش برنامه‌ریزی بیان ژن می‌باشد که از جمله روش‌های تکاملی است. این روش توانایی بالایی نسبت به روش‌های رگرسیونی و سایر روش‌های هوشمند دیگر داراست. بررسی منابع نشان می‌دهد پژوهش‌های زیادی در زمینه برآورد میزان رسوب با استفاده از روش‌های هیدرولوژیکی و هوشمند در داخل و خارج کشور انجام یافته است. AYTEK and KISI (2008) به مدل‌سازی رسوب معلق در مونتانا با رویکرد برنامه‌ریزی ژنتیکی پرداختند و نتایج تحقیق آن‌ها نشان داد که GP^۱ برای شبیه‌سازی، در مقایسه با منحنی‌های رتبه‌بندی رسوب و مدل‌های رگرسیون چندخطی بسیار خوب عمل کرده و روشی کاربردی می‌باشد. Dehghani et al. (2010) روش شبکه عصبی مصنوعی با ضریب تبیین ۰/۹۸، مربع خطای متوسط ۰/۰۰۱ را به‌عنوان روش برتر نسبت به روش منحنی سنجه با ضریب تبیین ۰/۷۷۸ و مربع خطای متوسط ۰/۰۴۳ معرفی کرده‌اند. Dehghani and Vafakhah (2013) با تخمین رسوب معلق روزانه، دریافتند که از بین روش‌های مختلف سنجه رسوب، روش دبی کلاسه و از بین شبکه‌های پرسپترون چندلایه^۲ و تابع پایه شعاعی^۳، شبکه پرسپترون چندلایه با ورودی‌های دبی همان روز، دبی ۱ و ۲ روز قبل، با دقت بالایی میزان رسوب را برآورد می‌نماید. Ramezanipour et al. (2017) در پژوهشی نشان دادند که مدل هیدرولوژیکی،

⁴Multilayer Perceptron-Stochastic Gradient Descent

⁵Genetic Algorithm- Adaptive Neuro-Fuzzy Inference System

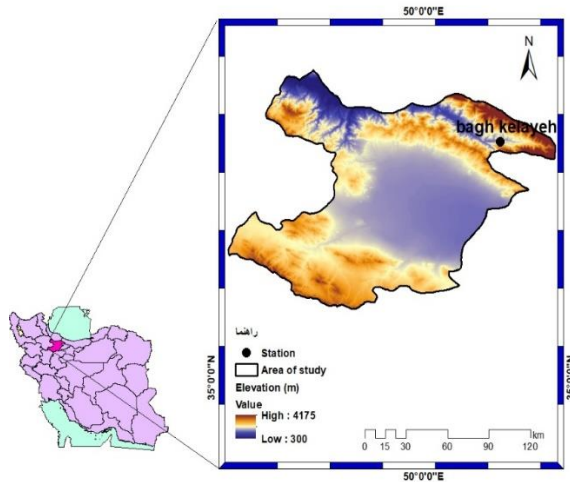
⁶Genetic Algorithm- Support Vector Regression

¹Genetic Programming

²Multilayer Perceptron

³Radial Basis Function





شکل ۱- موقعیت مکانی ایستگاه مورد مطالعه
Fig. 1 Location of the studied station

در پژوهش حاضر مجموعاً از ۴۱۷ داده مربوط به پارامتر دبی و رسوب (m^3/s) و رسوب ($tons/day$) در ایستگاه هیدرومتری باغ کلایه واقع در استان قزوین در دوره آماری ۲۰ yr (۱۳۹۹-۱۳۸۰) استفاده شد. مشخصات آماری متغیرهای مورد استفاده در جدول (۱) ارائه شده است.

جدول ۱- مشخصات آماری پارامترهای مورد استفاده برای ایستگاه باغ کلایه

Table 1 Statistical characteristics of implemented parameters for Bagh Kelaveh Station

Statistical Parameter	Q (m^3/s)	S ($tons/day$)
Mean	9.85	6633.75
Median	4.678	148.682
Max	89.39	474336.09
Min	1.45	0.697
Sd	11.42	36909.11
CV	1.16	5.56
Cs	2.817	9.393
Ck	10.752	98.829

طبق طبقه‌بندی ارائه شده توسط Warrick and Nielsen (1980)، (۱) $Cv < 0.12$ (نشان‌دهنده تغییرات کم، ۲) $0.60 < Cv < 0.12$ (متوسط و ۳) $Cv > 0.60$ زیاد می‌باشد، بنابراین طبق جدول (۱) رسوب و دبی هر دو دارای تغییرات کم می‌باشند. نزدیک بودن مقادیر میانگین و میانه به هم، همچنین نزدیک بودن ضریب چولگی به صفر و ضریب کشیدگی به سه، بیانگر توزیع متقارن بین داده‌ها است که عدم وجود شرایط مذکور برای داده‌های رسوب و دبی نشان‌دهنده نامتقارن بودن داده‌ها است. البته این شرایط با توجه به وضعیت اقلیمی و ماهیت ذاتی داده‌های رسوب دور انتظار نیست. همچنین برای تعیین داده‌های پرت از روش

دارای درجه عدم اطمینان مطلوبی در مدل‌سازی است معرفی کردند.

همچنان که در بررسی منابع ملاحظه گردید روش‌های جدید مبتنی بر داده عملکرد مطلوبی در تخمین میزان رسوب معلق در حوضه‌های مختلف با ویژگی‌های هیدرولوژیکی و هواشناختی متفاوت داشته‌اند. هدف پژوهش حاضر، برآورد میزان رسوب معلق در ایستگاه هیدرومتری باغ کلایه در استان قزوین با اقلیم نیمه‌خشک است. دوره آماری ۲۰ yr (۱۳۸۰-۱۳۹۹) در نظر گرفته شد. از روش‌های هیدرولوژیکی شامل Smearing و FAO، USBR و روش‌های داده‌مبنا شامل روش‌های رگرسیون خطی، یادگیری بر پایه نمونه K و برنامه‌ریزی بیان ژن استفاده شده است. در نهایت عملکرد این دو دسته روش مقایسه گردیده است.

۲- مواد و روش‌ها

۲-۱ منطقه مورد مطالعه و داده‌های مورد استفاده

استان قزوین در بخش شمال غربی ایران واقع شده و مساحت آن حدود 15820 km^2 می‌باشد، میانگین بارش سالانه در ۲۰ سال اخیر $302/51 \text{ mm}$ و متوسط دمای سالانه $15/21 \text{ }^\circ\text{C}$ بوده که براساس طبقه‌بندی دومارتن دارای اقلیم نیمه‌خشک و براساس اقلیم‌نمای آمبروزه دارای اقلیم خشک سرد می‌باشد. مرتفع‌ترین کوه‌های استان «سیاهلان، کی جکین، سفید کوه و سیاه کوه» بوده که حداکثر ارتفاع و کوه‌های شمالی سیاهلان 4175 m از سطح دریا می‌باشد. محدوده‌ی مرکزی و شرق استان را دشت تشکیل می‌دهد که شیب آن از شمال غرب به جنوب شرق امتداد دارد و در پایین‌ترین نقطه 1130 m است. حداقل ارتفاع آن در شمال غربی و در بخش طارم سفلی و کناره‌های دریاچه سفیدرود با ارتفاع 300 m از سطح دریا است (Anonymous 2021). باغ کلایه، روستایی از توابع بخش رودبار الموت شهرستان قزوین در استان قزوین است. ایستگاه هیدرومتری باغ کلایه در عرض جغرافیایی $38^\circ 23' 36''$ طول جغرافیایی $51^\circ 29' 50''$ و ارتفاع 1287 m از سطح دریا واقع شده است. میانگین بارش ۲۰ yr در این ایستگاه $423/06 \text{ mm}$ می‌باشد. موقعیت مکانی ایستگاه مورد مطالعه در شکل (۱) آورده شده است.

می‌دهد، افزایش اندک دبی جریان، قدرت فرساینده‌گی را به‌شدت افزایش می‌دهد (Dadashzadehasl et al. 2009).

۲-۲-۲- روش فائو

این روش، اصلاح‌شده روش USBR است. در آن شیب‌خط رگرسیون منحنی سنج رسوب ثابت فرض شده و برای نزدیک کردن مقادیر برآورد شده از منحنی سنج رسوب به مقادیر مشاهده شده توصیه می‌شود که در رابطه (۱) به‌جای ضریب a از ضریب a' که از رابطه (۴) محاسبه می‌شود، استفاده شود.

$$a' = \frac{\bar{Q}_s}{\bar{Q}_w^b} \quad (4)$$

که، \bar{Q}_s متوسط دبی رسوب، \bar{Q}_w متوسط دبی جریان متناظر آن و b همان ضریب معادله USBR است (Jones et al 1981).

۲-۲-۳- روش اصلاح‌گر

روشی غیرپارامتری است که با استفاده از رابطه (۵) محاسبه می‌شود.

$$L_s = L_{RC} \times \frac{\sum_{i=1}^N \exp(e_i)}{N} \quad (5)$$

که، L_s بیانگر رسوب محاسبه شده از روش اصلاح‌گر، L_{RC} رسوب محاسبه شده از روش USBR، e_i حداقل مربعات باقی‌مانده با استفاده از معادله منحنی سنج (اختلاف لگاریتم طبیعی رسوب مشاهده‌ای و محاسبه شده) است (Pavanelli and Bigi 2004).

۲-۲-۴- روش رگرسیون خطی

رگرسیون خطی یکی از روش‌های تحلیل رگرسیونی است. رگرسیون یک نوع مدل آماری برای پیش‌بینی یک متغیر از روی یک یا چند متغیر دیگر است. رگرسیون خطی نوعی تابع پیش‌بینی‌کننده خطی است که هرکدام از متغیرهای مستقل در ضریبی که در فرآیند تخمین برای آن متغیر به‌دست‌آمده ضرب می‌شود، جواب نهایی مجموع حاصل‌ضرب‌ها به‌علاوه یک مقدار ثابت خواهد بود که آن‌هم در فرآیند تخمین به‌دست‌آمده است. ساده‌ترین نوع رگرسیون خطی، رگرسیون خطی ساده است که برخلاف رگرسیون خطی چندگانه، تنها یک متغیر مستقل دارد. فرآیند تخمین سعی می‌کند ضرایب مدل رگرسیون خطی را به‌گونه‌ای انتخاب کند که با داده‌های موجود همخوانی داشته باشد،

گرایز بک، برای تعیین همگنی داده‌ها از آزمون همگنی نرمال استاندارد (SNHT)^۱ و برای تعیین روند از روش من کندانال در محیط نرم‌افزار Xlstat استفاده شد. در این مطالعه از روش‌های هیدرولوژیکی شامل منحنی سنج رسوب 2 (USBR)، فائو^۳ (FAO) و اصلاح‌گر^۴ (LS) و از روش‌های داده-مبنا شامل روش‌های رگرسیون خطی^۵ (LR)، برنامه‌ریزی بیان ژن^۶ (GEP) و یادگیری بر پایه نمونه^۷ K (IBK) برای برآورد میزان رسوب معلق استفاده شده است. در جریان مدل‌سازی، ۷۰٪ داده‌ها به واسنجی و ۳۰٪ بقیه داده‌ها به صحت‌سنجی اختصاص یافت. برای روش رگرسیون خطی و یادگیری بر پایه نمونه K از نرم‌افزار Weka و برای برنامه‌ریزی بیان ژن از نرم‌افزار Gene Xpro Tools استفاده شد.

۲-۲-۲- روش‌های مورد استفاده

۱-۲-۲- روش منحنی سنج رسوب

در این روش یک رابطه ریاضی (معمولاً غیرخطی) بین بار رسوبی معلق رودخانه و آبدهی آن برقرار می‌شود، در عمل با توجه به داده‌های دبی آب و دبی رسوب متناظر با آن، هر دو سری داده به صفحه مختصات لگاریتمی منتقل شده و خط بهترین برازش بر مبنای روش حداقل مربعات از میان آن‌ها عبور داده می‌شود و رابطه‌ای به‌صورت رابطه (۱) که به منحنی سنج رسوب معروف است بین دو متغیر برقرار می‌شود.

$$Q_s = aQ_w^b \quad (1)$$

که، Q_s بیانگر دبی رسوب، Q_w دبی جریان، a و b ضرایب ثابتی هستند که به روش کمترین مربعات خطا از روابط (۲) و (۳) محاسبه می‌شوند (Youssefvand 2004).

$$\log a = \frac{\sum \log Q_s - b \sum \log Q_w}{N} \quad (2)$$

$$b = \frac{N \sum \log Q_w \cdot \log Q_s - \sum \log Q_w \cdot \sum \log Q_s}{N \sum (\log Q_w)^2 - (\sum \log Q_w)^2} \quad (3)$$

ضریب a شاخصی از شدت فرسایش است، مقدار بالای آن نشان‌دهنده فرسایش بالای تشکیلات زمین‌شناسی است که انتقال می‌یابد. ضریب b توان فرساینده‌گی رودخانه را نشان

¹Standard Normal Homogeneity Test

²United State Bureau of Reclamation

³Food and Agricultural Organization

⁴Smearing

⁵Linear Regression

⁶Gene Expression Programming

⁷Instance Based Learning with Parameter K



افراد از جمعیت حاضر نگه‌داشته می‌شوند. (۶) بقیه افراد جمعیت حاضر براساس عملکردشان انتخاب می‌شوند. (۷) اصلاحات و بهبودهایی (تکثیر و جهش) روی جمعیت انتخاب‌شده صورت می‌گیرد، در نتیجه فرزندان با ویژگی‌های جدید تولید می‌شوند. (۸) فرزندان جدید، در یک چرخه، تحت همان فرایند توسعه قرار می‌گیرند و این فرایند برای تعداد معینی از نسل‌ها تکرار می‌شود تا راه‌حل مناسبی پیدا شود (Ferreira 2001). هنگام مدل‌سازی و برآورد میزان رسوب با روش برنامه‌ریزی بیان ژن، پس از سعی و خطاهای متعدد و بررسی تعداد ژن‌ها و کروموزوم‌های متفاوت، در نهایت بهترین عملکرد با تعداد ژن ۱۰ و کروموزوم ۲۰ حاصل شد.

۲-۲-۶- مدل یادگیری بر پایه نمونه K

روش‌های یادگیری بر پایه نمونه روش‌های ناپارامتریکی هستند که روابط از پیش تعیین‌شده برای پیش‌بینی متغیر مجهول استفاده نمی‌کنند و به‌جای آن از نمونه‌های مشخصی برای پیش‌بینی متغیر استفاده می‌کنند (Aha et al. 1991). روش IBK یک الگوریتم یادگیری براساس مشاهدات و نمونه‌ها می‌باشد (Shirzad et al. 2008). IBK یک رده‌بند با K همسایه نزدیک است که معیار فاصله ذکرشده را استفاده می‌کند. تعداد نزدیک‌ترین فاصله‌ها (پیش‌فرض K=1) می‌تواند به‌طور صریح در ویرایشگر شیء تعریف شود. پیش‌بینی‌های متعلق به بیش از یک همسایه می‌تواند براساس فاصله آن‌ها تا نمونه‌های آزمایش وزن‌دار گردد (Jalali and Homaei 2011). الگوریتم IB1 ساده‌ترین این الگوریتم‌ها و مشابه الگوریتم‌های نزدیک‌ترین همسایه هستند با این تفاوت که IB1 دامنه متغیرها را نرمال می‌کند و روشی ساده در برخورد با مقادیر گم‌شده دارد (Aha et al. 1991). اولین قدم برای استفاده از این الگوریتم، یافتن یک روش و رابطه برای محاسبه فاصله بین داده‌های آزمون و آموزش است.

$$d(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (10)$$

متغیر X داده‌های آموزش را با پارامترهای مشخص شده (X₁ تا X_n) و Y داده‌های آموزش را با همان پارامتر (Y₁ تا Y_n) نشان می‌دهد. بعدازاینکه فاصله اقلیدسی بین داده‌ها تعیین شد، نمونه‌های پایگاه داده به ترتیب صعودی از حداقل فاصله (حداکثر تشابه) تا حداکثر فاصله (حداقل تشابه) طبقه‌بندی می‌شوند. سپس تعداد نقاط (K) تعیین می‌شود. اگر K کوچک فرض شود نتایج به نقاط منفرد غیرمتعارف مدل

یعنی پیش‌بینی‌ها به مقادیر مشاهداتی نزدیک باشند. رابطه (۶) یک رابطه رگرسیون خطی را نشان می‌دهد.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_0 \quad (6)$$

که، ε_0 مقدار خطا، i شماره هر مشاهده، y_i متغیر وابسته و x_i متغیر مستقل می‌باشد. با داشتن مجموعه‌ای از این نقطه‌ها می‌توان مدل را به‌دست آورد (رابطه ۷).

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \quad (7)$$

که، e_i مانده نام دارد و تخمینی از اختلاف بین مقدار محاسبه شده و مقدار واقعی متغیر وابسته است. برای تخمین این مدل رگرسیون باید سه پارامتر $\hat{\beta}_0$ و $\hat{\beta}_1$ و e_i تخمین زده بشوند. ضرایب از روابط (۸) و (۹) به‌دست می‌آیند (Rencher and Christensen 2012).

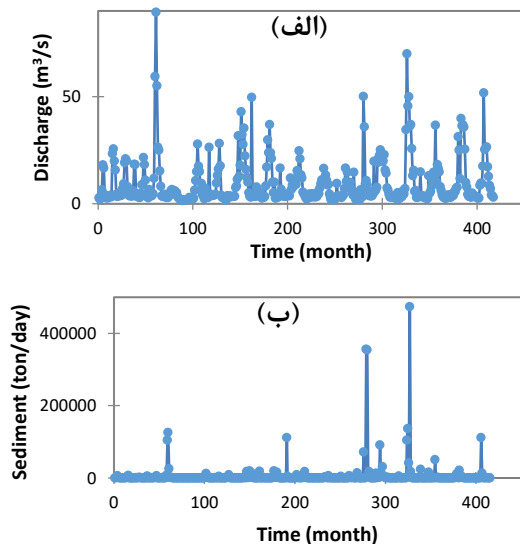
$$\hat{\beta}_0 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (8)$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_0 \bar{x} \quad (9)$$

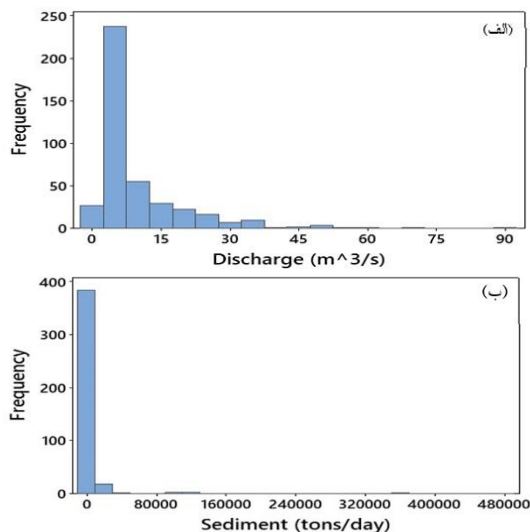
۲-۲-۵- روش برنامه‌ریزی بیان

روش برنامه‌ریزی بیان ژن که شکل توسعه‌یافته برنامه‌ریزی ژنتیک می‌باشد توسط Ferreira (2001) ارائه و روشی مناسب برای مدل‌سازی فرآیندهای غیرخطی و پیچیده است که جزء روش‌های الگوریتم گردشی محسوب می‌شود. مبنای آن براساس نظریه تکامل داروین استوار است. کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک ترکیب می‌شوند، ساختار خطی کروموزوم‌ها موجب می‌شود استفاده از عملگرهای ژنتیکی نظیر جهش و ترکیب همیشه ساختارهای صحیح و معتبری تولید کند. این الگوریتم به‌طور خودکار می‌تواند متغیرهای ورودی که در مدل بیشترین تأثیر را دارند، انتخاب کند. به‌طور کلی فرآیند گام‌به‌گام مدل برنامه‌ریزی بیان ژن به‌صورت زیر می‌باشد: (۱) فرآیند با تولید تصادفی کروموزوم‌ها از تعداد معینی جمعیت اولیه آغاز می‌شود. (۲) کروموزوم‌ها به‌صورت بیان درختی بیان می‌شوند. (۳) میزان مطلوبیت هر جواب بالقوه مسئله (کروموزوم) و درجه سازگاری آن ارزیابی می‌شود. (۴) اگر شرایط مطلوب حاصل شده باشد برنامه متوقف و جمعیت موجود جواب مطلوب را نشان می‌دهد در غیر این صورت: (۵) بهترین





شکل ۲- نمودار سری زمانی: الف) دبی و ب) رسوب
Fig. 2 Time series diagrams of: a) discharge and b) sediment



شکل ۳- هیستوگرام: الف) دبی و ب) رسوب

Fig. 3 Histogram of: a) discharge and b) sediment

جدول ۲- معیارهای آماری روش‌های هیدرولوژیکی و داده مینا
Table 2 Statistical criteria of hydrological and data-based methods

Model	R	RRMSE	NS
LR	0.8272	0.290834	0.224
IBK	0.9443	0.288822	0.242
GEP	0.85064	0.587604	0.652
FAO	0.6029	0.9381	0.113
USBR	0.603	0.9362	0.128
Ls	0.6029	0.9309	0.126

با توجه به جدول (۲) روش یادگیری بر پایه نمونه K به دلیل همبستگی بیشتر ($R=0.9443$) و ریشه نسبی خطای میانگین مربع کمتر ($RRMSE=0.2888$) بالاترین دقت را در برآورد میزان رسوب معلق در ایستگاه هیدرومتری باغ کلاهی داشته است. سایر روش‌های داده مبنای مورد مطالعه

حساس می‌شود و اگر فرض شود K قابل توجه است می‌توان مقداری از سایر کلاس‌ها را در محدوده مورد نظر قرار داد. بهترین مقدار برای K با استفاده از روش اعتبارسنجی متقابل به دست می‌آید (Wu et al. 2008). در این روش پس از انجام سعی و خطاهای متعدد مشخص گردید، نتیجه بهینه زمانی حاصل می‌شود که تعداد همسایگان مورد استفاده برابر با ۵ و تابع فاصله‌سنجی بهینه، تابع اقلیدسی در نظر گرفته شود.

۳-۲- معیارهای ارزیابی مدل‌ها

برای مقایسه دقت میزان رسوب برآورد شده از روش‌های مورد مطالعه با میزان رسوب مشاهده شده و انتخاب مدل مناسب، از شاخص‌های ضریب همبستگی^۱ (R)، ریشه نسبی خطای میانگین مربع^۲ (RRMSE) و نش-ساتکلیف^۳ (NS) استفاده شده است. فرمول‌های آماره‌های فوق به ترتیب در روابط (۱۱) تا (۱۳) ارائه شده است.

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (11)$$

$$RRMSE = \frac{RMSE}{Sd_{xi}} = \frac{\sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}}{Sd_{xi}} \quad (12)$$

$$NS = 1 - \left(\frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right) \quad (13)$$

که، x_i مقادیر مشاهداتی، y_i مقادیر محاسباتی و Sd مقدار انحراف معیار می‌باشد.

۳- یافته‌ها و بحث

نتایج آزمون همگنی نرمال استاندارد و آزمون من کندیال بیانگر همگن بودن داده‌ها و همچنین عدم وجود روند یعنی تصادفی بودن داده‌ها بود. نمودار سری زمانی دبی و رسوب در شکل (۲) و هیستوگرام در شکل (۳) نشان داده شده است. پس از انجام محاسبات، دقت روش‌های داده مینا و روش‌های هیدرولوژیکی براساس معیارهای آماری (روابط ۱۱ تا ۱۳) محاسبه شده و در جدول (۲) ارائه شده است.

¹Coefficient Of Correlation

²Relative Root Mean Square Error

³Nash Sutcliffe

نشان‌دهنده بالا بودن مقدار همبستگی و در قسمت ب نزدیک بودن رأس‌های ۶ ضلعی رسم شده به رأس‌های ۶ ضلعی بیرونی نشان‌دهنده بالا بودن خطا می‌باشد. با توجه به شکل ملاحظه می‌شود که روش یادگیری بر پایه نمونه K بالاترین همبستگی و کمترین خطا را داشته‌اند. با توجه به قسمت (ج) می‌توان نتیجه گرفت که روش برنامه‌ریزی بیان ژن به علت نزدیک بودن رأس ۶ ضلعی ترسیم‌شده به ۶ ضلعی بیرونی یعنی بیش‌تر بودن مقدار NS، عملکرد بهتری داشته است. بنابراین می‌توان نتیجه گرفت که عملکرد دو روش برنامه‌ریزی بیان ژن و یادگیری بر پایه نمونه K نسبت به سایر روش‌های مورد استفاده در پژوهش حاضر بهتر بوده است. نتایج دو مطالعه زیر به دلیل برتری روش‌های برنامه‌ریزی بیان ژن و K نزدیک‌ترین همسایگی با نتایج به‌دست‌آمده از این پژوهش همخوانی دارد.

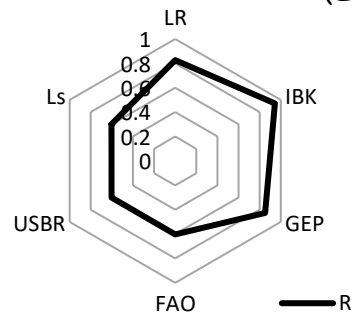
نتایج پژوهش (Emamgholizadeh et al. (2016 نشان داد که روش برنامه‌ریزی بیان ژن نسبت به روش منحنی سنجه عملکرد بهتری داشته و میزان خطای RMSE و MAE را به ترتیب ۹۱ و ۹۴٪ برای ایستگاه هیدرومتری جلویی و برای ایستگاه هیدرومتری پای پل ۶۰ و ۷۱٪ کاهش داده است. (Fathzadeh et al. (2017 به بررسی کارایی پارامترهای ژئومورفومتری و تکنیک‌های داده‌کاوی برای پیش‌بینی بار رسوب معلق در استان گیلان و لرستان پرداختند که از روش K نزدیک‌ترین همسایگی نیز استفاده کردند. مقدار معیارهای آماری همبستگی و نش ساتکلیف به‌دست‌آمده برای برآورد رسوب متوسط برابر با ۰/۶۳ و ۰/۴ بوده است. (Kargar et al (2019 برای مدل‌سازی انتقال رسوب در کانال‌های باز از تکنیک‌های برنامه‌نویسی عصبی فازی (NF) و بیان ژن استفاده کردند. آنها دریافتند که برخلاف پژوهش حاضر که GEP به‌عنوان روش برتر معرفی شد، روش NF را به‌عنوان روشی مؤثر برای پیش‌بینی رسوب معرفی کردند. یکی از دلایل این مغایرت می‌تواند تفاوت شرایط جریان در کانال و رودخانه باشد.

۴- نتیجه‌گیری

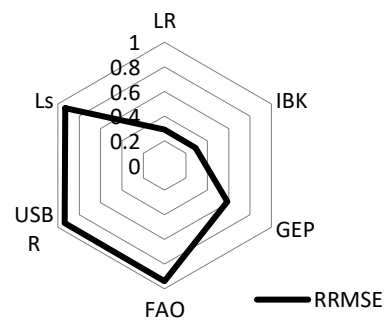
برآورد دبی بار رسوبات معلق رودخانه‌ها به دلیل تأثیرگذاری بر طراحی و مدیریت سازه‌های آبی، در مهندسی آب، هیدرولیک و محیط‌زیست مهم می‌باشد. در پژوهش حاضر با استفاده از داده‌های رسوب و دبی ماهانه، میزان رسوب معلق

نیز دقت قابل قبولی نسبت به روش‌های هیدرولوژیکی داشتند. روش‌های هیدرولوژیکی دارای دقت کم‌تر و خطای بیش‌تری بودند. در نتیجه انطباق کم‌تری بین مقادیر مشاهداتی و محاسباتی وجود خواهد داشت. در شکل (۴) معیارهای آماری مقایسه مدل‌های مورد استفاده در نمودارهای راداری به تصویر کشیده شده‌اند.

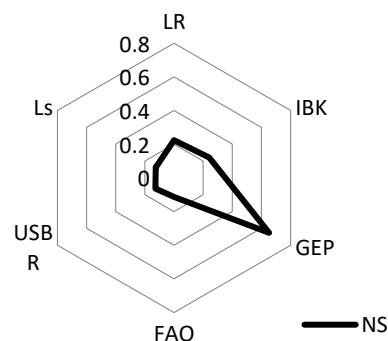
(الف)



(ب)



(ج)



شکل ۴- نمودار راداری معیارهای آماری روش‌های مورد مطالعه:

(الف) R، (ب) RRMSE و (ج) NS

Fig. 4 Radar chart statistical criteria of the studied methods: a) R, b) RRMSE, and c) NS

قسمت (الف)، ۶ ضلعی مربوط به R بین رسوب مشاهداتی و رسوب محاسبه شده، قسمت (ب) RRMSE و قسمت (ج) معیار NS را نشان می‌دهد. در قسمت الف نزدیک بودن رأس‌های ۶ ضلعی رسم شده به رأس‌های ۶ ضلعی بیرونی

¹Neuro Fuzzy

وجود حافظه بلندمدت با اطمینان بیشتری قابلیت استفاده را خواهند داشت.

با توجه به نتایج مذکور، پیشنهاد می‌شود روش‌های مورد استفاده در مطالعه حاضر در حوزه‌های آبخیز با اقلیم‌های مختلف و در مقیاس‌های زمانی متفاوت مورد بررسی قرار گرفته و برای هر اقلیم، بهترین روش مشخص شود. نتایج به‌دست‌آمده از این پژوهش قابل‌تعمیم به کلیه شرایط هیدرولوژیکی و مورفولوژیکی نبوده و صرفاً می‌تواند به‌عنوان افق دید جدیدی در حوضه‌های مشابه با منطقه مورد مطالعه مطرح کند.

دسترسی به داده‌ها

داده‌ها حسب درخواست، از طرف نویسنده مسئول قابل ارسال است.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که، هیچ‌گونه تضاد منافی در رابطه با نویسندگی و یا انتشار این مقاله ندارند.

References

- Aha, D. W., Kibler, D. and Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learn.*, 6, 37-66.
- Anonymous (2021). Geographical area of Qazvin city. Qazvin Municipality. <https://www.qazvin.ir/geographical-area>. [In Persian]
- Aytek, A. and Kisi, O. (2008). A genetic programming approach to suspended sediment modelling. *J. Hydrol.*, 351, 288-298. doi:10.1016/j.jhydrol.2007.12.005.
- Dadashzadeh, F., Mohseni, M., Ahmadi, H. and varvani, J. (2009). Evaluation and developing sediment rating curves models for estimating sediment yield of flood events (Case of study: Ghare Chay Basin). *Watershed Manage. Res.*, 84, 28-35 [In Persian].
- Dehghani, A. A., Malek Mohammadi, M. and Hezarjaribi, A. (2010). Estimation of Suspended Sediment Load in Behesht Abad River by Using Artificial Neural Network. *J. Water Soil Conserv.*, 17(1), 159-168 [In Persian].
- Dehghani, N. and Vafakhah, M. (2013). Comparison of daily suspended sediment load estimations by sediment rating curve and neural network models (Case Study: Ghazaghli Station in Golestan Province). *J. Water Soil Conserv.*, 20(2), 221-230 [In Persian]
- Emamgholizadeh, S., Karimi, R. and Azhdari, K. (2016). Comparison of conventional methods for estimating suspended sediment load of Karkheh river with gene expression planning method. *Quart. J. Geogra. Develop.*, 45, 121-140. [In Persian]
- Fathzadeh, A., Asadi, M. and Taghizadeh, R. (2017). Optimization of suspended load estimation models using morphological geometry parameters and feature reduction technique. *Iran. Soil Water Res.*, 3, 669-678. DOI:10.22059/ijswr.2017.210038.667483 [In Persian].
- Ferreira, C. (2001). Gene expression programming a new adaptive algorithm for solving problems. *Complex Sys.*, 13(2), 87-129.



- Iadanza, C. and Napolitano, F. (2006). Sediment transport time series in the Tiber River. *Phys. Chem. Earth*, 31, 1212-1227.
- Jalali, V. R. and Homaee, M. (2011). Introducing a nonparametric model using k-nearest neighbor technique for predicting soil bulk density. *J. Water Soil Sci.*, 15(56), 181-191.
- Jones, K. R., Berney, O., Carr, D. P. and Barret, E. C. (1981). *Arid zone hydrology for agricultural development*. FAO Irrig. Drain. Paper, 37, 271-284.
- Kargar, K., Sadeqh Safari, J., Mohammadi, A. and Samadianfard, S. (2019). Sediment transport modeling in open channels using neuro-fuzzy and gene expression programming techniques. *Water Sci. Technol.*, 79(12), 2318-2327. doi:10.2166/wst.2019.229
- Keihani, A. R., Mohammadi, A. and Fathian, H. (2021). Uncertainty analysis of SVM model parameters for estimating suspended and bed sediment load at Sierra station in Karaj by Monte-Carlo simulation method. *Iran. Soil Water Res.*, 52(1), 195-212. Doi: 10.22059/IJSWR.2020.308225.668704. [In Persian]
- Khosravi, K., Mao, L., Kisi, O., Yaseen, Z. and Shahid, Sh. (2018). Quantifying hourly suspended sediment load using data mining models: case study of a Glacierized Andean Catchment in Chile. *J. Hydrol.*, 567, 165-179. Doi: 10.1016/j.jhydrol.2018.10.015.
- Pavanelli, D. and Bigi, A. (2004). Suspended sediment concentration for three apennine monitored basins, particle size distribution and physical parameters. *Agro Environ. Congress, Venice, Italy*, 537 -544.
- Ramezanipour, E., Mosaedi, A. and Mesdaghi, M. (2017). Determination of the best model for estimation of suspended sediment by using statistical error criteria (Case study: some sub-watersheds of Kashaf Roud). *J. Watershed Res.*, 8(15), 112-124. Doi:10.29252/jwmr.8.15.112 [In Persian].
- Rencher, A. C. and Christensen William F. (2012). *Multivariate regression*. (3rd ed.), John Wiley & Sons, p. 19, ISBN 978-1-118-39167-9.
- Rezazadeh, A. and Sattari., M. T. (2017). Comparison of the efficiency of support vector regression methods and k-nearest neighborhood in estimating the amount of suspended sediment load in the river (Case study: Ligvan Chai River). *Iran. J. Nat. Resour.*, 2, 345-358 [In Persian].
- Roushangar, K. and Shahnazi, S. (2019). Evaluating the performance of data-driven methods for prediction of total sediment load in gravel-bed rivers. *Iran. Soil Water Res.*, 50(6), 1467-1477. Doi: 10.22059/ijswr.2019.253848.667867 [In Persian].
- Saghebian, M. (2021). Estimation of sediment suspended load using integrated intelligent methods taking into account model uncertainty. *J. Water Soil*, Doi:10.22067/JSW.2021.68665.1021 [In Persian].
- Shadkani, S., Abbaspour, A., Samadianfard, S., Hashemi, S., Mosavi, A. and Band, S.S. (2020). Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted trees for predicting daily suspended sediment load: the case study of the Mississippi River. *Int. J. Sediment Res.*, 36(4), 512-523. Doi.org/10.1016/j.ijsrc.2020.10.001.
- Shirzad, A., Soltani, F., and Zare Abyaneh, H. (2008). Simulation of scouring in accordance with energy-depleting structures using k-nearest closest neighbor (KNN) algorithm and non-fuzzy adaptive inference system (ANFIS). *First International Water Crisis Summit, University of Zabol* [In Persian]
- Warrick, A. W. and Nielsen, D. R. (1980). Spatial variability of soil physical properties in the field. In: *Encycl. Agrophys.*, 319-344. Doi: 10.1007/978-90-481-3585-1_163
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng A., Liu, B. and Philip, S. Y. (2008). Top 10 algorithms in data mining. *Know. Inform. Syst.*, 14, 1-37. Doi.org/10.1007/s10115-007-0114-2.
- Youssefvand, F. (2004). Suggestion of a method for estimation of suspended load in rivers (case study: Ghresoo river). M. Sc. Dissertation, Sari Agricultural Sciences and Natural Resources University. 138 pp. [In Persian]
- Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S. and Yaseen, Z.



(2020). On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in

Puerto Rico. *J. Hydrol.*, 585, 124759. Doi.org/10.1016/j.jhydrol.2020.124759.

How to cite this paper:

Javidan, S., Sattari, M. T., Karimzadeh, P. and Mehrabi, A. (2022). Performance analysis of hydrological and data-based models in estimation of suspended sediment rate. *Environ. Water Eng.*, 8(2), 468-480. DOI: 10.22034/JEWE.2021.281577.1552

