Performance assessment of Kernel-Based methods in estimation of suspended sediment loads (Case study: Maragheh SofiChay River)

Ali Rezazadeh Joudi[*]

Young Researchers and Elite Club, Maragheh Branch, Islamic Azad University, Maragheh, Iran

Mohammad-Taghi Sattari

Faculty Member, Department of Water Engineering, Agriculture Faculty, University of Tabriz, Iran

Extended Abstract

Introduction

Because of the importance of sediment transport in the efficient use of water resources and dam designing, estimation of the sediment load in rivers has been an essential interest to the engineers from long time ago. This leads to the design of various methods such as different empirical equations for solving the sediment transport. The error in most traditional experimental methods is common due to complexities in how this process works out and the vast amount of factors that cause this phenomenon. Therefore, achieving a suitable method that can accurately estimate the amount of sediment is very essential. In this study, the suspended sediment loads of SofiChay river have been estimated by modern data mining methods including Gaussian process and support vector machines that use the kernel functions that have a high ability to solve nonlinear problems. The results that obtained were compared with experimental methods such as sediment rating curve and seasonal method.

Materials and Methods

The study area

Sofi Chay catchment is up to 311 Km^3 in area and has been located in the south part of East Azerbaijan province and the northern city of Maragheh. Sofi Chay river is located within the geographical coordinates 37^0 '15 "2 to 37^0 '45 3" north latitude and 45^0 '56 " 31 to 46^0 '25 "5 east longitude.

Gaussian process regression

Gaussian processes are a fruitful way of defining prior distributions for flexible regression and

[*] E-mail: alijoudi66@gmail.com   Tel: +98 9397902966

classification models in which the regression or class probability functions are not limited to simple parametric forms. One attraction of Gaussian processes is the variety of covariance functions one can choose from. These lead to functions with different degrees of smoothness or different sorts of additive structures. When such a function defines the average response in a regression model with Gaussian errors, we can use matrix calculations to deduce that it is possible for data sets with more than a thousand samples. Gaussian processes in statistical modeling are very important because they are normal characteristics. Gaussian processes and related methods have been used in various contexts for many years. Despite this past usage, and despite the fundamental simplicity of the idea, Gaussian process models appear to have been little appreciated by most Bayesians. I speculate that this could be partly due to confusion between the properties one expects of the true function being modeled and those of the best predictor for this unknown function.

Support Vector Regression (SVR)
SVRs are a subset of SVMs that are particular learning systems that use a linear high dimensional hypothesis space called feature space. These systems are trained using a learning algorithm based on optimization theory. This method was introduced by Vapnik in 1995. SVMs have been employed for regression estimation, so called support vector regression (SVR), in which the real value functions are estimated. In this case, the aim of learning process is to find a function f(x) as an approximation of the value y(x) with minimum risk, and only based on the available independent and identically distributed data. Often in complex nonlinear problems, the original input space (predictor variable) is non-linearly related to the predicted variable (lateral spread displacement).

Results and Discussions
In this study, after required data related to the Sofi ChayRiver were collected, these data were examined by the standard normal homogeneity tests such as the Buishand range, Pettitt and Von Neumann ratio; and after refining the data, the drawing of sediment rating curve was developed. Then, the amount of sediment discharge of the river Sofi Chay was estimated using Gaussian process regression, support vector regression, sediment rating curve and seasonal methods. To achieve optimum results by the used data mining techniques, various scenarios including different types of kernel functions and different intervals of hyper parameters of kernel functions were defined. When Gaussian process regression, along with radial basis function kernel (Gaussian noise (ε) equal to 0.01 and gamma (Y) equal to 0.5), were used to estimate sediment discharge rate of the river Sofi Chay, it was observed that this method by presenting statistical indicators (correlation coefficient (R) equal to 0.977, Nash-Sutcliffe coefficient (NS) equal to 0.794, mean absolute error (MAE) equal to 77.4278 (tons/day) and root mean square error (RMSE) equal to 698.7455 (tons/day)), have the highest accuracy and lowest error among the methods investigated in this study. Also the both investigated data mining methods have far greater efficiency and accuracy in this area.

Conclusion
In this study, the amount of suspended sediment load was estimated using traditional methods

such as sediment rating curve and seasonal method in comparison with modern data mining methods based on kernel functions such as Gaussian process regression and support vector regression. The results indicated that seasonal method has better performance in this case rather than sediment rating curve. The comprehensive results show that both modern data mining methods examined in this study outperform rather than traditional methods. Among the Gaussian process regression and support vector regression results, we observed the higher ability of Gaussian process regression method with using radial basis function as a kernel function. Generally, use of Gaussian process regression method suggested in similar cases.

Keywords: Gaussian process regression, sediment rating curve, Sofi Chay River, support vector regression, suspended sediment load.