Accepted Manuscript

Daily Suspended Sediment Load Prediction Using Artificial Neural Networks and Support Vector Machines Machine

E. Kakaei Lafdani, A. Moghaddam Nia, A. Ahmadi

PII: S0022-1694(12)01027-X
DOI: http://dx.doi.org/10.1016/j.jhydrol.2012.11.048
Reference: HYDROL 18613

To appear in: Journal of Hydrology

Received Date: 23 February 2012
Revised Date: 18 September 2012
Accepted Date: 20 November 2012

Please cite this article as: Kakaei Lafdani, E., Moghaddam Nia, A., Ahmadi, A., Daily Suspended Sediment Load Prediction Using Artificial Neural Networks and Support Vector Machines Machine, Journal of Hydrology (2012), doi: http://dx.doi.org/10.1016/j.jhydrol.2012.11.048

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Daily Suspended Sediment Load Prediction Using Artificial Neural Networks and Support Vector Machines Machine

1Kakaei Lafdani, E., 2Moghaddam Nia, A. and 3Ahmadi, A.

1MSc Graduate of Watershed Management, Department of Range and Watershed Management, University of Zabol, Iran
2Associate Professor of Hydrology, Department of Range and Watershed Management, Faculty of Natural Resources, University of Tehran, Karaj, Iran
3Assistant Professor of Civil Engineering, Department of Civil Engineering, Isfahan University of Technology, Isfahan, Iran

Abstract:

Prediction of suspended sediment load is the most important factor in water resources management projects. In recent years decades, development of artificial intelligent intelligence techniques have been applied to predict suspended sediment load phenomenon due to its capability in modeling non-linear dynamic systems; as a predictor for hydrological phenomenon, has created a great change in predictions. This paper investigates the abilities of Support Vectors Machine (SVM) and Artificial Neural Network (ANN) models to predict daily suspended sediment load (SSL) in Doiraj River, located in Ilam province the west part of Iran. An 11-years period data (1994-2004) was applied for predicting suspended sediment load SSL. Streamflow and rainfall were used as the model inputs and SSL as the model output. The best input of SVM and ANN models was identified using combination of Gamma test and Genetic Algorithm (GT-GA). Its results accuracy was compared with the results of conventional correlation coefficient analysis between input and output variables and the best combination was identified. Also, the present study explores Gamma Test for the first time gamma test to identifying length of the training dataset and the best input combination of SVM and ANN models for suspended sediment load prediction identify the length of the training dataset. Finally, in order to predict SSL, we used the nu-SVR (using the four kernels including linear, polynomial, sigmoid and Radial Basis Function (RBF)) and ANN models (based on BFGS algorithm and Conjugate algorithm). The reliability of SVM and ANN models were evaluated based on performance criteria such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Efficiency Index (EI) and correlation coefficient ($R^2$). The obtained results show that the nu-SVR model with RBF kernel has the more capability in

* Corresponding Author: Alireza Moghaddam Nia, Associate Professor of Hydrology, Department of Range and Watershed Management, University of Tehran, Karaj, Iran . E-mail: ali.moghaddammia@live.com
prediction of suspended sediment load than the other models. ANN models and nu-SVR model using gamma test for input selection has better performance than regression combination. Also, the performance BFGS-ANN model were better than other models with RMSE value and $R^2$ equal to 0.34 (ton) and 0.99, respectively. The nu-SVR model with RBF kernel has more capability in prediction of SSL than the other kernels (RMSE=0.96 (ton) and $R^2=0.98$). In addition, the results show M-test can be used as a new method to determine the number of required data for network training for creating a smooth model by nu-SVR and ANN models.

**Keywords:** suspended sediment load Suspended Sediment Load; support vector machine Support Vector Machine; artificial neural network Artificial Neural Network; Gamma Test; Doiraj River basin Basin.

**INTRODUCTION**

The prediction of sediment load carried by a river, which influence influences hydraulic river structures, is very important for many studies on river engineering and dam engineering (Aytek and Kisi, 2008; Yang et al., 2009; Kisi et al., 2009; Kisi, 2010; Kisi et al., 2012). From the prospective of water resources management, sediment transport by river is raised as one of the biggest issues of surface water resources (Gericke and Venohr, 2012) exploitation in the world; that is the reason predicting SSL of river can be an appropriate index to estimate future conditions of water resources management. There are a number of some factors and parameters affecting on the SSL. In most studies, rainfall and streamflow have been reported as the main factors which affect influencing on the suspended load (Jie and Yu, 2011). Until now, many techniques have been proposed for prediction of suspended sediment load SSL. However, due to the complexity of SSL transportation (Nourani, 2009) mechanism and non-linear behavior (Rajaee, 2011) of effective hydrologic parameters, such techniques do not have enough precision.

For the two past In recent decades, development of artificial intelligence techniques (Cobaner et al., 2009) as a predictor for hydrological phenomenon has created a great change in predictions. Black box models like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have been successfully applied (Azamathulla et al. 2010, Kisi, 2005) as powerful tools in predicting of suspended sediment load successfully have used in various fields including water resources (Kisi, 2009). Recent experiences concerning hydrological
forecasts have shown that artificial neural networks and support vector machines can be a proper alternative to predict SSL carried by the river (Partal and Cigizoglu, 2008; Kisi, 2004; Kisi, 2005; Kisi et al., 2008; Nourani, 2009; Zhu et al., 2007; Rajaee et al., 2009; Cigizoglu, 2004). Many studies using Artificial Neural Networks and Support Vector Machines have conducted by researchers on suspended sediment load modeling, have been done about SSL modeling using Artificial Neural Networks and Support Vector Machines. Alp and Cigizoglu (2005), Nagy et al. (2002) and Agarwal et al. (2006) Alp and Cigizoglu, (2005); Nagy et al. (2002), Lin and Namin, (2005) and Agarwal et al. (2006) used ANN model in order to to prediction of sediment load concentration predict sediment load. Mustafa et al. (2011) have compared the performance of two artificial neural networks i.e. Multi Layer Feed Forward (MLFF) network and Radial Basis Function network in order to predict the suspended sediment discharge. The obtained results show that the RBF network model has better performance than the MLFF network model in suspended sediment discharge prediction. Bhattacharya et al. (2007) and Azamathulla et al. (2010) used machine learning approach to predict load transport. Jie and Yu. (2009) (2011) estimated suspended sediment load by using ANN and SVM models in Kaoping River basin located in southern Taiwan. The result shows that SVM outperforms the ANN model. Çimen. 2008 used support vector machine with Gaussian radial basis function kernel in order to estimate suspended sediment concentration for two rivers located in the USA. Results indicated SVM model can estimate sediment without producing negative sediment values.

One of the challenging issues for modeling natural phenomena by using intelligent techniques is model input selection. The questions on which inputs should be used and how long the training data should be for model development have been hard to solve in practice (Ahmadi et al., 2009). Pre-process input parameters in order to select an appropriate combination among them is of one the complex processes in non-linear system modeling. Despite multiple studies for phenomena modeling of suspended load rivers using intelligent methods, there are still a few fundamental questions concerning this subject including which inputs, what combination of input data and how long the training data should be used for model development (Ahmadi et al., 2009). Reviewing resources have indicated that in most of the references for modeling, the amount of suspended load of rivers using intelligent techniques have been investigated without using for input variables pre-processing methods, using trial and error among the limited combinations and then the best combination will be chosen. Nevertheless, lack of a
regular method to find the best combination among various input variables is felt. Also, one of the problems during the training process of black box modeling is over-fitting. Provided that the number of network parameters is extremely lower than the number of training data, the possibility of over-fitting is close to zero. If much data are collected and the size of training collection increased, no technique is needed to prevent from over-fitting. Unfortunately, determination the large size of network is difficult before solving this problem. Specifying the number of significant data is one of the objectives of this study to establish a smooth model for prediction of SSL using support vector machine and ANN models. The Gamma Test (GT) is a non-linear modeling and analysis tool, which allows us to examine the input/output in a non-linear dynamic system. The GT was briefly reported in konucar (1997) and Stefansson et al. (1997) and later discussed and used to determine the best combination of model inputs and the length of the training dataset by Chuzhanova et al. (1998); de Oliveira (1999); Tsui et al. (2002) and Jones et al. (2002). Remesan et al. (2008); Jamalizadeh et al. (2008); Ahmadi et al. (2009); Moghaddamnia et al. (2009); Remesan et al. (2009) and Noori et al. (2011). Lately, Gamma Test (GT) has been used by hydrologists as a new model to determine the number of required models for network training to create a smooth model (Moghaddamnia et al., 2009). The GT is a non-linear modeling and analysis tool. GT predicts the minimum achievable modeling error before the modeling. GT first was reported by Stefansson et al. (1997) and Končar. (1997) and later was discussed by many scientists and used to determine the best input combination (Chuzhanova et al., 1998; de Oliveira, 1999; Tsui et al., 2002; Jones et al., 2002; Remesan et al., 2008; Jamalizadeh et al., 2008; Moghaddamnia et al., 2009; Noori et al., 2011). In order to rainfall-runoff modeling, Remesan et al. (2009) used GT to identify data length and to confirm the GT result performed data partitioning approach. The result illustrated that GT may represent much better estimate of data length than the data partitioning approach. Nowadays, Genetic Algorithm (GA) utilize in many engineering fields such as optimization problems. GA uses Darwin's natural selection of genetic principles to find an optimum formula to predict or match the pattern. In many researches, GA has been applied as a powerful tool for optimization and modeling (James et al., 2002; Altunkaynak, 2009). Reza pour et al. (2011) in order to identify the relation between sediment load and stream flow discharge, used GA and compared performance of GA with Sediment Rating Curve (SRC) technique. Results show the GA model has better performance than SRC technique.
The main objective of this article is to assess the performance of two data-driven techniques of support vector machines and artificial neural networks to prediction of daily SSL based on the abilities of GT for dealing with tasks of selecting the best combination of model inputs and number of data points required for model training phase.

In this study, a duration period was considered from 1965 to 2005. For more accuracy in prediction, it was necessary to determine the most effective input variables. Thus, GT was utilized for selecting the most effective predicting variables. In order to decrease the complexity of the model and improving its accuracy, it was tried to determine the best combination among the mentioned variables in the previous section. Thus, GT-GA method was used and its accuracy was compared with the results of GT and conventional correlation coefficient analysis between input and output variables. The results of evaluating different input combinations provided through two methods; and the best combination was selected based on the results of test models. The suitable number of the data points required for training the model was determined using the M-test and to confirm the M-test result performed data partitioning approach. In all the cases, the models assessment criteria were Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Efficiency Index (EI) and correlation ($R^2$) between actual and predicted values. Then, support vector machine and ANN models were used for predicting SSL.

MATERIALS AND METHODSMaterials and Methodology

Gamma Test, V-ratio and M-test

The Gamma test was firstly reported by Končar (1997) and Agalbjörn, et al. (1997), and later enhanced and discussed in detail by many researchers (Chuzhanova et al. (1998); de Oliveira (1999); Tsui (1999); Durrant (2001); Tsui et al. (2002) and Jones et al. (2002)). Only a brief introduction on the Gamma Test is given here and the interested readers should consult the aforementioned papers for further details. Gamma test estimates the minimum mean square error which is achievable in continuous non-linear models with unseen data. Gamma test first was reported by Stefansson et al. (1997), končar. (1997) and Agalbjörn et al. (1997) and later was used and discussed by many scientists (Chuzhanova et al., 1998; de Oliveira, 1999; Tsui, 1999; Durrant, 2001; Tsui et al., 2002; Corcoran et al., 2003; Jones et al., 2002). This part provides a brief explanation about gamma test. The basic idea is quite distinct from the earlier attempts with nonlinear analysis. Suppose we have a set of observed data in the following form:
\[(x_i, y_i), 1 \leq i \leq M \]  \hspace{1cm} (1)

where the input vectors \(x_i \in \mathbb{R}^m\) are vectors confined to some closed bounded set \(C \subseteq \mathbb{R}^m\) and, without loss of generality, the corresponding outputs \(y_i \in \mathbb{R}\) are scalars. The vectors \(x\) contain predicatively useful factors influencing the output \(y\). The only assumption made is that the underlying relationship of the system is of the following form:

\[ y = f(x_1, \ldots, x_m) + r \]  \hspace{1cm} (2)

where \(f\) denotes a smooth function and \(r\) is a random variable that represents noise. Without loss of generality it can be assumed that the mean of the \(r\’s\) distribution is zero (since any constant bias can be subsumed into the unknown function \(f\)) and that the variance of the noise \(\text{Var}(r)\) is bounded. The domain of a possible model is now restricted to the class of smooth functions which have bounded first partial derivatives. The Gamma statistic \(\Gamma\) is an estimate of the model’s output variance that cannot be accounted for by a smooth data model. The Gamma Test is based on \(N[i,k]\), which are the \(k\)th \((1 \leq k \leq p)\) nearest neighbors for each vector \(x_i, (1 \leq k \leq p)\). Specifically, the Gamma Test is derived from the Delta function of the input vectors:

\[ \delta_M(k) = \frac{1}{M} \sum_{i=1}^{M} \left| x_{N(i,k)} - x_i \right|^2 \quad (1 \leq k \leq p) \]  \hspace{1cm} (3)

where \(\left| \ldots \right|\) denotes Euclidean distance, and the corresponding Gamma function of the output values,

\[ \gamma_M(k) = \frac{1}{2M} \sum_{i=1}^{M} \left| y_{N(i,k)} - y_i \right|^2 \quad (1 \leq k \leq p) \]  \hspace{1cm} (4)

where \(y_{N(i,k)}\) is the corresponding \(y\)-value for the \(k\)th nearest neighbor of \(x_i\) in Eq. (3). In order to compute \(\Gamma\) a least squares regression line is constructed for the \(p\) points \((\delta_M(k), \gamma_M(k))\).

\[ \gamma = A\delta + \Gamma \]  \hspace{1cm} (5)

The intercept on the vertical axis \((\delta = 0)\) is the \(\Gamma\) value, as can be shown,

\[ \gamma_M(k) \rightarrow \text{Var}(r) \text{ in probability as } \delta_M(k) \rightarrow 0 \]  \hspace{1cm} (6)

Calculating the regression line gradient can also provide helpful information on the complexity of the system under investigation. A formal mathematical justification of the method can be found in Evans and Jones (2002). The graphical output of this regression line (Eq.5) provides very useful information. First, it is remarkable that the vertical intercept \(\Gamma\) of
the y (or Gamma) axis offers an estimate of the best MSE achievable utilizing a modeling technique for unknown smooth functions of continuous variables (Evans and Jones, 2002). Second, the gradient offers an indication of model’s complexity (a steeper gradient indicates a model of greater complexity). The Gamma test is a non-parametric method and the results apply regardless of the particular techniques used to subsequently build a model of $f$. We can standardize the result by considering another term V-ratio, which returns a scale invariant noise estimate between zero and one. The V-ratio is be defined as

$$V_{ratio} = \frac{\Gamma}{\sigma^2(y)}$$

where, $\sigma^2(y)$ is the variance of output $y$, which allows a judgment to be formed independent of the output range as to how well the output can be modeled by a smooth function. A V-ratio close to zero indicates that there is a high degree of predictability of the given output $y$. We can also determine the reliability of $\Gamma$ statistic by running a series of Gamma test for increasing $M$, to establish the size of data set required to produce a stable asymptote. This is known as M-test. M-test result would help us to avoid the wasteful attempts of fitting the model beyond the stage where the MSE on the training data is smaller than $\text{Var}(r)$, which may lead to over fitting. The M-test also helps us to decide how much data we require to build a model with a mean squared error which approximates the estimated noise variance. In practice, the Gamma test can be achieved through winGammaTM software implementation (Durrant, 2001). Corcoran, et al. (2003), applied the Gamma Test as a method for crime incident forecasting by focusing upon geographical areas of concern that transcend traditional policing boundaries. The authors believed this technique was very effective and could be potentially used for water management including flood prediction and other hydrological nonlinear modeling (Corcoran et al. 2003).

The main idea was somewhat different from previous efforts for the nonlinear analysis. Suppose $X_i$ and $X_j$ are close to each other; therefore, $y_j$ and $y_i$ should also be close to each other. In gamma test, it is tried to make this view qualitative through mean distance between the nearest neighbor bounded set of $X_i$ and $X_j$ and mean length between the corresponding output points of $y_j$ and $y_i$ and achieve estimation for error value. Suppose there are a series of observations as the following form:

$$((x_1, ..., x_m), y) = (X, y)$$

(1)
In which $X = (x_1, \ldots, x_m)$ is the input vector at the range of $C \in \mathbb{R}^m$, and $y$ is the output vector.

The only assumption of this method is that the following equation is established between the systems:

$$y = f(x_1, \ldots, x_m) + r$$

(2)

In which $r$ is the random variable that illustrates noise of equation and must be determined. Without losing the generality of the function, it can be assumed that mean of this random variable is zero (as any constant bias might be subsumed into an unknown function) and its variance is bounded. Gamma test is based on $N[i, k]$ which includes a set of nearest neighbors from $k(1 \leq k \leq p)$ for each vector $X_i(1 \leq i \leq M)$. Delta function calculates the mean square of $k^{th}$ distance from the neighbor:

$$\delta_m(k) = \frac{1}{M} \sum_{i=1}^{M} |X_{N[i,k]} - X_i|^2$$

(3)

In which $| |$ indicates Euclidean distance, corresponding gamma function is as:

$$\gamma_m(k) = \frac{1}{2M} \sum_{i=1}^{M} (y_{N[i,k]} - y_i)^2$$

(4)

in which $y_{N[i,k]}$, $y$ is the corresponding value for the $K^{th}$ neighbor of $X_i$ in (3) equation. In order to calculate $\Gamma$, a linear regression line is built from P point on values of $\delta_m(k)$ and $\gamma_m(k)$.

$$\gamma = A \delta + \Gamma$$

(5)

Intercept of the vertical axis ($\delta = 0$) is the value of $\Gamma$ and $\gamma_m(k)$ is equal to variance errors. Drawing the regression line can provide useful data about complexity grade of the model. Vertical intercept of estimated line provides the best obtainable mean square error (Evans and Jones, 2002). Furthermore, the gradient of the line provides complexity of the model (high greater complexity models have steeper gradient). Gamma is a conceptual model and its results have nothing to do with used techniques for a model of $f$ function. These results can be standardized by considering term of $V_{ratio}$ which is defined as follows:
\[ V_{ratio} = \frac{\Gamma}{\sigma^2(y)} \]  

In which \( \sigma^2(y) \) is the \( y \) output variance that provides the power of judgment to be formed independent from the output range. When \( V_{ratio} \) is close to zero, there would be a higher degree of predictability of the required output of the model. The reliability of the gamma test can be done through implementation of it by increasing \( M \). This stage which is called \( M \), is used for reviewing the required number of data for producing an asymptote in gamma model in terms of \( M \) changes. These results help for the stage of education model and prevent from over-fitting of the network. Genetic algorithms act according to genetic principals; it makes some combinations and work on them to obtain the best inputs combination. These algorithms often are proper alternative for regression-based prediction techniques which provide a space for all the used combinations in genetic algorithm for obtaining the best combination. Practically, the Gamma-test, M-test and genetic algorithm can be obtained through winGamma™ Software (Durrant, 2001). For better determining the input combination, the winGamma software allows us 5 options:

1- Full embedding
2- genetic algorithm
3- Hill climbing
4- Sequential embedding
5- Increasing embedding

In this study we have utilized the Genetic Algorithm option for better determination of the input combination. Variables which can be modified for this purpose are: Population Size, Mutation Rate, Crossover Rate, Gradient Fitness, Intercept Fitness, Length Fitness, and Run Time. For managing this search, the parameters mentioned below can be utilized:

**Population Size** (default values=100): the utilized masks' population size.

**Mutation Rate** (default values=0.01): single bit mutation probability during reproduction.

**Crossover Rate** (default values=0.5): crossover occurrence probability during reproduction.
**Gradient Fitness** (default values=0.1): The GA fitness function weighting used for masks with low gradient in the Gamma Test. More simplicity can be achieved by increasing this weighting.

**Intercept Fitness** (default values=0.8): The GA fitness function weighting used for masks that have a low absolute value of the Gamma statistic. More accuracy can be achieved by increasing this weighting.

**Length Fitness** (default values=0.1): The GA fitness function weighting used for masks with a certain number of ‘1’s. Masks with fewer ‘1’s are more likely to be selected and simpler models can be achieved by increasing this weighting.

**Run Time** (default values=5 minutes): Selected GA performing (approximate) maximum time.

If a longer run time is allowed, the population can be set to be larger which can lead to a better final fitness of the best mask available. Run times as long as several hours are required by the GA for long masks (i.e. many inputs) and large data sets. For Gamma test, M test, and Genetic Algorithm Analysis, the winGamma™ software version 1.97 was used in this study.

**Support Vector Machines**

The foundation of the subject of Support Vector Machines (SVMs) has been developed principally by Vapnik and his collaborators (Vapnik, 1995; Vapnik, 1998). Their formulation embodies the Structural Risk Minimization (SRM) principle, which has been shown to be superior to the more traditional Empirical Risk Minimization (ERM) principle employed by many of the other modeling techniques (Osuna et al, 1997; Gunn, 1998). It is this difference that provides SVM with a greater ability to generalize, which is the goal in statistical learning. SVM has been proved to be effective in classification by many researchers in many different fields such as electric and electrical engineering, civil engineering, mechanical engineering, medical, financial and others (Vapnik, 1998).

**Statistical Learning Theory**

In statistical learning theory (Vapnik, 1995; Vapnik, 1998), the problem of learning an input-output relationship from a data set is generally viewed as that of choosing, from the given set of functions $f(x, \alpha), \alpha \in \Lambda$ (where $x \in \mathbb{R}^n$ is a random vector drawn independently from a...
fixed but unknown probability distribution function $P(x)$ and $\Lambda$ is a set of parameters), the one that best approximates the output value $y$ to every input vector $x$, according to a conditional distribution function $P(y \mid x)$, also fixed but unknown. The selection of the desired function is based on a training set of $l$ independent and identically distributed observations $(x_1, y_1), \ldots, (x_l, y_l)$ drawn according to $P(x, y) = P(x) P(y \mid x)$. If one considers the expected value of the loss due to classification or estimation errors, given by the risk functional,

$$R(\alpha) = \int L(y, f(x, \alpha))dP(x, y)$$  \tag{8}$$

Where $L(y, f(x, \alpha))$ is the discrepancy between the measured response $y$ and the response $f(x, \alpha)$ provided by the learning machine, the goal is to find the function $f(x, \alpha_0)$ that minimizes this risk functional $R(\alpha)$ in the situation where the only available information is the training set.

**Support Vector Regression**

SVMs can also be applied to regression problems by the introduction of an alternative loss function that is modified to include a distance measure (Smola, 1996). Let the observed variable $y$ be has real value, and let $f(x, \alpha), \alpha \in \Lambda$ be a set of real functions that contains the regression function $f(x, \alpha_0)$. Considering the problem of approximating the set of data, $\{(x_1, y_1), \ldots, (x_l, y_l), x \in \mathbb{R}^n, y \in \mathbb{R}\}$ with a linear function, $f(x, \alpha) = (w \cdot x) + b$, the optimal regression function is given by minimizing the empirical risk,

$$R_{\text{emp}}(w, b) = \frac{1}{l} \sum_{i=1}^{l} |y_i - f(x_i, \alpha)|$$  \tag{9}$$

With the most general loss function with $\varepsilon$-insensitive zone described as,

$$|y - f(x, \alpha)|_\varepsilon = \begin{cases} 
\varepsilon & \text{if } |y - f(x, \alpha)| \leq \varepsilon; \\
|y - f(x, \alpha)| & \text{otherwise} 
\end{cases}$$  \tag{10}$$

The objective is now to find a function $f(x, \alpha)$ that has at most a deviation of $\varepsilon$ from the actual observed targets $y_i$ for all the training data, and at the same time is as flat as possible. This is equivalent to minimizing the functional,

$$\Phi(w, \xi^*, \zeta) = \|w\|^2/2 + C \left( \sum \xi^* + \sum \zeta \right)$$  \tag{11}$$

Where $C$ is a pre-specified value and $\xi^*$, $\zeta$ are slack variables representing upper and lower constraints on the outputs of the system (Fig. 1), as follows,
\[ y_i - ((w \cdot x_i) + b) \leq \varepsilon + \xi_i \quad i = 1, 2, \ldots, l \]
\[ ((w \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, \ldots, l \]  
(12)

\[ \xi_i^* \geq 0 \quad \text{and} \quad \xi_i \geq 0 \]

Now the Lagrange function is constructed from both the objective function and the corresponding constraints by introducing a dual set of variables, as follows,
\[
L = \|w\|^2 / 2 + C \left( \sum_{i=1}^{l} (\xi_i + \xi_i^*) \right) - \sum_{i=1}^{l} \alpha_i [\varepsilon + \xi_i - y_i + (w \cdot x_i) + b] - \sum_{i=1}^{l} \alpha_i^* [\varepsilon + \xi_i^* + y_i - (w \cdot x_i) - b] - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*)
\]
(13)

It follows from the saddle point condition that the partial derivatives of \( L \) with respect to the primary variables (\( w, b, \xi_i, \xi_i^* \)) have to vanish for optimality. Substituting the results of this derivation into Equation (4) yields the dual optimisation problem,
\[
W(\alpha^*, \alpha) = -\varepsilon \sum_{i=1}^{l} (\alpha_i^*, \alpha_i)x_i + \sum_{i=1}^{l} y_i(\alpha_i^*, \alpha_i) - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\alpha_i^*, \alpha_j)(\alpha_j^*, \alpha_j)(x_i, x_j)
\]
(14)

That has to be maximized subject to the constraints:
\[
\sum_{i=1}^{l} \alpha_i^* = \sum_{i=1}^{l} \alpha_i; \quad 0 \leq \alpha_i^* \leq C \quad \text{and} \quad 0 \leq \alpha_i \leq C \quad \text{for} \quad i = 1, 2, \ldots, l
\]

Once the coefficients \( \alpha_i^* \)and \( \alpha_i \) are determined from Equation (14), the desired vectors can now be found as,
\[
w_0 = \sum_{\text{supportvectors}} (\alpha_i^*, \alpha_i)x_i \quad \text{and therefore} \quad f(x) = \sum_{\text{supportvectors}} (\alpha_i^*, \alpha_i)(x_i, x) + b_0
\]
(15)

Where \( b_0 = -(1/2)w_0, [x_i + x_i] \).

Once again, when linear regression is not appropriate, as in the case of most engineering applications, a non-linear mapping kernel \( K \) is used to map the data into a higher-dimensional feature space where linear regression is performed. Once the optimum values \( \alpha_i0 \) and \( \alpha_i0^* \) are obtained, then the regression function is given by:
\[
f(x) = w_0 \cdot x + b_0
\]
where
\[
w_0 \cdot x = \sum_{\text{supportvectors}} (\alpha_i^0 - \alpha_i)(x_i, x) \quad \text{and} \quad b_0 = -(1/2) \sum_{\text{supportvectors}} (\alpha_i^0 - \alpha_i)(K(x_i, x_i) + K(x_i, x_i))
\]
(16)

Recently, it has been extended to the domain of regression problems (Kecman, 2001). Dibike et al. (2001) presented some results showing that Radial Basis Function (RBF) is the best kernel function to be used in SVM models.
In recent years, modern tools regarding artificial intelligence called a support vector machine has had many applications in learning method machines (Cristianini et al., 1999). This method successfully has been used in information categorization and lately in regression problems. Mathematically, support vector machine is placed in classification and regression algorithms range which is formulated using the principals of statistical learning theory by Vapnik (1995). These model first was used for water resource management by (Sivapragasam et al., 2001; Dibike et al., 2001; Han and Yang, 2001) and its new model are called reference vector machines used by Han et al. 2002. SVMs can be applied to regression problems (Smola, 1996; kecman, 2001). A short explanation is given for these models as below. The main relationship for statistical learning process is as follows:

\[ y = f( X ) = \sum_{i=1}^{M} w_i \varphi ( X ) = W \varphi ( X ) \]  \hspace{1cm} (7)

Where the output of the model is the part of linear \( M \) and the converter is shown by the nonlinear model by \( \varphi ( ) \). This equation is converted as the below for using support vector machine model:

\[ y = f( X ) = \left\{ \sum_{i=1}^{N} w_i K( X_i, X ) \right\} - b \]  \hspace{1cm} (8)

Here \( K \) is the Kernel function, \( w_i \) and \( b \) are parameters of the model, \( N \) the total number of learning patterns and \( X_i \) data vector for network learning and \( X \) is an independent vector. The parameters of the model are determined with maximizing the objective of function. The general structure of these models is shown in Fig. 1.

Support vector machines use some of the specific Kernel functions which convert the input vector as the input data from nonlinear function in this model. Selection of an appropriate Kernel function is a complex stage and often is used from standard kernel function.

The aim of this linear regression model (equation 9) is to find the linear function which is the best interpolation for training point.

\[ y = f( x ) = \langle w . x \rangle + b \]  \hspace{1cm} (9)
According to the method by minimizing the sum of squares of obtained data \( \langle w, b \rangle \) parameters are determined (Cristianini and Shawe-Taylor, 2000).

\[
\sum_{i=1}^{l} (y_i - \langle w, x \rangle - b)^2
\]

(10)

In order to consider the error between actual values and modeling values, value \( \varepsilon \) is entering the limitations of the above model.

\[
\begin{align*}
    y_i - \langle w, x \rangle - b &< \varepsilon \\
    y_i - \langle w, x \rangle + b &\leq \varepsilon
\end{align*}
\]

(11)

It can be assumed that a band is placed around the function \( f(x) \) which causes training error for point out of this band and unless the covariate variable is called \( \xi_i \). This covariate variable is for point in zero band and increases exponentially for outside points. This regression method is called \( \varepsilon - SV \) which is the most common modeling method. In this model, the cost function which is shown in Fig. 2 is formulated as the following:

\[
|\xi|^e = |y - f(x)|_e = \begin{cases} 
0 & \text{if } |y - f(x)| \leq \varepsilon \\
|y - f(x)| - \varepsilon & \text{o.w.}
\end{cases}
\]

(12)

Two layers around layer of function \( f(x) \) should be determined in a way to maximize the boundary area which has an inverse relation with smooth Euclidean smooth vector \( \|w\|^2 \).

Therefore, smooth Euclidean smooth vector should be minimized considering to the cost function:

\[
\begin{align*}
    \text{Min} & \quad \frac{1}{2} \|w\|^2 + C \left( \sum_{i} \xi^*_i + \sum_{i} \xi_i \right) \\
    \text{Subject to} & : y_i - \langle w, x \rangle - b \leq \varepsilon + \xi_i \\
    & \langle w, x \rangle + b - y_i \leq \varepsilon + \xi^*_i \\
    & \xi_i, \xi^*_i \geq 0
\end{align*}
\]

(13)

where \( C \) is the cost factor. This equation can be solved using Lagrange multipliers. The obtained Lagrangian equation is as follows:
\[
\text{Min} \quad L = \frac{1}{2}\|v\|^2 + C \left( \sum_{i=1}^{l} \xi_i^* + \sum_{i=1}^{l} \tilde{\xi}_i \right) + \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
- \sum_{i=1}^{l} \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b)
\]

(14)

where \( \eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0 \) are the factors of Lagrange multipliers. Partial derivative of the above equation compared with initial variables \((w, b, \xi_i, \xi_i^*)\) is:

\[
\frac{\partial L}{\partial b} = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 
\]

(15)

\[
\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) x_i = 0 
\]

(16)

\[
\frac{\partial L}{\partial \xi_i^*} = C - (\alpha_i^* - \eta_i^*) = 0 
\]

(17)

where \( \eta_i^*, \xi_i^*, \alpha_i^* \) are corresponding with \( \eta_i, \xi_i, \alpha_i \) and \( \eta_i, \xi_i, \alpha_i \). By replacement of the equations 15 to 17, therefore we have:

\[
\text{Min} \quad -\frac{1}{2} \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*) 
\]

(18)

subject to: \( \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \)

(19)

\( \alpha_i, \alpha_i^* \in [0, C] \)

(20)

after rewriting the equation 16, it can be replaced in the equation.

\[
w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i 
\]

(21)

\[
f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b 
\]

(22)

This developed equation of support vectors is for linear model which is used for non-linear relationships. It is not proper for many hydrological analysis of linear regression for modeling and therefore, it is proper by converting Kernel for putting data in a space with more dimensions and then using the linear regression. Kernel function \( K(x, z) \) is \( \langle \phi(x), \phi(z) \rangle \).

Appropriate selection of Kernel function provides the possibility of using a non-linear
function in inputs space for changing to linear function in characteristics space. There are four standard conversion of Kernel function mostly used in regression and modeling including:

1- Linear Kernel

The simplest Kernel function as the below (Han et al, 2007):

$$K(x, z) = \langle x, z \rangle$$

(23)

2- Polynomial Kernel

Polynomial mapping is a common method for nonlinear modeling:

$$K(x, z) = \langle x, z \rangle^d$$

$$K(x, z) = \left(\langle x, z \rangle + 1\right)^d$$

(24)

Usually the second Kernel is preferred; because it solves the problems of Hessian so as to close zero.

3- Radial Basis Function (RBF)

A function based on Radial Basis Function more similar to Gaussian (bell shape) as below (Han et al, 2007):

$$K(x, z) = \exp\left(-\alpha |x - z|^2 \right)$$

(25)

In traditional methods RBF is used in order to determine subgroup at the center; in general, cluster method for selection of center’s subgroup. One of the features of SVM is that it is an unconditional selector i.e. any participated support vector in Gaussian function can achieve the center of the points. Considering that this feature can bring about a possibility to select the diagonal band function -S-, they use SRM rule (Vapnik, 1995).

4- Sigmoid Kernel

The Sigmoid Kernel function as the below:

$$K(x, z) = \tanh(-\alpha x^T z + c)$$

(26)
In the sigmoid kernel, the slope alpha and constant c should be determined. A common value for alpha is 1/n (n is the data dimension). In this study, modeling of SVM has been done through MATLAB R2008a.

 Artificial neural networks (ANNs)

The theory of ANNs was first proposed in the early 1940s when McCulloch and Pitts developed the first computational representation of a neuron (McCulloch and Pitts, 1943). Later Rosenblatt proposed the idea of perceptrons (Rosenblatt, 1962) in which single layer feedforward networks of McCulloch-Pitts neurons could carry out various computational tasks with the help of weights and training algorithm. The applications of ANNs are based on their ability to mimic the human mental and neural structure to construct a good approximation of functional relationships between past and future values of a time series. The supervised ANN is the most commonly used ANN, in which the input is presented to the network along with the desired output, and the weights are adjusted so that the network attempts to produce the desired output. There are different learning algorithms, and a popular algorithm is the back propagation algorithm, which employs gradient descent and gradient descent with momentum; these are often too slow for practical problems because they require low learning rates for stable learning. Algorithms like conjugate gradient, quasi-Newton, Levenberg-Marquardt (LM), etc., are faster algorithms that all make use of standard numerical optimization techniques. Minsky and Papert (1969) highlighted the weaknesses of single layer perceptrons as their ability to solve linearly separable problems only. In practice nowadays, it is usually most effective to use two hidden layers (Jones, 2004). In this study, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) neural network training algorithm (Fletcher, 1987), and conjugate gradient training algorithms along with a two layer architecture embedded in WinGamma software were used. The BFGS algorithm is a quasi-Newton method performed iteratively using successively improved approximations to the inverse Hessian matrix, instead of the true inverse.

The beginning of artificial neural network theory belongs to 1940s, when McCulloch famous psychologist and Walter Pitts mathematician founded it in 1943 (McCulloch and Pitts, 1943) and then Rosenblatt in 1962s proposed the idea of perceptrons (Rosenblatt, 1962). Mathematically, an artificial neural network is an approximation and its ability in approximating between the models of an issue causes to implement problems with high complexities such as pattern recognition, pattern classification, non-linear mapping and
associative memory and controlling Algorithms such as conjugate gradient, quasi-Newton, Levenberg-Marquardt (LM) and etc, are faster algorithms that all of them can use of standard numerical optimization methods There is a group of algorithms according to the Newton’s method need no calculation of second derivatives. Such methods are known as quasi-Newton methods which have had a successful performance in conducted researches which was designed by Broyden, Fletcher, Goldfarb and Shanno (BFGS). BFGS algorithm requires much more computation in each repetition and also higher storage than the conjugate gradient methods although it is less convergent in lower number of iterations. For smaller network, BFG training algorithm can be an effective training function. For more information please refer to conducted studies by Apostolopoulou et al. 2009; Minsky and Papert, 1969; Jones, 2004; Lawrence and Giles, 2000 and Nawi et al. 2008. In this study, we used two training algorithm including BFGS algorithm (Fletcher, 1987) and, conjugate gradient algorithm for prediction of suspended sediment load. For this purpose, we used winGamma software version 1.97.

**Evaluation Criteria for Model Performance**

The reliability of SVM and ANN models were evaluated based on the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Efficiency Index (EI) and correlation coefficient (R2). The RMSE represents the deviation between simulated values and observed values. The lower MAE values indicate more accurate estimations. EI measures are the ratio of the variability of the modeled values to the variability of the original data values. The equations of each criterion are as follows:

In order for assessing the accuracy of the models, various statistics have been developed and used the best known and most widely used will be presented in the following. These statistics are appropriately used in calibration phase to determine the parameter and structures

1- Root Mean Square Errors (RSME)

This statistic obtained through the following equation

\[
RMSE = \left( \frac{\sum_{i=1}^{N} (p_i - o_i)^2}{N} \right)^{1/2}
\]
\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Q_{(s,\text{predicted})i} - Q_{(s,\text{observed})i})^2} \]  

(27)

The less this statistic, the better is the function of the model.

2- Mean Absolute Error (MAE)

The mean absolute error is the predictions and observation obtained through the following equation

\[ MAE = \frac{\sum_{i=1}^{n} |p_i - o_i|}{n} \]  

(18)

\[ MAE = \frac{\sum_{i=1}^{n} |Q_{(s,\text{predicted})i} - Q_{(s,\text{observed})i}|}{n} \]  

(28)

The less this statistic, the better is the function of the model.

3- The correlation between predicted and actual values (R²):

\[ R^2 = \frac{\sum_{i=1}^{n} (o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^{n} (o_i - \bar{o})^2 \sum_{i=1}^{n} (p_i - \bar{p})^2}} \]  

(19)

\[ R^2 = \frac{\sum_{i=1}^{n} (Q_{(s,\text{predicted})i} - \overline{Q}_{\text{predicted}})(Q_{(s,\text{observed})i} - \overline{Q}_{\text{observed}})}{\sqrt{\sum_{i=1}^{n} (Q_{(s,\text{predicted})i} - \overline{Q}_{\text{predicted}})^2} (Q_{(s,\text{observed})i} - \overline{Q}_{\text{observed}})^2} \]  

(29)

4- Efficiency Index (EI)

\[ EI = 1 - \left( \frac{\sum_{i=1}^{n} (o_i - p_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2} \right) \]  

(20)

\[ EI = \frac{\sum_{i=1}^{n} (Q_{(s,\text{observed})i} - \overline{Q}_{\text{observed}})^2 - \sum_{i=1}^{n} (Q_{(s,\text{predicted})i} - Q_{(s,\text{observed})i})^2}{\sum_{i=1}^{n} (Q_{(s,\text{observed})i} - \overline{Q}_{\text{observed}})^2} \]  

(30)
where $o_i$ and $p_i$ are the observed and predicted values at time $i$, respectively; $\bar{o}$ and $\bar{p}$ is the mean of the observed and predicted values; and $n$ is the number of data points. Where $Q_{s, \text{observed}}$ and $Q_{s, \text{predicted}}$ are the observed and predicted values at time $i$, respectively; $\bar{Q}_{s, \text{observed}}$ and $\bar{Q}_{s, \text{predicted}}$ is the mean of the observed and predicted values; and $n$ is the number of data points. Moreover, all the input variables were scaled to fall in the range 0-1 following below equation:

$$x'_i = (x_i - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}})$$

(31)

**STUDY AREA AND DATA SET**

Doiraj River basin is located in Abdanan city in Ilam province Province at the western part of Iran between at 47° 16’ and 47° 40’ East latitude and 32° 34’ and 33° 05’ North longitude. The Doiraj River contains high level of sediment in its watershed. Data used in this study were daily river discharge, daily rainfall and daily suspended sediment load. These data were from Pol-e-Doiraj hydrometric station and Abdanan rain gauge station. The related stations used for the present study are given in Table 1.

Watershed elevation is varying from between 100 (m) to 2200 (m) above the sea level. The average rainfall in this basin is about 434.43 (mm) per year annum and occurs mostly in autumn and winter. The temperature is high in summers especially in southern parts of basin. The mean maximum daily temperature and the mean minimum daily temperature are 27.9 °c and 13.72 °c, respectively. Figure 2 3 shows location of Doiraj River basin in Iran.

**Results and discussion**

**Data Analysis and Model Input Selection based on Gamma Test Model Input Selection**

To prediction predict daily suspended sediment load SSL, we take took different lag time series of streamflow and rainfall ($Q_t$, $R_t$, $Q_{t-1}$, $R_{t-1}$, $Q_{t-2}$, $R_{t-2}$, $Q_{t-3}$, $R_{t-3}$, $Q_{t-4}$, $R_{t-4}$) as input data and suspended sediment load SSL discharge ($Q_{st}$) as output data.

Choosing proper component from input parameters Identifying best input combination is the most important step of any construction and designed of every modeling modeling. As indicated in Fig. 4, the complexities of the model including the higher number of inputs, more data for training model, a model with greater parameters, may have less prediction error; however, it is not necessarily ensure fewer errors at the test phase. In this condition, there is
an optimal condition in which prediction errors are minimized at the test phase (Bray and Han, 2004).

- Model input selection using GT and GT-GA

In this section, the parameters which increase the complexity of the model and those do not have a significant impact on the model’s results are identified and removed. Selection of an appropriate combination from input variables is the most important stage of construction process for any mathematical and intelligent modeling. The Gamma Test is able to provide the best mean square error that can possibly be achieved using any nonlinear smooth models.

There are $2^n - 1$ meaningful combinations of inputs; from which, the best one can be determined by observing the Gamma value, which indicates a measure of best MSE attainable using any modeling method for unseen smooth functions of continues variables (Remesan et al., 2008). Gamma test predicts the minimum achievable modeling error before the modeling.

Suppose $n$ is the variables influencing on occurrence of a phenomenon; $2^n - 1$ meaningful combination would be established from the input variables. In this study, Different combinations of input data were evaluated to assess their influence on the SSL prediction by GT as shown in Table 2. The gradient (A) is considered to be an indicator of model complexity and $V_{ratio}$ is a measure of the degree of predictability of given outputs using available inputs.

As shown in Figure 3 As the results indicated in Table 2, among 10 existence parameters, $Q_t$ has the greatest influence on suspended sediment discharge ($Q_{st}$). Removing this parameter from modeling causes the increase in amount of Gamma statistic ($\Gamma$) and SE. Because omitting this parameter from the modeling increases the Gamma value ($\Gamma$) and standard error (SE). Also, removing $Q_{t-3}$, $R_{t-3}$, $Q_{t-4}$ and $R_{t-4}$ has very slight effect on Gamma value. Furthermore, omitting the parameters $Q_{t-3}$, $R_{t-3}$, $Q_{t-4}$ and $R_{t-4}$ had no significant influence on gamma value. The minimum value of gamma static was observed when all available input data sets were used except $R_{t-4}$. After identifying the most effective variable, the best input combination should be determined for prediction.

In this study we were using Genetic Algorithm (GA) for finding best combinations The GA analysis results are shown in Figure 4. The best value of the Gamma statistic ($\Gamma$), gradient (A)
and $V_{\text{Ratio}}$ were observed when we used current rainfall ($R_t$), current discharge ($Q_t$), two step antecedent rainfall ($R_{t-2}$) and two step antecedent discharge ($Q_{t-2}$). On the other words, the embedding 1100110000 model was identified as the best combination because of its low $\Gamma$ value, the rapid decline of genetic algorithm SE graph, low $V_{\text{Ratio}}$ value, the regression line fit with slope $A=0.19956$ (low enough as a simple non-linear model with a minimum complexity) and good fit with $SE = 1.0833 \times 10^{-5}$ (as shown in Table 3). The Genetic Algorithm results are shown in Figure 4. In this study, we were using combination of gamma test and genetic algorithm for finding the best combinations and the accuracy, and reliability of its results have been compared with the results obtained from traditional correlation coefficient analysis. Figure 5 illustrates gamma changes and standard error in genetic algorithm test. According to the genetic algorithm method and given to the principals of this method, the best combination is the one with lowest gamma value and lowest value of SE as well. Meanwhile, we can see some of the cases with such criteria. Selecting the best case out of all these cases should be done precisely.

According to the results of GT-GA, the combination of 1100110000 (four inputs and an output) is selected as the best combination for inputs. The statistic values associated with the GT-GA in best combination is shown in Table 3. Small gamma value shows that data would achieve better possibility of results in modeling by the provided combination. Low SE value is also another reason for this claim. Low value of $V_{\text{ratio}}$ indicates that complexity of modeling in association with this combination is lower and better results can be expected.

Moreover, in order to review the reliability of the results obtained from the GT-GA method, various combinations from input parameters were evaluated using GT so as to determine the best combination among the remaining variables for predicting the SSL (note that in selecting the combinations, it has been tried to choose different combinations including parameters which have been recognized in prediction using gamma test as the most effective input parameters ($Q_t$, $R_t$, $Q_{t-1}$, $R_{t-1}$, $Q_{t-2}$, $R_{t-2}$)). These combinations are illustrated in Table 4 along with their gamma values and SE. The results indicate that the best input combination from the variables is when required rainfall and discharge are used along with rainfall and discharge with two delayed time ($Q_t$, $R_t$, $Q_{t-2}$, $R_{t-2}$). Low gamma value shows that data would achieve better possibility of results in modeling by the provided combination.
Therefore, the best input combination determined with GT-GA is similar to the best combination determined with GT.

- **Model input selection using correlation coefficient analysis**

The traditional correlation coefficient analysis between input and output variables was used in order to review the reliability of the results obtained from the GT and GT-GA methods. Correlation values of each of the input variables with sediment output are shown in Fig. 6. As you can see, \((Q_t)\) \((Q_{t-1})\) and \((R_t)\) have the highest correlation with sediment output. Given to the identified variables by this method, 4 different combinations (as indicated in Table 5) have been developed.

- **Identifying training data length**

The suitable number of the data points required for training the model was determined using the M-test. The results obtained from M-test analysis are shown in Fig. 8. As shown in the Figure 7, the M-test produced an asymptotic convergence of the gamma statistic to a value of \(2.82 \times 10^{-6}\) at around 2950 data points. The SE corresponding to \(M=2950\) is \(1.66 \times 10^{-5}\).

**Non-linear Modeling with SVMs SVM and ANNs**

In this study, ANNs and SVMs SVM models were utilized to predict suspended sediment load SSL. The ANN models are trained with the conjugate gradient algorithm and BFGS algorithm. The best combination of model inputs is a model with four inputs (as indicated by GT) and 2950 data points (as indicated by M-test) for training. Also, three types of SVM including one class SVM, epsilon-SVR and nu-SVR were compared in terms of performance criteria. The performances of these models were evaluated by four global statistics including \(R^2\), RMSE, EI, and MAE in training data set (\(M=2950\)). The model performance results for training and testing steps are given in Table 4. Given to the identified input variables by GT, GT-GA and correlation coefficient analysis, 5 BFGS-ANN models (as indicated in Table 6) and 5 CGNN models (as identified in Table 7) have been developed. Also, 2950 data points utilized for model’s training. Given to the results provided in Table 6 and Table 7, it can be see that determining the predictors through GT (or GT-GA) has been the best model in training and testing phase; because it had the minimum RMSE and maximum correlation between predicted and actual values.
Figs. 8 shows scatter plots of observed and predicted SSL by using ANN models. Curves of observed and predicted SSL by ANN models (Conjugate algorithm) are represented in Figs. 9.

In order to confirm the results achieved from the M-test and ensuring about lack of over-fitting occurrence during the training, 10 different scenarios have been defined and accordingly there have been developed 10 BFGS-ANN models. As it was illustrated in Table 8, during such scenarios (scenarios 1 to 10), the duration of training model increased from 1000 to 3500 data and performance of each scenario has been investigated. The results of Table 7 showed that during every 10 defined scenarios, RMSE values of the test phase were greater than RMSE in training phase. According to the results of the training phase, after scenario No. 3, (more than 2000 data for training) RMSE values decreased. Moreover, after the scenario No. 8, (more than 3000 data for the training model) during the test phase, RMSE values were increased and correlation values were decreased. Accuracy in this results demonstrated that the highest correlation and the lowest RMSE occurred during the training phase and testing phase between the scenario No. 7 (2900 data for the training) to scenario No. 8 (3000 data for the training) which was in accordance with the results of M-test (2950 data for the training). Curves of observed and predicted SSL by BFGS-ANN with the training and testing data sets are represented in Figs. 10.

In order to predict the SSL through SVR model, values of parameters $C$ and $\varepsilon$ should be optimized. The $C$ parameter keeps a balance between margin maximization and training error minimization. A smaller $C$ results in low pressure and a very large $C$ causes over-fitting of data training. It hardly has an impact on the prediction error. The kind of noise in the data, when determinable, directly influences the optimal value of $\varepsilon$. The number of resulting support vectors should be considered as well. The training set, if insensitive to $\varepsilon$, will not encounter the boundary conditions. The optimal $\gamma$ value should be determined through trial and error and be always maintained. When increased substantially, it results in over-fitting (i.e. the prediction of only the trained data). In this case the model becomes complex due to the need to consider the distances of all support vectors. When the value of $\gamma$ drops substantially, under-fitting occurs (i.e. the model being unable to predict the trained data) which is caused by the machine ignoring most of the support vectors. In this study, we used the nu-SVR model using the four kernels including linear, polynomial, sigmoid and Radial Basis Function and model’s performance were evaluated through the RMSE values. At any
stage of model, optimal values of the parameters were selected according to two important factors

1- The models in the test stage should have the lowest amount of RMSE using the selected parameter

2- Under-fitting and over-fitting must not occur during the prediction process of the model using the selected parameters

Figs. 11, 12, 13 and 14 shows the curve changes of predicted SSL using 3 values of $\varepsilon$ using linear, sigmoid, polynomial and RBF kernels respectively. As indicated here, with increasing $\varepsilon$ value, prediction error is increased and in $\varepsilon = 1$ this error is maximized and the SSL model can be predicted during the test period with the constant value.

Fig. 15 shows the RMSE changes of nu-SVR model with RBF kernel according to various $\varepsilon$ values. Best $\varepsilon$ values for linear, sigmoid, polynomial and RBF kernels are chosen 0.026, 0.063, 0.0027 and 0.009 respectively. Fig. 16 demonstrates the prediction errors changes of nu-SVR model with RBF kernel according to various C values. According to this figure, increase value of C from 2 to 9 can reduce the RMSE, from 4.34 (ton) to 4.05 (ton) and it minimizes its lowest value in $C = 9$ and thereafter with increase of C value, RMSE value would be increased and can be maximized with $C = 1000$ (RMSE=6.61 (ton)). Hence, $C = 9$ is chosen as the optimal value.

Furthermore, the RMSE changes of nu-SVR model with RBF kernel according to various $\gamma$ values are shown in Fig. 17. If $\gamma$ value is so small under-fitting occurs because most of support vectors are ignore. Also, the large value of $\gamma$ increases the complexity of model and will lead to over-fitting (Han et al, 2007). As you seen when $\gamma = 1$, nu-SVR model with RBF kernel has lowest RMSE. In order to evaluate the performance of GT-GA and GT in identifying the best input combination, SSL predict based on identified combination using correlation coefficient analysis by using nu-SVR model (with RBF kernel). These results are shown in Table 9.

As you can see, although nu-SVR model with combination model.3 (identified by regression method) has better performance in training phase, but in testing phase nu-SVR model by using best GT combination has lowest RMSE.
The performances of BFGS-ANN, CGNN and nu-SVR models were evaluated by four global statistics including $R^2$, RMSE, EI, and MAE. The model performance results for training and testing steps are given in Table 10. The results show, the BFGS algorithm based ANN model performed better on the testing data set than the conjugate gradient ANN model, with $R^2=0.99$, 0.97 and RMSE=0.0015, 0.0025 respectively. Also, epsilon-SVR results and nu-SVR results were better than one-class SVM model. The performances of epsilon-SVR and nu-SVR models are compared with four types kernel including linear, polynomial, Radial Basis Function (RBF) and sigmoid separately (Tables 5 and 6). Table 5 shows that the kernels of linear, sigmoid and RBF behave like each other in both training period and testing period with correlation coefficient values of 0.93, 0.94, 0.92 and RMSE values of 0.048, 0.031, 0.049, respectively. Among four types of kernel functions, epsilon-SVR model with polynomial kernel shows the worst performance for prediction. The results given in Tables 5 and 6 show that the nu-SVR model performed better on the training data set and testing data set than the epsilon-SVR model. Nu-SVR model with kernel function of Radial Basis Function (RBF) can predict suspended sediment load better than other kernels with values of correlation coefficient and RMSE 0.98 and 0.00423, respectively. Comparison between BFGS-based ANN model and nu-SVR model shows the performance BFGS-based ANN model is almost similar to the performance of nu-SVR model. On the other hand, the nu-SVR model and BFGS-based ANN model produces the lowest errors in terms of $R^2$ (0.98 and 0.99, respectively) and RMES (0.0042 and 0.0015, respectively). The results given in Table 4 indicate that the predictive capability of one-class SVM model is poor in comparison with the other used models in daily suspended sediment load modeling Figures 6 and 7 show scatter plots of observed and predicted suspended sediment loads by using ANN and SVM models with the testing data set. In addition, curves of observed and predicted suspended sediment loads by ANN models (BFGS algorithm and Conjugate algorithm) and SVM models (based on Radial Basis Function kernel) with the training and testing data sets are represented in Figures 8, 9, 10 and 11. The results show, the BFGS algorithm based ANN model performed better on the testing data set than the conjugate gradient ANN model, with $R^2=0.99$, 0.97 and RMSE=0.34, 0.58 respectively. The results given in Table 10 show that nu-SVR model with kernel function of Radial Basis Function (RBF) can predict SSL better than other kernels with values of
Comparison between BFGS-based ANN model and nu-SVR model shows the performance of BFGS-based ANN model is almost similar to the performance of nu-SVR model. On the other hand, the nu-SVR model and BFGS-based ANN model produces the lowest errors in terms of $R^2$ (0.98 and 0.99, respectively) and RMES 0.964 (ton) and 0.341 (ton) respectively. In addition, BFGS-based ANN model and nu-SVR model have the best EI values equal to 0.997 and 0.983 respectively. Also, the predictive capability of nu-SVR model with linear kernel is better than the polynomial kernel with RMSE 3.42 (ton) and 8.49 (ton) respectively. Sigmoid kernel has poor performance in comparison with the other kernels used in daily SSL modeling in study area. Figs. 18 shows scatter plots of observed and predicted SSL by using nu-SVR models with the testing data set. Curves of observed and predicted SSL by nu-SVR models (based on Radial Basis Function kernel) with the training and testing data sets are represented in Figs. 19.

**Conclusion**

In this study, artificial neural network and support vector machine models were applied to predict suspended sediment load in Doiraj River basin located in the west of Iran. This paper illustrates the first application of Gamma test for prediction of suspended sediment load. Streamflow and rainfall data were considered as input variables and length of the training data set and the best input combination of input variables for the models were identified by using GA. Finally, the daily suspended sediment load was predicted by the BFGS--based ANN model, Conjugate- based ANN model and one-class SVM, epsilon-SVR and nu-SVR models. nu-SVR model and performances of these models were compared in terms of four global statistics ($R^2$, RMSE, MAE and EI). In order to evaluate the performance of GT in identifying the best input combination, SSL predict based on identified combination using correlation coefficient analysis by using nu-SVR model (with RBF kernel), BFGS model and CGNN model. Results shown, although nu-SVR model with identified combination by regression method has better performance in training phase, but in testing phase nu-SVR model by using best GT combination has lowest RMSE. Also, BFGS-ANN and CGNN models has better performance in testing and training phase by using GT combination. In addition, partitioning approach was implemented in order to evaluate the performance of M-test in identifying the duration of the training model, and the performance of BFGS-ANN model was reviewed during 10 different scenarios. The results showed that GT can estimate the training data length of the model with a higher accuracy.
The results obtained from this study show that the BFGS-based ANN model and nu-SVR model produce the best performance in daily suspended sediment load prediction. In addition, RBF kernel function for SVM model represents the reasonable and promising results than the other kernel functions. Results shown performance of nu-SVR model with Linear kernel is almost similar to the RBF kernel. In other researchers' articles RBF kernel has been identified as best kernel for modeling using support vector machines (Dibike et al., 2001; Lin et al., 2006). Lin and Lin, (2003) statement behavior of RBF kernel and Sigmoid kernel are similar each other (for particular parameters). Keerthi and Lin, (2001) illustrate the ability of RBF kernel with some C and \( \gamma \) parameter is similar to linear kernel. In addition, polynomial kernel comparing RBF kernel has most numerical difficulties (because of more hyperparameters) (Lin et al., 2006). Therefore RBF kernel is chosen as the best kernel for modeling. Model input data selection for non-linear dynamic system such as suspended sediment load prediction is a complicated process. For non-linear systems such as SSL prediction, determination of best input combination is a complicated process. The present research aimed at answering the following three main questions:

1- Which one of input variables is the most important variable?

2- What combination of input variables should be considered for modeling?

3- How many data is necessary for training and testing of a data-driven model like SVM or ANN to predict suspended sediment load?

According to the research objectives, the application of state of the art techniques for selection of the best model input combination like Gamma test, as a less time-consuming procedure than trial and error method eases the selection of relevant variables in the construction of non-linear models for daily suspended sediment load prediction. In order to complete the current study, it is suggested that the results obtained from employing Gamma test are compared with those can be captured for the ANN and SVM models designed on the basis of other input selection techniques such as PCA and FS in predicting suspended sediment load.

SSL.

References


de Oliveria, A.G., 1999. Synchronisation of chaos and applications to secure communications. PhD thesis, Department of computing, Imperial College of Science, Technology and Medicine, University of London.


Han, D., Yang, Z., (2001). River flow modeling using Support Vector Machines. 29th IAHR Congress, Beijing, China, September, 17–21.


Tables list:

Table 1: List of the used station in Doiraj basin

Table 2: The Gamma Test results

Table 3. Gamma statistic values in best embedding

Table 4: Determination of best input combination using GT

Table 5: Identified combinations based on Correlation coefficient analysis

Table 6: comparison of results obtained from BFGS-ANN model based on the GT and correlation coefficient analysis combinations

Table 7: comparison of results obtained from CGNN model based on the GT and correlation coefficient analysis combinations

Table 8: Results of different scenarios for identifying training data length using BFGS-ANN model

Table 9: Results of different input combination obtained from the GA and Correlation method using nu-SVR model (RBF kernel)

Table 10. Comparison between nu-SVR and ANN models for prediction SSL
Figures list:

Fig. 1. Structure of SVM model (Vapnik, 1998)

Fig. 2. Cost function of SVM model (Lin et al., 2006)

Fig. 3. Doiraj River basin

Fig. 4. The impact of model complexity on accuracy of the results (Bray and Han, 2004)

Fig. 5. Variation of Gamma statistic for the data corresponding to different combination of inputs

Fig. 6. Correlation values between 10 input parameters and suspended sediment discharge

Fig. 7. The variation of Gamma static and standard error with unique data points

Fig. 8. Scatter plot of observed and predicted SSL by ANN models

Fig. 9. Observed and predicted curves of SSL by Conjugate ANN for training data set

Fig. 10. Observed and predicted curves of SSL by BFGS-ANN for testing data set

Fig. 11. Predicted SSL based on 3 different values using linear kernel

Fig. 12. Predicted SSL based on 3 different values using sigmoid kernel

Fig. 13. Predicted SSL based on 3 different values using polynomial kernel

Fig. 14. Predicted SSL based on 3 different values using RBF kernel

Fig. 15. Determination of optimal value using RBF kernel

Fig. 16. Determination of optimal C value using RBF kernel

Fig. 17. Determination of optimal gamma value using RBF kernel

Fig. 18. Scatter plot of observed and predicted SSL by nu-SVR model (RBF kernel)

Fig. 19. Observed and predicted curves of SSL by nu-SVR (RBF kernel)
Dear Editor:

The following paper is a product of MSc degree project in Watershed Management Engineering:

**Daily Suspended Sediment Load Prediction Using Artificial Neural Networks and Support Vector Machines**

Kakaei Lefdani, E., Moghaddam Nia, A. and Ahmadi, A.

The present manuscript has been written by the above authors as a research paper. We would be very happy if you could let us know for any constructive comments and revision in need.

We are looking forward to hearing from you any feedback and constructive comments.

Sincerely yours,

Alireza Moghaddam Nia
Associate Professor of Hydrology
Department of Range and Watershed Management
University of Zabol
Iran
Table 1: List of the used station in Doiraj basin

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Type of station</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol-e-Doiraj</td>
<td>Hydrometric station</td>
<td>45° 25´ 00</td>
<td>32° 59´ 00</td>
<td>940</td>
</tr>
<tr>
<td>Abdanan</td>
<td>rain gauge station</td>
<td>47° 24´ 00</td>
<td>32° 35´ 40´´</td>
<td>165</td>
</tr>
</tbody>
</table>

Table 2: The Gamma Test results

<table>
<thead>
<tr>
<th>Different combinations</th>
<th>Mask</th>
<th>Gamma(Γ)</th>
<th>Gradient(A)</th>
<th>SE</th>
<th>(V_{ratio})</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>111111111111</td>
<td>-0.000041</td>
<td>0.089963</td>
<td>-0.000038</td>
<td>-0.00994</td>
<td>4014</td>
</tr>
<tr>
<td>No (Q_i)</td>
<td>011111111111</td>
<td>0.0009575</td>
<td>0.067575</td>
<td>0.00000784</td>
<td>0.22912</td>
<td>4014</td>
</tr>
<tr>
<td>No (R_{i-1})</td>
<td>101111111111</td>
<td>0.0000252</td>
<td>0.072571</td>
<td>0.0000308</td>
<td>0.006049</td>
<td>4014</td>
</tr>
<tr>
<td>No (Q_{i-2})</td>
<td>110111111111</td>
<td>1.016</td>
<td>0.0000366</td>
<td>-0.00658</td>
<td>4014</td>
<td></td>
</tr>
<tr>
<td>No (R_{i-3})</td>
<td>111011111111</td>
<td>-0.0000111</td>
<td>0.085503</td>
<td>0.0000280</td>
<td>0.00282</td>
<td>4014</td>
</tr>
<tr>
<td>No (Q_{i-4})</td>
<td>111011111111</td>
<td>-0.000054</td>
<td>0.1079</td>
<td>0.0000294</td>
<td>0.01294</td>
<td>4014</td>
</tr>
<tr>
<td>No (R_{i-5})</td>
<td>111101111111</td>
<td>0.000050</td>
<td>0.0946</td>
<td>0.0004356</td>
<td>-0.01022</td>
<td>4014</td>
</tr>
<tr>
<td>No (Q_{i-6})</td>
<td>111110111111</td>
<td>-0.000042</td>
<td>0.10591</td>
<td>0.0000356</td>
<td>-0.01022</td>
<td>4014</td>
</tr>
<tr>
<td>No (R_{i-7})</td>
<td>111111011111</td>
<td>0.000050</td>
<td>0.10591</td>
<td>0.0000356</td>
<td>-0.01022</td>
<td>4014</td>
</tr>
<tr>
<td>No (Q_{i-8})</td>
<td>111111101111</td>
<td>0.000062</td>
<td>0.10483</td>
<td>0.0000391</td>
<td>-0.01487</td>
<td>4014</td>
</tr>
</tbody>
</table>

Table 3: Gamma statistic values in best embedding

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gamma(Γ)</th>
<th>Gradient(A)</th>
<th>SE</th>
<th>(V_{ratio})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>5.723×10^{-8}</td>
<td>0.19956</td>
<td>2.0833×10^{-5}</td>
<td>1.3694×10^{-5}</td>
</tr>
</tbody>
</table>

Table 4: Determination of best input combination using GT

<table>
<thead>
<tr>
<th>Input combination</th>
<th>Mask</th>
<th>Gamma (Γ)</th>
<th>Standard Error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q_i, P_i, Q_{i-2}, P_{i-2})</td>
<td>1100110000</td>
<td>5.723×10^{-8}</td>
<td>0.00002083</td>
</tr>
<tr>
<td>(Q_i, Q_{i-2}, P_{i-2})</td>
<td>1000110000</td>
<td>0.000003062</td>
<td>0.00001133</td>
</tr>
<tr>
<td>(Q_i, P_i, Q_{i-2})</td>
<td>1100010000</td>
<td>-0.00002105</td>
<td>0.000020297</td>
</tr>
<tr>
<td>(Q_i, P_i, Q_{i-2})</td>
<td>1100100000</td>
<td>-0.00000803</td>
<td>0.00001056</td>
</tr>
<tr>
<td>(Q_i, P_i)</td>
<td>1100000000</td>
<td>-0.000015673</td>
<td>0.00007402</td>
</tr>
<tr>
<td>(Q_i, P_i, Q_{i-1}, P_{i-1}, Q_{i-2}, P_{i-2})</td>
<td>1111110000</td>
<td>-0.000005622</td>
<td>0.00003213</td>
</tr>
<tr>
<td>(Q_i, P_i, Q_{i-1}, Q_{i-2}, P_{i-2})</td>
<td>1110110000</td>
<td>0.000001014</td>
<td>0.00003480</td>
</tr>
<tr>
<td>(Q_i, P_i, P_{i-1}, Q_{i-2}, P_{i-2})</td>
<td>1101110000</td>
<td>-0.00003431</td>
<td>0.000030962</td>
</tr>
</tbody>
</table>
Table 5: Identified combinations based on correlation coefficient analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>(Q_t, R_t, Q_{t-1}, R_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4})</td>
</tr>
<tr>
<td>Model 2</td>
<td>(Q_t, R_t, Q_{t-1})</td>
</tr>
<tr>
<td>Model 3</td>
<td>(Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4})</td>
</tr>
<tr>
<td>Model 4</td>
<td>(Q_t, R_t, Q_{t-1}, Q_{t-2})</td>
</tr>
</tbody>
</table>

Table 6: Comparison of results obtained from BFGS-ANN model based on the GT and correlation coefficient analysis combinations

<table>
<thead>
<tr>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient analysis</td>
<td>Correlation coefficient analysis</td>
</tr>
<tr>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.9867</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.7806</td>
</tr>
</tbody>
</table>

Table 7: Comparison of results obtained from CGNN model based on the GT and correlation coefficient analysis combinations

<table>
<thead>
<tr>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient analysis</td>
<td>Correlation coefficient analysis</td>
</tr>
<tr>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.9861</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.8115</td>
</tr>
</tbody>
</table>

Table 8: Results of different scenarios for identifying training data length using BFGS-ANN model

<table>
<thead>
<tr>
<th>scenarios</th>
<th>Training data length</th>
<th>Training period</th>
<th>Testing period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE (ton)</td>
<td>(R^2)</td>
<td>RMSE (ton)</td>
</tr>
<tr>
<td>1</td>
<td>1000</td>
<td>0.7681</td>
<td>0.9978</td>
</tr>
<tr>
<td>2</td>
<td>1500</td>
<td>1.4827</td>
<td>0.9992</td>
</tr>
<tr>
<td>3</td>
<td>2000</td>
<td>1.3815</td>
<td>0.9924</td>
</tr>
<tr>
<td>4</td>
<td>2250</td>
<td>0.865</td>
<td>0.9971</td>
</tr>
<tr>
<td>5</td>
<td>2500</td>
<td>0.8195</td>
<td>0.9971</td>
</tr>
<tr>
<td>6</td>
<td>2750</td>
<td>0.576</td>
<td>0.9986</td>
</tr>
<tr>
<td>7</td>
<td>2900</td>
<td>0.5047</td>
<td>0.9989</td>
</tr>
<tr>
<td>8</td>
<td>3000</td>
<td>0.3514</td>
<td>0.9995</td>
</tr>
<tr>
<td>9</td>
<td>3250</td>
<td>0.404</td>
<td>0.9993</td>
</tr>
<tr>
<td>10</td>
<td>3500</td>
<td>0.64</td>
<td>0.9982</td>
</tr>
</tbody>
</table>
Table 9: Results of different input combination obtained from the GA and Correlation method using nu-SVR model (RBF kernel)

<table>
<thead>
<tr>
<th></th>
<th>Training Correlation coefficient analysis</th>
<th></th>
<th>Testing Correlation coefficient analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
</tr>
<tr>
<td>R²</td>
<td>0.998</td>
<td>0.997</td>
<td><strong>0.999</strong></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.628</td>
<td>0.882</td>
<td><strong>0.6042</strong></td>
</tr>
</tbody>
</table>

Table 10: Comparison between nu-SVR and ANN models for prediction SSL

<table>
<thead>
<tr>
<th>Model</th>
<th>Training</th>
<th></th>
<th>Testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>RMSE</td>
<td>EI</td>
<td>MAE</td>
</tr>
<tr>
<td>BFGS-ANN</td>
<td>0.9998</td>
<td>0.0023</td>
<td>0.9999</td>
<td>0.0023</td>
</tr>
<tr>
<td>CGDNN</td>
<td>0.9997</td>
<td>0.2667</td>
<td>0.9996</td>
<td>0.22339</td>
</tr>
<tr>
<td>nu-SVR (Linear kernel)</td>
<td>0.9591</td>
<td>4.1236</td>
<td>0.9276</td>
<td>1.58425</td>
</tr>
<tr>
<td>nu-SVR (Polynomial kernel)</td>
<td>0.2933</td>
<td>13.3351</td>
<td>0.2428</td>
<td>7.70015</td>
</tr>
<tr>
<td>nu-SVR (Sigmoid kernel)</td>
<td>0.7800</td>
<td>11.395</td>
<td>0.4471</td>
<td>2.48238</td>
</tr>
<tr>
<td>nu-SVR (RBF kernel)</td>
<td>0.9975</td>
<td>0.7636</td>
<td>0.9975</td>
<td>0.5562</td>
</tr>
</tbody>
</table>
Figure 4

The diagram illustrates the relationship between prediction error and model complexity. The graph shows two curves:

- **High bias, Low variance**: This curve starts high on the left side and decreases as model complexity increases. It represents models that are too simple to capture the underlying patterns in the data.

- **Low bias, High variance**: This curve starts low on the left side and increases as model complexity increases. It represents models that are too complex, leading to overfitting.

The diagram also indicates the test set and training set, with a vertical dashed line marking the optimal model complexity where the model generalizes well to new data. The horizontal axis represents the model complexity, ranging from low to high.
Figure 9

Conjugate ANN (Training data set)
- Observed
- Predicted

Conjugate ANN (Testing data set)
- Observed
- Predicted

Daily SSL (ton) vs Days

Days

0 500 1000 1500 2000 2500

0 20 40 60 80 100 120

0 10 20 30 40 50

Days

0 250 500 750 1000

0 15 30 45 60 75 90 105 120

140 160 165
Figure 11:nu-SVR (Linear kernel)
nu-SVR (Polynomial kernel)
nu-SVR (Testing data set)

\[ y = 0.93x + 0.3889 \]

\[ R^2 = 0.9871 \]
> We identify the most effective variables on prediction in order of importance degree.
> We select the best combination of model inputs for predicting suspension load.
> We determine how many data is considered for training of ANN and SVM models.