Journal of Advanced Informatics in Water, Soil, and Structure 2025 2 (2) 187-201

Journal of Advanced Informatics in Water, Soil, and Structure

journal homepage: www.wss.torbath.ac.ir

Research Article

Analyzing the Uncertainty of Artificial Neural Network Models and Support Vector Machines in Estimating Saturated Hydraulic Conductivity

Nabizadeh Amir [a], *Khashe-Siuki Abbas* [b], *Afrasiab Peyman* [b], *Pourreza-Bilondi Mohsen* [b]

[a] Dept of Water Engineering, University of Zabol, Zabol, Iran.

[b] Dept of Water Engineering, University of Birjand, Birjand, Iran.

ARTICLE INFO

Received date: 31 Mar 2024
Accept date: 11 Jan 2025
Published date: 01 May 2025

Keywords:
ANN, LS-SVM, Soil, Suction Disc, Monte Carlo, Evaluation Criteria.

Abstract

Data-basis simulation models such as artificial neural network (ANN) and support vector machines (LS-SVM) are suitable substitute for expensive and laboratory methods of estimating soil hydraulic properties, including hydraulic saturation conductivity of soil. In this study, the objective is to predict the saturated hydraulic conductivity of soil in the study area, Sistan Dam, using ANN and LS-SVM. To this purpose, the saturated hydraulic conductivity of soil was measured in 112 points by the suction disc method. For selecting input variables among simple characteristics, the multistep regression method was used. In these models, the selection of different transfer and training functions represents a significant source of error. Consequently, a comprehensive analysis was conducted to identify uncertainty sources in simulating the saturated hydraulic conductivity of soil. To determine the best simulation model among neural network functions and support vector machines, a Monte Carlo method was employed to sample and evaluate their performance. This rigorous approach ensures a thorough examination of the strengths and limitations of these models, thereby enhancing the reliability and accuracy of soil conductivity predictions. Finally, the uncertainty in predicting the solution model can be analyzed to determine how much ANN models and support vector machines (LS-SVMs) are reliable. In general, when considering both linear and nonlinear scenarios in ANNs and LS-SVM, linear input methods may not be a suitable approach. This is primarily due to the complex and nonlinear nature of the soil properties and processes being modeled. Therefore, nonlinear input methods are typically preferred to accurately capture the intricate relationships and dynamics of soil systems. This highlights the importance of selecting appropriate input methods to ensure reliable and accurate modelling of soil properties and behaviour. In addition, among the ANN and LS-SVM models used in this study, ANN, with values of content ratio criteria (CR), (with the values of 90, 0.263, and 0.740) high degree of asymmetry index (S) and high and low asymmetry index (T), was of higher certainty and accuracy than the other models. Moreover, Logsig_trainlm and Tansig_trainbfg scenarios in predicting saturated hydraulic conductivity had a satisfactory process and less uncertainty.

Homepage: www.wss.torbath.ac.ir

*Corresponding Author:
Khashe-Siuki Abbas
Email: *abbaskhashei@birjand.ac.ir*
ORCID: *0000-0002-2863-8482*

https://wss.torbath.ac.ir/article_213145.html

How to cite this article:
Nabizadeh, A. , Khashei-Siuki, A. , Afrasiab, P., and Pourreza Bilondi, M. (2025). Analyzing the uncertainty of artificial neural network models and support vector machines in estimating saturated hydraulic conductivity. *Journal of Advanced Informatics in Water, Soil, and Structure*, *2*(2), 187-201.

© 2025 by the Authors, Published by University of Torbat Heydarieh. This article is an open access article distributed ns and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (http://creativecommons.org/licenses/by/4.0/).

Introduction

Reliable information on soil hydraulic properties plays a vital role in solving many water and soil management problems related to agriculture, environment, and environmental issues (Yao et al., 2015). Hydraulic conductivity is one of the key hydrodynamic properties in modeling water movement in soil, controlling surface runoffs, leaching agricultural fields, and transferring surface contamination into groundwater, which is essential in environmental assessment, solutions for groundwater pollution contamination, and water resources assessment. Hydraulic properties are also important for describing and predicting the transfer of water and solutes, as well as for the model of heat transfer near soil surface.)Cornelis et al., 2001). Several factors have an increasing or decreasing effect on the aforementioned properties, which can be attributed to soil physical and chemical properties, aggregate stability, climate, agricultural operations, water quality, land use, root dynamics, and the activity of living organisms in the soil (Fuentes et al., 2004). Tension penetrometer method, due to its ease of use, is one of the most popular tools for measuring the saturated and unsaturated water characteristics of a field (Moret & Arrue, 2007), which measures the hydraulic conductivity directly. Tension penetration method has a relatively fast application capability in the field with minimal soil surface corrosion. This device can also be located in the vicinity of zero-power matrix, where the soil pores have a high water activity for transferring water and solutes (Ankeny et al., 1991). However, the direct measurements of hydraulic conductivity are not implementable in a dense network of points due to being difficult and expensive. In addition, spatial variability also adds to the problems of these methods. For this reason, they are characterized as costly measured properties (Budiman et al., 1999; Pachepsky et al., 1996). This has prompted researchers to develop indirect reliable predictions of saturated hydraulic conductivity has been considered a lot since late nineteenth century (Rogiers et al., 2011).

Consequently, the focus has shifted towards the indirect estimation of hydraulic conductivity, leading to a growing interest in utilising pedotransfer (PTF) transfer functions for estimating soil's hydraulic properties. This increased attention underscores the significance of PTFs in advancing our understanding and predictive capabilities of soil behaviour. (Wagner et al., 2004; Stumpp et al., 2009; Arrington et al., 2013). One of the indirect methods for determining the saturated hydraulic conductivity is the application of data-basis simulation models such as artificial neural networks and support vector machines that can predict saturated hydraulic conductivity of soil at lower cost and higher efficiency using simple features such as soil texture, organic matter and organic carbon, predict latent properties.

In this regard, many studies have been conducted to estimate hydraulic conductivity using data-basis models. Xu et al. (2017) estimated the saturated hydraulic conductivity created by using three different types of PTFs with different inputs to estimate the saturated hydraulic conductivity using hydraulic guiding characteristics of 171 points in southwest of China and an artificial neural network model (ANN). Zhao et al. (2016) used linear regression (MLR) functions and artificial neural network (ANN) to estimate soil hydraulic conductivity in Loess Plateau of China.

Motaghian and Mohammadi (2011) estimated the saturated hydraulic conductivity space using the characteristics of the earth by means of regression, kriging, and artificial neural networks. Ferre Julià et al. (2004) estimated the hydraulic conductivity in terms of sand percent data and with the aid of transition functions for a wide range of climate and physiographic conditions in Spain. Merdun et al. (2006) compared artificial neural network and transfer regression functions to predict soil moisture and saturated hydraulic conductivity.

However, for environmental models that are appropriately used in a risk-based decision-making process and the obtained results are not certain, for the practical application of the results, it should be considered to estimate the uncertainty in the results of model (White et al., 2018), and having information about their uncertainty is essential, which, unlike the statistical models, is considered that in the neural network models. By examining references in this regard, there are limited methods for determining uncertainty in these models, including Monte Carlo method presented by Marcé et al. (2004). Monte Carlo simulation involves creating repetitions of random parameters from the distribution of probabilities, and then computing the output statistics (Dehghani et al., 2014).

In this method, randomly, the input database is restored for 1000 times without replacement, while maintaining the relationship between calibration (training and validation) and test sets (Noori et al 2010a). In this research, Monte Carlo method was considered due to the newness and proper function. The accurate analysis of environmental modeling often serves as a supporting tool in resource managers' decision making, which plays an important role in modeling (Gorelick and Zheng, 2015; Horne et al., 2016) and, on the other hand, analyzing the predictive range is very important to understand the uncertainty of modeling in a comprehensive and objective way. This uncertainty provides a clear understanding of the reliability of simulation results and management solutions that depend on the simulation results and considers a safety margin for the inadequacy of the model and its data (Anderson et al., 2015). Many methods have been proposed and tested for evaluating and expressing modeling uncertainty. By studying the available literature in this regard, according to several methods, the method proposed by Xiong et al. (2009), which gives a more precise analysis of the uncertainty range, was considered in this study. There is no study that uses these indexes for uncertainty issues of data-basis simulation models in relation to hydraulic conductivity and the effective parameters. The purpose of this research is to be able to determine and obtain the degree of uncertainty of the above models in estimating hydraulic conductivity.

Materials and methods

study area

The studied area is Sistan Dam land with an area of about 85 hectares. This area is located 25 kilometers south of Zabul in Sistan and Baluchistan province, which is one of the southeastern provinces of Iran. At the beginning of the study, field studies and soil samples were conducted at 112 points of the Sistan Dam research farm. To this purpose, 112 points were firstly identified with a mean distance of 91 m from each other on Sistan Dam field. Hydraulic conductivity was measured using a tension disk penetrometer. Then, intact and handled soil samples was collected from the surface depth (0-20 cm) and were used to laboratory was transferred to measure the physical and chemical properties of the soil. So simple soil characteristics including soil texture components (silt, sand, and clay percentages), bulk density, true density, organic matter, organic carbon, Ec, and pH were evaluated.

Field and laboratory practices

Hydraulic conductivity measurements were carried out using a disc diffusion and soil sampling device in about 110 points of the Sistan Dam research farm. For this purpose, at first, about 110 points were determined with a mean distance of 91 meters from each other at Sistan dam field (Figure 1).

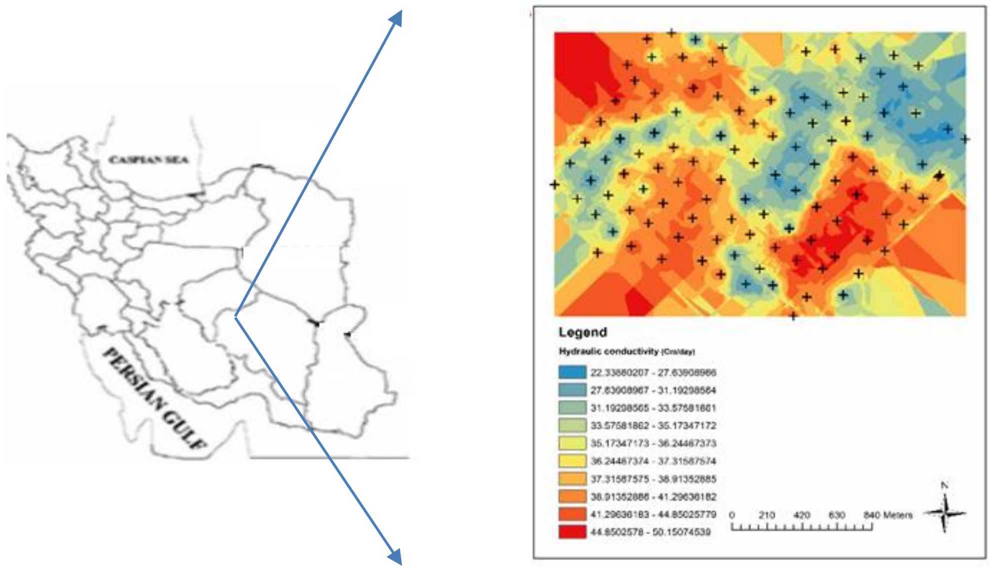

Figure 1- Zoning sampled points on Sistan Dam filed

Data-Basis Forecast Models

In order to implement data-basis models of artificial neural network and support vector machines in this research, the Artificial Neural Network Toolkit and support vector machines were used in Matlab. Also, the data sets were standardized for the first time in the range of 0-1 in order to obtain the stability of the model (Equation 1) and then were returned to the initial values.

$$X_{norm} = \left[\frac{X - X_{min}}{X_{max} - X_{min}} \right] \times 0.8 + 0.1 \qquad (1)$$

Where *X* is the initial value, *Xmin* and *Xmax* are the minimum and maximum values respectively, *Xnorm* is the normalized value and 0.8 and 0.1 are evaluation values.

The first and the most important step in making data-basis models is to select appropriate inputs. Choosing appropriate inputs is important for the following reasons:

A. By increasing the number of input variables, the complexity of computing and memory requirements of the model increases rapidly.

B- By entering a large number of variables, training the model gets challenging and inefficient.

C- Entering outlier variables leads to an increase in the number of least localized at the error level, which can lead to non-convergence and, consequently, low accuracy of the model (Body et al., 2005).

In this research, the stepwise regression method was used to determine proper variables. Gravel, sand, clay, pH, Ec, bulk density, true density, organic matter, organic carbon, and porosity as independent variables and hydraulic conductivity as dependent variable (Table 1) were considered for obtaining regression models. Based on the results obtained from this step (Table 2), the results indicate that eliminating most of the data does not affect the main variable and does not create meaningful differences. While by removing clay, Ec, organic carbon, and organic matter, the amount of R^2 decreases from 0.535 to 0.382, which indicates the significant effect of these four variables on the saturated hydraulic conductivity.

Table 1- Summary of descriptive statistics of soil parameters used to estimate saturated hydraulic conductivity

Parameter	Unit	Mean	Minimum	Maximum	Range	STD.d	CV	Skewness	Kurtosis
Sand	%	70.65	57.86	94.52	36.66	7.38	10.45	0.373	-0.20
Clay	%	3.99	0.09	13.2	13.11	3.07	76.87	0.926	0.08
Silt	%	25.36	5.39	37.39	32	5.27	20.79	-0.565	0.91
pH	-	7.72	7.08	8.13	1.05	0.23	2.93	-0.448	0.15
Ec	ds/m	2.15	0.71	7	6.29	1.66	77.2	1.525	1.22
Bulk Density	g/cm3	1.55	0.56	2.35	1.79	0.26	16.77	0.043	1.84
Real Density	g/cm3	2.61	2.04	2.94	0.9	0.23	8.65	-0.564	-0.49
Organic carbon (O.C)	%	0.63	0.21	1.11	0.9	1.66	25.4	0.244	0.63
Organic Matter (O.M)	%	1.09	0.27	1.92	1.55	0.26	25.45	0.243	0.6
Porosity (n)	%	40.28	0.74	79.84	79.1	0.23	28.62	-0.316	1.11
Hydraulic conductivity (Tension Disk)	cm/day	36.924	18.02	52.14	34.12	0.16	26.787	-0.558	-0.827

Table 2- Validation values for selecting effective parameters by means of stepwise regression

Model	RMSE(cm/day)	Radj	R2
All parameter	7.069	0.489	0.535
Without Ec and O.M	7.75	0.425	0.467
Without Ec, O.M and O.C	7.993	0.346	0.382

In this research, 64% of the data points (72 out of 112 samples) were allocated for training, 18% (20 sample sets) for validation, and another 18% (20 sample sets) for testing. Initially, the training dataset was utilized. As shown in Table 2, clay content, electrical conductivity (Ec), organic carbon, and organic matter were used as input variables, while the saturated hydraulic conductivity measured by the tension disk system was the output variable. The computational process in MATLAB is illustrated in Figure 2 as a flowchart.

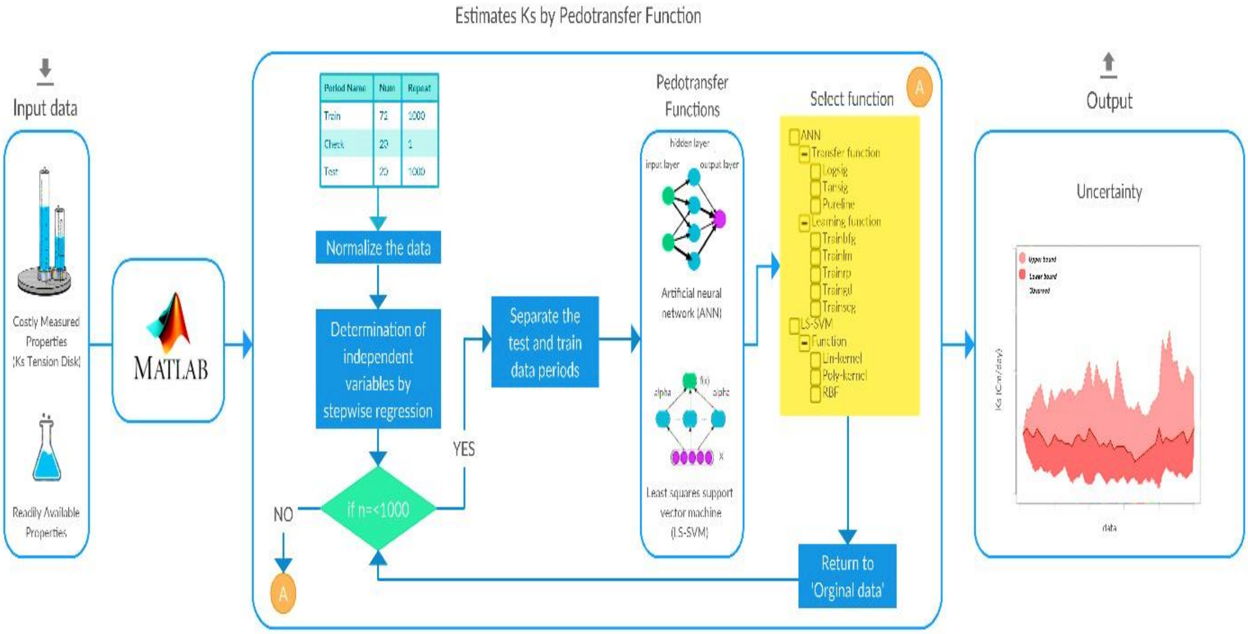

Figure 2- Schematic map of data-basis prediction models in MATLAB

Uncertainty of data-basis models

In the context of the research, various data-based models are utilized, leading to distinct uncertainty analysis approaches compared to conceptual models. Understanding and applying these unique methods is crucial when utilizing and advancing these models (Akbarzadeh et al., 2010). Within such models, it is evident that predictions are not absolute, emphasizing the importance of conducting thorough analyses to enhance the efficiency of predictive model outcomes. Uncertainty analysis involves randomly selecting input values and inputting them into the model to generate output distributions.

Data-basis models of neural network and support vector machines were used as a Monte Carlo sampling. The basis of Monte Carlo simulation is also based on using random numbers and the acquisition of a probability distribution function of a model (Eckhardt et al., 2003). Therefore, 1000 series of samples were produced. Estimation of 95% confidence interval is more consistent with reality (Noori et al., 2010b) because the confidence interval provides more information than the other statistics related to the range of model predictions (Noori et al., 2010b). The 95% confidence interval is determined by finding 2.5% and 97.5% of the buildable distribution (Noori et al., 2009).

Uncertainty of support vector machines

In this research, the least squares model of support vector machines (LS-SVM) was used in MATLAB using LS-SVM toolbox. The LS-SVM algorithm was applied to two stages training-validating and testing. In addition, employing this model to predict hydraulic conductivity has not been reported yet.

The process of employing the method is described step-by-step in the following:

1. Data split for training, validation, and test courses
2. Introducing the simple variables defined by the stepwise regression method as the input vector and the saturated hydraulic conductivity measured by a tension disk as the output.
3. In this research step 1 was repeated 1000 times and randomly, using Monte Carlo method, so that the training and validation data were different in all 1000 samples and only data related to the test range was assumed to be identical throughout the repetitions.
4. Training the relationships between input and output vectors in the training stage after initial weighing by using the least squares regression support vector machines using linear kernel, polynomial kernel, and radial basic kernel functions
5. Providing the data of the test period with the results of training weights and estimating the predicted values of saturated hydraulic conductivity during this period
6. Determining the 95% probability band for the model output
7. Investigating and concluding the uncertainty of the network for various compounds based on 95% confidence level and tension disk data as

the observed data

Uncertainty of the artificial neural network

In this research, the multi-layer perceptron network was employed for modeling and post-error algorithm using Tansig, Logsig and Pureline transfer functions for both network layer was identical and Trainbfg, Traingd, Trainlm, Trainrp, and Trainscg training functions were utilized in MATLAB.

The process is as described step-by-step as follows:

1- Dividing data for training, validation, and test courses
2- Introducing the simple variables defined by the stepwise regression method as the input vector and the saturated hydraulic conductivity measured by the tension disk as the output vector
3- In this research step 1 was repeated 1000 times and randomly, using Monte Carlo method, so that the training and validation data were different in all 1000 samples, and only data related to the test range was assumed to be identical throughout the repetition.
4- Training the relationships between input and output vectors in the training stage after initial weighing by using the transfer and training functions employed in the network
5- Providing the data of the test period with the results of training weights and estimating the predicted values of saturated hydraulic conductivity during this period
6- Determining the 95% probability band for the model output
7- Investigating and concluding the uncertainty of network for various compounds based on 95% confidence level and the tension disk data as the observed data

Evaluation criteria

For uncertainty analysis, some criteria are always required for analyzing the produced confidence limits and the quality of the boundaries. Therefore, the following criteria are defined and applied to 95% confidence interval:

Containing ratio (CR): This criterion is defined as the ratio of output to the uncertainty range, which is defined as a percentage. In this research, the output are the data of the tension disk in cm/d.

Bandwidth (B): The width of the uncertainty range for the total predicted output calculated according to equation (1):

$$B = \frac{1}{N}\sum_{i=1}^{N} b_i \qquad (2)$$

$$where\; b_i = q_i^u - q_i^l$$

Where b_i is the width of the uncertainty range for the i^{th} output, q_i^u is the upper limit of the confidence range for the output of the i^{th} time, q_i^u is the lower limit of the confidence limit for the output of the i^{th} time and N is the total number of measurements.

Two indicators of the mean asymmetry grade (S[1] and T[2]): Two indicators that are considered for estimating the geometric structure of the range formed from the high and low limits of uncertainty and investigating the symmetry of the uncertainty range. The first indicator, which is represented by S, is described in Equation 2:

$$S = \frac{1}{N}\sum_{i=1}^{N} s_i \qquad (3)$$

where

$$s_i = \left|h_i - 0.5\right|,\; h_i = \frac{q_i^u - Q_i}{q_i^u - q_i^l} = \frac{q_i^u - Q_i}{b_i}$$

Where s_i represents the degree of asymmetry of the upper limit of the output obtained in the range of uncertainty associated with h_i function for the i^{th} output.

The second indicator for evaluating the mean asymmetry degree of the uncertainty range with respect to the output graph is defined as Eq. (3), and is represented by T.

$$T = \frac{1}{N}\sum_{i=1}^{N} t_i \qquad (3)$$

where

$$t_i = \left(\frac{\left|\left(q_i^u - Q_i\right)^3 + \left(q_i^l - Q_i\right)^3\right|}{\left(q_i^u - q_i^l\right)^3}\right)^{1/3}$$

In which the variations of t_i depend on the location of the obtained output graph with respect to the uncertainty range, indicating the status of the degree of asymmetry of the upper and lower bounds.

Deviation-amplitude (D): To calculate the actual

[1] Symmetry

[2] Two- Symmetry

difference between the range consisting of the intermediate points of the uncertainty range and the output graph, the lateral index of standard deviation domain is used, which is shown by D and is defined as equation 4:

$$D = \frac{1}{N}\sum_{i=1}^{N} d_i \qquad (4)$$

where

$$d_i = \left|q_i^m - Q_i\right| = \left|\frac{1}{2}\left(q_i^u - q_i^l\right) - Q_i\right|$$

Where d_i is the output deviation obtained from the midpoint of the range of uncertainty, resulting from the bandwidth *bi* and the degree of asymmetry *si*.

Results and discussion

In this research, the uncertainty assessment related to the structure of neural network modeling and support vector machines was analyzed. In this section, the results of different performances with scenarios that were obtained by combining transfer and training functions in MATLAB with different sampling of the training course achieved by Monte Carlo sampling are provided in this section.

The results of repeating the neural network model 1000 times with 72 different data sets of the neural network and support vector machines obtained by Monte Carlo sampling for each of the 15 scenarios (the combination of three transfer functions with 5 training functions) were obtained in the neural network and three support vector models, which are discussed conclusively in results and discussion section.

The results of the uncertainty of neural network

Figures 3 to 5 depict a confidence interval of 95% of the model output for the 15 available scenarios. Figure 3 shows the transfer function of sigmoid logarithm, Figure 4 shows the linear transmission function, and Figure 5 shows the sigmoid tangent transfer function, which each were implemented with 5 different training functions including Trainbfg, Trainscg, Trainrp, Trainlm, and Traingd.

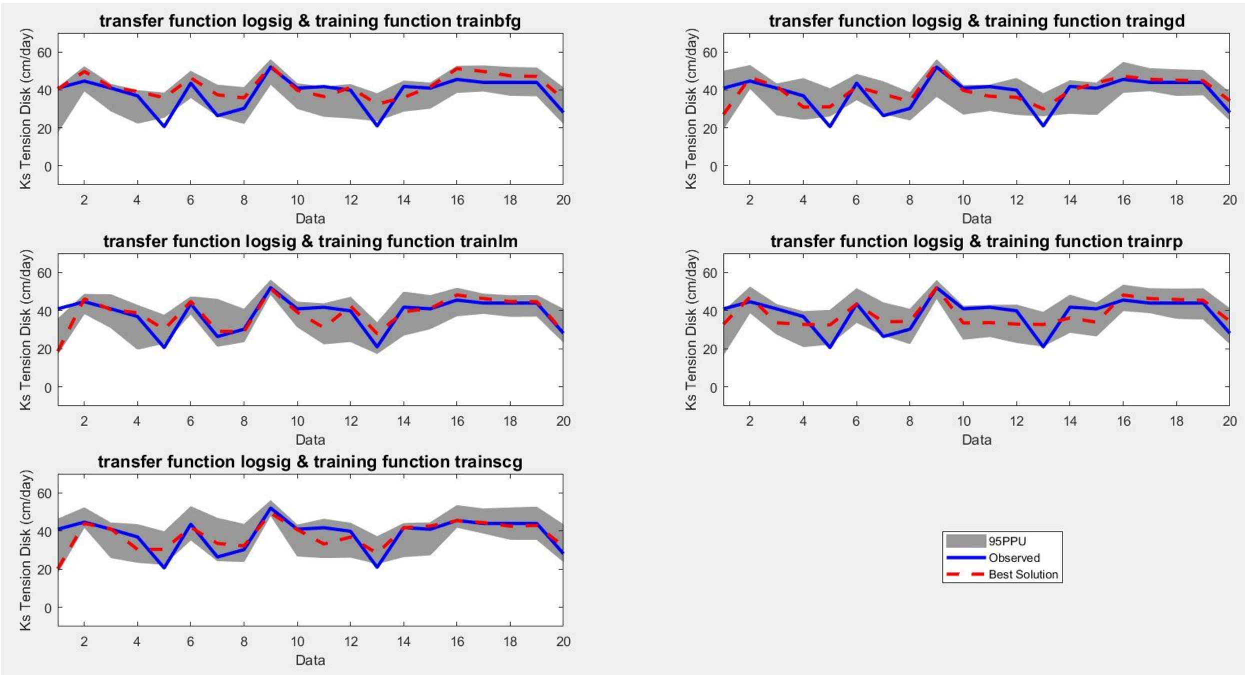

Figure 3- 95% confidence interval of output by the combination of using the Logsig transfer function and various training functions

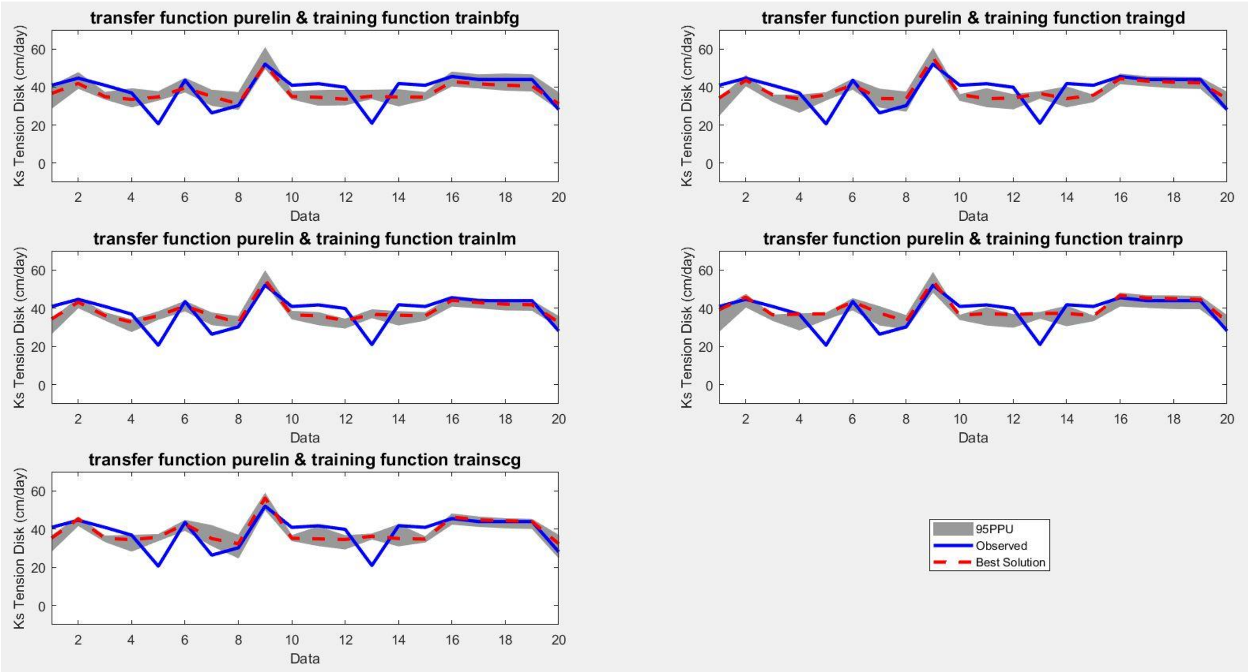

Figure 4- 95% confidence interval of output by the combination of using Pureline transfer function and various training functions

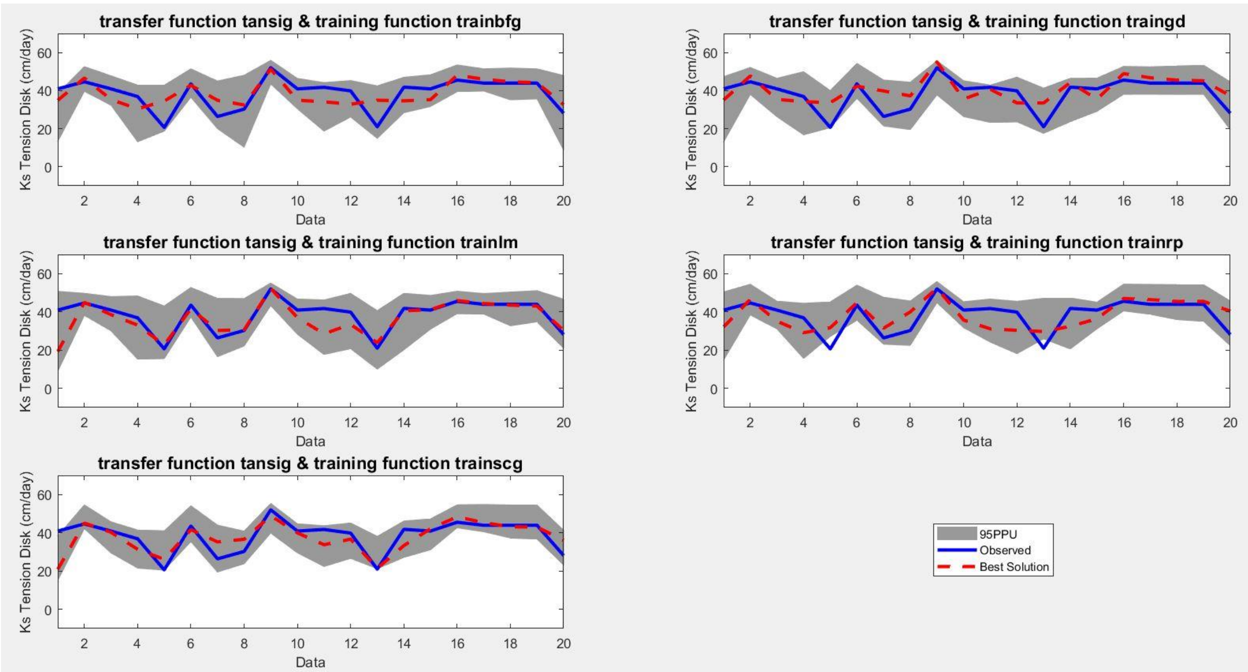

Figure 5- 95% confidence interval of output by the combination of using Tansig transfer function and various training functions

The values of the uncertainty criteria in evaluating different scenarios of the neural network are shown in Table 3. According to Figures 3, 4, and 5, it is observed that the content ratio (CR) of Tansig_trainlm and Tansig_traingd scenarios, with 100% of observed data being placed in the confidence range (CR=100%), Tansig_trainbfg scenario with 95% of observed data in the confidence range (CR=95%), Logsig_trainlm, Logsig_trainscg, Tansig_trainscg and Tansig_trainrp scenarios with 90% of observed data in the confidence level (CR=90%), Logsig_traingd and Logsig_trainbfg scenarios, with 85% of the observation data within the confidence interval

(CR=85%), and Logsig_trainrp scenario due to placing 80% of the observational data on the confidence level (80% = CR) are much more favorable than the other 5 scenarios.

As shown in Figure 4, the compatibility of the predicted results by the linear transfer function and the confidence range is very low and it means that the linear transfer function has very low certainty. Apart from the scenarios related to the linear transfer function, among the other scenarios Logsig_trainbfg, Logsig_traingd, Logsig_trainlm, Logsig_trainrp, Logsig_trainscg, and Tansig_trainscg have a better homogenous narrower bandwidth and a lower relative average width, respectively.

The best scenario can be selected by using the other four indicators. This is worth noting that the scenarios related to the linear transfer function did not fall within the appropriate symmetry limit and are far away from the suitable symmetry conditions (0≤T<1 and S<0.5). Therefore Logsig_trainlm scenario with the degree of symmetry indices (S, T), standard deviation range (D), and mean standard deviation range (RD) (S=0.218, T=0.67, D=3.825, RD=0.118) and Tangsig_trainbfg (Figure 5-a) with the degree of symmetry indices (S, T), standard deviation range (D) and mean standard deviation range (RD) less (S=0.178, T=0.598, D=3.980, RD=0.123) are of more certainty. Hence, in this study these scenarios were chosen as the best scenario used in the artificial neural network to predict the hydraulic conductivity.

Table 3- Values of uncertainty criteria in assessing different neural network scenarios

Model	Transfer Function	Learning Function	CR%	B	RB	S	T	D	RD
1	Logsig	Trainbfg	85	15.424	0.430	0.292	0.736	4.554	0.144
2	Logsig	Traingd	85	16.070	0.440	0.292	0.745	4.547	0.146
3	Logsig	Trainlm	90	15.990	0.455	0.218	0.670	3.825	0.118
4	Logsig	Trainrp	80	16.732	0.472	0.268	0.721	4.769	0.149
5	Logsig	Trainscg	90	17.048	0.480	0.252	0.685	4.486	0.143
6	Pureline	Trainbfg	45	7.549	0.203	0.865	1.413	5.128	0.168
7	Pureline	Traingd	40	6.667	0.184	1.086	1.670	5.801	0.183
8	Pureline	Trainlm	35	5.349	0.146	1.155	1.732	5.878	0.189
9	Pureline	Trainrp	45	6.699	0.179	1.119	1.715	5.082	0.170
10	Pureline	Trainscg	50	6.647	0.184	1.103	1.687	4.921	0.164
11	Tansig	Trainbfg	95	20.996	0.621	0.178	0.598	4.980	0.123
12	Tansig	Traingd	100	20.910	0.594	0.205	0.655	4.490	0.137
13	Tansig	Trainlm	100	22.558	0.659	0.160	0.609	4.089	0.123
14	Tansig	Trainrp	90	20.347	0.575	0.250	0.716	5.298	0.176
15	Tansig	Trainscg	90	17.794	0.503	0.263	0.740	4.819	0.145

After obtaining the results of the 15 scenarios, their results were analyzed cumulatively. So that the results of implementing all of the training functions (5 training functions) were combined together and then extracted with 5,000 runs within the 95% confidence interval separately for each of the transfer functions as before, and are depicted in Figure 6. The same process of aggregation of results and the estimation of the confidence interval was applied to different transfer functions and separately for training functions and is shown in Figure 7. In this case, each of the graphs in Figure 5 is due to 3000 runs related to the three transfer functions. Table 4 shows the values of uncertainty criteria in assessing different scenarios.

In Table 4, the content ratio (CR) got slightly better for Traingd, Trainlm, and Trainbfg training functions, however, for all of the employed training functions, the content ratio (CR), bandwidth (B), and average bandwidth (RB) are relatively the same. By studying the degrees of symmetry indices, two Trainlm and Trainbfg training functions have more suitable symmetry and the output deviation obtained from the midpoint of the prediction range of Trainbfg training function is slightly lower. But due to the very small difference, Trainbfg and Trainlm training functions are of more certainty and can be chosen as the best training functions for estimating saturated hydraulic saturation of soil.

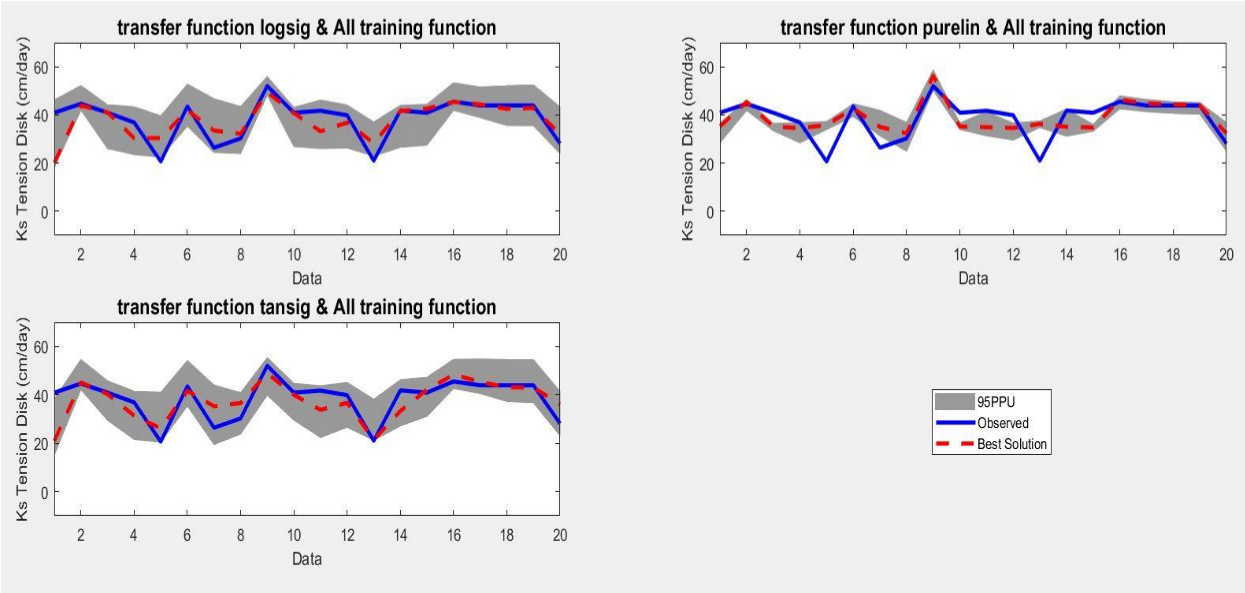

Figure 6- 95% confidence interval in the various training functions separately for each of the transfer functions

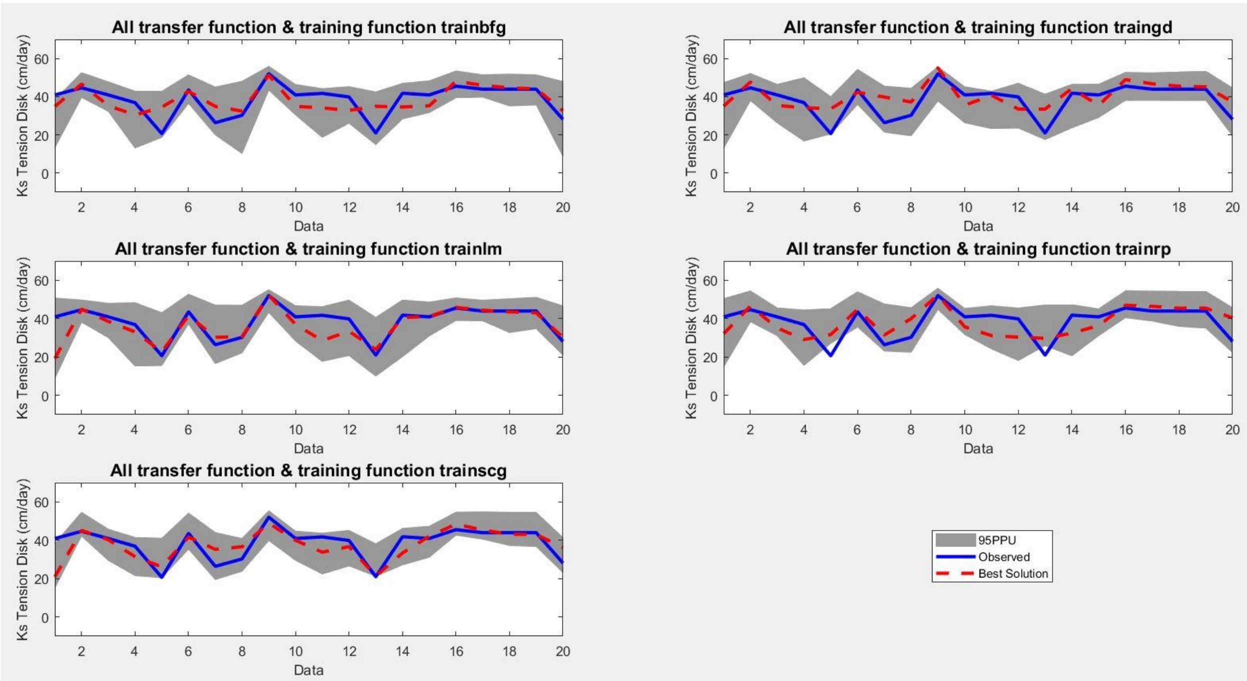

Figure 7- 95% output confidence interval in different transfer functions separately for each training function

Also, according to Table 4, the CR for each of the sigmoid logarithm and sigmoid tangent transfer functions is equal to 90%, but this value is much lower for the linear transfer function (CR=50%), as mentioned earlier it is noted that in the linear transfer function, a slight correlation is found between observational data and the desired confidence range. Since the content ratio has a very strong linear correlation with both bandwidth (B) and average relative bandwidth (RB), the resulting bandwidth (B) and average relative bandwidth (RB) for the linear transfer function are less than the two employed transfer functions. Moreover, the values of these two indicators for the sigmoid logarithm and the sigmoid tangent transmission are relatively equal. By examining the indices of symmetry degree (S, T), standard deviation range (D), and mean standard deviation range (RD) it is observed that sigmoid logarithm transfer function has a better symmetry and less deviation than the other two transmission functions, and in the second place, sigmoid tangent transfer function has more favorable conditions, but the linear transfer function is not within the symmetry range ($0 \leq T < 1$ and $S < 0.5$) and has an over-predicted asymmetric. As a result, sigmoid logarithm transfer function is of more certainty.

Table 4- Values of uncertainty criteria in evaluating various aggregation scenarios for the neural network

Number of simulations	Function Transfer	Function Learning	CR%	B	RB	S	T	D	RD
5000	Logsig	Overall	90	17.048	0.480	0.252	0.685	4.486	0.143
5000	Pureline	Overall	50	6.647	0.184	1.103	1.687	4.921	0.164
5000	Tansig	Overall	90	17.794	0.503	0.263	0.740	4.819	0.145
3000	Overall	Trainbfg	95	20.996	0.621	0.178	0.598	4.980	0.123
3000	Overall	Traingd	100	20.910	0.594	0.205	0.655	4.490	0.137
3000	Overall	Trainlm	100	20.558	0.659	0.161	0.609	4.089	0.123
3000	Overall	Trainrp	90	20.348	0.575	0.250	0.716	4.298	0.176
3000	Overall	Trainscg	90	17.794	0.503	0.263	0.740	4.819	0.145

Results obtained from the uncertainty of support vector

Figure 8 shows a confidence interval of 95% of the support vector output for linear, polynomial, and RBF kernel functions used in this research. The pale lines show 95% interval, the dark lines are the observed data of saturated hydraulic conductivity, and the dashed lines represent the best response from kernel function.

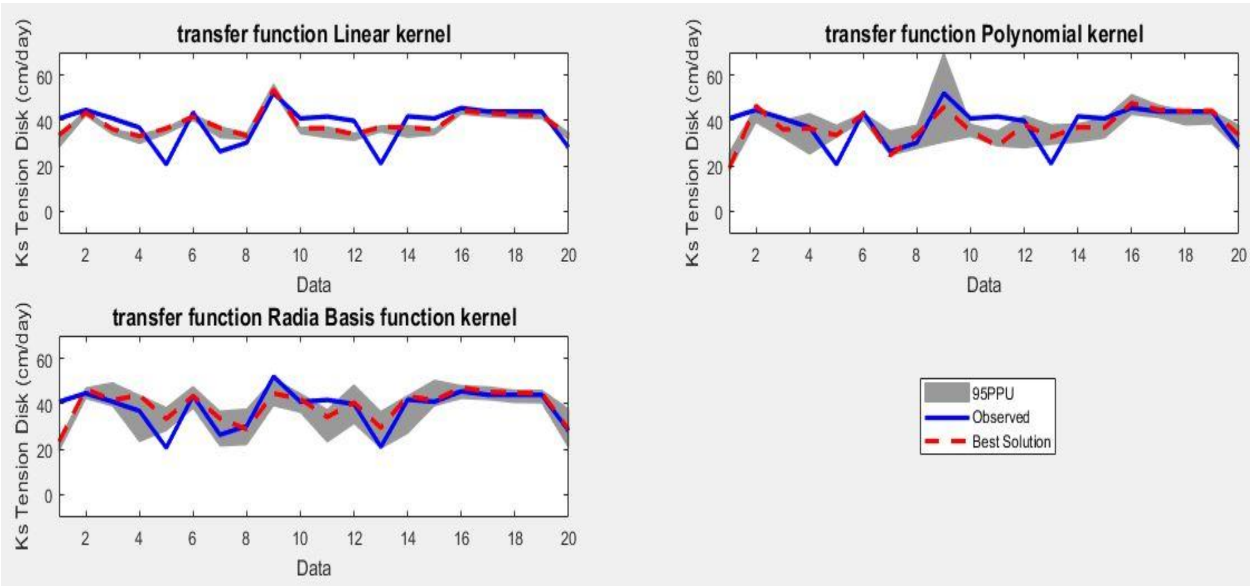

Figure 8 - a) 95% confidence interval of the output of linear kernel function of the support vector, b) Figure 7 - 95% confidence interval of the output of polynomial kernel function of the support vector c) Figure 7 - 95% confidence interval of the output of RBF kernel function of the support vector

Table 5- Uncertainty assessment criteria of the support vector uncertainty

Function	CR%	B	RB	S	T	D	RD
Lin-kernel	5	3.261	0.090	1.738	2.409	5.577	0.181
Poly-kernel	60	9.936	0.265	0.728	1.246	5.416	0.169
RBF	80	11.263	0.329	0.468	0.931	4.146	0.126

According to Table 5, the content ratios (CR) for Lin_kernel, Poly_kernel, and RBF_kernel functions are 5, 60, and 80%, respectively, which indicate that Lin_kernel function has the least degree of certainty and the match between observation data values with a confidence interval (Figure 8a), but Poly_kernel and RBF_kernel functions have better certainty and compatibility (Figure 8b and 8c). According to Figure 8, Poly_kernel function has less bandwidth (B) and relative RB bandwidth (RB) than RBF_kernel function.

But by examining the symmetry indices (S, T), standard deviation range (D) and mean standard deviation range (RD), not only RBF_kernel function has better results, but also Poly_kernel function is not in the range of symmetry conditions ($S< 0.5$ and $0 \leq T <1$) and has an over-predicted asymmetric value. As a result, RBF_kernel function has more certainty and was selected as the best used function of support vector machines for predicting hydraulic conductivity in this research.

Comparing the uncertainty assessment of support vector and neural network

By reviewing Table 6, the results indicate the superiority of the neural network functions to the support vector, so that apart from the two indicators of bandwidth (B) and average relative bandwidth (RB), which have a very strong linear correlation with the content ratio (CR), the support vector functions, unlike the neural network, are far away from ideal values (CR=100%). Also, in the remaining indices, the support vector functions have weaker results, so that in the study of the indices of symmetry degree (S,T) are over-predicted. On the other hand, neural network functions generally have more desirable symmetry and less deviation. Comparing the superior scenario of the neural network extracted from Table 3 and Figure 9 and the superior function of vector machines extracted from Table 5 and Figure 10, the results indicate a superiority of the neural network scenario.

Conclusion

In this research, the amount of uncertainty of the transmission functions of artificial neural network and support vector machines in estimating the saturated hydraulic conductivity obtained by tension disk was studied. First, four variables including Ec, sand, organic matter, and organic carbon were selected by means of multi-stage regression stepwise as a simple characteristic to predict saturated hydraulic conductivity. Afterwards, by examining the uncertainty of each of the scenarios used in the transmission functions of neural network and support vector machines, the logsig_trainlm scenarios combined with the sigmoid logarithm transfer functions and Trainlm and Tansig_trainbfg training methods combined with sigmoid Tangent transfer functions and Trainbfg training method were determined as the best combination of transfer functions and training functions for artificial neural network and RBF_kernel function as the best function used for support vector in this research.

Table 6- Values of uncertainty criteria in assessing the aggregation of all scenarios

Overlay Function	CR%	B	RB	S	T	D	RD
ANN	90	17.794	0.503	0.263	0.740	4.819	0.145
LS-SVM	60	9.936	0.265	0.728	1.246	5.416	0.169

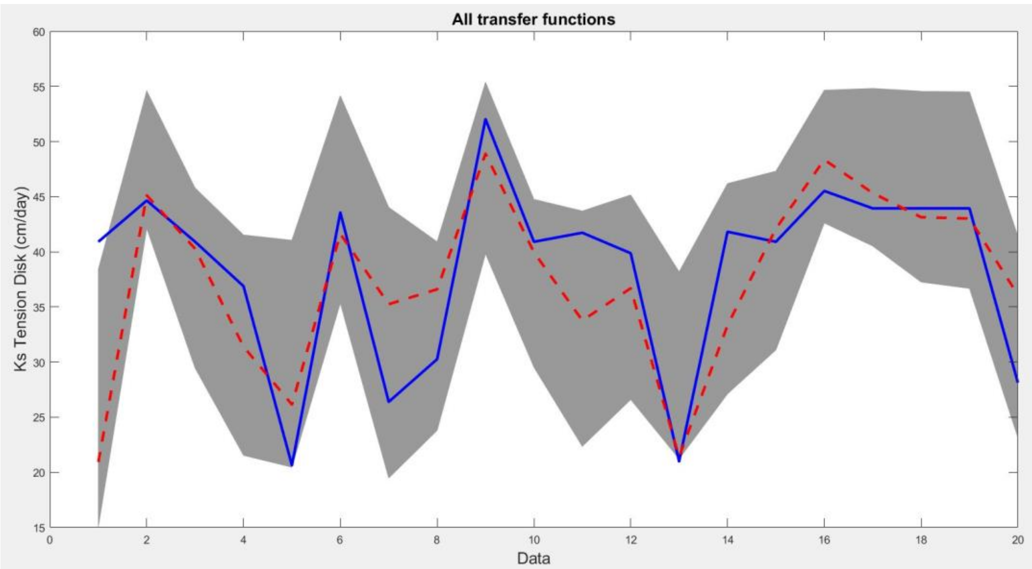

Figure 9- 95% output confidence interval of training and transmission functions of artificial neural network

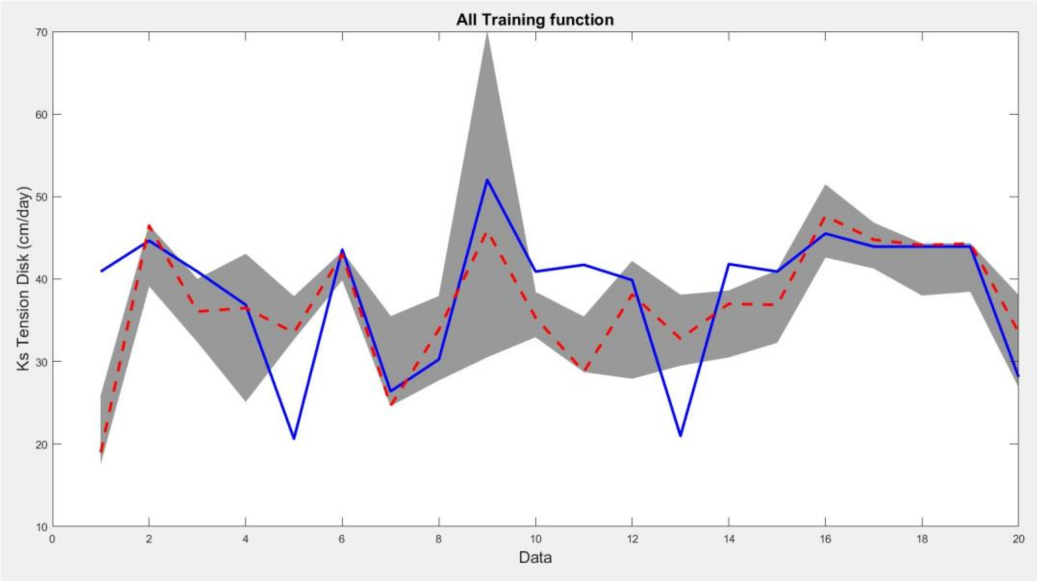

Figure 10- 95% output confidence interval of regression training and transmission functions of support vector

Also, by aggregating transfer functions with each other and training functions with each other in the neural network, sigmoid logarithm transfer function were determined as the best neural network transmission function and Trainlm and Trainbfg functions were selected as the best neural network training functions. While the same intervals of training and testing, as well as a similar samplings were used for the artificial neural network and support vector machines, but very different results were reported. So that by using the two models of neural network simulation and support vector in this research, the neural network model is of higher certainty and accuracy with the values of 90, 0.263, and 0.740 for content ratio (CR), high asymmetric degree index (S) and high and low asymmetry index (T), respectively. This means that the value of saturated hydraulic conductivity is very unfavorable through using support vector machines techniques in the study area.

Funding sources

No funding was received for conducting this study.

Declaration of competing interest

The authors declare no conflict of interest in preparing this article.

Authors contribution statement

N.A. *Conceptualization: evolution of overarching research goals and aims. Programming, software development; designing computer programs.*

K.S.A. *Conceptualization: evolution of research team to reach aims. Project administration: management and coordination responsibility for the research activity planning and execution. Supervision: oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.*

A.P. *Conceptualization: evolution of overarching research goals and aims. Programming, software development; designing computer programs.*

P.B.M. *Writing: original draft. Preparation and creation of the published work. Formal analysis.*

Date Availability

The data supporting this study's findings are available from the corresponding author upon reasonable request.

References

Akbarzadeh, A., Nouri, R., Farrokh Nia, A., Khakpur, A., Sabahi, M. S. (2000). Analysis of accuracy and uncertainty of intelligent models in predicting longitudinal propagation coefficient of rivers. *Quarterly Journal of Water and Sewage*, 21(3), 99–107.

Anderson, M. P., Woessner, W. W., & Hunt, R. J. (2015). *Applied Groundwater Modeling* (2nd ed.). Academic Press.

Ankeny, M. D., Ahmed, M., Kaspar, T. C., & Horton, R. (1991). Simple field method determining unsaturated hydraulic conductivity. *Soil Science Society of America Journal*, 55, 467–470.

Arrington, K. E., Ventura, S. J., & Norman, J. M. (2013). Predicting saturated hydraulic conductivity for estimating maximum soil infiltration rates. *Soil Science Society of America Journal*, 77, 748–758.

Bowden, G. J., Dandy, G. C., & Maier, H. R. (2005). Input determination for neural network models in water resources applications, Part 1: Background and methodology. *Journal of Hydrology*, 301(1), 75–92.

Budiman, M., McBratney, A. B., & Bristow, K. L. (1999). Comparison of different ... [title truncated in original] *[Please provide full title and source if available]*

Cornelis, W. M., Ronsyn, J., Van Meirvenne, M., & Hartmann, R. (2001). Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Science Society of America Journal*, 65, 638–648.

Dehghani, M., Saghafian, B., Nasiri Saleh, F., Farokhnia, A., & Noori, R. (2014). Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte–Carlo simulation. *International Journal of Climatology*, 34, 1169–1180.

Eckhardt, K., Breuer, L., & Frede, H. G. (2003). Parameter uncertainty and the significance of simulated land use change effects. *Journal of Hydrology*, 273(1), 164–176.

Ferrer, J. M., Estrela Monreal, T., Sánchez del Corral Jiménez, A., & García Meléndez, E. (2004). Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction. *Geoderma*, 123, 257–277.

Fuentes, J. P., Flury, M., & Bezdicek, D. F. (2004). Hydraulic properties in a silt loam soil under natural prairie, conventional tillage, and no-till. *Soil Science Society of America Journal*, 68, 1679–1688.

Gorelick, S. M., & Zheng, C. (2015). Global change and the groundwater management challenge. *Water Resources Research*, 51(5), 3031–3051.

Horne, A., Szemis, J. M., Kaur, S., Webb, J. A., Stewardson, M. J., Costa, A., & Boland, N. (2016). Optimization tools for environmental water decisions: A review of strengths, weaknesses, and opportunities to improve adoption. *Environmental Modelling & Software*, 84, 326–338.

Marcé, R., Comerma, M., Garcia, J. C., & Armengol, J. (2004). A neuro-fuzzy modeling tool to estimate fluvial nutrient loads in watersheds under time-varying human impact. *Limnology and Oceanography: Methods*, 2(11), 342–355.

Merdun, H., Çınar, Ö., Meral, R., & Apan, M. (2006). Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil & Tillage Research*, 90, 108–116.

Merdun, H., Ozer, C., Meral, R., & Apan, M. (2006). Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil Tillage Research*, 90, 108–116.

Alemi, M. H. (1981). *Water and soil*. Tehran: Tehran University Press. (In Persian)

Moret, D., & Arrúe, J. L. (2007). Dynamics of soil hydraulic properties during fallow as affected by tillage. *Soil and Tillage Research*, 96, 103–113.

Motaghian, H. R., & Mohammadi, J. (2011). Spatial estimation of saturated hydraulic conductivity from terrain attributes using regression, kriging, and artificial neural networks. *Pedosphere*, 21(2), 170–177.

Noori, R., Abdoli, M. A., Farokhnia, A., & Abbasi, M. (2009). Results uncertainty of solid waste generation forecasting by hybrid of wavelet transform–ANFIS and wavelet transform–neural network. *Expert Systems with Applications*, 36(6), 9991–9999.

Noori, R., Khakpour, A., Omidvar, B., & Farokhnia, A. (2010a). Comparison of ANN and principal component analysis–multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications*, 37(8), 5856–5862.

Noori, R., Hoshiyaripour, G. A., Ashrafi, K., & Araabi, B. N. (2010b). Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *Atmospheric Environment*, 44(4), 476–482.

Pachepsky, Y. A., Tim Lin, D., & Varallyay, G. (1996). Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal*, 60, 727–733.

Pourreza Bilandi, Mohsen, Khasiei Sayuki, Abbas. (2015). Uncertainty analysis of the neural network output in simulating saturated hydraulic saturation of soil. *Irrigation and Drainage Journal of Iran*, 4, 9, 655–664.

Rezaei, E., Khasiei Sayuki, A., Shahidi, A. (2014). Designing groundwater monitoring network using LS-SVM minimum squared model. *Iran Water and Soil Research*, 45(4), 389–396.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., & Dassargues, A. (2011). Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences*, 25.

Sobieraj, J. A., Elsenbeer, H., & Vertessy, R. A. (2001). Pedotransfer functions for estimating saturated hydraulic conductivity. *Journal of Hydrology and Hydromechanics*, 251, 202–220.

Stumpp, C., Engelhardt, S., Hofmann, M., & Huwe, B. (2009). Evaluation of pedotransfer functions for estimating soil hydraulic properties of prevalent soils in a catchment of the Bavarian Alps. *European Journal of Forest Research*, 128, 609–620.

Wagner, B., Tarnawski, V. R., & Stöckl, M. (2004). Evaluation of pedotransfer functions predicting hydraulic properties of soils and deeper sediments. *Journal of Plant Nutrition and Soil Science*, 167, 236–245.

White, J. T., Fienen, M. N., Barlow, P. M., & Welter, D. E. (2018). A tool for efficient, model-independent management optimization under uncertainty. *Environmental Modelling & Software*, 100, 213–221.

Xiong, L., Wan, M., Wei, X., & O'Connor, K. (2009). Indices for assessing the prediction bounds of hydrological models and application by generalized likelihood uncertainty estimation. *Hydrological Sciences Journal*, 54(5).

Xu, C., Xu, X., Liu, M., Liu, W., Yang, J., Luo, W., Zhang, R., & Kiely, G. (2017). Enhancing pedotransfer functions (PTFs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern China. *Catena*, 158, 350–356.

Yao, R. J., Yang, J. S., Wu, D. H., Li, F. R., Gao, P., & Wang, X. P. (2015). Evaluation of pedotransfer functions for estimating saturated hydraulic conductivity in coastal salt-affected mud farmland. *Journal of Soils and Sediments*, 15, 1–15.

Zhao, C., Shao, M., Jia, X., Nasir, M., & Zhang, C. (2016). Using pedotransfer functions to estimate soil hydraulic conductivity in the Loess Plateau of China. *Catena*, 143, 1–6.