DESERT

Desert
Online at http://desert.ut.ac.ir

Desert 25-2 (2020) 147-154

Soil texture fractions modeling and mapping using LS-SVR algorithm

M. Jeihouni[a], S.K. Alavipanah[ab*], A. Toomanian[a], A.A. Jafarzadeh[c]

[a] *Dept. of Remote Sensing and GIS, Faculty of Geography, University of Tehran, Tehran, Iran*
[b] *Dept. of Geography, Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany*
[c] *Dept. of Soil Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran*

Received: 21 September 2019; Received in revised form: 12 February 2020; Accepted: 9 March 2020

Abstract

Soil texture is variable through space and controls most of the soil's Physico-chemical, biological and hydrological characteristics and governs agricultural production and yield. Therefore, determining its variability and generating accurate soil texture maps have a key role in soil management and sustainable agriculture. The purpose of this study is to introduce a numerical algorithm named Least Square Support Vector Machine for Regression (LS-SVR) as a predictive model in Digital Soil Mapping (DSM) of soil texture fractions and evaluating its performances based on modeling evaluation criteria. In this study, the soil texture data of 49 soil profiles in Tabriz plain, Iran, was used. The important covariates were selected using Genetic Algorithm (GA). The model evaluation results based on ME, MAE, RMSE, and R^2 indicate the high performance of LS-SVR in predicting soil texture components. The prediction RMSE for sand, silt, and clay was 6.82, 5.08 and 6.06, respectively. Silt prediction had the highest ME and the lowest MAE and RSME values. The algorithm simulated the complex spatial patterns of soil texture fractions and provided high accuracy predictions and maps. Therefore, the LS-SVR algorithm has the capability to be used as predictive models in soil texture digital mapping. This study highlighted the potential of the LS-SVR algorithm in high precision soil mapping. The generated maps can be used as basic information for environmental management and modeling.

Keywords: Digital Soil Mapping, Soil Texture, Spatial Variability, Soil Management, Soil Physical property

1. Introduction

Over the past decades, the need for soil maps and information about soil has increased. Soil maps are broadly used as basic information for ecological evaluations, land/agricultural management, and natural protections (Van der Ploeg and Vlijm, 1978; Cámara *et al*., 2017; Pinheiro *et al*., 2018). Soil texture fractions or soil mineral particle size is highly variable through space and is the most important physical property that governs soil's Physico-chemical, biological and hydrological characteristics and processes (Adhikari *et al*., 2013). It influences agricultural production, crop yield, soil fertility, and moisture retention (Silva Chagas *et al*., 2016). So, one of the major challenges in soil management and sustainable agriculture is the need for accurate soil texture maps.

The classical method of soil properties mapping was based on assigning the mean value of each soil property per map unit (polygon). But this method had many disadvantages, for example it was not practical for highly variable soil properties, needs many soil samples, time-consuming, and expensive (Pahlavan-Rad and Akbarimoghaddam, 2018). Therefore, the need for new mapping methods arises, which can provide accurate and high-resolution (pixel-based) soil properties maps. Here, the key role of Digital Soil Mapping (DSM) is highlighted as a successful technique to convert discrete observation points to a continuous surface. DSM employs field observations, laboratory measurements, digital elevation model (DEM),

* Corresponding author. Tel.: +98 912 3207202
Fax: +98 21 61112591
E-mail address: salavipa@ut.ac.ir

and satellite imagery derivates as inputs for building mathematical/statistical (quantitative) models to map spatial patterns of soil properties through the area (Minasny and McBratney, 2016). For more details and explanations on DSM readers can refer to McBratney *et al*. (2003), and Minasny and McBratney (2016).

In recent years, DSM has been used to map different soil properties such as soil organic carbon, clay, and nitrogen (Minasny *et al*., 2013; Lin *et al*., 2016; Sindayihebura *et al*., 2017) with different predictive models. Silva Chagas *et al*. (2016) used Multiple Linear Regressions (MLR) and Random Forest (RF) as predictive models to map soil texture. Pahlavan-Rad and Akbarimoghaddam (2018) employed RF as a predictive model to estimate soil texture and pH. Shahbazi *et al*. (2019) used RF to map the spatial distribution of clay. MLR and RF are used as general predictive models in DSM. But there is a need to introduce new predictive models such as machine learning algorithms to model the spatial patterns of soil properties.

Machine learning approaches such as Support Vector Machine for Regression (SVR) and Least Square Support Vector Machine for Regression (LS-SVR) have been successfully applied in many fields (Ballabio, 2009; Pasolli *et al*., 2011). LS-SVR differs from the standard SVR, it is a powerful method for solving non-linear problems and function estimation (Kumar and Kar, 2009), and this machine learning method has been used in numerous studies. LS-SVR employed for modeling of stream-flow prediction (Bhagwat and Maity, 2013, Kisi 2015a), rainfall downscaling (Pham *et al*., 2019), solar power output (Lin and Pai, 2016), drought forecasting (Deo *et al*., 2017), pan evaporation (Goyal *et al*., 2014; Kisi, 2015b) and modeling of soil moisture retention curve (Achieng, 2019). The LS-SVR has not been applied as a predictive model in DSM as of yet.

The present study's aims are: (1) assessing the potential of LS-SVR techniques as predictive model for soil texture fraction prediction, and (2) generating digital soil texture maps. This paper presents a novel application of LS-SVR as predictive models in DSM.

2. Materials and Methods

2.1. Study area and data

The study area is part of Tabriz plain in the north-west of Iran, which lies between Latitude 38° 00′ to 38° 16′ N and Longitude 45° 51′ to 46° 10′ E with the area of 424 km^2 (Figure 1). The average annual precipitation is about 280mm and the climate is semi-arid. The surface soil texture (sand, silt, and clay fractions) data of 49 soil profiles in the study area was used and the spatial distribution of sample points and their locations are shown in Figure 1. The environmental covariates as spatial coordinates, topographical wetness index (TWI), slope, elevation, and remote sensing covariates from Landsat images such as spectral bands, spectrally derived indexes, PCA and tasseled cap transformations were used as well.

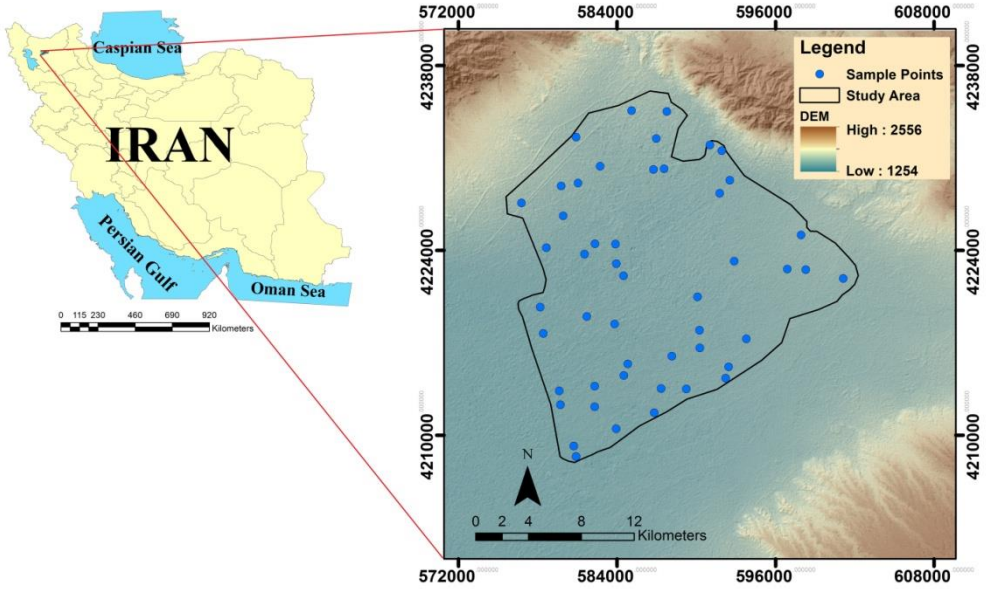

Fig. 1. Study area and locations of soil samples

2.2. LS-SVR

The LS-SVR as a developed version of the standard SVR is a powerful algorithm for solving non-linear problems and function estimation (Kumar and Kar, 2009; Kisi, 2015b). LS-SVR is a kernel-based machine learning method, first proposed by Suykens and Vandewalle (1999) and has higher computational performance than standard SVR (Bhagwat and Maity 2013). Because in training phase formulation of LS-SVR a set of linear equations should be solved instead of a quadratic programming problem in the standard SVR (Suykens and Vandewalle 1999). In solving an LS-SVR formulation one less parameter is required to optimize the model, which is the advantage of LS-SVR over SVR (Goyal *et al*., 2014).

The mathematics of the LS-SVR algorithm is represented based on Bhagwat and Maity (2013) and Kisi (2015b):

The aim is the building of function *y=f(x)*, which characterizes the dependence of the inputs x_i (independent variables) and output y_i (target variable). The form of the nonlinear function can be expressed as (1):

$$y = w^T \varphi(x) + b \tag{1}$$

Where w is weight vector, f is non-linear mapping function and b is the bias term (Cao *et al*., 2008).

Using the function estimation error, the optimization problem is defined as

$$\begin{cases} \text{Min } J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \\ \text{such that } y_i = w^T \varphi(x_i) + b + e_i \quad i = 1,...,N \end{cases} \tag{2}$$

Where γ is the regularization constant and e_i is the training random error for x_i.

To find the optimal parameters of w and e, that minimize the prediction error of the regression model, the Lagrange multiplier optimal programming method is applied to solve the Eq (2). The objective can be determined by changing the constraint problem into an unconstraint problem. The Lagrange function can be expressed as

$$L(w,b,e,\alpha) = J(w,e) - \sum_{i=1}^{N} \alpha_i \{ w^T \varphi(x_i) + b + e_i - y_i \} \tag{3}$$

Where α_i is the Lagrange multiplier.

The optimal conditions can be obtained by taking the partial derivatives of Eq. (3) with respect to *w*, *b*, *e* and α by taking into account the Karush–Kuhn–Tucker (Flecher, 1980), respectively as

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \tag{4}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \tag{5}$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \quad i = 1,...,N \tag{6}$$

$$\frac{\partial L}{\partial x_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0, \quad i = 1,...,N \tag{7}$$

From the set of Eqs. (4-7), *w* and *e* can be eliminated. Therefore the linear equations can be derived as

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8}$$

where $Y=y_1,..., y_m$, $Z= \varphi(x_1)^T y_1,..., \varphi(x_m)^T y_m$, $I=[1,...,1]$, $\alpha=[\alpha_1,..., \alpha_l]$.

By defining kernel function as Eq. (9) which is satisfied with mercer's condition.

$$K(x,x_i) = \varphi(x)^T \varphi(x_i), \quad i = 1,...,N. \tag{9}$$

The resulting LS-SVR model can be expressed as:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x,x_i) + b \tag{10}$$

The radial basis function (RBF) is a broadly used kernel function (Bhagwat and Maity, 2013; Kisi, 2015b) that was employed in the current study. The RBF kernel is defined as:

$$K(x,x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\partial^2} \right) \tag{11}$$

2.3. Feature selection

The features (important covariates) were selected using Genetic Algorithm (GA) which is a broadly used technique for feature selection (Yang *et al*., 2019) in models that lacking the interior feature section as LS-SVR. Accordingly, to improve the LS-SVR model, in

the programming and coding phases, the GA entered as an optimizer component in the main LS-SVR code to optimize the combinations of the features selected as important covariates for each soil variable prediction. After feature selection by GA, the selected features used as input covariates for implementing the LS-SVR model.

2.4. Methodology

In the present study, the RBF based LS-SVR algorithm employed as a predictive model for estimating soil texture fractions (sand, silt, and clay) using 49 samples, DEM, and Lansat5 image derivates. The accuracy of the model for each soil particle size prediction was evaluated through the leave-on-out cross-validation technique. The model performance for each soil texture component evaluated based on four statistical criteria: mean error (ME), mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination (R^2). Then the digital maps of soil texture fractions generated for each texture component.

3. Results and Discussion

3.1. Descriptive statistics

Descriptive statistics of soil texture fractions are presented in Table 1. The mean sand, silt, and clay fractions (%) were 25.35, 45.25, and 29.40%, respectively. Sand fractions ranged from 5 to 81.20%, silt ranged from 8 to 65.60%, and clay ranged from 5.10 to 64.55%. Sand has the widest range compared to silt and clay. The distribution of soil texture classes was plotted on a USDA texture triangle diagram, which is graphically shown in Figure 2. The soil samples had different soil texture classes range from loamy sand to clay. The soil texture classes in the study area are loamy sand, sandy loam, loam, silt loam, clay loam, silty clay loam, silty clay, and clay. With respect to the soil texture diagram, soil texture classes in the study area are mostly distributed in loamy classes.

Table 1. Descriptive statistics of soil soil texture fractions; sand, silt and clay

Variable	Min	Max	Mean	Median	SD
Sand (%)	5.00	81.20	25.35	24.60	16.28
Silt (%)	8.00	65.60	45.25	45.30	10.23
Clay (%)	5.10	64.55	29.40	28.40	13.67

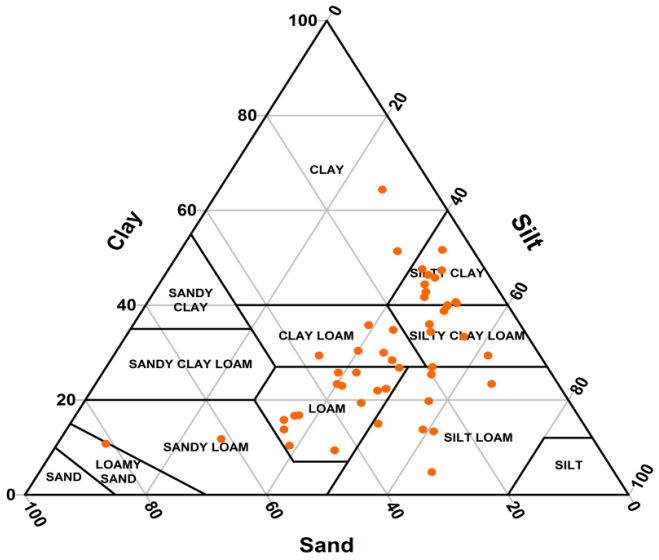

Fig. 2. Soil texture triangle diagram with a sampled texture classes

3.2. Feature selection and model performance

To predict soil texture fractions, first the most important features defined for each soil variable based on GA. Table 2 indicates the selected features for each variable. The selected features used as important covariates to predict the soil texture fractions. During the feature selection phase; Band 5, Band7, wetness, and X coordinate were selected as important covariates for predicting all soil texture fraction (sand, silt, and clay). However, other covariates were selected for predicting each soil variable as well. Silt prediction with five important

covariates (i.e., Band 5, Band7, greenness, wetness, and X coordinate) had the minimum number of important covariates. Sand and clay were used 13 covariates for predicting each variable. In all predictions, spatial coordinates were marked as important covariates in modeling soil texture fraction. This indicates that LS-SVR can model the spatial trends of soil texture fractions and relate the spatial dependence with point's spatial coordinates.

Table 2. The selected covariates for soil texture prediction

	Selected covariates
Sand	B1, B3, B4, B5, B7, PC2, PC3, PC4, PC5, PC6, Wetness, X, Y
Silt	B5, B7, Greenness, Wetness, X
Clay	B1, B4, B5, B7, PC1, PC2, PC4, PC5,Brightness, Greenness, Wetness, X, Y

B1: band1; B2: band2; B3: band2; B4: band4; B5: band5; B7: band7; X: X coordinates; Y: Y coordinates

The LS-SVR model was used for predicting the soil texture fractions based on the input combinations as indicated in Table 2. The models' performance evaluation results are presented in Table 3. The ME values were ranged from -2.73 to -2.47 that belonged to silt and clay, respectively. All predictions had negative ME which indicates that all prediction were underestimated the observed values. The maximum and minimum MAE was observed for sand and silt with values of 4.58 and 3.49, respectively. The LS-SVR had predicted sand, silt and clay with the RMSE values of 6.82, 5.08, and 6.06, respectively.

The RMSE and MAE values for sand prediction were high because sand content has a wide range compared to silt and clay in the study area. Conversely, the silt predictions had lowest RMSE and MAE values regarding its narrow range.

The scatter plots of LS-SVR for observed and predicted soil particle fractions are shown in Fig. 3. The plots indicate the relations between predicted and observed values. Based on the models' R^2 values, the algorithm was modelled the soil texture components with R^2 of 0.86, 0.83, and 0.84 for sand, silt, and clay, respectively. In all predictions, the regression lines were close to 1:1 lines which indicate the models' accuracy. Regarding models performance results presented in Table 3 and Fig. 3, the LS-SVR algorithm had high performance and predictive capability in the prediction of all soil texture fractions and provided accurate predictions.

Table 3. Performance evaluation of the model for soil texture fractions

Evaluation criteria	Parameter		
	Sand	Silt	Clay
ME	-2.62	-2.73	-2.47
MAE	4.58	3.49	4.50
RMSE	6.82	5.08	6.06
R^2	0.86	0.83	0.84

ME: mean error; MAE: mean absolute error; RMSE: root mean squared error; R^2: coefficient of determination

3.3. Digital soil maps

The soil texture fractions (sand, silt, and clay) distribution maps were generated and presented in Fig. 4. Regarding the generated prediction maps, the north, north-west, and west of the area have a high fraction of sand and the least sand amounts were predicted in the central, eastern, and southern parts of the area. The highest amounts of silt are identified in the north, center, and south parts of the area, where the low amounts are highlighted in the west and north-western parts. The clay map indicates a high amount of clay in the north-east, east, center, and south-east of the area. The areas with lower clay content were located in the north.

The LS-SVR model could find the general trends in spatial distributions of soil texture fractions particles in the study area. The high sand content in the north and north-west of the area is regarding the proximity to the mountainous area. The higher silt and clay contents are predicted in eastern, central, and southern areas which are the lowest parts of the Tabriz plain.

The results indicate that the LS-SVR has high performance and could model the complex spatial patterns of soil texture fractions by providing high accuracy predictions and maps. Accordingly, the LS-SVR algorithm has the potential to be used as predictive models in DSM for soil texture fractions mapping.

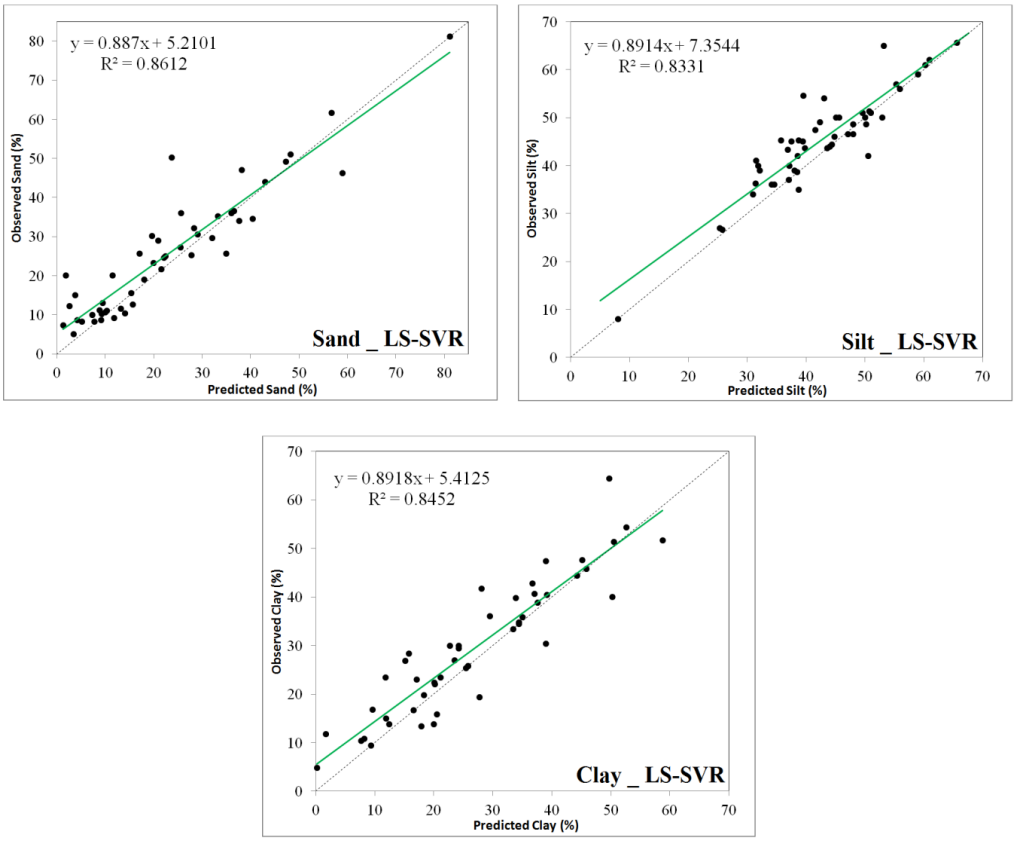

Fig. 3. Scatter plots between the observed and estimated values for each soil particle size using the LS-SVR model

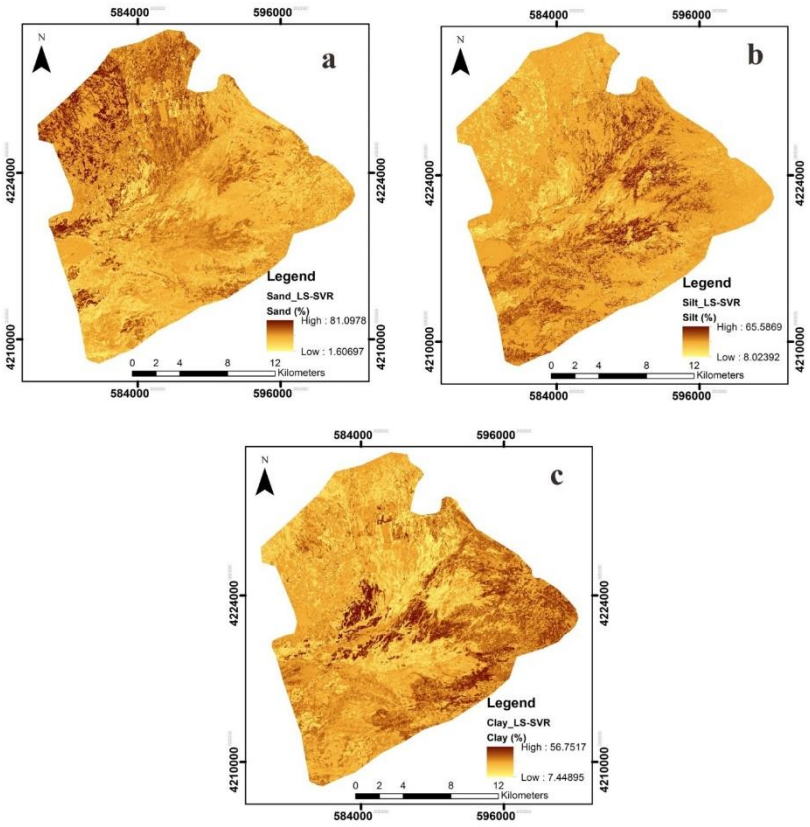

Fig. 4. The generated soil texture fractions maps by LS-SVR algorithm. Sand (a), silt (b) and clay (c)

4. Conclusion

The goals of this study were to investigating the potential of a numerical algorithm named LS-SAVR as a predictive model in digital mapping of soil texture fractions (soil mineral particle size), and generating soil texture distribution maps using this model in a semi-arid environment. The GA used to select important covariates. According to the results, remote sensing covariates were successfully used in the digital mapping of soil texture fractions. The validation results in terms of ME, MAE, RMSE, and R^2, indicate the high accuracy and capability of LS-SVR in predicting and mapping of soil texture fractions. Moreover, this algorithm could model the complex spatial patterns of soil texture components using spatial coordinates. This study highlights the potential of the LS-SVR algorithm in the DSM of soil texture. The LS-SVR model is recommended for the digital mapping of soil texture, which has a key role in soil management and sustainable agriculture. Finally, the generated high-resolution digital soil texture maps as a result of this study can be taken into account by the ministry of agriculture for appropriate decision making in land/agricultural management and natural protections.

References

Achieng, K. O., 2019. Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. Computers & Geosciences, 133, 104320.

Adhikari, K., R.B. Kheir, M.B.Greve, P.K.Bøcher, B.P. Malone, B. Minasny, A.B. McBratney, M.H. Greve, 2013. High-resolution 3-D mapping of soil texture in Denmark. Science Society of America Journal, 77(3); 860-876.

Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma, 151(3-4); 338-350.

Bhagwat, P.P. R. Maity, 2013. Hydroclimatic streamflow prediction using least square-support vector regression. ISH Journal of Hydraulic Engineering, 19(3); 320-328.

Cámara, J., V. Gómez-Miguel, M.Á. Martín, 2017. Lithologic control on soil texture heterogeneity. Geoderma, 287; 157-163.

Cao, S. G., Y. B. Liu, Y. P. Wang, 2008. A forecasting and forewarning model for methane hazard in working face of coal mine based on LS-SVM. Journal of China University of Mining and Technology, 18(2); 172-176.

Deo, R. C., O. Kisi, V. P. Singh, 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmospheric Research, 184; 149-175.

Fletcher, R., 1980. Practical Methods of Optimization: Vol. 1 Unconstrained Optimization. John Wiley & Sons.

Goyal, M.K., B. Bharti, J. Quilty, J. Adamowski, A. Pandey, 2014. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert systems with applications, 41(11); 5267-5276.

Kisi, O., 2015a. Streamflow forecasting and estimation using least square support vector regression and adaptive neuro-fuzzy embedded fuzzy c-means clustering. Water Resources Management, 29(14); 5109-5127.

Kisi, O., 2015b. Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. Journal of Hydrology, 528; 312-320.

Kumar, M., I.N. Kar, 2009. Non-linear HVAC computations using least square support vector machines. Energy Conversion and Management, 50(6); 1411-1418.

Lin, K. P., P. F. Pai, 2016. Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. Journal of Cleaner Production, 134; 456-462.

Lin, Y., S.E. Prentice III, T. Tran, N.L. Bingham, J.Y. King, O.A. Chadwick, 2016. Modeling deep soil properties on California grassland hillslopes using LiDAR digital elevation models. Geoderma regional, 7(1); 67-75.

McBratney, A.B., M.M. Santos, B. Minasny, 2003. On digital soil mapping. Geoderma, 117(1-2); 3-52.

Minasny, B. A.B. McBratney, 2016. Digital soil mapping: A brief history and some lessons. Geoderma, 264; 301-311.

Minasny, B., A.B. McBratney, B.P. Malone, I. Wheeler, 2013. Digital mapping of soil carbon. Advances in Agronomy, 118; 1-47.

Pahlavan-Rad, M.R. A. Akbarimoghaddam, 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). Catena, 160; 275-281.

Pasolli, L., C. Notarnicola, L. Bruzzone, 2011. Estimating soil moisture with the support vector regression technique. IEEE Geoscience and remote sensing letters, 8(6); 1080-1084.

Pham, Q. B., T. C. Yang, C. M. Kuo, H. W. Tseng, P. S. Yu, 2019. Combing random forest and least square support vector regression for improving extreme rainfall downscaling. Water, 11(3); 451.

Pinheiro, H.S.K., W.D. Carvalho Junior, C.D.S. Chagas, L.H.C.D. Anjos, P.R. Owens, 2018. Prediction of topsoil texture through regression trees and multiple linear regressions. Revista Brasileira de Ciência do Solo, *42*. doi.org/10.1590/18069657rbcs20170167

Shahbazi, F., A. McBratney, B. Malone, S. Oustan, B. Minasny, 2019. Retrospective monitoring of the spatial variability of crystalline iron in soils of the east shore of Urmia Lake, Iran using remotely sensed data and digital maps. Geoderma, 337; 1196-1207.

Silva Chagas, C. D., W. Carvalho Junior, S.B. Bhering, B. Calderano Filho, 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena, 139; 232-240.

Sindayihebura, A., S. Ottoy, S. Dondeyne, M. Van Meirvenne, J. Van Orshoven, 2017. Comparing digital soil mapping techniques for organic carbon and clay content: Case study in Burundi's central plateaus. Catena, 156; 161-175.

Suykens, J.A., J. Vandewalle, 1999. Least squares support vector machine classifiers. Neural processing letters, 9(3); 293-300.

Van der Ploeg, S.W.F., L. Vlijm, 1978. Ecological evaluation, nature conservation and land use planning with particular reference to methods used in the Netherlands. Biological conservation, 14(3); 197-221.

Yang, Y., R. A. V. Rossel, S. Li, A. Bissett, J. Lee, Z. Shi, L. Court, 2019. Soil bacterial abundance and diversity better explained and predicted with spectro-transfer functions. Soil Biology and Biochemistry, 129; 29-38.