



Article

Spatiotemporal Assessment of Soil Organic Carbon Change Using Machine-Learning in Arid Regions

Hassan Fathizad ^{1,*}, Ruhollah Taghizadeh-Mehrjardi ^{2,3,4,*} , Mohammad Ali Hakimzadeh Ardakani ¹ ,
Mojtaba Zeraatpisheh ^{5,6} , Brandon Heung ⁷ and Thomas Scholten ^{2,3,4}

- ¹ Department of Arid and Desert Regions Management, School of Natural Resources & Desert Studies, Yazd University, Yazd 89195741, Iran; hakim@yazd.ac.ir
- ² Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, 72070 Tuebingen, Germany; thomas.scholten@uni-tuebingen.de
- ³ CRC 1070 Resource Cultures, University of Tübingen, Gartenstraße 29, 72070 Tuebingen, Germany
- ⁴ DFG Cluster of Excellence “Machine Learning”, University of Tübingen, 72074 Tuebingen, Germany
- ⁵ Henan Key Laboratory of Earth System Observation and Modeling, Henan University, Kaifeng 475004, China; mojtaba.zeraatpisheh@henu.edu.cn
- ⁶ College of Geography and Environmental Science, Henan University, Kaifeng 475004, China
- ⁷ Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Halifax, NS B3H 4R2, Canada; brandon.heung@dal.ca
- * Correspondence: h.fathizad@stu.yazd.ac.ir (H.F.); ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de (R.T.-M.)



Citation: Fathizad, H.; Taghizadeh-Mehrjardi, R.; Hakimzadeh Ardakani, M.A.; Zeraatpisheh, M.; Heung, B.; Scholten, T. Spatiotemporal Assessment of Soil Organic Carbon Change Using Machine-Learning in Arid Regions. *Agronomy* **2022**, *12*, 628. <https://doi.org/10.3390/agronomy12030628>

Academic Editors: Enrico Borgogno Mondino and Emanuele Lugato

Received: 10 January 2022

Accepted: 1 March 2022

Published: 4 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Soil organic carbon (SOC) is an essential property of soil, and understanding its spatial patterns is critical to understanding vegetation management, soil degradation, and environmental issues. This study applies a framework using remote sensing data and digital soil mapping techniques to examine the spatiotemporal dynamics of SOC for the Yazd-Ardakan Plain, Iran, from 1986 to 2016. Here, a conditioned Latin hypercube sampling method was used to select 201 sampling sites. A set of 37 environmental predictors were obtained from Landsat imagery taken in 1986, 1999, 2010 and 2016. Here, SOC was modeled for 2016 using the Random Forest (RF), support vector regression (SVR), and artificial neural networks (ANN) machine-learners by correlating environmental predictors with soil data. The results showed that RF yielded the highest accuracy ($R^2 = 0.53$), compared to the other two learners. By performing a variable importance analysis of the RF model, normalized difference vegetation index, modified vegetation index, and ground-adjusted vegetation index were determined to be the most important environmental predictors. By applying the model calibrated from 2016 data to 1986, 1999 and 2010, the results showed a substantial decrease in SOC; these decreases in SOC were mainly attributed to land use changes and agricultural activities.

Keywords: random forest; machine learning; spatial distribution; variable importance analysis; vegetation index; temporal change

1. Introduction

Soil organic carbon (SOC) has a significant impact on many soil functions, such as the production of food and other biomass; and the provisioning of biological habitats and genetic resources. It is an important indicator for assessing and managing soil fertility, soil quality, and soil degradation [1,2]; hence, accurate information on the spatiotemporal patterns of SOC are required to support sustainable land use and management. Furthermore, information on the spatiotemporal variability of SOC is particularly important in the context of climate change at the local- [3], regional- [4], and global-scales [5]. As a result, methods for measuring, modelling, and monitoring SOC are continuously evolving around the world; however, methods of direct sampling and soil analysis using laboratory or field measurements are generally expensive and time-consuming to perform and therefore impractical to monitor the SOC changes over large spatial extents [6–8].

The recent development of digital soil mapping methods (DSM) provides a framework for characterizing the spatiotemporal patterns of soil properties [9]. DSM approaches involve the creation and operation of terrestrial spatial information systems obtained from field and laboratory observations, combined with spatial and non-spatial inference systems to generate raster-based map prediction and their respective uncertainty estimates [10–12]. These approaches apply statistical tools to quantify the relationships between soil properties and environmental predictors [8]. Compared to traditional approaches of soil mapping, DSM methods provide a more accurate representation of soil variability [7,11] and provide a suite of map products for supporting decision-making processes that are designed to address agricultural and other environmental issues [8]. When developing predictive models of SOC, previous studies have applied the use of machine learners, such as artificial neural networks (ANN), Cubist model trees [13,14], decision trees [15], support vector regression (SVR), and geostatistical approaches via kriging, inverse distance weighted interpolation, and spline approaches [16–18]. Additionally, the Random Forest (RF) learner has become increasingly popular in DSM research [13,19,20].

Many studies have predicted SOC using DSM techniques for a single point in time, but do not account for its temporal dynamics in models [6,13,21,22]. To address this issue, Fathizad et al. [23] provided a framework, which involved training and validating their predictive models using current data to map soils (i.e., soil salinity) and applied the model to historical environmental predictors acquired from remote sensing [23]. A similar approach has been applied in Taghizadeh-Mehrjardi et al. [24] to map historical patterns of heavy metal in soil samples from Iran [24].

The arid and semi-arid regions of Iran, as well as its deserts, are vulnerable ecosystems that are inhabited by large populations. These landscapes cover most of the terrestrial land-base of Iran; hence, methods for monitoring SOC in these areas are particularly important—especially given the presently low concentrations of SOC in these regions. Because there are considerable environmental challenges of conducting field studies in these areas of Iran, there is very little spatial information about soils and how they change over time. Hence, the objectives of this study were (1) to compare the performance of three ML models (RF, SVR, and ANN) to predict the spatial distribution of SOC and (2) to study the changes of SOC from 1986 to 2016 in the Yazd-Ardakan plain.

2. Materials and Methods

The methodological framework that was used to carry out this study is summarized in Figure 1.

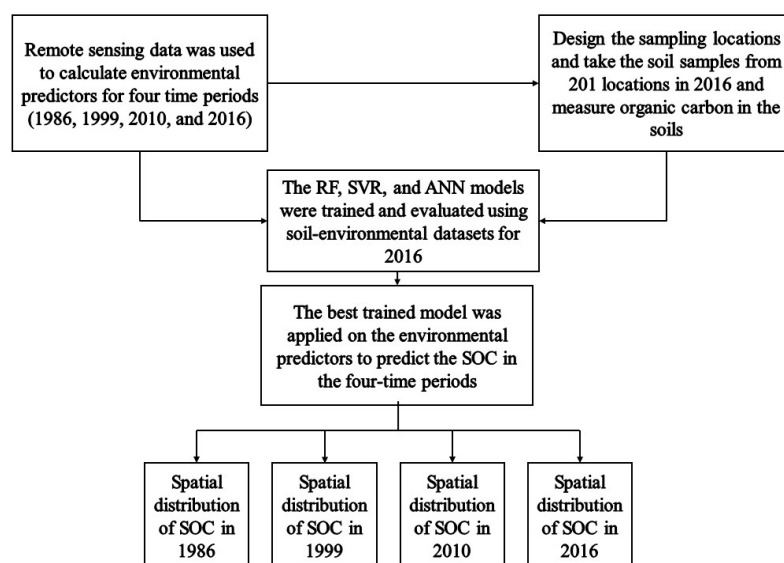


Figure 1. A methodological framework for predicting the spatiotemporal dynamics of soil organic carbon (SOC) using Random Forest (RF), support vector regression (SVR), and artificial neural networks (ANN).

2.1. Study Area & Soil Sampling

The Yazd-Ardakan Plain is located on the central plateau of Iran and in the central part of Yazd Province ($53^{\circ}45'14.6''$ E to $54^{\circ}48'18.1''$ E longitude; and $31^{\circ}51'0.6''$ N to $32^{\circ}26'37.5''$ N longitude) and has an area of 4829 km^2 . The study area has an elevation range of 977 m to 2684 m above mean sea level; the rainfall in this area is low and irregular (about 118 mm/year), and the evaporation rate ranges from 2200 to 3200 mm/year. According to the United States Classification of Soil Classification, the soil in the area is mainly described as Entisols and Aridisols soil. Geologically, the Yazd-Ardakan Plain is mainly composed of intrusive rocks, especially Neogene formations, and is covered by Quaternary alluvial deposits and agglomerations. The geological formations in the study area are varied with the youngest formations originating from the Quaternary period, which covers most of the territory.

This study used a conditioned Latin hypercube sampling method to select 201 sampling locations [25]. A global positioning system was used to find the geographic location of the sampling points, where soil samples were taken from the topsoil layer (i.e., 0–20 cm depth increment; Figure 2). SOC was measured using the wet oxidation (combustion) method [26].

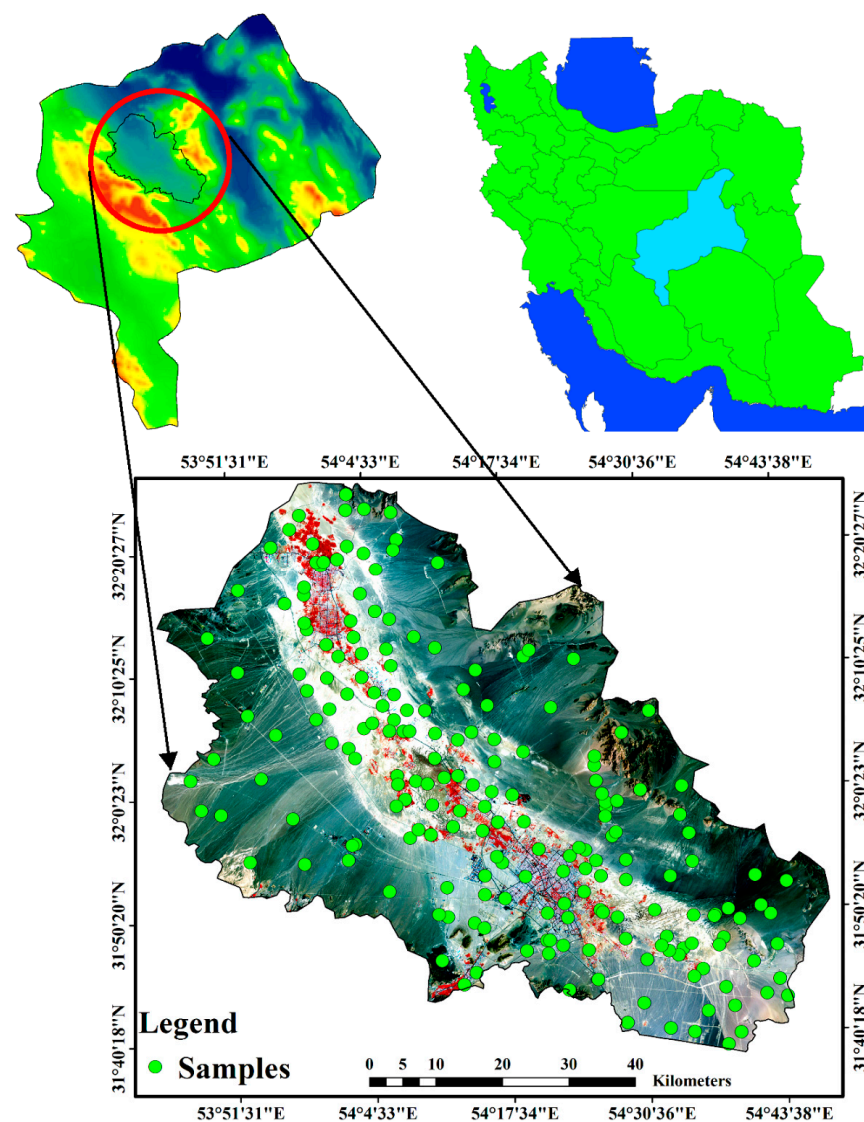


Figure 2. Location of the Yazd-Ardakan Plain and sampling points overlaid on a false color composite Landsat image.

2.2. Environmental Predictors

Satellite images obtained by Landsat 5 in 1986, Landsat 5 in 1999, Landsat 7 in 2010, and Landsat 8 in 2016 were used to calculate vegetation indices and other remote sensing indices. After carrying out preprocessing of the satellite images, a total of 37 vegetation indices were calculated. These indices included Normalized Differential Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Ratio Vegetation Index (RVI), Distinct Vegetation Index (DVI), and the Principal Components (PC) of the spectral bands. A complete list of the vegetation indices used in this study is presented in Table 1.

Table 1. List of environmental predictors derived from Landsat imagery.

Covariates	Definition	Reference
Difference vegetation index (DVI)	$\text{NIR} - \text{Red}$	[27]
Enhanced vegetation index (EVI)	$G \times (\text{NIR} - \text{Red}) / (\text{NIR} + c1 \times \text{Red} - c2 \times \text{Blue} + L)$	[28]
Global vegetation index (GVI)	$-0.29 \times (G) - 0.56 \times (\text{Red}) + 0.6 (\text{IR}) + 0.49 (\text{NIR})$	[29]
Infrared percentage vegetation index (IPVI)	$\text{NIR} / (\text{NIR} + \text{Red})$	[30]
Normalized difference vegetation index (NDVI)	$(\text{Red} - \text{NIR}) / (\text{Red} + \text{NIR})$	[31]
Blue	Reflectance value of Landsat satellite band	Landsat satellite
Green	Reflectance value of Landsat satellite band	Landsat satellite
Red	Reflectance value of Landsat satellite band	Landsat satellite
Near-infrared (NIR)	Reflectance value of Landsat satellite band	Landsat satellite
Shortwave infrared (SWIR)	Reflectance value of Landsat satellite band	Landsat satellite
Principal components of Landsat bands	PC1, PC2, PC3, and PC4	[32]
Normalized-NDVI	$(\text{NIR} - (\text{TM1} + \text{Green})) / (\text{NIR} + (\text{TM1} + \text{Green}))$	[31]
Optimized soil-adjusted vegetation index (OSAVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.16)$	[33]
PD 311	$\text{Red} - \text{TM1}$	[34]
PD 312	$(\text{Red} - \text{Blue}) / (\text{Red} + \text{TM1})$	[34]
PD 321	$\text{Red} - \text{Green}$	[34]
PD 322	$(\text{Red} - \text{Green}) / (\text{Red} + \text{Green})$	[34]
Ratio-Based	$\text{NIR} / (\text{Blue} + \text{Green})$	[31]
Ratio vegetation index (RVI)	$(\text{NIR} / \text{Red})$	[35]
Soil-adjusted vegetation index (SAVI)	$[(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + L)] \times (1 + L)$	[28]
Stress-related	$(\text{TM1} \times \text{Green}) / \text{Red}$	[31]
Transformed vegetation index (TVI)	$(\text{SWIR} - \text{Red}) / (\text{SWIR} + \text{Red})$	[36]
VIT01	$\text{Red} / \text{Thermal}$	[37]
VIT02	$\text{Thermal} / (\text{Red} + \text{SWIR})$	[37]
VIT03	$\text{Thermal} / \text{Red}$	[37]
VIT04	$\text{Thermal} / (\text{SWIR} + \text{Green})$	[37]
Brightness index	$\text{BI} = ((\text{Red} \times \text{Red}) + (\text{NIR} \times \text{NIR}))^{0.5}$	[38]
Normalized difference moisture index (NDMI)	$\text{NDMI} = (\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR})$	[39]
Normalized difference snow index (NDSI)	$\text{NDSI} = (\text{Red} - \text{NIR}) / (\text{Red} + \text{NIR})$	[40]
Salinity index ₁ (S ₁)	$S1 = \text{Blue} / \text{Red}$	[41]
Salinity index ₂ (S ₂)	$S2 = (\text{Blue} - \text{Red}) / (\text{Blue} + \text{Red})$	[42]
Salinity index ₃ (S ₃)	$S3 = (\text{Green} \times \text{Red}) / \text{Blue}$	[42]
Salinity index ₄ (S ₄)	$S4 = (\text{Blue} \times \text{Red}) / \text{Green}$	[42]
Salinity index ₅ (S ₅)	$S5 = (\text{Red} \times \text{NIR}) / \text{Green}$	[42]
Salinity index ₆ (S ₆)	$S6 = (\text{Blue} \times \text{Red})^{0.5}$	[38]
Salinity index ₇ (S ₇)	$S7 = (\text{Green} \times \text{Red})^{0.5}$	[43]
Salinity index ₈ (S ₈)	$S8 = (\text{Blue}^2 \times \text{Green}^2 \times \text{Red}^2)^{0.5}$	[43]

In addition, Figure 3 shows examples of the spatial distribution of selected vegetation indices that were used to predict SOC for the study area. These indices were calculated using the ArcGIS 10, Idrisi Selva, and ENVI 4.8 software. To evaluate the relationships

between remote sensing data and SOC, a Pearson correlation analysis was performed between SOC and each predictor.

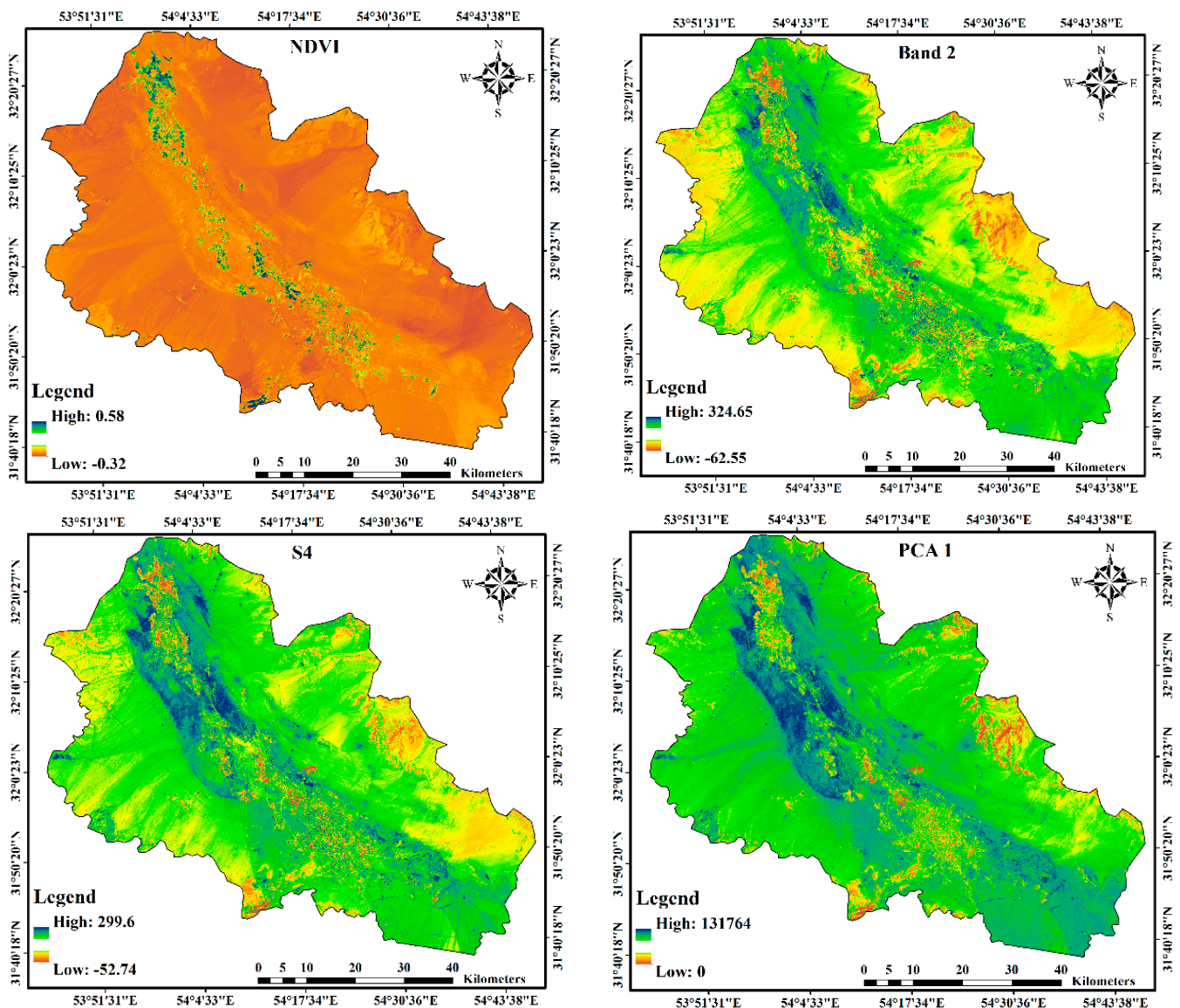


Figure 3. Map of environmental predictors used for predicting SOC in the Yazd-Ardakan Plain (see Table 1 for codes).

2.3. Machine Learning

In this study, the three ML models (RF, SVR, and ANN) were fitted and validated using the 2016 data. To evaluate the historical changes in the SOC, the best fitting ML model was applied to the remote sensing data collected for the periods 1986, 1999, and 2010.

2.3.1. Random Forest (RF)

Although a large variety of ML techniques have been tested in DSM [20], the RF learner [44] has recently seen great utility in Iran for predicting the spatial patterns of soils [20,24,45,46]. The RF learner is an example of an ensemble model, which is based on a unique set of CART-like decision tree models developed from a random bootstrap sample of the training dataset. The use of an ensemble model aims to minimize the effect of overfitting the model, which is a common problem with hierarchical, nonlinear models. Furthermore, additional randomness is introduced into the model whereby a random subset of the predictors is tested when generating the node-splitting rules at each node.

Here, the goal of the node-splitting rules is to maximize the uniformity within the nodes and the heterogeneity between the nodes with respect to the training data. When creating a set of individual decision trees, the trees are aggregated into a single predictive model. The main hyperparameters in the RF model include n_{tree} , which specifies the number of decision trees within the ensemble; and m_{try} , which specifies the number of predictors that are tested when generating each node-splitting rule. The default settings for these hyperparameters are $n_{tree} = 1000$ trees and $m_{try} = p^{0.5}$, where p is the number of predictors. Based on these two parameters, decision trees are grown as large as possible and without pruning [20,25]. Since not all training observations are used to generate each decision tree, the out-of-bag samples can be used to perform a permutation-based, variable importance analysis and calculate the percent increase in mean square error (%IncMSE). A higher %IncMSE represents greater variable importance.

2.3.2. Support Vector Regression (SVR)

Support vector machines were proposed in the late 1960s by Vapnik and Lerner [47]. This supervised learning approach may be used for classification and regression purposes to perform dichotomy classification of multidimensional feature-vectors [47,48]. Support vector machines were originally created for classification purposes, where it seeks an optimal hyperplane to ensure the largest margin between classes, resulting in a higher likelihood of generalization. In regression (i.e., SVR), the model looks for a function that meets the error criteria, ignoring points near and distant from the decision boundary. These are the “low error” sites, with minimal residuals. Points outside the margin are permitted with a penalty weight. The penalty balances the effect of outliers by allowing points outside the regression function [48]. The ability of SVR to generalize is dependent on the tuning of hyperparameters.

2.3.3. Artificial Neural Networks (ANN)

Artificial neural networks (ANN) are highly adaptable computer networks that may be used to model complicated nonlinear interactions between variables [49]. The algorithm is based on a set of algorithm fitting functions that make no assumptions regarding error distribution [49,50]. When applied to large regions, it may provide the benefit of abstraction. An ANN model is developed in three stages: data production, optimum configuration selection, and validation on an independent data set.

2.4. Accuracy and Uncertainty Assessment

To evaluate the performance of the RF, SVR, and ANN models for SOC prediction, a 10-fold cross-validation procedure was used. The measures of accuracy used in this study included mean absolute error (MAE), coefficient of determination (R^2), and root mean square error (RMSE). These indicators are formulated as follows:

$$MAE = \frac{\sum_{i=1}^n |X_i - Y_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^n (X_i^* - Y_i^*)^2}{\sum_{i=1}^n (X_i - Y_i)^2} \quad (3)$$

where X_i and Y_i correspond to the measured and predicted values, respectively; and X_i^* and Y_i^* correspond to the common of the measured and predicted values, respectively. The modeling and validation were carried out by using the *randomForest* and *caret* packages of the R 3.5.1 statistical software.

To assess the uncertainty of the three models, a leave-one-out cross-validation method was used. This method resulted in 201 predicted SOC maps. Based on the predicted maps,

the mean and standard deviation (SD) of SOC for each pixel were calculated. Then, the proportion of measured SOC that fell within the 90% prediction interval (i.e., prediction interval coverage probability; PICP) and mean prediction interval (MPI: upper prediction limit minus the lower prediction limit) were calculated to measure the quality of the uncertainty estimates.

3. Results and Discussion

3.1. Summary Statistics

A summary of the SOC data is shown in Figure 4. The minimum and maximum amount of SOC in the study area were 0.02% and 1.01%, respectively. Overall, the SOC contents had a mean value of 0.32% with a median of 0.28%, indicating that the study area had a low SOC. Furthermore, the low coefficient of variation of 1.24, indicates low spatial variability in SOC, with most of the study area having SOC values $> 0.5\%$. These results were consistent with another study that found similarly low SOC in the Herat Plain in the Yazd Province of Iran [50].

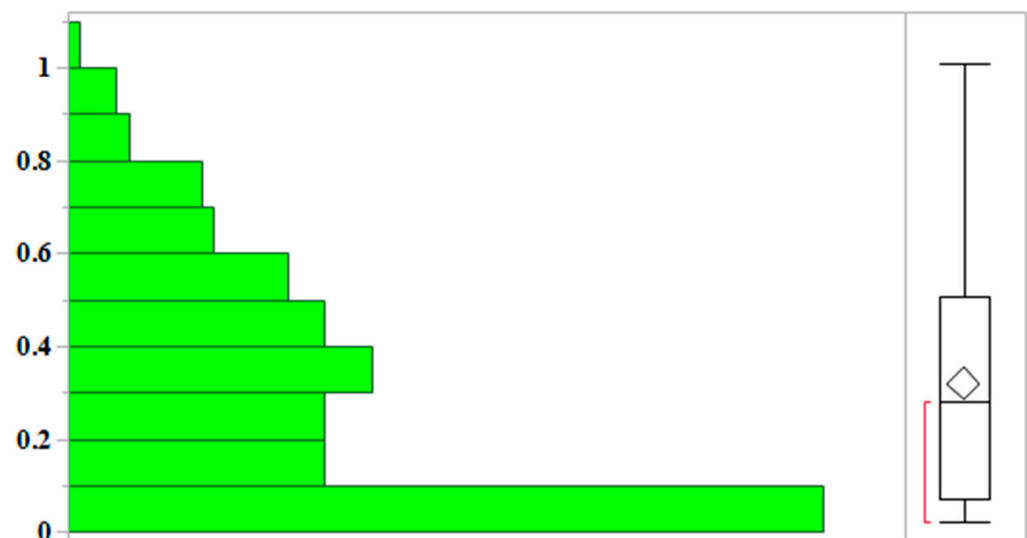


Figure 4. Distribution of SOC contents of the Yazd-Ardakan Plain.

The relationships between the remote sensing and SOC data were evaluated using Pearson's correlation coefficient analysis (Table 2). Here, 16 of the 37 environmental predictors were identified to be significantly correlated with SOC contents at p value < 0.01 , while the other four predictors were significantly correlated at p value < 0.05 . Several studies around the world have also reported a high correlation between SOC content and remote sensing predictors [13,50,51].

3.2. Accuracy and Uncertainty Assessments

The SVR, RF, and ANN models were tested using a 10-fold cross-validation process to model SOC. The evaluation results of the models are presented in Table 3. The results showed higher accuracy of the RF model ($R^2 = 0.54$; $RMSE = 0.08\%$; $MAE = 0.06$) compared to SVR and ANN. Furthermore, we assessed the uncertainty of the models using PICP and MPI (Table 3). Theoretically, 90% of the observations should fall within the defined prediction interval with a confidence level of 90% while the MPI should be as narrow as possible. The results indicated that RF achieved the highest PICP (81%) and the lowest MPI (0.14) in predicting SOC, compared to the SVR and ANN models. Similarly, Pahlavan-Rad et al. [19] applied the RF model for SOC predictions for the low relief landscapes of eastern Iran and reported an $RMSE = 0.16\%$ and $MAE = 0.21$. Other studies have also described the effectiveness of RF learners as being dependable in predicting SOC and other soil properties [20,52]. Therefore, the RF model was selected for predicting the spatiotemporal patterns of SOC.

Table 2. Pearson's Correlation Coefficient (PCC) between environmental predictors and soil organic carbon content (codes refer to Table 1).

Index	PCC	Index	PCC
CTVI	0.37 **	S6	−0.09
DVI	0.38 **	S7	−0.07
EVI	−0.37 **	S8	0.27 **
GVI	0.33 **	SAVI	0.37 **
IPVI	−0.11	Stress-related	−0.04
NDSI	−0.37 **	TTVI	0.37 **
NDVI	0.37 **	TVI	0.37 **
NIR	0.36 **	VIT01	−0.06
OSAVI	0.37 **	VIT02	−0.14 *
PD311	−0.09	VIT03	0.12
PD312	−0.12	VIT04	−0.08
PD321	−0.15 *	PCA1	−0.16 *
PD322	−0.20 **	PCA2	0.03
Ratio Based	−0.08	PCA3	−0.31 **
RVI	−0.37 **	PCA4	0.26 **
S1	0.13	Band1	−0.11
S2	0.12	Band2	−0.10
S3	−0.08	Band3	−0.08
S4	−0.14 *	Band4	−0.10
S5	0.07		

** Correlation is significant at p value < 0.01 (2-tailed). * Correlation is significant at p value < 0.05 (2-tailed).

Table 3. Accuracy and uncertainty assessments of models to predict SOC.

Model	R^2	RMSE	MAE	MPI	PICP
RF	0.536	0.082	0.058	0.14	81
MLP	0.501	0.142	0.110	0.19	75
SVR	0.407	0.181	0.134	0.21	67

3.3. Variable Importance Analysis

To further understand the soil-environmental relationships, variable importance analysis was carried out on the RF model (Figure 5). The analysis showed that NDVI was the most effective predictor of SOC with %IncMSE = 11.11%. In addition, CTVI, SAVI, and S3 were also highly ranked variables with a %IncMSE of 10.55%, 8.54%, and 6.77%, respectively. Similar studies showed that vegetation indices could increase the accuracy of modeling when predicting SOC in northern Iran [50]. In addition, the identification of NDVI as the most important predictor was also consistent with several other studies [46,50,53]. Similarly, Falahatkar et al. (2016) [54] also found that remote sensing indices are generally good predictors of SOC content, while Hong et al. (2002) [55] also showed that PCA2 and PCA4 were strongly correlated with soil chemical properties, such as soil organic matter. In contrast, Gomez et al. (2008) show that the SOC content was not related to the NDVI [56].

Because SOC directly influences soil color and its reflectance, remote sensing data, such as NDVI, CTVI and SAVI can characterize SOC variability, especially in undisturbed ecosystems [51]. Furthermore, soil salinity, which is influenced by agricultural expansion, land use, and the land cover type, can directly impact SOC input and turnover and may be characterized using the salinity indices used in this study. Thus, among the environmental covariates, three soil salinity indices (S3, S4, S5) were effective in explaining the variability of SOC.

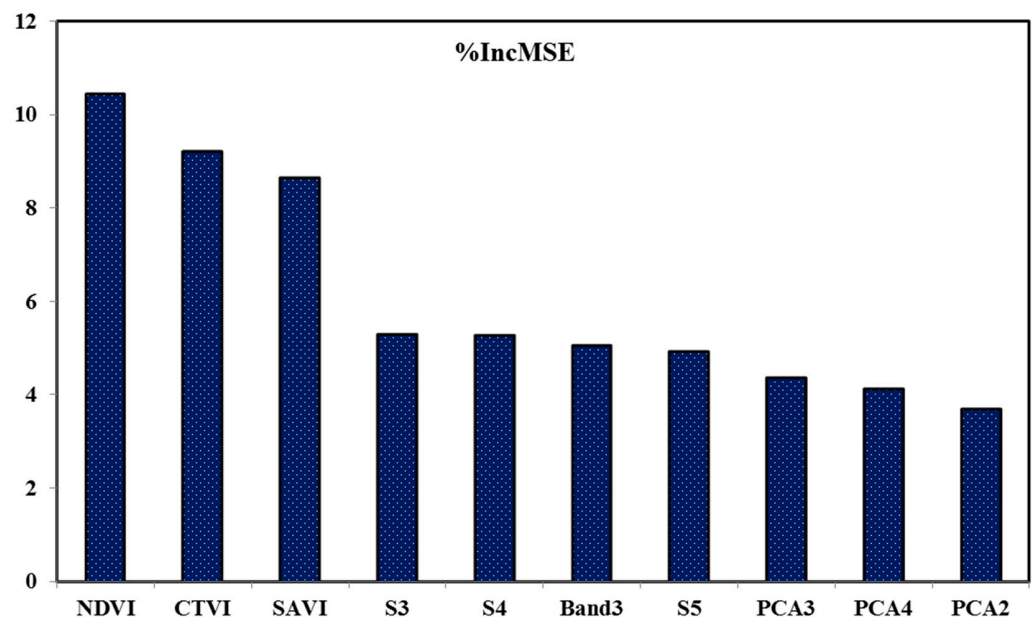


Figure 5. Variable importance analysis using the percent increase in mean square error (%IncMSE) for predicting soil organic carbon content using the Random Forest model.

3.4. Soil Organic Carbon Trends for 1986, 1999, 2010, and 2016

After selecting the fitted RF model as the best model, a SOC map was generated for 2016 at a 30 m spatial resolution for the Yazd-Ardakan Plain in Iran. Using historical remote sensing data, the RF model generated from the 2016 data was then applied to the 1986, 1999, and 2010 datasets (Figure 6). To evaluate the spatial-temporal changes in SOC and aid in interpreting the results, the SOC maps were reclassified into three classes in Figure 6, where $\text{SOC} < 0.3\%$, SOC is between $0.3\text{--}0.6\%$, and $\text{SOC} > 0.6\%$. When comparing the NDVI (Figure 3) and SOC map for 2016 (Figure 6), higher SOC corresponded with higher NDVI values. For example, NDVI values were as high as 0.58 in the central part of the study area, where there was abundant vegetation due to the presence of agricultural lands and orchards, and therefore, the highest SOC values were predicted in the same region. In comparison, the surrounding regions had substantially lower NDVI, with values reaching as low as -0.32 ; hence, the SOC values were low as well. Similar to Zhao et al. (2015), this study attributed the spatial patterns of SOC to the presence of croplands and agriculture, suggesting that land use is a key control of SOC [57].

To further examine the SOC changes across these four periods, the area of the SOC classes is presented in Figure 7. The results show that from 1986 to 2016, the classes with SOC content $> 0.6\%$ and $0.3\text{--}0.6\%$ decreased by 25,888 ha (5.26%) and 138,272 ha (28.63%), respectively. In comparison, the area of the $\text{SOC} < 0.3\%$ class increased by 164,160 ha (33.99%) and thus indicating a general decrease in SOC from 1986 to 2016 (Figure 7). These changes could be linked to climate change, decreased groundwater quality, and decreased rainfall from 1986 to 2016. In addition, because of the increased soil salinity in the region, the cultivated lands also drastically decreased from 1986 to 2016. For example, Fathizad et al. (2020) reported that an increase in soil salinity was attributed to the expansion of agricultural lands, increased number of wells, and the overexploitation of groundwater resources [23]. Thus, agricultural activities in arid and semi-arid regions of Iran are the most important controls of SOC.

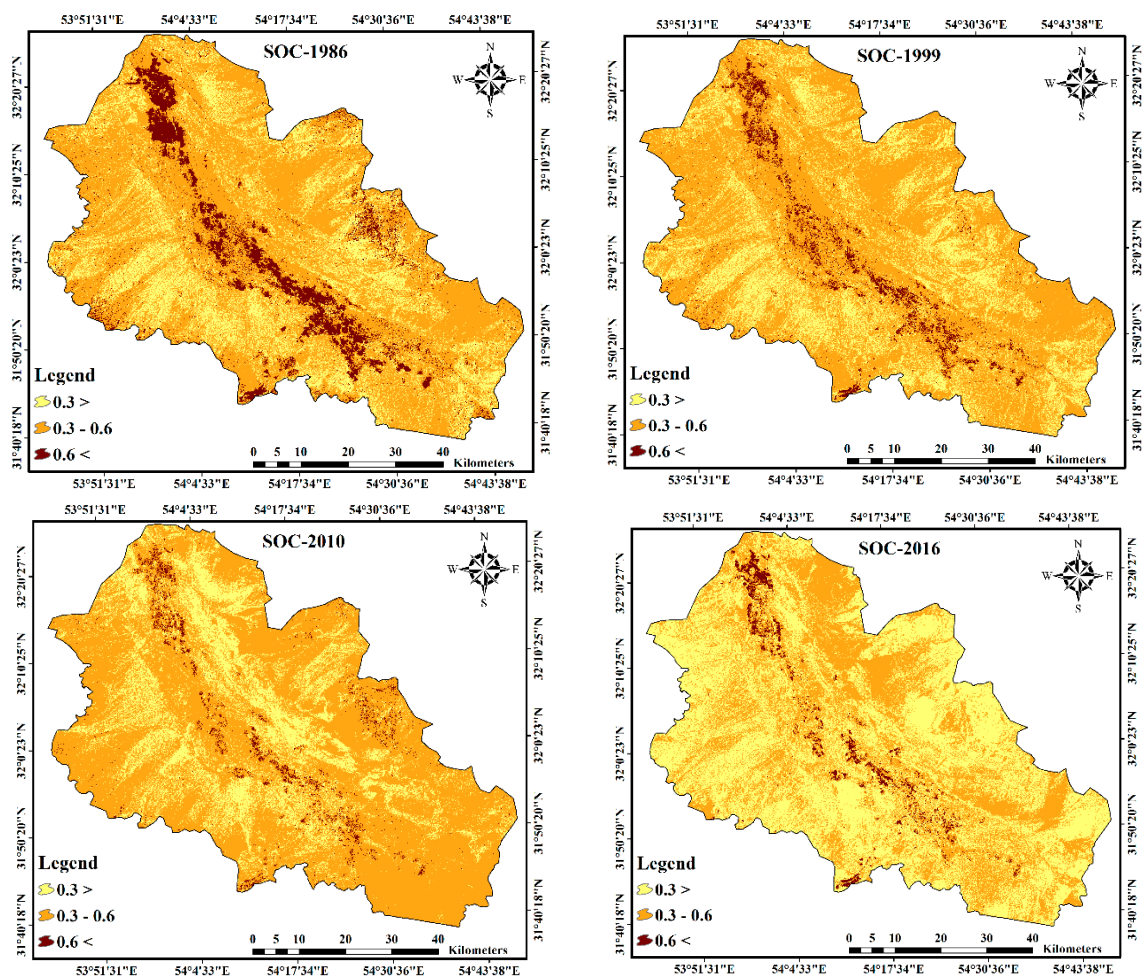


Figure 6. Predicted soil organic carbon (SOC) maps of the Yazd-Ardakan Plain for 1986, 1999, 2010, and 2016.

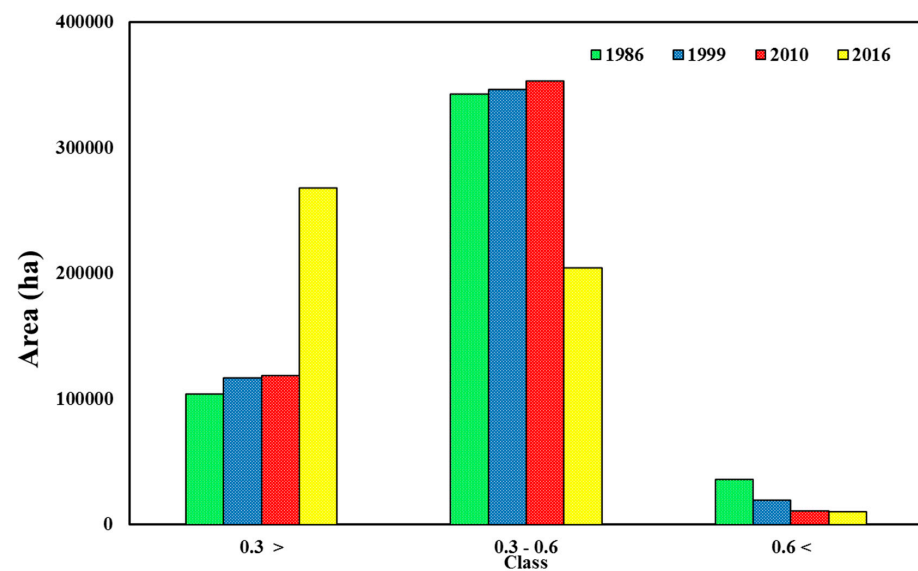


Figure 7. Areal extent of each soil organic carbon (SOC) class for the Yazd-Ardakan Plain for 1986, 1999, 2010, and 2016.

4. Conclusions

This study demonstrated the effectiveness of the RF model for predicting the spatiotemporal patterns of SOC content of the oasis and arid-agroecosystem area which the approach may be utilized in other similarly arid conditions. In general, the results showed that the RF model could be used for mapping the spatiotemporal dynamics of SOC content. The results revealed alarming changes in SOC content where areas with SOC = 0.3–0.6% decreased by 25,888 ha, and areas with SOC < 0.3% increased by 164,160 ha. These drastic changes resulted from reduction in agricultural activity and cultivatable areas from 1986 to 2016. During this period, the area of agriculture lands decreased by ~1.18%, while the areas of barren lands and sandy hills increased 5.16% and 0.09%, respectively. It can be concluded that the mismanagement of the lands, not only by the replacement of agricultural lands with residential areas, but also by declining water quality, as reported by Fathizad et al. [23], reduced agricultural activity and SOC. Soil dynamics, in addition to soil formation and evolution, are strongly influenced by soil management. Therefore, future research should be conducted with the focus of obtaining other environmental predictors to further investigate changes in other soil properties based on the use of DSM and machine-learning techniques. It is also suggested that RF models and other environmental predictors, such as land class and land cover data and satellite images with higher resolution be used in future studies. This study provides a method for understanding the spatiotemporal dynamics of SOC and the methods may be adapted to the monitoring of other soil properties.

Author Contributions: Conceptualization, H.F. and R.T.-M.; methodology, H.F. and R.T.-M.; software, H.F. and R.T.-M.; validation, H.F., R.T.-M., M.A.H.A., M.Z., B.H. and T.S.; formal analysis, H.F., R.T.-M. and M.A.H.A.; investigation, H.F., M.A.H.A., M.Z. and R.T.-M.; resources, H.F.; data curation, H.F. and M.A.H.A.; writing—original draft preparation, H.F., M.A.H.A., R.T.-M., M.Z., B.H. and T.S.; writing—review and editing, H.F., M.A.H.A., R.T.-M., M.Z., B.H. and T.S.; visualization, H.F. and R.T.-M.; supervision, M.A.H.A. and R.T.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: Ruhollah Taghizadeh-Mehrjardi and Thomas Scholten have been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC number 2064/1—Project number 390727645, and collaborative research center SFB 1070 'ResourceCultures'—Project number 215859406. Mojtaba Zeraatpisheh's postdoctoral program at Henan University, China, has been supported by the National Key Research and Development Program of China, grant numbers 2017YFA0604302 and 2018YFA0606500.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Manlay, R.J.; Feller, C.; Swift, M. Historical evolution of soil organic matter concepts and their relationships with the fertility and sustainability of cropping systems. *Agric. Ecosyst. Environ.* **2007**, *119*, 217–233. [[CrossRef](#)]
2. Ayoubi, S.; Karchegani, P.M.; Mosaddeghi, M.R.; Honarjoo, N. Soil aggregation and organic carbon as affected by topography and land use change in western Iran. *Soil Tillage Res.* **2012**, *121*, 18–26. [[CrossRef](#)]
3. Malone, B.P.; Minasny, B.; McBratney, A.B. *Using R for Digital Soil Mapping*; Springer International Publishing: Cham, Switzerland, 2017.
4. Guo, P.-T.; Li, M.-F.; Luo, W.; Tang, Q.-F.; Liu, Z.-W.; Lin, Z.-M. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* **2015**, *237–238*, 49–59. [[CrossRef](#)]
5. Arrouays, D.; McBratney, A.; Minasny, B.; Hempel, J.; Heuvelink, G.; Macmillan, R.; Hartemink, A.; Lagacherie, P.; McKenzie, N. The GlobalSoilMap project specifications. *GlobalSoilMap* **2014**, *494*, 9–12. [[CrossRef](#)]
6. Kempen, B.; Brus, D.; Stoorvogel, J. Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma* **2011**, *162*, 107–123. [[CrossRef](#)]
7. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Finke, P. Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology* **2017**, *285*, 186–204. [[CrossRef](#)]
8. McBratney, A.; Santos, M.M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]

9. Grunwald, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* **2009**, *152*, 195–207. [\[CrossRef\]](#)
10. Behrens, T.; Scholten, T. Chapter 25 A Comparison of Data-Mining Techniques in Predictive Soil Mapping. *Dev. Soil Sci.* **2006**, *31*, 353–617. [\[CrossRef\]](#)
11. Zhu, A.X.; Liu, J.; Du, F.; Zhang, S.J.; Qin, C.-Z.; Burt, J.; Behrens, T.; Scholten, T. Predictive soil mapping with limited sample data. *Eur. J. Soil Sci.* **2015**, *66*, 535–547. [\[CrossRef\]](#)
12. Stumpf, F.; Schmidt, K.; Goebes, P.; Behrens, T.; Schönbrodt-Stitt, S.; Wadoux, A.; Xiang, W.; Scholten, T. Uncertainty-guided sampling to improve digital soil maps. *Catena* **2017**, *153*, 30–38. [\[CrossRef\]](#)
13. Taghizadeh, R.; Nabiollahi, K.; Kerry, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* **2016**, *266*, 98–110. [\[CrossRef\]](#)
14. Taghizadeh, R.; Toomanian, N.; Khavaninzhadeh, A.R.; Jafari, A.; Triantafyllis, J. Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. *Eur. J. Soil Sci.* **2016**, *67*, 707–725. [\[CrossRef\]](#)
15. Taghizadeh, R.; Minasny, B.; Sarmadian, F.; Malone, B. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* **2014**, *213*, 15–28. [\[CrossRef\]](#)
16. Liu, D.; Wang, Z.; Zhang, B.; Song, K.; Li, X.; Li, J.; Li, F.; Duan, H. Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. *Agric. Ecosyst. Environ.* **2006**, *113*, 73–81. [\[CrossRef\]](#)
17. Chai, X.; Shen, C.; Yuan, X.; Huang, Y. Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. *Geoderma* **2008**, *148*, 159–166. [\[CrossRef\]](#)
18. Sumfleth, K.; Duttman, R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecol. Indic.* **2008**, *8*, 485–501. [\[CrossRef\]](#)
19. Pahlavan-Rad, M.R.; Dahmardeh, K.; Brungard, C. Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. *Geoderma Reg.* **2018**, *15*, e00195. [\[CrossRef\]](#)
20. Zeraatpisheh, M.; Jafari, A.; Bodaghabadi, M.B.; Ayoubi, S.; Taghizadeh-Mehrjardi, R.; Toomanian, N.; Kerry, R.; Xu, M. Conventional and digital soil mapping in Iran: Past, present, and future. *Catena* **2019**, *188*, 104424. [\[CrossRef\]](#)
21. McBratney, A.B.; Stockmann, U.; Angers, D.A.; Minasny, B.; Field, D.J. Challenges for Soil Organic Carbon Research. In *Soil Carbon*; Springer: Cham, Switzerland, 2014; pp. 3–16. [\[CrossRef\]](#)
22. Liu, F.; Zhang, G.-L.; Sun, Y.-J.; Zhao, Y.-G.; Li, D.-C. Mapping the Three-Dimensional Distribution of Soil Organic Matter across a Subtropical Hilly Landscape. *Soil Sci. Soc. Am. J.* **2013**, *77*, 1241–1253. [\[CrossRef\]](#)
23. Fathizad, H.; Ardakani, M.A.H.; Sodaiezhadeh, H.; Kerry, R.; Taghizadeh-Mehrjardi, R. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma* **2020**, *365*, 114233. [\[CrossRef\]](#)
24. Taghizadeh-Mehrjardi, R.; Fathizad, H.; Ardakani, M.A.H.; Sodaiezhadeh, H.; Kerry, R.; Heung, B.; Scholten, T. Spatio-Temporal Analysis of Heavy Metals in Arid Soils at the Catchment Scale Using Digital Soil Assessment and a Random Forest Model. *Remote Sens.* **2021**, *13*, 1698. [\[CrossRef\]](#)
25. Minasny, B.; McBratney, A. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* **2006**, *32*, 1378–1388. [\[CrossRef\]](#)
26. Walkley, A.; Black, I.A. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [\[CrossRef\]](#)
27. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [\[CrossRef\]](#)
28. Huete, A. Extension of soil spectra to the satellite: Atmosphere, geometric, and sensor considerations. *Photo Interpretation* **1996**, *34*, 101–118.
29. Leblon, B. Soil and vegetation optical properties. Faculty of Forestry and Environmental Management University of New Brunswick, Fredericton (NB) Canada. *Int. J. For. Eng.* **1997**, *13*, 57.
30. Crippen, R. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [\[CrossRef\]](#)
31. Foody, G.M.; Cutler, M.; McMorro, J.; Pelz, D.; Tangki, H.; Boyd, D.; Douglas, I. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Glob. Ecol. Biogeogr.* **2001**, *10*, 379–387. [\[CrossRef\]](#)
32. Nield, S.J.; Boettinger, J.L.; Ramsey, R.D. Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data. *Soil Sci. Soc. Am. J.* **2007**, *71*, 245–252. [\[CrossRef\]](#)
33. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [\[CrossRef\]](#)
34. Arzani, H.; King, G.W. *Application of Remote Sensing (Landsat TM Data) for Vegetation Parameters Measurement in Western Division of NSW*; International Grassland Congress: Hohhot, China, 2008.
35. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [\[CrossRef\]](#)
36. Pettorelli, N.; Vik, J.O.; Mysterud, A.; Gaillard, J.-M.; Tucker, C.J.; Stenseth, N.C. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* **2005**, *20*, 503–510, Erratum in *Trends Ecol. Evol.* **2006**, *21*, 11. [\[CrossRef\]](#)
37. Kullberg, E.G.; DeJonge, K.C.; Chávez, J.L. Evaluation of thermal remote sensing indices to estimate crop evapotranspiration coefficients. *Agric. Water Manag.* **2017**, *179*, 64–73. [\[CrossRef\]](#)

38. Khan, N.M.; Rastoskuev, V.V.; Sato, Y.; Shiozawa, S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* **2005**, *77*, 96–109. [\[CrossRef\]](#)
39. Wilson, E.H.; Sader, S.A. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [\[CrossRef\]](#)
40. Major, D.J.; Baret, F.; Guyot, G. A ratio vegetation index adjusted for soil brightness. *Int. J. Remote Sens.* **1990**, *11*, 727–740. [\[CrossRef\]](#)
41. Bannari, A.; Guedon, A.M.; El-Harti, A.; Cherkaoui, F.Z.; El-Ghmari, A. Characterization of Slightly and Moderately Saline and Sodic Soils in Irrigated Agricultural Land using Simulated Data of Advanced Land Imaging (EO-1) Sensor. *Commun. Soil Sci. Plant Anal.* **2008**, *39*, 2795–2811. [\[CrossRef\]](#)
42. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Toomanian, N.; Heung, B.; Behrens, T.; Mosavi, A.; Band, S.B.; Amirian-Chakan, A.; Fathabadi, A.; Scholten, T. Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. *Geoderma* **2021**, *383*, 114793. [\[CrossRef\]](#)
43. Douaoui, A.E.K.; Nicolas, H.; Walter, C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* **2006**, *134*, 217–230. [\[CrossRef\]](#)
44. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
45. Nabiollahi, K.; Taghizadeh-Mehrjardi, R.; Shahabi, A.; Heung, B.; Amirian-Chakan, A.; Davari, M.; Scholten, T. Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. *Geoderma* **2021**, *385*, 114858. [\[CrossRef\]](#)
46. Taghizadeh-Mehrjardi, R.; Hamzehpour, N.; Hassanzadeh, M.; Heung, B.; Goydaragh, M.G.; Schmidt, K.; Scholten, T. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. *Geoderma* **2021**, *399*, 115108. [\[CrossRef\]](#)
47. Vapnik, V.N.; Lerner, A. Pattern recognition using generalized portrait method. *Autom. Remote Control.* **1963**, *24*, 774–780.
48. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [\[CrossRef\]](#)
49. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, NY, USA, 1966.
50. Tajik, S.; Ayoubi, S.; Zeraatpisheh, M. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. *Geoderma Reg.* **2020**, *20*, e00256. [\[CrossRef\]](#)
51. Wang, S.; Wang, Q.; Adhikari, K.; Jia, S.; Jin, X.; Liu, H. Spatial-Temporal Changes of Soil Organic Carbon Content in Wafangdian, China. *Sustainability* **2016**, *8*, 1154. [\[CrossRef\]](#)
52. Lamichhane, S.; Kumar, L.; Wilson, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* **2019**, *352*, 395–413. [\[CrossRef\]](#)
53. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [\[CrossRef\]](#)
54. Falahatkar, S.; Hosseini, S.M.; Ayoubi, S.; Salmanmahiny, A. Predicting soil organic carbon density using auxiliary environmental variables in northern Iran. *Arch. Agron. Soil Sci.* **2015**, *62*, 375–393. [\[CrossRef\]](#)
55. Hong, S.Y.; Sudduth, K.A.; Kitchen, N.R.; Drummond, S.T.; Palm, H.L.; Wiebold, W.J. Estimating within-field variations in soil properties from airborne hyperspectral images. In Proceedings of the Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 2002, Denver, CO, USA, 10–15 November 2002.
56. Gomez, C.; Rossel, R.V.; McBratney, A. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma* **2008**, *146*, 403–411. [\[CrossRef\]](#)
57. Zhao, M.-S.; Zhang, G.-L.; Wu, Y.-J.; Li, D.-C.; Zhao, Y.-G. Driving forces of soil organic matter change in Jiangsu Province of China. *Soil Use Manag.* **2015**, *31*, 440–449. [\[CrossRef\]](#)