Dimension reduction of the remote sensing data to estimate soil organic carbon

Niusha Mozafarinahr[1], Hadiseh Sadat Hasani[2], Marzieh Jafari[2*]

[1]M.Sc Graduate, Department of Geodesy and Surveying Engineering, Tafresh University, Tafresh, Iran
[2]Assistant Professor, Department of Geodesy and Surveying Engineering, Tafresh University, Tafresh, Iran

ABSTRACT

Soil is a very complex phenomenon that includes organic materials, minerals, water and air. The distribution of organic matter in the soil has a profound effect on biological activity, nutrient availability, soil and soil seed structure, and water holding capacity, and soil management in general. In this research, the relation between soil spectral reflectance using the Landsat 8 satellite data as well as the SRTM Elevation data and soil organic carbon has been investigated. In the proposed method, spectral reflection of data in the main bands of the Landsat 8 satellite is investigated and processed. In addition to the main bands, vegetation and lighting indices and topographic features have been studied. In this study, a method for selecting effective indexes in increasing the accuracy of soil organic carbon modelling is presented. For this purpose, in the first step of modelling, Linear regression, Support Vector Machine regression and Neural Network methods have been used for the connection between remote sensing data and soil organic carbon. To implement the proposed method, 100 soil samples in East Azerbaijan province have been used. According to RMSE and R2 statistical indices, which are the basis for evaluating the models, the neural network model was selected as the final model and with the values of RMSE = 0.404, R2= 0.254 and RRMSE=46.597 is more accurate than the regression method. Due to the importance of dimensionality to increase accuracy and reduce the complexity of calculations, a genetic algorithm was proposed in this study. This efficient algorithm increases the accuracy of soil organic carbon modelling and eliminates additional indicators. After applying the genetic algorithm (GA) to the neural network model, we were able to achieve better accuracy and the values of the baseline statistical indices were changed to RMSE = 0.279, R2 = 0.718 and RRMSE=27.116. Also, to check the efficiency of the genetic algorithm, the PCA algorithm was also implemented on the data and the comparison results showed that the genetic algorithm was successful in reducing dimensions along with increasing accuracy.

KEYWORDS

Remote Sensing, Soil Organic Carbon, Neural Network, Dimension Reduction, Genetic Algorithm.

1. Introduction

To determine soil organic carbon, field observing and remote sensing methods are used. Considering the time-

* Corresponding Author: Email: jafari@tafreshu.ac.ir

consuming field operations and laboratory investigations and the unavailability of some points, we need indirect, fast and low-cost methods. The use of remote sensing data, digital terrain model and topographic features in the estimation of soil indices is a reliable, cheap and accurate method, which is based on the surface reflection of light by phenomena [1, 2]. Recent studies show that the measurement data of two indices extracted from it is a suitable source of information in determining soil organic carbon [3]. Considering the importance of soil organic carbon, its modelling is very important for unavailable areas. Due to the progress of remote sensing systems, a huge range of different data is available to everyone. Due to the high dimensions of the data as a result of the complexity of the calculations, it is very important to choose the correct data [4].

Height, slope, direction of slope, shape and position of slope and natural conditions of soil drainage are factors that are attributed to topography [5]. Since topography affects the movement of water and the distribution of materials carried by water flow and thus controls the intensity and type of different processes in the soil, topography is one of the most important indices affecting soil organic carbon changes and its spatial distribution [6, 7]. Studies show that a large number of soil characteristics depend on the slope and the direction of the slope [8]. The relationship between soil characteristics and several slope factors such as degree, curvature, distance from the flat part, longitudinal direction and height compared to the flat part has been investigated and the results of recent research show a significant relationship between different slope values and soil characteristics [9].

Besides the topographic data, soil spectral reflectance with different wavelengths can also be used to estimate soil organic carbon. Measurement of soil spectral reflectance is done by three methods laboratory, field and remote sensing. Since the spectral reflectance of the soil is affected by its constituent materials, various factors such as moisture, percentage of organic matter, percentage of iron oxide, relative percentage of clay, silt and roughness coefficient of the soil surface affect the spectral reflectance of the soil [10, 11].

Due to the impact of spectral and topographical information in the estimation of soil organic carbon and the availability of different remote sensing data, data processing faces challenges. Despite the valuable results of previous research, there are limitations due to the large amount of data in conducting the research. Finding the features of effective remote sensing in soil organic carbon modelling can avoid wasting time and money. Feature selection techniques can simplify modelling by reducing the number of input variables and improving prediction accuracy. To remove additional and unimportant features, several feature selection and extraction methods have been developed, including the most important of them are Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Artificial Bee Colony (ABC) algorithm and Principal Component Analysis (PCA).

In this research, in the first step, soil organic carbon modelling has been done using an artificial neural network based on spectral and topographical data. In the next step, the selection of optimal and effective indices is done using the genetic algorithm, and with the help of the optimal indices, we will try to increase the accuracy of modelling. In other words, the innovation of this research is in applying the genetic algorithm to the modelling process using remote sensing data to avoid wasting time and money in addition to increasing the accuracy of the model.

2. Methodology

The process of implementing the proposed method takes place in four main steps. In the first step, after preparing the desired satellite images, the necessary pre-processing will be done. In the second step, the desired indices are extracted from the images and the feature vector is determined for specific points. In the third step, soil organic carbon modelling is done using the feature vector of points, and in the last step, using the dimension reduction technique, we try to reduce and eliminate some features to achieve higher accuracy for modelling. The flowchart of the proposed method is shown in "Figure 1".

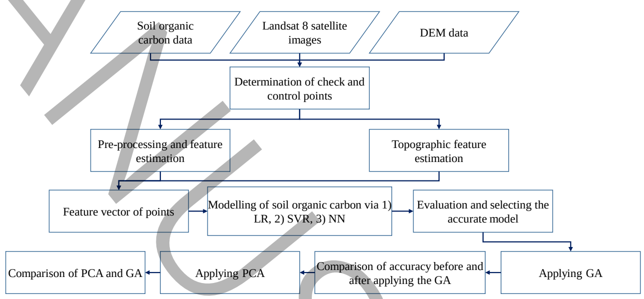

Figure 1. The flowchart of the proposed method.

3. Results and Discussion

The studied area is located in the northwest of Iran and East Azarbaijan province, between longitude 45.49 to 48.02 and latitude 36.93 to 39.30. The region area is about 47830 m^2. According to the size of the area and the existing conditions, 100 points with latitude and longitude coordinates along with the known soil organic carbon were provided by the Soil and Water Research Institute of Iran. To implement the method, the Landsat 8 satellite images were selected for this research due to their availability, and spectral and spatial resolution [12].

In addition, the topographical data were extracted from the SRTM[2] model.

For all data coordinates, soil organic carbon, spectral indices, values in the main band and topographical indices are calculated. Then, we modelled soil organic carbon based on points with known coordinates and organic carbon, as well as based on 21 features collected, using three methods: Linear Regression (LR), Support Vector linear Regression (SVR), and Neural Network (NN). According to the numerical results, the evaluation criteria for NN modelling are superior to the other two models. So NN modelling is more accurate and predicts soil organic carbon better.

After implementing the proposed method based on combining the GA with the soil organic carbon modelling method, we will evaluate the modelling results. To the effectiveness of the GA, the PCA method was implemented on the data so that the results can be compared with the results obtained from the GA. According to the eigenvalues, the first 10 principal components were used in this study. The comparison of the desired evaluation indices can be seen in "Table 1".

Table 1. Comparison of accuracy evaluation criteria resulting from dimension reduction based on the GA and PCA.

Method	RMSE	R^2	RRMSE
GA	0.279	0.718	27.116
PCA	0.417	0.372	40.469

4. Conclusions

The obtained results indicate that the GA is more capable of dimension reduction of the input data compared to the PCA method. According to the presented results the efficiency of the GA to increase the accuracy is clear. Therefore, by applying this algorithm, you can reach results closer to reality.

According to the results of the GA application, among the 31 input features that include the spectral-topographic indices obtained along with the main bands of the Landsat satellite, the application of the 15 spectral-topographic features (Coastal, Green, NIR, SWIR1, SWIR2, MSR, OSAVI, NDSI, SI1, SI2, SI32, SI4, SI5, Slope, Aspect) has increased the accuracy of the modelling; about the rest of the indexes, it is better to exclude them from the modelling process.

5. References

[1] F. Chen, D.E. Kissel, L.T. West, W. Adkins, Field-scale mapping of surface soil organic carbon using remotely sensed imagery, Soil Science Society of America Journal, 64(2) (2000) 746-753.

[2] M. Ladoni, H.A. Bahrami, S.K. Alavipanah, A.A. Norouzi, Estimating soil organic carbon from soil reflectance: a review, Precision Agriculture, 11 (2010) 82-99.

[3] X. He, L. Yang, A. Li, L. Zhang, F. Shen, Y. Cai, C. Zhou, Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images, Catena, 205 (2021) 105442.

[4] G.S. Bhunia, P. Kumar Shit, H.R. Pourghasemi, Soil organic carbon mapping using remote sensing techniques and multivariate regression model, Geocarto International, 34(2) (2019) 215-226.

[5] H. Emami, Investigation the effects of aspect and degree of slope on soil quality in the South East of Mashhad, Journal of Water and Soil Conservation, 23(2) (2016) 301-310, (in Persian).

[6] J. Kumhálová, F. Kumhála, M. Kroulík, Š. Matějková, The impact of topography on soil properties and yield and the effects of weather conditions, Precision Agriculture, 12 (2011) 813-830.

[7] S. Singh, Understanding the role of slope aspect in shaping the vegetation attributes and soil properties in Montane ecosystems, Tropical Ecology, 59(3) (2018) 417-430.

[8] S. Jakšić, J. Ninkov, S. Milić, J. Vasin, M. Živanov, D. Jakšić, V. Komlen, Influence of slope gradient and aspect on soil organic carbon content in the region of Niš, Serbia, Sustainability, 13(15) (2021) 8332.

[9] P.H. Walker, Contributions to the understanding of soil and landscape relationships, Soil Research, 27(4) (1989) 589-605.

[10] H. Aïchi, Y. Fouad, C. Walter, R.V. Rossel, Z.L. Chabaane, M. Sanaa, Regional predictions of soil organic carbon content from spectral reflectance measurements, Biosystems engineering, 104(3) (2009) 442-446.

[11] H. Bartholomeus, M.E. Schaepman, L. Kooistra, A. Stevens, W. Hoogmoed, O. Spaargaren, Spectral reflectance based indices for soil organic carbon quantification, Geoderma, 145(1-2) (2008) 28-36.

[12] T.R. Loveland, J.R. Irons, Landsat 8: The plans, the reality, and the legacy, Remote Sensing of Environment, 185 (2016) 1-6.

[2] Shuttle Radar Topography Mission