Advanced Geomatics

http://publish.mersin.edu.tr/index.php/geomatics/index

e-ISSN: 2791-8637

Vis-NIR Spectroscopy Coupled with Machine Learning Algorithms to Predict and Identify the Key Wavelengths of Soil Gypsum Content in Fars Province, Southern Iran

Monireh Mina[1], Mahrooz Rezaei*[2], Leila Hossein Abadi [3], Abdolmajid Sameni[1]

[1]Shiraz University, School of agriculture, Department of soil science, shiraz, Iran
[2]Wageningen University & Research P.O. Box 47, AA, 6700, Meteorology and Air Quality Department, Wageningen, the Netherlands
[3] Shahid Beheshti University, Remote sensing and GIS center, Tehran, Iran

Keywords

Key wavelength,
PLSR model,
Savitzky-Golay filter,
Spectral reflectance.

Research Article

Received : 11.09.2022
Revised : 16.02.2023
Accepted : 09.03.2023
Published : 31.03.2023

Abstract

The use of soil spectral reflectance, which has been introduced as a new method in soil science, is widely used in estimating the physicochemical properties of soil. The purpose behind this research was estimating the amount of gypsum in surface soils of Fars province. Based on random sampling method, 100 soil samples were collected and measured by standard method. Spectral analysis of soil samples was performed using a spectrophotometer between the range of 2500-400 nm. After this stage, various preprocessing methods were evaluated and finally the percentage of soil gypsum was modeled using two models of partial least squares regression (PLSR) and support vector regression (SVR). Our results illustrated that best results for estimating the percentage of soil gypsum are related to the SVR model with Preprocessing Savitzky- Golay Filter with the first derivative. Also, according to RPIQ statistics, the estimation of PLSR model for the percentage of soil gypsum in the weak class is 1.02% and for the SVR model in the moderate class is 1.54%. In the present study, key wavelengths were defined as wavelengths which ranged around 750, 1400, 1570, 1750-1800, 2100, 2200 and 2338 nm and showed the highest correlation with gypsum content in soil.

1. Introduction

The use of visible-near-infrared spectroscopy has been introduced as a fast, inexpensive and non-destructive method that has a remarkable capability in estimating different soil properties (Cambou et al., 2016). Among various soil characteristics soil gypsum has great importance. Gypsum has more solubility than carbonates and therefore, is under the influence of leaching process, and this resulted in less amount in the soil (Chaternour et al., 2020). The amount of gypsum has significant effect on soil properties namely soil water retention, aggregate stability and soil structure; More than 25% gypsum content in soil has a negative effect on plant growth and soil resilience (Smith and Robertson., 1962). Due to the cost, time and difficulty of direct measurement of soil gypsum, the use of indirect methods such as soil spectral behavior and spectroscopy has become common (khayamim et al., 2015). So far, many studies have been done in this field, most of which have been researched on soil particle size, CaCO3 (Gomez et al., 2008), soil organic matter (SOM) (Ostovari et al., 2018) and soil moisture (Mina et al., 2021). In these studies, methods such as support vector Regression (SVR), partial least squares regression (PLSR), and principal component regression (PCR) have been utilized for assessing the corelation between soil properties and spectral data (Farifteh et al., 2007). The SVR method makes the use of the SVM's principles and rules for regression problems. Being famous for its supervision and non reliablility on parameters, SVR is a statestic-based learning method (Vapnik, 1996). such method has enough sufficiency in generalizing models which were trained to unseable data with a decent accuracy (Gholizadeh et al., 2013).on the one hand, The strength of SVR models are their effectiveness in working with data from many variables in high-dimensional (Karatzoglou et al., 2006) and their great ability to assess how variables are dependent on each other, so they can be predicted by a hyperplane which was fitted optimally

* **Corresponding Author**

(monireh.mina@gmail.com)
*(mahrooz.rezaei@wur.nl)
(leilahosseinabadi1993@gmail.com)
(majid.baba@gmail.com)

Cite this;

Mina, M., Rezaei, M., Abadi, L. H. & Sameni, A. (2023). Vis-NIR spectroscopy coupled with machine learning algorithms to predict and identify the key wavelengths of soil gypsum content in Fars province, southern Iran. Advanced Geomatics, 3(1), 09-15.

to training datasets (Wu, W., et al 2018). On the other hand, there are some difficulties in selecting among various algorithims and choosing the optimized parameters for the purpose of improving prediction outcomes. The utilization of such models for optimizing predictive outcomes in soil properties is crucial (Deiss, L. et al 2020). In SVR models there are two features which need optimization. Firstly, we need to select the kernel function (in algorithm) and secondly, it is the noise tolerance in the epsilon loss function for each kernel. SVR models use various kernel functions and it includes :, polynomial, sigmoid , radial, and linear, and each kernel has parameters which should be optimized, and their potential suitability needs to be considered (Deiss, L. et al 2020).

One of the important steps in estimating the soil properties is using pre-processing spectroscopy (Ostovari et al., 2018). The preprocessing methods, using mathematical functions, corrected the nonlinear relations created in relation to the amount of light absorption, and by eliminating the noise, the clarity of the absorption characteristics improves and ultimately the calibration become better (Mina et al., 2021). Derivative methods are the most widely used pre-processing methods in spectrophotometric studies, amplified in poorly-recorded signal derivative methods and lead to improvement of soil properties (Stenberg et al., 2010). In most studies, they used First Derivative (FD) and Second Derivative (SD) methodology (Martins, 1989) and Standard Normal Variate (SNV) (Barnes, 1989) as an Inseparable technique in spectral modeling. Akbarifazli et al., 2021 conducted an experiment for the purpose of accuracy evaluation of visible and near-infrared spectroscopy in estimating SOM and its total neutralization value. They have used PLSR and SVR methods to estimate these parameters then compared their results. it illustrated that PLSR model outperformed SVR in estimation of the mentioned parameters. In another study Gholizadeh et al., 2013 explored the suitableness of Vis-NIR, which is between the range of 350 and 2500 nm, and mid-IR spectroscopy as an effective method for determining SOM quantity. Their result showed that spectroscopy method, specifically the mid-IR method which made use of the Least Squares Support Vector Machine (LS-SVM) algorithm can be a valuable tool to determine SOM quantity and quality. It should be noted that, they have used Savitzky–Golay in their research.

Also, studies have been reported in estimating the soil gypsum content using spectral reflections. Among these studies, Chaternor et al., 2020, has conducted a research in Khuzestan province which determined soil gypsum key wavelength. in this study soil spectrum was preprocessed using various methods including the Savitzky-Golay filter and two multivariate regression models including PLSR and SVR. Utilizing these methods allowed them to compare the estimated performance of soil gypsum content which revealed that SVR model presented better performance compared with PLSR. In another study, Khayamim et al., 2015 have used vis-NIR spectroscopy to evaluate this method's efficiency in quantification of gypsum content and carbonate content in soil. They compared the result of PLSR method with the routine standard laboratory technique and another feature-specific method using Continuum Removal (CR). Their results have shown that PLSR outperformed the other method in both mentioned soil properties. Also, the properties of calcium carbonate and gypsum were obtained from optimal accuracy with coefficients of determination of 0.45 and 0.8 respectively. (Khayamim et al., 2015). Hassani et al. (2014) estimated the properties of gypsum (RPD = 2.65), organic matter (RPD = 1.64) and calcium carbonate (RPD = 2.86) using spectral reflections. Gohari et al. (2017) used vis-NIR spectroscopy for the estimation of gypsum content, organic matter and carbonates in soil. their research showed better result in the percentage of gypsum and organic matter in the good class, while it showed worse result in the carbonates in the weak class due to the relative deviation of the model prediction.

Due to the importance of gypsum content, this research was done to estimate the percentage of soil gypsum with PLSR and SVR models using spectral data by applying Savitzky -Glaye filters with the first derivative.

2. Method

2.1. Study Area

Fars province was our chosen study area and it is situated in the south central region in Iran with coordinates 27° 2' to 31° 42' lat. and 50° 42' to 55° 36' long. The mean annual rainfall varies from 100 mm in the southern part to nearly 400 mm in the northern part.

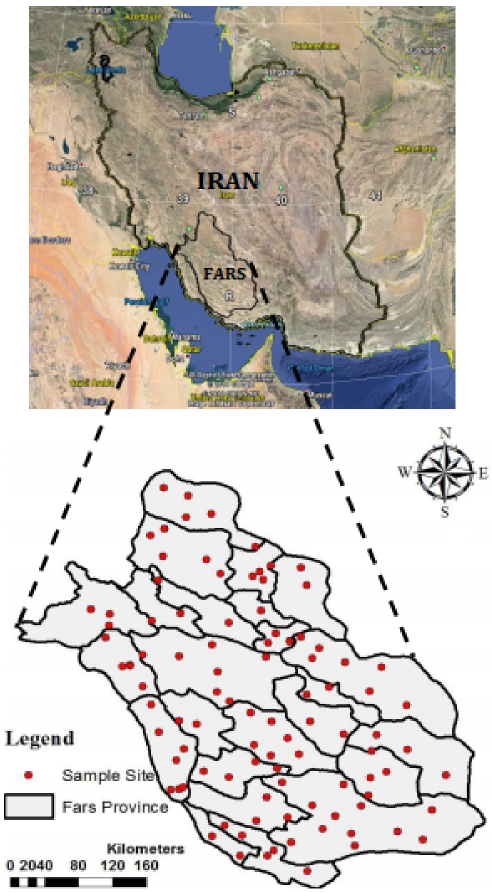

Figure.1. The sampling area map in Fars province in Iran.

2.2. Soil Sampling and Analysis

Sampling of soil was conducted randomly in 0 -10 cm depth. 100 soil samples were chosen and collected and then the samples were transferred into laboratory, they were air dried first and after that passed through a sieve which was 2 mm. Gypsum content was determined by acetone method (Nelson,1982).

2.3. Spectral Reflectance Measurement

Soil spectral data were determined using a spectroscopy device (NIRS-XDS) in the range of vis-NIR wavelengths (2500-400 nm). 20 g of each sample of air-dried soil with a size of less than 2 mm was placed in a special container and then 5 scans were performed on them (Figure 1, a). Due to the high noise at the beginning and end of the spectral data, the range of 449-400 and 2500-2451 nm was removed from the modeling process and then for the purpose of eliminating turbulence and environmental factors, and increasing the data's quality and obtaining better results, preprocessing of the Savitzky and Glaye filters with the first derivative was applied to the spectral data of soil samples (Figure 1, B).

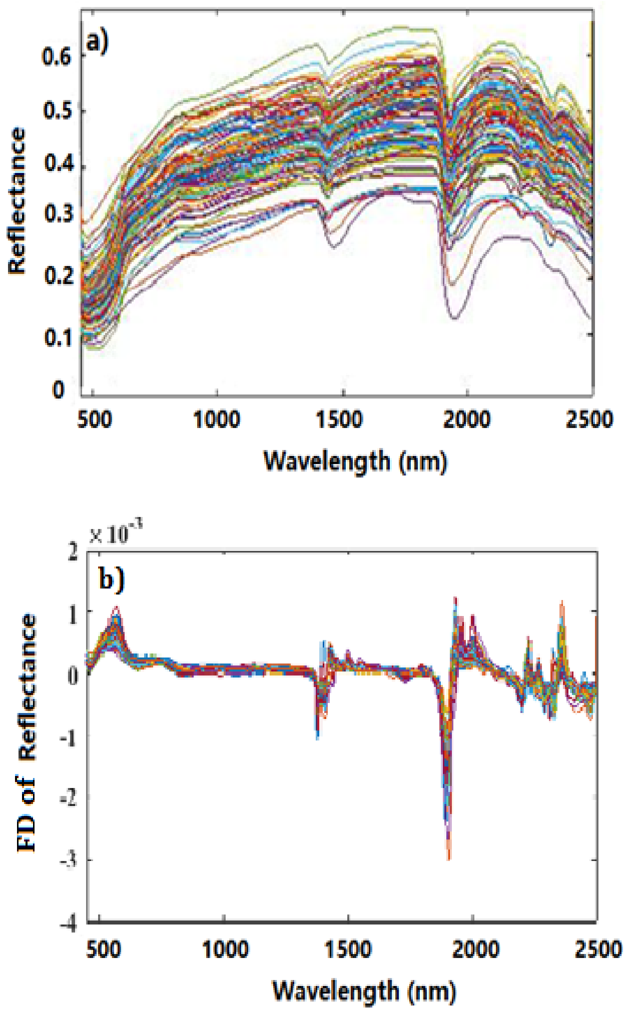

Figure. 2. The a) raw and b) preprocessed spectral reflectance data.

2.4. Modeling

To predict soil gypsum by utilizing the spectra, a multivariate regression and a Machine Learning (ML) technique were carried out, namely PLSR method (Haaland and Thomas, 1988) and SVR (Vapnik, 1995). PLSR is utilizes linear least squares regression which instead of using original input data uses new components and it can cope with multidimensional data (Mirzaei et al., 2022). On the contrary SVR owns special features that can handle data which might be complex and multidimensional. Easily extending them to a nonlinearl modified feature space. In current research, we have set the kernel and a specific type of SVR, namely linear function and epsilon-SVR, respectively; and Cost parameter (C) were fine-tuned by implementing a grid search systematic technique, and optimal parameters were selected by the minimum of RMSE which was acquired through holdout cross-validation. The Unscrambler X v. 10.4 and MATLAB 2019b software were used for spectral data processing and statistical analysis and modelling.

2.5. Model Evaluation

To predict the percentage of soil gypsum based on soil spectral reflectance, PLSR and SVR were applied. To evaluate the accuracy of the models, three statistical criteria including coefficient of determination (R2), Root Mean Square Error (RMSE) and Ratio of Performance to the Interquartile range (RPIQ) were utilized (Mina et al., 2022).

$$R^2 = \frac{\left[\sum_{i=1}^{n}(O_i - \overline{O_i})(P_i - \overline{P_i})\right]^2}{\sum_{i=1}^{n}(O_i - \overline{O_i})^2 \sum_{i=1}^{n}(P_i - \overline{P_i})^2} \tag{1}$$

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}\right)} \tag{2}$$

$$RPIQ = \frac{IQ}{RMSE} \qquad IQ = Q_3 - Q_1 \tag{3}$$

3. Results

For the purpose of modeling, at first, the data were divided into two different sets including training data (70% of data) and also testing data (30% of data) randomly. Using t-test, there was no significant difference between the two datasets. Table 1 depicts a summary of statistics in the measured soil gypsum content in train and test datasets. All soil samples contained a low gypsum content with a mean of 0.97% and 0.99% for train and test datasets, respectively.

Table 1. Statistical analysis of gypsum content in soil, Range (including min and max), mean values, Standard Deviation (SD) and Coefficient of Variation (CV). a- significant difference ($p< 0.05$)

soil parameter	Gypsum	
	Train	Test
Unit	%	%
Range	0.2-3.98	0-3.90
Mean ± SD	$0.97^a \pm 0.64$	$0.99^a \pm 0.68$
CV (%)	65.97	68.68

The values of R^2, RMSE and RPIQ from modeling in estimating soil gypsum based on soil spectral reflections are illustrated in Table 2. The outcomes of Table 2 showed that SVR model has a higher performance in estimating the amount of soil gypsum than the PLSR model. SVR has the highest R2 (0.85, 0.73%) and RMSE (0.22, 0.39%) in both training and testing stages, respectively. In addition to RMSE, the accuracy of the model predicted by RPIQ was also evaluated. Classification is done by Lacerda et al., 2016 into 6 classes: Very Poor with RPIQ < 1, weak with RPIQ = 1 – 1.4, Moderate with RPIQ = 1.4-1.8, Good with RPIQ = 1.8-2, Very Good with RPIQ = 2-2.5 and Excellent with RPIQ >2.5. The SVR model has a moderate performance using spectral reflectance with Savitzky- Glaye filter with the first derivative, and PLSR has a poor performance in estimating the amount of soil gypsum.

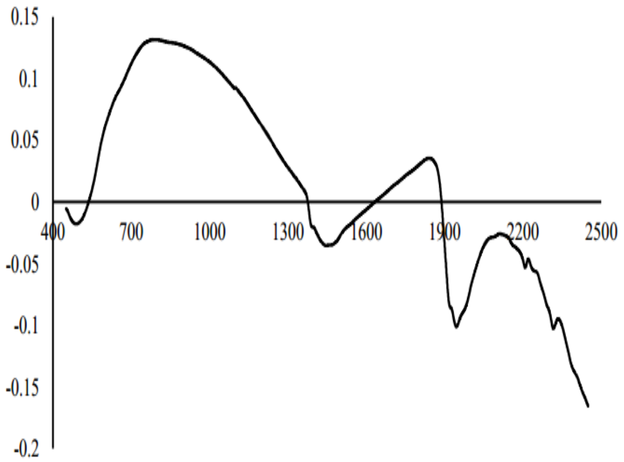

Figure. 3. Pearson's correlation coefficient between row spectral reflectance values across the range of Vis_NIR and soil gypsum

Table 2. Prediction result for gypsum using PLSR and SVR models. R^2, RMSE, and RPIQ.

Method	Train			Test		
Model	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ
PLSR	0.74	0.30	1.97	0.57	0.48	1.02
SVR	0.85	0.22	2.10	0.73	0.39	1.54

Figure 2 depicts the measured gypsum versus the predicted gypsum using PLSR and SVR models in the mentioned datasets. In both of them, the points show a well-scattered look around 1:1 line for SVR comparing with PLSR, that demonstrates better performance of SVR.

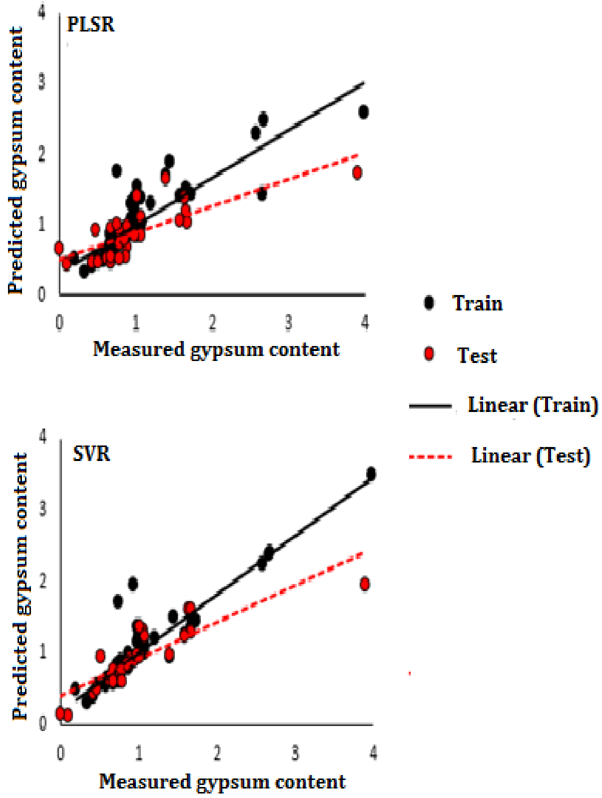

Figure 2. Scatter plots of predicted comparing with measured gypsum by PLSR and SVR. a) Train set (N=70), b) Test set (N=30). PLSR and SVR models.

Soil spectra representative depicts three specific absorption bands including 1414, 1915, 2212, and 2341 nm (Figure. 1). In Figure. 3 the Pearson correlation coefficient between the spectral reflectance and soil gypsum content ranging from 400 to 2500 nm is presented. It can be useful in identifying the bands which are considered to be the most influential ones. As it is conspicuous in Figure. 3, a relative high correlation between the measured values of soil gypsum and the soil spectral reflectance is clearly observed.

overall, the relationship between spectra and soil gypsum can be clearly seen in specific bands of near 1400, 1900, 2200, and 2340nm.

Absorption characteristics also demonstrate hydroxyl groups at 1915 nm, the bonding of hydroxides with iron, clay mineral networks, free and hygroscopic water at 1414 nm, magnesium and aluminum metals at 2212 nm (Clark et al., 1990). The peaks of absorption which happen nearly at 2341 nm are known to be related to CO3 groups in carbonate minerals in soil (Lagacherie et al., 2008, Gomez et al., 2008). The soil gypsum depicted the highest and most significant correlation ($P < 0.05$) with spectral bands including 750, 1400, 1570, 1750-1800, 2100, 2200 and 2338 nm.

4. Discussion

Among soil gypsum in the train and test datasets the t-test enumerated no significant difference. In the table 1 statistical description of soil properties in two sets of training and testing is presented. The standard deviation of the amount of gypsum in training set is 0.64 and in the test set is 0.68. This clearly shows that the test dataset is a good representation of the dataset.

It has been used as an input parameter in PLSR and SVR models to estimate soil gypsum using wavelengths of the visible-near-infrared range (400-2500 nm). In predicting the percentage of gypsum by two models PLSR and SVR, the highest value of R^2 and the lowest value of RMSE were obtained in each training and test sets (Table 2). The results of this study are consistent with the research which was done by Chaternor et al., 2020. The results clearly show that the SVR model is better for estimating soil gypsum than PLSR model. Research has been conducted using PLSR and SVR models in 72 soil spectral samples in Iran. Their results evidenced that SVR model has the highest performance in estimating soil CEC.

In another study, Khayamim et al. (2015) obtained an excellent yield (RPIQ> 2) for the amount of soil gypsum using the PLSR model.

Also, SVR had the shortest distance from the line (1: 1) and the best fit (Figure 2).

In general, according to Table 2 and Figure 2, the results clearly show that the performance of SVR is much better than PLSR for estimation of soil gypsum content. Therefore, it can be concluded that SVR is a more suitable multivariate method for soil spectral data.

According to Nawar et al. 2016 research, the range of changes in the concentration of soil properties has an undeniable role in the accuracy assessment of the regression model and with an increase in changes and data breadth and also an increase in range, the model's accuracy estimation increases.

Also according to the Wilding (1985), the extent in which the data with CV in the range of > 35 is considered as a large extent. In the present study, the CV for train and test datasets are 65.97% and 68.68 % respectively, which indicates the appropriate breadth in the Collected data and has improved the accuracy of gypsum estimation in both models.

In spectral curves there is a peak in the range of 700-750 nm which can be attributed to the highest reflectance owning to bright mineral's presence namely calcites and carbonates. Moreover, Iron oxides and organic matter greatly affect the reflectance in visible region, and their existence cause soil color to be darken and spectral reflectance to decrease (Hant, 1970). Chaternor et al. (2020) reported a noticeable correlation in wavelengths of 1450, 1550, 1700, 2100, 2200 and 2400 nm with gypsum. Hassani et al. (2014) similarly presented wavelengths of 1100-1200, 1450, 1500, 1550, 1650, 1950 and 2200 nm, too. Also, Harrison et al. (2012) mentioned the 1750, 1945, 2100-2200 and 2400 nm wavelengths which are related to SO4 group of gypsum. Also, Hunt (1970) have reported for gypsum wavelengths of 1100 and 680 nm, which were proportional with our results. Khayamim et al., 2015, have shown that the correlation between gypsum content and 1578 nm wavelength can be associated to gypsum amount in soil. Mina et al. (2021) introduced 1827 and around 2300 nm wavelengths in their research and Viscarra Rossel and McBratney (1998) presented Wavelengths around 1600, 2000, and 2100 nm in a model with the aim of estimating the mineral lattice. Wavelengths which are around 2340 nm can also be a representative of illite or the presence of muscovite minerals mixtures (Post and Noble, 1993).

5. Conclusion

In the current research, we have done an exploration in reflectance spectroscopy ability with the aim of estimating gypsum content in soil. All in all, the presented results clarified that a correlation between the soil gypsum content and soil spectral reflectance exist. Wavelengths nearly at 1900, 2338, 2200, and 1400 nm were identified as key spectral bands for gypsum content assessment. Choosing between the two models, the ML algorithm resulted in better performance comparing with the commonly used PLSR method. Our outcomes have proven that spectral reflectance is an undeniable efficient tool for sufficiently assessing large areas.

Therefore, it can be said that soil spectral reflections can be used as a rapid and alternative method in soil.

To obtain a comprehensible knowledge of ML method's performances in soil science studies, we suggest a comparison of other data mining techniques namely Random Forest (RF) and Artificial Neural Networks (ANN), for the future studies.

Acknowledgement

We would like to thank the reviewers for their constructive comments and suggestions.

Author contributions

Monire Mina: Methodology, Software, Formal analysis, Investigation, Resource, Writing – Orginal Draft, Visualization,
Mahrooz Rezaei: Coceptulization, Review & Editing, Supervision, Funding acquisition,
leila Hossein Abadi: Writing – Review & Editing, Software,
Abdolmajid Sameni: Methodology, Review, Funding acquisition, Supervision.

Conflicts of interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

Research and publication ethics were complied with in the study.

References

Akbarifazli, R., Babaeinejad, T., Ghanavati, N., Hasani, A., & Askari, M. S. (2021). Assessing the Visible–Near-Infrared Spectroscopy Method and PLSR and SVMR Methods in Modeling Organic Carbon and Total Neutralizing Value of Soil. Iranian Journal of Soil and Water Research, 52(4), 1011-1023.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., & Barthès, B. G. (2016). Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. *Geoderma*. *261*, 151-159.

Chatrenor, M., Landi, A., Farrokhian Firouzi, A., Noroozi, A., & Bahrami, H. A. (2020). Application of hyperspectral images in Quantification of soil gypsum in center areas of Khuzestan province prone to dust generation. *Applied Soil Research*, *8*(3), 1-13.

Clark, R. N., Swayze, G. A., Singer, R. B., & Pollack, J. B., 1990. High-resolution reflectance spectra of Mars in the 2.3-μm region: Evidence for the mineral scapolite. *Journal of Geophysical Research: Solid Earth*. *95*(B9), 14463-14480.

Deiss, L., Margenot, A. J., Culman, S. W., & Demyan, M. S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, *365*, 114227.

Farifteh, J., Van der Meer, F., Atzberger, C., & Carranza, E. J. M. (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*. *110*(1), 59-78.

Gholizadeh, A., Borůvka, L., Saberioon, M., & Vašát, R. (2013). Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Applied spectroscopy*, *67*(12), 1349-1362.

Gomez, C., Lagacherie, P., & Coulouma, G., 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*. *148*(2), 141-148.

Harrison, T. N. (2012). Experimental VNIR reflectance spectroscopy of gypsum dehydration: Investigating the gypsum to bassanite transition. *American Mineralogist*, *97*(4), 598-609.

Hassani, A., Bahrami, H., Noroozi, A., & Oustan, S. (2014). Visible-near infrared reflectance spectroscopy for assessment of soil properties in gypseous and calcareous soils. *Watershed Engineering and Management*, *6*(2), 125-138.

Hunt, G. R. (1970). Visible and near-infrared spectra of minerals and rocks: I silicate minerals. *Modern geology*, *1*, 283-300.

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of statistical software*, *15*, 1-28.

Khayamim, F., Wetterlind, J., Khademi, H., Robertson, A. J., Cano, A. F., & Stenberg, B. (2015). Using visible and near infrared spectroscopy to estimate carbonates and gypsum in soils in arid and subhumid regions of Isfahan, Iran. *Journal of Near Infrared Spectroscopy*, *23*(3), 155-165.

Lacerda, M. P., Demattê, J. A., Sato, M. V., Fongaro, C. T., Gallo, B. C., & Souza, A. B. (2016). Tropical texture determination by proximal sensing using a regional spectral library and its relationship with soil classification. *Remote Sensing*, *8*(9), 701.

Lagacherie, P., Baret, F., Feret, J. B., Netto, J. M., & Robbez-Masson, J. M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*. *112*(3), 825-835.

Mehrabi Gohari, E., Matinfar, H. R., Jafari, A., Taghizadeh-Mehrjardi, R., & Khayamim, F. (2020). Comparing Different Statistical Models and Pre-processing Techniques for Estimation several chemical properties of the soil Using VNIR/SWIR Spectrum. *Iranian Journal of Remote Sensing & GIS*, *11*(4), 47-60.

Mina, M., Rezaei, M., Sameni, A., Moosavi, A. A., & FALLAH SHAMSI, R. A. S. H. I. D. (2022). Using Soil Pedotransfer and Spectrotransfer Functions to Estimate Cation Exchange Capacity in Calcareous Soils, Fars Province. *Iranian Journal of Soil and Water Research*, *52*(11), 2911-2922.

Mina, M., Rezaei, M., Sameni, A., Moosavi, A. A., & Ritsema, C. (2021). Vis-NIR spectroscopy predicts threshold velocity of wind erosion in calcareous soils. *Geoderma*, *401*, 115163.

Mirzaei, S., Boloorani, A. D., Bahrami, H. A., Alavipanah, S. K., Mousivand, A., & Mouazen, A. M. 2022. Minimising the effect of moisture on soil property prediction accuracy using external parameter orthogonalization. *Soil and Tillage Research*, *215*, 105225.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., & Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research*, *155*, 510-522.

Nelson, R.E. (1982). Carbonate and gypsum. In: Page, A.L. (Ed.), Methods of Soil Analysis:Part1Agronomy Handbook 9. *American Society of Agronomy and Soil Science Society of America, Madison (WI)*. 6, 181–197.

Ostovari, Y., Ghorbani-Dashtaki, S., Bahrami, H. A., Abbasi, M., Dematte, J. A. M., Arthur, E., & Panagos, P. (2018). Towards prediction of soil erodibility, SOM and CaCO3 using laboratory Vis-NIR spectra: A case study in a semi-arid region of Iran. *Geoderma*, *314*, 102-112.

Post, J. L., & Noble, P. N. (1993). The near-infrared combination band frequencies of dioctahedral smectites, micas, and illites. *Clays and clay minerals*, *41*(6), 639-644.

Rossel, R. V., & McBratney, A. B. (1998). Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma*, *85*(1), 19-39.

Smith, P. A., & Robertson, J. E. (1962). Some factors affecting the site of alkylation of oxime salts. *Journal of the American Chemical Society*, *84*(7), 1197-1204.

Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, *9*.

Wilding, L. P. (1985). Spatial variability: its documentation, accomodation and implication to soil surveys. In *Soil spatial variability, Las Vegas NV, 30 November-1 December 1984* (pp. 166-194).

Wu, W., Zucca, C., Muhaimeed, A. S., Al-Shafie, W. M., Fadhil Al-Quraishi, A. M., Nangia, V., ... & Liu, G. (2018). Soil salinity prediction and mapping by machine learning regression in C entral M esopotamia, I raq. *Land degradation & development*, *29*(11), 4005-4014.