



## Spectral behavior modeling of soil texture over dust center of Khuzestan Province using hyperspectral images and Random Forest (RF) model

Mansour Chaternour<sup>1</sup>, Ahmad Landi<sup>1\*</sup>, Ahmad Farrokhian Firouzi<sup>1</sup>, Aliakbar Noroozi<sup>2</sup>, Hosseinali Bahrami<sup>3</sup>

1-Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

2-Soil Conservation and Watershed Management Research Institute, Tehran, Iran

3-Department of Soil Science, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran

**Keywords:** Random forest model, Second derivative filter, Continuum removal filter, Key wavelength, Main factors

### 1-Introduction

In recent years, some areas of Khuzestan Province have experienced excessive drought and have become dust production centers due to the decrease in soil resistance and loss of vegetation. Combination of these factors along with poor management, has led the soil erosion resistance to reduce against winds. Among soil properties, texture as an essential characteristic plays a crucial role in soil resistance to wind and rain erosive factors and affects water movement and soil fertility (Hillel, 1980).

Since traditional methods of soil texture measurement are costly and time-consuming, in recent decades researchers, have used novel methods such as Remote Sensing and reflectance spectroscopy to estimate soil properties, especially for large areas (Ben-Dor et al., 2009; Curcio et al., 2013).

One of the limiting factors in the evaluation of soil properties by spectroscopy is the identification of pre-processing methods in noise and error elimination and determining the appropriate regression model to estimate these properties (Mohamed et al., 2017). Therefore, proper regression and pre-processing methods are required to determine soil properties using soil reflectance. Xuemei and Jianshe (2013) used PLSR and LS-SVM models to investigate soil properties. Based on their results, the LS-SVM model estimated the properties of organic matter, nitrogen, phosphorus, and potassium with determination coefficients of 0.87, 0.82, 0.76, and 0.73, respectively. Silva et al. (2016) used the PLSR model spectroscopy and second derivative methods to determine the soil texture of southwestern Greece. Based on evaluation of coefficients of determination and root mean square error, the results of their study showed the accuracy percentage of sand, silt and clay were 0.3, 0.59 and 0.69 and 5.47, 5.18, 5.39 respectively in 100 grams of soil. Wang et al. (2018) used partial least squares regression (PLSR) and random forest (RF) methods to estimate soil salinity and based on their research results, the RF model had better performance than the PLSR model.

In the context of considering the complicated relationship between soil properties and their reflectance, it is necessary to use the spectroscopic method to determine the best statistical model for spectral analysis in different regions. In this regard, the objectives of this study are 1-Estimation of clay, silt, and sand percentages characteristics of productive dust soil of Khuzestan province by PLS-RF model, 2- Comparison of PLS-RF model performance and accuracy in 6 spectral methods including: Main Spectrum, Savitzky-Golay filter, first derivative with the Savitzky-Golay filter (FD-SG), second derivative with the Savitzky-Golay filter (FD-SG), Standard Normalization Method (SNV) and Continuous Removal Method (CR), and 3-Determining the key wavelengths of soil texture in these areas.

### 2-Material and methods

The area under study was in a geographical coordinate range between 30°24' to 31°19' and 49°21' to 49°28'. The area is located in south and southeast of Ahvaz toward the south of Khuzestan province which is about 110,000 hectares. One hundred forty-two soil samples were collected from 0 to 5 cm depth by the systematic-random method. Hydrometer method was also used to determine soil texture.

\* Corresponding author: landi@scu.ac.ir

DOI: 10.22055/aag.2019.29642.1990

Received 2019-05-21

Accepted 2019-10-12



Spectroscopy: ASD FieldSpec3 laboratory spectrometer was used to determine soil reflectance. At the laboratory, a small amount of soil sample was transferred to a Petri dish with a diameter of 8 cm and a depth of 2 cm. Spectral measurements were carried out using three separate detector types in the range 2500–3500 nm, which is the range from visible to near-infrared. For each sample, ten reflectance spectra were measured, and ten replicates for each soil sample were averaged using ViewSpect software then stored as a spectral library in the spectral library.

Spectrum analysis: Pre-processing was performed on the primary spectrum, including types of filters using software and then modeling and estimation of soil texture by PLS-RF method.

### 3-Findings

The results of PLS-RF model showed that for clay, the highest accuracy belonged to the continuum removal method and the least accuracy belonged to the primary spectrum. For sand percentage, second derivative and SG methods had the highest and least accuracy, respectively. Finally, for the silt, the least accuracy was obtained rather than clay and sand percentages. According to the results, the highest and least estimation accuracy was observed in continuous removal (CR) and Savitzki-Golay straightening methods, respectively.

### 4-Conclusion

In this study, the performance of PLS-RF model in the primary spectrum and five pre-processing methods was compared to estimate the silt, sand, and clay properties in the soil of dust center areas of Khuzestan province. According to comparison with pre-processing methods in estimating soil properties, it was observed that continuum removal method in both clay and silt percentage had the best performance and for the sand percentage, the second derivative method showed the best accuracy estimation.

Based on the results of spectral correlation in this study, the key wavelengths for clay percentage were in the wavelength range 1200-1210, 1800 and 2200 nm, for sand percentage in the wavelength range 1400-150, 1930-1930, 2200 and 2220 nm and for soil silt percentage The wavelengths of 1320, 1615 and 2200 nm were observed. Therefore, the use of PLS regression method and determination of significant components in each spectral group resulted in reducing computational effort, increasing the speed of computational processing and finally obtained the optimum performance in estimating soil properties.

### References

- Hillel, D., 1980. Applications of soil physics. Academic Press, San Diego.
- Ben-Dor, E., Chabrilat, S., Demattê, J., Taylor, G., Hill, J., Whiting, M., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment* 113, 38-55.
- Curcio, D., Ciraolo, G., D'Asaro, F., Minacapilli, M., 2013. Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences* 19, 494-503.
- Mohamed, E., Saleh, A., Belal, A., Gad, A.A., 2017. Application of near-infrared reflectance for quantitative assessment of soil properties. *The Egyptian Journal of Remote Sensing and Space Science* 21, 1-14.
- Xuemei, L., Jianshe, L., 2013. Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration. *Measurement* 46(10), 3808-3814.
- Silva, E.B., Ten Caten, A., Dalmolin, R.S.D., Dotto, A.C., Silva, W.C., Giasson, E., 2016. Estimating Soil Texture from a Limited Region of the Visible/Near-Infrared Spectrum. *Digital Soil Morphometrics*, Springer, pp. 73-87.
- Wang, J., Ding, J., Abulimiti, A., Cai, L., 2018. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. *Peerj* 6, 4703.

## مدل سازی رفتار طیفی بافت خاک در کانون های ریزگرد استان خوزستان با استفاده از تصاویر ابر طیفی و مدل جنگل تصادفی

منصور چترنور

گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

احمد لندی\*

گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران  
مرکز پژوهشی منطقه ای ریزگردها، دانشگاه شهید چمران اهواز، اهواز، ایران

احمد فرخیان فیروزی

گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

علی اکبر نوروزی

پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

حسینعلی بهرامی

گروه خاکشناسی، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران

تاریخ پذیرش: ۱۳۹۸/۰۷/۲۰

تاریخ دریافت: ۱۳۹۸/۰۲/۳۱

\*landi@scu.ac.ir

### چکیده

بافت خاک نقش مهمی در مقاومت خاک به فرسایش بادی دارد. تصاویر ابر طیفی با مزیت هزینه پایین و سرعت عمل بالا، ابزار مناسبی برای بررسی ویژگی های خاک از جمله بافت محسوب می شوند. هدف این مطالعه ارزیابی رفتار طیفی درصد رس، شن و سیلت در خاک های مستعد تولید ریز گرد استان خوزستان با استفاده از مدل PLS-RF است. در ابتدا فاکتورهای اصلی با مدل رگرسیون حداقل مربعات جزئی تعیین و سپس مدل جنگل تصادفی روی فاکتورهای تعیین شده اجرا گردید. در مرحله بعد عملکرد طیف اصلی و پیش پردازش های: فیلتر ساویتزکی و گلای، فیلتر ساویتزکی و گلای به همراه مشتق اول، فیلتر ساویتزکی و گلای به همراه مشتق دوم، روش نرمال سازی استاندارد و روش حذف پیوستار در حذف نویز و افزایش دقت مدل PLS-RF مقایسه شد. نتایج نشان داد که روش حذف پیوستار در دو ویژگی درصد رس ( $RPDCAL = 1/98$ ) و درصد سیلت ( $RPDCAL = 1/65$ ) و روش مشتق دوم برای درصد شن ( $RPDCAL = 1/97$ )، بهترین عملکرد را داشته اند. همچنین طول موج کلیدی برای درصد رس در طول موج های ۱۲۰۰-۱۲۱۰، ۱۸۰۰ و ۲۲۰۰ نانومتر، برای درصد شن در محدوده طول موج های ۱۴۰۰-۱۴۵۰، ۱۹۱۰-۱۹۳۰، ۲۲۰۰ و ۲۲۲۰ نانومتر و برای درصد سیلت خاک در محدوده طول موج های ۱۳۲۰، ۱۶۱۵ و ۲۲۰۰ نانومتر مشاهده گردید.

**کلمات کلیدی:** فیلتر ساویتزکی و گلای، فیلتر مشتق دوم، فیلتر حذف پیوستار، طول موج کلیدی، فاکتور اصلی

### ۱- مقدمه

خصوصیت مهم و اثرگذار در محیط زیست شناخته می شود. بافت خاک نقش مهم و کلیدی در مقاومت خاک در برابر عوامل فرساینده آبی و بادی دارد و در حرکت آب، باروی و حاصلخیزی خاک اثرگذار است (Hillel, 1980).

در روش های سنتی، برای بررسی دانه بندی و بافت، نیاز به جمع آوری نمونه و تجزیه و تحلیل آزمایشگاهی است. به دلیل هزینه زیاد و زمان بر بودن این روش ها، در دهه های اخیر پژوهشگران از روش های نوینی از جمله دورسنجی و طیفسنجی بازتابی در برآورد خصوصیات خاک، به ویژه در سطوح وسیع استفاده می کنند (Ben-Dor et al., 2009; Curcio et al., 2013). از عوامل محدودکننده در ارزیابی ویژگی های خاک با روش

شوری و سدیمی شدن خاک از جمله مهم ترین مسائل زیست محیطی در مناطق خشک و نیمه خشک محسوب می شود (Ranjbar and Jalali, 2016). در سال های اخیر پدیده خشک سالی سبب تجمع نمک، کاهش حاصلخیزی و تغییرات فیزیکی و مکانیکی در خاک سطحی شده است. مناطقی از استان خوزستان با این پدیده مواجه شده و با کاهش مقاومت خاک و از بین رفتن پوشش گیاهی، به کانون های تولید ریزگرد تبدیل شده اند. مجموعه این عوامل تحت تأثیر مدیریت نامطلوب، سبب کاهش مقاومت خاک در برابر باد فرساینده شده است. در بین خصوصیات خاک، بافت به عنوان یک

عملکرد بهتری ( $R^2 = 0.93$ ), ( $RMSE = 4.57$ ) و ( $RPD \leq 2.50$ ) داشت.

مناطق مستعد تولید ریز گرد استان خوزستان سطح گسترده‌ای دارند و در بعضی نواحی، ناهمگونی در ویژگی‌های خاک مشاهده می‌شود. در بین ویژگی‌های خاک، بافت خاک تأثیر قابل توجهی در مقاومت خاک به باد فرساینده و تولید ریز گرد دارد. استفاده از روش‌های نوین مانند طیف سنجی می‌تواند نقش مهمی را در ارزیابی بافت خاک در این مناطق داشته باشد. با توجه به رابطه پیچیده ویژگی‌های خاک با بازتاب آنها نیاز است روش طیف بینی برای تعیین بهترین مدل آماری در تجزیه و تحلیل طیف در مناطق مختلف استفاده گردد.

در این راستا اهداف این پژوهش عبارت‌اند از: ۱- برآورد ویژگی‌های درصد رس، سیلت و شن در خاک‌های مستعد تولید ریز گرد استان خوزستان با روش مدل PLS-RF، ۲- مقایسه کارایی و دقت مدل PLS-RF در ۶ روش طیفی شامل: طیف اصلی، فیلتر ساویتزکی و گلای، فیلتر ساویتزکی و گلای به همراه مشتق اول (FD-SG)، فیلتر ساویتزکی و گلای به همراه مشتق دوم (SD-SG)، روش نرمال سازی استاندارد (SNV) و روش حذف پیوستار (CR) و ۳- تعیین طول موج‌های کلیدی بافت خاک در این مناطق.

## ۲- مواد و روش‌ها

### ۲-۱- منطقه مورد مطالعه

منطقه مورد مطالعه در محدوده‌ای با مختصات جغرافیایی: بین طول جغرافیایی  $31^{\circ} 19'$  تا  $30^{\circ} 24'$  و عرض جغرافیایی  $49^{\circ} 21'$  تا  $49^{\circ} 28'$  بود، این منطقه در جنوب و جنوب شرق اهواز به سمت جنوب استان خوزستان است و مساحت حدود ۱۱۰ هزار هکتار را شامل می‌شود. فصل گرما از اواسط اردیبهشت شروع و تا اواخر مهرماه ادامه دارد، متوسط درجه حرارت  $31/2$  درجه سانتی‌گراد و میانگین بارش ۲۶۱ میلی‌متر دارد. از نظر زمین‌شناسی در گروه دشت‌های رسوبی و زمین‌های پست و شور قرار دارد و به دلیل صعود مویینه، املاح در سطح خاک تجمع یافته است. میانگین شوری این مناطق  $34/8$  دسی‌زیمنس بر متر است و در مناطقی شوری به ۹۶ دسی‌زیمنس بر متر رسیده و بلورهای نمک در سطح خاک نمایان شده است. همچنین در نقاطی از سطح زمین ویژگی سدیمی بروز و رنگ قهوه‌ای نمایان شده است.

### ۲-۲- نمونه برداری

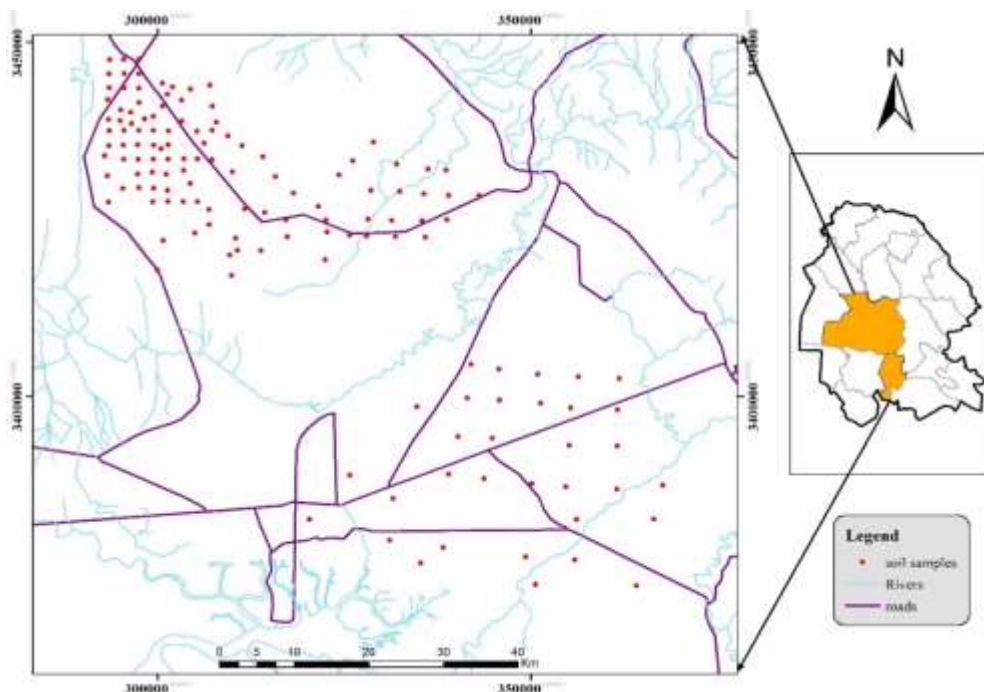
در ابتدا منطقه به صورت شبکه‌های  $2 \times 2$  کیلومتری تقسیم‌بندی و نمونه‌برداری به صورت سیستماتیک و تصادفی انجام شد. برای نمونه برداری خاک، ۱۴۲ نمونه خاک از عمق صفر تا ۵ سانتیمتری جمع‌آوری شد. موقعیت نقاط نمونه برداری در شکل ۱ ارائه شده است. برای تعیین بافت خاک در آزمایشگاه، نمونه‌های خاک هوا خشک، سپس با کوبیدن از الک دو میلی‌متری عبور داده شد. در آزمایشگاه با استفاده از روش هیدر متر بافت خاک اندازه‌گیری شد.

طیف‌سنجی، شناسایی روش‌های پیش‌پردازش در حذف نویز و خطا و تعیین مدل رگرسیونی مناسب در برآورد این ویژگی‌ها است (Mohamed et al., 2017) بنابراین به روش رگرسیونی و پیش‌پردازش مناسبی نیاز است تا بتوان ویژگی‌های خاک با استفاده از بازتاب خاک را تعیین کند.

در دو دهه اخیر پژوهش‌های گسترده‌ای در ارزیابی خصوصیات خاک با روش طیف‌سنجی (۴۰۰-۲۵۰۰ نانومتر) و انواع مختلف روش‌های آماری چند متغیره انجام شده است. بیلگلی و همکاران (Bilgili et al., 2010) از روش طیف‌سنجی برای بررسی خصوصیات خاک‌های شمال ترکیه با دو مدل (PLSR) و (MARS<sup>1</sup>) استفاده کردند. در پژوهش آن‌ها مدل MARS برآورد قابل قبولی از ویژگی‌های ماده آلی، کربنات کلسیم، درصد رس و شن، کلسیم و منیزیم تبدالی و ظرفیت تبادل کاتیونی را با ضرایب تعیین ( $R^2$ ) به ترتیب  $0.73$ ،  $0.71$ ،  $0.186$ ،  $0.180$ ،  $0.169$ ،  $0.163$  و  $0.173$  نشان داد. خویمی و جانشی (Xuemei and Jianshe, 2013) از مدل‌های PLSR و LS-SVM در بررسی خصوصیات خاک استفاده کردند. در پژوهش آن‌ها دو مدل عملکرد مناسبی نشان داد و مدل LS-SVM ماده آلی، نیتروژن، فسفر و پتاسیم را به ترتیب با ضرایب تعیین  $0.87$ ،  $0.82$ ،  $0.76$  و  $0.73$  برآورد کرد. کورسیو و همکاران (Curcio et al., 2013) در روش طیف‌سنجی مرئی - مادون قرمز از روش پیش‌پردازش حذف پیوستار<sup>۲</sup> و مدل PLSR برای برآورد بافت خاک استفاده کردند. نتایج مطالعه آن‌ها مشخص کرد که مدل PLSR دقت بالایی در برآورد درصد رس ( $RMSE = 0.5/8$  و  $R^2 = 0.87$ )، شن و سیلت به ترتیب  $RMSE = 7/7$  و  $RMSE = 7/2$  و  $R^2 = 0.80$  و  $R^2 = 0.60$  دارد. دوتو و همکاران (Dotto et al., 2014) از روش طیف بینی مرئی و مادون قرمز نزدیک و میانی و رگرسیون چندگانه (SWIR) برای برآورد بافت خاک استفاده کردند. برای شن و ماسه ضرایب تعیین بین  $0.72$  و  $0.77$  متغیر بود و مشخص شد که روش طیف‌سنجی دقت قابل قبولی در شناسایی بافت خاک منطقه دارد. در بررسی ویژگی‌های خاک استان اصفهان، ماده آلی، کربنات کلسیم و گچ را با دقت قابل قبولی با ضرایب تعیین ( $R^2$ )  $0.61$ ،  $0.45$  و  $0.180$  برآورد شده است. با از روش طیف‌سنجی برای برآورد ویژگی‌های کربنات کلسیم و گچ ( $R^2 > 0.5$  و  $RPD > 1.8$ ) عملکرد عالی و برای ماده آلی عملکرد قابل قبولی ( $R^2 > 0.5$  و  $RPD > 1.4$ ) مشاهده شده است. سیلوا و همکاران (Silva et al., 2016) برای تعیین بافت خاک جنوب غربی یونان از روش طیف‌سنجی با مدل PLSR و روش مشتقات دوم استفاده کردند. نتایج مطالعه آن‌ها نشان داد که براساس ارزیابی ضرایب تعیین و ریشه میانگین مربعات خطا، دقت شناسایی درصد ذرات شن، سیلت و رس به ترتیب  $0.3$ ،  $0.59$  و  $0.69$  و  $5/47$ ،  $5/18$ ،  $5/39$ ، در  $100$  گرم خاک است. وانگ و همکاران (Wang et al., 2018) برای برآورد شوری خاک، از دو روش رگرسیون حداقل مربعات جزئی (PLSR) و جنگل تصادفی (RF) با پیش‌پردازش مشتق جزئی (در محدوده صفر تا ۲ و با فاصله  $0.1$ ) استفاده کردند. در نتایج پژوهش آن‌ها مدل RF نسبت به مدل PLSR

1- Multivariate adaptive regression splines

2- Continuum Removal



شکل ۱- موقعیت کانون ریز گرد و محل‌های نمونه‌برداری (نقاط قرمز رنگ).

Fig. 1. Dust center situation and soil samples location (red color points).

گردید. به علاوه دو وقفه حاصل از تغییر آشکارساز<sup>۲</sup> در محدوده ۹۰۰ و ۱۷۰۰ نانومتر نیز حذف شد (Rossel et al., 2009).

در ادامه پیش پردازش‌هایی بر روی طیف اصلی انجام گرفت که شامل: فیلتر Savitzky-Golay (به صورت تابع چندجمله‌ای درجه دوم و نیز تعداد ۲۳ نقطه هموارساز)، فیلتر Savitzky-Golay به همراه مشتق اول (FD-SG) و مشتق دوم (SD-SG)، روش نرمال استاندارد به همراه فیلتر Savitzky-Golay با استفاده از نرم‌افزار The Unscrambler انجام شد (CAMO, 1998). در ادامه با استفاده از نرم‌افزار Matlab روش حذف پیوستار اجرا شد.

مشتق‌گیری با اختلافات دیفرانسیلی نقاط عطف و بازتاب‌های حداکثر را با سرعت و وضوح بیشتری نشان می‌دهد. حساسیت داده‌های طیفی نوپزدار با مشتقات کاهش و سبب افزایش وضوح ویژگی‌های جذب می‌شود. رابطه مشتق اول و دوم به صورت زیر است (He et al., 2009):

$$R'(\lambda_i) = [R(\lambda_i) - R(\lambda_{i-1})] / 2 \times \Delta\lambda \quad (1)$$

$$R''(\lambda_i) = [R'(\lambda_i) - R'(\lambda_{i-1})] / 2\Delta\lambda \\ = [R(\lambda_{i+1}) - 2R(\lambda_i) + R(\lambda_{i-1}))] / \Delta\lambda^2 \quad (2)$$

### ۳-۲- طیف‌سنجی

برای تهیه بازتاب خاک نمونه‌ها از طیف‌سنج آزمایشگاهی ASD FieldSpec3 استفاده شد. در آزمایشگاه، مقداری خاک از هر نمونه به پتری دیش با قطر ۸ سانتیمتر و عمق ۲ سانتیمتر منتقل گردید. اندازه‌گیری طیفی با سه نوع آشکارساز جداگانه در محدوده ۳۵۰۰-۲۵۰۰ نانومتر، که شامل محدوده مرئی تا مادون قرمز نزدیک است؛ انجام شد. نمونه‌های خاک در فاصله ۲۰ سانتیمتری و زاویه ۴۵ درجه از لامپ هالوژنی دستگاه قرار گرفتند. اندازه‌گیری طیفی به روش غیر تماسی و در ارتفاع ۳ سانتی‌متر از بالای نمونه‌ها انجام شد. زمان ادغام، یک ثانیه تنظیم گردید. برای واسنجی نوری دستگاه طیف‌سنج از یک صفحه سفید با ضریب انعکاس معلوم و برابر یک، استفاده شد. تمام اندازه‌گیری‌های طیفی در یک اتاق کاملاً تاریک انجام شد تا از تداخل نور جانبی و ایجاد خطا جلوگیری شود. برای هر نمونه ۱۰ طیف بازتابی اندازه‌گیری شد، ۱۰ طیف تکراری برای میانگین‌گیری با نرم‌افزار ViewSpect انجام شد و به عنوان طیف نمونه خاک در کتابخانه طیفی ذخیره گردید. هر منحنی در محدوده ۳۵۰-۲۵۰۰ نانومتر طیف ۲۰۵۰ بازتاب طیفی (۲۰۵۰ طول موج) را شامل می‌شود.

### ۴-۲- روش‌های پیش‌پردازش و کاهش حجم داده‌های طیفی

به دلیل وجود خطا و وجود نویز احتمالی در بازتاب خاک، طیف اصلی باید با روش‌های پیش‌پردازش اصلاح شود. هدف اصلی از پیش‌پردازش، حذف آشفتنگی طیفی (نویز)، ارتقاء کیفی داده‌ها و استفاده بهینه از داده‌های طیفی در برآورد خصوصیات خاک است (Fearn et al., 2009). برای انجام پیش‌پردازش، ابتدا دو بخش نویز دار در ابتدا و انتهای طیف که به ترتیب در محدوده بین ۳۵۰ تا ۴۵۰ و ۲۵۰۰-۲۴۵۰ نانومتر قرار دارند حذف



۸ بیانگر رابطه نهایی مدل است (Sjöström et al., Ji et al., 2016; 1983):

$$\hat{y} = \hat{b}_i x_i + \hat{b}_0 \quad (8)$$

$\hat{y}$ : ویژگی برآورد شده با مدل،  $\hat{b}_0$ : مرز جداکننده و  $\hat{b}_i$ : بردارهای رگرسیونی هستند. مدل جنگل تصادفی (RF) یک جنگل تصادفی، از درخت‌های هرس نشده تشکیل شده است و هر درخت از روش الگوریتم افزایشی بازگشتی<sup>۱</sup> تعیین می‌گردد.

در ساخت جنگل در ابتدا یک نمونه تصادفی از تعداد  $n$  مشاهده انتخاب می‌گردد، در نمونه انتخاب شده، از الگوریتم افزایشی بازگشتی استفاده می‌شود تا یک درخت رشد کند، در هر گره افزار از  $m$  متغیر تصادفی با  $P$  متغیر برآوردی، پیش‌بینی انجام می‌شود.

الگوریتم هنگامی که هر درخت به بزرگترین اندازه خود رسید، متوقف می‌شود. اجرای مدل به تعداد  $T$  دفعه است و تعداد دفعات  $T$  با توجه به نوع مطالعه در حدود ۱۰۰۰ مرتبه در نظر گرفته می‌شود (Genauer et al., 2008). الگوریتم تطبیقی این مدل (Breiman, 1999) سبب کاهش انتخاب گزینشی داده می‌شود و روشی موثر در طبقه‌بندی و رگرسیون است. هرچه مجموعه داده ورودی بزرگ‌تر و روند آن‌ها تصادفی باشد، میزان دقت مدل بیشتر است (Breiman, 2001).

این مدل قابلیت برآورد خصوصیات مختلف با نویز بالا را دارد (Chen and Liu, 2005). مدل‌سازی مجموعه داده بزرگ، قدرت تفکیک بالا، نشان دادن طول موج‌های متوالی و تنوعی از اطلاعات طیفی از ویژگی‌های دیگر این مدل است (Wang et al., 2017). رابطه کلی مدل به صورت زیر است:

$$F(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (9)$$

$$D = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (10)$$

در روابط ۹ و ۱۰  $F(x)$ : مدل نهایی یا میانگین درخت‌های تصادفی،  $D$ : مجموعه داده‌های اصلی (طیف و EC) ورودی به مدل  $B$  مجموعه درخت‌های تصمیم که از داده‌های اصلی ( $D$ ) به روش تصادفی نمونه‌گیری با جایگزینی تعیین می‌شود و  $T_b$  بهترین متغیر تصادفی یا درخت تصادفی است که به دفعات انتخاب می‌شود.

در رابطه ۲  $\lambda_i$  طول موج برای هر باند،  $R'(\lambda_i)$  و  $R''(\lambda_i)$  به ترتیب مشتق اول و دوم طیف‌های مختلف برای طول موج  $\lambda_i$ ،  $\Delta\lambda$ : فاصله بین طول موج  $\lambda_i$  و  $\lambda_{i-1}$ . با افزایش میزان  $\Delta\lambda$ ، منحنی دیفرانسیل طیف به سمت صاف شدن بیشتر تمایل دارد، که سبب حذف بسیاری از ویژگی‌های ظریف طیفی می‌شود. در این مطالعه،  $\Delta\lambda = 10$  نانومتر انتخاب شده است. واریانس نرمال استاندارد (SNV) نوعی تغییر جهت‌دار در منحنی طیف است و یک داده را در جهتی تغییر می‌دهد که درون یک ستون یا دسته از داده متمرکز شود (CAMO, 1998).

روش حذف پیوسته سبب حذف نویز جانبی طیف می‌شود و با تغییر شکل منحنی طیف از حالت بازتابی به جذبی، ویژگی‌های جذبی ضعیف را در منحنی نشان می‌دهد (Wenjun et al., 2014). در این روش معمولاً ویژگی‌های جذبی انتخابی مورد استفاده قرار می‌گیرد و ویژگی‌های جذب منفرد بر اساس یک خط مشترک مورد مقایسه قرار می‌گیرد. این روش از نسبت مقدار بازتاب در انحنای جذبی ( $R(\lambda_i)$ ) به سطح بازتاب خط پیوستار در طول موج متناظر ( $R_c(\lambda_i)$ ) به دست می‌آید (Mutanga and Skidmore, 2003).

$$R'(\lambda_i) = \frac{R(\lambda_i)}{R_c(\lambda_i)} \quad (3)$$

تعداد ۱۴۲ نمونه خاک به منظور مدل‌سازی انتخاب شد. در مرحله بعد از روش رگرسیون حداقل مربعات جزئی (PLSR) برای تعیین فاکتورهای معنادار در هر دسته از طیف‌های اصلی و پیش‌پردازش‌ها استفاده گردید. با انتخاب و جداسازی فاکتورهای معنادار در هر گروه، سایر فاکتورها از فرایند مدل‌سازی حذف گردید. فاکتورهای معنادار محاسبه شده تا ۹۷ درصد واریانس کل داده‌ها را تشکیل می‌دهند. مدل جنگل تصادفی بر اساس همبستگی بین خصوصیات درصد رس، شن و سیلت در این فاکتورها مورد استفاده قرار گرفت.

## ۲-۵- روش PLSR

مدل PLSR دارای رابطه خطی و قابلیت برقراری رابطه بین ویژگی‌های خالص خاک و بازتاب طیف در محدوده VIS-NIR است (Farifteh et al., 2007; Nawar et al., 2014). رابطه مدل به صورت زیر است:

$$X = TP^T + E \quad (4)$$

$$y = Tq + f \quad (5)$$

$$X = [x_1, x_2, \dots, x_n] \quad (6)$$

$$T = \omega^T X \quad (7)$$

در رابطه ۴،  $X$ : داده‌های طیفی (متغیر مستقل)،  $y$ : ویژگی خاک (متغیر وابسته)،  $T$ : متغیر پنهانی که ترکیبی خطی و از طیف محاسبه می‌شود،  $\omega$  مقیاس وزنی و از روش بردارهای ویژه  $X'yy'X$  محاسبه می‌شود.  $P$ : بار طیف و  $q$  بار خصوصیت شیمیایی خاک و میزان وابستگی طیف و خصوصیات خاک را به متغیر پنهانی  $T$  نشان می‌دهند.  $E$  و  $f$  مقادیر باقیمانده که نشان‌دهنده نویز یا متغیرهای بی‌ربط در مدل هستند. رابطه

**۲-۶- ضریب همبستگی**

در این پژوهش ضریب همبستگی بین بازتاب طیفی با ویژگی‌های خاک از رابطه ۱۱ و ۱۲ محاسبه شد (He et al., 2009).

$$r_i = \frac{\text{cov}(R, S)}{\sqrt{D(R)} \cdot \sqrt{D(S)}} \quad (11)$$

$$r_i = \frac{\sum_{n=1}^N (R_{ni} - \bar{R})(OM_n - \bar{S})}{\sqrt{\sum_{n=1}^N (R_{ni} - \bar{R})^2 \sum_{n=1}^N (OM_n - \bar{S})^2}} \quad (12)$$

در این رابطه S ویژگی وابسته (ویژگی خاک مورد بررسی)، IR ضریب همبستگی یکنواخت بین مقدار ویژگی خاک و طیف بازتابی و یا تغییرات آن، R<sub>ni</sub>: بازتاب بازتاب نمونه nام در طول موج یا باند موج Nام،  $\bar{R}$ : میانگین بازتاب طیفی یا بازتاب طیفی تغییر یافته برای N نمونه در i موج باند یا طول موج، S: غلظت، درصد یا مقدار ویژگی خاک اندازه‌گیری شده در نمونه n خاک،  $\bar{S}$ : میانگین ویژگی خاک، N تعداد کل نمونه خاک است.

**۲-۷- معیارهای ارزیابی**

برای ارزیابی دقت مدل‌ها از سه معیار ارزیابی مجذور میانگین مربعات خطا (RMSE)، ضریب تعیین (R<sup>2</sup>) و نسبت عملکرد به انحراف (RPD) در دو گروه واسنجی (۷۰ درصد داده‌ها) و اعتبار سنجی (۳۰ درصد داده‌ها) استفاده شد (Ji et al., 2016; Chang et al., 2001).

$$R^2 = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (14)$$

$$PRD = \frac{SD}{RMSE} \quad (15)$$

$$SD = \sqrt{\frac{\sum_{i=1}^n (O_i - \bar{O})^2}{n-1}} \quad (16)$$

n تعداد داده‌ها، O<sub>i</sub> داده‌های مشاهداتی، P<sub>i</sub> داده‌های پیش‌بینی شده و علامت بار بیانگر متوسط متغیرها است. آماره RPD عملکرد مدل را ارزیابی می‌کند. مقادیر کمتر از ۱/۴، بین ۱/۴-۲ و بیشتر از ۲ به ترتیب نشان‌دهنده ضعیف، قابل قبول و عالی هستند. دقت گروه قابل قبول (مقادیر ۱/۴-۲) با به کارگیری روش‌های مختلف واسنجی افزایش می‌یابد؛ اما اگر RPD کوچک‌تر از ۱/۴ باشد مدل قابلیت برآورد ندارد (Chang et al., 2001).

**۳- نتایج و بحث**

پارامترهای آماری ویژگی‌های درصد رس، شن و سیلت نمونه‌های خاک برداشت شده در جدول ۱ ارائه شده است. رس دارای بیشترین میانگین (۳۷/۵) و کمترین میانگین مربوط به درصد سیلت (۲۵/۵) است. بیشترین و کمترین شاخص پراکندگی (۰/۴۳) به ترتیب به رس و شن تعلق دارد. درصد رس دارای چولگی ۰/۹۷ است که این پارامتر از نظر پراکندگی، به سمت مقادیر بزرگ‌تر متمایل است، همچنین مقدار کشیدگی ۰/۱۸ نشان می‌دهد که ارتفاع قله در منحنی توزیع داده‌ها به مقدار جزئی بالاتر از ارتفاع قله توزیع نرمال است. چولگی درصد شن و سیلت به ترتیب برابر ۰/۴۶ و ۰/۳۶ است؛ که نشان‌دهنده توزیع تقریباً نرمال است. درصد شن و سیلت با مقدار کشیدگی به ترتیب ۰/۷۹- و ۰/۸۴- قله توزیع پایین‌تری نسبت به منحنی توزیع نرمال دارند. همبستگی پیرسون بین درصد ذرات رس، شن و سیلت در جدول ۲ نشان داده شده است. همان‌طور که مشاهده می‌شود همبستگی مثبت و منفی بین خصوصیات خاک وجود دارد. درصد رس با شن خاک دارای همبستگی معنی‌دار ۰/۷۶- در سطح یک درصد و درصد سیلت با رس و شن همبستگی معنادار ۰/۳۰- و ۰/۳۳- در سطح سه درصد نشان داد.

جدول ۱- پارامترهای آماری رس، شن و سیلت ۱۴۲ نمونه جمع‌آوری شده خاک.

Table 2. Statistical description of clay, sand and silt in 142 soil samples.

Soil properties	Max (%)	Med (%)	Min (%)	MEAN (%)	CV	ST.DEV	SKEW	KURT
Clay (%)	72.2	35.3	19	37.7	0.43	13.5	0.97	<b>0.18</b>
Sand (%)	66.5	34.4	17	36.5	0.36	13.2	0.46	<b>-0.79</b>
Silt (%)	40.2	26.8	10.8	25.7	0.27	7.01	-0.36	<b>-0.84</b>

جدول ۱- همبستگی پیرسون بین ویژگی های خاک.

Table 2. Pearson Correlation between soil properties.

	Clay (%)	Sand (%)	Silt (%)
Clay (%)	1		
Sand (%)	-0.76**	1	
Silt (%)	-0.30**	-0.33*	1

### ۳-۱- روابط رگرسیونی در مدل PLSR

و  $R^2_{CAL} = 0.77$ ،  $RMSE_{CAL} = 6/31$  تغییرات بیشتری در روند طیف صورت گرفت و دقت مدل افزایش یافت. روش صاف سازی ساویتری و گولای باعث افزایش دقت برآورد طیف شد، بنابراین می توان نتیجه گرفت که طیف خاک تحت تأثیر عوامل نویز محیطی قرار گرفته است و صاف سازی سبب تعدیل طیف و کاهش خطای برآورد نسبت به طیف اصلی شده است. برای درصد شن نتایج گروه واسنجی همانند درصد رس بوده و نوع روش های پیش پردازش در دقت برآورد مدل اثر داشته است. همچنین هر شش طیف بازتابی دقت برآورد قابل قبول تا عالی نشان داده اند. چهار روش پیش پردازش مشتق اول (FD-SG)  $(R^2_{CAL} = 0.79)$  و  $RPD_{CAL} = 6/07$ ،  $RMSE_{CAL} = 5/87$  و  $RPD_{CAL} = 1/97$  (SD-SG) مشتق دوم  $(R^2_{CAL} = 0.73)$ ،  $RMSE_{CAL} = 5/87$  و  $RPD_{CAL} = 1/84$  و واریانس استاندارد (SNV)  $(R^2_{CAL} = 0.80)$ ،  $RMSE_{CAL} = 5/93$  و  $RPD_{CAL} = 1/84$  و حذف پیوسته (CR)  $(R^2_{CAL} = 0.75)$ ،  $RMSE_{CAL} = 6/98$  و  $RPD_{CAL} = 1/68$  نسبت به طیف اصلی سبب افزایش دقت برآورد مدل PLS-RF شدند. بیشترین دقت برآورد در روش پیش پردازش مشتق دوم (SD-SG)  $(R^2_{CAL} = 0.79)$  و  $RPD_{CAL} = 6/07$  مشاهده شد و روش SG کمترین دقت  $(R^2_{CAL} = 0.66)$  و  $RPD_{CAL} = 7/68$  را نشان داد و می توان نتیجه گرفت که روش SG سبب تعدیل و کاهش نویز طیف اصلی نشده است.

جدول ۳ روابط رگرسیونی بین درصد رس خاک و فاکتور معنادار حاصل از مدل PLSR، در روش های طیف اصلی، طیف صاف شده با ساویتری گولای، پیش پردازش مشتق اول و مشتق دوم، واریانس نرمال استاندارد و روش حذف پیوستار ارائه شده است. فاکتورهای اولیه حاوی اطلاعات اصلی هستند. درحالی که فاکتورهای بعدی نویزها را هم شامل می شود (Camo, 1998). ارزیابی آماری خصوصیات خاک نشان داد روشی که بیشترین دقت دارد بیشترین فاکتور معنادار را نیز نشان داده است. روش حذف پیوستار و واریانس نرمال استاندارد بیشترین فاکتور معنادار را نشان دادند. نتایج مدل PLS-RF در برآورد درصد رس خاک در جدول ۴ ارائه شده است. نتایج نشان داد که نوع روش پیش پردازش در دقت برآورد مدل اثر داشت و هر شش روش در مرحله واسنجی دارای دقت برآورد قابل قبول تا عالی دارند. پنج روش پیش پردازش ساویتری و گولای، مشتق اول، مشتق دوم، واریانس استاندارد نرمال و حذف پیوسته سبب افزایش دقت برآورد طیف اصلی شدند. بیشترین دقت برآورد مربوط به روش پیش پردازش حذف پیوسته (CR)  $(R^2_{CAL} = 0.85)$ ،  $RMSE_{CAL} = 5/93$  و  $RPD_{CAL} = 1/98$  و کمترین دقت برآورد در روش طیف اصلی  $(R^2_{CAL} = 0.64)$  و  $RPD_{CAL} = 1/65$  مشاهده شد. مشتق گیری سبب افزایش دقت برآورد مدل گردید به نحوی که با افزایش ضریب مشتق از مشتق اول  $(R^2_{CAL} = 0.76)$  و  $RPD_{CAL} = 6/91$  و  $RMSE_{CAL} = 6/91$  به مشتق دوم  $(R^2_{CAL} = 0.74)$  و  $RPD_{CAL} = 1/83$

جدول ۳- رابطه رگرسیونی بین رس خاک با فاکتورهای معنادار در هر طیف.

Table 3. Regression relationship between soil clay content and significant factors in each spectrum.

SPECTRA	Regression equation
Row	$Y = 11.83 + 1.6X_1 - 7.42X_2 - 3.62X_3$
SG	$Y = 12.84 + 1.34X_1 + 7.23X_2 + 9.43X_3$
FD-SG	$Y = 8.074 - 1.23X_1 + 13.9X_2 - 3.12X_3$
SD-SG	$Y = 8.07 + 1.66X_1 + 11.05X_2 + 10.52X_3$
SNV	$Y = 14.83 + 1.4X_1 - 7.32X_2 - 4.92X_3 + 6.05X_4$
CR	$Y = 14.84 + 1.54X_1 + 7.13X_2 + 8.43X_3 + 10.29X_4 + 9.06X_5$



بوده‌اند. در شکل ۲ شیب خط رگرسیونی در نمودارهای مقایسه داده‌های اندازه‌گیری شده و برآورد شده درصد رس خاک نشان داد که روش پیش‌پردازش حذف پیوستار کمترین فاصله را از خط (۱:۱) دارد و بهترین برازش و توزیع داده‌ها در اطراف این خط دارد. روش طیف اصلی نسبت به روش‌های دیگر بیش‌ترین فاصله نسبی را از خط (۱:۱) نشان داد.

نتایج گروه اعتبارسنجی مدل PLS-RF در برآورد درصد رس خاک در جدول ۴ ارائه شده است. مشاهده شد که نوع روش پیش‌پردازش در دقت برآورد مدل اثر داشت و هر دو روش طیف اصلی، طیف نرم شده، واریانس استاندارد نرمال و حذف پیوستار دقت برآورد قابل قبول ( $1/65 < RPDval < 1/4$ ) نشان دادند. در برآورد درصد رس تمامی روش‌ها دقت قابل قبولی ( $1/4 < RPDval < 1/65$ ) را ارائه نمودند. در برآورد درصد سیلت دو روش مشتق دوم و حذف پیوستار دقت قابل قبول ( $1/65 < RPDval < 1/4$ ) داشتند و بقیه روش‌ها دقت برآورد ضعیف ( $RPDval < 1/4$ ) را ارائه کردند. در گروه اعتبارسنجی نیز روش حذف پیوستار در دو ویژگی درصد رس و سیلت بهترین دقت و در درصد رس بهترین دقت مربوط به مشتق دوم داشت و نتایج مشابه گروه واسنجی بود.

بر اساس نتایج ارائه‌شده در جدول ۴ مشخص شد که دقت برآورد درصد سیلت نسبت به درصد رس و شن کمتر به دست آمد. روش‌های طیف اصلی (ROW) ( $R^2_{CAL} = 0/53$ ،  $RMSE_{CAL} = 4/50$  و  $RPD_{CAL} = 1/50$ )، مشتق دوم (SD-SG) ( $R^2_{CAL} = 0/63$ ،  $RMSE_{CAL} = 4/53$  و  $RPD_{CAL} = 1/31$ ) و حذف پیوستار (CR) ( $R^2_{CAL} = 4/46$  و  $RPD_{CAL} = 1/65$ ) دقت قابل قبولی را در برآورد درصد سیلت خاک نشان دادند ولی روش‌های صاف سازی ساویتزی گولای (SG) ( $R^2_{CAL} = 0/69$ ،  $RMSE_{CAL} = 5/21$  و  $RPD_{CAL} = 1/3$ )، مشتق اول (FD-) ( $R^2_{CAL} = 0/56$ ،  $RMSE_{CAL} = 5/21$  و  $RPD_{CAL} = 1/32$ )، (SG) ( $R^2_{CAL} = 0/57$ ،  $RMSE_{CAL} = 5/15$  و  $RPD_{CAL} = 1/32$ ) و استاندارد (SNV) ( $R^2_{CAL} = 0/69$ ،  $RMSE_{CAL} = 4/46$  و  $RPD_{CAL} = 1/65$ ) در برآورد درصد سیلت خاک ناتوان بود. بیش‌ترین دقت برآورد در روش حذف پیوستار (CR) و کمترین دقت برآورد در روش صاف سازی ساویتزی گولای مشاهده شد. نتایج برآورد درصد سیلت خاک نیز نشان داد که دو روش پیش‌پردازش مشتق دوم، و حذف پیوستار سبب افزایش دقت برآورد مدل PLS-RF گردید درحالی‌که روش‌های دیگر دقت کمتری را نسبت به طیف اصلی داشتند؛ که می‌توان نتیجه گرفت که در حذف نویز طیف اصلی ناتوان

جدول ۴- نتایج معیارهای ارزیابی روش‌های PLS - RF در مدل‌سازی ویژگی‌های خاک مراحل واسنجی و اعتبارسنجی.

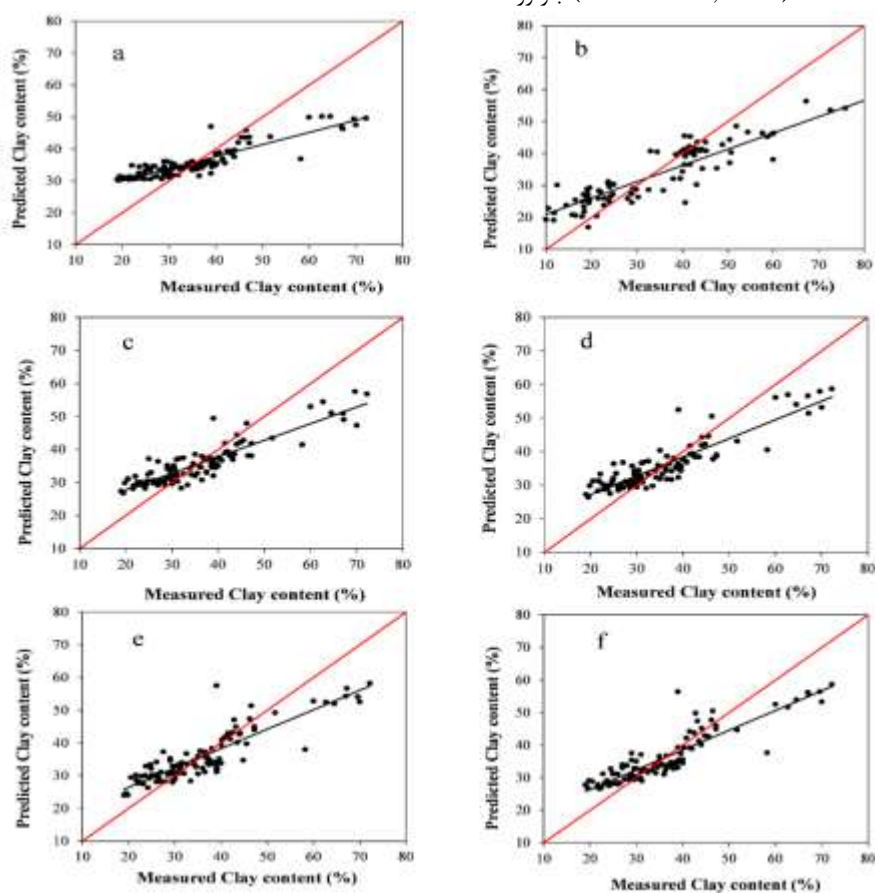
Table 4. Results of evaluation criteria for soil properties modeling by PLS-RF model in calibration and validation steps.

EVALUATION CRITERIA							
Soil Properties	Preprocessing Method	Calibration			VALIDATION		
		R <sup>2</sup>	RMSE	RPD	R <sup>2</sup>	RMSE	RPD
Clay (%)	Row	0.69	8.64	1.65	0.61	9.88	<b>1.40</b>
	SG	0.70	7.22	1.69	0.66	8.82	<b>1.55</b>
	FD-SG	0.74	6.91	1.76	0.56	1.04	<b>1.26</b>
	SD-SG	0.77	6.31	1.83	0.60	1.46	<b>1.30</b>
	SNV	0.80	6.16	1.91	0.62	9.61	<b>1.53</b>
	CR	0.85	5.93	1.98	0.69	7.76	<b>1.57</b>
Sand (%)	Row	0.65	7.94	1.60	0.60	1.69	<b>1.42</b>
	SG	0.67	7.68	1.66	0.58	1.83	<b>1.40</b>
	FD-SG	0.73	6.07	1.79	0.67	8.95	<b>1.59</b>
	SD-SG	0.80	5.87	1.97	0.71	7.35	<b>1.83</b>
	SNV	0.78	5.93	1.84	0.67	9.24	<b>1.55</b>
	CR	0.75	6.89	1.68	0.66	7.38	<b>1.48</b>
Silt (%)	Row	0.53	4.50	1.50	0.55	5.43	<b>1.39</b>
	SG	0.56	5.21	1.3	0.60	4.90	<b>1.25</b>
	FD-SG	0.57	5.15	1.32	0.52	5.63	<b>1.09</b>
	SD-SG	0.63	4.53	1.44	0.48	5.75	<b>1.59</b>
	SNV	0.46	5.40	1.31	0.46	6.02	<b>1.35</b>
	CR	0.69	4.46	1.65	0.51	5.70	<b>1.63</b>

کمتری را نشان داد. جز اصلی در شن خاک کانی کوارتز است و با وجود روشنایی زیاد، ویژگی جذب مشخصی را محدوده مادون قرمز و مرئی ندارد. بنابراین ممکن است درصد شن دقت نسبتاً پایین تری نسبت به درصد رس داشته باشد (Mohamed et al., 2018). بنابراین با توجه به نوع کانی غالب دقت برآورد ذرات رس، شن و سیلت متفاوت است. در خاک‌های Oxisols با توجه به وجود کانی‌های رسی، دقت برآورد رس در محدوده (۴۰۰-۹۸۰ نانومتر) بیشتر از شن و سیلت است. کانی کلسیت، جذب در طول موج بین باندهای ۱۸۰۰-۲۴۰۰ دارد. امواج ضعیف در ۱۹۰۰ نانومتر در کائولینیت و جذب قوی در این طول موج نشان‌دهنده مونت موریلونایت است. کانی‌های دیگر مانند فلدسپارها و انواع گرانیات ویژگی جذب قابل‌توجهی در محدوده مرئی و مادون قرمز ندارند. بنابراین با توجه به نوع کانی غالب در خاک دقت برآورد متفاوت برای ذرات رس، شن و سیلت می‌تواند متفاوت باشد.

در اغلب مطالعاتی که در زمینه مقایسه روش‌های پیش‌پردازش و مدل‌سازی خصوصیات خاک انجام شده است روش پیش‌پردازش CR که ویژگی‌های جذبی را در طیف خاک نشان می‌دهد دقت قابل قبولی را نشان داده است (Vasques et al., 2008; Nawar et al., 2014). در پژوهش حاضر نیز روش CR دقت برآورد قابل قبولی برای درصد رس و سیلت داشت. با این وجود برای درصد شن روش مشتق دوم دقت بالاتری را نشان داد. نتایج این پژوهش با دقت قابل قبولی در زمینه برآورد بافت خاک با روش طیف سنجی با نتایج مطالعات کورسیو همکاران (Curcio et al., 2013)، سیلوا و همکاران (Silva et al., 2016) و بیلگی و همکاران (Bilgili et al., 2010) مطابقت دارد.

نتایج برآورد درصد سیلت نسبت به درصد شن و رس دقت کمتری نشان داد، از آنجا که درصد سیلت خاک در اغلب پژوهش‌ها از اختلاف مجموع درصد رس و شن از ۱۰۰ به دست می‌آید و ویژگی جذبی رایجی را در پژوهش‌های پیشین نشان نداده است (Curcio et al., 2013). برآورد



شکل ۲- مدل‌سازی درصد رس خاک با استفاده از داده‌های واسنجی (calibration) برازش یافته بر روی مدل PLS - RF. (a) طیف اصلی (ROW)، (b) طیف صاف شده با ساویتزی گولای، (c) مشتق اول طیف نرم شده با ساویتزی گولای (FD-SG)، (d) مشتق دوم طیف نرم شده با ساویتزی گولای (SD-SG)، (e) روش واریانس نرمال استاندارد (SNV) و (f) حذف پیوستار (CR).

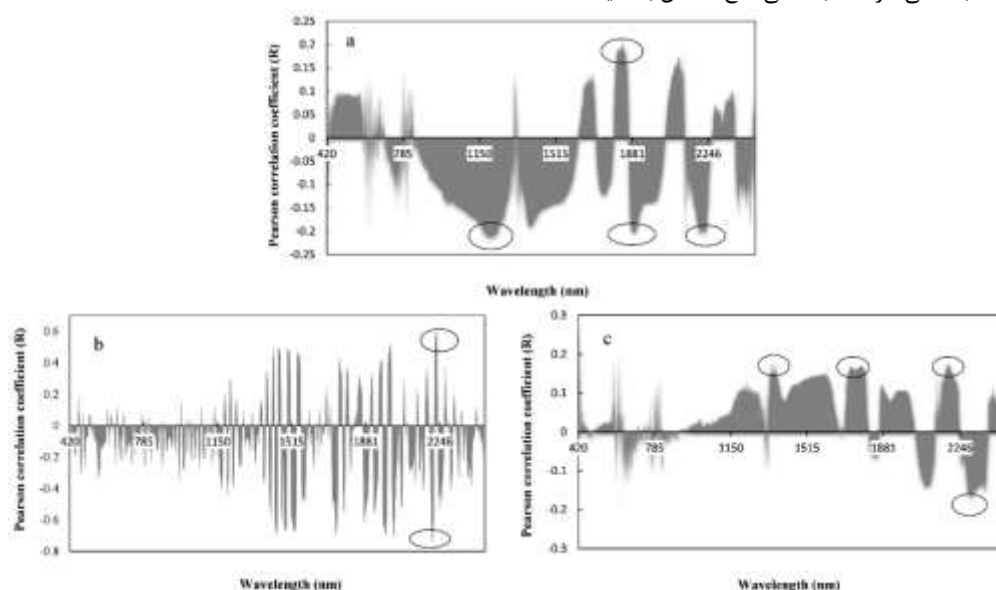
Fig. 2. Modeling of soil clay content fitted on PLS - RF model, using calibration data. a: main spectra (ROW), (b) Savitzky-Golay filter (SG), (c) the first derivative with the Savitzky-Golay filter (FD-SG), (d) the second derivative with the Savitzky-Golay filter (SD-SG), (e) the standard normal variant (SNV), and (f) the continuum removal method (CR).

### ۳-۲- ویژگی‌های طیفی خاک در روش‌های پیش‌پردازش

در شکل ۳ همبستگی بین طیف خاک و روش‌های پیش‌پردازش طیفی در حالتی که بیش‌ترین دقت برآورد ویژگی خاک بر اساس مدل PLS-RF مشاهده گردید ارائه شده است. برای درصد رس و سیلت روش حذف پیوسته و برای درصد شن، روش مشتق اول بهترین دقت برآورد را نشان داد. برای درصد رس خاک همبستگی قوی منفی در محدوده طول موج ۱۲۰۰-۱۲۱۰، ۱۹۰۰ و ۲۲۰۰ نانومتر و همبستگی مثبت در ۱۸۰۰ نانومتر مشاهده گردید. برای درصد شن همبستگی مثبت و منفی در محدوده ۱۴۰۰-۱۴۵۰ و ۱۹۱۰-۱۹۳۰ نانومتر که همبستگی قوی در ۲۲۰۰ و ۲۲۲۰ نانومتر مشاهده شد. برای درصد سیلت خاک همبستگی مثبت در طول موج ۱۳۲۰، ۱۶۱۵، ۲۲۰۰ نانومتر و همبستگی منفی در محدوده ۲۲۵۰ نانومتر نتیجه‌گیری شد. شکل ۴ نتایج طیف اصلی و پنج پیش‌پردازش استفاده‌شده در این پژوهش را نشان می‌دهد. ویژگی‌های ماده آلی و گچ خاک با تأثیر در بازتاب خاک، بر روی بروز سایر خصوصیات خاک اثر دارند. در مطالعه ناوار و همکاران (Nawar et al., 2016) بیش‌ترین همبستگی ماده آلی در طیف اصلی و طول موج ۶۰۰ نانومتر و برای طیف‌های پیش‌پردازش شده با روش‌های حذف پیوستار، مشتق اول و مشتق دوم طول موج در طول موج ۱۹۰۰ نانومتر مشاهده شد. همچنین این پژوهشگران برای درصد رس خاک نیز همبستگی قوی با طیف اصلی در محدوده ۶۰۰ نانومتر مشاهده کردند. نوکیتا و همکاران (Nocita et al., 2014) در مطالعه خود نشان دادند که طول موج‌هایی در محدوده ۶۸۰-۵۵۰ نانومتر طیف اصلی به محتوای مواد آلی خاک حساس است. این پژوهشگران گزارش کردند که طول موج ۶۰۰ نانومتر با کربن آلی خاک همبستگی دارد. همبستگی گچ خالص با طیف

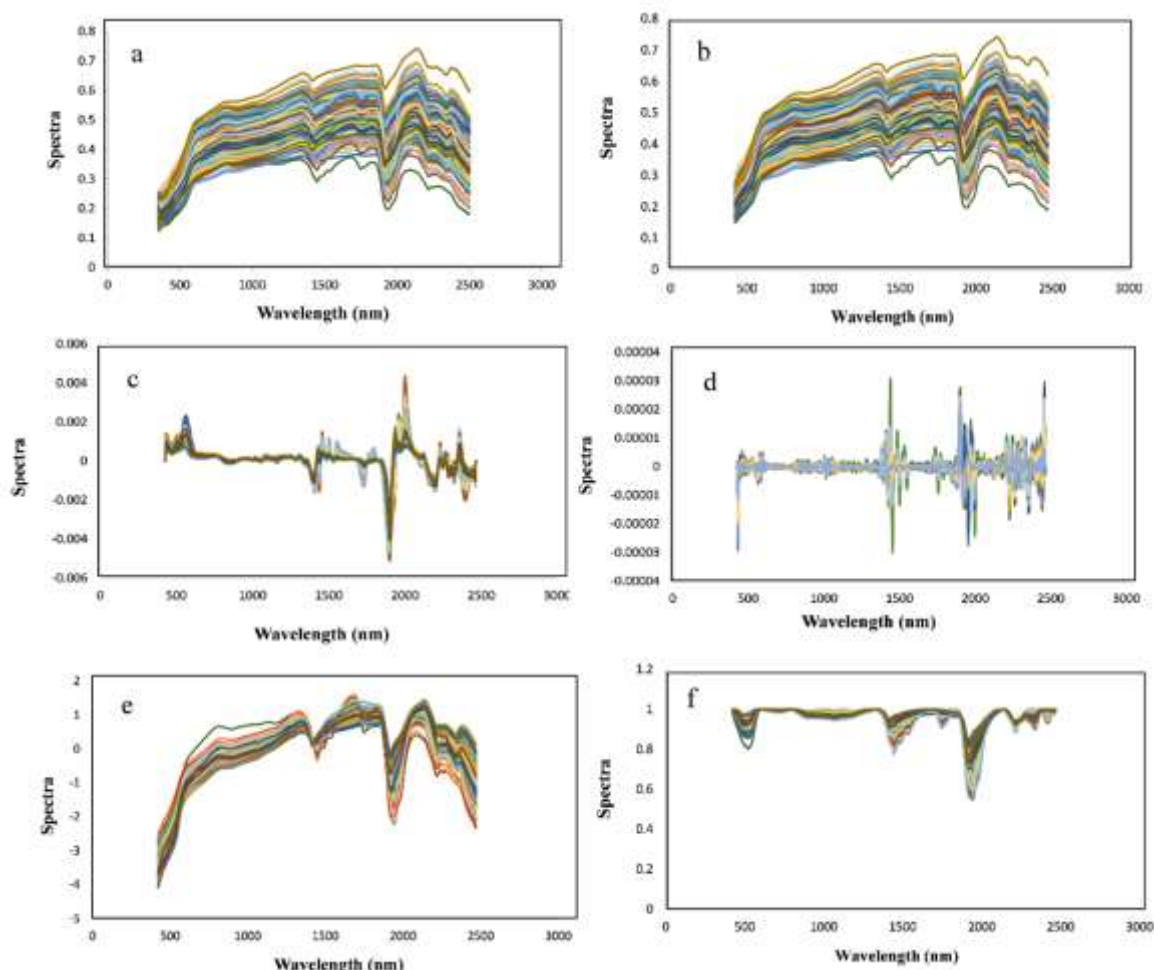
خاک در طول موج‌های ۱۲۰۰، ۱۴۰۰-۱۶۰۰، ۱۷۵۰، ۱۹۴۵، ۲۱۰۰-۲۲۰۰ و ۲۴۰۰ توسط هریسون گزارش شده است (Harrison, 2012). ویژگی‌های جذبی طیف خاک در طول موج ۱۶۰۰، ۲۳۵۰، ۲۵۰۰، ۲۰۰۰ و ۱۹۰۰ می‌تواند ناشی از انواع کربنات‌ها باشد. طول موج ۵۴۰، ۵۷۰، ۱۴۳۰، ۱۴۶۰، ۱۷۴۰، ۱۸۷۰، ۱۹۴۰، ۲۰۱۰، ۲۳۵۰ و ۲۴۱۰ با شوری خاک ناشی از انواع کاتیون‌ها و آنیون‌ها همبستگی دارد همچنین در اغلب پژوهش‌ها مشخص شده است طول موج‌های ۵۴۰، ۱۷۴۰، ۲۰۱۰ و ۲۳۵۰ نانومتر به درجه شوری خاک حساسیت بیشتری دارند (Xu et al., 2016).

خصوصیات جذبی در اثر برهمکنش بین ویژگی خاک با طیف بازتابی است و مولفه‌های مولکولی مانند گروه‌های عاملی در بروز این ویژگی‌ها اثر می‌گذارند نوع اثرگذاری به ترکیبی که گروه عاملی در آن وجود دارد بستگی دارد. در پژوهش‌های قبلی به انواعی از گروه‌های عاملی و میزان اثر آنها در بازتاب خاک با توجه به هر ویژگی خاک، مانند درصد رس، ماده آلی و غیره اشاره شده است. مشاهده شده که طول موج‌های ۱۴۰۰، ۱۹۰۰ و ۲۲۰۰ نانومتر به ویژگی جذبی در اثر ارتعاش گروه‌های عاملی OH آب آزاد (Shi et al., 2014)، در طول موج ۱۹۰۰ و ۲۲۰۰ نانومتر برای گروه‌های عاملی Al-OH و Mg-OH (Wang et al., 2018; Pu et al., 2003)، و در طول موج‌های ۴۶۰، ۵۵۰ و ۷۴۰ در اثر ارتعاش ناشی از گروه‌های عاملی N-H و C-H + C-H، C-H + C-C است (Xuemei and Jianshe, 2013). خصوصیات جذبی در طول موج ۲۲۸۰ نانومتر در اثر کانی مونت موریلونایت است که آهن تبادل‌ی در گروه عاملی Fe-OH دارد و در طول موج ۲۳۴۰ نانومتر در اثر جذب کانی‌های ایلایت و مسکویت است (Postand Noble, 1993).



شکل ۳- همبستگی بین طیف با ویژگی‌های خاک با طیف‌های مختلف، (a) درصد رس خاک با روش حذف پیوستار (CR)، (b) درصد شن با روش مشتق دوم طیف نرم شده (SD-SG) و (c) درصد سیلت با حذف پیوسته (CR)، در محدوده طول موج کلیدی (دایره‌های توخالی سیاه رنگ).

Fig. 3. Diagram of correlation soil properties with different spectra including, (a) soil clay percentage with continuum removal method (CR), (b) sand percentage with the second derivative with the Savitzky-Golay filter (SD-SG), and (c) percentage silt with continuum removal (CR), in the key wavelength range (Hollow black circles).



شکل ۴- طیف اصلی خاک و انواع روش‌های پیش‌پردازش شامل (a) طیف اصلی (ROW)، (b) طیف صاف شده با سائوتزی گولای، (c) مشتق اول طیف نرم شده با سائوتزی گولای (FD-SG)، (d) مشتق دوم طیف نرم شده با سائوتزی گولای (SD-SG)، (e) روش واریانس نرمال استاندارد (SNV) و (f) حذف پیوستار (CR).

Fig. 4. (a) View of the original spectrum and variety of preprocessing methods include (b) Savitzky-Golay filter (SG), (c) the first derivative with the Savitzky-Golay filter (FD-SG), (d) the second derivative with the Savitzky-Golay filter (SD-SG), (e) the standard normal variant (SNV), and (f) the continuum removal method (CR).

خاک با طول موج‌های ۴۶۵، ۵۴۷، ۶۲۷، و ۷۵۲ همبستگی مثبت و با طول موج ۴۹۲، ۵۱۲، ۵۸۷، ۶۶۲ و ۸۶۷ همبستگی منفی دارد. کورسیا و همکاران (Curcio et al., 2013) نتیجه گرفت که طول موج‌های ۱۲۰۰، ۱۴۰۰ و ۱۹۰۰ دارای همبستگی بالا بافت خاک است. راسل و همکاران (Rossel et al., 2006) با روش CR طول موج‌های کلیدی رس خاک را در ۲۲۰۶ و ۲۲۰۷ نانومتر معرفی کردند. چابریلت و همکاران (Chabrilat et al., 2002) طی پژوهشی در شناسایی کانی‌های رسی ایلات، کائولینایت و اسمکتایت در خاک از طول موج ۲۲۰۰ به عنوان طول موج کلیدی استفاده کردند.

گرد و غبار استان خوزستان مورد مقایسه قرار گرفت. نتایج عبارتند از

(۱) مقایسه روش‌های پیش‌پردازش در برآورد ویژگی‌های خاک نشان داد که روش حذف پیوستار در دو ویژگی درصد رس ( $RPD_{CAL} = 1/98$  و  $R^2_{CAL} = 0.93$ ) و درصد سیلت ( $RPD_{CAL} = 1/97$  و  $R^2_{CAL} = 0.85$ )، بهترین عملکرد داشت، در صورتی که برای

استفاده از پنج روش پیش‌پردازش با حذف نویز در برآورد ویژگی خاک موثر بوده است. پیش‌پردازش CR خصوصیات جذبی در ناحیه مادون قرمز را به خوبی نشان می‌دهد و طول موج در ناحیه ۱۴۰۰ و ۱۹۰۰ می‌تواند نشان‌دهنده آب کریستاله و هیدراته در انواع کانی‌ها و طول موج ۱۴۰۰ می‌تواند از ارتعاش گروه هیدروکسیل در آب پیوندی در سیلیکات‌ها باشد. از این ویژگی جذبی در تشخیص غیرمستقیم شن استفاده می‌شود. ناحیه ۲۲۰۰ نانومتر ناشی از گروه عاملی OH-Al و می‌تواند ناشی از حضور رس در خاک باشد (Rossel; Curcio et al., 2013) (Silva et al., 2016) نشان داد که رس

#### ۴- نتیجه‌گیری

در این پژوهش عملکرد مدل PLS-RF در طیف اصلی و پنج روش پیش‌پردازش: مشتق اول و دوم، واریانس استاندارد نرمال و حذف پیوسته به منظور برآورد ویژگی‌های درصد سیلت، شن و رس در خاک‌های مستعد تولید

گردید.  
۴) با توجه به تغییرپذیری مکانی ویژگی‌های خاک و از طرفی هزینه زیاد و زمان بر بودن اندازه‌گیری در سطح وسیع کانون‌های ریز گرد استان خوزستان، روش طیف‌سنجی ابزاری مفید برای جایگزینی یا تکمیل کننده روش‌های سنتی تجزیه و تحلیل آزمایشگاهی خاک و کاهش تعداد نمونه‌ها، جهت ارزیابی وضعیت اراضی خواهد بود.  
۵) استفاده از روش طیف‌سنجی و تعیین طول موج کلیدی، در ترکیب با دور سنجی سبب افزایش دقت نقشه بافت خاک مناطق مستعد تولید ریز گرد استان خوزستان می‌شود.

درصد شن، روش مشتق دوم بهترین عملکرد (RPD<sub>CAL</sub> = ۱/۶۵ و ۴/۴۶ = R<sup>2</sup> CAL = ۰/۶۹۰، RMSE<sub>CAL</sub>) را نشان داد  
۲) در این پژوهش طول موج کلیدی برای درصد رس در محدوده طول موج‌های ۱۲۰۰-۱۲۱۰، ۱۸۰۰ و ۲۲۰۰ نانومتر، برای درصد شن در محدوده طول موج‌های ۱۴۰۰-۱۴۵۰، ۱۹۱۰-۱۹۳۰، ۲۲۰۰ و ۲۲۲۰ نانومتر و برای درصد سیلت خاک در محدوده طول موج‌های ۱۳۲۰، ۱۶۱۵، ۲۲۰۰ و نانومتر مشاهده گردید.  
۳) استفاده از روش رگرسیون PLS و تعیین مؤلفه‌های معنادار در هر دسته طیفی سبب کاهش حجم محاسبات، افزایش سرعت عمل پردازش در کامپیوتر محاسباتی و در نهایت دست آوردن عملکرد مطلوب در برآورد ویژگی‌های خاک

## منابع

- Ben-Dor, E., Chabrilat, S., Demattê, J., Taylor, G., Hill, J., Whiting, M., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment* 113, 38-55.
- Bilgili, A.V., Van Es, H., Akbas, F., Durak, A., Hively, W., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments* 74(2), 229-238.
- Breiman, L., 1999. Using adaptive bagging to debias regressions. Technical Report 547, Statistics Dept. UCB.
- Breiman, L., 2001. Random forests. *Machine learning* 45(1), 5-32.
- CAMO, A., 1998. The Unscrambler User Manual. CAMO ASA Norway.
- Chabrilat, S., Goetz, A.F., Krosley, L., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sensing of Environment* 82(2-3), 431-445.
- Chang, C-W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65(2), 480-490.
- Chen, X-W., Liu, M., 2005. Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics* 21(24), 4394-4400.
- Curcio, D., Ciralo, G., D'Asaro, F., Minacapilli, M., 2013. Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences* 19, 494-503.
- Daempanah, R., Haghnia, Gh., Alizadeh, A., Karimi, A., 2011. Mapping Salinity and Sodicity of Surface Soil by Remote Sensing and Geostatistic Methods in South Side of Mah Valat County. *Journal of Water and Soil* 25, 498-508.
- Dotto, A.C., Dalmolin, R.S.D., Pedron, F.d.A., Caten, A.T., Ruiz, L.F. C., 2014. Digital mapping of soil properties: particle size and soil organic matter by diffuse reflectance spectroscopy. *Revista Brasileira de Ciência do Solo* 38(6), 1663-1671.
- Farifteh, J., Van der Meer, F., Atzberger, C., Carranza, E. 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.
- Fearn, T., Riccioli, C., Garrido-Varo, A., Guerrero-Ginel, J.E., 2009. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems* 96(1), 22-26.
- Genuer, R., Poggi, J-M., Tuleau, C., 2008. Random Forests: some methodological insights. arXiv preprint arXiv:0811.3619.
- 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. *Journal of Watershed Engineering and Management* 6, 125-138.
- He, T., Wang, J., Lin, Z., Cheng, Y. 2009. Spectral features of soil organic matter. *Geo-spatial Information Science* 12(1), 33-40.
- Heidarian, P., Joudaki, M., Darvishi Khatoni, j., Shahbazi, R., 2015. Recognized Dust Sources in Khuzestan Province. *Geological Survey and Mineral Exploration of Iran*.
- Hillel, D., 1980. Applications of soil physics. Academic Press, San Diego.
- Ji, W., Adamchuk, V.I., Biswas, A., Dhawale, N.M., Sudarsan, B., Zhang, Y., Rossel, R.A.V., Shi, Z. 2016. Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering* 152, 14-27.
- Khayamim, F., Khademi, H., Stenberg, B., Wetterlind, J., 2015. Capability of vis-NIR Spectroscopy to Predict Selected Chemical Soil Properties in Isfahan Province. *JWSS-Isfahan University of Technology* 19(72), 81-92.
- Mohamed, E., Saleh, A., Belal, A., Gad, A.A., 2018. Application of near-infrared reflectance for quantitative assessment of soil properties. *The Egyptian Journal of Remote Sensing and Space Science*, 21(1), 1-14.
- Mutanga, O., Skidmore, A.K., 2003. Continuum-removed absorption features estimate tropical savanna grass quality in situ. Paper presented at the Earsel workshop on imaging spectroscopy.



- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., 2014. Modeling and mapping of soil salinity with reflectance spectroscopy and landsat data using two quantitative methods (PLSR and MARS). *Remote Sensing* 6(11), 10813-10834.
- Nocita, M., Stevens, A., Toth, G., Panagos, P., Van Wesemael, B., Montanarella, L., 2014. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology and Biochemistry* 68, 337-347.
- 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.
- Pu, R., Ge, S., Kelly, N., Gong, P., 2003. Spectral absorption features as indicators of water status in coast live oak (*Quercus agrifolia*) leaves. *International Journal of Remote Sensing* 24(9), 1799-1810.
- Ranjbar, F., Jalali, M., 2016. The combination of geostatistics and geochemical simulation for the site-specific management of soil salinity and sodicity. *Computers and Electronics in Agriculture* 121, 301-312.
- Rossel, R.V., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* 150(3-4), 253-266.
- Rossel, R.V., Walvoort, D., McBratney, A., Janik, L.J., Skjemstad, J., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1-2), 59-75.
- Shi, Z., Wang, Q., Peng, J., Ji, W., Liu, H., Li, X., Rossel, R.A.V., 2014. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. *Science China Earth Sciences* 57(7), 1671-1680.
- Silva, E.B., Ten Caten, A., Dalmolin, R.S.D., Dotto, A.C., Silva, W.C., Giasson, E., 2016. Estimating Soil Texture from a Limited Region of the Visible/Near-Infrared Spectrum Digital Soil Morphometrics, Springer, pp. 73-87.
- Sjöström, M., Wold, S., Lindberg, W., Persson, J.-Å., Martens, H., 1983. A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Analytica Chimica Acta* 150, 61-70.
- Vasques, G., Grunwald, S., Sickman, J., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146(1-2), 14-25.
- Wang, J., Ding, J., Abulimiti, A., Cai, L., 2018. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS-NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. *PeerJ* 6, e4703.
- Wang, J., Tiyip, T., Ding, J., Zhang, D., Liu, W., Wang, F., 2017. Quantitative estimation of organic matter content in arid soil using vis-NIR spectroscopy preprocessed by fractional derivative. *Journal of Spectroscopy*.
- Wenjun, J., Zhou, S., Jingyi, H., Shuo, L., 2014. In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *Plos one* 9(8), 105708.
- Xu, C., Zeng, W., Huang, J., Wu, J., van Leeuwen, W., 2016. Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sensing* 8(1), 42.
- Xuemei, L., Jianshe, L., 2013. Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration. *Measurement* 46(10), 3808-3814.