



## Evaluation of PLSR and bagging-PLSR methods in estimating soil texture, calcium carbonate, and pH using spectral data

Aram Shahabi<sup>1</sup> | Masoud Davari<sup>2</sup> | Kamal Nabiollahi<sup>3</sup> | Rohullah Taghizadeh Mehrjardi<sup>4</sup>

1. Department of Soil Science and Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran. E-mail: [aramshahabi1373@gmail.com](mailto:aramshahabi1373@gmail.com)
2. Department of Soil Science and Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran. E-mail: [m.davari@uok.ac.ir](mailto:m.davari@uok.ac.ir)
3. Corresponding Author, Department of Soil Science and Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran. E-mail: [k.nabiollahi@uok.ac.ir](mailto:k.nabiollahi@uok.ac.ir)
4. Department of Arid and Desert Regions Management, Faculty of Agriculture and Natural Resources, Ardakan University, Ardakan, Iran. E-mail: [rh\\_taghizade@yahoo.com](mailto:rh_taghizade@yahoo.com)

### Article Info

**Article type:** Research Article

**Article history:**

**Received:** July. 17, 2023

**Revised:** Aug. 25, 2023

**Accepted:** Sep. 2, 2023

**Published online:** Oct. 23, 2023

**Keywords:**

CCE,  
Soil Properties,  
Partial Least-Squares Regression,  
Visible and Near-Infrared  
Spectroscopy.

### ABSTRACT

The rapid, accurate, and low-cost determination of soil properties has particularly important for land planning and management. The objective of this study was to evaluate the Vis-NIR spectral reflectance of soils, as a rapid, cost-effective, and non-destructive technique, for estimating some soil properties [sand, silt, clay, pH, and calcium carbonate equivalent (CCE)] by partial least-square regression (PLSR) and bagging-PLSR methods. For this purpose, a total of 220 composite soil samples were collected from 0-20 cm depth in Ghorveh Plain, Kurdistan province, in September 2019. The selected soil properties were measured by standard laboratory methods. The proximal spectral reflectance of soil samples was also measured within the 350-2500 nm range (Vis-NIR) using a handheld Spectroradiometer. Different pre-processing methods were assessed after recording the spectra. The results indicated that the R<sup>2</sup> values for the PLSR method ranged from 0.58 to 0.76, while the bagging-PLSR produced R<sup>2</sup> values between 0.59 and 0.74. The RMSE values obtained for sand, silt, clay, CCE, and pH were 17.43, 7.65, 7.83, 7.94, and 0.66, respectively for the PLSR, and 16.66, 7.63, 8.13, 7.71, and 0.45 for the bagging-PLSR. Based on the ratio of prediction to deviation (RPD) values, the bagging-PLSR model achieved the best performance in predicting sand and CCE. However, for clay and pH prediction, the PLSR model was the most accurate. Both the PLSR and bagging-PLSR models yielded identical predictions for silt content, with an RPD value of 1.53. Overall, the results showed that PLSR and bagging-PLSR models have acceptable accuracy for estimating the proposed properties of the soils.

Cite this article: Shahabi, A., Davari, M., Nabiollahi, K., & Taghizadeh Mehrjardi, R. (2023). Evaluation of PLSR and bagging-PLSR methods in estimating soil texture fractions, pH, and calcium carbonate using spectral data, *Iranian Journal of Soil and Water Research*, 54 (8), 1215-1231. <https://doi.org/10.22059/ijswr.2023.362342.669533>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2023.362342.669533>



## ارزیابی روش‌های PLSR و bagging-PLSR در برآورد اجزای بافت خاک، کربنات کلسیم و pH خاک با استفاده از داده‌های طیفی

آرام شهبابی<sup>۱</sup> | مسعود داوری<sup>۲</sup> | کمال نبی‌اللهی<sup>۳</sup> | روح‌الله تقی‌زاده مهرجردی<sup>۴</sup>

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه کردستان، سنندج، ایران. رایانامه: [aramshahabi1373@gmail.com](mailto:aramshahabi1373@gmail.com)

۲. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه کردستان، سنندج، ایران. رایانامه: [m.davari@uok.ac.ir](mailto:m.davari@uok.ac.ir)

۳. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه کردستان، سنندج، ایران. رایانامه: [k.nabiollahi@uok.ac.ir](mailto:k.nabiollahi@uok.ac.ir)

۴. گروه مدیریت مناطق خشک و بیابانی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، ایران. رایانامه: [rh\\_taghizade@yahoo.com](mailto:rh_taghizade@yahoo.com)

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	تعیین سریع، دقیق و کم‌هزینه ویژگی‌های خاکی در برنامه‌ریزی و مدیریت اراضی از اهمیت قابل توجهی برخوردار است. هدف از این پژوهش ارزیابی بازتاب طیفی نزدیک، به‌عنوان تکنیکی سریع، مقرون به صرفه و غیرمخرب در تخمین برخی ویژگی‌های خاک (شن، سیلت، رس، pH و کربنات کلسیم معادل (CCE)) با استفاده از روش‌های رگرسیون حداقل مربعات جزئی (PLSR) و رگرسیون حداقل مربعات جزئی توأم با باز نمونه‌گیری (bagging-PLSR) بود. بدین منظور، ۲۲۰ نمونه مرکب خاک از عمق ۰ تا ۲۰ سانتی‌متری در شهر یور ۱۳۹۸ از دشت قروه استان کردستان جمع‌آوری شد. سپس ویژگی‌های انتخابی خاک با روش‌های آزمایشگاهی استاندارد اندازه‌گیری شد. بازتاب طیفی نزدیک نمونه‌ی خاک‌ها در محدوده ۳۵۰ تا ۲۵۰۰ نانومتر (Vis-NIR) با بهره‌گیری از دستگاه اسپکترورادایومتر آزمایشگاهی اندازه‌گیری شد. پس از ثبت طیف‌ها، انواع مختلف روش‌های پیش‌پردازش مورد ارزیابی قرار گرفت. نتایج نشان داد که روش PLSR مقادیر $R^2$ بین ۰/۵۸ تا ۰/۷۶ به‌دست می‌دهد، درحالی‌که این مقادیر در bagging-PLSR بین ۰/۵۹ تا ۰/۷۴ متغیر می‌باشد. مقادیر RMSE معادل با ۱۷/۴۳، ۷/۶۵، ۷/۸۳، ۷/۹۴ و ۰/۶۶ در روش PLSR و همچنین مقادیر ۱۶/۶۶، ۷/۶۳، ۸/۱۳، ۷/۷۱ و ۰/۴۵ در روش bagging-PLSR به‌ترتیب برای شن، سیلت، رس، CCE و pH خاک به‌دست آمد. بر اساس مقادیر RPD (نسبت برآورد به انحراف)، در برآورد مقدار شن و CCE بهترین عملکرد توسط مدل bagging-PLSR به‌دست آمد؛ این درحالی‌است که در برآورد رس و pH خاک مدل PLSR دقیق‌ترین بود. نتایج پیش‌بینی هر دو مدل برای سیلت یکسان بود ( $RPD = 1/53$ ). در کل نتایج نشان داد که مدل‌های PLSR و bagging-PLSR در برآورد ویژگی‌های خاکی مورد مطالعه از دقت قابل قبولی برخوردار می‌باشند.
واژه‌های کلیدی: کربنات کلسیم معادل، طیف‌سنجی مرئی-مادون قرمز نزدیک، رگرسیون حداقل مربعات جزئی، ویژگی‌های خاکی.	
تاریخ دریافت: ۱۴۰۲/۴/۲۶	
تاریخ بازنگری: ۱۴۰۲/۶/۳	
تاریخ پذیرش: ۱۴۰۲/۶/۱۱	
تاریخ انتشار: ۱۴۰۲/۸/۱	

استناد: شهبابی، آرام، داوری، مسعود، نبی‌اللهی، کمال، و تقی‌زاده مهرجردی، روح‌الله. (۱۴۰۲). ارزیابی روش‌های PLSR و bagging-PLSR در برآورد اجزای بافت خاک، pH و کربنات کلسیم خاک با استفاده از داده‌های طیفی، مجله تحقیقات آب و خاک ایران، ۵۴ (۸)، ۱۲۳۱-۱۲۱۵.



<https://doi.org/10.22059/ijswr.2023.362342.669533>

© نویسنده‌گان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2023.362342.669533>

## مقدمه

خاک یکی از بنیادی‌ترین منابع طبیعی است که حیات انسان به آن وابسته است. خاک در زیست‌بوم کارکردهای چندگانه‌ای دارد که از جمله آن‌ها می‌توان به (۱) تنظیم‌کننده جریان، زهکشی و ذخیره آب، (۲) فیلتری برای مواد مغذی، فلزات سنگین و سایر آلاینده‌ها، (۳) زیستگاه بیولوژیک و ذخیره ژنی، (۴) مخزنی بالقوه برای ذخیره کربن (کمک به تقلیل گرمایش جهانی) و (۵) شالوده و مصالح سازنده ساختمان‌ها و سازه‌های بشری اشاره کرد (Stenberg *et al.*, 2010). توانایی خاک برای انجام این عملکردها به ترکیب، ساختار و وضعیت و ماهیت ویژگی‌های خاک بستگی دارد (Karlen *et al.*, 1997). از نظر فیزیکی، خاک ماتریس پیچیده‌ای از ذرات آلی و معدنی، آب، هوا و موجودات زنده است که با هم یک سیستم سه فازی از جامدات، مایعات و گازها را تشکیل می‌دهند (Bhattacharyya and Pal, 2015). فرآیندها و مکانیسم‌ها در خاک، به‌عنوان یک محیط ناهمگن، متخلخل و سه فازی، بسیار پیچیده هستند و از این رو درک کامل آن‌ها دشوار است (Mohamed *et al.*, 2018). رویکردهای تجزیه و تحلیل سنتی بسیاری در آزمایشگاه و مزرعه برای کمی کردن رابطه بین این ویژگی‌های خاک (فیزیکی، شیمیایی و بیولوژیک) و اجزای اصلی آن توسعه یافته‌اند. از نظر تاریخی، دانش بشر از این سیستم پیچیده خاک و ارزیابی ویژگی‌ها و عملکردهای آن از طریق این نوع تجزیه و تحلیل‌های آزمایشگاهی و مزرعه‌ای به‌دست آمده است (Viscarra Rossel and McBratney, 1998).

از آنجایی که ویژگی‌های خاک در مزرعه و/یا مقیاس‌های بزرگتر هم از نظر مکانی و هم زمانی متغیر است؛ لذا تعداد نمونه‌های جمع‌آوری و تجزیه و تحلیل شده بایستی به گونه‌ای باشد که بتواند این تنوع مکانی را نشان داده و ویژگی‌های خاک را به درستی مشخص کند. واقعیت این است که نمونه‌برداری مرسوم خاک و اندازه‌گیری قابل اعتماد ویژگی‌های خاک، به‌ویژه در مقیاس‌های جغرافیایی بزرگ، ممکن است کاری پرزحمت، زمان‌بر و پرهزینه بوده و نیاز به استفاده از مواد شیمیایی مضر داشته باشد. این در حالی است که دسترسی به مقادیر زیادی از داده‌های دقیق خاک برای مدیریت بهتر آن ضروری است. از این‌رو، نیاز به روش‌هایی جدید که در آن محدودیت‌های معمول تجزیه و تحلیل خاک‌ها وجود نداشته باشد و توانایی ارائه اطلاعاتی دقیق و با کیفیت از خاک را به‌صورت ساده و در سریع‌ترین زمان ممکن و با حداقل هزینه داشته باشد به یکی از اولویت‌های مهم پژوهش‌های محققین علوم خاک تبدیل شده است (Zhu *et al.*, 2001). در این راستا، تکنیک‌های سنجش از راه دور و نزدیک<sup>۱</sup> به عنوان گزینه‌های ممکن برای افزایش، تکمیل و یا جایگزینی روش‌های سنتی اندازه‌گیری خاک در نظر گرفته می‌شوند (Viscarra *et al.*, 2006). در چند دهه گذشته کارایی طیف‌سنجی بازتابی مرئی<sup>۲</sup> (Vis) (۳۵۰-۷۰۰ نانومتر) - مادون قرمز نزدیک<sup>۳</sup> (NIR) (۷۰۰-۲۵۰۰ نانومتر)، به‌عنوان یک روش سنجش از نزدیک، در برآورد برخی از ویژگی‌های خاک توجهات فراوانی را به خود جلب کرده است (Bellon-Maurel and McBratney, 2011). استفاده روزافزون از داده‌های طیفی در علوم خاک و محیط‌زیست را می‌توان از سویی به وجود اطلاعات فراوان نهفته در این طیف‌ها و از سویی دیگر به توسعه و ایجاد روش‌های مختلف داده‌کاوی برای مدل کردن آن‌ها نسبت داد. سادگی، سریع بودن، ارزان بودن، حداقل بودن آماده‌سازی نمونه‌ها و غیر مخرب بودن آن‌ها برای زیست‌بوم از مهمترین مزیت‌های طیف‌سنجی مرئی - مادون قرمز نزدیک است (Ramírez *et al.*, 2023). (Khayamim *et al.*, 2015). طبق تعریف، طیف‌سنجی به تهیه منحنی‌های بازتابی پدیده یا پدیده‌های موردنظر در دامنه‌ای از طول موج‌های مشخص گفته می‌شود. در طیف‌سنجی انرژی الکترومغناطیسی گسیل شده از منبع نور به پدیده برخورد کرده، قسمتی از آن منعکس، بخشی جذب و قسمتی دیگر عبور می‌کند. از این‌رو، طیف‌سنجی کمی کردن انعکاس، جذب و یا عبور انرژی الکترومغناطیسی از یک پدیده است. طیف‌سنجی باتوجه به منحنی‌های طیفی هر پدیده، اطلاعات ترکیبی و ساختار مولکول‌ها را در دامنه طول موج‌های Vis-NIR بررسی می‌کند؛ شایان ذکر است که فرکانس‌های پایدار مولکول‌های مربوط به اجزای سازنده خاک‌ها بیشتر در این دو دامنه واقع شده‌اند (Stenberg *et al.*, 2010). به لحاظ ویژگی‌های خاص ترکیب مولکولی و همچنین ساختاری، اجزای تشکیل دهنده خاک رفتار طیفی ویژه‌ای از خود نشان می‌دهند. از این رو با بهره‌گیری از منحنی‌های طیفی خاک‌ها و داده‌های نهفته در آنها می‌توان اطلاعات مختلفی از ویژگی‌های فیزیکی و شیمیایی خاک‌ها استخراج کرد. ویژگی‌هایی از خاک‌ها همچون بافت، ساختمان، کانی‌شناسی، اکسید و هیدروکسیدهای آهن و منگنز، رنگ، رطوبت، ماده آلی، کربنات‌ها و شوری بر روی انعکاس طیفی خاک‌ها اثر می‌گذارند (Chen *et al.*, 2014). افزایش مقدار رس می‌تواند منجر به کاهش بازتاب طیفی خاک شود. موضوع کاهش بازتاب طیفی خاک با کاهش اندازه ذرات خاک توسط Mousavi *et al.*

1 Proximal and remote sensing techniques

2 Visible

3 Near-Infrared

(2020) نیز گزارش شده است. این در حالی است که خاک‌های درشت‌دانه بازتاب طیفی بیشتری دارند که می‌تواند ناشی از رنگ سفید کوارتز و کانی‌های فلدسپار پتاسیم تشکیل دهنده این خاک‌ها باشد. این موضوع همچنین می‌تواند به دلیل نگره‌داری کمتر رطوبت در خاک‌های درشت بافت در مقایسه با خاک‌های ریزبافت‌ها باشد. کانی‌هایی کلسیت و دولومیت که در خاک‌های مناطق خشک و نیمه‌خشک غالب بوده اثر قابل توجهی روی بازتاب طیفی خاک دارند. واقعیت این است که وجود کانی‌هایی روشن همچون سیلیکات‌ها و کربنات‌ها در خاک، می‌تواند باعث افزایش بازتاب طیفی خاک شود. راجع به pH خاک نیز انتظار نمی‌رود که پاسخ طیفی مستقیمی داشته باشد، با این حال در پژوهش‌های متعددی کم و بیش به خوبی با استفاده از داده‌های طیفی برآورد شده است (Stenberg et al., 2010). Chang et al. (2001) بیان کردند که این می‌تواند به دلیل همبستگی pH با ترکیبات فعال طیفی خاک همچون ماده آلی و رس باشد.

## پیشینه پژوهش

غالب پژوهش‌های صورت گرفته راجع به کارایی طیف‌سنجی Vis-NIR در پیش‌بینی ویژگی‌های مبنایی خاک همچون ماده آلی (Zhao et al., 2022; Nocita et al., 2013)، آهک خاک (Khayamim et al., 2015; Gomez et al., 2008)، هدایت الکتریکی و ظرفیت تبادل کاتیونی خاک (Rasooli et al., 2023; Savides et al., 2010) و همچنین تعیین بعضی از عناصر غذایی ماکرو و میکرو (Kodaira et al., 2013) بوده است. در این راستا، رسولی و همکاران (۱۳۹۷) برخی ویژگی‌های خاکی منطقه بردسیر کرمان را با استفاده از روش طیف‌سنجی Vis-NIR پیش‌بینی کردند. این محققین گزارش کردند که با توجه به مقدار آماره  $RPD^1$  (نسبت انحراف معیار مقادیر اندازه گیری شده به حداقل مربعات خطا) پیش‌بینی مدل رگرسیون حداقل مربعات جزئی  $PLSR^2$  برای آهک مناسب و برای مقادیر سیلت و شن کاملاً مناسب می‌باشد. طیبی و همکاران (۱۳۹۷) در پژوهشی با استفاده از داده‌های طیفی به برآورد اجزای تشکیل دهنده بافت خاک پرداختند. نتایج حاصل از این تحقیق بیان‌گر کارایی مدل PLSR در پیش‌بینی درصد رس و سیلت بود. Chatrenour et al. (2022) در پژوهشی کارایی مدل‌های خطی [شامل PLSR و شاخص نسبت ساده  $SRI^3$ ] و غیرخطی [(شامل رگرسیون حداقل مربعات جزئی - شبکه عصبی پس انتشار  $PLS-BPNN^4$ ) و رگرسیون حداقل مربعات جزئی - جنگل تصادفی  $PLS-RF^5$ ] را در برآورد مقدار رس و هدایت الکتریکی منابع گردوغبار بررسی کردند. با توجه به نتایج این پژوهش، مدل‌های PLSR و PLS-BPNN بهترین کارایی را در برآورد مقدار رس و هدایت الکتریکی خاک داشتند. این در حالی است که روش SRI در برآورد مقدار رس و شوری خاکی از دقت پایینی برخوردار بود. Dotto et al. (2017) نشان دادند که از یک کتابخانه طیفی منطقه‌ای Vis-NIR و رگرسیون‌های PLS و ماشین بردار پشتیبان (SVMR) می‌توان برای پیش‌بینی رس، سیلت و شن استفاده کرد. Adeline et al. (2017) توانایی پیش‌بینی pH سطحی خاک را از یک کتابخانه Vis-NIR در مقیاس میدانی با استفاده از PLSR گزارش کردند. Khayamim et al. (2015) بیان کردند که برآوردهای مدل Vis-NIR PLSR از ماده آلی قابل قبول و برای گچ نیز کاملاً مناسب است؛ حال آن‌که برآوردهای این مدل برای کربنات‌های خاک ضعیف می‌باشد. Shiferaw and Hergarten (2014) در تحقیقی به برآورد کربن آلی خاک با استفاده از داده‌های طیفی و مدل PLSR و  $PCR^6$  پرداختند. مقادیر  $R^2$  و RPD تخمین کربن آلی خاک در مدل PLSR به ترتیب برابر با ۰/۶ و ۱/۵ و در مدل PCR نیز معادل ۰/۵ و ۱/۵ بود که این نتایج نشان از برتری مدل PLSR نسبت به مدل PCR داشت. Qiu (2010) از داده‌های طیفی ۱۹۸ نمونه خاک به برآورد تغییرات رس، pH، کربن آلی، کربن غیرآلی و CEC با استفاده از مدل PLSR پرداختند. این محققین نشان دادند که با بهره‌گیری از داده‌های طیفی و مدل PLSR امکان جداسازی ویژگی‌های مهم این نمونه‌ها با مواد مادری متفاوت وجود دارد.

برای رسیدن به برآوردهایی دقیق از یک کتابخانه طیفی Vis-NIR، انتخاب یک رویکرد مدل‌سازی مناسب برای طیف‌ها همراه با ویژگی‌های خاک ضروری است. تاکنون مدل‌های مختلفی برای بررسی رابطه بین داده‌های طیفی خاک با ویژگی‌های فیزیکوشیمیایی خاک ارائه شده است (Zhao et al., 2021). از بین این مدل‌ها، افزون بر رگرسیون خطی چندگانه  $MLR^7$ ، می‌توان به رگرسیون مؤلفه‌های اصلی (PCR) و رگرسیون حداقل مربعات جزئی (PLSR) اشاره کرد (Breiman, 1996). رگرسیون PLS از جمله روش‌های نوین آماری چند متغیره است که از داده‌های طیفی در تمامی دامنه طول موجی بهره گرفته و با گنجاندن آن‌ها در تعداد کمتری فاکتورهای

1 Ratio of predicted deviation

2 Partial least-squares regression

3 Simple ratio index

4 Partial least squares-backpropagation neural network

5 Partial least squares-random forest

6 Principal component regression

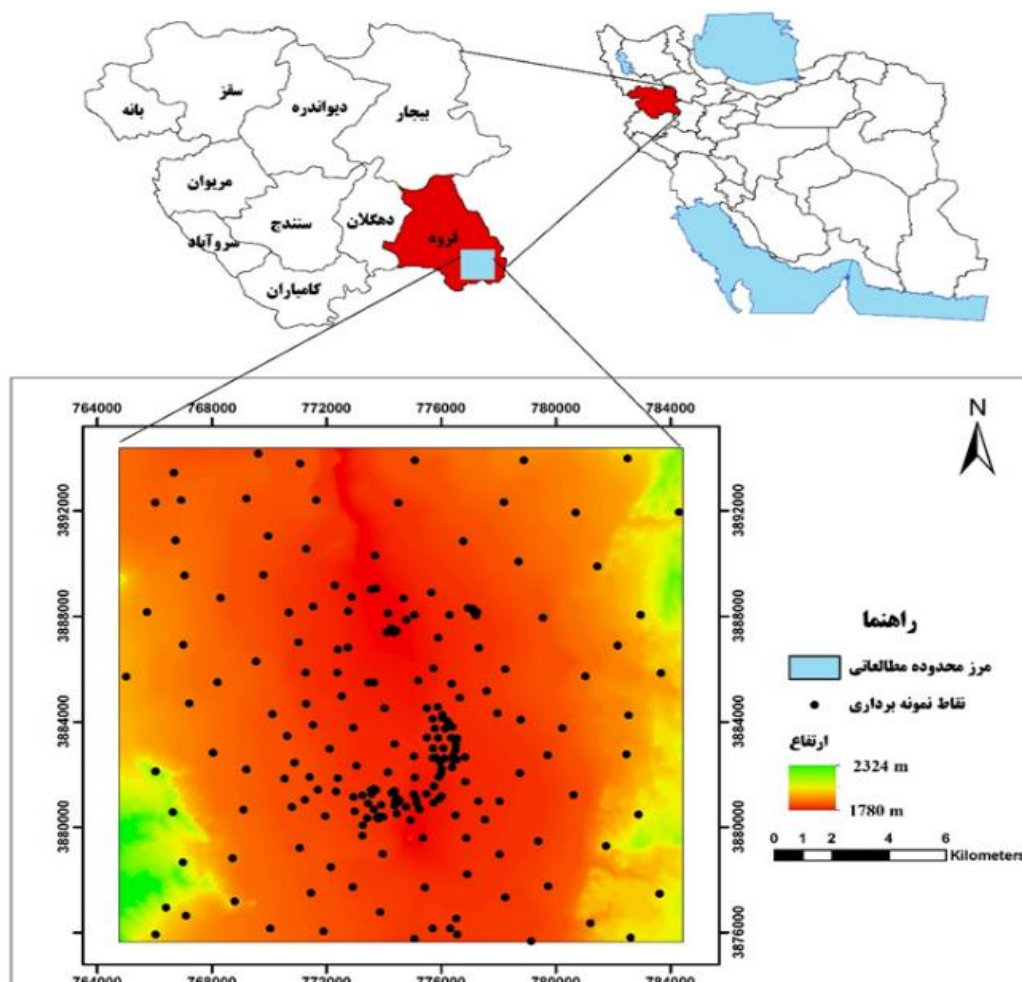
7 Multiple linear regression

تخمین‌گر و اشتقاق یک رابطه خطی، ویژگی‌های خاکی موردنظر را برآورد می‌کند (Wold et al., 2001). این روش در بسیاری از پژوهش‌های برآورد ویژگی‌های خاکی از دقت قابل‌قبولی برخوردار بوده است (Esbensen, 2006). می‌توان گفت که Vis-NIR PLSR متداول‌ترین روشی است که در برآورد خصوصیات خاک با استفاده از داده‌های طیفی به‌کار گرفته می‌شود (Zhao et al., 2021). این در حالی است که نتایج حاصل از برخی پژوهش‌ها بیان‌گر آن است که استفاده ترکیبی از رگرسیون حداقل مربعات جزئی توأم با بازنمونه‌گیری (bagging-PLSR) می‌تواند در بهبود دقت مقادیر برآوردی مؤثرتر باشد (بابائیان و جلالی، ۱۳۹۵؛ Zhao et al., 2021). افزایش دقت نتایج مدل‌های بیش‌برآورد و ارائه عدم اطمینان مقادیر برآوردی، از مزیت‌های bagging-PLSR به‌شمار می‌رود. بنابراین، هدف از انجام این پژوهش مطالعه رفتار طیفی خاک‌ها جهت برآورد مقادیر ذرات شن، سیلت، رس، pH و کربنات کلسیم معادل خاک با بهره‌گیری از روش PLSR و bagging-PLSR بود.

### روش‌شناسی پژوهش

#### مشخصات منطقه مورد مطالعه

منطقه مورد مطالعه در بیست کیلومتری شهرستان قروه در استان کردستان به طرف استان همدان قرار گرفته است (شکل ۱). اراضی محدوده مطالعاتی در حد فاصل عرض شمالی  $33^{\circ}38'07.5''$  تا  $35^{\circ}38'09.4''$  و طول شرقی  $47^{\circ}56'05.9''$  تا  $44^{\circ}39'78.0''$  واقع شده است. بیشینه و کمینه ارتفاع این منطقه به ترتیب برابر با ۲۳۲۴ متر و ۱۷۸۰ متر از سطح دریا است. وسعت تقریبی این منطقه ۳۷۰۰۰ هکتار و میانگین بارش و دمای سالانه آن به ترتیب ۳۶۹/۸ میلی‌متر و ۱۰/۲۰ درجه سانتی‌گراد می‌باشند. زراعت (عمدتاً گندم و جو)، اراضی بایر و مرتع، کاربری‌های اراضی اصلی منطقه است. رژیم رطوبتی و حرارتی این منطقه به ترتیب زیریک و مزیک می‌باشد. مطابق سیستم طبقه‌بندی مبنای جهانی (WRB) (IUSS, 2015)، خاک‌های منطقه مورد مطالعه عمدتاً تحت عنوان Cambisols و Leptosols طبقه‌بندی می‌شوند.



شکل ۱. موقعیت منطقه مطالعاتی در کشور و استان و توزیع مکانی نقاط نمونه‌برداری



### نمونه‌برداری و اندازه‌گیری ویژگی‌های خاکی

در شهریور ۱۳۹۸، ۲۲۰ نمونه خاک از عمق ۲۰-۰ سانتی‌متری با استفاده از روش نمونه‌برداری تصادفی طبقه‌بندی شده (براساس مقدار هدایت الکتریکی و شیب) جمع‌آوری شد (شکل ۱). با انتقال نمونه‌ها به آزمایشگاه و هوا خشک کرن آنها، نمونه‌ها به آرامی کوبیده شده و از الک دو میلی‌متری عبور داده شدند. سپس ویژگی‌هایی همچون بافت خاک با استفاده از روش هیدرومتری، واکنش خاک (pH) در عصاره اشباع با استفاده از دستگاه pH متر و کربنات کلسیم معادل خاک (CCE) با روش تیتراسیون برگشتی با اسید کلریدریک مورد اندازه‌گیری قرار گرفت (Gee and Or, 2002; Sparkes *et al.*, 1996).

### اندازه‌گیری خصوصیات طیفی خاک

بازتاب‌های طیفی نمونه‌های خاک مورد مطالعه با استفاده از دستگاه اسپکتروادیومتر (FieldSpec@3, ASD, FR, USA) آزمایشگاه خاکشناسی دانشگاه تربیت مدرس ثبت گردید (شکل ۲). بدین منظور حدود پنجاه گرم خاک هوا خشک عبور داده شده از الک دو میلی‌متری در پتری دیشی با قطر ده سانتی‌متر و ضخامت یک سانتی‌متر قرار داده شد. سپس با بهره‌گیری از یک پراب تماسی، انعکاس طیفی تمامی نمونه‌ها در دامنه ۲۵۰۰-۳۵۰ نانومتر با درجه وضوح یک نانومتر اندازه‌گیری شد. جدار شیشه‌ای پراب به صورت قائم در تماس با نمونه خاک‌ها قرار گرفته و بازتاب‌ها به وسیله یک سنسور ثبت گردید. برای اطمینان از دقت طیف‌های ثبت شده پنج تکرار از هر نمونه قرائت شد (Viscarra Rossel, 2008). جهت افزایش دقت اندازه‌گیری‌ها و پاسخ آشکارساز، با بهره‌گیری از یک صفحه سفید استاندارد<sup>۱</sup> (بازتاب ۱۰۰ درصد) دستگاه واسنجی گردید. ثبت منحنی‌های طیفی هر کدام از نمونه خاک‌ها با استفاده از نرم‌افزار RS3 موجود بر روی رایانه متصل به دستگاه اسپکتروادیومتر انجام شد. از نرم‌افزار View Spec Pro 6.0.15 جهت میانگین گرفتن از پنج قرائت برای هر کدام از نمونه‌ها استفاده شد و یک منحنی طیفی به ازای هر نمونه خاک بدست آمد. پس از تأیید درستی بازتاب‌های طیفی اولیه در نرم‌افزار View Spec Pro 6.0.15، منحنی‌های طیفی به صورت کتابخانه‌های طیفی و در قالب فایل‌های متنی ذخیره شدند و سپس جهت اعمال روش‌های مختلف پیش پردازش و اشتقاق مدل‌های رگرسیونی به نرم‌افزار ParLeS v3.1 منتقل شدند.



شکل ۲. دستگاه طیف‌سنج زمینی (FieldSpec@3, ASD, FR, USA)

### پیش‌پردازش داده‌های طیفی خاک

انجام تبدیل‌های طیفی و فرایندهای پیش‌پردازش داده‌ها جهت مناسب‌تر کردن آن‌ها، از اهمیت قابل توجهی در بهبود واسنجی مدل‌های پیشنهادی برخوردار است (Stenberg *et al.*, 2010). جهت حذف آشفتگی داده‌های طیفی (نویز)، از بین بردن عوامل محیطی، تصحیح اثرات پس‌زمینه، حذف غیرخطی‌ها در طیف‌ها و همچنین ارتقاء کیفیت داده‌ها و در نهایت دستیابی به نتایج بهتر واسنجی مدل‌ها، از روش‌های مختلف پیش‌پردازش و پیش‌تیمار طیف‌ها استفاده می‌شود. در این راستا و به منظور پیش‌پردازش داده‌های طیفی، الگوریتم‌های مختلفی همچون فیلتر ساویتزکی و گلائی (Savitzky - Golay) و فیلتر میانه (برای حذف نویز و صاف کردن طیف‌ها)، با استفاده از مشتق‌های اول یا دوم (برای حذف اثرات پس‌زمینه)، همچنین با استفاده از متغیر نرمال استاندارد و تصحیح پخشیده چندگانه (برای حذف پراکندگی نور و تصحیح خط پایه) و تبدیل‌های جذب و حذف پیوستار (برای حذف غیرخطی‌ها در طیف‌ها) پیشنهاد شده است (Viscarra Rossel *et al.*, 2010; Esbensen, 2006). در این پژوهش، فرایند پیش‌پردازش داده‌های طیفی با استفاده از الگوریتم‌های مختلف تعبیه شده در نرم‌افزار ParLeS v3.1 انجام شد. محدوده‌های طیفی ۳۵۰-۴۵۰ و ۲۴۵۰-۲۵۰۰ نانومتر نیز به دلیل انحراف معیار بالا از فرایند مدل‌سازی حذف شدند (Chatrenour *et al.*, 2023). نتایج حاصل از اعتبارسنجی متقابل و ارزیابی نتایج مدل‌ها نشان دهنده مناسب

بودن فیلتر ساویتزکی و گلای نسبت به دیگر انواع رویکردهای پیش‌پردازش بود. منحنی‌های حذف پیوستار نیز، براساس یک خط محدب فرضی که دو انتهای یک مشخصه جذب طیفی را بهم متصل می‌کنند، با بهره‌گیری از نرم‌افزار ENVI 4.8 و طبق رابطه زیر محاسبه شدند:

$$R_{CR}(\lambda_i) = \frac{R_{\lambda_i}}{R_{CL}(\lambda_i)} \quad \text{رابطه (۱)}$$

که در آن،  $R_{CL}(\lambda_i)$  نشان‌دهنده میزان بازتاب در منحنی حذف پیوستار به ازای طول موج  $\lambda_i$ ، مقدار بازتاب طول موج متناظر در منحنی طیفی خام و  $R_{CR}(\lambda_i)$  مقدار بازتاب منحنی فرضی در طول موج‌های متناظر است. مقادیر منحنی‌های حذف پیوستار همواره در محدوده صفر تا یک بوده و این منحنی‌ها موجب بهبود و آشکارسازی تفاوت‌های موجود در طیف‌های خاک می‌شوند.

### روش رگرسیون حداقل مربعات جزئی (PLSR)

در این روش، متغیر پاسخ با استفاده از متغیرهای پنهان با ماتریس متغیر تخمین‌گر ارتباط داده می‌شود:

$$X = t_1 p_1 + \dots + t_a p_a + E_a \quad \text{رابطه (۲)}$$

$$Y = t_1 q_1 + \dots + t_a q_a + f_a$$

که در آن، پارامتر  $t$  بیان‌گر بردار رتبه‌ها، فاکتور  $p$  نشان‌دهنده بارگذاری‌های طیفی، پارامتر  $q$  بیان‌گر بارگذاری‌های ویژگی‌های خاکی، پارامترهای  $E$  و  $f$  به ترتیب نشان‌دهنده خطای باقیمانده متغیرهای تخمین‌گر و نتایج حاصل از اثر تخمینی پارامتر  $a$  ( $a = 1, 2, \dots, A$ ) است.

$$t_a = X_{a-1} \cdot W_a \quad \text{رابطه (۳)}$$

در رابطه (۳)، پارامتر  $W_a$  نشان‌دهنده ضریب وزنی مقیاس‌سازی شده و پارامتر  $c$  نشان‌دهنده فاکتور مقیاس‌می‌باشد. ضرایب مدل رگرسیون ( $b$  و  $b_0$ ) براساس روابط ۴ و ۵ محاسبه و در مدل برآوردگر (رابطه ۶) بر مبنای تعداد فاکتورهای بهینه اتخاذ شده، برای برآورد متغیر مورد نظر ( $\hat{y}_i$ ) مورد استفاده قرار می‌گیرد. در نهایت با بهره‌گیری از رابطه ۶ مقادیر برآوردی متغیر مورد نظر  $\hat{f}_{bag}(y)$  محاسبه می‌شود.

$$b = w(p'w)^{-1} \cdot q \quad \text{رابطه (۴)}$$

$$b_0 = \bar{y} - \bar{x}'b \quad \text{رابطه (۵)}$$

$$\hat{y}_i = b_0 + \hat{x}_i \quad \text{رابطه (۶)}$$

لازم به ذکر است که در روابط فوق  $X_i$  باند طیفی است.

### روش رگرسیون حداقل مربعات جزئی توأم با بازنمونه‌گیری (bagging-PLSR)

روش بازنمونه‌گیری مورد استفاده در bagging-PLSR، روش نمونه‌گیر bootstrap است. در این روش به صورت تصادفی نمونه‌برداری از داده‌ها انجام شده، سپس از طریق همان داده نمونه‌برداری شده جایگزینی دوباره داده‌ها صورت می‌گیرد. در صورتی که سری داده‌های واسنجی به صورت  $Z_N = (z_1, z_2, \dots, z_n)$  لحاظ شود که در آن  $z_n = (x_n, y_n)$  و  $N$  تعداد داده‌ها باشد و نمونه‌برداری از این سری  $B$  مرتبه تکرار شود ( $b = 1, 2, \dots, B$ ) تعداد  $B$  سری داده راه‌انداز شده ( $Z_N^{*b}$ ) تشکیل خواهد شد. هر یک از سری داده‌های حاصل حدود ۶۳ درصد داده‌ها را پوشش داده و منجر به ایجاد سری‌هایی هم‌تای و منطبق با تعداد داده‌های اولیه می‌شود. سپس بهترین مدل PLSR به هر یک از این سری‌های جدید داده‌ها برازش داده شده و مقادیر برآوردی  $\hat{f}^{*b}(y)$  حاصل می‌شود. در نهایت با استفاده از رابطه زیر مقادیر برآوردی متغیر مورد نظر محاسبه می‌گردد.

$$\hat{f}_{bag}(y) = \frac{1}{B} \sum_{b=1}^N \hat{f}^{*b}(y) \quad \text{رابطه (۷)}$$

### ارزیابی کارایی مدل‌ها

جهت ارزیابی دقت مدل‌های پیشنهادی در هر یک از مراحل واسنجی و اعتبارسنجی از آماره‌های مختلفی همچون ضریب تبیین ( $R^2$ )، ریشه میانگین مربعات خطا ( $RMSE$ )، نسبت انحراف دقت (RPD)، میانگین خطا  $(ME)^2$  و شاخص انطباق (d) استفاده می‌گردد. شاخص انطباق نشان‌دهنده درجه‌ای از دقت بوده که در آن مقادیر تخمینی به دست آمده است. در مقایسه با ضریب تبیین که بیانگر همبستگی بین

1 Root mean square error

2 Mean error



مقادیر تخمینی و مشاهده‌ای است، شاخص انطباق درجه‌ای را بیان می‌کند که در آن پیش‌بینی‌های مدل عاری از خطا می‌باشند.

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2} \quad \text{رابطه ۸}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{رابطه ۹}$$

$$RPD = \frac{SD}{RMSE} \quad \text{رابطه ۱۰}$$

$$ME = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \quad \text{رابطه ۱۱}$$

$$d = 1 - \frac{\sum_{i=1}^n (\theta_i^p - \theta_i^o)^2}{\sum_{i=1}^n [ |(\theta_i^p - \bar{\theta}^o)| + |(\theta_i^o - \bar{\theta}^o)| ]^2} \quad \text{رابطه ۱۲}$$

$y_i$  و  $\hat{y}_i$  به ترتیب بیانگر مقادیر مشاهده‌ای و میانگین آنها،  $\bar{y}$  و  $\bar{\hat{y}}$  بیانگر مقادیر تخمینی و میانگین آنها،  $N$  تعداد داده‌ها،  $m$  تعداد فاکتورهای مدل و  $SD$  انحراف معیار داده‌ها است. انتخاب تعداد فاکتورهای مناسب (متغیرهای تخمینگر) در مدل PLSR به کمک آماره  $RMSE$  انجام شد. آماره  $RPD$  یکی از برجسته‌ترین شاخص‌های ارزیابی کارایی مدل‌ها در پژوهش‌های مربوط به طیف‌سنجی بوده که از نسبت خطای استاندارد داده‌ها به ریشه میانگین مربعات خطا ( $RMSE$ ) مدل برآوردگر به دست می‌آید. براساس مطالعات انجام شده، مقادیر  $RPD > 2$  بیانگر دقت بسیار بالا،  $RPD = 1.4-2$  بیانگر دقت خوب تا متوسط و مقادیر  $RPD < 1.4$  بیانگر دقت ضعیف فرآیند تخمین است.

## یافته‌های پژوهش و بحث

### توصیف آماری ویژگی‌های خاک

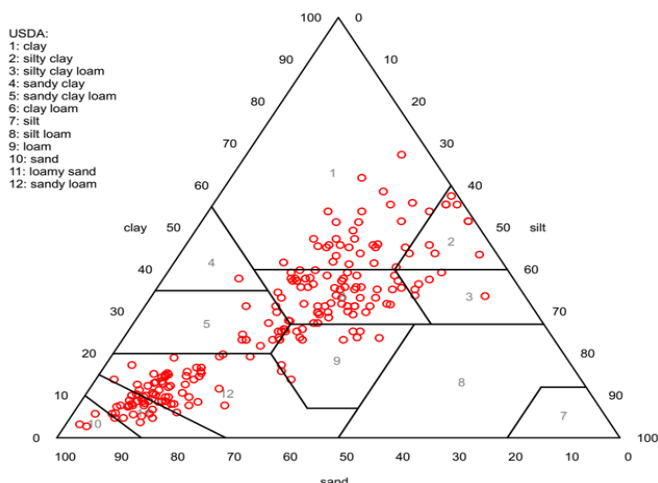
در جدول (۱)، پارامترهای آماری مربوط به ویژگی‌های خاک‌های مورد مطالعه برای سری داده‌های اعتبارسنجی و واسنجی نشان داده شده است. همان‌طور که در این جدول مشاهده می‌شود مقادیر pH با میانگینی برابر با ۸/۱۰ محدودده‌ای از ۶/۴۴ تا ۱۰/۷۳ را در برمی‌گیرد که خود بیانگر غالب بودن شرایط قلیایی در منطقه مورد مطالعه است. کربنات کلسیم معادل خاک (CCE) نیز با میانگینی برابر با ۲۲/۸۷ درصد دارای کمینه و بیشینه‌ای به ترتیب برابر با ۱/۵۰ و ۶۹/۷۵ درصد می‌باشد.

جدول ۱. توصیف آماری ویژگی‌های خاک در دو سری داده‌ها واسنجی و اعتبارسنجی

ویژگی‌های خاک	داده‌های واسنجی مدل (۱۵۴ نمونه)				داده‌های اعتبارسنجی مدل (۶۶ نمونه)			
	انحراف معیار	حداکثر	حداقل	میانگین	انحراف معیار	حداکثر	حداقل	میانگین
شن	۲۴/۴۰	۹۴/۳۲	۱/۱۴	۵۱/۳۳	۲۴/۹۹	۹۳/۳۲	۱/۱۴	۴۴/۳۶
سیلت	۱۱/۱۳	۴۸/۷۲	۳/۲۸	۲۴/۲۱	۱۱/۵۹	۴۷/۳۶	۸	۲۵/۸۵
رس	۱۴/۴۲	۶۷/۳۲	۲/۶۷	۲۶/۵۷	۱۶/۰۱	۵۷/۵۰	۳/۱۷	۲۵/۸۲
pH	۰/۸۴	۱۰/۷۳	۶/۴۴	۸/۱۰	۰/۸۶	۱۰/۵۸	۶/۶۷	۸/۱۰
کربنات کلسیم معادل	۱۲/۹۶	۶۹/۷۵	۱/۵۰	۲۱/۶۹	۱۳/۳۴	۶۸/۵۰	۴/۷۵	۲۴/۰۶

مقادیر ذرات رس، سیلت و شن به ترتیب محدوده‌هایی از ۲/۶۷ تا ۶۷/۳۲، ۳/۲۸ تا ۴۸/۷۲ و از ۱/۱۴ تا ۹۴/۳۲ درصد را در بر می‌گیرد که نشان دهنده گستردگی کلاس‌های بافتی خاک‌های مورد مطالعه می‌باشد. شکل (۳) نیز تأییدی بر نتایج جدول (۱) است که خاک‌های مورد بررسی از نقطه نظر کلاس‌های بافتی بسیار متنوع و گسترده بوده و در مجموع شامل ۱۰ کلاس بافتی شنی، لوم شنی، شنی لومی، لوم، لوم رسی، لوم رس شنی، رس شنی، لوم رس سیلتی، رس سیلتی و رسی می‌باشند.



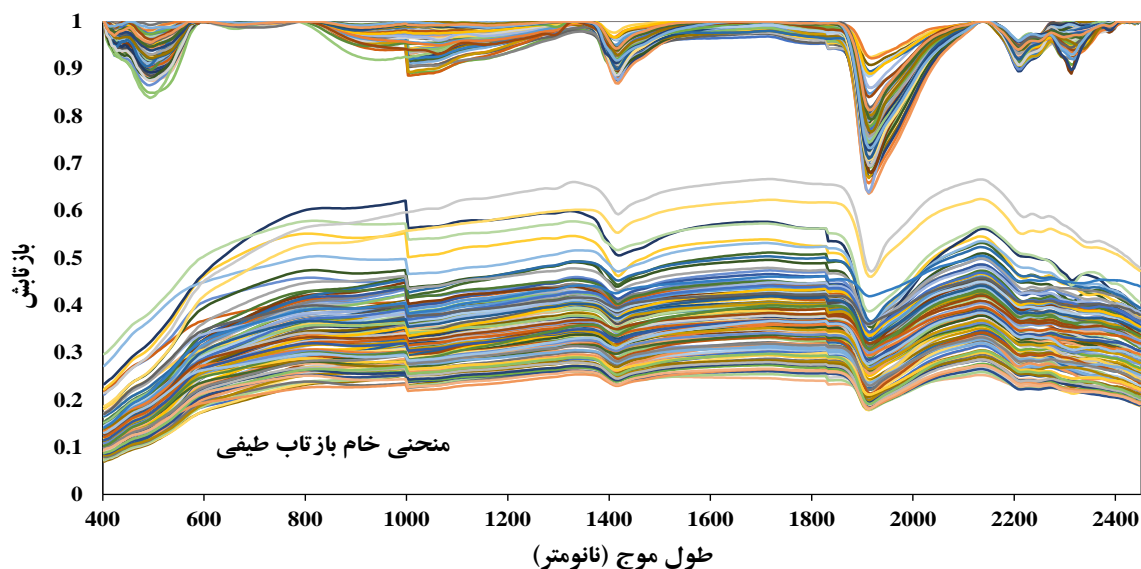


شکل ۳. توزیع کلاس بافتی خاک‌های مورد مطالعه

### طیف‌سنجی

#### منحنی‌های بازتاب طیفی

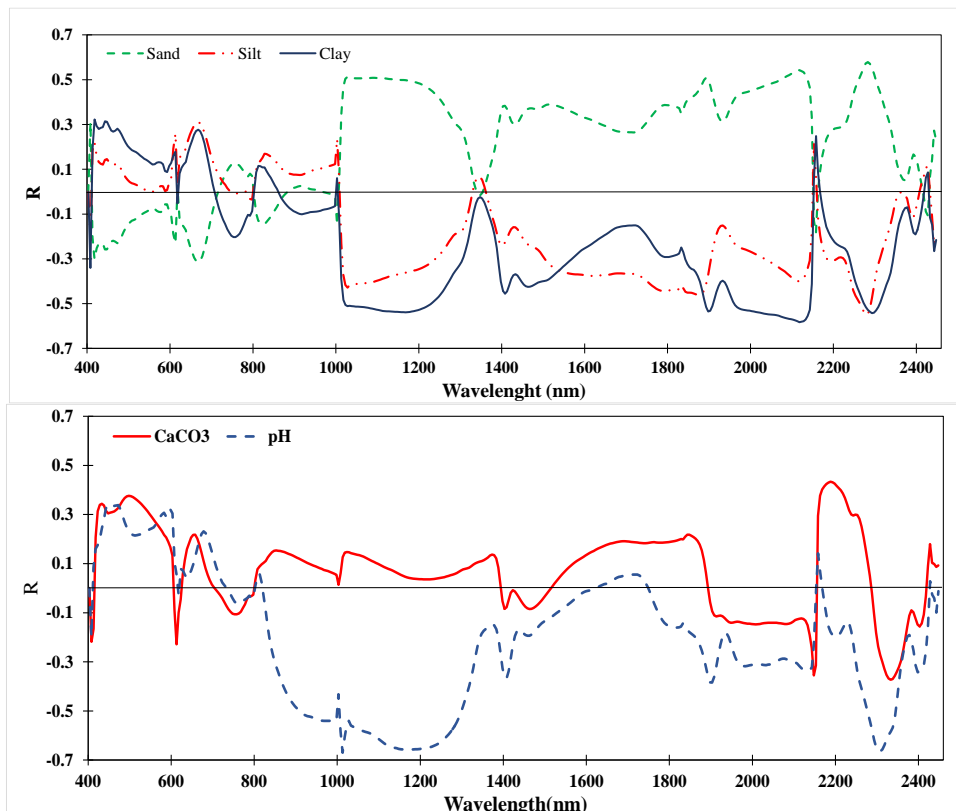
در شکل (۲) تصویر منحنی‌های انعکاس طیفی شامل منحنی‌های خام طیفی و منحنی‌های حذف پیوستار نمونه‌های خاک مورد مطالعه ارائه شده است. همان‌گونه که در این شکل مشاهده می‌شود نمای کلی طیف‌ها در همه خاک‌ها تقریباً مشابه است، هرچند در میزان شدت بازتاب اختلافاتی وجود دارد. با توجه به این شکل، منحنی‌های خام طیفی در محدوده ۶۰۰-۵۰۰ نانومتر دارای یک برآمدگی و در طول-موج‌های ۱۴۱۴، ۱۹۱۵ و ۲۲۱۲ نانومتر دارای مشخصه‌هایی طیفی هستند. شایان ذکر است که در منحنی‌های حذف پیوستار، این مشخصه‌های طیفی به صورت نمایان‌تری نشان داده شده است (شکل ۴). با مروری بر منابع مشخص می‌شود که مشخصه‌های طیفی در طول موج‌های ۱۴۱۴ و ۱۹۱۵ نانومتر می‌تواند به دلیل ارتعاشات ترکیبی آب پیوندی در شبکه بین‌لایه‌ای به‌عنوان کاتیون‌های هیدراته و آب جذب شده روی سطوح ذرات باشد. علاوه بر این، باند ۱۴۱۴ نانومتر می‌تواند همچنین ناشی از کشش O-H ساختاری در لایه اکتاهیدرال باشد. این در حالی است که ویژگی جذبی در طول موج ۲۲۱۲ نانومتر به دلیل ترکیبی از خمش پیوند Al-OH و کشش پیوند O-H می‌باشد (Stenberg *et al.*, 2010). با توجه به این شکل، کاهش میزان بازتاب طیفی در محدوده Vis قابل مشاهده بوده و بیشینه بازتاب در این محدوده ۰/۶ می‌باشد. در محدوده NIR میزان بازتاب‌های طیفی بیشتر بوده که بیشینه ۰/۷۷ را نشان می‌دهد. در این پژوهش کمینه و بیشینه مقدار بازتاب طیفی بدست آمده معادل ۰/۰۸ و ۰/۷۷ می‌باشد. کریمی و همکاران (۱۳۹۶) نیز مقادیر ۰/۰۳ و ۰/۴۷ را به عنوان کمینه و بیشینه مقدار بازتاب طیفی نمونه خاک‌های مورد بررسی گزارش کردند.



شکل ۴. میانگین منحنی‌های خام و حذف پیوستار بازتاب طیفی خاک‌های مورد مطالعه

### بررسی ضریب همبستگی پیرسون بین ویژگی‌های فیزیکی و شیمیایی مورد مطالعه و داده‌های طیفی

همبستگی بین ویژگی‌های خاکی مورد مطالعه با داده‌های طیفی محدوده طول موج‌های ۲۴۵۰-۴۰۰ نانومتر (Vis-NIR) در شکل (۵) ارائه شده است. با توجه به شکل (۵)، ویژگی‌های خاکی مورد بررسی با داده‌های طیفی در محدوده طول موج‌های Vis-NIR دارای همبستگی بالایی هستند. بررسی ضریب همبستگی پیرسون (R) بین داده‌های طیفی و ویژگی‌های فیزیکی و شیمیایی مورد مطالعه، بیان‌گر همبستگی‌های منفی و مثبت معنی‌داری بین آن‌ها در طول موج‌های مختلف می‌باشد. نتایج نشان داد که بیشینه ضریب همبستگی منفی و مثبت برای pH خاک به ترتیب معادل  $-0/66$  و  $0/33$  است که به ترتیب در طول موج‌های ۱۰۱۳ و ۴۴۸ نانومتر اتفاق می‌افتد. همچنین بررسی رابطه بین CCE و داده‌های طیفی نیز بیان‌گر همبستگی بالا و معنی‌دار این پارامتر با داده‌های طیفی در محدوده‌ای گسترده از طول موج‌های مختلف می‌باشد. بیشینه ضریب همبستگی مثبت برای CCE در طول موج ۲۱۸۸ نانومتر رخ داده که معادل  $0/43$  است؛ بیشینه ضریب همبستگی منفی این پارامتر که معادل  $-0/37$  می‌باشد در طول موج ۲۳۳۳ نانومتر اتفاق می‌افتد. در مجموع می‌توان اینگونه بیان کرد که بیشترین همبستگی CCE در طول موج‌های ۲۱۵۰ تا ۲۳۵۰ نانومتر و برای pH در طول موج‌های ۹۰۰ تا ۱۳۰۰ و ۲۳۰۰ تا ۲۳۹۷ نانومتر رخ می‌دهد. عزیززی و همکاران (۱۳۹۷) بیشینه همبستگی منفی و مثبت را برای pH خاک به ترتیب در طول موج‌های ۲۲۸۹ و ۲۳۹۷ نانومتر و برای CCE به ترتیب در طول موج‌های ۱۳۷۹ و ۲۰۹۰ نانومتر گزارش کرده‌اند. پژوهش‌های پیشین نشان داده‌اند که مشخصه‌های جذبی قوی نزدیک به طول موج ۲۳۴۱ نانومتر، می‌تواند ناشی از وجود گروه‌های دارای بنیان  $CO_3$  در کانی‌های کربناته باشد. نتایج همچنین نشان داد توزیع اندازه ذرات خاک در طول موج‌های مختلف دارای همبستگی‌های مثبت و منفی معنی‌داری می‌باشد. بیشینه همبستگی مثبت برای رس و سیلت به ترتیب  $0/32$  و  $0/31$  بوده که در طول موج‌های ۴۱۸ و ۶۶۸ نانومتر رخ می‌دهد. افزون بر این، بیشینه همبستگی منفی ذرات رس و سیلت نیز برابر  $-0/58$  و  $-0/54$  می‌باشد که به ترتیب در طول موج‌های ۲۱۱۸ و ۲۲۷۸ نانومتر به دست آمد. بیشینه همبستگی منفی و مثبت برای شن به ترتیب معادل  $-0/31$  و  $0/57$  می‌باشد که به ترتیب در طول موج‌های ۶۶۸ و ۲۲۸۳ نانومتر اتفاق می‌افتد. در مجموع بیشترین همبستگی بین بازتاب طیفی و بافت خاک در محدوده‌ای حدود ۱۰۰۰ تا ۲۳۰۰ نانومتر طیف الکترومغناطیسی رخ داده است.



شکل ۵. ضریب همبستگی پیرسون بین پارامترهای فیزیکی و شیمیایی مورد مطالعه با مقادیر بازتاب طیفی خاک در طول موج‌های مختلف

### اشتقاق مدل‌هایی جهت برآورد ویژگی‌های خاکی مورد مطالعه

پیش از واسنجی مدل‌های پیشنهادی جهت برآورد ویژگی‌های خاکی مورد مطالعه، تأثیر روش‌های مختلف پیش‌پردازش داده‌های طیفی بر

روی دقت مدل به روش PLSR مورد ارزیابی قرار گرفت (داده‌ها ارائه نشده است). با توجه به نتایج حاصل از آماره‌های ارزیابی، مشتق اول + فیلتر ساویتزکی - گلائی در مقایسه با سایر روش‌های پیش‌پردازش از دقت بالاتری برخوردار بود. در جدول ۲ آماره‌های حاصل از مدل‌سازی پارامترهای شن، سیلت، رس، pH و کربنات کلسیم معادل خاک توسط رگرسیون حداقل مربعات جزئی (PLSR) ارائه شده است. با توجه به نتایج، مقادیر ضریب تبیین ( $R^2$ ) برای ذرات شن، سیلت، رس، pH و  $\text{CaCO}_3$  در محدوده‌ای از ۰/۴۹ تا ۰/۷۵ متغیر بوده و مقادیر ریشه حداقل مربعات خطای (RMSE) آنها نیز به ترتیب معادل ۱۴/۴۸، ۸/۱۸، ۷/۲۸ درصد، ۰/۴۳ و ۹/۲۶ درصد می‌باشد. نتایج جدول (۲) همچنین بیان‌گر است که همه مدل‌های اشتقاق یافته با مقادیر ME بیشتر از صفر برای پارامترهای مورد مطالعه دارای بیش‌برآوردی هستند. آماره ضریب انطباق (d) نیز از ۰/۹۲ برای شن تا ۰/۹۸ برای pH متغیر است. در مجموع یافته‌های مرحله واسنجی نشان داد که نتایج مدل‌سازی برای رس، pH و شن خوب و برای سیلت و کربنات کلسیم متوسط می‌باشد. با توجه به مناسب‌ترین روش پیش‌پردازش استفاده شده در این پژوهش، بهترین مدل PLSR برای تخمین مقدار شن، سیلت، رس، pH و کربنات کلسیم معادل به ترتیب دارای تعداد فاکتورهای ۱۴، ۷، ۲۳، ۱۰ و ۱۸ است. شایان ذکر است که مقدار مرجع برای تعیین نمونه پرت به‌گونه‌ای تعریف شد که اختلاف بین مقادیر پیش‌بینی و اندازه‌گیری شده هر ویژگی بیشتر از سه برابر  $\text{RMSECV} / \text{RMSEP}$  باشد (Islam et al., 2003). برای این مبنا برای مرحله واسنجی، تعداد نمونه‌هایی که به‌عنوان داده پرت شناسایی و حذف شدند برای شن، سیلت، رس، pH و کربنات کلسیم به ترتیب ۴، ۷، ۴ و ۵ عدد بودند (جدول ۲).

جدول ۲. آماره‌های حاصل از برازش مدل PLSR بر ویژگی‌های خاکی مورد مطالعه

پارامتر	روش کالیبراسیون	ME	RMSE	$R^2$	d	تعداد فاکتور	تعداد داده پرت حذف شده
شن	PLSRCal	۲۰۹/۶۸	۱۴/۴۸	۰/۶۵	۰/۹۲	۱۴	۷
سیلت		۶۶/۹۰	۸/۱۸	۰/۴۶	۰/۹۷	۷	۴
رس		۵۲/۹۶	۷/۲۸	۰/۷۵	۰/۹۳	۲۳	۷
pH		۰/۱۸	۰/۴۳	۰/۷۴	۰/۹۸	۱۰	۴
کربنات کلسیم معادل		۸۵/۶۶	۹/۲۶	۰/۴۹	۰/۹۲	۱۸	۵

در جدول (۳) نتایج حاصل از ارزیابی دو مدل رگرسیونی PLSR و bagging-PLSR برای تمامی ویژگی‌های خاکی مورد مطالعه نشان داده شده است. در مدل PLSR مقادیر ضریب تبیین ( $R^2$ ) برای ذرات شن، سیلت، رس، pH و کربنات کلسیم معادل خاک در دامنه‌ای از ۰/۵۸ تا ۰/۷۶ متغیر است. این در حالی است که در مدل bagging-PLSR مقادیر  $R^2$  برای این خصوصیات در دامنه‌ای از ۰/۵۹ تا ۰/۷۴ به‌دست آمد. Curcio et al. (2013) نتایج برآوردهای VNIR-SWIR PLSR را برای رس دارای دقت بالا ( $\text{RMSE} = 5.8\%$ )،  $R^2 = 0.87$  و برای سیلت و شن رضایت‌بخش (به ترتیب دارای  $\text{RMSE} = 7.2\%$ ،  $R^2 = 0.60$  و  $\text{RMSE} = 7.7\%$ ،  $R^2 = 0.80$ ) گزارش کردند. با توجه به نتایج، مقادیر RMSE در مدل PLSR برای این ویژگی‌ها به ترتیب برابر با ۱۷/۴۳، ۷/۶۵ و ۷/۸۳ درصد و ۷/۹۴ درصد و در مدل bagging-PLSR این مقادیر به ترتیب ۱۶/۶۶، ۷/۶۳ و ۸/۱۳ درصد و ۷/۷۱ درصد و ۰/۴۵ محاسبه گردید. (Lacerda et al. 2016) در محدوده ۲۵۰۰ - ۴۰۰ nm توسط مدل PLSR مقادیر شن و رس را به ترتیب با ضریب تبیینی معادل ۰/۹۶ و ۰/۹۳ پیش‌بینی کردند. Hassani et al. (2014) نیز نتایج دقت بالایی را برای برآورد کربنات کلسیم معادل توسط مدل PLSR در محدوده Vis-NIR گزارش کردند ( $\text{RMSE} = 17.92\%$ ،  $R^2 = 0.65$ ). این در حالی است که نتایج آن‌ها در برآورد pH چندان رضایت‌بخش نبود ( $\text{RMSE} = 0.14\%$ ،  $R^2 = 0.35$ ). علاوه بر آماره‌های  $R^2$  و RMSE، دقت برآورد مدل توسط آماره RPD که یکی از شاخص‌های ارزیابی دقت در مطالعات طیف‌سنجی است، نیز ارزیابی می‌گردد (Viscarra Rossel et al., 2010). با توجه به نتایج جدول ۳، مقادیر RPD برای ذرات شن و کربنات کلسیم معادل خاک در مدل bagging-PLSR نسبت به مدل PLSR بالاتر بوده و در مورد ذرات سیلت هر دو مدل دارای نتایج یکسانی هستند. برای ذرات رس و pH خاک مدل PLSR نسبت به مدل bagging-PLSR دارای RPD بالاتری است. در مجموع با توجه به آماره RPD، برآوردهای هر دو مدل برای رس دارای دقت بسیار بالا و برای دیگر ویژگی‌ها شامل شن، سیلت، pH و کربنات کلسیم معادل خاک دارای دقت خوب تا متوسطی بود. (Jaconi et al. 2019) در پژوهشی بافت خاک بخشی از خاک‌های آلمان را با بهره‌گیری از طیف‌سنجی NIR، به‌عنوان یک روش آسان، سریع و کم‌هزینه، ارزیابی کردند. یافته‌های حاصل از پژوهش آن‌ها نشان



داد که مقادیر  $R^2$  به ترتیب برای ذرات رس، سیلت و شن ۰/۹۷، ۰/۹۴ و ۰/۹۶ و مقادیر RPD بدست آمده برای این ذرات ۶/۲، ۴/۲ و ۴/۸ می‌باشد. این پژوهشگران بیان کردند که تکنیک طیف‌سنجی در تجزیه و تحلیل بافت خاک در مقیاس‌های بزرگ و با داده‌های فراوان مفید می‌تواند روشی کم‌هزینه و مطلوب باشد. Silva et al. (2016) در پژوهشی کارایی داده‌هایی طیفی با استفاده از مدل PLSR را در پیش‌بینی ذرات بافت خاک گزارش کردند. بابائیان و جلالی (۱۳۹۵) در پژوهشی نشان دادند که مدل bagging-PLSR نسبت به مدل PLSR دارای دقت بالاتری در برآورد کربن آلی خاک می‌باشد.

جدول ۳. نتایج دقت آزمایشی مدل‌های PLSR و bagging-PLSR در برآورد ویژگی‌های خاکی مورد مطالعه

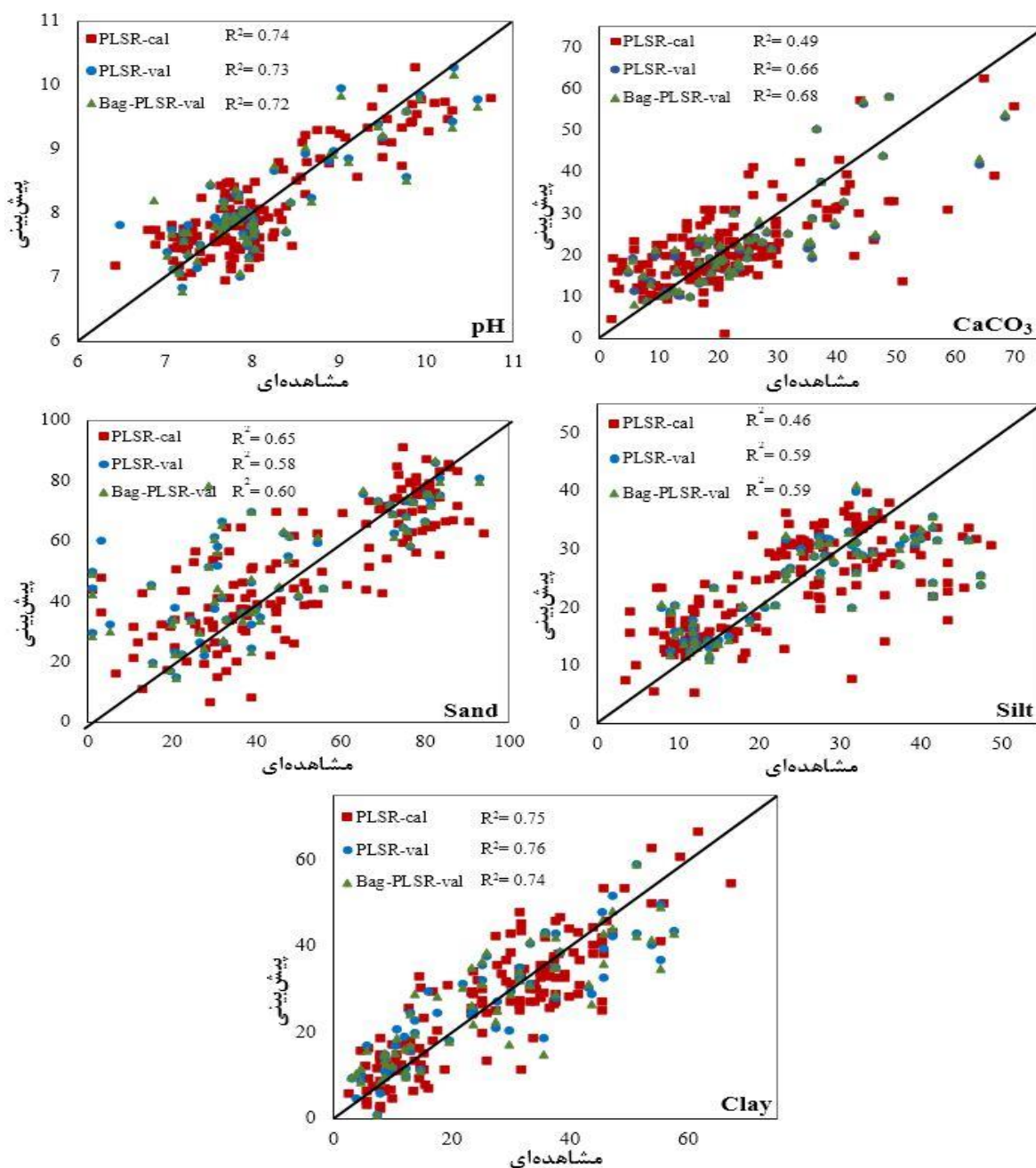
پارامتر	روش‌های مختلف	ME (%)	RMSE (%)	RPD	$R^2$
شن	PLSR <sub>Val</sub>	۳۰۳/۹۰	۱۷/۴۳	۱/۴۶	۰/۵۸
	Bagging-PLSR <sub>Val</sub>	۲۷۷/۷۰	۱۶/۶۶	۱/۵۰	۰/۶۰
سیلت	PLSR <sub>Val</sub>	۵۸/۵۹	۷/۶۵	۱/۵۳	۰/۵۹
	Bagging-PLSR <sub>Val</sub>	۵۸/۲۵	۷/۶۳	۱/۵۳	۰/۵۹
رس	PLSR <sub>Val</sub>	۶۱/۲۸	۷/۸۳	۲/۰۶	۰/۷۶
	Bagging-PLSR <sub>Val</sub>	۶۶/۱۰	۸/۱۳	۱/۹۹	۰/۷۴
pH	PLSR <sub>Val</sub>	۰/۲۱	۰/۴۶	۱/۹۲	۰/۷۳
	Bagging-PLSR <sub>Val</sub>	۰/۲۱	۰/۴۵	۱/۹۱	۰/۷۲
کربنات کلسیم معادل	PLSR <sub>Val</sub>	۶۳/۰۹	۷/۹۴	۱/۶۹	۰/۶۶
	Bagging-PLSR <sub>Val</sub>	۵۹/۳۹	۷/۷۱	۱/۷۵	۰/۶۸

در شکل (۶) نمودار پراکنش ۱:۱ مقادیر برآوردی در برابر مقادیر مشاهده‌ای ویژگی‌های خاکی مورد بررسی در گروه داده‌های اعتبارسنجی و واسنجی ارائه شده است. نتایج حاصل از این شکل و جدول ۳ مؤید این موضوع است که در برآورد مقدار شن و کربنات کلسیم معادل خاک بهترین عملکرد توسط مدل bagging-PLSR به دست آمد. این در حالی است که در برآورد مقدار رس و pH خاک، برغم اینکه اختلاف بین دو مدل PLSR و bagging-PLSR چند معنی‌دار نبود، اما مدل PLSR دارای عملکرد به نسبت بهتری بود. شایان ذکر است که برآوردهای هر دو مدل از مقدار سیلت تقریباً مشابه بود. در پژوهشی مشابه بابائیان و جلالی (۱۳۹۵) برتری bagging-PLSR را نسبت به PLSR در برآورد کربن آلی خاک گزارش کردند. Zhao et al. (2021) نیز بیان کردند که مدل bagging-PLSR در مقایسه با مدل Cubist از دقت بیشتری در برآورد رس، سیلت و شن برخوردار است. این در حالی است که در برآورد ظرفیت تبادل کاتیونی خاک، Cubist نسبت به bagging-PLSR قوی‌تر بود.

## نتیجه‌گیری

این پژوهش با هدف ارزیابی عملکرد مدل‌های PLSR و bagging-PLSR در برآورد اجزای بافت، pH و کربنات کلسیم معادل خاک با استفاده از داده‌های طیفی Vis-NIR در برخی از خاک‌های استان کردستان انجام شد. بر این اساس، یافته‌های این پژوهش نشان داد که داده‌های طیفی خاک در محدوده Vis-NIR در برآورد مقادیر ذرات شن، سیلت، رس، pH و کربنات کلسیم معادل خاک کارایی بالایی دارند. نتایج حاصل از اعمال روش‌های مختلف پیش‌پردازش بر روی داده‌های طیفی نیز بیان‌گر آن بود که مشتق اول + فیلتر ساویتزکی و گالای بهترین روش از بین روش‌های پیش‌پردازش داده‌های طیفی است. این بدین معنی است که پیش‌تیمار داده‌های طیفی خاک و گزینش بهترین روش پیش‌پردازش آن‌ها، یکی از عوامل مهم و تأثیرگذار در دقت مدل‌های PLSR و bagging-PLSR است. با توجه به نتایج، مدل bagging-PLSR در برآورد ذرات شن و کربنات کلسیم معادل خاک و مدل PLSR در برآورد ذرات رس و pH خاک دارای عملکرد بهتری بودند. لازم به ذکر است که هر دو مدل در برآورد مقدار سیلت نتایج تقریباً مشابهی داشتند. هر دو مدل PLSR و bagging-PLSR در برآورد رس از دقت بسیار بالا و در برآورد شن، سیلت، pH و کربنات کلسیم معادل خاک دارای دقت خوب تا متوسطی بودند. در کل می‌توان این‌گونه بیان کرد که استفاده از داده‌های طیفی Vis-NIR می‌تواند به‌عنوان روشی جایگزین برای روش‌های رایج آزمایشگاهی مطرح باشد و به کاهش تعداد نمونه‌ها در مطالعات خاک‌شناسی و زیست‌محیطی کمک کند. شایان ذکر است که مدل‌های برآوردگر مذکور بایستی در سایر مناطق و خاک‌های متنوع‌تر ارزیابی و بهینه گردند، آن‌گاه به‌طور گسترده در مطالعات خاک مورد استفاده قرار گیرند. پیشنهاد می‌شود هنگام توسعه مدل‌های برآوردگر در پژوهش‌های آتی، تعداد بیشتری نمونه خاک مورد تجزیه و تحلیل قرار گرفته و در فرایند

مدل‌سازی استفاده گردد. افزون بر این، ویژگی‌های خاکی مورد مطالعه در پژوهش کنونی را می‌توان در مقیاس بزرگ با استفاده از تصاویر ماهواره‌ای یا هوابرد ابرطیفی نقشه کرد.



شکل ۶. نمودار پراکنش ۱:۱ مقادیر پیش‌بینی و اندازه‌گیری شده ویژگی‌های خاکی مورد مطالعه در مجموعه داده‌های واسنجی و اعتبارسنجی

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

## منابع

- بابائیان، ابراهیم و جلالیان، حمیدرضا (۱۳۹۵). برآورد مقدار کربن آلی خاک با استفاده از داده‌های ابرطیفی در گستره VIS-NIR-SWIR. نشریه مدیریت خاک و تولید پایدار، ۶ (۲): ۸۲-۶۵.
- رسولی، نجمه؛ فرپور، محمدهادی؛ خیامیم، فاطمه و رنجبر، حجت‌الله (۱۳۹۷). پیش‌بینی برخی ویژگی‌های خاک به روش طیف‌سنجی مرئی - مادون قرمز نزدیک در منطقه بردسیر کرمان. نشریه پژوهش‌های خاک (علوم خاک و آب)، الف ۳۲ (۴): ۲۳۳-۲۳۱.





طیبه، محبوبه؛ نادری، مهدی؛ محمدی، جهانگرد و حسینجانی‌زاده، مهدیه (۱۳۹۷). مقایسه روش‌های مختلف آماری در برآورد اجزای بافت خاک با استفاده از داده‌های طیفی در محدوده مرئی - فروسرخ نزدیک و کوتاه. نشریه آب و خاک (علوم و صنایع کشاورزی)، ۳۲ (۱): ۷۳-۸۵.

عزیزی، کامران؛ نبی‌اللهی، کمال و داوری، مسعود (۱۳۹۷). ارزیابی قابلیت روش طیف‌سنجی در تخمین برخی ویژگی‌های خاک‌های مبتلا به نمک. مهندسی زراعی (مجله علمی کشاورزی)، ۴۱ (۳): ۱-۱۶.

کریمی، صلاح‌الدین؛ داوری، مسعود؛ بهرامی، حسینعلی؛ بابائیان، ابراهیم و حسینی، سید محمدطاهر (۱۳۹۶). برآورد برخی ویژگی‌های مبنایی خاک توسط طیف‌سنجی مرئی - مادون قرمز نزدیک در استان کردستان. تحقیقات آب و خاک ایران، ۴۸ (۳): ۵۷۳-۵۸۵.

## REFERENCES

- Adeline, K.R.M., Gomez, C., Gorretta, N., & Roger, J.M. (2017). Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*, 288, 143-153. <https://doi.org/10.1016/j.geoderma.2016.11.010>.
- Akpa, S. I. C., Odeh, I. O. A., & Bishop, T. F. A. (2014). Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, 78: 1953-1966. <http://doi.org/10.2136/sssaj.2014.05.0202>.
- Akumu, C. E., Johnson, J. A., Etheridge, D., Uhlig, P., Woods, M., Pitt, D. G., & McMurray, S. (2015). GIS-fuzzy logic based approach in modeling soil texture: Using parts of the Clay Belt and Hornepayne region in Ontario Canada as a case study. *Geoderma*, 239-240: 13-24. <https://doi.org/10.1016/j.geoderma.2014.09.021>.
- Azizi, K., Nabiollahi, K. & Davari, M. 2017. Evaluation of spectroscopic capability in estimating some properties of salt-affected soils. *Agricultural Engineering (Agricultural Scientific Journal)*, 41(3): 1-16. <https://doi.org/10.22055/AGEN.2019.25763.1427>. (In Persian).
- Breiman, L. (1996). Bagging predictors. *Machine Learning*. 24: 2. 123-140. <https://doi.org/10.1007/BF00058655>
- Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils critical review and research perspectives. *Soil Biology and Biochemical*, 43: 1398-1410. <https://doi.org/10.1016/j.soilbio.2011.02.019>.
- Babaeian, A., & Jalali, V. (2015). Estimation of soil organic carbon using hyperspectral data in VIS-NIR-SWIR range. *Journal of Soil Management and Sustainable Production*, 6(2), 65-82. <https://doi.org/10.22069/EJSMS.2016.3143>. (In Persian).
- Chatrenour, M., Landi, A., Bahrami, H.A. & Mirzaei, S. (2023). Dust source clay content and salinity estimation using VNIR spectrometry. *Arid Land Research and Management*, 37(3), 369-388. <https://doi.org/10.1080/15324982.2023.2170837>.
- 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, 480-490. <https://doi.org/10.2136/sssaj2001.652480x>.
- Chen, L., Sheng-lu, Z., Shao-hua, W., Qing, Z., & Qi, D. (2014). Spectral Response of Different Eroded Soils in Subtropical China: A Case Study in Changting County, China. *Journal of Materials Science*, 11: 697-707. <https://doi.org/10.1007/s11629-013-2780-8>.
- Curcio D., Ciraolo G., D'Asaro F., & Minacapillia M. (2013). Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences*, 19: 494 - 503. <https://doi.org/10.1016/j.proenv.2013.06.056>.
- Dotto, A.C., Dalmolin, R.S.D., Grunwald, S., ten Caten, A., & Pereira Filho, W. (2017). Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil Tillage and Research*, 172, 59-68. <https://doi.org/10.1016/j.still.2017.05.008>.
- Esbensen, K.H. (2006). *Multivariate Data Analysis -In practice*. CAMO Software AS. 5th Edition, 589 pages.
- Gee, G.W., & Or, D. (2002). 2.4 Particle-Size Analysis. In: Dane, J.H., Topp, C.G. (Eds.), *Methods of Soil Analysis: Part 4 Physical Methods*. Soil Science Society of America, Madison, WI, pp. 255-293.
- 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: 141-148. <https://doi.org/10.1016/j.geoderma.2008.09.016>.
- Hartemink, A. E., & McBratney, A. B. (2008). A soil science renaissance. *Geoderma*, 148: 123-129. <https://doi.org/10.1016/j.geoderma.2008.10.006>.
- Islam, K., Singh, B., & McBratney, A. (2003). Simultaneous estimation of several soil properties by ultra-



- violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research*. 41:1101-1114. <https://doi.org/10.1071/SR02137>.
- IUSS Working GWRB. (2015). World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports. 106: 1-192
- Jaconi, A., Vos, C., & Don, A. (2019). Near infrared spectroscopy as an easy and precise method to estimate soil texture. *Geoderma*, 337: 906-913. <https://doi.org/10.1016/j.geoderma.2018.10.038>.
- Kodaira, M., & Shibusawa, S. (2013). Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. *Geoderma*, 199: 64-79. <https://doi.org/10.1016/j.geoderma.2012.09.007>.
- 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. <https://doi.org/10.1255/jnirs.1157>.
- Karimi, S.A., Davari, M., Bahrami, H., Babaian, A., & Hosseini, S.M.T. (2017). Derivation and evaluation of spectral transfer function and soil transfer function in order to estimate cation exchange capacity. *Journal of Soil Research (Soil and Water Sciences)*, A, 31(4): 573-585. <https://doi.org/10.22092/IJSR.2018.115957>. (In Persian).
- Lacerda M.P.C., Demattê J.A.M., 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(701):1-20. <https://doi.org/10.3390/rs8090701>.
- Mousavi, F., Abdi, E., Ghalandarzadeh, A., Bahrami, H.A., Majnounian, B., & Ziadi, N. (2020). Diffuse reflectance spectroscopy for rapid estimation of soil Atterberg limits. *Geoderma*, 361, 114083. <https://doi.org/10.1016/j.geoderma.2019.114083>.
- Nocita, M., Stevens, A., Noon, C., & van Wesemael, B. (2013). Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma*, 199: 37-42. <https://doi.org/10.1016/j.geoderma.2012.07.020>.
- Qiu, H. (2010). Studies on the potential ecological risk and homology correlation of heavy metal in the surface soil. *Journal of Agricultural Science*, 2:1916-9760. <https://doi.org/10.5539/jas.v2n2p194>.
- Ramírez, P. B., Calderón, F. J., Jastrow, J. D., Chien-Lu Ping, Ch. & Matamala, R. (2023). Applying NIR and MIR spectroscopy for C and soil property prediction in northern cold-region ecosystems. Which approach works better? *Geoderma Regional*, 32, e00617. <https://doi.org/10.1016/j.geodrs.2023.e00617>.
- Rasouli, N., Farpour, M. H., Khayamim, F., & Ranjbar, H. (2017). Prediction of selected soil properties using visible and near infrared spectroscopy in Bardsir area, Kerman Province. *Iranian Journal of Soil Research*, 32(2): 243-231. <https://doi.org/20.1001.1.22287124.1397.32.2.8.9>. (In Persian).
- Rasooli, N., Farpour, M. J., Mahmoodabadi, M., & Esfandiarpour-Boroujeni, I. (2023). Vis-NIR spectroscopy as an eco-friendly method for monitoring pedoenvironmental variations and pedological assessments in Lut Watershed, Central Iran. *Soil and Tillage Research*, 233: 105808. <https://doi.org/10.1016/j.still.2023.105808>.
- Sparks, D.L., Page, A.L., Helmke, P.A., & Loeppert, R.H. (1996). Methods of Soil Analysis Part 3—Chemical Methods. Soil Science Society of America, American Society of Agronomy, Madison, WI, SSSA Book Series.
- Stenberg, B., Viscarra Rossel, R., Mouazen, A., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107: 163-215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7).
- Savvides, A., Corstanje, R., Baxter, S. J., Rawlins, B. G., & Lark, R. M. (2010). The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent. *Geoderma* 154: 353-358. <https://doi.org/10.1016/j.geoderma.2009.11.007>.
- Shiferaw A., & Hergarten Ch. (2014). Visible near infra-red (VisNIR) spectroscopy for predicting soil organic carbon in Ethiopia. *Journal of Ecology and the Natural Environment*, 6:126-139. <https://doi.org/10.5897/JENE2013.0374>.
- Silva E. B., ten Caten, 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..p. 73-87. In A.E. Hartemink and B. Minasny (eds). Digital Soil Morphometr. Springer International Publishing, Switzerland.
- Tayibi, M., Naderi, M., Mohammadi, J., & Hosseinjanizadeh, M. (2017). Comparison of different statistical methods in estimating soil texture components using spectral data in the visible-near and short-infrared



- range. *Water and Soil Journal (Agricultural Sciences and Technology)*, 32(1): 73-85. <https://10.22067/jsw.v32i1.63618>. (In Persian).
- Viscarrá Rossel, R.A.V. (2008). ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems*, 90, 72–83. <https://doi.org/10.1016/j.chemolab.2007.06.006>.
- Viscarrá Rossel, R., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158: 46-54. <https://doi.org/10.1016/j.geoderma.2009.12.025>.
- Wold S., Sjöström M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- Zhu, A.X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert Knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65: 1463-1472. <https://doi.org/10.2136/sssaj2001.6551463x>.
- Zhao, D., Arshad, M., Li, N., & Triantafyllis, J. (2021). Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. *Catena*, 196: 104938. <https://doi.org/10.1016/j.catena.2020.104938>.
- Zhao, X., Zhao, D., Wang, J., & Triantafyllis, J. (2022). Soil organic carbon (SOC) prediction in Australian sugarcane fields using Vis-NIR spectroscopy with different model setting approaches. *Geoderma Regional*, 30, e00566. <https://doi.org/10.1016/j.geodrs.2022.e00566>.

## Evaluation of PLSR and bagging-PLSR methods in estimating soil texture, pH, and calcium carbonate using spectral data

### EXTENDED ABSTRACT

#### Introduction

The rapid, accurate, and low-cost determination of soil properties has particularly important for land planning and management. However, conventional soil sampling and reliable measurement of soil properties, especially on a large geographic scale, can be a laborious task, time-consuming, expensive, and require quantities of harmful chemicals substance for performing experiments. It is, perhaps, for these reasons that the proximal and remote sensing techniques are being considered as possible alternatives to enhance, complement or substitute traditional soil analysis methods. During the last few decades, the use of visible (Vis) and near-infrared (NIR) diffuse reflectance spectroscopy, as a proximal sensing technique, has attracted tremendous attention for assessing soil properties. Therefore, the objective of this study was to evaluate the Vis-NIR spectral reflectance of soils, as a rapid, cost-effective and non-destructive technique, for estimating some soil properties by PLSR and bagging-PLSR methods

#### Materials and Methods

A total of 220 composite soil samples were collected from the 0–20 cm depth in Ghorveh Plain, Kurdistan province, in September 2019. These soil samples were transported to the laboratory, air dried, grounded, and then sieved to a size fraction of smaller than 2-mm. The selected soil properties including sand, silt, clay, pH, and calcium carbonate equivalent (CCE) were measured by standard laboratory methods. In addition, the proximal spectral reflectance of soil samples was also measured within the 350-2500 nm range (Vis-NIR) using a handheld spectroradiometer. To minimize the impact of random noise and improve calibration models, different pre-processing methods were assessed after recording the spectra.

#### Results

The study found that the first derivative of the Savitzky-Golay smoothing filter was the most effective pre-processing technique for calibrating the PLS regression. The optimal number of factors for predicting sand, silt, clay, pH, and CCE were 14, 7, 23, 10, and 18, respectively, using the PLSR method. The  $R^2$  values for the PLSR method ranged from 0.58 to 0.76, while the bagging-PLSR produced  $R^2$  values between 0.59 and 0.74. The RMSE values obtained for sand, silt, clay, CCE, and pH were 17.43, 7.65, 7.83, 7.94, and 0.66, respectively for the PLSR, and 16.66, 7.63, 8.13, 7.71, and 0.45 for the bagging-PLSR. Based on the ratio of prediction to deviation (RPD) values, the bagging-PLSR model achieved the best performance in predicting sand and CCE. However, for clay and pH prediction, the PLSR model was the most accurate. Both the PLSR and bagging-PLSR models yielded identical predictions for silt content, with an RPD value of 1.53.

#### Conclusion

It is concluded that the bagging-PLSR outperformed PLSR in predicting sand content and CCE. However, PLSR was more effective for predicting clay content and pH. Both models produced similar results for silt content. In all, the accuracy levels of both PLSR and bagging-PLSR were high for clay content and moderate to good for sand and silt contents, pH, and CCE. These findings suggest that the Vis-NIR spectroscopy, as a complement or replacement approach to laboratory conventional methods, can be used for rapid and cost-efficient assessment of soil properties.

**Keywords:** CCE, Soil Properties, Partial Least-Squares Regression, Visible and Near-Infrared Spectroscopy.