

مقایسه روش‌های مختلف آماری در برآورد اجزای بافت خاک با استفاده از داده‌های طیفی در محدوده مرئی - فروسرخ نزدیک و کوتاه

محبوبه طیبی^{۱*} - مهدی نادری^۲ - جهانگرد محمدی^۳ - مهدیه حسینیجانی زاده^۴

تاریخ دریافت: ۱۳۹۶/۰۲/۱۸

تاریخ پذیرش: ۱۳۹۶/۱۱/۰۹

چکیده

استفاده از روش‌های نوین از جمله طیف‌سنجی در محدوده مرئی و فروسرخ نزدیک و فروسرخ کوتاه (۴۰۰-۲۵۰۰ نانومتر) به عنوان یک روش سریع، آسان و کم هزینه در پیش‌بینی ویژگی‌های خاک می‌تواند بسیار موثر باشد. این مطالعه با هدف بررسی توانایی داده‌های طیفی در محدوده مرئی، فروسرخ نزدیک و فروسرخ کوتاه (۴۰۰-۲۵۰۰ نانومتر) در برآورد اندازه ذرات خاک با استفاده از روش‌های رگرسیون حداقل مربعات جزئی (PLSR) و رگرسیون مؤلفه اصلی (PCR) انجام شد. برای این منظور ۱۲۰ نمونه خاک از منطقه کفه مور، استان کرمان برداشته شد. جهت ارزیابی مدل ۸۰ درصد داده‌ها برای کالیبراسیون مدل و ۲۰ درصد برای صحت‌سنجی مدل به صورت تصادفی انتخاب شدند. همچنین جهت اعتبارسنجی از روش حذف هر بار یک نمونه (Leave one out-cross validation) استفاده شد نتایج نشان داد بیشترین مقدار R^2 و کمترین مقدار RMSE برای داده‌های کالیبراسیون و اعتبارسنجی برای لگاریتم پارامترهای رس و شن در روش PLSR همراه با پیش‌پردازش مشتق دوم و برای لگاریتم سیلت در روش PLSR همراه با پیش‌پردازش مشتق اول به دست آمد. با توجه به مقادیر انحراف پیش‌بینی باقیمانده (RPD) پیش‌بینی مدل برای درصد رس و سیلت قابل قبول و برای درصد شن ضعیف می‌باشد. براساس نتایج این مطالعه طیف‌سنجی می‌تواند به عنوان یک روش سریع، آسان و غیرمخرب در برآورد اجزای بافت خاک مورد استفاده قرار گیرد.

واژه‌های کلیدی: رگرسیون حداقل مربعات جزئی (PLSR)، رگرسیون مؤلفه اصلی (PCR)، طیف‌سنجی

نموده‌اند (۱۲).

مقدمه

از آنجا که خاک نیز همانند سایر اجسام موجود در طبیعت براساس ترکیب و ساختار مولکولی خود، بازتاب مشخصی در نواحی مختلف امواج الکترومغناطیس و یا طول موج‌های گوناگون نشان می‌دهد، دارای سیمای طیفی مخصوص به خود است (۳۲). بنابراین می‌توان با کمک طیف‌های بازتابی اطلاعات فیزیکی و شیمیایی خاک را بدست آورد. برخی از ویژگی‌های خاک، مانند رنگ، بافت، ساختمان، کانی‌شناسی، ماده آلی، کربنات‌های آزاد، شوری، رطوبت و اکسید و هیدروکسیدهای آهن و منگنز در انعکاس بازتابی آن اثر می‌گذارند و سبب شکل‌گیری الگوی سیمای طیفی خاصی می‌شوند (۹).

بر اساس بررسی‌های انجام شده امکان استفاده از طیف‌سنجی به عنوان یک روش سریع، آسان، کم‌هزینه و پیشرفته برای پیش‌بینی ویژگی‌های خاک وجود دارد (۱، ۲۲ و ۴۰). در این رابطه گزارش‌هایی وجود دارد که نشان می‌دهد توزیع اندازه ذرات خاک (۱۲ و ۱۵)، مقدار اکسیدهای آهن (۱۹)، کربنات‌ها (۴ و ۲۴)، ماده آلی (۲۶)، مقدار

بافت خاک یکی از ویژگی‌های مهم فیزیکی خاک است که اهمیت بسیاری در مطالعات خاک، آب، محیط زیست و کشاورزی دارد. بافت خاک نقش کلیدی در فرآیندهای تخریب خاک و انتقال توسط آب، کنترل کیفیت خاک و بهره‌وری آن دارد بنابراین به عنوان یکی از پارامترهای مهم محیطی خاک شناخته می‌شود (۱۶). تعیین بافت خاک و تهیه نقشه‌های آن در مقیاس‌های مختلف و سطوح وسیع نیازمند حجم بالای نمونه‌برداری، صرف وقت و هزینه زیادی می‌باشد. از این رو محققین سعی در توسعه و استفاده از ابزارها و فن‌آوری‌های نوین علمی از جمله طیف‌سنجی در تعیین و برآورد بافت خاک

۱، ۲ و ۳- به ترتیب دانش آموخته دکتری، دانشیار و استاد گروه علوم خاک، دانشکده کشاورزی، دانشگاه شهرکرد

۴- استادیار گروه اکولوژی، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان

(Email: mahtayebi@yahoo.com

*) نویسنده مسئول:

درصد رس، سیلت و شن خاک با استفاده از روش‌های PLSR و PCR و (۲) ارزیابی انواع روش‌های پیش‌پردازش داده‌های طیفی در مدل‌سازی انجام شد.

مواد و روش‌ها

منطقه مورد پژوهش و روش نمونه‌برداری

منطقه کفه مور با مساحت ۲۹۵۳۳ هکتار در ۶۰ کیلومتری جنوب غرب سیرجان در مسیر جاده شیراز و بین طول‌های جغرافیایی ۱۵° تا ۵۵° تا ۲۵° و عرض‌های جغرافیایی ۵۱° تا ۱۱° قرار گرفته است. به طوری که ناحیه معدنی سنگ آهن گل‌گهر در شمال آن واقع شده است (شکل ۱). معدن سنگ آهن گل‌گهر در حال حاضر یکی از بزرگترین معادن روباز در سطح کشور بوده و به طور گسترده و وسیع مواد معدنی آن از عمقی بسیار پایین‌تر از تراز سطح آب زیرزمینی منطقه، استخراج می‌گردد. در این ناحیه با توجه به معدن‌کاری‌های گسترده، تخلیه‌ی پساب‌های معادن به درون حوضچه‌ها و سدهای رسوب‌گیر، احتمال آلودگی خاک‌های مجاور، تغییر در ویژگی‌های فیزیکی و شیمیایی آن‌ها و در نتیجه تخریب خاک به طور گسترده وجود دارد (۱۸). میانگین بارندگی سالانه ۱۷۲ میلی‌متر، حداقل درجه حرارت ۱۲۰°C- و حداکثر آن ۴۲°C می‌باشد. رژیم حرارتی منطقه ترمیک و رژیم رطوبتی آن اریدیک است. این ناحیه در لبه شمال شرقی زون سنج - سیرجان قرار گرفته است.

در این پژوهش محل نقاط نمونه‌برداری به روش هایپرکیوب لاتین (Latin Hypercube) با استفاده از نقشه‌های زمین‌شناسی، کاربری اراضی و مدل رقومی ارتفاع (DEM) در محیط نرم افزار R مشخص گردید (۳۸). ۱۲۰ نمونه خاک به صورت مرکب و از اختلاط ۴ نمونه خاک از عمق صفر تا ۱۰ سانتی‌متری از سطح خاک برداشته شد. به منظور تعیین بافت خاک و همچنین طیف‌سنجی، ابتدا نمونه‌های خاک هوا خشک و از الک ۲ میلی‌متری عبور داده شدند. برای تعیین بافت خاک از روش هیدرومتر (۳۴) استفاده شد. در این روش پس از زودده شدن مواد سیمان‌کننده و پراکنش دانه‌های خاک به کمک محلول کالگون و همزن برقی، با قرائت عدد هیدرومتر پس از ۴۰ ثانیه و ۲ ساعت درصد رس، سیلت و شن خاک محاسبه شد و به وسیله مثلث بافت خاک، کلاس بافت خاک تعیین شد.

برای اندازه‌گیری بازتاب طیفی خاک‌های مورد نظر در محدوده مرئی و فرورسرخ نزدیک و کوتاه (۳۵۰-۲۵۰۰ نانومتر) در آزمایشگاه، از دستگاه رادیواسپکترومتر (Analytical Spectral Devices, ASDFieldspec[®]3) استفاده شد. به گونه‌ای که حدود ۱۰۰ گرم از هر نمونه (با ضخامت حدود ۴ cm و قطر حدود ۶ cm) برای آنالیز مورد استفاده قرار گرفت. رادیواسپکترومتر قبل از اولین استفاده و در فواصل منظم در طول قرائت (به ازای هر ۵ نمونه) با استفاده از

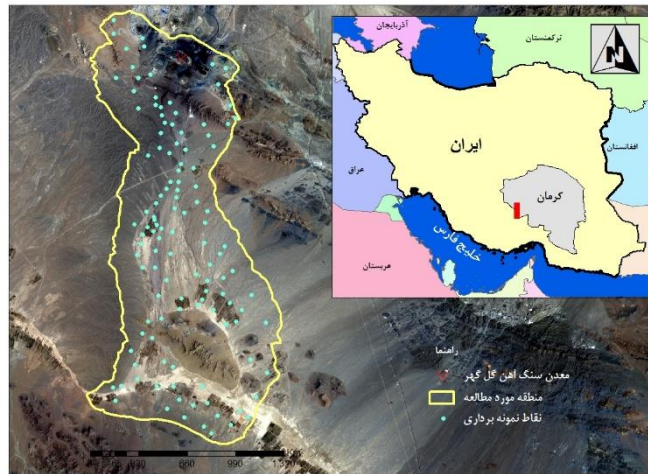
pH (۴۶)، رطوبت خاک (۱۳ و ۲۱) با استفاده از طیف‌سنجی در محدوده مرئی و فرورسرخ نزدیک و فرورسرخ کوتاه (۴۰۰-۲۵۰۰ نانومتر) برآورد شده است.

برای بررسی ارتباط بین پارامترهای طیفی ثبت شده خاک‌ها به وسیله اسپکترومتر^۱ و خصوصیات فیزیکوشیمیایی خاک‌ها از مدل‌های مختلفی استفاده می‌شود (۴۰) که از آن جمله می‌توان رگرسیون حداقل مربعات جزئی (Partial Least Squares Regression, PLSR) (۴۹) و رگرسیون مؤلفه اصلی (Principal Component Regression, PCR) (۳۱) را نام برد. کیوئی (۳۷) از بازتاب طیف مادون قرمز میانی و روش PLSR جهت تخمین تغییرات کربن آلی، کربن غیر آلی، CEC، pH و رس ۱۹۸ نمونه خاک مدیترانه‌ای ایتالیا (جزایر غرب مدیترانه) استفاده نمودند. نتایج نشان دادند که با استفاده از طیف‌سنجی و روش PLSR امکان تفکیک خصوصیات مهم این نمونه‌ها با مواد مادری متفاوت وجود دارد. سامرز و همکاران (۴۵) از روش PLSR و میگزبان انعکاس در محدوده VNIR-SWIR برای آزمون همبستگی بین کانی‌شناسی خاک، رس، کربن آلی استفاده کردند و گزارش نمودند که مقادیر R^۲ برای تمامی پارامترها بیش از ۰/۵۰ و مقادیر درصد انحراف نسبی (Relative percent deviation, RPD) قابل قبول بوده است. شیفر و هرگاتن (۴۲) داده‌های طیفی در محدوده VIS/NIR و روش‌های PLSR و PCR را برای تخمین کربن آلی خاک در ایتوپوی به کار بردند. نتایج نشان داد مقادیر R^۲ و RPD مربوط به مدل PLSR به ترتیب ۰/۷، ۰/۶ و ۱/۵ و مقادیر مربوط به مدل PCR به ترتیب ۰/۷، ۰/۵ و ۱/۲ بودند و روش PLSR نسبت به روش PCR برتری داشته است.

از آنجا که عوامل متعددی از جمله پخشیدگی نور، غیرخطی بودن و بسیاری عوامل محیطی می‌توانند میزان توانایی مدل‌های ایجاد شده با داده‌های طیفی را تحت تأثیر قرار دهند (۳۹ و ۵۰) بنابراین در بیشتر مطالعات از روش‌های پیش‌پردازش مشتق اول (First Derivative, FD) و مشتق دوم (Second Derivative, SD) (۱۴) و تصحیح پخشیده‌چندگانه (Multiplicative Scatter Correction, MSC) (۲۹ و ۳۹) و متغیر نرمال استاندارد (Standard Normal Variate, SNV) (۲) به عنوان تکنیک‌های جداناپذیر در مدلسازی طیفی استفاده می‌شود. با توجه به اینکه تاکنون مطالعه کمی در خصوص برآورد اجزای بافت خاک به کمک داده‌های طیف‌سنجی و مدل‌های PLSR و PCR و همچنین مقایسه‌ی انواع تکنیک‌های پیش‌پردازش در کشور صورت نگرفته است، این پژوهش با هدف (۱) بررسی توانایی داده‌های طیفی در محدوده مرئی و فرورسرخ نزدیک و فرورسرخ کوتاه (۴۰۰-۲۵۰۰ نانومتر) در برآورد

شد و میانگین آن‌ها با استفاده از نرم افزار RS^۳ محاسبه و در تجزیه و تحلیل‌های طیفی در نظر گرفته شد.

صفحه سفید مبنا (White panel Spectralon) واسنجی شد. برای هر نمونه خاک ۲۰ اسکن به منظور کاهش میزان نویز در نظر گرفته



شکل ۱- موقعیت جغرافیایی منطقه مورد مطالعه در کشور و استان کرمان.
Figure 1- Location of the Kafemoor basin in Kerman province, Iran

است که مؤلفه‌های اصلی تنها با مقادیر متغیرهای X تعیین می‌شود، در صورتی که در روش PLSR، مقادیر هر دو متغیرهای X و Y در ساخت مؤلفه‌ها تأثیر می‌گذارند (۷).

برای ارزیابی مدل‌های ایجاد شده توسط دو مدل ۸۰ درصد داده‌ها (تعداد ۹۶ نمونه) برای کالیبراسیون مدل و ۲۰ درصد (تعداد ۲۴ نمونه) برای صحت‌سنجی مدل به صورت تصادفی انتخاب شدند. همچنین جهت اعتبارسنجی مدل برآزش داده شده در گروه کالیبراسیون از روش اعتبارسنجی هر بار یک نمونه (Leave one out-cross validation) استفاده شد.

با استفاده از آماره‌های ضریب تبیین (R^2) (رابطه ۱)، ریشه میانگین مربعات خطا (RMSE) (رابطه ۲) و انحراف پیش‌بینی باقیمانده (Residual Prediction Deviation, RPD) (رابطه ۳) دقت و صحت مدل‌ها مورد ارزیابی قرار گرفتند (۳۳).

$$R^2 = \frac{\left[\sum_{i=1}^n (O_i - \bar{O}_i)(P_i - \bar{P}_i) \right]^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2} \quad \text{رابطه (۱)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad \text{رابطه (۲)}$$

$$RPD = \frac{SD}{RMSE} \quad \text{رابطه (۳)}$$

طول موج‌های بین ۳۵۰ تا ۴۰۰ نانومتر و ۲۴۵۰ تا ۲۵۰۰ که دارای آشفتگی‌های زیاد بودند حذف گردیدند و به منظور حذف هر گونه اطلاعات اضافی، انواع روش‌های پیش‌پردازش از جمله FD^۱ و SD^۲ به همراه فیلتر ساویتزکی-گلای (Savitzky-Golay)، MSC^۳ و SNV^۴ بر روی داده‌های طیفی به کمک نرم‌افزار 10.3 Unscambler انجام شد. بررسی توزیع آماری داده‌ها به وسیله آزمون کولموگروف-اسمیرنوف توسط نرم‌افزار ۹ SAS انجام گردید. ابتدا با تبدیل داده‌ها به لگاریتم نرمال‌سازی داده‌ها انجام، سپس از روش‌های PLSR و PCR برای برآورد درصد رس، سیلت و شن خاک استفاده شد. در روش PLSR، معادلات ساختاری مبتنی بر مؤلفه‌های اصلی (PCA) می‌باشند با این تفاوت که در روش رگرسیون PLSR متغیرهای مستقل X نیز هم زمان با متغیرهای وابسته Y به تعدادی مؤلفه اصلی تبدیل می‌شود به نحوی که بیشترین درصد تغییرات، در مؤلفه‌های اولیه ارائه می‌شود (۳۶). در واقع هدف اصلی رگرسیون PLSR ساختن یک مدل خطی براساس رابطه $Y=AX+B$ است که در آن Y متغیر برآورد شده، X متغیر پیش‌بینی کننده می‌باشد. ضمناً A ماتریسی شامل ضرایب رگرسیون و B ماتریسی از خطاهای باقیمانده است (۵ و ۱۰). رگرسیون مؤلفه اصلی ترکیبی از رگرسیون خطی چندگانه (MLR) و تجزیه به مؤلفه‌های اصلی (PCA) است. تفاوت عمده PLSR با PCR این

- 1- First Derivative
- 2- Second Derivative
- 3- Multiplicative Scatter Correction
- 4- Standard Normal Variate

P_i و O_i به ترتیب مقادیر مشاهداتی و تخمین زده شده، SD، انحراف معیار مقادیر مشاهداتی و n تعداد داده‌ها می‌باشد.

نتایج و بحث

جدول (۱) خلاصه آماری برخی خصوصیات فیزیکی و شیمیایی خاک‌های منطقه مورد مطالعه را نشان می‌دهد. نتایج نشان می‌دهد میزان رس بین ۴/۸۸ تا ۳۲/۵۲ درصد، میزان سیلت بین ۲ تا ۷۶ درصد و میزان شن بین ۳/۴۸ تا ۹۱/۸۴ درصد متغیر است. براساس توزیع کلاس‌های بافتی، خاک‌های منطقه در ۷ کلاس بافتی خاک شامل شن، لوم شنی، لوم رسی شنی، لوم، سیلت لوم، شنی لومی و لوم رسی سیلتی قرار می‌گیرند. چگالی ظاهری خاک از ۱/۱۵ تا ۲/۲۱ مگا گرم بر مترمکعب و چگالی حقیقی خاک از ۲/۰۵ تا ۲/۹۲ مگا

جدول ۱- خلاصه آماری خصوصیات فیزیکی و شیمیایی خاک

Table 1- Summary statistics of soil physical and chemical properties

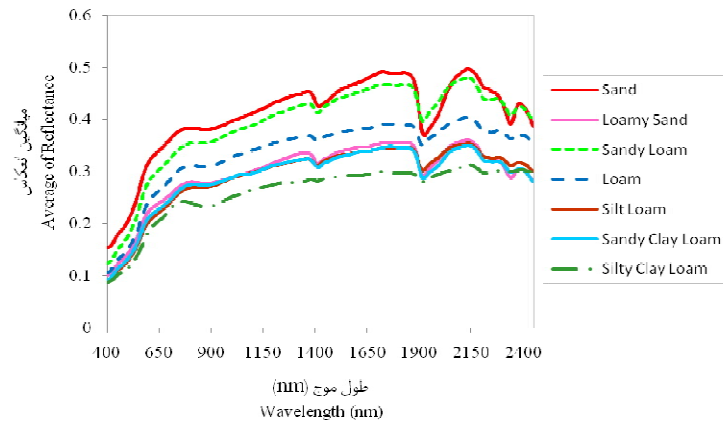
متغیر	واحد	حداقل	حداکثر	میانگین	انحراف	ضریب
رس	%	4.88	32.52	12.01	4.65	38.71
سیلت	%	2	76	29.57	17.09	57.79
شن	%	3.48	91.84	58.42	18.91	32.37
چگالی ظاهری	Mgm ⁻³	1.15	2.21	1.66	0.14	8.35
چگالی حقیقی	Mgm ⁻³	2.05	2.92	2.55	0.12	4.7
تخلخل	%	0.21	0.44	0.35	0.03	7.35
Fe	(mg kg ⁻¹)	4436.25	271375	24040	45119	187.68
O.M	(%)	0.03	0.69	0.24	0.14	58.33
pH	-	7.03	8.87	7.74	0.57	7.40
EC	(ds m ⁻¹)	0.36	295.06	32.14	61.03	189.88

می‌توان به وجود مقادیر زیاد آهن در نمونه‌های خاک و افزایش تیرگی رنگ خاک نسبت داد. از آنجا که هر چه اندازه ذرات خاک کوچکتر باشند به دلیل سطح ویژه بالاتر و همچنین وجود کانی‌های رسی و اکسیدهای آهن، منگنز و آلومینیوم دارای توانایی بالاتری در نگهداری فلزات سنگین به خصوص آهن می‌باشند (۶ و ۲۸). به طور کلی طیف خاک در محدوده VNIR/SWIR تحت تاثیر مجموعه‌ای از عوامل مختلف از جمله بافت، رنگ، ساختمان، کانی‌شناسی، ماده آلی، کربنات‌های آزاد، شوری، رطوبت و اکسید و هیدروکسیدهای آهن و منگنز قرار می‌گیرد (۸).

لوکاس و گالگی (۲۷) رنگ خاک را یکی از عوامل مهم در تغییر بازتابش در خاک دانستند. جونز و واگان (۲۰) گزارش کردند که سایه‌ی ذرات درشت، زبری سطحی و رطوبت بالا عوامل تاثیرگذار بر میزان بازتابش خاک هستند. شکل (۳) طیف انعکاسی نمونه‌های خاک بعد از اعمال پیش‌پردازش‌های FD و SD به همراه فیلتر ساویتزکی و گلای، MSC و SNV را نشان می‌دهند.

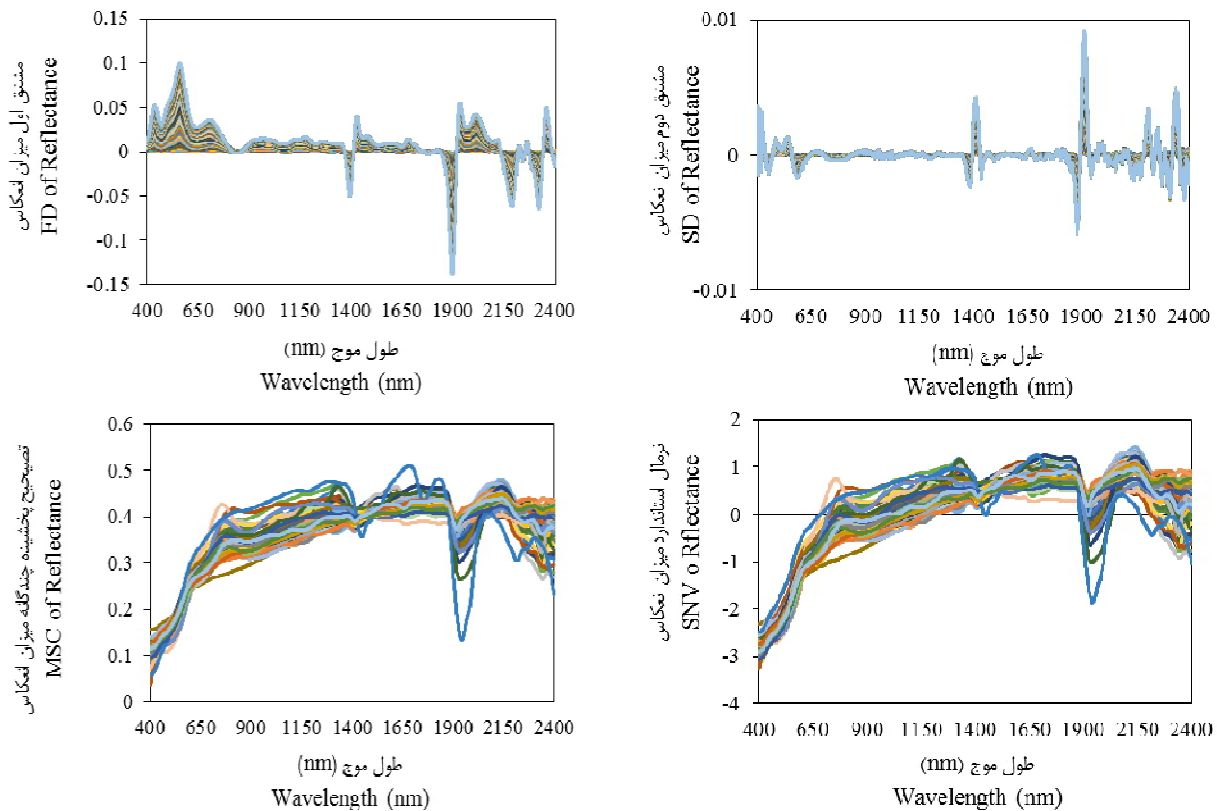
ویژگی‌های طیفی خاک

در شکل ۲ منحنی‌های طیفی میانگین ۷ کلاس بافت خاک آورده شده است. بر اساس تحقیقات هانت و اشلی (۱۷) تمام خاک‌ها دارای سیمای طیفی تقریباً مشابه‌ای هستند و ویژگی‌های جذب آب در ۱۴۰۰ و ۱۹۰۰ نانومتر و ویژگی‌های جذب آب در ۲۳۰۰ نانومتر را نشان می‌دهند. مقایسه منحنی‌های طیفی در کلاس‌های مختلف بافتی نشان داد که بافت شن، شنی لومی و لوم دارای بیشترین بازتابش و کلاس لوم رسی سیلتی کمترین بازتابش را دارد. هر چند بر طبق نتایج محققین مختلف مشخص شده است که هر چه اندازه ذرات کوچکتر باشد خاک نرم‌تر و میزان بازتابش بیشتر است. نتایج معکوس ما و افزایش شدت بازتابش در بافت‌های درشت‌تر را می‌توان به وجود مقادیر بیشتر شن و در نتیجه ترکیبات کوارتز که از مهمترین فاکتورهای شدت بازتابش است، نسبت داد (۳ و ۴۴). در بافت لوم رسی سیلتی کاهش میزان بازتابش به دلیل وجود رس و فیلسیلیکات‌ها است. علت این نتایج در خاک‌های منطقه را نیز



شکل ۲- منحنی طیفی خاک‌ها در محدوده مرئی-فروسرخ نزدیک و کوتاه براساس میانگین انعکاس در ۷ کلاس بافتی

Figure 2- Reflectance spectra of soil samples in the visible- near-infrared and shortwave infrared regions based on average of reflectance for 7 soil texture classes



شکل ۳- طیف انعکاسی ۱۲۰ نمونه خاک بعد از اعمال پیش پردازش‌های مشتق اول (FD)، مشتق دوم (SD)، تصحیح پخشیده چندگانه (MSC) و متغیر نرمال استاندارد (SNV) در محدوده مرئی- فروسرخ نزدیک و کوتاه (۴۰۰-۲۴۵۰ نانومتر).

Figure 3- Reflection spectrum of 120 soil samples after applying First (FD) and Second Derivative of reflectance (SD), multiplicative scatter correction (MSC) and standard normal variate (SNV) pre-processing in the visible- near-infrared and shortwave infrared regions

برآورد ذرات رس، سیلت و شن با استفاده از داده‌های طیفی در محدوده مرئی و مادون قرمز نزدیک

جدول‌های ۲ تا ۴ مقادیر R^2 ، RMSE و RPD حاصل از مدل‌سازی با استفاده از روش‌های PLSR و PCR و انواع روش‌های پیش‌پردازش در برآورد ذرات رس، سیلت و شن را نشان می‌دهند. در روش PLSR نتایج برآورد لگاریتم درصد رس نشان داد که مقادیر R^2 از ۰/۲۲ تا ۰/۷۲ در گروه کالیبراسیون و از ۰/۱۲ تا ۰/۵۴ در گروه اعتبارسنجی متغیر است. مقدار RMSE در دو گروه کالیبراسیون و اعتبارسنجی به ترتیب از ۰/۰۸ تا ۰/۱۳ و در گروه اعتبارسنجی از ۰/۱۰ تا ۰/۱۴ متغیر است. در برآورد لگاریتم درصد رس با استفاده از روش PLSR، بیشترین مقدار R^2 و کمترین مقدار RMSE با پیش‌پردازش SD و کمترین مقدار R^2 و بیشترین مقدار RMSE بدون استفاده از پیش‌پردازش در هر دو گروه کالیبراسیون و اعتبارسنجی حاصل شد (جدول ۲). شاخص RPD از دیگر معیارهای سنجش مدل در مطالعات طیف‌سنجی می‌باشد که به ۶ کلاس خیلی ضعیف ($RPD < 1$)، ضعیف ($1-1/4$)، متوسط ($1/4-1/8$)، خوب

($1/8-2$)، خیلی خوب ($2-2/5$) و عالی ($> 2/5$) طبقه‌بندی می‌شوند (۷ و ۳۳). مقادیر RPD برای داده‌های کالیبراسیون (RPD_{CV}) برابر با ۱/۹۲ و برای داده‌های اعتبارسنجی (RPD_{VAL}) برابر با ۱/۵۴ محاسبه شد. RPD_{CV} و RPD_{VAL} براساس طبقه‌بندی چانگ و همکاران (۸) و موازن و همکاران (۳۵) در کلاس خوب قرار می‌گیرند. در روش PCR تنها پیش‌پردازش SD قادر به ایجاد ارتباط بین داده‌های طیفی و لگاریتم درصد رس بود هرچند که در کلاس ضعیف قرار می‌گیرد. سایر روش‌های پیش‌پردازش نیز نتایج خوبی را ارائه ندادند (جدول ۲).

براساس آزمون عدم قطعیت مارتینز طیف‌های مهم در مدل‌سازی براساس روش‌های اعتبارسنجی تلاقی (cross validation)، جک نایف و نقاط پایدار، مشخص شدند (۷ و ۳۰). براساس این روش ۶۵ طیف به عنوان طیف‌های مهم در تعیین لگاریتم درصد رس هستند که مهمترین آن‌ها براساس بیشترین ضریب رگرسیونی در روش PLSR با پیش‌پردازش SD برابر با ۲۰۸۷ نانومتر است.

جدول ۲- نتایج آماری روش‌های PLSR و PCR و روش‌های مختلف پیش‌پردازش طیفی در تعیین لگاریتم رس (%)

Table 2- Statistics results of PLSR and PCR methods with different pre-processing techniques for estimating of LogClay (%)

مدل	پیش‌پردازش	R^2_{CV}	RMSE _{CV}	R^2_{VAL}	RMSE _{VAL}	RPD _{CV}	RPD _{VAL}
PLSR	No	0.22	0.13	0.12	0.14	1.18	1.10
	FD	0.43	0.11	0.22	0.11	1.39	1.39
	SD	0.72	0.08	0.54	0.10	1.92	1.54
	MSC	0.27	0.13	0.16	0.14	1.19	1.10
	SNV	0.28	0.12	0.15	0.14	1.28	1.10
PCR	No	-	-	-	-	-	-
	FD	-	-	-	-	-	-
	SD	0.21	0.13	0.12	0.14	1.20	1.12
	MSC	-	-	-	-	-	-
	SNV	-	-	-	-	-	-

لگاریتم درصد سیلت هستند که مهمترین آن‌ها براساس بیشترین ضریب رگرسیونی در روش PLSR با پیش‌پردازش FD برابر با ۲۳۵۷ نانومتر است.

در برآورد لگاریتم درصد شن با استفاده از PLSR، مقدار R^2 در دو گروه کالیبراسیون و اعتبارسنجی به ترتیب از ۰/۱۱ تا ۰/۵۶ و ۰/۰۸ تا ۰/۴۷ برای بدون پیش‌پردازش و پیش‌پردازش SD محاسبه شد. کمترین مقدار RMSE در پیش‌پردازش SD و بیشترین مقدار RMSE برای بدون پیش‌پردازش در دو گروه کالیبراسیون و اعتبارسنجی بدست آمد. بیشترین شاخص RPD در هر دو گروه کالیبراسیون و اعتبارسنجی در پیش‌پردازش SD بدست آمد که نشان دهنده کارایی خوب در گروه کالیبراسیون و کارایی ضعیف در گروه اعتبارسنجی براساس طبقه‌بندی چانگ و همکاران (۸) و موازن و

در برآورد لگاریتم درصد سیلت با استفاده از روش PLSR، مقدار R^2 در گروه کالیبراسیون و اعتبارسنجی به ترتیب از ۰/۳۴ تا ۰/۷۳ و ۰/۲۷ تا ۰/۵۸ متغیر است. کمترین مقدار RMSE و بیشترین مقدار شاخص RPD برای روش FD در دو گروه کالیبراسیون و اعتبارسنجی بدست آمد (جدول ۳). مقادیر RPD_{CV} برابر با ۲/۰۷ و برای RPD_{VAL} برابر با ۱/۵۲ محاسبه شد. RPD_{CV} و RPD_{VAL} براساس طبقه‌بندی چانگ و همکاران (۸) و موازن و همکاران (۳۵) به ترتیب در کلاس خیلی خوب و خوب قرار می‌گیرند. در روش PCR پیش‌پردازش MSC، بیشترین مقادیر R^2 و کمترین مقادیر RMSE را برای داده‌های کالیبراسیون و اعتبارسنجی نشان داد که کارایی مدل در هر دو حالت ضعیف است (جدول ۳). براساس آزمون عدم قطعیت مارتینز (۷ و ۳۰) طیف به عنوان طیف‌های مهم در تعیین

همکاران (۳۵) است (جدول ۴). در روش PCR و پیش‌پردازش MSC، بیشترین مقادیر R^2 و کمترین مقادیر RMSE برای داده‌های کالیبراسیون و اعتبارسنجی حاصل شد که کارایی مدل در هر دو حالت ضعیف است (جدول ۴).

جدول ۳- نتایج آماری روش‌های PLSR و PCR و روش‌های مختلف پیش‌پردازش طیفی در تعیین لگاریتم سیلت (%)

Table 3- Statistics results of PLSR and PCR methods with different pre-processing techniques for estimating of LogSilt (%)

مدل	پیش‌پردازش	R^2_{CV}	RMSE _{CV}	R^2_{VAL}	RMSE _{VAL}	RPD _{CV}	RDP _{VAL}
PLSR	No	0.34	0.23	0.27	0.24	1.16	1.24
	FD	0.73	0.14	0.54	0.19	2.07	1.52
	SD	0.70	0.15	0.58	0.18	1.93	1.61
	MSC	0.49	0.20	0.33	0.24	1.45	1.21
	SNV	0.46	0.21	0.35	0.23	1.38	1.26
PCR	No	0.37	0.22	0.25	0.25	1.31	1.16
	FD	0.34	0.23	0.24	0.25	1.20	1.10
	SD	0.27	0.24	0.16	0.26	1.17	1.08
	MSC	0.47	0.20	0.37	0.23	1.31	1.25
	SNV	0.40	0.22	0.35	0.23	1.30	1.24

جدول ۴- نتایج آماری روش‌های PLSR و PCR و روش‌های مختلف پیش‌پردازش طیفی در تعیین لگاریتم شن (%)

Table 4- Statistics results of PLSR and PCR methods with different pre-processing techniques for estimating of LogSand (%)

مدل	پیش‌پردازش	R^2_{CV}	RMSE _{CV}	R^2_{VAL}	RMSE _{VAL}	RPD _{CV}	RDP _{VAL}
PLSR	No	0.11	0.20	0.08	0.21	1.11	1.06
	FD	0.13	0.20	0.09	0.21	1.12	1.06
	SD	0.56	0.14	0.47	0.16	1.59	1.39
	MSC	0.22	0.19	0.12	0.21	1.17	1.06
	SNV	0.21	0.19	0.13	0.20	1.18	1.12
PCR	No	0.11	0.21	0.07	0.21	1.06	1.06
	FD	0.15	0.20	0.11	0.21	1.12	1.07
	SD	0.24	0.19	0.17	0.20	1.18	1.12
	MSC	0.29	0.18	0.13	0.21	1.25	1.07
	SNV	0.13	0.18	0.07	0.19	1.11	1.05

R^2 و ریشه مربعات خطا (RMSEP) برای ۲۰ درصد داده‌ها که از قبل جدا شده بودند، تعیین گردید (جدول ۵). برای تعیین لگاریتم درصد رس و شن روش PLSR همراه با پیش‌پردازش SD و برای تعیین لگاریتم درصد سیلت روش PLSR همراه با پیش‌پردازش مشتق اول به عنوان روش‌های مناسب‌تری انتخاب شدند. نتایج صحت سنجی نشان داد مقادیر R^2_{PRE} برای لگاریتم درصد رس، سیلت و شن به ترتیب برابر با ۰/۶۲، ۰/۸۲ و ۰/۲۵ و مقادیر RMSEP به ترتیب برابر با ۰/۰۸، ۰/۱۱ و ۰/۱۴ بودند.

براساس آزمون عدم قطعیت مارتینز (۷ و ۳۰) ۱۰۶ طیف به عنوان طیف‌های قابل قبول برای تعیین درصد شن بودند که مناسب ترین آن‌ها براساس بیشترین ضریب رگرسیونی در روش PLSR با پیش‌پردازش SD برابر با ۲۳۵۷ نانومتر بود.

ارزیابی صحت داده‌های طیفی در برآورد ذرات رس، سیلت و شن

برای صحت‌سنجی مدل‌های ایجاد شده توسط روش‌های PLSR و PCR با روش‌های مختلف پیش‌پردازش مقادیر ضریب تبیین (R^2_{PRE})

جدول ۵- نتایج صحت سنجی مدل‌سازی برای لگاریتم رس، سیلت و شن (%)

Table 5- Results of accuracy for modeling of LogClay, LogSilt and LogSand (%)

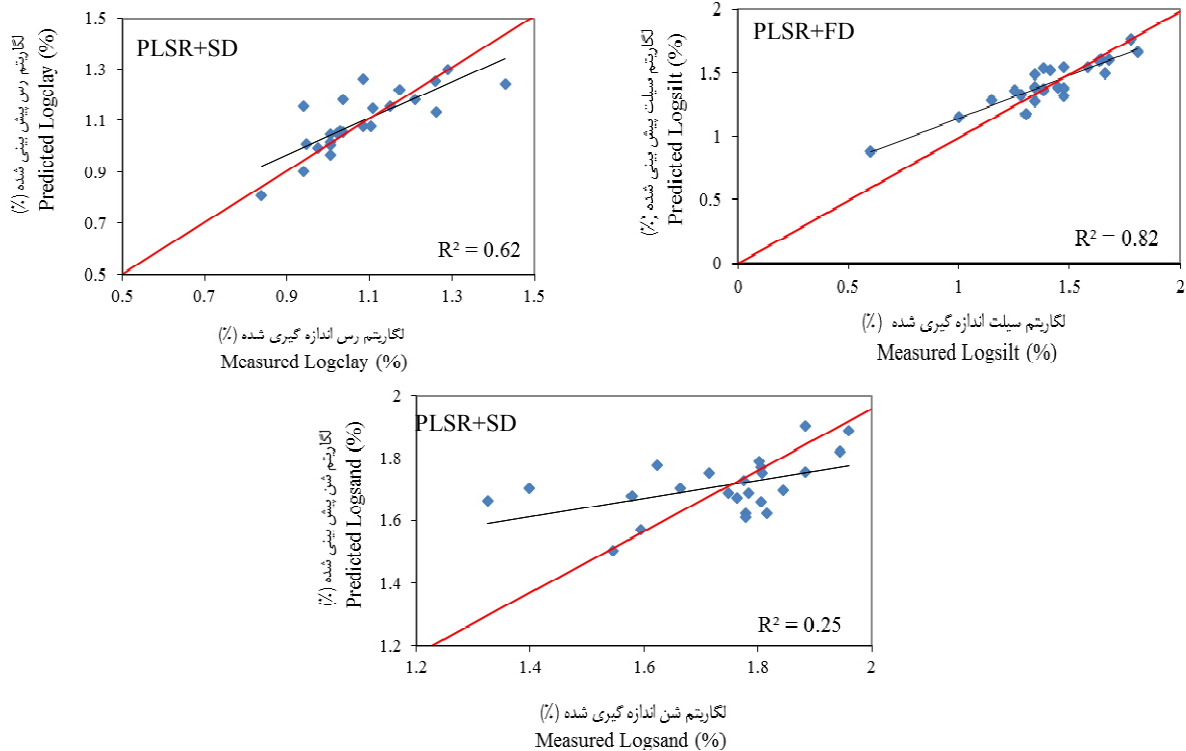
متغیر	واحد	مدل	پیش‌پردازش	R^2_{PRE}	RMSEP
لگاریتم رس	%	PLSR	SD	0.62	0.08
لگاریتم سیلت	%	PLSR	FD	0.82	0.11
لگاریتم شن	%	PLSR	SD	0.25	0.14

درصد رس، سیلت و شن را نشان می‌دهد. در مورد صحت معادلات،

شکل (۵) رابطه مقادیر اندازه‌گیری شده و پیش‌بینی شده لگاریتم

estimation) و در درصد‌های بالاتر از ۱۸ درصد میزان سیلت اندازه‌گیری شده و برآورد شده نزدیک خط ۱:۱ قرار دارند که نشانه دقت بیشتر در پیش‌بینی مقادیر بالاتر سیلت است. در مورد شن در مقادیر مختلف هم کم برآوردی و هم بیش برآوردی وجود دارد که نشانه دقت کم در پیش‌بینی میزان شن خاک است.

شکل (۵) نشان می‌دهد که تقریباً برای ذرات رس پایین‌تر از ۱۹/۵ درصد ($\text{LogClay}=1/29$) میزان رس اندازه‌گیری شده و برآورد شده نزدیک خط ۱:۱ قرار دارند و معادله توانسته است تخمین اندازه‌گیری شده را با دقت بالایی انجام دهد. در مورد سیلت برای درصد‌های پایین‌تر از ۱۸ درصد ($\text{LogSilt}=1/25$) معادله بیش برآوردی (Over



شکل ۵- رابطه مقادیر اندازه‌گیری شده و پیش‌بینی شده لگاریتم رس، سیلت و شن (%)
 Figure 5- Measured vs. Predicted LogClay, LogSilt and LogSand (%)

به عنوان روشی کارآمد در مطالعات زمین‌آماري و ایجاد نقشه‌های شن، رس و سیلت خاک‌های جنگلی در جنوب ایتالیا دانستند. سیلوا و همکاران (۴۳) از روش PLSR و داده‌های طیفی ناحیه ۱۰۷۵-۳۲۵ نانومتر برای تعیین اجزای بافت خاک استفاده نمودند. نتایج نشان داد که این روش برای رس و سیلت نتایج قابل قبولی دارد در حالی که برای شن نتایج ضعیف است. لاسردا و همکاران (۲۳) از روش PLSR برای پیش‌بینی بافت خاک استفاده نمودند. نتایج نشان داد مقدار R^2 برای شن و رس به ترتیب برابر با ۰/۹۶ و ۰/۹۳ است و پیش‌بینی بافت خاک با استفاده از طیف‌سنجی ۱/۲ درصد با مقادیر واقعی اختلاف دارد. آن‌ها اظهار داشتند این روش می‌تواند به توسعه برنامه‌های کامپیوتری در ارتباط با کلاس‌بندی بافت خاک و نقشه برداری دیجیتالی خاک در مقیاس دقیق کمک کند.

همچنین مقایسه روش‌های مختلف پیش‌پردازش برای رس، سیلت و شن نشان داد که دقت مدل‌سازی با به کارگیری انواع

به طور کلی براساس نتایج به دست آمده، روش PLSR در مقایسه با روش PCR دارای کارایی بالاتری است و بهترین نتایج مدل‌سازی با پیش‌پردازش SD برای برآورد رس و شن و با پیش‌پردازش FD برای برآورد سیلت به دست آمد. دیگر پژوهشگران نیز نتایج مشابهی مبنی بر نتایج بهتر روش PLSR در پیش‌بینی ویژگی‌های خاک با استفاده از داده‌های طیفی در محدوده مرئی و فرورسرخ نزدیک و فرورسرخ کوتاه (۴۰۰-۲۵۰۰ نانومتر) بدست آوردند (۲۵، ۴۲، ۴۷). برای مثال کیورسیو و همکاران (۱۲) از روش PLSR و داده‌های طیف‌سنجی برای تعیین بافت خاک استفاده نمودند. نتایج آن‌ها نشان داد مقدار R^2 برای رس، سیلت و شن به ترتیب برابر با ۰/۸۷، ۰/۸۰ و ۰/۶۰ و مقدار RMSE به ترتیب برابر با ۵/۸، ۷/۷ و ۷/۲ درصد است. دیوا و همکاران (۱۳) نیز از روش طیف‌سنجی برای تعیین بافت خاک استفاده نمودند. کانفورتی و همکاران (۱۱) استفاده از داده‌های طیف‌سنجی در دامنه VNIR/SWIR و روش PLSR را

شن ضعیف بود ولی با داشتن مقادیر رس و سیلت به راحتی می‌توان مقادیر شن خاک را محاسبه نمود. براساس نتایج به دست آمده، روش PLSR در مقایسه با روش PCR دارای کارایی بالاتری است و بهترین نتایج مدل‌سازی با پیش‌پردازش SD برای برآورد رس و شن و با پیش‌پردازش FD برای برآورد سیلت به دست آمد. استفاده از روش‌های مختلف پیش‌پردازش نتایج پیش‌بینی درصد اجزای بافت خاک را بهبود بخشید بنابراین یکی از مهمترین عوامل تأثیرگذار بر دقت مدل‌سازی با استفاده از داده‌های طیفی، انجام پیش‌پردازش بر روی داده‌های طیفی خاک و انتخاب مناسب‌ترین روش پیش‌پردازش است. به طور کلی هرچند بافت خاک از پارامترهای زودبافت خاک می‌باشد، اما در سطوح وسیع به دلیل حجم بالای نمونه‌برداری، نیازمند به صرف وقت و هزینه زیادی می‌باشند. از این‌رو استفاده از ابزارها و فن‌آوری‌های نوین علمی می‌تواند در تعیین و برآورد آن‌ها در مقیاس بزرگ مؤثر باشد.

روش‌های پیش‌پردازش بر روی داده‌های طیفی می‌تواند افزایش یافت. قلی‌زاده (۱۴) و شرما و همکاران (۴۱) نیز نتایج مشابه‌ای بدست آوردند. به طور کلی طیف‌سنجی می‌تواند برای تعیین بافت خاک به کار رود.

نتیجه‌گیری کلی

بررسی اثر کلاس‌های بافتی مختلف بر منحنی طیفی نشان داد علاوه بر بافت خاک، مجموعه‌ای از عوامل مختلف از جمله رنگ، ساختمان، کانی‌شناسی، ماده آلی و... می‌تواند بر میزان بازتابش در محدوده VNIR/SWIR مؤثر باشد. نتایج این مطالعه نشان داد با استفاده از داده‌های طیفی در محدوده مرئی و فرورسرخ نزدیک و فرورسرخ کوتاه (۴۰۰-۲۴۵۰ نانومتر) می‌توان درصد اجزای بافت خاک را تعیین نمود. در واقع طیف‌سنجی به عنوان یک روش سریع، آسان و کم هزینه می‌تواند در پیش‌بینی درصد رس و سیلت خاک مورد استفاده قرار گیرد. هر چند نتایج طیف‌سنجی برای برآورد درصد

منابع

- 1- Araújo S.R., Demattê J.A.M., Franceschini M.H.D., Rizzo R., Stenberg B., and Wetterlind J. 2013. Improving the predictive performance of a national vis-NIR spectroscopic library by comparing clustering data transformation, and data-mining calibration techniques. p. 2431-2440. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto, 13 - 18 April. 2013. Foz do Iguaçu, PR, Brazil.
- 2- Barnes R.J., Dhanoa M.S., and Lister S.J. 1989. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43:772-777.
- 3- Bellinaso H., Demattê J.A.M., and Araújo S.R. 2010. Soil spectral library and its use in soil classification. *Rev. Bras.Ciênc. Solo*, 34:861-870.
- 4- Ben-Dor E., and Banin A. 1995. Near-infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. *Soil Science*, 159:259-270.
- 5- Bureau S., Ruiz D., Reich M., Gouble B., Bertrand D., Audergon J.M., and Renard C.M.G.C. 2009. Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. *Food Chemistry*, 113: 1323-1328.
- 6- Cai Y., Cabrera J.C., Georgiadis M., and Jayachandran K. 2002. Assessment of arsenic mobility in the soils of some golf courses in South Florida. *Science of the Total Environment*, 291:123-134.
- 7- CAMO AS Press. 2006. The Unscrambler Tutorials. CAMO Software Research & Development Team, CAMO Software, NedreVollgate 8, N-0158, Oslo, Norway, retrieved on March 20/2012 from <http://www.camo.com/>
- 8- Chang C.W., Laird D.A., Mausbach M.J., and Hurburgh C.R. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties *Soil Science Society of America Journal*, 65:480-490.
- 9- Chen L., Sheng-lu Z., Shao-hua W., Qing Z., and 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.
- 10- Cho M.A., Skidmore A., Corsi F., van Wieren S.E., and Sobhan I. 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9: 414-424.
- 11- Conforti M., Froio R., Matteucci G., and Buttafuoco G. 2015. Visible and near infrared spectroscopy for predicting texture in forest soil: an application in southern Italy. *iForest*, 8: 339-347
- 12- Curcio D., Ciralo G., D'Asaro F., and Minacapillia M. 2013. Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences*, 19:494 - 503.
- 13- Divya Y., Sanjeevi S., and Ilamparuthi K. 2014. A study on the hyperspectral signatures of sandy soils with varying texture and water content. *Arabian Journal of Geosciences*, 7:3537-3545.
- 14- Gholizadeh A., Boruvkai L., Saberioon M.M., Kozaki J., Vasati R., and Nemeki K. 2015. Comparing Different Data Preprocessing. *Soil & Water Res.*, 10:218-227.

- 15- Gomez C., Rossel V., and McBratney B. 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, 146:403-411.
- 16- Hillel D. *Applications of Soil Physics*. Academic Press Inc. 1980.
- 17- Hunt G.R., and Ashley R.P. 1979. Spectra of altered rocks in the visible and near infrared. *Economic Geology*, 74:1613-1629.
- 18- Jahanshahi R., and Zare, M, 2015. Assessment of heavy metals pollution in groundwater of Golgohar iron ore mine area, Iran. *Environmental Earth Science*, 74:505-520.
- 19- Ji J.F., Balsam W.L., Chen J., and Liu L.W. 2002. Rapid and quantitative measurement of hematite and goethite in the Chinese Loess-Paleosol sequence by diffuse reflectance spectroscopy. *Clay Mineral*, 50:208-216.
- 20- Jones H.G., and Vaughan R.A. 2010. *Remote sensing of vegetation (principles, techniques, and applications)*. Oxford University Press, New York.
- 21- Kaletia A.L., Tian L.F., and Hirschi M.C. 2005. Relationship between soil moisture content and soil surface reflectance. *Transactions of the ASAE*, 48:1979-1986.
- 22- Klement A., Jaksik O., Kodesova R., Drabek O., and Boruvka L. 2013. Application of VNIR diffuse reflectance spectroscopy to estimate soil organic carbon content, and content of different forms of iron and manganese. *Geophysical Research Abstracts*, 15:10846-1.
- 23- Lacerda M.P.C., Demattê J.A.M., Sato M.V., Fongaro C.T., Gallo B.C., and 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.
- 24- Lagacherie P., Baret F., Feret J.B., Netto J.M., and Robbez-Masson J.M. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, 112:825-835.
- 25- Lee K.S., Lee D.H., Sudduth K.A., Chung S.O., Kitchen N.R., Drummond S.T. 2009. Wavelength identification and diffuse reflectance estimation for surface and profile soil properties. *American Society of Agricultural and Biological Engineers*, 52:683-695
- 26- Liu Y., Jiang Q., Fei T., Wang J., Shi T., Guo K., Li X., and Chen Y. 2014. Transferability of a Visible and Near-Infrared Model for Soil Organic Matter Estimation in Riparian Landscapes. *Remote Sensing*, 6:4305-4322
- 27- Lucas Y., and Gagelli J. 2002. Hyperspectral detection of sand. p. 12-17. 7th International Conference on Remote Sensing for Marine and Coastal Environments, 20-22 May. 2002. Miami, Florida. U.S
- 28- Luo X., Yu S., and Li X. 2011. Distribution, availability, and sources of trace metals in different particle size fractions of urban soils in Hong Kong: Implications for assessing the risk to human health. *Environmental Pollution*, 159:1317-1326
- 29- Martens H., Jensen S.A., Geladi P. 1983. Multivariate linearity transformations for near infrared reflectance spectroscopy. In: Christie O.H.J. *Proc. Nordic Symp. Stokkland Forlag, Norway: Applied Statistics*; p. 205-234.
- 30- Martens H., Martens M. 2000. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (pls). *Food Qual Prefer*, 11:5-16.
- 31- Martens H., Næs T. 1989. *Multivariate Calibration*. John Wiley & Sons: Chichester, United Kingdom.
- 32- McCoy R.M. 2005. *Field Methods in Remote Sensing*. A Division of Guilford Publications, Inc. Spring, New York, U.S, pp. 67-87.
- 33- McDowell M.L., Bruland G.L., Deenik J.L., Grunwald S., and Knox N.M. 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma*, 189:312-320
- 34- Miller R.H., and Keeny D.R. 1992. *Methods of Soil Analysis*. p. 65-98. Part 1, 2, Physical, Chemical and Mineralogical properties. Soil Science Society of America, Madison, Wisconsin, USA.
- 35- Mouazen A.M., Kuang B., De Baerdemaeker J., and Ramon H. 2010. Comparison between principal components, partial least squares and artificial neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared 455 spectroscopy. *Geoderma*, 158:23-31.
- 36- Nørgaard L., Saudland A., Wagner J., Nielsen J.P., Munck L., and Engelsen S.B. 2000. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Applied Spectroscopy*, 54:413-419.
- 37- 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.
- 38- R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- 39- Rinnan A., Berg F.V.D., and Engelsen S.B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28: 1201-1222.
- 40- Schwartz G., Eshel G., and Ben-Dor E. 2011. Reflectance spectroscopy as a tool for monitoring contaminated soils. p. 67-90. In S. Pascucci (ed.) *Soil Contamination, In Technology*.

- 41- Sharma S., Goodarzi M., Ramon H., and Saeys W. 2014. Performance evaluation of preprocessing techniques utilizing expert information in multivariate calibration. *Talanta*, 121:105-12.
- 42- Shiferaw A., and 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.
- 43- Silva E. B., ten Caten, Dalmolin R.S.D., Dotto A.C., Silva W.C., and 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.
- 44- Stoner E.R., Baumgardner M.F. 1981. Characteristics variations in reflectance of surface soils. *Soil Science Society of America Journal*, 45: 1161–1165.
- 45- Summers D., Lewis M., Ostendorf B., and Chittleborough D. 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators*, 11:123-131.
- 46- ViscaraRossel R., and Behrens T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158:46-54.
- 47- Viscarra Rossel R.A., Taylor H.J., and McBratney A.B. 2007. Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science*, 58:343-353.
- 48- Wilding L. 1985. Spatial variability. Its documentation, accommodation, and implication to soil surveys. In D.R. Nielson and J. Bouma (eds) *Soil Variability*, Pudo, Wagenigen, the Netherlands.
- 49- Wold S., Martens H., and Wold H. 1983. The multivariate calibration problem in chemistry solved by the PLS method. *Matrix Pencils*.
- 50- Wold S., Sjöström M., and Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130.

Comparing Different Statistical Models and Pre-processing Techniques for Estimation of Soil Particles Using VNIR/SWIR Spectrum

M. Tayebi^{1*} - M. Naderi² - J. Mohammadi³ - M. Hosseinjani Zadeh⁴

Received: 05-08-2017

Accepted: 29-01-2018

Introduction: Soil texture is one of the major physical properties of soils that plays important roles in water holding capacity, soil fertility, environmental quality and agricultural developments. Measurement of soil texture elements in large scales is time consuming and costly due to the high volume of sampling and laboratory analysis. Therefore, assessing and using simple, quick, low-cost and advanced methods such as soil spectroscopy can be useful. The objectives of this study were to examine two statistical models of Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) to estimate soil texture elements using Visible and Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) reflectance spectroscopy (400-2450nm).

Materials and Methods: A total of 120 composite soil samples (0-10 cm) were collected from the Kafemoor basin (55° 15' - 55° 25' E; 28° 51' - 29° 11' N), Sirjan, Iran. The samples were air dried and passed through a 2 mm sieve and soil texture components were determined by the hydrometer method (Miller and Keeny 1992). Reflectance spectra of all samples were measured using an ASD field-portable spectrometer in the laboratory. Soil samples were divided into two random groups (80% and 20%) for calibration and validation of models. PLSR and PCR models and different pre-processing methods i.e. First (FD) and Second Derivatives (SD), Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) were applied and compared to estimate texture elements. The cross-validation method was used to evaluate calibration and validation sets in the first part (80%) and coefficient of determination (R^2), Root Mean Square Error (RMSE) and Residual Prediction Deviation (RPD) were also calculated. For testing predictive models, the second part of data (20%) was used and R^2 and RMSE of predictive accuracy were calculated.

Results and Discussion: The results of applying two statistical models for estimating LogClay (%) showed that R^2 of calibration (R^2_{CV}) and validation (R^2_{VAL}) datasets ranged from 0.22 to 0.72 and 0.12 to 0.54, respectively. The lowest RMSE was computed for PLSR model with SD pre-processing. The highest RPD of calibration (RPDCV) and validation (RPDVAL) were obtained for PLSR with SD pre-processing technique which was classified as a very good and good model, respectively. The results indicated possible prediction of soil clay content by using PCR model with SD pre-processing techniques. In addition, the PCR predicted soil texture elements poorly according to RPD values while the PLSR model with SD pre-processing was the best model for predicting soil clay content. The R^2_{CV} and R^2_{VAL} of PLSR models for LogSilt (%) varied from 0.34 to 0.73 and 0.27 to 0.58, respectively. The RMSECV varied from 0.14 for FD pre-processing to 0.23 for no-preprocessing and the RMSEVAL ranged between 0.18 and 0.24. The highest RPDCV (2.07) and RPDVAL (1.59) were obtained for PLSR with FD pre-processing which were classified as very good and good models, respectively. The results of PCR model developments for estimating LogSilt (%) indicated that the highest RPDCV and RPDVAL were, respectively, 1.31 and 1.25 for MSC pre-processing techniques which were rated as poor models. On the contrary to PLSR models, PCR models were not reliable for predicting LogSilt (%). The results of PLSR models for estimating LogSand (%) revealed that the highest R^2_{CV} and R^2_{VAL} were 0.56 and 0.47, respectively and the lowest RMSECV and RMSEVAL were 0.14 and 0.16, respectively which were obtained for SD pre-processing. The RPDCV and RPDVAL values for SD pre-processing in PLSR model were 1.59 and 1.39 which were rated as good and poor performance of predictions, respectively. The highest RPDCV and RPDVAL for PCR models were obtained with the MSC pre-processing indicating poor model. Therefore, PLSR model with SD pre-processing techniques was superior model for estimation of LogSand (%). Overall, PLSR model with SD pre-processing techniques performed better in estimating clay and

1, 2 and 3- Ph.D. educated, Associate Professor and Professor of Soil Science Department, College of Agriculture Shahrekord University, Iran

4- Assistant Professor, Department of Ecology, Research Institute of Sciences and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran

(* - Corresponding Author Email: mahtayebi@yahoo.com)

sand and PLSR model with FD pre-processing gave better estimate of silt content.

Conclusions: Our finding indicated that clay and silt content can be estimated by using electromagnetic spectrum between VNIR-SWIR region. Further, spectroscopy could be considered as a simple, fast and low cost method in predicting soil texture and PLSR model with SD and FD pre-processing seems to be more robust algorithm to estimate LogClay and LogSilt, respectively.

Keywords: Partial Least Squares Regression (PLSR), Principal Component Regression (PCR), Spectroscopy

