



مقایسه روش‌های PLSR و PCR در برآورد برخی فلزات سنگین با استفاده از داده‌های طیف‌سنجی در دامنه VNIR/SWIR

محبوبه طیبی*^۱، مهدی نادری^۲ و جهانگرد محمدی^۳

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

* Email: mahtayebi@yahoo.com

چکیده

این مطالعه با هدف بررسی توانایی داده‌های طیفی در محدوده مرئی، فروسرخ نزدیک و فروسرخ کوتاه (۴۰۰-۲۴۵۰ نانومتر) در برآورد فلزات آهن، منگنز، نیکل و کروم با استفاده از روش‌های رگرسیون حداقل مربعات جزئی (PLSR) و رگرسیون مؤلفه اصلی (PCR) انجام شد. ارزیابی انواع روش‌های پیش‌پردازش داده‌های طیفی از جمله مشتق اول (FD)، مشتق دوم (SD)، تصحیح پخشیده چندگانه (MSC) و متغیر نرمال استاندارد (SNV) در مدل‌سازی انجام شد. جهت ارزیابی مدل ۸۰ درصد داده‌ها برای کالیبراسیون مدل و ۲۰ درصد برای صحت‌سنجی مدل به صورت تصادفی تقسیم شدند. همچنین جهت اعتبارسنجی از روش Leave one out-cross validation استفاده شد و مقادیر ضریب تبیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و انحراف پیش‌بینی باقیمانده (RPD) تعیین گردید. نتایج نشان داد برای فلزات آهن، منگنز و کروم روش PLSR همراه با پیش‌پردازش SD نتایج کارآمد و قابل قبولی دارد در حالی که برای نیکل نتایج مدل‌های ایجاد شده ضعیف می‌باشد.

واژه‌های کلیدی: رگرسیون حداقل مربعات (PLSR)، رگرسیون مؤلفه اصلی (PCR)، طیف‌سنجی در محدوده VNIR/SWIR.

مقدمه

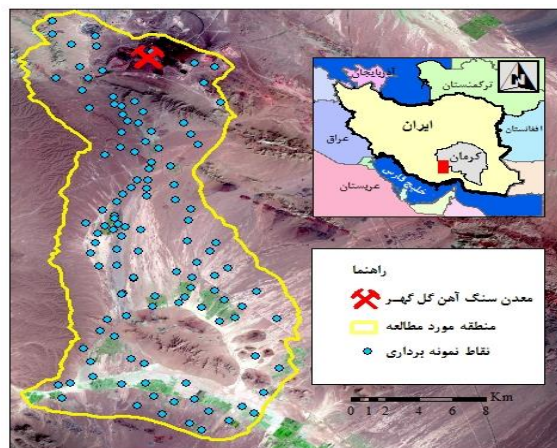
با توجه به این‌که روش‌های متعارف در تعیین و برآورد ویژگی‌های خاک بسیار پیچیده، وقت‌گیر و پرهزینه و نیازمند نمونه‌برداری، تجزیه شیمیایی و تهیه نقشه می‌باشند (کمپر و سومر، ۲۰۰۲) از این‌رو استفاده از ابزارها و فن‌آوری‌های نوین از جمله طیف‌سنجی می‌تواند به عنوان یک روش سریع، غیر مخرب و پیشرفته در پیش‌بینی ویژگی‌های خاک مؤثر باشند. با استفاده از طیف‌سنجی در محدوده مرئی و فروسرخ نزدیک و فروسرخ کوتاه (VNIR/SWIR) می‌توان برخی از ترکیبات خاک که دارای ویژگی‌های طیفی می‌باشند مانند اکسیدهای آهن، کربنات‌ها، ماده آلی، رس‌ها و... را تعیین نمود (اسچوارتز و همکاران، ۲۰۱۱). علاوه بر این از طریق همبستگی بین ترکیبات طیفی فعال خاک و غلظت عناصر، می‌توان با استفاده از طیف بازتابی خاک غلظت این عناصر در نمونه‌های خاک را به طور غیرمستقیم بدست آورد (ویو و همکاران، ۲۰۰۵). برای بررسی ارتباط بین پارامترهای طیفی ثبت شده خاک‌ها به وسیله اسپکترومتر و خصوصیات شیمیایی خاک‌ها از روش‌های مختلفی از جمله رگرسیون حداقل مربعات جزئی (Partial Least Squares Regression, PLSR) و رگرسیون مؤلفه اصلی (Principal Component Regression, PCR) استفاده می‌شود (اسچوارتز و همکاران، ۲۰۱۱).

پژوهشگران بسیاری از روش‌های طیف‌سنجی برای تعیین آلودگی خاک‌ها استفاده نمودند به عنوان مثال ویو و همکاران (۲۰۰۵) از روش PCR و داده‌های طیفی برای پیش‌بینی میزان آلودگی حاصل از نقره در خاک استفاده نمودند. سانگ و همکاران (۲۰۱۲) طی پژوهشی جهت تعیین پتانسیل سمی فلزات سنگین در خاک‌های کشاورزی چین از انتشار بازتاب‌های طیفی در ناحیه مرئی فروسرخ نزدیک و فروسرخ میانی و روش PLSR برای ارزیابی فرم‌های پیوندی و پیش‌بینی غلظت فلزات استفاده کردند. سریانو- دیسلا و همکاران (۲۰۱۳) برای مطالعه امکان پیش‌بینی مقادیر عناصر خاک براساس طیف انعکاسی مادون قرمز میانی از روش PLSR استفاده نمودند. بسیاری از محققین برای حذف هر گونه اطلاعات اضافی که در اثر عواملی چون پخشیدگی نور، غیرخطی بودن و ... ایجاد می‌شوند و برای ایجاد مدل‌های قوی و با دقت بیشتر، از روش‌های مختلف

پیش‌پردازش از جمله مشتق اول (First Derivative, FD) و مشتق دوم (Second Derivative, SD)، تصحیح پخشیده چندگانه (Multiplicative Scatter Correction, MSC) و متغیر نرمال استاندارد (Standard Normal Variate, SNV) استفاده می‌نمایند (رینان و همکاران، ۲۰۰۹). با توجه به اینکه مطالعات کمی در خصوص برآورد فلزات سنگین خاک به کمک داده‌های طیف‌سنجی و مدل‌های PLSR و PCR و همچنین مقایسه‌ی انواع تکنیک‌های پیش‌پردازش در کشور صورت گرفته است، این پژوهش با هدف (۱) بررسی توانایی داده‌های طیفی در محدوده VNIR/SWIR در برآورد فلزات آهن، منگنز، نیکل و کروم با استفاده از روش‌های PLSR و PCR، (۲) ارزیابی انواع روش‌های پیش‌پردازش داده‌های طیفی در مدل‌سازی انجام شد.

مواد و روش‌ها

منطقه مورد مطالعه با مساحت ۲۹۵۳۳ هکتار در ۶۰ کیلومتری جنوب غرب سیرجان در استان کرمان بین طول‌های جغرافیایی ۵۵ درجه و ۱۵ دقیقه تا ۵۵ درجه و ۲۵ دقیقه و عرض‌های جغرافیایی ۲۸ درجه و ۵۱ دقیقه تا ۲۹ درجه و ۱۱ دقیقه قرار گرفته است. به طوری که ناحیه معدنی سنگ آهن گل‌گهر در شمال آن واقع شده است (شکل ۱). این منطقه جز مناطق با آب و هوای خشک و بیابانی است که میانگین بارندگی سالانه آن ۱۷۲ میلی‌متر و حداقل و حداکثر درجه حرارت آن به ترتیب 12°C و 42°C می‌باشد.



شکل ۱- موقعیت جغرافیایی منطقه مورد مطالعه در کشور و استان کرمان.

در این پژوهش با پردازش تصاویر ماهواره‌ای لندست ۸ و با استفاده از نقشه‌های زمین‌شناسی، کاربری اراضی، توپوگرافی و مدل رقومی ارتفاع (DEM) محل نقاط نمونه‌برداری به روش هایپرکیوب لاتین (Latin Hypercube) در محیط نرم افزار R مشخص گردید. 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 غلظت فلزات سنگین و همچنین طیف‌سنجی، ابتدا نمونه‌های خاک هوا خشک و از الک ۲ میلی‌متری عبور داده شدند. غلظت عناصر سنگین آهن، منگنز، نیکل و کروم پس از هضم نمونه‌های خاک با اسید نیتریک ۴ نرمال به روش اسپوزیتو و همکاران (۱۹۸۲) با استفاده از دستگاه جذب اتمی تعیین شدند. برای اندازه‌گیری بازتاب طیفی خاک‌های مورد نظر در محدوده VNIR/SWIR از دستگاه طیف‌سنج زمینی در آزمایشگاه استفاده شد. به گونه‌ای که حدود ۱۰۰ گرم از هر نمونه 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 صفحه سفید مبنا واسنجی شد. برای هر نمونه خاک ۲۰ اسکن به منظور کاهش میزان نویز در نظر گرفته شد و میانگین آن‌ها با استفاده از نرم افزار RS^3 محاسبه و در تجزیه و تحلیل‌های طیفی در نظر گرفته شد. برای تصحیح پرش و هموارسازی طیف از نرم افزار SAMS V3.2 استفاده گردید. طول موج‌های بین ۳۵۰ تا ۴۰۰

نانومتر و ۲۴۵۰ تا ۲۵۰۰ که دارای آشفستگی‌های زیاد بودند حذف گردیدند و انواع روش‌های پیش‌پردازش از جمله FD و SD به همراه فیلتر ساویتزکی- گلائی، MSC و SNV بر روی داده‌های طیفی به کمک نرم‌افزار 10.3 Unscamler انجام شد. بررسی توزیع آماری داده‌ها به وسیله آزمون کولموگروف اسمیرنوف توسط نرم‌افزار ۹ SAS انجام گردید. بعد از نرمال‌سازی داده‌ها (تبدیل لگاریتمی) از روش‌های PLSR و PCR برای برآورد غلظت فلزات سنگین استفاده شد. برای این منظور 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 برازش 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 (Coefficient of Determination, R²) (رابطه ۱)، ریشه میانگین مربعات خطا (Root mean square error, RMSE) (رابطه ۲) و انحراف پیش‌بینی باقیمانده (Residual Prediction Deviation, RPD) (رابطه ۳) دقت و صحت مدل‌ها مورد ارزیابی قرار گرفتند (مک‌داول و همکاران، ۲۰۱۲).

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

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

$$RPD = \frac{SD}{RMSE} \quad (3)$$

در این معادله، O_i و P_i به ترتیب مقادیر مشاهداتی و تخمین زده شده، SD انحراف معیار مقادیر مشاهداتی و n تعداد داده‌ها می‌باشد. شاخص RPD از معیارهای سنجش میزان کارایی مدل در مطالعات طیف‌سنجی می‌باشد که به ۳ کلاس ضیف (RPD < ۱/۵)، قابل قبول (۱/۵-۲)، کاملاً مناسب (RPD > ۲) طبقه‌بندی می‌شود (ویلیامز، ۲۰۰۱؛ سامرز و لوئیس، ۲۰۱۱).

نتایج و بحث

خلاصه آماری غلظت فلزات آهن، منگنز، نیکل و کروم در خاک‌های منطقه مورد مطالعه در جدول (۱) آورده شده است. نتایج نشان می‌دهد که میانگین غلظت آهن، منگنز، نیکل و کروم به ترتیب برابر با ۲۴۰۴۰، ۲۷۳/۱، ۳۴/۹۹ و ۲۸۷۹ میلی‌گرم بر کیلوگرم است. آهن و کروم دارای چولگی مثبت و توزیع غیر نرمال هستند در حالی که منگنز و نیکل دارای توزیع نرمال می‌باشند. برای غلظت آهن، نیکل و کروم ضریب تغییرات بالاتر از ۳۵ درصد می‌باشد که نشانگر وجود تغییرات زیاد در منطقه مورد مطالعه می‌باشد. ضریب تغییرات برای منگنز بین ۱۵ تا ۳۵ درصد است که بیانگر وجود تغییرات متوسط است.

جدول ۱- خلاصه آماری غلظت فلزات سنگین خاک (mg/kg)

متغیر	واحد	حداقل	حداکثر	میانگین	چولگی	انحراف استاندارد	ضریب تغییرات
Fe	(mg/kg)	۴۴۳۶/۲۵	۲۷۱۳۷۵	۲۴۰۴۰	۴/۱۰	۴۵۱۱۹	۱۸۷/۶۸
Mn	(mg/kg)	۱۲۰/۴۵	۵۱۰/۶۸	۲۷۳/۱۰	۰/۲۲	۶۳/۹۴	۲۳/۴۱
Ni	(mg/kg)	۶/۶۵	۷۶	۳۴/۹۹	۰/۴۲	۱۳/۲۶	۳۷/۸۹
Cr	(mg/kg)	۹۲۲/۵۰	۱۲۹۰۰	۲۸۷۹	۳/۵۹	۱۷۶۲	۶۱/۲۰

جدول (۲) و (۳) مقادیر R²، RMSE و RPD حاصل از مدل‌سازی با استفاده از روش PLSR، PCR و انواع روش‌های پیش‌پردازش در برآورد غلظت فلزات سنگین در خاک را نشان می‌دهند. همان‌طور که مشاهده می‌شود روش‌های مختلف پیش‌پردازش سبب بهبود نتایج مدل‌سازی گردیده است. روش FD با حذف خط مبنا، روش SD با حذف خط مبنا و روند خطی،

روش MSC با نرمال سازی برای تصحیح پراکندگی نور و روش SNV با تصحیح پراکندگی نور منجر به افزایش کارایی مدل می گردند (قلی زاده و همکاران، ۲۰۱۵). در روش PLSR نتایج برآورد لگاریتم غلظت آهن، منگنز، نیکل و لگاریتم کروم نشان داد که بیشترین مقدار R^2 و کمترین مقدار RMSE در هر دو گروه کالیبراسیون و اعتبارسنجی با استفاده از روش PLSR همراه با پیش پردازش SD حاصل شد (جدول ۲). به طوری که مقدار R^2 در گروه کالیبراسیون برای لگاریتم آهن، منگنز، نیکل و لگاریتم کروم به ترتیب برابر با ۰/۸۸، ۰/۹۷، ۰/۴۲ و ۰/۸۱ و در گروه اعتبارسنجی به ترتیب برابر ۰/۶۳، ۰/۶۲، ۰/۰۴ و ۰/۴۱ به دست آمد. برای لگاریتم آهن و منگنز مقدار RPD برای داده های کالیبراسیون (RPD_{CV}) بزرگتر از ۲ و مقدار RPD برای داده های اعتبارسنجی (RPD_{VAL}) بین ۱/۵ تا ۲ محاسبه شد که براساس کلاس بندی ویلیامز و همکاران (۲۰۰۱) و سامرز و لوئیس (۲۰۱۱)، به ترتیب در کلاس کاملاً مناسب و قابل قبول قرار می گیرند. برای نیکل مقادیر RPD_{CV} و RPD_{VAL} کوچکتر از ۱/۵ محاسبه شدند و در کلاس ضعیف قرار گرفتند. برای کروم مقدار RPD_{CV} برابر با ۲/۳۶ و مقدار RPD_{VAL} برابر با ۱/۳۸ محاسبه شد که به ترتیب در کلاس کاملاً مناسب و ضعیف قرار می گیرند (جدول ۲).

جدول ۲- تاثیر روش های مختلف پیش پردازش طیفی بر دقت تعیین غلظت فلزات سنگین به روش PLSR

متغیر	پیش پردازش	R^2_{CV}	$RMSE_{CV}$	R^2_{VAL}	$RMSE_{VAL}$	RPD_{CV}	RPD_{VAL}
logFe (mg/kg)	No	۰/۵۵	۰/۲۷	۰/۴۹	۰/۲۹	۱/۴۹	۱/۳۹
	FD	۰/۸۷	۰/۱۵	۰/۵۹	۰/۲۷	۲/۶۹	۱/۵۰
	SD	۰/۸۸	۰/۱۴	۰/۶۳	۰/۲۴	۲/۸۹	۱/۶۸
	MSC	۰/۶۵	۰/۲۴	۰/۵۳	۰/۲۸	۱/۶۸	۱/۴۴
	SNV	۰/۵۶	۰/۲۶	۰/۴۲	۰/۳۰	۱/۵۰	۱/۳۱
Mn (mg/kg)	No	۰/۲۵	۵۰/۸۷	۰/۱۶	۵۵/۱۴	۱/۱۶	۱/۰۷
	FD	۰/۷۷	۲۷/۹۸	۰/۲۶	۵۱/۳۵	۲/۱۲	۱/۱۵
	SD	۰/۹۷	۱۰/۰۸	۰/۶۲	۳۴/۶۰	۵/۸۷	۱/۷۱
	MSC	۰/۶۲	۳۶/۱۵	۰/۳۹	۴۵/۹۵	۱/۶۳	۱/۲۹
	SNV	۰/۵۱	۴۱/۰۴	۰/۲۹	۴۸/۹۲	۱/۴۴	۱/۱۸
Ni (mg/kg)	No	-	-	-	-	-	-
	FD	۰/۰۵	۱۱/۹۳	۰/۰۱	۱۲/۳۴	۱/۰۳	۰/۹۹
	SD	۰/۴۲	۹/۳۲	۰/۰۴	۱۲/۰۷	۱/۳۱	۱/۰۲
	MSC	۰/۲۶	۱۰/۵۱	۰/۰۶	۱۲/۰۲	۱/۱۷	۱/۰۲
	SNV	۰/۲۹	۱۰/۲۹	۰/۱۹	۱۱/۰۳	۱/۱۹	۱/۱۲
logCr(mg/kg)	No	۰/۳۰	۰/۱۳	۰/۲۳	۰/۱۴	۱/۲۷	۱/۱۸
	FD	۰/۳۶	۰/۱۳	۰/۲۷	۰/۱۴	۱/۲۷	۱/۱۸
	SD	۰/۸۱	۰/۰۷	۰/۴۱	۰/۱۲	۲/۳۶	۱/۳۸
	MSC	۰/۳۴	۰/۱۳	۰/۲۶	۰/۱۴	۱/۲۷	۱/۱۸
	SNV	۰/۳۵	۰/۱۳	۰/۲۴	۰/۱۴	۱/۲۷	۱/۱۸

در روش PCR نتایج برآورد لگاریتم آهن، منگنز و لگاریتم کروم نشان داد که بیشترین مقادیر R^2 و کمترین مقادیر RMSE با پیش پردازش های MSC برای داده های کالیبراسیون و اعتبارسنجی حاصل شد (جدول ۳). کارایی این مدل ها برای لگاریتم آهن در هر دو گروه کالیبراسیون و اعتبارسنجی به ترتیب در کلاس قابل قبول و ضعیف و برای منگنز و لگاریتم کروم در هر دو گروه در کلاس ضعیف قرار می گیرند. در حالی که در برآورد نیکل با استفاده از روش PCR، بیشترین مقادیر R^2 و کمترین مقادیر RMSE با پیش پردازش SNV ایجاد شد که کارایی این مدل در کلاس ضعیف قرار می گیرد. به طور کلی نتایج مدل سازی با استفاده از روش PCR نشان داد که این روش کارایی کمی در برآورد غلظت فلزات سنگین دارد. مقایسه روش های PLSR و PCR در برآورد غلظت فلزات سنگین نشان داد روش PLSR همراه با پیش پردازش SD نتایج مطلوب تری را ارائه می دهند. که علت آن را می توان به استفاده همزمان میزان بازتابش (متغیرهای مستقل X) و غلظت عنصر (متغیرهای وابسته Y) در تعیین مولفه اصلی در روش PLSR نسبت داد. در حالی که در روش PCR تنها میزان بازتابش (متغیرهای مستقل X) در تعیین مولفه اصلی به کار می رود.

برای صحت‌سنجی مدل‌های ایجاد شده توسط روش‌های PLSR و PCR با روش‌های مختلف پیش‌پردازش مقادیر ضریب تبیین (R^2_{PRE}) و ریشه مربعات خطا (RMSEP) برای ۲۰ درصد داده‌ها که از قبل جدا شده بودند، تعیین گردید. جدول (۴) مقادیر R^2_{PRE} و RMSEP حاصل از صحت‌سنجی بهترین مدل‌های ایجاد شده در برآورد فلزات سنگین خاک را نشان می‌دهد. برای برآورد تمامی فلزات مورد مطالعه روش PLSR با پیش‌پردازش SD به عنوان روش مناسب‌تر انتخاب شدند. نتایج صحت‌سنجی نشان داد مقادیر R^2_{PRE} برای لگاریتم آهن، منگنز، نیکل و لگاریتم کروم به ترتیب برابر با ۰/۵۴، ۰/۶۰، ۰/۱۱ و ۰/۵۱ و مقادیر RMSEP به ترتیب برابر با ۰/۲۱، ۴۳/۹۵، ۱۹/۳۴ و ۰/۱۳ بودند. به طور کلی نتایج صحت‌سنجی نشان داد لگاریتم آهن، منگنز و لگاریتم کروم دارای R^2_{PRE} بالاتر از ۰/۵۰ و نیکل دارای R^2_{PRE} بسیار پایینی هستند.

جدول ۳- تاثیر روش‌های مختلف پیش‌پردازش طیفی بر دقت تعیین غلظت فلزات سنگین به روش PCR

متغیر	پیش‌پردازش	R^2_{CV}	RMSE _{CV}	R^2_{VAL}	RMSE _{VAL}	RPD _{CV}	RDP _{VAL}
logFe (mg/kg)	No	۰/۵۴	۰/۲۷	۰/۴۸	۰/۳۰	۱/۴۹	۱/۳۹
	FD	۰/۵۴	۰/۲۷	۰/۴۹	۰/۲۹	۱/۵۰	۱/۳۹
	SD	۰/۴۶	۰/۲۹	۰/۴۰	۰/۳۲	۱/۳۹	۱/۲۶
	MSC	۰/۶۱	۰/۲۴	۰/۵۳	۰/۲۸	۱/۶۸	۱/۴۴
	SNV	۰/۵۸	۰/۲۶	۰/۴۴	۰/۳۰	۱/۵۵	۱/۳۱
Mn (mg/kg)	No	۰/۲۴	۵۱/۱۳	۰/۱۳	۵۵/۰۲	۱/۱۵	۱/۱۳
	FD	۰/۱۲	۵۵/۱۴	۰/۱۰	۵۶/۲۳	۱/۰۷	۱/۰۵
	SD	۰/۱۱	۵۵/۴۷	۰/۰۴	۵۸/۶۹	۱/۰۶	۱/۰۱
	MSC	۰/۳۹	۴۵/۷۶	۰/۲۵	۵۰/۹۱	۱/۲۹	۱/۱۶
	SNV	۰/۲۷	۵۰/۰۲	۰/۲۴	۵۱/۶۹	۱/۱۴	۱/۱۸
Ni (mg/kg)	No	۰/۱۹	۱۰/۹۶	۰/۰۷	۱۲	۱/۱۲	۱/۰۳
	FD	-	-	-	-	-	-
	SD	-	-	-	-	-	-
	MSC	۰/۰۲	۱۲/۱۲	۰/۰۰۴	۱۲/۳۴	۱/۰۱	۰/۹۹
	SNV	۰/۲۸	۱۰/۳۷	۰/۲۰	۱۱/۰۳	۱/۱۸	۱/۱۲
logCr(mg/kg)	No	۰/۲۹	۰/۱۴	۰/۱۹	۰/۱۵	۱/۱۸	۱/۱۰
	FD	۰/۳۳	۰/۱۳	۰/۲۸	۰/۱۴	۱/۲۷	۱/۱۸
	SD	۰/۳۲	۰/۱۳	۰/۲۷	۰/۱۴	۱/۲۷	۱/۱۸
	MSC	۰/۳۶	۰/۱۳	۰/۲۷	۰/۱۴	۱/۲۷	۱/۱۸
	SNV	۰/۳۲	۰/۱۳	۰/۲۶	۰/۱۴	۱/۲۷	۱/۱۸

جدول ۴- نتایج صحت‌سنجی برای بهترین مدل‌های ایجاد شده در برآورد فلزات سنگین خاک

متغیر	واحد	مدل	پیش‌پردازش	R^2_{PRE}	RMSEP
LogFe	(mg kg ⁻¹)	PLSR	SD	۰/۵۴	۰/۲۱
Mn	(mg kg ⁻¹)	PLSR	SD	۰/۶۰	۴۳/۹۵
Ni	(mg kg ⁻¹)	PLSR	SD	۰/۱۱	۱۹/۳۴
LogCr	(mg kg ⁻¹)	PLSR	SD	۰/۵۱	۰/۱۳

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 فلزات سنگین به دست آمد. هر چند کارایی مدل برای نیکل ضعیف است ولی این روش نسبت به روش PCR و سایر روش‌های پیش‌پردازش دارای کارایی بالاتری بود. محققین بسیاری نیز نتایج مشابهی را گزارش کردند و از روش PLSR به عنوان یکی از روش‌های مناسب در تعیین فلزات سنگین خاک نام برده‌اند. به عنوان مثال سانگ و همکاران (۲۰۱۲) نیز نشان دادند که روش PLSR در محدوده VNIR/SWIR نتایج بهتری در برآورد نقره، کادمیوم، کروم، مس و سرب دارد و مقدار R^2 بیشتر از ۰/۵۰ است. سریانو- دیسلا و همکاران (۲۰۱۳) با مطالعه امکان پیش‌بینی مقادیر فلزات خاک براساس طیف انعکاسی مادون قرمز میانی و روش PLSR نشان دادند که برای فلزات سنگینی چون آهن، نیکل، کروم، روی، مس و منگنز میزان R^2 بیشتر از ۰/۶۰ و نتایج خوب است و برای سایر فلزات از جمله کادمیوم، مولیبدوم، سرب و ... میزان



R^2 کمتر از ۰/۵۰ و نتایج خوب نیست. استفاده از روش‌های مختلف پیش‌پردازش نتایج پیش‌بینی را بهبود بخشید که با نتایج پژوهشگران مختلفی از جمله قلی‌زاده و همکاران (۲۰۱۵)، شارما و همکاران (۲۰۱۴) مطابقت دارد.

منابع

- Gholizadeh A., Boruvkai L., Saberioon M.M., Kozaki J., Vasati R. and Nemeki K. 2015. Comparing Different Data Preprocessing. *Soil and Water Research*, 10: 218–227.
- Kemper T. and Sommer S. 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environmental Science and Technology*, 36: 2742–2747.
- 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.
- 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.
- 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*.
- Sharma S., Goodarzi M., Ramon H. and Saeyns W. 2014. Performance evaluation of preprocessing techniques utilizing expert information in multivariate calibration. *121: 105-112*.
- Song Y., Fengling L., Zhongfang Y., Godwin A.A., Ray L.F. and Junfeng J. 2012. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. *Applied Clay Science*, 64: 75-85.
- Soriano-Disla J.M., Janik L., McLaughlin M.J., Forrester S., Kirby J., Reimann C. and The EuroGeoSurveys GEMAS Project Team. 2013. The use of diffuse reflectance mid-infrared spectroscopy for the prediction of the concentration of chemical elements estimated by X-ray fluorescence in agricultural and grazing European soils. *Applied Geochemistry*, 29: 135–143.
- Sposito G.L., Lund J., and Chang A.C. 1982. Trace metal chemistry in arid-zone field soils amended with sewage sludge: I. Fractionation of Ni, Cu, Zn, Cd, and Pb in solid phases. *Soil Science Society of America Journal*, 46: 260-265.
- Summers D., Lewis M., Ostendorf B. and Chittleborough D. 2011. Visible near-infrared reflectance spectroscopy visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research*, 41: 101-114.
- Williams P.C. 2001. Implementation of near-infrared technology. Pp. 145-169. *In: Williams P., Norris K. (Eds.), Near-infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists Inc., St. Paul, MN*
- Wu Y., Chen J., Wu X., Tian Q., Ji J. and Qin Z. 2005. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Applied Geochemistry*, 20: 1051–1059.

Comparing PLSR and PCR Models for Assessing Soil Heavy Metals based on VNIR/SWIR Spectral data

M.Tayebi^{1,*}, Mehdi Naderi², J. Mohammadi³

1, 2 and 3 -PhD student, , Professor, Department of Soil Sciences, Faculty of Agriculture, Shahrekord University.

* Email: mahtayebi@yahoo.com

Abstract

Two statistical models including Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) were comparatively performed to determine prediction accuracy of Visible and Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) reflectance spectroscopy (400-2450 nm) in quantifying Fe, Mn, Ni and Cr concentration in contaminated soils. Different spectral pre-processing techniques such as First and Second Derivatives (FD and SD), Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) were also applied to select the best model. The ability of predictive models was evaluated by splitting soil samples into two random groups (80% and 20%). The first part (80%) was used to evaluate calibration and validation sets using cross-validation method and the second part (20%) was applied to test models. Coefficient of determination (R^2), Root Mean Square Error (RMSE) and Residual Prediction Deviation (RPD) were also calculated to evaluate models. The results indicated that the PLSR model with Second derivative pre-processing was more accurate technique for predicting Fe, Mn and Cr concentration than PCR. Whereas for Ni, the predictive models didn't give acceptable and appropriate prediction performance.

Keywords: Partial Least Squares Regression (PLSR), Principal Component Regression (PCR), VNIR/SWIR spectroscopy