

Modeling Soil Salinity in Khuzestan Lands Susceptible for Dust Production Using Visible-Near Infrared Spectroscopic Method

MANSOUR CHATERNOUR¹, AHMAD LANDI^{1*}, AHMAD FARROKHIAN FIROUZI¹, ALIAKBAR NOROOZI²,
HOSSEINALI BAHRAMI³

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

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

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

(Received: March. 30, 2019- Revised: May. 13, 2019- Accepted: May. 14, 2019)

ABSTRACT

A broad area of saline and semi-saline lands of Khuzestan province have changed into centers susceptible to dust production due to eroded wind and lack of surface coating and low soil resistance. The objective of this study was to model the soil salinity of sensitive areas to dust production in Khuzestan Province using regression method of visible and near-infrared wavelengths (2500-350 nm). The least square multivariate regression model, artificial neural network and random forest model were used to estimate soil salinity. The main soil spectrum was determined using the FieldSpect machine. Also, preprocessing methods including Savitzky-Golay filter, the first derivative with the Savitzky-Golay filter (FD-SG), the second derivative with the Savitzky-Golay filter (SD-SG), the standard normalization method (SNV), and the continuum remove method (CR) were used to eliminate the noise and to increase the accuracy of the multivariate model. The results showed that the combined model partial least squares-artificial neural network model with assessment criteria (RPDcal = 3.40-2.65) has high accuracy for salinity estimation. In contrast, the combined model of least squares - random forest showed the lowest accuracy (RPDcal = 0.85-1.98). Preprocess of the main spectrum in two models (neural network and partial least squares regression) increased the relative accuracy of the model; while in the random forest model, preprocess reduced the accuracy of the model compared to the main spectrum. The ranges of 1800, 1900, 2000, 2300 and 1500 nm were recognized as "the key wavelengths" impressed by soil salinity. The key wavelengths can be used in remote sensing studies and mapping of soil salinity in areas sensitive to dust production in Khuzestan province.

Keywords: Partial least squares regression, Preprocessing, Savitzky-Golay filter, Key wavelengths, Random forest model

استفاده از روش طیف‌سنجی مرئی-مادون قرمز نزدیک در مدل‌سازی شوری خاک اراضی مستعد تولید ریزگرد استان خوزستان

منصور چترنور^۱، احمد لندی^{۱*}، احمد فرخیان فیروزی^۱، علی اکبر نوروزی^۲، حسینعلی بهرامی^۳

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

۲. پژوهشکده حفاظت خاک و آبخیزداری، تهران، ایران

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

(تاریخ دریافت: ۱۳۹۸/۱/۱۰ - تاریخ بازنگری: ۱۳۹۸/۲/۲۳ - تاریخ تصویب: ۱۳۹۸/۲/۲۴)

چکیده

سطح وسیعی از اراضی شور و نیمه شور استان خوزستان به علت عدم پوشش سطحی و مقاومت کم خاک در برابر باد فرساینده به کانون‌های مستعد تولید ریزگرد تبدیل شده‌اند. هدف از این پژوهش مدل‌سازی شوری خاک مناطق حساس به تولید ریزگرد استان خوزستان با روش طیف‌سنجی امواج مرئی و مادون قرمز نزدیک (۳۵۰-۲۵۰۰ نانومتر) بود. از مدل‌های چند متغیره رگرسیون حداقل مربعات جزئی، شبکه عصبی مصنوعی و مدل جنگل تصادفی برای مدل‌سازی شوری خاک به کار گرفته شد. طیف بازتابی خاک با دستگاه طیف‌سنج زمینی (FieldSpec) تعیین شد. همچنین روش‌های پیش‌پردازش فیلتر ساویتزی گولای، مشتق اول به همراه فیلتر ساویتزی گولای (FD-SG)، مشتق دوم به همراه فیلتر ساویتزی گولای (SD-SG)، روش نرمال‌سازی استاندارد (SNV) و روش حذف پیوستار (CR)، جهت حذف نویز و افزایش دقت مدل‌های چند متغیره مورد استفاده قرار گرفت. نتایج نشان داد که مدل ترکیبی حداقل مربعات جزئی- شبکه عصبی مصنوعی با معیارهای ارزیابی $(R^2_{cal} = 0.74 - 0.65)$ در برآورد شوری خاک دقت مناسبی دارد. در مقابل مدل ترکیبی حداقل مربعات - جنگل تصادفی نیز کمترین دقت $(R^2_{cal} = 0.11 - 0.08)$ را نشان داد. پیش‌پردازش طیف اصلی در دو مدل شبکه عصبی و رگرسیون حداقل مربعات جزئی سبب افزایش نسبی دقت مدل شد در حالی که در مدل جنگل تصادفی پیش‌پردازش سبب کاهش دقت برآورد مدل، نسبت به طیف اصلی شد. محدوده ۱۹۰۰، ۱۸۰۰، ۲۰۰۰، ۲۳۰۰ و ۱۵۰۰ نانومتر به عنوان طول موج کلیدی متأثر از شوری خاک شناسایی شد. از طول موج‌های کلیدی به دست آمده، می‌توان در مطالعات دورسنجی و تهیه نقشه شوری مناطق حساس به تولید گرد و غبار استان خوزستان استفاده کرد.

واژه‌های کلیدی: رگرسیون حداقل مربعات جزئی، پیش‌پردازش، فیلتر ساویتزی گولای، طول موج کلیدی، مدل جنگل تصادفی

مقدمه

مقرون به صرفه و غیر مخرب برای برآورد ویژگی‌های خاک استفاده شده است. در شناسایی و ارزیابی شوری خاک استفاده از روش طیف‌سنجی مفید است (Zeng et al., 2018).

برای برآورد خصوصیات خاک به روش طیف‌سنجی نیاز به مدل‌های چند متغیره رگرسیونی است که رابطه طیف را با ویژگی‌های خاک به خوبی نشان دهد. مطالعات نشان داده است که به دلیل توزیع غیریکنواخت داده‌ها و رفتار بازتابی غیرخطی این ویژگی‌های خاک استفاده از مدل‌های دارای رفتار خطی مانند PLSR محدودیت‌هایی دارد (Nawar et al., 2015; Nawar et al., 2016). در صورت استفاده از روش‌های غیرخطی با توجه به رفتار پیچیده و غیرخطی طیف خاک، انتظار می‌رود که محدودیت مدل کاهش یافته و دقت مدل‌سازی تا حد معناداری افزایش یابد.

سطح وسیعی از اراضی شور و نیمه شور استان خوزستان به کانون‌های مستعد تولید گرد و غبار تبدیل شده است. افزایش شوری و سدیم خاک منجر به افزایش منافذ و کاهش ضخامت لایه پخشیده، کاهش مقاومت خاک و افزایش حساسیت در برابر فرسایش بادی می‌شود. با توجه به سطح وسیع (حدود ۳۵۰ هزار هکتار) این مناطق، ارزیابی شوری خاک با روش‌های مرسوم و سنتی پرهزینه و زمان‌بر است. علاوه بر این در روش‌های قدیمی تعمیم نتایج به کل منطقه دارای محدودیت‌هایی است (Dehaan & Taylor, 2002). در سال‌های اخیر از تصاویر ابر طیفی در دامنه‌های مرئی، مادون قرمز نزدیک و مادون قرمز میانی (طول موج بین ۳۵۰ تا ۲۵۰۰ نانومتر) به عنوان روشی سریع،

Ji *et al.* (2016) برای برآورد خصوصیات رطوبت، ماده آلی، ظرفیت تبادل کاتیونی و عناصر تبدالی خاک از رفتار طیفی استفاده کردند. نتایج این مطالعه نشان داد که مدل PLSR عملکرد مناسبی در برآورد ظرفیت تبادل کاتیونی (CEC)، کلسیم و منگنز با شاخص‌های عملکرد بالاتر ($R^2 > 0.5$ و $RPD > 1/4$) داشت، اما در برآورد خصوصیات Cu، Fe، فسفر، نیتروژن نیتراتی، Na و K تبدالی عملکرد ضعیف‌تر را ($R^2 < 0.5$ و $RPD < 1/4$) نشان داد. در مطالعه‌ای دیگر Wang *et al.* (2018) برای برآورد شوری خاک، از دو روش رگرسیون حداقل مربعات جزئی (PLSR) و جنگل تصادفی (RF) با پیش‌پردازش مشتق جزئی (در محدوده صفر تا ۲ و با فاصله ۰/۱) استفاده کردند. اعتبارسنجی مدل‌ها نشان داد که در برآورد شوری خاک مدل RF نسبت به روش PLSR عملکرد بهتری ($R^2 = 0.93$)، ($RMSE = 4/57$) و RPD (۲/۷۸ \leq ۲/۵۰) دارد.

با توجه به محدودیت مدل‌های خطی در برآورد شوری خاک، پیچیدگی رفتار نمک در خاک و نیاز به مطالعه رفتار آن در کانون‌های تولید ریزگرد در استان خوزستان، اهداف این پژوهش عبارت‌اند از: ۱- مقایسه کارایی و دقت روش‌های PLSR، شبکه عصبی ANN و جنگل تصادفی در مدل‌سازی شوری خاک جنوب شرق اهواز، ۲- مقایسه روش‌های مختلف پیش‌پردازش شامل: فیلتر ساویتزکی و گلای، فیلتر ساویتزکی و گلای به همراه مشتق اول (FD-SG)، فیلتر ساویتزکی و گلای به همراه مشتق دوم (SD-SG)، روش نرمال‌سازی استاندارد (SNV) و روش حذف پیوستار (CR). ۲- تعیین طول موج‌های کلیدی مربوط به شوری خاک این مناطق هدف دیگر این پژوهش بود.

مواد و روش‌ها

منطقه مورد مطالعه

این مطالعه در منطقه‌ای واقع در جنوب و مرکز استان خوزستان با مساحت حدود ۱۱۰ هزار هکتار و مختصات $31^{\circ} 10'$ و E (۴۱ $^{\circ} 48'$) انجام شد. این منطقه، از نظر زمین‌شناسی در گروه دشت‌های رسوبی و زمین‌های پست و شور قرار دارد (شکل ۱). میانگین بلند مدت بارش و دمای منطقه به ترتیب ۲۱۸ میلی‌متر و ۲۶ سانتی‌گراد است (irimo, 2019). شوری اراضی به دلیل انتقال املاح از زمین‌های بالادست و صعود موینگی به دلیل بالا بودن سطح آب زیرزمینی است. کاهش بارش در سال‌های اخیر پدیده شور شدن را تشدید کرده است. این مناطق دارای خاک‌های عمیق، شور و سدیمی است.

به دلیل پیچیدگی ترکیبات نمک، ناهمگنی خاک و شرایط منطقه، اطلاعات طیفی در مناطق مختلف حتی در شوری‌های مشابه متفاوت است (Wang *et al.*, 2014). در مورد شوری خاک طول موج خالص یا محدوده خاصی که مورد تأکید همه پژوهشگران باشد وجود ندارد و یا محدوده طیفی موردنظر در ترکیب با خصوصیات دیگر مانند رطوبت خاک است (Farifteh *et al.*, 2008). بنابراین، نیاز است که دقت مدل‌ها در مناطق مختلف جهت برآورد شوری خاک ارزیابی شود.

در مطالعات گذشته برآورد خصوصیات خاک با روش طیف‌سنجی بازتابی زمینی، اغلب از بعضی ویژگی‌های خاک با نشانه‌های بارز طیفی در محدوده امواج الکترومغناطیسی از قبیل ماده آلی (Ji *et al.*, 2009; He *et al.*, 2016)، کربنات کلسیم (Gomez *et al.*, 2008)، بافت (Curcio *et al.*, 2013) و در مواردی از شوری و عناصر محلول (Silva *et al.*, 2016) استفاده شده است. Farifteh *et al.* (2008) پژوهشی را برای بررسی رابطه بین غلظت نمک خاک و پاسخ طیفی با استفاده از مدل رگرسیونی خطی انجام دادند. در این مطالعه از سه بافت خاک مختلف و شش نوع مواد معدنی نمکی استفاده شد. نتایج نشان داد که برآورد مدل برای نمک NaCl، دارای دقت قابل قبول ($R^2 CV = 0.65$) با $PRMSECV = 7.50/3\%$ و برای نمک‌های KCl و Na_2SO_4 دقت مدل‌سازی کاهش ($R^2 CV > 0.5$) یافت. Zeng *et al.* (2018) با هدف مدل‌سازی شوری خاک از تصاویر ابر طیفی پیوسته و چند طیفی ماهواره لندست استفاده کردند. آنها در برآورد شوری با روش تصاویر ماهواره‌ای، از شاخص شوری استفاده کردند. روش دیگر استفاده از طیف‌سنجی و تجزیه تحلیل طیف، بر اساس مدل‌های رگرسیون حداقل مربعات جزئی (PLSR)، ماشین بردار پشتیبانی (SVM) و یادگیری عمیق (DL) بود. نتایج مطالعه آن‌ها نشان داد مدل PLSR با شاخص شوری، عملکرد بهتری (۰/۷۱ تا $R^2 = 0.86$) نسبت به SVM و DL دارد. همچنین در این مطالعه مشخص شد که مدل PLSR اثرات منفی نویزها را کاهش داد.

در پژوهشی Khayamim *et al.* (2015) خصوصیات شیمیایی خاک‌های استان اصفهان را با طیف‌سنجی بررسی کردند و ویژگی‌های ماده آلی، کربنات کلسیم و گچ را با دقت مناسبی با ضریب تعیین (R^2) ۰/۶۱، ۰/۴۵ و ۰/۸۰ برآورد کردند. Hassani *et al.* (2014) از روش طیف‌سنجی جهت برآورد ویژگی‌های کربنات کلسیم و گچ ($R^2 > 0.5$ و $RPD > 1/8$) عملکرد عالی و برای ماده آلی عملکرد قابل قبولی ($R^2 > 0.5$ و $RPD > 1/4$) را مشاهده کردند.

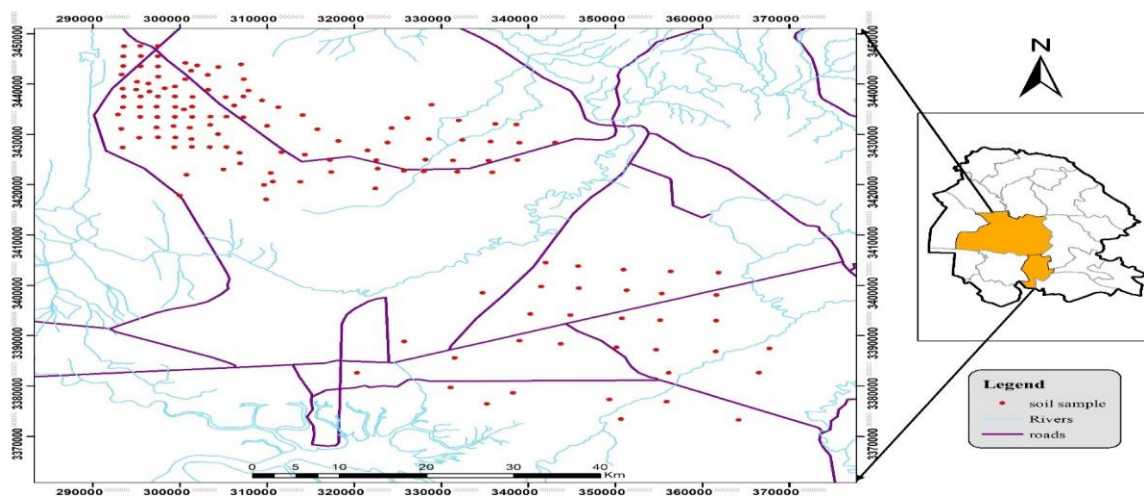
نمونه برداری

برای نمونه برداری، ابتدا منطقه به صورت شبکه های دو کیلومتری تقسیم بندی و نمونه برداری به صورت سیستماتیک و تصادفی انجام شد. ۱۴۲ نمونه خاک از عمق صفر تا ۵ سانتیمتری خاک سطحی جمع آوری شد و پس از هوا خشک شدن، کوبیدن و عبور

از الک دو میلی متری جهت اندازه گیری شوری و عناصر محلول خاک و طیف سنجی به آزمایشگاه منتقل شد. هدایت الکتریکی عصاره اشباع نمونه ها (EC) (جدول ۱) از مخلوط ۱:۵ خاک و آب با دستگاه شوری سنجی DDS-11A در دمای ۲۵ درجه سانتی گراد اندازه گیری شد (Zeng et al., 2018).

جدول ۱. توصیف آماری شوری نمونه های خاک

	آماره						
	بیشترین	میانه	کمترین	میانگین	ضریب تغییرات	انحراف معیار	چولگی
گروه واسنجی	۹۶/۵۵	۲۹/۵۸	۲/۸	۳۴/۱۴۰	۰/۵۹	۲۰/۱۴	۰/۵۹
گروه اعتبار سنجی	۸۴/۳۶	۳۳/۴۴	۱/۶	۳۵/۰۸	۰/۴۴	۱۷/۴۹	۰/۵۹



شکل ۱. موقعیت منطقه مورد مطالعه در استان خوزستان و کانون های ریزگرد و نقاط نمونه برداری

طیف سنجی

برای طیف سنجی نمونه ها، مقداری خاک به پتری دیش با قطر ۸ سانتی متر و عمق ۲ سانتی متر منتقل شد و قبل از آزمایش به درون اتاق تاریک منتقل شدند. در ابتدای آزمایش با استفاده از یک صفحه سفید (با ضریب انعکاس معلوم و برابر با یک) عمل واسنجی نوری دستگاه طیف سنج انجام گرفت. بازتاب خاک با استفاده از طیف سنج آزمایشگاهی ASD FieldSpec3 (ساخت شرکت Analytical Spectral Devices, Inc) انجام شد. در اندازه گیری از سه نوع آشکارساز VNIR (۱۰۵۰-۳۵۰ نانومتر)، SWIR1 (۱۸۰۰-۱۰۰۰ نانومتر) و طیف سنج SWIR2 (۱۸۰۰-۲۵۰۰ نانومتر) استفاده شد. طیف سنجی با فاصله طول موج یک نانومتر صورت پذیرفت. نمونه خاک در فاصله ۲۰ سانتیمتری و زاویه ۴۵ درجه از لامپ هالوژنی دستگاه قرار گرفت. اندازه گیری طیفی در ارتفاع ۳ سانتی متر از بالای نمونه و در ۱ ثانیه انجام گرفت. برای هر نمونه ۱۰ طیف بازتابی اندازه گیری گردید. از ۱۰ طیف تکراری جهت کاهش نویز میانگین گیری شد و به عنوان طیف اصلی خاک ثبت گردید.

پیش پردازش ها

جهت حذف نویز تصادفی، خطای پس زمینه ای نمونه ها و تغییرات مسیر طیفی از چند روش پیش پردازش برای اصلاح طیف اصلی استفاده شد (Fearn et al., 2009). دو بخش دارای نویز در ابتدا و انتهای طیف ها در محدوده بین ۳۵۰ تا ۴۵۰ و ۲۵۰۰-۲۴۵۰ نانومتر حذف گردیدند. در مواردی نیز دو طول موج در محدوده ۹۰۰ و ۱۷۰۰ نانومتر به دلیل وقفه حاصل از تغییر آشکارساز حذف شدند (Rosset et al., 2009). با استفاده از نرم افزار Unscrambler 10.4 و MATLAB، پنج روش پیش پردازش Savitzky-Golay، مشتق اول با Savitzky-Golay (FD-SG)، مشتق دوم با Savitzky-Golay (SD-SG)، روش نرمال سازی استاندارد (SNV) و روش حذف پیوسته (CR) انجام شد.

الگوریتم Savitzky-Golay با استفاده از رگرسیون خطی حداقل مربعات، یک چند جمله ای در اطراف یک نقطه در طیف برآزش می دهد و با این روش اطلاعات را نرم یا صاف می کند. این روش با انتخاب محدوده صاف کردن انجام می شود. در این روش تعداد نقاط سمت راست و چپ هر نقطه به صورت جداگانه انتخاب

به صورت زیر است:

$$X = TP^T + E \quad (\text{رابطه ۱})$$

$$y = Tq + f \quad (\text{رابطه ۲})$$

$$X = [x_1, x_2, \dots, x_n] \quad (\text{رابطه ۳})$$

$$T = \omega^T X \quad (\text{رابطه ۴})$$

در رابطه (۴)، X: داده‌های طیفی (متغیر مستقل)، y: ویژگی خاک (متغیر وابسته)، T: متغیر پنهانی که ترکیبی خطی و از طیف محاسبه می‌شود، ω مقیاس وزنی و از روش بردارهای ویژه $x'y'x$ محاسبه می‌شود. P: بار طیف و q بار خصوصیت شیمیایی خاک و میزان وابستگی طیف و خصوصیات خاک را به متغیر پنهانی T نشان می‌دهند. E و f مقادیر باقیمانده که نشان‌دهنده نویز یا متغیرهای بی‌ربط در مدل هستند. رابطه (۵) بیانگر رابطه نهایی مدل است (Sjöström et al., 2016; Ji, et al., 1983

$$\hat{y} = \hat{b}_i x_i + \hat{b}_0 \quad (\text{رابطه ۵})$$

ی: ویژگی برآورد شده با مدل، \hat{b}_0 : مرز جداکننده و \hat{b}_i : بردارهای رگرسیونی هستند.

مدل جنگل تصادفی (RF) روشی بر اساس ماشین یادگیری و توانایی حل مسائل دسته‌بندی، رگرسیون و سایر مسائل در زمینه‌های گوناگون است. این مدل قابلیت پیش‌بینی انفرادی و همبستگی بین مقادیر پیش‌بینی شده را دارد. الگوریتم تطبیقی این مدل (Breiman, 1999) سبب کاهش انتخاب گزینشی داده‌ها شده و روشی مؤثر در طبقه‌بندی و رگرسیون است. هرچه مجموعه داده ورودی بزرگ‌تر و روند آن‌ها تصادفی باشد، میزان دقت مدل بیشتر است (Breiman, 2001). این مدل قابلیت برآورد خصوصیات مختلف خاک در بازتاب طیفی با وجود نویز بالا را دارد (Chen & Liu, 2005). رابطه کلی مدل به صورت زیر است:

$$F(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (\text{رابطه ۶})$$

$$D = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (\text{رابطه ۷})$$

در روابط (۶ و ۷) F(x): مدل نهایی یا میانگین درخت‌های تصادفی، D: مجموعه داده‌های اصلی (طیف و EC) ورودی به مدل، B: مجموعه درخت‌های تصمیم که از داده‌های اصلی (D) به روش تصادفی نمونه‌گیری با جایگزینی تعیین می‌شود و T_b : بهترین متغیر تصادفی یا درخت تصادفی که به دفعات انتخاب می‌شود.

مدل شبکه عصبی پرسپترون چند لایه (MLP) سیستمی محاسباتی با تعدادی از عناصر متصل به هم و پاسخ‌های پویای ورودی به خروجی است (Caudill, 1987). مدل شبکه‌های عصبی

می‌شود. آزمون‌های مشتق اول و دوم با نشان دادن جزئیات منحنی یک طیف، روش‌هایی را برای تشخیص نقاط بیشینه و کمینه نسبی فراهم می‌کنند. واریانس نرمال استاندارد (SNV) نوعی تغییر ردیفی و جهت‌دار است و بر اساس الگوریتم خود، یک داده را در جهتی تغییر می‌دهد که درون یک ستون یا دسته از داده متمرکز شود (CAMO, 1998).

روش حذف پیوستار برای جدا کردن و تحلیل عارضه‌ها در طیف بازتابی به کار گرفته می‌شود و برای داده‌های طیف‌سنج‌های زمینی، به عنوان بخشی از تحلیل‌ها برای توزیع کانی‌ها و گیاهان، در طیف‌های مرجع کتابخانه‌ای استفاده شده است. پیوستار برآوردی از دیگر پدیده‌های جذبی موجود در طیف است که تنها شامل پدیده جذبی موردنظر نیست (Clark, 1999). بنابراین، حذف پیوستار اغلب بر روی همه پدیده‌های جذبی طیف، اجرا شده و برای مطالعه روابط بین عوارض جذبی و ترکیبات بیوشیمیایی و محتوی آب پوشش گیاهی مفید است (Huang et al., 2008; Huber et al., 2004). همچنین حذف پیوستار اثرات نامطلوب خاک زمینه و جذب‌های اتمسفری را روی آنالیز طیفی به حداقل می‌رساند (Kokaly and Clark, 1999). با انجام این تکنیک مقادیر جذب‌های ناخواسته عوامل نامشخص در این منحنی، حذف و یا به حداقل می‌رسد. در این چنین منحنی‌هایی، واریانس مطلق حذف و حداکثرهای جذب بارزتر می‌شوند (Curran et al., 2001).. در واقع روش حذف پیوستار در کل دامنه طیف (۴۵۰-۲۴۵۰ نانومتر) انجام شد و منحنی طیف‌ها بر اساس الگوریتم این روش نرمال گردید. طیف نرمال شده به عنوان یک روش پیش‌پردازش و ورودی به مدل PLSR استفاده شد.

از روش رگرسیون حداقل مربعات جزئی (PLSR) برای تعیین فاکتورهای معنادار در هر دسته از طیف‌های اصلی و پیش‌پردازش‌ها استفاده گردید. اولین فاکتور بیش‌ترین اطلاعات و بیش‌ترین واریانس را نشان می‌دهد. فاکتور دوم حداکثر اطلاعاتی که در فاکتور اول توضیح داده نشده است را بیان می‌کند. در هنگامی که فاکتورهای معنادار در هر گروه انتخاب شوند سایر فاکتورها از فرایند مدل‌سازی حذف می‌شوند. مدل‌های ANN و RF بر اساس همبستگی بین EC خاک و مؤلفه اصلی اجرا گردید.

روش‌های مدل‌سازی (PLSR, ANN و RF)

مدل PLSR دارای قابلیت مدل‌سازی چندین متغیر پاسخ همزمان مستقل و نویزدار است (Wang et al., 2014). این روش توانایی نتیجه‌گیری دارد و می‌تواند رابطه خطی بالقوه بین ویژگی‌های خالص خاک و بازتاب طیف در محدوده VIS-NIR را شبیه‌سازی کند (Farifteh et al., 2007; Nawar et al., 2014). رابطه مدل

نمونه i ، y_i مقدار اندازه‌گیری شده در نمونه i ، n تعداد نمونه‌ها در سری واسنجی و اعتبارسنجی، m تعداد تکرار و \bar{y} میانگین مقادیر اندازه‌گیری شده است. در رابطه (۱۲) SD انحراف معیار نمونه است و در رابطه (۱۳) محاسبه می‌شود.

آماره RPD عملکرد مدل را ارزیابی می‌کند. مقادیر کمتر از ۱/۴، بین ۱/۴-۲ و بیشتر از ۲ به ترتیب نشان‌دهنده ضعیف، قابل قبول و عالی هستند. دقت گروه قابل قبول (مقادیر ۲-۱/۴) با به‌کارگیری روش‌های مختلف واسنجی افزایش می‌یابد؛ اما اگر RPD کوچک‌تر از ۱/۴ باشد مدل، پیش‌بینی قابل قبولی را نشان نداده است (Chang et al., 2001).

نتایج و بحث

در شکل (۲) ضریب رگرسیون PLS بین هدایت الکتریکی و طیف خاک هر طول موج نشان داده شده است. همچنین همبستگی بین هدایت الکتریکی و طیف خاک در حالت‌های طیف اصلی و پنج پیش‌پردازش ارائه شده است. دایره‌های سیاه‌رنگ متغیرهای مهم (طول موج کلیدی) برای پیش‌بینی یک متغیر خاص را نشان می‌دهد. ضرایب همبستگی مثبت و منفی بین خصوصیات خاک و طیف مشاهده شد. بیش‌ترین همبستگی با طیف اصلی در ۴۰۰ و ۲۳۰۰ نانومتر وجود داشت. در پیش‌پردازش مشتق اول بیش‌ترین همبستگی منفی در ۱۸۰۰ و همبستگی مثبت در ۱۹۰۰ مشاهده گردید. همبستگی مثبت در محدوده ۱۹۲۰ و همبستگی منفی در ۱۹۵۰ برای مشتق دوم مشاهده شد. در طیف فیلتر شده با SG همبستگی قوی منفی در نزدیکی ۲۰۰۰ نانومتر و همبستگی مثبت در ۲۳۵۰ نانومتر مشاهده شد. در دسته CR همبستگی قوی منفی در ۵۰۰ و ۲۰۰۰ نانومتر و همبستگی مثبت در ۲۳۵۰ نانومتر نتیجه‌گیری شد. در طیف SNV همبستگی قوی مثبت در ۶۰۰ نانومتر و همبستگی منفی در ۱۰۰۰، ۲۲۰۰، ۲۳۰۰ نانومتر مشاهده گردید.

طول موج کلیدی، طول موجی از طیف است که بیش‌ترین تأثیر را بر خصوصیت خاک نشان می‌دهد. طول موج‌های کلیدی در هر پیش‌پردازش به این صورت مشاهده گردید: در طیف اصلی ۹۰۰-۵۰۰، ۱۹۰۰-۱۴۰۰ نانومتر نشان‌دهنده وجود EC است اما بیش‌ترین همبستگی مثبت در (۱۸۸۵-۱۸۸۲) نانومتر و همبستگی منفی در (۱۹۵۰-۱۹۳۵) نانومتر مشاهده شد. برای مشتق اول: طول موج‌های ۶۰۰، ۹۰۰-۱۴۰۰، ۱۵۰۰-۱۶۵۰، ۱۹۰۰ و ۲۰۰۰ نانومتر دارای همبستگی با EC خاک شناسایی شد ولی بیش‌ترین همبستگی مثبت در (۱۹۳۴-۱۹۳۷) نانومتر و همبستگی منفی در (۱۸۹۶-۱۹۰۱) نانومتر مشاهده گردید. در مشتق دوم محدوده‌های ضعیفی با تراکم کم در ۵۰۰ - ۶۰۰

معمولاً در لایه‌ها سازمان‌دهی می‌شوند. لایه‌ها از تعدادی گره‌های متصل به هم تشکیل شده که دارای یک تابع فعال است (Mohamed et al., 2018). الگوریتم پس‌انتشار یک الگوریتم بسیار رایج است که برای آموزش شبکه MLP مورد استفاده قرار می‌گیرد (Demuth & Beale, 1998). این الگوریتم شامل دو مرحله است. در مرحله اول، محاسبات رو به سمت جلو به طرف لایه خروجی انجام شده و بعد از این که خطاها محاسبه شد، در مرحله دوم خطاهای لایه خروجی با انجام اصلاحات لازم با در نظر گرفتن وزن بر اساس رابطه زیر به سمت عقب انجام می‌شود (Rahmati et al., 2018).

در رابطه (۸) $\Delta w_{ij}(n)$ و $\Delta w_{ij}(n-1)$ به ترتیب افزایش وزن بین گره‌های i ام و j ام، E خطا، و α ضریب مومنتم و ضریب یادگیری (مقدار آن‌ها بین ۰ تا ۱) است. خروجی هر نرون توسط رابطه زیر تعریف می‌شود:

$$\Delta w_{ij}(n) = -\varepsilon \times \frac{\partial E}{\partial w_{ij}} + \alpha \times \Delta w_{ij}(n-1) \quad (\text{رابطه ۸})$$

$$y = f(u) = f\left(\sum_{i=1}^n p_i W_{j,i} + b_j\right) \quad (\text{رابطه ۹})$$

$W_{j,i}$ مقدار وزن اتصال بین نرون j ام لایه مذکور با نرون i ام لایه قبل، b_j وزن مربوط به خطا (Bias) برای نرون j ام، P_i مقدار خروجی از نرون i ام لایه‌ی قبل، a مقدار خروجی از نرون j ام، f تابع آستانه نرون j ام است.

معیارهای ارزیابی

برای ارزیابی دقت مدل‌ها در دو گروه واسنجی (۷۰ درصد داده‌ها) و اعتبارسنجی (۳۰ درصد داده‌ها) از سه شاخص آماری ریشه میانگین مربعات خطا (RMSE)، ضریب تبیین (R^2) و نسبت عملکرد به انحراف (RPD) استفاده شد (Chang et al., 2001; Ji et al., 2016)

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(y_{ij} - \bar{y}_{ij})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (y_{ij} - \bar{y}_{ij})^2}} \quad (\text{رابطه ۱۰})$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - y_i)^2}{mn}} \quad (\text{رابطه ۱۱})$$

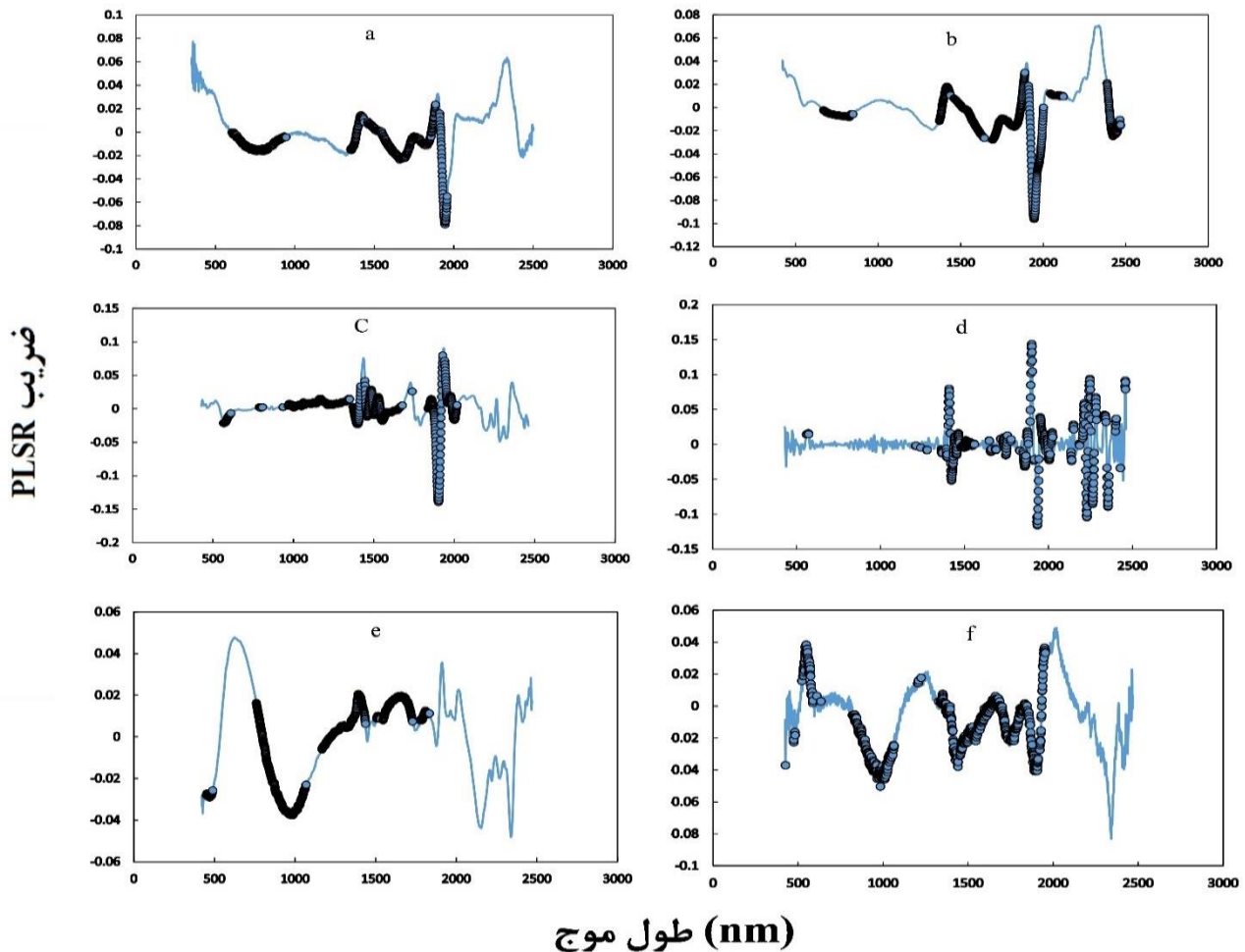
$$PRD = \frac{SD}{RMSE} \quad (\text{رابطه ۱۲})$$

$$SD = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad (\text{رابطه ۱۳})$$

در رابطه (۱۰) \hat{y}_{ij} مقدار برآورد شده برای تکرار j در

پیش‌پردازش CR محدوده‌های با تراکم کم در ۸۰۰-۱۰۰۰، ۱۸۰۰-۱۶۰۰ نانومتر وجود دارد و بیش‌ترین همبستگی مثبت در (۱۵۵۶-۱۵۶۰) نانومتر و همبستگی منفی در (۱۹۱۰-۱۹۱۸) نانومتر مشاهده شد. در پیش‌پردازش SNV طول موج‌های کلیدی محدوده طیفی (۲۲۰۸-۲۲۳۶) نانومتر و (۲۳۴۴-۲۳۴۷) نانومتر وجود دارد. شناسایی ویژگی‌های خاک بر اساس کشش و ارتعاش گروه‌های عاملی آزاد و پیوندی در بازتاب خاک انجام می‌شود.

۸۰۰-۱۱۰۰، ۱۵۰۰-۱۸۰۰ نانومتر وجود دارد و بیش‌ترین همبستگی مثبت در (۱۹۰۱-۱۸۹۶) نانومتر و همبستگی منفی در (۲۳۵۴-۲۳۵۲) نانومتر نتیجه‌گیری گردید. در پیش‌پردازش SG محدوده خالی با تراکم زیاد و فراوانی بیشتر در ۶۰۰-۷۰۰، ۱۹۰۰-۱۴۰۰، ۲۱۰۰، ۲۵۰۰ نانومتر مشاهده شد و بیش‌ترین همبستگی مثبت در (۲۰۱۸۰، ۲۱۸۱، ۲۲۰۵، ۲۲۰۶) نانومتر و همبستگی منفی در (۲۳۴۱-۲۳۴۷) نانومتر مشاهده گردید. در



شکل ۲. همبستگی EC با طیف خاک در هر طول موج. a: طیف اصلی (خام)، b: فیلتر سازی طیف با روش ساویتزی گولای (SG)، c: مشتق اول طیف نرم شده با ساویتزی گولای (FD-SG)، d: مشتق دوم طیف فیلتر شده با ساویتزی گولای (SD-SG)، e: روش واریانس استاندارد نرمال (SNV)، f: روش حذف پیوسته (CR)

اندازه بلور و نظم ساختاری تفاوت‌هایی دارند (Drake, 1995; Farifteh et al., 2008). برای انواعی از نمک‌ها مشاهده شد که Halit (NaCl) دارای خصوصیات جذبی در ۱۴۴۰ و ۱۹۳۳ نانومتر و طول موج‌های ۷۹۳، ۹۹۹، ۱۲۴۰، ۱۴۹۰، ۱۶۳۱، ۱۷۶۰ و ۱۹۴۶ مربوط به نمک Epsomite (MgSo4 7H2O) است (Farifteh et al., 2008). Xu (۲۰۱۶) با استفاده از روش پیش‌پردازش مشتق دوم، خصوصیات جذبی را برای نمک‌های خاک در طول موج ۱۹۰۰، ۱۳۹۰، ۱۴۳۰، ۱۴۷۰، ۱۷۴۰، ۴۴۰، ۲۰۱۰ و ۲۲۷۰ نانومتر نشان داد.

محدوده‌های فعال طیفی در طول موج‌های ۴۶۰، ۵۵۰ و ۷۴۰ بر اثر کشش و ارتعاش گروه‌های عاملی C-H، C-H، C-H، N-H و C-C، ایجاد می‌شود (Xuemei & Jianshe, 2013). ناحیه طیفی ۱۹۰۰ و ۲۲۰۰ نانومتر به علت کشش و ارتعاش ناشی از گروه‌های عاملی Al-OH و Mg-OH است (Pu et al., 2003; Wang et al., 2018). شوری در اثر کاتیون‌ها و آنیون‌های محلول ایجاد می‌شود. هر نمک با توجه منشأ خود اثر متفاوتی را در خصوصیات جذبی خاک نشان می‌دهد (Xu et al., 2016). ویژگی‌های جذبی انواع نمک‌ها به دلایل تغییر در خلوص نمک،

جدول ۲ نتایج معیارهای ارزیابی روش‌های PLSR، PLS-ANN، PLS-RF در مدل‌سازی شوری خاک در گروه‌های واسنجی و اعتبارسنجی

روش‌های مدل‌سازی و پیش‌پردازش		معیارهای ارزیابی					
مدل	پیش‌پردازش	واسنجی (۱۰۳ نمونه)			اعتبارسنجی (۳۹ نمونه)		
		R ²	RMSE	RPD	R ²	RMSE	RPD
PLSR	R	۰/۸۲	۷/۸۷	۲/۳۵	۰/۶۶	۸/۷۴	۱/۷۲
	SG	۰/۸۴	۷/۸۲	۲/۳۷	۰/۷۰	۸/۴۰	۱/۸۱
	FD-SG	۰/۸۴	۷/۸۲	۲/۳۷	۰/۶۷	۸/۷۴	۱/۷۲
	SD-SG	۰/۹۰	۶/۱۴	۳/۱۲	۰/۶۵	۹/۱۵	۱/۵۹
	SNV	۰/۸۵	۷/۸۰	۲/۴۴	۰/۶۷	۸/۷۴	۱/۷۲
	CR	۰/۸۱	۷/۶۶	۲/۱۰	۰/۶۷	۸/۷۷	۱/۵۹
PLS-ANN	R	۰/۸۹	۶/۵۵	۳/۰۰	۰/۸۴	۸/۴۱	۲/۴۷
	SG	۰/۹۱	۵/۹۳	۳/۴۰	۰/۸۶	۷/۳۶	۲/۶۰
	FD-SG	۰/۸۸	۶/۸۲	۲/۸۹	۰/۸۷	۸/۳۹	۲/۵۱
	SD-SG	۰/۸۸	۶/۲۸	۲/۸۷	۰/۸۹	۷/۰۸	۳/۰۴
	SNV	۰/۸۶	۵/۲۳	۲/۶۵	۰/۸۷	۸/۴۵	۲/۶۶
	CR	۰/۸۶	۶/۵۱	۲/۷۱	۰/۸۸	۷/۹۹	۲/۸۷
PLS-RF	R	۰/۸۹	۷/۰۳	۱/۹۸	۰/۷۵	۱۰/۳۶	۱/۳۹
	SG	۰/۸۳	۷/۲۷	۱/۸۸	۰/۷۲	۱۱/۰۲	۱/۲۵
	FD-SG	۰/۸۸	۸/۳۶	۱/۴۰	۰/۷۴	۱۱/۴۸	۱/۰۹
	SD-SG	۰/۸۸	۷/۱۸	۱/۹۵	۰/۸۳	۹/۲۹	۱/۵۹
	SNV	۰/۸۵	۷/۹۲	۱/۸۴	۰/۷۵	۱۰/۵	۱/۳۵
	CR	۰/۸۷	۷/۲۸	۱/۹۵	۰/۸۰	۹/۴۳	۱/۶۳

داد. ضعیف‌ترین عملکرد در مدل PLS-RF و برای پیش‌پردازش SNV ($R^2_{CAL} = 0/84$ ، $RMSE_{CAL} = 7/92$ و $RPD_{CAL} = 0/84$) مشاهده شد.

نتایج مدل گروه اعتبارسنجی در جدول (۱) و شکل (۳) ارائه شده است. در مدل PLSR طیف‌های پیش‌پردازش شده عملکرد قابل‌قبول تا عالی را نشان داد. روش SG ($R^2_{VAL} = 1/81$) و RPD بهترین عملکرد و روش مشتق اول (FD-SG) و (SNV); کمترین عملکرد نسبی ($R^2_{VAL} = 0/67$ ، $RMSE_{VAL} = 8/74$ و $RPD_{VAL} = 1/72$) را داشتند. در مدل PLS-ANN سری اعتبارسنجی همانند گروه واسنجی عملکرد نسبی بهتری را نسبت به دو مدل دیگر نشان داد. در این مدل بیشترین عملکرد مربوط به روش مشتق دوم (SD-SG) ($R^2_{VAL} = 0/89$ ، $RMSE_{VAL} = 7/08$ و $RPD_{VAL} = 3/04$) و کمترین عملکرد نسبی در طیف اصلی ($R^2_{VAL} = 2/47$) و RPD ($R^2_{VAL} = 0/84$ ، $RMSE_{VAL} = 8/41$) بود و تمامی روش‌های طیفی عملکرد عالی را نشان دادند. در مدل PLS-RF دو روش

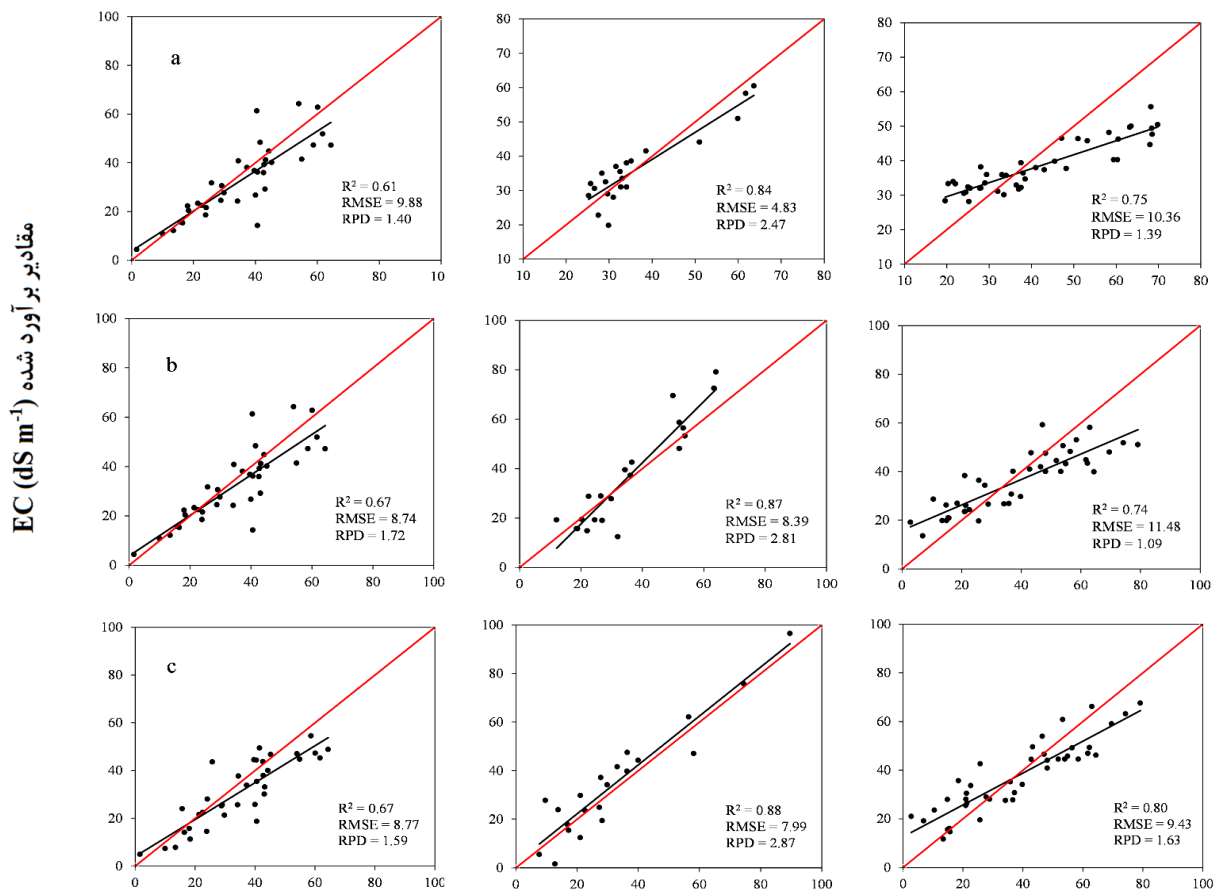
اثر روش پیش‌پردازش بر عملکرد مدل‌سازی

نتایج مدل PLSR با توجه به شش گروه طیف نشان داد که نوع روش‌های پیش‌پردازش در دقت برآورد مدل مؤثر بوده و هر شش روش دقت برآورد خوب تا عالی را نشان دادند (جدول ۱). در مدل PLSR طیف‌های پیش‌پردازش شده دقت بیشتری نسبت به طیف اصلی بدون پیش‌پردازش نشان دادند. در پیش‌پردازش مشتق دوم بیشترین دقت برای (SD-SG) ($R^2_{CAL} = 0/90$ ، $RMSE_{CAL} = 6/14$ و $RPD_{CAL} = 3/12$) و کمترین دقت برای CR ($R^2_{CAL} = 0/66$ ، $RMSE_{CAL} = 8/74$ و $RPD_{CAL} = 2/10$) مشاهده شد. چهار روش پیش‌پردازش SG ($R^2_{CAL} = 0/84$)، SNV ($R^2_{CAL} = 0/85$)، SD- و SG ($R^2 = 0/90$) و FD-SG ($R^2 = 0/84$)، دقت برآورد مدل را افزایش داد و فقط برای روش CR ($R^2 = 0/81$) دقت برآورد مدل کاهش یافت. بیشترین دقت برای مدل PLS-ANN در پیش‌پردازش SG ($R^2_{CAL} = 0/91$ ، $RMSE_{CAL} = 5/39$ و $RPD_{CAL} = 3/40$) به دست آمد. مدل PLS-ANN در طیف اصلی، FD-، SNV و CR دقت بالاتری را نسبت به دو مدل دیگر نشان

(RMSE) (به ویژه در مدل RF) میزان عملکرد کاهش یافته است. استفاده از روش‌های پیش‌پردازش SG, FD-SG, SD-SG و CR و سبب افزایش دقت و عملکرد برآورد مدل PLSR شد. در پیش‌پردازش CR برای EC عملکرد برآورد مدل کاهش یافت. عمل صاف‌سازی طیف در دو روش FD-SG و SD-SG نیز مورد استفاده قرار گرفت اما در SNV و CR انجام نشد. عدم حذف نویز و صاف‌سازی طیف در روش SNV تأثیر معنی‌داری در کاهش دقت برآورد مدل PLSR ایجاد نکرد. در روش CR به ویژه برای EC دقت برآورد کاهش یافت. بنابراین می‌توان نتیجه گرفت که در این پژوهش روش CR نسبت به SNV حساسیت بیشتری به نویز طیف دارد. Weng و همکاران (۲۰۰۸) از مدل PLSR برای برآورد شوری خاک استفاده کردند و عملکرد مدل را با پیش‌پردازش CR افزایش دادند. در پژوهش حاضر نیز پیش‌پردازش CR دقت مناسبی را به خصوص در دو مدل PLS-ANN و PLS-RF داشت ولی در هر دو مدل به مقدار جزئی دقت کمتری نسبت به پیش‌پردازش‌های دیگر حتی طیف اصلی نشان داده است.

پیش‌پردازش مشتق اول و حذف پیوستار دقت قابل قبول ($RPD > 1/4$) را نشان دادند و روش‌های دیگر در برآورد شوری خاک ناتوان بودند ($RPD < 1/4$).

در صورتی که محدوده غلظت خصوصیت خاک گسترده‌تر باشد، دقت برآورد مدل آماری بیشتر می‌شود. پراکندگی بیشتر داده‌ها سبب افزایش SD و RPD می‌شود (Nawar et al., 2016). در این مطالعه نیز EC خاک دارای مقادیر گسترده ($1-6 \text{ ds.m}^{-1}$) و انحراف معیار $19/02$ است. شاخص پراکندگی بالای مقدار EC در دقت برآورد دو مدل PLSR و PLS-ANN تأثیر داشته و دقت را افزایش داده است. اما در مدل PLS-RF به دلیل افزایش خطای برآورد و مقدار RMSE، با وجود پراکندگی و ضریب تبیین بالا، شاخص RPD و عملکرد مدل کاهش یافته است. در این مطالعه با توجه به ضریب تبیین، دقت برآورد سری داده‌های واسنجی مدل‌ها در تمامی روش‌های پیش‌پردازش برای هدایت الکتریکی خاک از $0/8$ بیشتر بود؛ ولی شاخص RPD تا $0/84$ کاهش یافت که نشان‌دهنده عملکرد ضعیف مدل است. کاهش RPD نشان می‌دهد به علت افزایش خطای برآورد و آماره



مقادیر اندازه‌گیری شده EC (dS m⁻¹)

شکل ۳. مدل‌سازی شوری خاک با استفاده از داده‌های اعتبارسنجی (calibration) برازش یافته بر روی مدل‌های PLSR (چپ)، PLS-ANN (وسط)، و PLS-RF (راست). a: طیف اصلی (ROW)، b: مشتق اول طیف فیلتر شده با سوابتزی گولای (FD-SG)، c: روش حذف پیوستار (CR)

نتیجه‌گیری کلی

این مطالعه در خاک‌های شور و سدیمی مناطق حساس به تولید ریزگرد استان خوزستان انجام شد. عملکرد سه مدل PLSR، PLS-ANN و PLS-RF در شش دسته طیف اصلی (ROW)، و پیش‌پردازش‌های SG، SD-SG، FD-SG، SNV و CR در برآورد هدایت الکتریکی EC خاک انجام شد. یافته‌های پژوهش عبارت است از:

بر اساس شاخص ضریب تبیین (R^2) هر سه مدل عملکرد خوبی را در برآورد خصوصیات خاک نشان دادند اما با توجه به شاخص RPD عملکرد مدل PLS-ANN نسبتاً بیشتر بود و مدل PLS-RF کمترین عملکرد را در برآورد EC خاک با روش طیف‌سنجی نشان داد.

عملکرد برآورد EC خاک، بیشتر تحت تأثیر نوع مدل قرار گرفت. در هر مدل یک روش پیش‌پردازش عملکرد بهتری نسبت به سایر روش‌ها نشان داد. در مدل‌های PLSR، PLS-ANN و PLS-RF به ترتیب روش‌های مشتق اول، SG و طیف اصلی بهترین عملکرد را نشان دادند. استفاده از روش‌های پیش‌پردازش طیفی مثل مشتق‌گیری طیفی می‌تواند دقت مدل‌های چند متغیره را افزایش دهد.

استفاده از روش رگرسیون PLS و تعیین مؤلفه‌های معنادار در هر دسته طیفی سبب کاهش حجم محاسبات و افزایش سرعت عمل نرم‌افزار متلب گردید. با وجود اینکه استفاده از روش مؤلفه‌های معنادار در مدل PLS-ANN عملکرد بسیار خوبی را از خود نشان داد اما مدل PLS-RF عملکرد پایین‌تری را نسبت به PLSR نشان داد.

طول موج‌هایی در محدوده ۱۹۰۰، ۱۸۰۰، ۱۹۰۰، ۲۰۰۰، ۲۳۰۰، ۱۵۰۰، ۱۹۰۰ نانومتر تحت عنوان طول موج کلیدی برای EC خاک شناسایی شد. استفاده از طول موج‌های کلیدی در ترکیب با تصاویر ماهواره‌ای و روش سنجش از دور سبب افزایش دقت نقشه شوری خاک می‌شود.

نتایج برآورد سه مدل PLSR, PLS-ANN, PLS-RF

کارایی مدل‌های PLSR، PLS-ANN، PLS-RF در مدل‌سازی هدایت الکتریکی خاک بر اساس آماره‌های ضریب تبیین (R^2)، RMSE و RPD در جدول (۲) نشان داده شده است. در هر پیش‌پردازش هر دو مدل PLSR و PLS-ANN عملکرد بسیار خوب تا خوب را در برآورد خصوصیات خاک از خود نشان دادند. مدل PLS-RF با عملکرد خوب تا ضعیف شناسایی شد. در مقایسه دو سری داده‌های واسنجی و اعتبارسنجی، در اکثر پیش‌پردازش‌ها برای هر مدل، دقت واسنجی نسبت به اعتبارسنجی بیشتر بود.

روش PLS-ANN دارای حداقل اختلاف عملکرد در دوره واسنجی و اعتبارسنجی بود. در مقابل روش PLSR بیش‌ترین اختلاف را نشان داد. در اعتبارسنجی، روش PLS-ANN و پیش‌پردازش SD-SG بهترین عملکرد ($R^2_{\text{Val}}=89$ ، $\text{RMSE}_{\text{Val}}=7.08$ ، $\text{RPD}_{\text{Val}}=3/04$)= 7.08 ، $\text{RMSE}_{\text{Val}}=7/08$ ، $\text{RPD}_{\text{Val}}=3/04$)= 7.08 مدل‌سازی را نشان داد. در طیف اصلی و پیش‌پردازش‌های FD-SG، SNV و CR نیز بهترین عملکرد مربوط به مدل PLS-ANN با ($\text{RPD}_{\text{Val}}=2/3-47/04$)= 7.08 ، $\text{RMSE}_{\text{Val}}=11/02$ و $\text{RPD}_{\text{Val}}=1/25$)= 7.08 ، $\text{RMSE}_{\text{Val}}=0/72$ مشاهده شد. شیب مدل برآورد داده‌ها (PLS-ANN) در طیف اصلی و پیش‌پردازش‌های SG، FD-SG، SD-SG، SNV و CR توزیع مناسبی را در اطراف خط (۱:۱) نشان داد. در واقع شیب مدل PLS-ANN در دو سری داده‌های واسنجی و اعتبارسنجی نسبت به دو مدل دیگر، هم‌راستایی بیشتری با خط (۱:۱) دارد و توزیع داده‌ها در اطراف این خط نشان‌دهنده برتری نسبی این مدل است. با مقایسه دو مدل دیگر در هر پیش‌پردازش، مدل PLSR نسبت به PLS-RF داده‌های سری واسنجی و اعتبارسنجی توزیع بهتری در اطراف خط (۱:۱) دارند. این روند در سری داده‌های اعتبارسنجی در تمام پیش‌پردازش‌های SG، FD-SG، SD-SG، SNV و CR برای EC خاک بارزتر است (شکل ۳).

REFERENCES

- Breiman, L. (1999). Using adaptive bagging to debias regressions: Technical Report 547, Statistics Dept. UCB
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32
- CAMO, A. (1998). The Unscrambler User Manual. CAMO ASA Norway
- Caudill, M. (1987) Neural networks primer, part I. *AI expert*, 2(12), 46-52
- Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2), 480-490
- Chen, X.-W., & Liu, M. (2005). Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics*, 21(24), 4394-4400
- Clark, R.N., 1999. Chapter 1: Spectroscopy of Rocks and Minerals and Principles of Spectroscopy, Manual of Remote Sensing. (A.N. Rencz, ed.) John Wiley and Sons, New York, p 3-58, 1999. (Invited book chapter) www.SID.ir

<http://speclab.cr.usgs.gov>.

- Curcio, D., Ciraolo, G., D'Asaro, F., & Minacapilli, M. (2013). Prediction of soil texture distributions using VNIR-SWIR reflectance spectroscopy. *Procedia Environmental Sciences*, 19, 494-503
- Curran, P.J., Dungan, J.L., Peterson, D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry, testing the Kokaly and Clark methodologies. *Rem. Sens. Environ.* 76: 349-359.
- Demuth, H., & Beale, M. (1998). *Neural network toolbox: For use with MATLAB*, Natick, MA: The Math Works. Inc.. *OpenURL*
- Drake, N. (1995). Reflectance spectra of evaporite minerals (400-2500 nm): applications for remote sensing. *International Journal of Remote Sensing*, 16(14), 2555-2571.
- Farifteh, J., Van der Meer, F., Atzberger, C., & Carranza, E. (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, 110(1), 59-78
- Farifteh, J., Van der Meer, F., Van der Meijde, M., & Atzberger, C. (2008). Spectral characteristics of salt-affected soils: A laboratory experiment. *Geoderma*, 145(3-4), 196-206
- Fearn, T., Riccioli, C., Garrido-Varo, A., & Guerrero-Ginel, J. E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, 96(1), 22-26
- Gomez, C., Lagacherie, P., & Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, 148(2), 141-148
- Hassani, A., Bahrami, H., Noroozi, A., & Oustan, S. (2014). Visible-near infrared reflectance spectroscopy for assessment of soil properties in gypseous and calcareous soils. *Journal of Watershed Engineering and Management*
- He, T., Wang, J., Lin, Z., & Cheng, Y. (2009). Spectral features of soil organic matter. *Geo-spatial Information Science*, 12(1), 33-40
- Huber, S., Kneubuhler, M., Psomas, A., Itten, K., Zimmermann, N.E. 2008. Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. *For. Ecol. Manage.* 256: 491-501.
- Huang, Z., Turner, Brian, J., Dury, Stephen J., Wallis, Ian R. Foley, William, J. 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, 93(1): 18-29
- Iran Meteorological Organization(
<http://www.irimo.ir/far/index.php>)
- Ji, W., Adamchuk, V. I., Biswas, A., Dhawale, N. M., Sudarsan, B., Zhang, Y., . . . Shi, Z. (2016). Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *biosystems engineering*, 152, 14-27
- Ji, W., Li, S., Chen, S., Shi, Z., Rossel, R. A. V., & Mouazen, A. M. (2016). Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil and Tillage Research*, 155, 492-500
- Khayamim, F., Khademi, H., Stenberg, B., & Wetterlind, J. (2015). Capability of vis-NIR Spectroscopy to Predict Selected Chemical Soil Properties in Isfahan Province. *JWSS-Isfahan University of Technology*, 19(72), 81-92.
- Kokaly, R.F. 2011. PRISM: Processing routines in IDL for spectroscopic measurements (installation manual and user's guide, version 1.0): U.S. Geological Survey Open-File Report 2011-1155, 431p.
- Mohamed, E., Saleh, A., Belal, A., & Gad, A. A. (2018). Application of near-infrared reflectance for quantitative assessment of soil properties. *The Egyptian Journal of Remote Sensing and Space Science*, 21(1), 1-14
- Nawar, S., Buddenbaum, H., & Hill, J. (2015). Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: a case study from Egypt. *Arabian Journal of Geosciences*, (7) 8, 5127-5140.
- Nawar, S., Buddenbaum, H., Hill, J., & Kozak, J. (2014). Modeling and mapping of soil salinity with reflectance spectroscopy and landsat data using two quantitative methods (PLSR and MARS). *Remote Sensing*, 6(11), 10813-10834
- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., & Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research*, 155, 510-522
- Pu, R., Ge, S., Kelly, N., & Gong, P. (2003). Spectral absorption features as indicators of water status in coast live oak (*Quercus agrifolia*) leaves. *International Journal of Remote Sensing*, 24(9), 1799-1810
- Rahmati, H., Gholizadeh, S., & Ansari, H. (2018). Estimation runoff of Bara-Ariye basin using WetSpa and artificial neural network models. *journal of geography and planning*, 21(62), 95-115 (In Farsi).
- Rossel, R. V., Cattle, S. R., Ortega, A., & Fouad, Y. (2009). In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*, 150(3-4), 253-266
- Silva, E. B., ten Caten, A., Dalmolin, R. S. D., Dotto, A. C., Silva, W. C., & Giasson, E. (2016). Estimating Soil Texture from a Limited Region of the Visible/Near-Infrared Spectrum *Digital Soil Morphometrics* (pp. 73-87): Springer
- Sjöström, M., Wold, S., Lindberg, W., Persson, J.-Å., & Martens, H. (1983). A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Analytica Chimica Acta*, 150, 61-70
- Dehaan, R., & Taylor, G. (2002). Field-derived spectra of salinized soils and vegetation as indicators of

- irrigation-induced soil salinization. *Remote Sensing of Environment*, 80(3), 406-417
- Wang, J., Ding, J., Abulimiti, A., & Cai, L. (2018). Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS-NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. *PeerJ*, 6e4703
- Wang, J., Li, Z., Qin, X., Yang, X., Gao, Z., & Qin, Q. (2014). *Hyperspectral predicting model of soil salinity in Tianjin costal area using partial least square regression*. Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International
- Weng, Y., Gong, P., & Zhu, Z. (2008). Soil salt content estimation in the Yellow River delta with satellite hyperspectral data. *Canadian Journal of Remote Sensing*, 34(3), 259-270
- Wenjun, J., Zhou, S., Jingyi, H., & Shuo, L. (2014). In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PloS one*, 9(8), e105708
- Xu, C., Zeng, W., Huang, J., Wu, J., & van Leeuwen, W. (2016). Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sensing*, 8(1), 42
- Xuemei, L., & Jianshe, L. (2013). Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration. *Measurement*, 46(10), 3808-3814
- Zeng, W., Zhang, D., Fang, Y., Wu, J., & Huang, J. (2018). Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data. *Journal of Applied Remote Sensing*, 12(2), 022204