

مدل‌سازی آماری شوری خاک در پهنه‌های گسترده

یوسف هاشمی نژاد^۱ - مهدی همایی^{۲*} - علی اکبر نوروزی^۳

تاریخ دریافت: ۱۳۹۵/۰۵/۰۴

تاریخ پذیرش: ۱۳۹۵/۱۲/۱۷

چکیده

شور شدن خاک‌ها در جهان به گونه‌ای روزافزون روبه گسترش است و در نتیجه تولید محصولات کشاورزی در مواجهه با این تنش کاهش می‌یابد. سیاست‌گذاران و تصمیم‌سازان در راستای برنامه‌ریزی برای تطبیق با تغییرات اقلیمی و افزایش نیاز به غذا نیازمند پایش کمی مستمر شوری خاک می‌باشند. شاخص‌های طیفی حاصل از سنجنده‌های ماهواره‌ای و یا سنجنده‌های نزدیک به سطح زمین به‌طور روزافزونی برای پایش شوری خاک مورد استفاده قرار می‌گیرند به نحوی که تا کنون تعداد زیادی شاخص برای پایش شوری خاک معرفی شده‌اند. برای مدل‌سازی و سنجش اعتبار مدل حاصله روش‌های رگرسیونی مختلفی مورد استفاده قرار گرفته که مهم‌ترین آن‌ها رگرسیون خطی چندگانه (شامل رگرسیون گام‌به‌گام، انتخاب رو به جلو و حذف رو به عقب) و رگرسیون حداقل مربعات جزئی است. در این پژوهش به منظور ارزیابی این دو روش در مدل‌سازی تغییرات شوری خاک از اندازه‌گیری‌های آزمایشگاهی و الکترومغناطیسی شوری خاک مربوط به ۹۷ نقطه در سال ۱۳۹۲ و ۲۲۵ نقطه در سال ۱۳۹۳ در بخشی از دشت سبزوار - داورزن به مساحت حدود ۵۰ هزار هکتار استفاده شد. تعداد ۲۳ شاخص طیفی از تصاویر ماهواره لندست ۸ مربوط به تاریخ‌های نمونه‌برداری استخراج و به همراه مدل رقومی ارتفاع به‌عنوان متغیر مستقل مورد استفاده قرار گرفت. روش‌های مختلف رگرسیون خطی چندمتغیره با استفاده از داده‌های سال اول به‌عنوان آموزش و سال دوم به‌عنوان آزمون و بالعکس هرچند ضریب تبیین بین حدود ۲۲ تا ۸۸ درصد ایجاد کرد، ولی این همبستگی در دسته اعتبار سنجی از ۲۹ درصد تجاوز نکرد. به علت وجود هم‌راستایی خطی چندگانه در بین متغیرهای مستقل روش رگرسیون خطی چندگانه برای تمام متغیرها قابل کاربرد نبود. حذف متغیرهای دارای هم‌راستایی خطی، تبدیل لگاریتمی و تصادفی کردن کل داده‌ها در دو دسته آموزش و آزمون، ضریب رگرسیون مدل و اعتبار آن را به‌طور قابل قبولی افزایش داد. استفاده از رگرسیون حداقل مربعات جزئی با استفاده از داده‌های اصلی و تبدیل لگاریتمی شده سال اول و دوم به‌عنوان آموزش و آزمون و بالعکس نیز در دسته آموزش ضریب تبیین بین ۳۹ تا ۸۵ درصد ایجاد کرد، ولی از برآورد در دسته آزمون ناتوان بود. تصادفی کردن داده‌ها و تقسیم مجدد آن‌ها به دو دسته آموزش و آزمون موجب ارتقای چشمگیر ضریب تبیین در دسته اعتبارسنجی شد. تکرار عملیات تصادفی کردن نشان داد که روش از ثبات لازم برای برآورد ضرایب متغیرها برخوردار است.

واژه‌های کلیدی: اعتبار سنجی، رگرسیون خطی چندمتغیره، رگرسیون حداقل مربعات جزئی، سنجش‌ازدور، شاخص‌های طیفی

مقدمه

(۳۴۰ میلیون هکتار) از اراضی تحت کشت دنیا شور، و ۳۷ درصد (۵۶۰ میلیون هکتار) از این اراضی سدیمی باشند (۳۳). در واقع این تخمین‌ها در بهترین حالت برآوردی تجربی هستند چراکه به دلیل ذات پیچیده و پویای زمانی و مکانی شوری خاک هیچ سیاهه قابل اعتمادی از آن وجود ندارد (۳۰). نقشه‌های مقیاس منطقه‌ای موجود عموماً یا کیفی هستند و یا غیرقابل اتکا (۱۸ و ۱۹) و در نتیجه اطلاعات مفید اندکی را در اختیار تولیدکنندگان، مدیران، مروجان، سیاست‌گذاران و تصمیم‌سازان قرار می‌دهد. برای تصمیم‌سازان در مواجهه با تغییر الگوهای اقلیمی و افزایش نیاز به غذا لازم است که روش‌های کمی پایش شوری خاک در مقیاس منطقه‌ای و ملی ایجاد شود تا به آن‌ها اطلاعات موردنیاز برای انجام توصیه‌های جایگزین را

شور شدن خاک باروری خاک‌ها را کاهش می‌دهد (۱۱). اصلاح محلول یا از طریق افزایش مکش اسمزی خاک، یا از طریق سمیت ویژه یون‌ها و یا از طریق اختلال در فرآیندهای متابولیکی گیاه باعث کاهش عملکرد گیاه می‌شوند (۱۴). تخمین زده می‌شود که ۲۳ درصد

۱ - دانشجوی دکتری گروه خاکشناسی، دانشگاه تربیت مدرس

۲ - استاد گروه آبیاری و زهکشی، دانشگاه تربیت مدرس

* - نویسنده مسئول: (Email: mhomaee@modares.ac.ir)

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

ارائه نماید (۳۰).

بخش کشاورزی ایران به عنوان هیجدهمین کشور بزرگ دنیا تولیدکننده ۲۰ درصد تولید ناخالص ملی است درحالی که چالش‌های بزرگی فراروی آن قرار دارد. علاوه بر خشکی که ۹۰ درصد مساحت کشور را در بر گرفته است از خشکسالی، شوری و فرسایش خاک به عنوان مهمترین معضلات کشاورزی ایران یاد می‌شود (۲۶). هرچند که شوری مفهومی وابسته به گیاه است ولی مقدار آن در زمان، مکان و عمق خاک در منطقه ریشه بسیار متغیر است (۱۵).

ارزیابی شوری به‌طور مرسوم از طریق تفسیر نتایج تجزیه آزمایشگاهی بر روی نمونه‌های خاک انجام می‌شود. لیکن این روش - های ارزیابی خاک نیازمند صرف وقت و هزینه زیادی می‌باشند (۳۶). خوشبختانه در دهه‌های اخیر به دلیل پیشرفت‌های به‌وجود آمده در فناوری سنجش‌ازدور، پیشرفت‌های فراوانی در این زمینه ایجاد شده است که در نتیجه امکان ارزیابی سریع‌تر و به‌صرفه‌تر شوری خاک را فراهم آورده است. در واقع، از دهه ۱۹۷۰ برخی پژوهشگران (۵، ۱۳، ۲۳، ۲۷، ۲۰، ۲۱، ۳۱، ۴، ۷، ۹، ۲، ۲۹، ۶، ۱۰، ۲۵، ۳، ۳۲) رابطه بین شوری خاک و ویژگی‌های طیفی و رادار را بررسی کرده‌اند. در نتیجه این تلاش‌ها روابط مختلفی بین شوری خاک با اندیس‌های مختلف مستخرج از تصاویر به‌دست آمده است. برای ارتباط دادن این اندیس‌ها با شوری خاک معمولاً از روش‌های رگرسیونی استفاده می‌شود.

روش مرسوم مدل کردن ماتریس Y (معمولاً بردار شوری خاک) در مقابل ماتریس X (متغیرهای وابسته) از طریق روش رگرسیون چندگانه خطی^۱ (MLR) بوده ولی این روش تا زمانی درست عمل می‌کند که تعداد متغیرهای مستقل کم و کاملاً غیروابسته باشند. در نتیجه استفاده از این فن برای مدل کردن شوری خاک در مقابل اعداد دیجیتال باندهای مختلف تصاویر ماهواره‌ای و نیز شاخص‌های گوناگون مستخرج از آن‌ها که عمدتاً دارای همبستگی داخلی با یکدیگر نیز هستند با محدودیت روبرو خواهد شد.

روش رگرسیون جزئی حداقل مربعات^۲ (PLSR) روش نسبتاً جدیدی است که در دهه ۸۰ و ۹۰ میلادی توسعه یافته است و این قابلیت را دارد که تعداد زیادی فاکتور پاسخ مختلف را که شدیداً با یکدیگر هم‌راستایی خطی دارند^۳ و نیز داده‌های مغشوش^۴ و حتی نامتوازن هم در ماتریس X و هم در ماتریس Y را به‌صورت کارآمدی در مدل حفظ کند. روش رگرسیون جزئی حداقل مربعات (PLSR)^۵ دارای این قابلیت مطلوب است که دقت پارامترهای مدل با افزایش تعداد متغیرهای همبسته و نیز تعداد مشاهدات افزایش می‌یابد (۳۵).

روش‌های مبتنی بر رگرسیون متعددی برای تخمین شوری خاک به کار رفته که روش رگرسیون جزئی حداقل مربعات کارآمدترین روش آن است (۲۴). روش رگرسیون جزئی حداقل مربعات دارای قابلیت‌هایی است که برای مدل کردن رابطه خطی محتمل بین شوری خاک و طیف انعکاسی مفید واقع می‌شود. از این روش برای تخمین برخی خصوصیات شیمیایی خاک به کمک طیف‌سنجی استفاده شده است (۳۴). نتایج پژوهشی که در آن دو روش رگرسیون جزئی حداقل مربعات و شبکه عصبی مصنوعی را برای مدل کردن شوری خاک استفاده شد (۸) حاکی از همبستگی مناسب مدل ایجاد شده برای برآورد شوری خاک در سه مقیاس کوچک (آزمایشگاهی)، متوسط (مزرعه) تا بزرگ (تصویر ماهواره‌ای) بود ($R^2 > 0.81$). از این فن برای ارزیابی شوری خاک به کمک تلفیق داده‌های سنجش شده از نزدیک سطح زمین و سنجش شده از دور در غرب تگزاس ایالات متحده آمریکا استفاده شده (۱) و روش تلفیقی جدیدی معرفی شده است. با جمع‌بندی بررسی منابع مشخص شد که روش رگرسیونی مرسوم در برآورد شوری خاک در مقیاس وسیع چندان موفق نبوده‌اند، روش رگرسیون جزئی حداقل مربعات نیز عمدتاً برای مطالعات طیف‌سنجی و در مقیاس آزمایشگاهی استفاده شده است. در نتیجه این پژوهش با هدف بررسی روش‌های مختلف مدل‌سازی تغییرات شوری در مقیاس وسیع و با استفاده از تصاویر ماهواره‌ای انجام شد.

مواد و روش‌ها

محدوده مورد مطالعه بخشی از دشت سبزوار - داورزن به وسعت تقریبی ۵۰ هزار هکتار در جنوب غربی شهرستان سبزوار و حاشیه رودخانه فصلی و شور کالشور است. برای تعیین موقعیت نقاط نمونه - برداری خاک از روش فرامکعب لاتین (میناسنی و مک براتنی^۶، ۲۰۰۶) استفاده شد. در این روش، از توزیع احتمال تجمع‌ی عوامل کمکی برای تعیین مناسب‌ترین نقاط نمونه‌برداری استفاده می‌شود. بر این اساس موقعیت ۱۰۰ نقطه نمونه‌برداری برای سال اول تحقیق تعیین شد. موقعیت نقاط بر روی سامانه موقعیت یاب جهانی (GPS) منتقل و سپس موقعیت نقاط در منطقه دقیقاً مشخص و اندازه‌گیری‌های مربوطه انجام شد. از ۱۰۰ نقطه سال اول ۸۶ نقطه هم برای نمونه‌برداری و هم قرائت دستگاه EM38 انجام شد، ۱۵ نقطه به دلیل خشکی زیاد خاک قابل نمونه‌برداری نبودند و برای آن‌ها فقط قرائت دستگاه EM38 انجام شد. برای قرائت دستگاه EM38 پس از کالیبراسیون مزرعه‌ای دستگاه در محل، دستگاه در وضعیت افقی^۷ روی سطح خاک قرار داده شد و اندازه‌گیری انجام گرفت.

- 1- Multivariate Linear Regression (MLR)
- 2- Partial least square Regression (PLSR)
- 3- Colinearity
- 4 - Noisy
- 5- Partial Least Squares Regression

6- Minasny and McBratney

7- Horizontal Orientation

۳۰ متر و باند ۸ دارای قدرت تفکیک مکانی ۱۵ متر است. باندهای ۱۰ و ۱۱ از سنجنده TRIS دارای قدرت تفکیک مکانی ۱۰۰ متر هستند ولی با نمونه‌برداری مجدد^۴ به صورت ۳۰ متری ارائه می‌شوند.^۵

شاخص‌های طیفی

شاخص‌های طیفی استفاده‌شده در این تحقیق شامل ۱۱ باند تصویر لندست ۸، ۶ شاخص شوری، رطوبت و روشنایی (S11, S12, PCA1, PCA2, SI3, NDMI, NDSI, BI) مؤلفه اصلی (PCA3) شاخص انتقال طیفی (Tasseled Cap 1, 2, 3) به همراه مدل رقومی ارتفاع^۶ (DEM) بود (۲۵). به این ترتیب در مجموع ۲۴ شاخص طیفی از تصویر ماهواره‌ای لندست ۸ با استفاده از نرم‌افزار ARC GIS برای نقاط نمونه‌برداری استخراج شد.

مدل‌سازی آماری

شاخص‌های طیفی استخراج‌شده به‌عنوان متغیرهای مستقل در مقابل مقادیر متوسط شوری عصاره اشباع خاک (EC_e) تا عمق ۹۰ سانتی‌متری و همچنین قرائت دستگاه EM38 در وضعیت افقی (EMh) به‌عنوان متغیرهای وابسته برای مدل‌سازی آماری مورد استفاده قرار گرفتند. مدل‌های آماری مورد استفاده شامل رگرسیون چندمتغیره خطی (MLR) و رگرسیون حداقل مربعات جزئی بودند (PLSR) و در نهایت مناسب‌ترین مدل انتخاب و ارزیابی شد.

نتایج و بحث

نتایج مدل‌سازی و اعتبار سنجی به سه روش رگرسیون خطی چندگانه شامل گام‌به‌گام^۷ انتخاب رو به جلو^۸ و حذف رو به عقب^۹ در جدول‌های (۱) و (۲) نشان داده شده است. در این روش‌ها مرحله‌به‌مرحله، تأثیر حذف و یا اضافه کردن متغیرها بر مدل رگرسیون خطی چندگانه بررسی می‌شود و در صورت بهبود رگرسیون، متغیر مربوطه در مدل حفظ می‌شود.

در این جدول‌ها، یک‌بار داده‌های سال اول برای استخراج (آموزش) مدل و سال دوم برای اعتبارسنجی (آزمون) مدل استفاده شده‌اند (جدول ۱) و یک‌بار عکس این عمل انجام شده است (جدول ۲).

نمونه‌های خاک در فواصل ۳۰ سانتی‌متری تا عمق ۹۰ سانتی‌متری برداشت شدند. نمونه‌ها به آزمایشگاه منتقل و پس از خشک و الک کردن هدایت الکتریکی، درصد شن، سیلت، رس و رطوبت خاک اندازه‌گیری شد. هدایت الکتریکی در عصاره اشباع از طریق تهیه گل اشباع و استخراج عصاره پس از ۲۴ ساعت به وسیله پمپ مکش (۲۸) اندازه‌گیری شد. درصد ذرات شن، سیلت و رس به روش هیدرومتری و رطوبت خاک از طریق خشک کردن در دمای ۱۰۵ درجه سانتی‌گراد به مدت ۲۴ ساعت اندازه‌گیری شد (۱۷).

رابطه موجود بین شوری عصاره اشباع خاک و قرائت دستگاه EM38 به دست آمد. روابط حاصله با استفاده از رگرسیون چند متغیره به کمک متوسط درصد رس و متوسط رطوبت خاک اصلاح شده و ارتقا یافتند.

نمونه‌برداری سال دوم با تراکم ۲۵ نقطه انجام شد. به طریق مشابه نقاط نمونه‌برداری با استفاده از فن فرامکعب تعیین و با استفاده از GPS موقعیت نقاط پیدا شدند. در هر موقعیت نمونه‌برداری در سال دوم افزون بر موقعیت مرکز نقطه که هم قرائت دستگاه EM38 و هم نمونه‌برداری خاک انجام می‌شد، در هشت نقطه در اطراف مرکز نیز فقط قرائت دستگاه EM38 انجام می‌شد. این هشت نقطه در امتداد چهار جهت جغرافیایی اصلی و در فواصل ۳۰ و ۶۰ متری نسبت به نقطه مرکز قرار داشتند. در مجموع در سال دوم تعداد ۲۲۵ نقطه قرائت وجود داشت که از ۲۵ نقطه در سه عمق نمونه‌های خاک اخذ شدند. با استفاده از روابط استخراج‌شده از سال اول شوری خاک در نقاط فاقد نمونه خاک برآورد شده و سپس رابطه بین مقادیر شوری و مقادیر قرائت دستگاه EM38 تعیین شد.

تصاویر ماهواره‌ای

در این پژوهش از تصاویر سنجنده^۱ OLI و سنجنده^۲ TRIS از ماهواره لندست ۸ در ردیف ۳۵ و گذر ۱۶۱ از آرشیو تارنمای سازمان زمین‌شناسی ایالات متحده آمریکا^۳ (USGS) استفاده شد. سنجنده OLI شامل باندهای قدیمی خالص سازی شده به همراه سه باند جدید یعنی باند آبی تیره برای مطالعات مناطق ساحلی/ریزگردها، یک باند مادون قرمز موج کوتاه برای تعیین سیروس و یک باند ارزیابی کیفیت است. سنجنده TRIS دو باند حرارتی را ارائه می‌دهد. این دو سنجنده اعداد دیجیتال را در فضای ۱۲ بیتی (۴۰۹۶ سطح خاکستری مختلف) در مقایسه با فضای هشت بیتی قبلی (۲۵۶ سطح خاکستری) ارائه می‌کنند. تصاویر به یک فضای ۱۶ بیتی ارسال می‌شوند و در نتیجه ۵۵۰۰۰ سطح خاکستری مختلف به صورت بالقوه ممکن است موجود باشد. باندهای سنجنده OLI به‌جز باند ۸ دارای قدرت تفکیک مکانی

4- Resampling
5-http://landsat.usgs.gov/band_designations_landsat_satellite.php
6- Digital Elevation Model
7- Stepwise
8- Forward Selection
9- Backward Elimination

1- Operational Land Imager (OLI)
2- Thermal Infrared Sensor (TRIS)
3- United States Geological Survey (USGS)

جدول ۱- نتایج مدل‌سازی (داده‌های سال اول) و اعتبارسنجی شوری (داده‌های سال دوم) عصاره اشباع (EC_e) و القای الکترومغناطیس (EM_h) در مقابل متغیرهای مستقل به روش رگرسیون چندگانه خطی

Table 1- Modeling (first year data) and verification (second year data) of soil saturated extract salinity (EC_e) and Electromagnetic induction (EM_h) versus independent variables using multiple linear regressions.

روش Method	متغیر وابسته Dependent Variable	R^2	R^2_{adj}	مدل Model	R^2 اعتبارسنجی R^2 verification
گام به گام Stepwise	EC_e	24.83	22.41	$EC_e = -1052 + 0.1894 b_1 - 0.1714 b_3 + 0.0269 PCA_1$	29.93
	EM_h	31.56	28.58	$EM_h = -7511 + 1.949 b_2 - 1.960 b_3 - 0.233 b_6 + 0.357 PCA_1$	24.17
انتخاب رو به جلو Forward Selection	EC_e	39.12	32.82	$EC_e = -122 - 0.070 b_1 + 0.187 b_2 - 0.179 b_3 - 0.1247 b_6 - 0.0283 b_8 + 0.2739 b_{10} - 0.3233 b_{11} + 0.1339 PCA_1 - 0.0394 SI_2$	15.48
	EM_h	49.42	42.87	$EM_h = -26088 - 0.824 b_1 + 1.88 b_2 + 10.02 b_3 - 0.706 b_6 + 2.164 b_{10} - 2.771 b_{11} - 0.22 PCA_1 + 1.172 PCA_3 + 19.42 BI - 0.000059 SI_1 - 20.42 SI_2$	38.89
حذف رو به عقب Backward Elimination	EC_e	41.76	33.44	$EC_e = -689675 - 13.06 b_1 - 10.11 b_2 - 6.32 b_3 + 16.16 b_5 + 5.37 b_6 - 8.89 b_9 + 0.2313 b_{10} - 0.257 b_{11} + 17.18 PCA_2 + 19.00 PCA_3 - 0.765 SI_2 + 3.48 SI_3$	2.06
	EM_h	53.79	46.55	$EM_h = -3428733 - 16.27 b_1 + 183.9 b_5 + 49.6 b_6 - 25.62 b_9 + 2.118 b_{10} - 2.470 b_{11} - 127.6 PCA_1 + 33.65 PCA_2 + 157.0 PCA_3 + 52.1 BI + 120055 NDSI - 61.4 SI_2 + 93.8 SI_3$	0.23

جدول ۲- نتایج مدل‌سازی (داده‌های سال دوم) و اعتبارسنجی شوری (داده‌های سال اول) عصاره اشباع (EC_e) و القای الکترومغناطیس (EM_h) در مقابل متغیرهای مستقل به روش رگرسیون چندگانه خطی

Table 2- Modeling (second year data) and verification (first year data) of electrical conductivity of saturated extract (EC_e) and electromagnetic induction (EM_h) data against independent variables using multiple linear regression method

روش Method	متغیر وابسته Independent variable	R^2	R^2_{adj}	مدل Model	R^2 اعتبارسنجی R^2 verification
گام به گام Stepwise	EC_e	88.74	88.15	$EC_e = -2797 - 0.0926 b_3 + 1.640 b_4 - 0.00896 b_8 + 0.287 b_9 + 0.1106 b_{10} - 0.1146 b_{11} - 0.9541 TC_1 - 1.178 TC_2 + 5127 NDMI - 0.000011 SI_1 - 0.3736 DEM$	3.41
	EM_h	88.74	88.15	$EM_h = -22151 - 0.732 b_3 + 12.96 b_4 - 0.0708 b_8 + 2.269 b_9 + 0.874 b_{10} - 0.906 b_{11} - 7.542 TC_1 - 9.312 TC_2 + 40529 NDMI - 0.000084 SI_1 - 2.954 DEM$	1.64
انتخاب رو به جلو Forward Selection	EC_e	89.02	88.16	$EC_e = -2628 + 31.1 b_1 + 33.0 b_2 + 31.0 b_3 + 53.5 b_4 - 0.00955 b_8 + 0.272 b_9 + 0.1053 b_{10} - 0.1064 b_{11} - 0.000063 PCA_1 - 72.1 TC_1 - 0.628 TC_2 - 21.3 TC_3 + 5077 NDMI - 0.000010 SI_1 - 1.65 SI_3 - 0.3865 DEM$	0.98
	EM_h	89.02	88.16	$EM_h = -20816 + 246 b_1 + 261 b_2 + 245 b_3 + 423 b_4 - 0.0755 b_8 + 2.150 b_9 + 0.833 b_{10} - 0.841 b_{11} - 0.00050 PCA_1 - 570 TC_1 - 4.97 TC_2 - 168.8 TC_3 + 40136 NDMI - 0.000075 SI_1 - 13.01 SI_3 - 3.055 DEM$	11.96
حذف رو به عقب Backward Elimination	EC_e	88.65	87.99	$EC_e = -2774 + 0.0857 b_1 + 0.0848 b_3 - 0.3667 b_5 + 0.1523 b_6 - 0.00990 b_8 + 0.282 b_9 + 0.1132 b_{10} - 0.1206 b_{11} + 0.2404 b_i + 4935 NDMI - 0.000011 SI_1 - 0.3607 DEM$	3.97
	EM_h	88.65	87.99	$EM_h = -21969 + 0.677 b_1 + 0.671 b_3 - 2.898 b_5 + 1.204 b_6 - 0.0783 b_8 + 2.227 b_9 + 0.895 b_{10} - 0.953 b_{11} + 1.901 b_i + 39008 NDMI - 0.000086 SI_1 - 2.851 DEM$	3.05

شناخته‌شده‌ای مانند NDMI و NDSI وجود دارند که از دو یا چند باند تصویر ماهواره‌ای استفاده می‌کنند. ولی این شاخص‌ها نیز هم با خود و هم با باندهای تصویر همبستگی نشان می‌دهند. با بررسی ماتریس همبستگی تعداد شش متغیر مستقل که فاقد هم‌راستایی خطی بودند انتخاب شدند. برای نرمال کردن توزیع فراوانی از متغیرهای فاقد مقادیر منفی لگاریتم در مبنای ۱۰ گرفته شد. این متغیرها شامل لگاریتم اعداد دیجیتال باندهای ۲، ۹، ۱۱ و مدل رقومی ارتفاع به همراه دو شاخص NDMI و NDSI بودند که رابطه آن‌ها با لگاریتم EC_e و لگاریتم EMh بررسی شد.

با توجه به این که همبستگی بین متغیرهای مستقل و لگاریتم EMh بیش از لگاریتم EC_e بود در این بخش تنها به بررسی و صحت سنجی روابط در مورد این متغیر وابسته پرداخته می‌شود. جدول (۴) نتایج رگرسیون گام‌به‌گام بین شش متغیر مستقل باقیمانده و لگاریتم EMh را نشان می‌دهد.

در حالی که داده‌های سال نخست برای مدل‌سازی و سال دوم برای اعتبار سنجی به کار رفتند رگرسیون در هر دو دسته اندک بود. در حالی که داده‌های سال دوم برای مدل‌سازی و سال نخست برای اعتبار سنجی به کار رفتند، ضریب رگرسیون مدل قابل قبول بود لیکن در دسته اعتبار سنجی فاقد اعتبار بودند. در حالی که کل داده‌های حاصل از دو سال به صورت تصادفی به ۲۱۲ نقطه مدل‌سازی و ۱۰۶ نقطه اعتبار سنجی تقسیم شدند، مقداری از قدرت رگرسیون در دسته مدل‌سازی کاسته شده و به‌طور فراوانی اعتبار مدل در دسته اعتبار سنجی افزایش یافت. تکرار عملیات تصادفی کردن نیز نتایج کمابیش مشابهی از لحاظ ضرایب متغیرها و نیز ضرایب رگرسیون ایجاد کرد.

جدول (۵) نتیجه مدل‌سازی به روش حداقل مربعات جزئی را نشان می‌دهد. این جدول حاوی ضرایب متغیرهای مستقل در مدل رگرسیون حداقل مربعات جزئی در دو حالت است که متغیر وابسته شوری عصاره اشباع خاک (EC_e) و یا قرائت القای الکترومغناطیس در وضعیت افقی (EMh) باشد. همچنین، ضرایب همبستگی (R^2)، همبستگی برآورد (R^2 - pred) و R^2 اعتبار سنجی برای هر دو حالت در جدول درج شده است.

در این جدول، مدل‌های ۱ و ۲ به ترتیب نشان‌دهنده حالت‌هایی هستند که در آن‌ها نتایج سال اول برای آموزش و سال دوم برای آزمون مدل استفاده شده‌اند و برعکس. در این میان مدل شماره ۲ در مقایسه با مدل شماره ۱ هم بر مبنای EC_e و هم بر مبنای EMh از قوت بیشتری در دسته آموزش برای برآورد شوری خاک برخوردار است. علت این افزایش همبستگی تعداد بیشتر مشاهدات در سال دوم است. ولی هر دو مدل در برآورد شوری خاک در دسته آزمون (اعتبار سنجی) ناتوان بوده‌اند.

همچنین، در این جدول هم مقدار متوسط شوری عصاره اشباع خاک (EC_e) و هم قرائت دستگاه القاگر الکترومغناطیس در وضعیت افقی (EMh) به عنوان متغیر مستقل استفاده شده‌اند. در سال دوم اندازه‌گیری EMh از ۲۲۵ نقطه با توجه به رابطه کالیبراسیون حاصل از سال اول (رابطه ۱) به عدد EC_e تبدیل شد:

$$EC_e = 0.1265EMh + 5.58 \quad R^2 = 0.78 \quad (1)$$

به همین دلیل ضرایب همبستگی در سال دوم برای EC_e و EMh کاملاً مشابه هم است. به‌طور کلی ضرایب همبستگی برای سال دوم بیشتر از سال اول و برای EMh بالاتر از EC_e است. با توجه به این که EMh سنجشی از انعکاس طیف الکترومغناطیس است لذا انتظار می‌رود همبستگی آن با بقیه دامنه طیف انعکاسی که به‌وسیله سنجنده ماهواره‌ای ثبت می‌شود، بیشتر باشد. همچنین، با افزایش تعداد متغیرهای مستقل قدرت همبستگی افزایش می‌یابد.

نکته مهم‌تر در مورد این جدول همبستگی پایین یا بسیار پایین مدل‌های حاصله در دسته اعتبار سنجی است. در بررسی علت این موضوع باید به پیش‌فرض‌های رگرسیون چندگانه خطی توجه شود. یکی از مهم‌ترین این پیش‌فرض‌ها که معمولاً نادیده گرفته می‌شود عدم وجود هم خطی چندگانه^۱ بین متغیرهای مستقل است. برای ارزیابی وجود یا عدم وجود هم خطی چندگانه از چند شاخص استفاده می‌شود. ماتریس همبستگی بین متغیرهای مستقل و نیز شاخص‌های تحمل^۲ و فاکتور تورم واریانس^۳ چند شاخص برای بررسی احتمال وجود هم‌راستایی خطی بین متغیرها است در صورتی که درجه همبستگی بین دو متغیر مستقل بیش از ۷۰ درصد باشد و یا این که شاخص تحمل زیر ۰/۱ و یا فاکتور تورم واریانس بالای ۱۰ باشد احتمال بروز هم‌راستایی خطی افزایش می‌یابد.

جدول ۳ ماتریس همبستگی بین متغیرهای مستقل و شاخص‌های تحمل و تورم واریانس را نشان می‌دهد. بر این اساس، بین متغیرهای مستقل هم‌راستایی خطی چندگانه وجود دارد و در نتیجه نمی‌توان از رگرسیون چندگانه خطی برای مدل کردن همه این متغیرها استفاده کرد. موضوع وجود هم‌راستایی خطی در مورد متغیرهای مستقل در برآورد توابع انتقالی خاک نیز مشاهده شده است. در استفاده از سه پارامتر درصد ذرات شن، سیلت و رس برای اشتقاق توابع انتقالی منحنی رطوبتی خاک و توابع هیدرولیکی غیراشباع خاک مشاهده شده است که بین درصد این ذرات هم‌راستایی خطی وجود دارد (۱۲ و ۱۶). در نتیجه به‌جای استفاده از درصد این ذرات از نسبت بین ذرات شن و سیلت و نیز درصد رس استفاده کردند و به این ترتیب هم‌راستایی حذف شد. در مورد شاخص‌های طیفی، نسبت‌های

1- Multi co-linearity

2- Tolerance

3- Variance Inflation Factor (VIF)

جدول ۳- ماتریس همبستگی بین متغیرهای مستقل و شاخص های تحمل (To) و تورم واریانس (VIF). همبستگی بالای ۰/۷ در صد، شاخص تحمل زیر ۱/۰ و تورم واریانس بالای ۱۰ نشانگر وجود همبستگی خطی است.

Table 3- Correlation matrix of independent variables, tolerance (To) and variance inflation factor (VIF). Correlation higher than 0.7 percent, tolerance less than 0.1 and VIF higher than 10 indicates colinearity.

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	PCA1	PCA2	PCA3	TC1	TC2	TC3	BI	NDMI	NDSI	SI1	SI2	SI3	DEM	
B1	1.00																								
B2	1.00	1.00																							
B3	0.97	0.99	1.00																						
B4	0.94	0.96	0.99	1.00																					
B5	0.89	0.90	0.93	0.93	1.00																				
B6	0.85	0.88	0.94	0.96	0.93	1.00																			
B7	0.78	0.82	0.90	0.94	0.87	0.98	1.00																		
B8	0.94	0.96	0.98	0.98	0.92	0.94	0.91	1.00																	
B9	0.26	0.26	0.28	0.28	0.34	0.34	0.32	0.29	1.00																
B10	0.70	0.66	0.68	0.66	0.72	0.69	0.61	0.66	0.37	1.00															
B11	0.69	0.66	0.67	0.65	0.71	0.68	0.61	0.65	0.36	1.00															
PCA1	-0.38	-0.35	-0.38	-0.38	-0.53	-0.49	-0.43	-0.39	-0.45	-0.83	1.00														
PCA2	-0.62	-0.57	-0.51	-0.47	-0.54	-0.41	-0.30	-0.48	-0.18	-0.78	-0.43	1.00													
PCA3	0.74	0.73	0.68	0.66	0.63	0.57	0.50	0.66	0.15	0.48	0.47	-0.12	-0.68	1.00											
TC1	0.96	0.97	1.00	0.99	0.95	0.96	0.91	0.98	0.30	0.69	0.69	-0.42	-0.52	0.68	1.00										
TC2	-0.97	-0.97	-0.96	-0.92	-0.92	-0.89	-0.82	-0.92	-0.33	-0.73	-0.73	0.48	0.61	-0.70	-0.95	1.00									
TC3	-0.82	-0.84	-0.91	-0.92	-0.98	-0.97	-0.92	-0.90	-0.36	-0.70	-0.70	0.55	0.46	-0.57	-0.93	0.88	1.00								
BI	0.15	0.09	0.00	-0.06	0.20	-0.15	-0.29	-0.03	0.01	0.09	0.09	-0.48	-0.52	0.65	0.99	-0.94	-0.97	1.00							
NDSI	0.15	0.09	0.00	-0.06	0.20	-0.15	-0.29	-0.03	0.01	0.09	0.09	-0.48	-0.52	0.65	0.99	-0.94	-0.97	1.00							
NDMI	0.15	0.09	0.00	-0.06	0.20	-0.15	-0.29	-0.03	0.01	0.09	0.09	-0.48	-0.52	0.65	0.99	-0.94	-0.97	1.00							
SI1	0.95	0.97	0.99	0.99	0.93	0.95	0.91	0.98	0.28	0.68	0.67	-0.38	-0.52	0.66	0.99	-0.94	-0.92	0.97	-0.02	0.00	1.00				
SI2	0.94	0.95	0.98	0.98	0.98	0.96	0.91	0.97	0.32	0.70	0.70	-0.46	-0.52	0.66	0.99	-0.95	-0.96	1.00	0.09	-0.16	0.98	1.00			
SI3	0.95	0.97	1.00	1.00	0.93	0.96	0.92	0.99	0.28	0.67	0.66	-0.39	-0.49	0.67	1.00	-0.93	-0.92	0.97	-0.04	0.01	1.00	0.98	1.00		
DEM	-0.29	-0.29	-0.29	-0.30	-0.31	-0.26	-0.25	-0.30	0.25	0.06	0.07	-0.17	0.06	-0.31	-0.30	0.25	0.29	-0.31	-0.17	0.11	-0.30	-0.31	-0.30	1.00	
To	0.001	0	0	0	0	0	0.01	0.03	0.63	0.00	0.00	0.04	0.02	0.04	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.43	
VIF	1431	2898		7015		2493	108	40	2	1416	1327	23	45	20	1166	1744	90	107	953					2	

جدول ۴- مدل سازی و اعتبار سنجی داده های فاقد هم راستایی و تبدیل شده در مقابل لگاریتم القای الکترومغناطیس (L-EMH)

Table 4- Modeling and verification of non- collinear and transformed data against log- EM (L-EMH)

دسته آموزش Train Set	دسته آزمون Test Set	مدل Model	R ²	R ² _{adj}	R ² دسته آزمون R ² Test Set
سال اول First Year	سال دوم Second Year	L-EMH = -259 + 5.15 L-b2 + 75.2 L-b9 - 13.03 L-dem	38.70	36.72	0
سال دوم Second Year	سال اول First Year	L-EMH = -179.2 + 2.853 L-b2 + 57.1 L-b9 + 6.391 ndmi + 7.054 ndsi - 13.79 L-dem	69.17	68.45	40
تصادفی ۲۱۲ نقطه Randomly 212 points	تصادفی ۱۰۶ نقطه Randomly 106 points	L-EMH = -108.6 + 3.328 L-b2 + 40.5 L-b9 - 3.859 L-b11 + 7.03 ndmi + 7.15 ndsi - 11.83 L-dem	54.54	52.95	63.87
تصادفی ۲۱۲ نقطه Randomly 212 points	تصادفی ۱۰۶ نقطه Randomly 106 points	L-EMH = -168.6 + 3.086 L-b2 + 57.5 L-b9 - 3.433 L-b11 + 5.448 ndmi + 6.34 ndsi - 13.19 L-dem	62.17	61.02	51.56
تصادفی ۲۱۲ نقطه Randomly 212 points	تصادفی ۱۰۶ نقطه Randomly 106 points	L-EMH = -165.6 + 2.533 L-b2 + 56.8 L-b9 - 3.247 L-b11 + 7.664 ndmi + 8.06 ndsi - 12.84 L-dem	61.68	60.56	49.54
تصادفی ۲۱۲ نقطه Randomly 212 points	تصادفی ۱۰۶ نقطه Randomly 106 points	L-EMH = -122.4 + 3.525 L-b2 + 44.7 L-b9 - 4.114 L-b11 + 5.41 ndmi + 6.11 ndsi - 12.32 L-dem	61.22	60.09	49.26
تصادفی ۲۱۲ نقطه Randomly 212 points	تصادفی ۱۰۶ نقطه Randomly 106 points	L-EMH = -159.3 + 2.768 L-b2 + 55.1 L-b9 - 3.610 L-b11 + 6.68 ndmi + 7.35 ndsi - 12.64 L-dem	58.34	57.12	59.66

اعداد آموزش معتبر هستند و در دسته آزمون معتبر نخواهند بود. برای رفع این مشکل نتایج تبدیل لگاریتمی شده دو سال به صورت تصادفی مخلوط شده و دوسوم آن برای آموزش و یک سوم برای آزمون مدل استفاده شد.

مدل ۵ نتیجه رگرسیون حداقل مربعات جزئی به این روش را نشان می دهد. ضریب تعیین مقادیر EC_e و EMh برآورد شده با مقادیر اندازه گیری شده در دسته آزمون به ترتیب حدود ۵۴ و ۶۱ درصد است. در نتیجه به این روش می توان شوری خاک را در نمونه های مجهول برآورد کرد. نگاهی به ضرایب متغیرها در مدل ۵ مشخص می کند که غیر از عرض از مبدأ بیشترین تأثیرات مربوط به باند ۹ و سپس مدل رقومی ارتفاع است. همچنین مدل رقومی ارتفاع با ضریب منفی آشکار شده است که این مورد با واقعیت زمینی تطبیق دارد زیرا شورترین نقاط، در کم ارتفاع ترین مناطق از محدوده مورد مطالعه آشکار شده اند.

شکل (۱) مقدار خطای باقیمانده مدل ۵ در حالتی که متغیر وابسته لگاریتم EC_e و یا لگاریتم EMh است را نشان می دهد. درصد تجمعی نرمال خطای استاندارد شده، هیستوگرام فراوانی خطای استاندارد شده، مقدار خطای باقیمانده به ازای مقادیر مختلف برآزش داده شده و نیز مقدار خطای باقیمانده در مشاهدات مختلف در این شکل نشان داده شده است. بر این اساس در حالتی که متغیر وابسته

یکی دیگر از پیش فرض های رگرسیون چندگانه خطی نرمال بودن توزیع احتمال متغیر مستقل است. بر این اساس هر چند روش های مختلفی برای تبدیل داده های غیر نرمال به نرمال وجود دارد، لیکن معادلات به کار رفته برای نرمال سازی عمدتاً فقط در دامنه اعداد اولیه قابل کاربرد هستند و ممکن است در دامنه اعداد جدیدی که در آینده وارد مدل خواهند شد، معتبر نباشند.

در نتیجه، در این پژوهش برای نرمال کردن داده ها از لگاریتم در مبنای ۱۰ داده استفاده شد. به این ترتیب از همه متغیرها به جز TC2، NDMI و TC3، که حاوی اعداد منفی بودند، لگاریتم در مبنای ۱۰ گرفته شده است.

مدل های ۳ و ۴ در جدول (۵) به ترتیب نشان دهنده حالت هایی هستند که داده های تبدیل شده سال نخست برای آموزش و سال دوم برای آزمون مدل رگرسیون حداقل مربعات جزئی استفاده شده اند و برعکس. بر این اساس هر چند ضرایب همبستگی دسته آموزش در مدل ۳ در مقایسه با مدل یک افزایش یافته است لیکن مجدداً هر دو مدل در پیش بینی شوری خاک در دسته آزمون ناتوان بوده اند.

با توجه به این که دامنه اعداد متغیرهای مستقل و وابسته در دو سال مقداری متفاوت است و به خصوص برای شاخص های طیفی دامنه اعداد وابسته به عوامل خطای متعددی مانند ابرناکی، وجود ریزگرد، اثر سایه و ... است. در نتیجه مدل های حاصله فقط در دامنه

نتیجه گیری کلی

به عنوان نتیجه گیری نهایی، مروری بر نتایج پژوهش نشان می-دهد که به دلیل وجود هم‌راستایی خطی چندگانه بین داده‌های متغیر مستقل امکان استفاده از تمام متغیرها در روش رگرسیون خطی چندگانه وجود ندارد. هرچند با استفاده از این روش ضریب تعیین در داده‌های آموزش مدل به گونه‌ای چشمگیر زیاد است، لیکن مدل حاصله، قادر به برآورد شوری خاک در دسته آزمون نیست. حذف متغیرهای دارای هم‌راستایی خطی، تبدیل لگاریتمی متغیرهای غیرمنفی و نیز تصادفی کردن داده‌ها در دو دسته مدل‌سازی و اعتبارسنجی هم قدرت مدل و هم اعتبار آن را افزایش داد. در مورد رگرسیون حداقل مربعات جزئی نیز، دو اقدام برای بهبود قدرت رگرسیون ضروری به نظر می‌رسد:

EMh است، در مقایسه با EC_e هم مقدار خطای باقی‌مانده کمتر است و هم به توزیع نرمال نزدیک‌تر است.

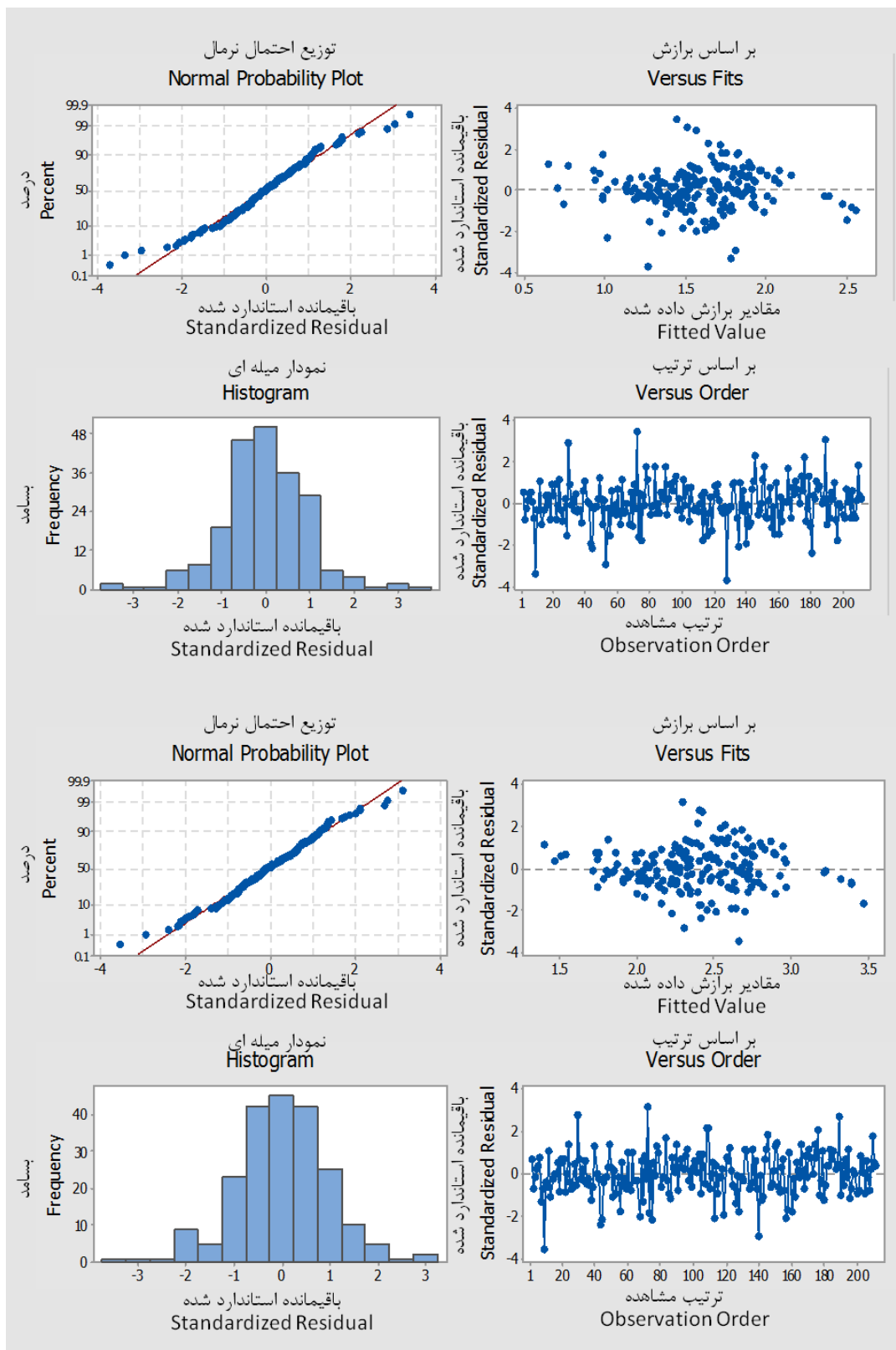
با توجه به این‌که داده‌ها به صورت تصادفی به دو بخش تقسیم شدند، در نتیجه این امکان وجود داشت که در صورت تکرار عملیات تصادفی کردن نتایج متفاوتی هم از لحاظ ضرایب رگرسیون و هم ضریب تعیین به دست آید. در نتیجه، در مدل‌های ۶ تا ۱۰ عملیات تصادفی کردن جداگانه تکرار و داده‌ها به دو دسته آموزش و آزمون تقسیم‌بندی و مدل مربوطه استخراج و در جدول (۶) به صورت خلاصه نشان داده شده است.

در همه این مدل‌ها، شاخص‌های طیفی TC2 و TC3 با ضریب صفر و یا نزدیک به صفر وارد مدل شده‌اند. همچنین، در همه موارد، DEM با ضریب منفی وارد مدل شده است. همچنین، دامنه تغییرات ضرایب مدل‌ها زیاد نیست در نتیجه می‌توان از این ضرایب برای تعیین شوری در نقاط مجهول استفاده کرد.

جدول ۵- نتایج مدل‌سازی و اعتبارسنجی شوری عصاره اشباع (EC_e) و القای الکترومغناطیس (EMh) در مقابل متغیرهای مستقل به روش رگرسیون حداقل مربعات جزئی

Table 5- Modeling and Verification of electrical conductivity of saturated extract (EC_e) and electromagnetic induction (EMh) against independent variables using partial least square regression

تخمینگر Predictor	ضرایب تخمینگر (Predictors Coefficients)									
	پاسخ مدل ۱		پاسخ مدل ۲		پاسخ مدل ۳		پاسخ مدل ۴		پاسخ مدل ۵	
	EC_e	EMh	EC_e	EMh	EC_e	EMh	EC_e	EMh	EC_e	EMh
Constant	499.99	7409.21	-2081.63	-16499.8	-85.291	-114.014	-220.826	-272.043	-176.924	-265.691
B1	0.11	0.83	0.02	0.2	15.697	13.762	1.609	1.822	3.374	3.332
B2	0.08	0.64	0.01	0.1	10.064	11.752	6.151	7.241	2.478	2.374
B3	-0.04	-0.36	-0.01	-0.1	-9.16	-11.394	0.466	0.533	-2.06	-2.22
B4	0	0.02	0	0	2.435	3.854	-0.105	-0.159	-0.393	-0.281
B5	0.03	0.2	0	0	5.091	5.922	-0.411	-0.482	-0.804	-0.833
B6	-0.01	-0.11	0.01	0.1	-4.386	-3.983	-2.148	-2.486	-2.319	-2.436
B7	0.03	0.23	0.02	0.1	10.249	11.016	6.414	7.641	8.837	9.241
B8	-0.04	-0.13	-0.01	-0.1	-3.087	-0.624	-0.275	-0.317	1.149	1.285
B9	-0.13	-0.83	0.37	3	-7.776	11.515	48.348	60.141	45.171	70.663
B10	0.3	2.55	0.15	1.2	108.9	143.985	76.179	90.606	4.205	4.571
B11	-0.37	-3.23	-0.15	-1.2	-120.258	-170.792	-54.67	-64.484	-4.124	-4.019
PCA1	0.01	0.04	0	0	2.323	2.248	-17.884	-21.221	3.097	3.152
PCA2	-0.01	-0.03	0	0	-0.717	3.709	-0.829	-0.963	-0.025	-0.141
PCA3	0.01	0.06	0.01	0	17.568	18.362	5.86	6.936	-0.47	-0.351
TC 1	0	0.03	0	0	0.573	0.701	-0.27	-0.342	-0.596	-0.613
TC 2	0	0.03	-0.02	-0.2	0	0.001	0	0	0	0
TC 3	0.04	0.42	0	0	0	0	0	0	0	0
BI	0	-0.02	0	0	0.617	0.062	-0.617	-0.744	-0.447	-0.418
NDMI	0	0	2237.31	17686.3	15.336	15.919	9.217	10.646	11.141	11.8
NDSI	1172.09	8143.83	1725.43	13639.7	-2.934	-2.066	1.23	1.244	4.201	5.218
SI1	1029.54	8443.04	0	0	-0.961	-0.929	0.064	0.061	-0.547	-0.547
SI2	0	0	0	0	-1.98	-3.127	-0.813	-0.979	-0.785	-0.795
SI3	-0.01	-0.05	0	0	-1.609	-1.516	-0.23	-0.303	-1.066	-1.053
DEM	-0.01	-0.08	-0.39	-3.1	-2.699	-3.771	-8.321	-10.44	-8.447	-11
R^2	39.32	46.90	85.21	85.21	53.44	64.56	82.59	80.85	59.47	62.58
R^2_{Pred}	22.49	29.28	82.95	82.95	39.64	51.18	80.40	78.51	53.01	56.67
R^2 اعتبارسنجی R^2 verification	0	0	0	0	0	0	0	0	53.98	60.83



شکل ۱- ارزیابی توزیع خطای باقیمانده برای وضعیتی که متغیر مستقل لگاریتم ECE (بالا) و یا لگاریتم EMh (پایین) است.
 Figure 1- Distribution of residuals when predictor is \log of ECE (top) and \log of EMh (bottom)

جدول ۶- نتایج مدل‌سازی و اعتبار سنجی شوری عصاره اشباع (EC_e) و القای الکترومغناطیس (EMh) در مقابل متغیرهای مستقل به روش رگرسیون حداقل مربعات جزئی

Table 6- Modeling and Verification of electrical conductivity of saturated extract (EC_e) and electromagnetic induction (EMh) against independent variables using partial least square regression

تخمینگر Predictor	ضرایب تخمینگر (Predictors Coefficients)									
	پاسخ مدل ۶		پاسخ مدل ۷		پاسخ مدل ۸		پاسخ مدل ۹		پاسخ مدل ۱۰	
	EC _e	EMh	EC _e	EMh	EC _e	EMh	EC _e	EMh	EC _e	EMh
Constant	-116.7	-	-	-	-	-	-	-	-	-
B1	4.308	185.522	117.819	170.953	250.833	312.803	135.906	208.304	274.88	224.345
B2	3.274	4.283	4.034	3.659	4.946	4.691	5.182	5.165	3.447	3.469
B3	3.274	3.142	4.189	3.723	3.733	3.455	3.101	3.053	4.326	4.468
B4	-1.08	-1.204	-1.093	-1.061	-1.728	-1.651	-1.917	-1.801	-0.849	-0.863
B5	-0.056	0.189	0.392	0.683	-0.217	0.107	-0.132	0.017	0.091	-0.071
B6	-0.897	-0.86	-0.474	-0.341	-0.801	-0.693	-0.754	-0.713	-0.872	-0.886
B7	-2.672	-2.872	-1.813	-1.823	-2.761	-2.799	-2.629	-2.678	-2.808	-2.729
B8	9.555	10.008	7.096	6.649	9.654	9.291	8.673	8.269	7.864	8.098
B9	-3.272	-3.312	-3.33	-2.953	-2.826	-2.546	0.09	0.274	-3.052	-3.262
B10	30.659	50.164	36.777	52.408	57.78	75.611	27.947	49.436	70.132	54.033
B11	3.924	4.41	1.957	1.901	6.342	6.353	6.175	5.94	6.623	6.565
PCA1	-3.846	-3.552	-6.29	-5.715	-2.549	-2.102	-3.423	-3.23	-2.226	-2.537
PCA2	2.242	2.153	0.807	0.821	4.306	4.292	3.671	3.463	3.065	3.165
PCA3	0.199	0.109	-0.401	-0.46	0.475	0.336	0.42	0.416	0.516	0.561
TC 1	-0.315	-0.321	-0.411	-0.476	0.017	-0.172	-0.128	-0.109	0.09	0.15
TC 2	-0.315	-0.271	0.092	0.196	-0.383	-0.247	-0.529	-0.434	-0.21	-0.266
TC 3	0	0	0	0	0	0	0	0	0	0
BI	-0.636	-0.52	-0.447	-0.233	-0.25	-0.065	-0.418	-0.334	-0.607	-0.681
NDMI	11.608	13.146	9.148	10.134	11.229	12.104	11.462	12.032	11.975	11.396
NDSI	4.929	6.031	4.471	5.209	4.211	5.164	4.046	4.549	5.504	4.746
SI1	-0.243	-0.196	-0.109	-0.015	-0.426	-0.313	-0.437	-0.368	-0.149	-0.201
SI2	-0.803	-0.733	-0.64	-0.451	-0.605	-0.431	-0.815	-0.718	-0.739	-0.806
SI3	-0.522	-0.41	-0.239	-0.029	-0.882	-0.636	-0.87	-0.727	-0.356	-0.47
DEM	-10.11	-12.562	-9.143	-11.55	-9.07	-11.349	-8.969	-11.432	-	-10.274
R ²	59.65	63.21	59.34	63.86	56.51	61.88	63.61	68.17	57.77	61.81
R ² (Pred)	55.07	59.32	53.21	59.90	50.58	57.12	57.38	62.98	51.43	57.74
R ²										
اعتبارسنجی R ² verification	53.47	59.54	52.84	56.89	60.13	60.44	44.06	50.21	53.97	60.14

چشمگیری بر ضریب تعیین در دسته آموزش و آزمون و نیز ضرایب متغیر مستقل نداشته است.

مقایسه بین دو متغیر وابسته متوسط شوری عصاره اشباع خاک (EC_e) و قرائت دستگاه القاگر الکترومغناطیس در وضعیت افقی (EMh) در اکثریت مدل‌های مورد بررسی نشان داد که ضریب تعیین بین متغیرهای مستقل و وابسته در حالتی که EMh به‌عنوان متغیر وابسته استفاده می‌شود افزایش می‌یابد.

تبدیل داده‌ها به منظور نزدیک کردن آن‌ها به توزیع نرمال و نیز لحاظ کردن کل دامنه تغییرات متغیر مستقل در مدل‌سازی برای افزایش قدرت انعطاف مدل در برابر تغییرات این متغیرها. برای تبدیل داده‌ها و نزدیک کردن توزیع آن به توزیع نرمال در این پژوهش از تبدیل لگاریتمی استفاده شد و نیز برای لحاظ کردن کل دامنه تغییرات متغیر در دسته آموزش و آزمون این داده‌ها به صورت تصادفی با هم مخلوط و دو سوم داده‌ها برای آموزش و یک سوم برای آزمون استفاده شدند. نتایج پژوهش نشان داد که تکرار تصادفی کردن تأثیر

1. Aldabaa A.A., Weindorf D.C., Chakraborty S., Sharma A. and Lid B. 2015. Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geoderma*, 239–240: 34–46.
2. Brunner P., Li H.T., Kinzelbach W. and Li W.P. 2007. Generating soil electrical conductivity maps at regional level by integrating measurements on the ground and remote sensing data. *International Journal of Remote Sensing*, 28, 3341–3361.
3. Corwin D.L. and Lesch S.M. 2014. A simplified regional-scale electromagnetic induction — Salinity calibration model using ANOCOVA modeling techniques. *Geoderma*, 230–231: 288–295
4. Douaoui A.E.K., Nicolas H. and Walter C. 2006. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134: 217–230.
5. Driessen P.M. and Schoorl R. 1973. Mineralogy and morphology of salt efflorescences on saline soils in the Great Konya Basin. *Turkey Journal of Soil Science*, 24: 436–442.
6. Eldeiry A.A. and Garcia L.A. 2010. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using Landsat images. *Journal of Irrigation and Drainage Engineering*, 136: 355–364.
7. Farifteh J., Farshad A. and George R.J. 2006. Assessing salt-affected soils using remote sensing, solute modelling, and geophysics. *Geoderma*, 130: 191–206.
8. Farifteh J., van der Meer F., Atzberger C. and Carranza E. 2007. Quantitative analysis of salt affected soil reflectance spectra: a comparison of two adaptive methods (PLSR and ANN). *Remote Sensing and Environment*, 110: 59–78.
9. Fernández-Buces N., Siebe C., Cram S. and Palacio J.L. 2006. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: a case study in the former lake Texcoco, Mexico. *Journal of Arid Environments*, 65: 644–667.
10. Furby S., Caccetta P. and Wallace J. 2010. Salinity monitoring in Western Australia using remotely sensed and other spatial data. *Journal of Environmental Quality*, 39: 16–25.
11. Ghassemi F., Jakeman A.J. and Nix H.A. 1995. Salinisation of land and water resources: Human causes, management and case studies. University of New South Wales Press, Sydney, Australia.
12. Ghorbani Dashtaki S., Homae M. and Khodaverdiloo H. 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. *Soil Use and Management*. 26(1): 68-74.
13. Golovina N.N., Minskiy D., Pankova Y. and Solovyev D.A. 1992. Automated air photo interpretation in the mapping of soil salinization in cotton-growing zones. *Mapping Sciences and Remote Sensing*, 29: 262–268.
14. Homae M., Feddes R.A. and Dirksen C. 2002. A macroscopic water extraction model for nonuniform transient salinity and water stress. *Soil Science Society of America Journal*, 66 (6): 1764- 1772.
15. Homae M. and Schmidhalter U. 2008. Water integration by plants root under non-uniform soil salinity. *Irrigation Science*, 27(1):83-95.
16. Khodaverdiloo H., Homae M. van Genuchten M.T. and Ghorbani Dashtaki S. 2011. Deriving and validating pedotransfer functions for some calcareous soils. *Journal of Hydrology*, 399(1): 93-99.
17. Klute A. 1986. Methods of soil analysis. Part 1. Physical and mineralogical methods. CAB Direct. 1188 pp.
18. Lal R., Iivari T. and Kimble J.M. 2004. *Soil Degradation in the United States: Extent, Severity, and Trends*. CRC Press, Boca Raton, FL, USA.
19. Lobell D.B. 2010. Remote sensing of soil degradation: introduction. *Journal of Environmental Quality*, 39: 1-4.
20. Metternicht G. 1998. Analysing the relationship between ground based reflectance and environmental indicators of salinity processes in the Cochabamba Valleys (Bolivia). *International Journal of Ecology and Environmental Sciences* 24: 359–370.
21. Metternicht G.I. and Zinck J.A. 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing and Environment*, 85: 1–20.

22. Minasny B. and McBratney A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32: 1378–1388.
23. Mougnot B., Pouget M. and Epema G. 1993. Remote sensing of salt-affected soils. *Remote Sensing Reviews*, 7: 241–259.
24. Nawar S., Buddenbaum H. Hill J. and 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.
25. Noroozi A.A., Homae M. and Farshad A. 2012. Integrated Application of Remote Sensing and Spatial Statistical Models to the Identification of Soil Salinity: A Case Study from Garmsar Plain, Iran. *Environmental Sciences*, 9(1): 59-74.
26. Qadir M., Qurshi A.S. and Cheraghi S.A.M. 2007. Extent and characterization of salt-affected soils in Iran and strategies for their amelioration and management. *Land Degradation and Development*, 19: 214-227.
27. Rao B., Sankar T., Dwivedi R., Thammappa S., Venkataratnam L., Sharma R. and Das S. 1995. Spectral behaviour of salt-affected soils. *International Journal of Remote Sensing*, 16: 2125–2136.
28. Richards L.A. 1954. Diagnosis and improvement of saline and alkali soils. *Agricultural Handbook no. 60*, USDA.
29. Rodríguez P.G., González M.E.P. and Zaballos A.G. 2007. Mapping of salt affected soils using TM images. *International Journal of Remote Sensing*, 28: 2713–2722.
30. Scudiero E., Skaggs T.H. and Corwin D.L. 2014. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional*. 2-3: 82-90.
31. Shao Y., Hu Q., Guo H., Lu Y., Dong Q. and Han C. 2003. Effect of dielectric properties of moist salinized soils on backscattering coefficients extracted from RADARSAT image. *IEEE Trans. Geosciences and Remote Sensing*, 41: 1879–1888.
32. Taghizadeh-Mehrjardi R., Minasny B., Sarmadian F. and Malone B. 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* 213: 15–28.
33. Tanji K.K. and Wallender W.W. 2012. Nature and extent of agricultural salinity and sodicity. In: Wallender W.W., Tanji K.K. (eds.) *Agricultural Salinity Assessment and Management*. ASCE Manuals and Reports on Engineering Practices No. 71. ASCE, Reston, VA, USA, pp. 10-25.
34. Udelhoven T., Emmerling C. and Jarmer T. 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant and Soil* 251: 319–329.
35. Wold S., Sjostrom M. and Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58: 109–130
36. Wu W., Mhaimed A.S., Al-Shafie W.M., Ziadat F., Dhehibi B., Nangia V. and De Pauwa E. 2014. Mapping soil salinity changes using remote sensing in Central Iraq. *Geoderma Regional*, 2–3: 21–31.

Statistical Modeling of Soil Salinity on Large Scale

Y. Hasheminejad¹ - M. Homae^{2*} - A. Noroozi³

Received: 25-07-2016

Accepted: 07-03-2017

Introduction: Soil salinization is increasing across developing world countries and agricultural production is decreasing as a result of this stress. Climate change could adversely affect soil salinization trend through the decrease in rainfall and increased evapotranspiration in arid regions. Policy and decision makers require continuous and quantitative monitoring of soil salinity to adapt with the adverse effects of climate change and increasing need for food. Indices derived from near surface or satellite based sensors are increasingly applied for monitoring of soil salinity so a considerable number of these indices are introduced already for soil salinity monitoring. Different regression methods have been already used for modeling and verification of developed models amongst them multiple linear regression (including stepwise, forward selection and backward elimination) and partial least square regression are the most important methods.

Materials and Methods: To evaluate different approaches for modeling soil salinity against remotely sensed data, an area of about 50000 ha was selected in Sabzevar- Davarzan plain during 2013 and 2014 years. The locations of sampling points were determined using Latin Hypercube Sampling (LHS) strategy. Sampling density was 97 points for 2013 and 25 points for 2014. All points were sampled down to 90 cm depth in 30 cm increments. Totally 366 soil samples were analyzed in the laboratory for electrical conductivity of saturated extract. Electromagnetic induction device (EM38) was also used to measure bulk soil electrical conductivity for the sampling points at the first year and sampling points and 8 points around it at the second year. Totally 97 and 225 EM measurements were also recorded for first and second years respectively. Mean measured soil EC data were calibrated against the EM measurements. Finding the fair correlations, the EM and EC data could be converted to each other. 23 spectral indices derived from Landsat 8 images in the sampling dates along with DEM were used as independent variables. Multiple Linear Regression (MLR) and Partial Least Square Regression (PLSR) methods were evaluated for their fitness in predicting soil salinity from independent variables in different calibration and verification datasets.

Results and Discussion: Different multiple linear regression approaches using the first year data for training and second year data for testing the models and vice versa were evaluated which produced determination coefficients of about 22 to 88 percent in the training dataset but this regression did not reach to 29 percent in the test dataset. Due to the multiple co-linearity amongst the independent variables the multiple linear regression methods were not applicable to all variables. Excluding the co-linear variables, log- transforming and randomizing them into train and test datasets improved the determination coefficient of model and its validation at an acceptable level. Application of partial least square regression using the original and log- transformed data of first and second years as train and test datasets and vice versa introduced determination coefficients of about 39 to 85 percent in the training dataset but were not able to predict in the test dataset. Random dividing of all data into train and test datasets considerably increased the determination coefficient in the verification dataset. Repeating the randomization showed that the approach has the required consistency for predicting the coefficients of variables.

Conclusions: Wide range of independent variable could be used for predicting soil salinity from remotely sensed data and indices. On the other hand the independent variables generally show multi-colinearity amongst themselves. Correlation matrix, variance inflation factor and tolerance indices could be used to identify multi-colinearity. Removing or scaling the variable with high colinearity could improve the regression. Different data transformation methods including log- transformation could also significantly improve the strength of regression. In this research EM data showed more significant correlations with spectral indices in comparison with laboratorial measured EC data. As the EM38 device measures the reflectance in special range of spectrum this higher correlation could be expected. Such models should be calibrated and verified against ground truth data.

1 - PhD Student, Soil Science Department, Tarbiat Modares University

2 - Professor, Irrigation and Drainage Department, Tarbiat Modares University

(* - Corresponding Author Email: mhomaee@modares.ac.ir)

3 - Assistant Professor, Soil Conservation and Watershed Management Research Center (SCWMRC), Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran

Generally a part of data set is used for calibrating (making the model) and the remained for verifying (testing the model). Random dividing of the total data of 2 years into calibration (2/3 of data) and verification (1/3 of data) could significantly improve the regression in the verification data set. This procedure increases the range of variability for data used for calibration and verification and prevents outlier predictions.

Keywords: Multiple linear regression, Partial least square regression, Remote sensing, Spectral indices, Verification