مقایسه روش‌های زمین‌آمار و درختان تصمیم‌گیری تصادفی در پهنه‌بندی شوری خاک در زهک دشت سیستان

محمدرضا پهلوان‌راد[1]، علیرضا اکبری مقدم، خداداد ده‌مرده، غلامعلی کیخا، ناصر دواتگر و محمد مهدی طهرانی

استادیار بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، زابل، ایران؛ **pahlavanrad@gmail.com**

کارشناس ارشد بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش و کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، زابل، ایران؛ **alirezam.248@gmail.com**

کارشناس ارشد بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش و کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، زابل، ایران؛ **Dahmardehh1@yahoo.com**

کارشناس مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی؛ **keykha309@yahoo.com**

استادیار موسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی کرج، ایران ؛ **n_davatgar@yahoo.com**

استادیار موسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی کرج، ایران ؛ **mtehrani2000@yahoo.com**

دریافت: **95/5/25** و پذیرش: **96/3/3**

چکیده

نقشه‌های خاک از منابع عمده اطلاعات برای مدیریت اراضی، منابع طبیعی و محیط زیست هستند. شوری خاک یکی از عوامل مهم محدود کننده تولید در اراضی دشت سیستان می‌باشد. در این تحقیق روش‌های زمین آمار و درختان تصمیم-گیری تصادفی[2] جهت تولید نقشه‌های شوری خاک مقایسه شدند. تعداد 460 نمونه خاک از عمق 30-0 سانتیمتر با فواصل 750 متری در سطح 41000 هکتار از اراضی زهک دشت سیستان به روش مرکب برداشت شد و هدایت الکتریکی عصاره اشباع خاک اندازه‌گیری گردید. 361 نمونه برای آموزش و 99 نمونه برای اعتبارسنجی مستقل استفاده شدند. در روش زمین آمار انواع مختلف نیم‌تغییرنماهای دایره‌ای، کروی، نمایی و قوسی و روش‌های مختلف میانیابی معکوس فاصله، کریجبینگ ساده، کریجینگ معمولی، کریجینگ عام و کوکریجینگ برازش و بهترین مدل‌ها انتخاب شدند. در مدل درختان تصمیم‌گیری تصادفی از روش نقشه‌برداری رقومی خاک استفاده شد و خصوصیات مختلف اراضی با استفاده از مدل رقومی ارتفاع و تصاویر ماهواره‌ای استخراج شدند. نتایج روش زمین‌آمار نشان داد که همبستگی مکانی شوری خاک در منطقه مورد مطالعه متوسط است و بهترین مدل نیم‌تغییرنما و میانیابی به ترتیب کروی و کریجینگ معمولی بودند. در روش درختان تصمیم‌گیری تصادفی شاخص پوشش گیاهی، جهت شیب و شاخص شوری دارای بیشترین اهمیت در پیش‌بینی شوری خاک بودند. نتایج مقادیر ریشه میانگین مربعات خطا و خطای میانگین برای داده‌های آموزش و تست نشان داد که روش درختان تصمیم‌گیری تصادفی نسبت به روش زمین‌آمار اندکی برتری داشت. استفاده از سایر متغیرهای کمکی مانند نقشه‌های کاربری اراضی و سری خاک می‌تواند دقت نقشه تولیدی با روش درختان تصمیم‌گیری تصادفی را بهبود بخشد.

واژه‌های کلیدی: اراضی مسطح، نقشه‌برداری رقومی خاک، نقشه شوری خاک، زمین آمار

---

[1]. نویسنده مسئول، آدرس: زابل، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی سیستان، بخش تحقیقات خاک و آب

[2]. Random forest

مقدمه

وجود نقشه‌های خاک از نیازهای اساسی در مطالعات علوم زمین می‌باشد و از منابع عمده اطلاعات برای مدیریت اراضی، منابع طبیعی و محیط زیست هستند. تهیه نقشه شوری خاک به عنوان یکی از عوامل محدود کننده رشد، یکی از ضروریات لازم در مناطق خشک می‌باشد. به‌طور کلی دو روش سنتی و نوین برای تهیه نقشه‌های خاک وجود دارد. در روش‌های سنتی، خاک به صورت ناپیوسته مد نظر قرار می‌گیرد و اصل بر تعمیم نتایج نقاط مطالعاتی به کل واحد نقشه است؛ ولی در روش‌های نوین، خاک به‌عنوان یک موجودیت پیوسته در نظر گرفته می‌شود که این روش‌ها سعی دارند این تغییرات را تا حد ممکن نشان دهند (ژو و همکاران،2003). مهم‌ترین محدودیت نقشه‌برداری سنتی خاک، تعمیم نتایج حاصل از خاکرخ‌های شاهد به کل واحد نقشه، بدون در نظر گرفتن تغییرپذیری مکانی (و یا زمانی) خاک می‌باشد (یوتست و همکاران، 2002). از روش‌های نوین برای تهیه نقشه‌های خصوصیات و کلاس اراضی می‌توان به مدل‌هایی مانند زمین آمار، مدل‌های آماری تجربی، طبقه‌بندی فازی، تصمیم گیری درختی و شبکه‌های عصبی مصنوعی استفاده کرد (مک براتنی و همکاران، 2003).

مدل زمین‌آمار

از اواسط قرن بیستم شاخه‌ای از علم آمار به نام زمین‌آمار پا به عرصه علوم نهاد که امکان پردازش داده‌ها و توصیف مکانی آنها را به وجود آورد. در زمین‌آمار وابستگی مکانی پارامترهای اندازه‌گیری شده به وسیله نیم-تغییرنما رسم می‌شود. نیم‌تغییرنما، تغییرات فاصله‌ای یا ساختار تغییرپذیری یک متغیّر خاص را نشان داده و از ابزارهای اساسی زمین‌آمار برای بررسی تغییرات مکانی خصوصیات خاک است (ایساکس و سری واستاوا، 1989). در نیم‌تغییرنما سه پارامتر اساسی اثر قطعه‌ای، حد آستانه و دامنهٔ تأثیر بیان شده است. با استفاده از پارامترهای نیم‌تغییرنما در زمین‌آمار می‌توان مقدار یک کمیت در نقاطی که نمونه برداری نشده‌اند و دارای مختصات معلوم هستند را با استفاده از مقادیر سایر نقاط که نمونه‌برداری شده‌اند و دارای مختصات معلوم هستند با میانیابی به دست آورد. انواع مختلفی از روش‌های میانیابی در زمین آمار وجود دارند که مهمترین آنها معکوس فاصله، سطوح روند، اسپلین، کریجینگ و کوکریجینگ هستند (محمدی، 1385). سکوتی اسکوئی و همکاران (1386) پراکنش شوری خاک را با استفاده از روش زمین-آمار مورد ارزیابی قرار دادند و مشاهده کردند که روش کریجینگ دقت بالائی در برآورد شوری در نقاطی که نمونه‌برداری نشده است دارد. یانگ و همکاران (2005) برای تخمین شوری خاک از سه روش کریجینگ معمولی، کوکریجینگ و کریجینگ – رگرسیون استفاده کردند. صادقی و همکاران (1393) برای تهیه نقشه عناصر کم مصرف در استان گلستان از روش زمین‌آمار استفاده کردند و مشاهده کردند روش‌های معکوس چند ربعی و کریجینگ مناسب‌ترین روش برای تخمین منگنز و روی بودند. یویان و کای (2010) از روش‌های زمین‌آماری به عنوان یک ابزار مناسب تصمیم‌گیری در فرآیندهای هیدرولوژی، کمّی‌سازی تغییرات مکانی و پهنه‌بندی غلظت نیترات در آب زیرزمینی استفاده کردند.

مدل درختان تصمیم گیری تصادفی (RF[1])

مدل توسعه یافته از مدل طبقه‌بندی و رگرسیون درختی (CART[2]) می‌باشد. روش طبقه‌بندی و رگرسیون درختی (برایمن‌وهمکاران،1984) روشی است که داده‌ها را به طور تکراری برای بدست آوردن ارتباط بین متغیر پاسخ و متغیرهای مستقل و انجام تخمین جداسازی می‌کند. در روش درختان تصمیم‌گیری تصادفی به جای رشد یک درخت، صدها یا هزاران درخت طبقه‌بندی تولید می‌شود (بریمن و کاتلر، 2004). تکنیک درختان تصمیم-گیری تصادفی همچنین اهمیت متغیرها در مدل‌سازی را نیز تعیین می‌کند.

برونگارد (2009) تکنیک درختان تصمیم‌گیری تصادفی را برای پیش بینی مکانی کلاس‌ها و خصوصیات خاک استفاده کرد وی دریافت که این تکنیک یک روش مفید برای پیش بینی کلاس‌های خاک، خصوصیات خاک و توزیع پوشش گیاهی است. گریم و همکاران (2008) روش درختان تصمیم‌گیری تصادفی را برای پیش بینی و منشاء کربن آلی خاک در محیط تروپیکال استفاده کردند و آن را یک ابزار نیرومند برای تخمین مکانی و درک ارتباطات خاک- زمین یافتند. لایب و همکاران (2012) روش‌های درختان تصمیم گیری تصادفی و درخت رگرسیون(RT[3])را برای پیش بینی توزیع بافت خاک استفاده کرده و مشاهده کردند که روش درختان تصمیم-گیری تصادفی دارای دقت بالاتری بود. پهلوان راد و همکاران (2014) روش درختان تصمیم‌گیری تصادفی را روش مناسبی برای به‌روزرسانی نقشه‌های خاک در استان گلستان یافتند. وری و همکاران (2015) از روش‌های مختلف تهیه نقشه رقومی خاک برای نقشه‌برداری ذخیره

1. Random Forest
2. Classification and regression trees
3. Regression tree

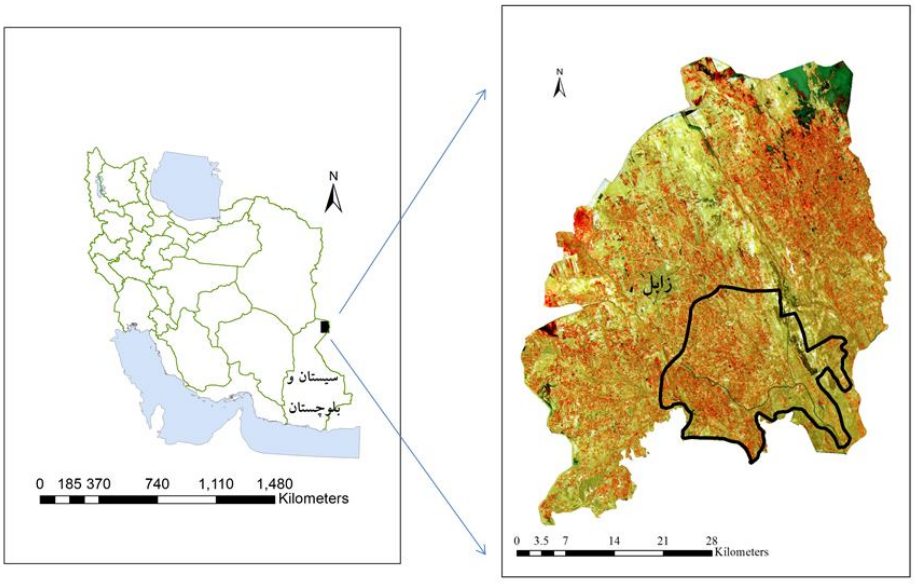
کربن آلی خاک، استفاده نمودند. نتایج آن‌ها دقت بالای نقشه‌های تهیه شده با این روش را نشان داد و تأکید و تأییدی بر استفاده از روش‌های مختلف نقشه‌برداری رقومی در تهیه نقشه خصوصیات خاک بود. هنگل و همکاران (2015) مشاهده کردند که استفاده از روش درختان تصمیم‌گیری تصادفی دارای دقت بیشتری در مقایسه با روش‌های رگرسیون خطی در پیش‌بینی خصوصیات خاک در آفریقا بود.

شوری از عوامل مهم محدودکننده رشد در اراضی منطقه سیستان می‌باشد. با توجه به در پیش رو بودن طرح بزرگ انتقال آب با لوله به اراضی دشت سیستان، اطلاع دقیق از وضعیت پراکنش شوری اراضی این دشت، نقش مهمی در انتخاب اراضی دارای محدودیت‌های کمتر برای اجرای این پروژه و مدیریت بهینه این اراضی دارد. از این‌رو تحقیق حاضر جهت ارزیابی دو مدل زمین‌آمار به عنوان یک روش مرسوم در تهیه نقشه خصوصیات خاک در کشور و مدل درختان تصمیم‌گیری تصادفی به عنوان روش جدید، برای تخمین شوری خاک در نقاط فاقد نمونه‌برداری، در اراضی شهرستان زهک دشت سیستان انجام شد.

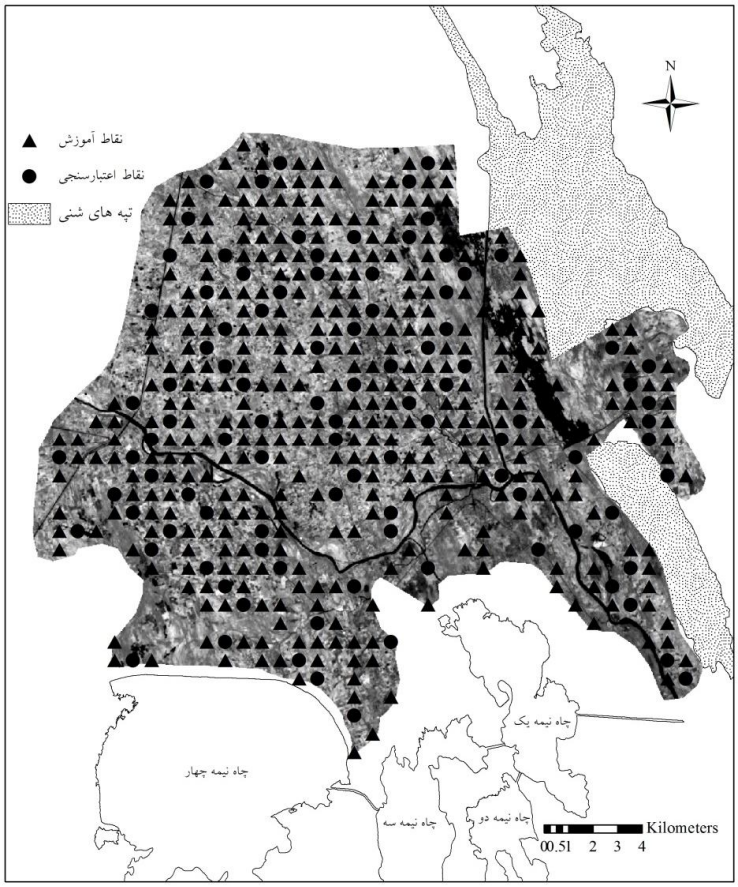
مواد و روش‌ها

وضعیت عمومی منطقه و روش نمونه‌برداری

محدوده مورد مطالعه دارای مساحت حدود 41000 هکتار در محدوده اراضی زهک دشت سیستان در شمال استان سیستان و بلوچستان واقع شده است (شکل 1). مواد مادری منطقه به‌طور عمده رسوبات آبرفتی رودخانه هیرمند هستند. رژیم رطوبتی منطقه اریدیک و رژیم حرارتی آن هایپرترمیک است. از لحاظ شیب، منطقه تقریباً مسطح و با شیب 2-1 درصد می‌باشد. نمونه‌برداری با فواصل 750 متری و به تعداد 460 نمونه انجام شد. پس از مشخص‌شدن محل جغرافیایی هر نطقه با استفاده از دستگاه جی پی اس در عملیات میدانی، یک نمونه از آن محل و سه نمونه به شعاع 10 متر از اطراف آن از عمق 30-0 سانتیمتری برداشت و با هم مخلوط و یک نمونه مرکب تهیه شد و پس آماده‌سازی و عبور از الک دو میلیمتری، هدایت الکتریکی عصاره اشباع نمونه‌های برداشت شده تعیین گردید. 361 نمونه برای آموزش دو مدل و از 99 نمونه برای اعتبار سنجی استفاده شد (شکل 2).



شکل 1- محل منطقه مورد مطالعه



شکل 2- محل نقاط آموزش و تست در منطقه مطالعاتی

زمین آمار

تجزیه و تحلیل ساختار مکانی داده‌ها و رسم نیم‌تغییرنما با استفاده از ArcGIS 10.3 انجام گردید. پس از بررسی داده‌ها و مناسب بودن آنها برای آنالیز زمین-آماری، انواع مختلف مدل‌های نیم‌تغییر نما شامل کروی، گوسی، نمایی و خطی وهمچنین روش‌های مختلف میانیابی شامل معکوس فاصله، پلی نومیال عام، پلی نومیال موضعی و کریجینگ‌ها برازش و بهترین مدل‌ها بر اساس کمترین ریشه میانگین مربعات خطا و خطای میانگین انتخاب شدند.

نیم‌تغییرنمای تجربی عبارت از متوسط مجذور اختلافات بین دو مشاهده ($Z(x+h)$, $Z(x)$) در دو موقعیت مکانی واقع در فضای نمونه‌برداری است که توسط آرایه h از هم جدا می‌شوند (معادله 1) (محمدی، 1385).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i - h))^2 \quad (1)$$

که در آن، $\gamma(h)$ مقدار واریوگرام برای جفت نقاطی که به فاصلهٔ h از هم قرار دارند، $N(h)$ تعداد زوج نقاطی که به فاصله h از هم قرار دارند، $Z(x_i)$ مقدار مشاهده شده متغیر x و $Z(x_i+h)$ مقدار مشاهده شده متغیر که به فاصله h از x قرار دارد. یک نیم تغییرنمای ایده‌آل دارای سه پارامتر شامل اثر قطعه‌ای (Nugget effect)، حد آستانه (Sill) و دامنه (Range) می‌باشد. اثر قطعه‌ای بیان کننده مولفه غیرساختاری (تصادفی) واریانس می‌باشد. حد آستانه تقریبی از واریانس کل را ارائه نموده و مقدار دامنه بیان‌گر فاصله‌ای است که در ماورای آن نمونه‌ها را می‌توان مستقل از یگدیگر به حساب (محمدی، 1385). نسبت واریانس قطعه‌ای به آستانه ($\frac{C_0}{C_0+C}$) شاخصی از قدرت ساختار متغیّرهای مکانی است. اگر این نسبت کمتر از 0/25 باشد، متغیّر از ساختار مکانی قوی و اگر نسبت بین 0/25 تا 0/75 قرار گیرد ساختار مکانی آن متوسط و اگر این نسبت بیش از 0/75 باشد، ساختار مکانی آن ضعیف خواهد بود. (محمدی، 1385).

روش درختان تصمیم‌گیری تصادفی

در استفاده از تکنیک درختان تصمیم‌گیری تصادفی از روش نقشه‌برداری رقومی خاک (DSM[1]) استفاده شد. نقشه‌برداری رقومی خاک شامل روش‌ها و مدل‌هایی است که بین توزیع خاک (کلاس‌ها یا خصوصیات خاک) و داده‌هایی که به آسانی و با قیمت ارزان از طریق روش‌های سنجش‌ازدور، تصاویر و عکس-های ماهواره‌ای و داده‌های ژئومورفومتری بدست می‌آیند و تحت عنوان متغیرهای کمکی محیطی[2] نامیده می‌شوند ارتباط برقرار می‌کند. اگر چه تکنیک‌های نقشه‌برداری رقومی خاک بر روی محاسبات دیجیتالی استوار است ولی اساس آن بر روی معادلات تشکیل خاک (بنی، 1941؛ مک براتنی و همکاران، 2003) قرار دارد. بنی (1941) مدل معروف خود را که بر اساس آن خاک بر روی یک زمین‌نما تابعی از 5 فاکتور محیطی است معرفی کرد. این مدل به صورت معادله $S= f (CL,O, R, P, T)$ تعریف شده است، که S : خاک، CL: اقلیم، O: ارگانیسم که بیشتر به صورت پوشش گیاهی بیان می‌شود، R: توپوگرافی یا پستی و بلندی، P: مواد مادری و T: زمان می‌باشد.

مک براتنی و همکاران (2003) مدل پدولوژیکی ClORPT را بازبینی کردند و پارامترهای خاک (کلاس‌ها و یا خصوصیات خاک) و موقعیت مکانی را به آن اضافه نموده و مدل جدیدی تحت عنوان مدل SCORPAN را برای کمی کردن ارتباط بین داده‌های دقیق مکانی و خاک ارائه کردند. اجزای این مدل تحت عنوان متغیرهای کمکی محیطی معرفی شده‌اند. مدل جدید به صورت $S= f (s, c, o, r, p, a, n)$ است که فاکتورهای s داده‌های صحرایی و یا آزمایشگاهی خاک در نقاط نمونه‌برداری یا در نقشه‌های خاک هستند و n موقعیت مکانی نقاط نمونه‌برداری می‌باشند که به پارامترهای مدل بنی (1941) اضافه شدند.

از مدل رقومی ارتفاع (DEM[3]) با اندازه پیکسل30 متری، خصوصیات مختلف زمین شامل ارتفاع، شیب، جهت شیب، انحنای سطح ، انحنای نیمرخ، شاخص خیسی، شاخص توپوگرافی LS (حال‌ضرب طول شیب و جهت شیب) و سطح پایه شبکه کانال‌ها با استفاده از نرم افزار SAGA استخراج گردیدند (ویلسون و گالانت، 2000). از تصاویر ماهواره لندست +ETM 8، شاخص پوشش گیاهی تفاضلی نرمال شده (NDVI[4]) و شاخص شوری تفاضلی نرمال شده (NDSI[5]) در مدل سازی استفاده شد.

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (2)$$

$$NDSI = \frac{Red-NIR}{Red+NIR} \quad (3)$$

در این معادله NIR و Red مقادیر انعکاس امواج مادون قرمز نزدیک و قرمز هستند.

لایه‌های مختلف متغیرهای تولید شده با اندازه پیکسل‌های 30 متری و مقادیر مختلف شوری خاک وارد نرم افزار R 3.0.1 (2013) گردید و مدل درختان تصمیم گیری تصادفی اجرا گردید. در این تکنیک اهمیت متغیرها نیز تعیین می‌شود. مقادیر درست متغیرها با مقادیر که به طور تصادفی برای هر درخت تولید شده است جایگزین می‌شود و اثر این تغییر را روی طبقه‌بندی اندازه‌گیری می‌کند؛ اگر این جایگزینی اثری روی خطای اندازه‌گیری نداشته باشد اهمیت آن متغیر کم است و اگر مقدار خطای اندازه‌گیری افزایش یابد آن متغیر مهم می‌باشد (بریمن و کاتلر، 2004). در مدل نهایی متغیرهایی که اهمیت بیشتری داشتند استفاده شدند. در مدل RF تعداد درختان در جنگل و تعداد متغیرهای محیطی در گره هر درخت توسط کاربر مشخص می‌شود و مناسب‌ترین مقدار این دو پارامتر با روش سعی و خطا جهت بدست آوردن کمترین مقدار خطا بدست آمد.

مقایسه زمین‌آمار و درختان تصمیم‌گیری تصادفی

به منظور ارزیابی دقت مدل‌ها ریشه میانگین مربعات خطا (RMSE) (بیان‌گر دقت میانگین پیش‌بینی می‌باشد و شدت خطا منتظره را نشان می‌دهد) و خطای میانگین (ME) (برای تعیین اریب و تمایل به کم یا بیش برآورد) برای داده‌های آموزش و تست محاسبه گردید.

$$ME = \frac{1}{N}\sum_{i=1}^{N}[z(x_i) - z^*(x_i)] \quad (4)$$

$$RMSE = \left[\frac{\sum_{i=1}^{N}[z(x_i) - z^*(x_i)]^2}{N}\right]^{1/2} \quad (5)$$

که $z^*(x_i)$ بیان‌گر مقادیر تخمینی، $z(x_i)$ مقادیر واقعی متغیرها و N تعداد داده‌ها می‌باشد.

1. Digital Soil Mapping
2. Environmental covariates
3. Digital elevation model
4. Normalized difference vegetation index
5. Normalized difference salinity index

نتایج و بحث

توصیف آماری هدایت الکتریکی عصاره اشباع خاک در محدوده مطالعاتی در جدول 1 نشان داده شده است. براساس نتایج این جدول میانگین شوری خاک در اراضی مورد مطالعه 17/3 دسی زیمنس بر متر با حداقل شوری 0/7 و حداکثر 117 دسی زیمنس بر متر بود که نشان می‌دهد اراضی مورد مطالعه تحت تأثیر شوری قرار دارند.

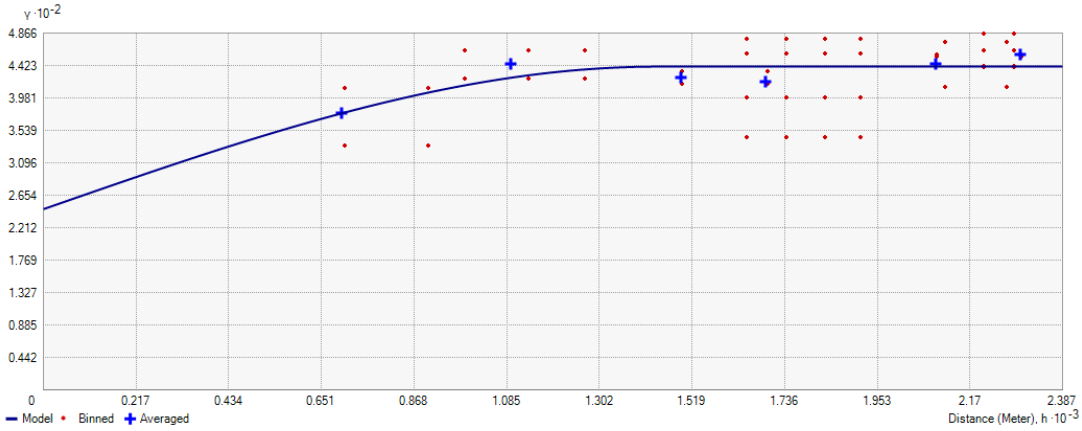
جدول 1- آماره‌های توصیفی هدایت الکتریکی خاک

تعداد	واحد	میانگین	انحراف معیار	میانه	حداقل	حداکثر	چولگی	افراشتگی
460	$ds\ m^{-1}$	17/3	22/7	6/5	0/7	117	1/95	6/8

اجرای مدل زمین آمار

بررسی هیستوگرام‌های فراوانی داده‌های اندازه-گیری شده، نشان‌دهنده غیرنرمال بودن داده‌ها و چولگی شدید آنها به چپ است. با توجه به اینکه نرمال بودن داده‌ها در تجزیه تحلیل‌های زمین‌آماری نتایج مطلوبتری را در پی خواهد داشت (محمدی، 1385)، اقدام به نرمال-سازی توزیع داده‌ها با استفاده از تبدیل‌گر لگاریتمی گردید. با انجام سعی و خطا بین روش‌های مختلف بهترین مدل میانیابی، روش کریجینگ معمولی با نیم‌تغییر نمای کروی بدست آمد (شکل 3) که دارای کمترین میانگین مربعات خطا (22/9) بود. جدول 3 مقادیر پارامترهای مختلف نیم‌تغییرنمای انتخاب شده را نشان می‌دهد. مقدار وابستگی مکانی هدایت الکتریکی عصاره اشباع خاک در منطقه متوسط و دامنه تأثیر 1436متر بود. کم بودن وابستگی مکانی هدایت الکتریکی اشباع خاک در منطقه مطالعاتی را می‌توان به عوامل ایجاد شوری در این منطقه مربوط دانست. تبخیر زیاد در منطقه (حدود 5 متر در سال)، نفوذپذیری کم اراضی و وجود لایه‌های محدود کننده در اعماق کمتر از 5 متر که مانع از شستشوی نمک به اعماق پایین‌تر می‌شوند و همچنین رها کردن بعضی اراضی به صورت بایر به علت کمبود آب در اثر خشکسالی که منجر به تجمع نمک در اثر تبخیر زیاد در این اراضی می‌گردد، از عوامل ایجاد شوری در این منطقه می‌باشد که سبب می‌گردد مقدار توزیع شوری در اراضی این منطقه دارای وابستگی مکانی بالایی نباشد.

دیانی و همکاران (1391) در پهنه‌بندی شوری و سدیمی بودن اراضی غرب کارون مشاهده کردند که شوری و رس خاک هر دو دارای واریوگرام کروی بوده و بهترین روش میانیابی، روش کریجینگ بود. کاظمی پشت مساری و همکاران (1391) در تخمین و پهنه نیتروژن و فسفر خاک مشاهده کردند که روش کریجینگ از سایر روش‌های زمین آمار مناسب‌تر و هر دو خصوصیت خاک دارای واریوگرام نمایی بودند.



شکل 3- نیم تغییر نمای هدایت الکتریکی خاک

جدول 2- پارامترهای نیم تغییرنمای هدایت الکتریکی خاک در منطقه مورد مطالعه

مدل	دامنه (متر)	اثر قطعه‌ای	حد آستانه	درصد وابستگی مکانی	کلاس وابستگی مکانی
کروی	1436	247	193	56	متوسط

اجرای مدل‌درختان تصمیم‌گیری تصادفی

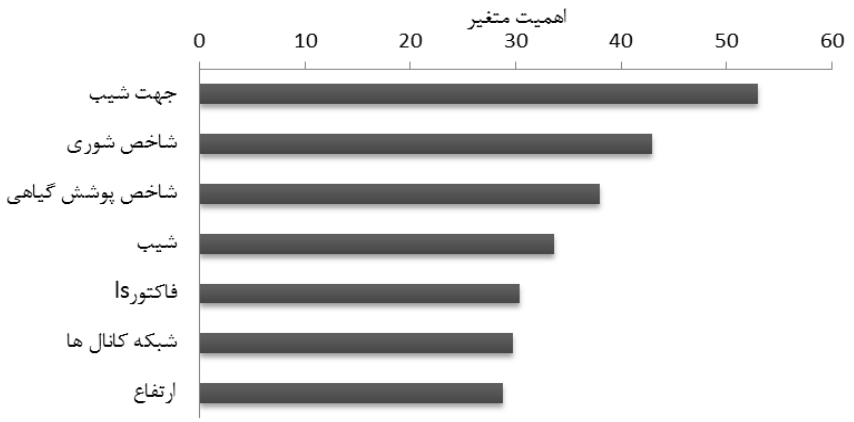
بهترین مدل، با روش سعی و خطا با ایجاد 5000 درخت و استفاده از 7 متغیر در مدل به‌دست آمد. این متغیرها دارای اهمیت بیشتری در تخمین شوری خاک بودند که شامل شیب، جهت شیب، ارتفاع، سطح پایه شبکه کانال‌ها، شاخص LS، شاخص پوشش گیاهی و شاخص شوری بودند. سایر متغیرها تأثیر کمی در مدل سازی داشتند، از این رو، در مدل نهایی از این پنج متغیر استفاده شد.

اهمیت متغیرها

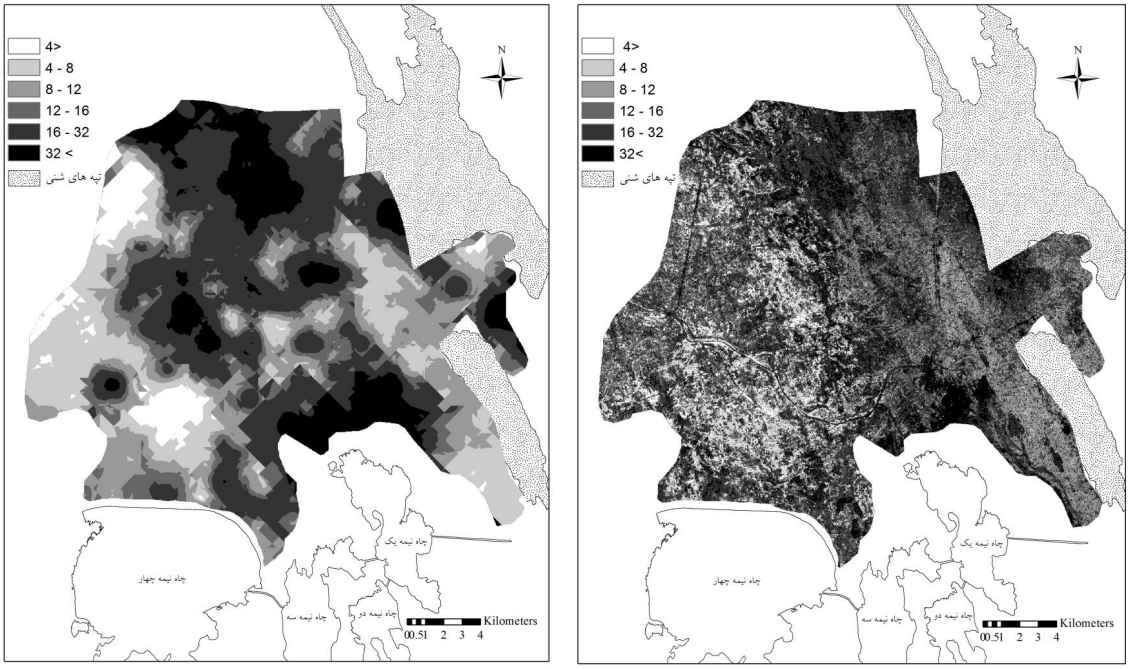
بین متغیرهای انتخاب شده برای مدل‌سازی شاخص جهت شیب، شاخص شوری، شاخص پوشش گیاهی و شیب دارای بیشترین اهمیت در تخمین نقشه تولیدی بودند (شکل 4). در مناطقی که شوری خاک کمتر بوده است پوشش گیاهی بیشتری وجود داشته و توانسته است در پیش‌بینی شوری خاک اهمیت بیشتری داشته باشد و شاخص شوری هم توانسته به خوبی در تخمین نقشه شوری خاک مؤثر باشد. گریم و همکاران (2008) مشاهده کردند که توپوگرافی مهمترین متغیر در تخمین ذخیره کربن آلی در مطالعه آنان بوده است. پهلوان راد و همکاران (1393) در تخمین کلاس‌های خاک در اراضی استان گلستان مشاهده کردند که شاخص پوشش گیاهی و ژئومورفولوژی از مهمترین عوامل در تخمین کلاس‌های خاک بود. تقی‌زاده مهرجردی و همکاران (1393) شاخص خیسی و ژئومورفولوژی را به عنوان متغیرها در تخمین شوری خاک در منطقه اردکان یزد معرفی کردند. دهنی و لوئیس (2012) کارایی بالای تصاویر +ETM را در تهیه نقشه شوری مشاهده کردند.

پهنه‌بندی شوری خاک

نقشه‌های تولید شده با دو روش در شکل 5 ارائه شده است که نشان می‌دهد در هر دو روش بیشترین شوری خاک در قسمت جنوبی و شمالی منطقه مطالعاتی می‌باشد و کمترین مقدار شوری خاک در قسمت جنوب شرقی و غرب منطقه مطالعاتی جایی که اراضی در فاصله کمتری از رودخانه سیستان که از شاخه‌های منشعب رودخانه هیرمند در دشت سیستان است قرار دارند و با کمترین جریان آب در آن آبیاری می‌شوند و همچنین با چاه‌هایی که در فاصله نزدیکتری به رودخانه حفر شده آبیاری می‌گردد که کیفیت بهتری نسبت به سایر نقاط اراضی زهک دارند.



شکل 4- اهمیت متغیرها در تخمین هدایت الکتریکی خاک



شکل 5- نقشه‌های تولید شده هدایت الکتریکی خاک با مدل‌های درختان تصمیم‌گیری تصادفی (سمت راست) و زمین‌آمار (سمت چپ) مقایسه مدل‌ها

مقایسه مقدار جذر میانگین مربعات خطای آزمایش در داده‌های آموزش (جدول 3) نشان می‌دهد که روش درختان تصمیم‌گیری تصادفی دارای خطای خیلی کمتری نسبت به روش زمین‌آمار است. مقدار تفاوت میانگین خطا در دو روش دارای اختلاف کمی است. مقایسه نتایج داده‌های تست نشان داد که مقدار جذر میانگین مربعات خطای دو روش با اختلاف کمی در روش درختان تصمیم‌گیری تصادفی (19/5) کمتر از روش زمین‌آمار (20/3) بود. مقدار جذر میانگین مربعات خطای و میانگین خطای مطلق در روش زمین‌آمار اندکی کمتر از روش درختان تصمیم‌گیری بود (جدول 3). بنابراین روش درختان تصمیم‌گیری تصادفی دارای اندکی دقت بالاتری نسبت به روش زمین‌آمار بود. هنگل و همکاران (2015) مشاهده کردند که استفاده از روش درختان تصمیم‌گیری تصادفی دارای دقت بیشتری در مقایسه با روش کریجینگ در پیش‌بینی خصوصیات خاک در آفریقا داشت. کمپن (2013) در تخمین کربن آلی افق سطحی و ضخامت پیت با دو روش رگرسیون درختی و زمین‌آمار در سطح 14000 هکتار در کشور هلند مشاهده کرد که روش کریجینگ دارای دقت بهتری بود. عدم قطعیت موجود در روش درختان تصمیم‌گیری تصادفی می‌تواند به دلیل مسطح بودن اراضی منطقه باشد که متغیرهای استخراج شده از نقشه رقومی ارتفاع نتوانسته‌اند تغییرات خاک را بیان کنند. برخی مطالعات نیز نشان داده که فاکتور توپوگرافی در مدل‌سازی اهمیت زیادی نداشته است (ری‌ورو[1] و همکاران، 2007) این مساله به‌خصوص در مناطق مسطح ممکن است اتفاق بیفتد. اکرمخانوف[2] و همکاران (2011) ارتباط معنی‌داری بین خصوصیات اراضی و داده‌های شوری خاک به علت کم بودن ارتفاع منطقه مشاهده نکردند.

استفاده از سایر متغیرهایی که در این مطالعه استفاده نشد مانند نقشه‌های کاربری اراضی، فاصله از رودخانه و همچنین استفاده از متغیرهایی با بزرگنمایی 10 متر می تواند دقت مدل را افزایش دهد و سبب بهبود دقت نقشه تولیدی گردد.

1. Rivero
2. Akramkhanov

جدول 3- مقادیر میانگین جذر مربعات خطا و میانگین خطا برای داده‌های آموزش

پارامتر مدل	درختان تصمیم‌گیری تصادفی	زمین‌آمار
RMSE	10/0	21/3
ME	-0/3	0/1

جدول 4- مقادیر میانگین جذر مربعات خطا و میانگین خطا برای داده‌های تست

پارامتر مدل	درختان تصمیم‌گیری تصادفی	زمین‌آمار
RMSE	19/5	20/3
ME	-4/1	-2/6

نتیجه‌گیری

خاک‌های سطحی اراضی شهرستان زهک دشت سیستان با میانگین شوری 17/3 دسی زیمنس بر متر به دلایل تبخیر زیاد در منطقه (حدود 5 متر در سال)، نفوذپذیری کم اراضی و وجود لایه‌های محدود کننده در اعماق کمتر از 5 متر که مانع از شستشوی نمک به اعماق پایین‌تر می‌شوند و همچنین رها کردن بعضی اراضی به صورت بایر به علت کمبود آب در اثر خشکسالی که منجربه تجمع نمک در اثر تبخیر زیاد در این اراضی می‌گردد، متأثر از شوری هستند. مقایسه روش‌های تکنیک درختان تصمیم‌گیری تصادفی که در آن از روش نقشه-برداری رقومی خاک استفاده شد و زمین‌آماری در اراضی زهک دشت سیستان نشان داد که روش درختان تصمیم-گیری تصادفی نسبت به روش زمین‌آمار برتری اندکی در پیش‌بینی توزیع مکانی شوری خاک داشت که علت آن احتمالاً مربوط به پستی و بلندی کم اراضی منطقه مطالعاتی باشد. نتایج مدل زمین‌آمار نشان داد که وابستگی مکانی هدایت الکتریکی عصاره اشباع خاک در منطقه متوسط و دارای دامنه تأثیر 1436متر بود. تکنیک درختان تصمیم‌گیری تصادفی نشان داد که جهت شیب و شاخص شوری از مهمترین متغیرهای تأثیرگذار در تغییرات شوری خاک در منطقه مطالعاتی بودند. استفاده از سایر متغیرهایی که در این تحقیق استفاده نشد، مانند نقشه‌های کاربری اراضی، فاصله از رودخانه، نقشه بافت و سری‌های خاک، می‌تواند دقت نقشه تهیه شده با روش درختان تصمیم-گیری تصادفی را افزایش دهد. پیشنهاد می‌شود تکنیک درختان تصمیم‌گیری تصادفی در سایر مناطق کشور برای پیش‌بینی خصوصیات خاک ارزیابی‌گردد.

فهرست منابع:

1. پهلوان راد، م.ر.، خرمالی، ف.، تومانیان، ن.، کیانی، ف.، و کمکی، چ.ب.پ.، 1393. پهنه‌بندی رقومی واحدهای خاک با استفاده از مدل درختان تصمیم‌گیری تصادفی در استان گلستان. پژوهش‌های حفاظت آب و خاک. جلد 21 . شماره 6: 73-93.
2. تقی‌زاده مهرجردی، ر.ا.، سرمدیان، ف.، روستا، م.ج.، رحیمیان، م.ح.، امید، م.، و تومانیان، ن. 1393.پهنه‌بندی رقومی قابلیت هدایت الکتریکی ظاهری با استفاده از رگرسیون کریجینگ و واریوگرام محلی در منطقه اردکان. مدیریت خاک و تولید پایدار. جلد چهار. 1:29.
3. دیانی، م.، جعفری، س.، خلیل مقدم، ب.، دهقانی، ا. 1391. پهنه‌بندی خطر شوری و سدیمی شدن خاک سطحی با استفاده از زمین‌آمار (مطالعه موردی: اراضی غرب رودخانه کارون در استان خوزستان). *پژوهش و سازندگی*. شماره 94: 86-95.
4. سکوتی اسکوئی، ر؛ مهدیان، م؛ محمودی، ش، 1386 ،مقایسه کارایی برخی روشهای زمین آماری برای پیش‌بینی پراکنش مکانی شوری خاک، مطالعه موردی دشت ارومیه، مجله پژوهش و سازندگی، شماره74 :90-98.

5. صادقی، س.، کاظمی پشت مساری، ح.، طهماسبی سروستانی. 1393. ارزیابی روش‌های درونیابی جهت تعیین پراکندگی برخی عناصر کم‌مصرف در اراضی کشاورزی استان گلستان. مدیریت خاک و تولید پایدار.دوره 4 شماره 3: 323-337.
6. کاظمی پشت مساری، ح.، طهماسبی سروستانی، ز.ا.، بهنام، ک.، شتایی4، ش.، صادقی، س. 1391. ارزیابی روشهای زمین آمار جهت تخمین وپهنه بندی عناصر غذایی پرمصرف اولیه در برخی اراضی کشاورزی استان گلستان. دانش آب و خاک. جلد 22 شماره 1: 201-219.
7. محمدی، ع. 1385. پدومتری (آمار مکانی)، انتشارات پلک. 453 صفحه.
8. Akramkhanov, A., Martius, C., Park, S.J., and Hendrickx, J.M. 2011. Environmental factors of spatial distribution of soil salinity on flat irrigated terrain. Geoderma. 163: 55-62.
9. Breiman, L., and Cutler, A. 2004. Random Forests. Department of Statistics, University of Berkeley. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
10. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and Regression Trees. Chapman and Hall, New York. 368 pages.
11. Brungard, C.W. 2009. Alternative Sampling and Analysis Methods for Digital Soil Mapping in Southwestern Utah. Thesis for Master of Science. Utah State University. USA.
12. Dehni, A., and Lounis, M. 2012. Remote sensing techniques for salt affected soil mapping: application to the Oran region of Algeria. Pro. Engi. 33: 188-198.
13. Grimm, R., Behrens, T., Marker, M., and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using random forests analysis. Geoderma. 146(1-2):102-113.
14. Hengl, T., Heuvelink, B. M., Kempen, B., Leenaars, J.G. B., Walsh., M. G., Shepherd, K.D., Sila, A., MacMillan, R.A., Jesus, J. M., Tamene, L., and Tondoh, J.E. 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLOS ONE 10 (6): e0125814. doi:10.1371/journal.pone.0125814.
15. Isaaks, H.E. and Srivastava, R.M. 1989. An introduction to applied geostatistics. Oxford University Press, NY.
16. Jenny, H. 1941. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York.
17. Kempen, B. 2013. Comparison of universal kriging and regression tree modeling for soil property mapping. Geophysical Research Abstracts. Vol. 15, EGU2013-6899.
18. Ließ, M., Glaser, B., and Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. Groderma. 170: 70-79.
19. McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. Geoderma. 117(1–2):3-52.
20. Pahlavan Rad, M.R., Khormali, F., Toomanian, N, Kiani, F and Komaki, C.B. 2015. Digital soil mapping using Random Forest model in Golestan province. Journal of Water and Soil Conservation. 21(6): 73-93.
21. Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard , C.W., Komaki, B and Bogaert. P., 2014. Mapping and Updating Soil Series Using Random Forest and Conditioned Latin Hypercube Sampling in the Loess Soils of Northern Iran. Geoderma. (232-234): 97-106
22. R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
23. Rivero, R.G., Grunwald, S., and Bruland, G.L. 2007. Incorporation of spectral data into multivariate geostatistical models to map soil phosphorus variability in a Florida wetland. Geoderma. 140: 428– 443.

24. Taghizadeh-Mehrjardi, R., Sarmadian, F., Rousta, M.J. Rahimian, M.H., Omid, M. and Toomanian N. 2015. Digital mapping of apparent electrical conductivity using regression kriging and local variogram in Ardakan region. Journal of Soil Management and Sustainable. 4(4): 1-29.
25. Utest, A., Lopez, T. and Diaz L. 2002. A comparison of soil maps, kriging and a combined method for spatially predicting bulk density and field capacity of Ferralsols in the Havana-Matanzas Plain. Geoderma. 96: 199-213.
26. Uyan, M. and Cay, T. 2010. Geostatistical methods for mapping groundwater nitrate concentrations. 3rd International conference on cartography and GIS, 15-20 June, Nessebar, Bulgaria.
27. Were, K., Bui, D. T., Disk, B., Singl, B. R. 2015. Acomparative assessment of support vector regression, arttifical nenural networks, and random forest for peredicting soil organic carbon stocks across an afromonkane land scape. Ecological indicator. 394-403
28. Wilson, J.P., and Gallant, J.C. 2000. Terrain Analysis: Principles and Applications. In: G.J. Wilson JP (Ed.), Digital terrain analysis. John Wiley, New York. 478 pages.
29. Yang, L. I., Zho, S. H., Fang, W. U., Hong, Y. L. I., and Feng, L. I. 2005. Improved prediction of sampling density for soil salinity by different geostatistical methods. Agricultural Sciences in China. 6 (7): 632 – 841.
30. Zhu, A.X., Hudson, B., Burt, J.E., and Lubich K. 2001. Soil mapping using GIS, expert knowledge and fuzzy logic. Soil Science Society of America Journal. 65:1463-1472.

Comparison of Geostatistical and Random Forest Methods in Mapping Soil Salinity in the Lands of Zahak County of Sistan Plain

M. R. Pahlavan Rad[1], A. A. Moghadam, K. d. Dahmardeh, G. A. Keykha, N. Davatgar and M. M. Tehrani

Assistant Professor, Soil and Water Research Department, Sistan Agricultural and Natural Resources Research and Education Center, AREEO, Zabol, Iran; E-mail: pahlavanrad@gmail.com
MSc., Soil and Water Research Department, Sistan Agricultural and Natural Resources Research and Education Center, AREEO, Zabol, Iran; E-mail: alirezam.248@gmail.com
MSc., Soil and Water Research Department, Sistan Agricultural and Natural Resources Research and Education Center, AREEO, Zabol, Iran; E-mail: Dahmardehh1@yahoo.com
MSc., Soil and Water Research Department, Sistan Agricultural and Natural Resources Research and Education Center, AREEO, Zabol, Iran; E-mail: keykha309@yahoo.com
Assistant Professor, Soil and Water Research Institute, Agricultural Research, Education and Extension Organization(AREEO), Karaj, Iran; E-mail: n_davatgar@yahoo.com
Assistant Professor, Soil and Water Research Institute, Agricultural Research, Education and Extension Organization(AREEO), Karaj, Iran; E-mail: mtehrani2000@yahoo.com

Received: August, 2016 and Accepted: May, 2017

Abstract

Soil maps are the major source of information for land management, natural resources, and environment. Soil salinity is one of the most important factors limiting crop production in Sistan plain. In this research, geostatistical and random forest methods were compared to produce soil salinity maps. Preparation of accurate maps for soil salinity conditions is of great help in proper management of lands in this area. For this purpose, 460 composite soil samples were collected from 0-30 cm depth in 41000 ha of Zahak region of Sistan plain, using 750 m grid network. Then, the electrical conductivity of saturation paste was measured. Afterwards, 361 samples were used for training and 99 for testing. In the geostatistical model, different semi-variogram including circular, spherical, exponential, and Gaussian and different interpolation methods including inverse distance weighting, ordinary kriging, simple kriging, universal kriging and co-kriging were fitted and the best models were selected. In random forest model, digital soil mapping technique was used and environmental covariates were derived from digital elevation model (DEM) map and A Landsat 8 ETM+ image. The results of geostatistical method showed that soil salinity had a medium spatial correlation in the study area and the best semi-variogram and interpolation model were spherical and ordinary kriging, respectively. In random forest model, aspect, NDVI, and NDSI were the most important covariate in predicting soil salinity. The results Root Mean Square Error and Mean Error for the training and testing data showed that random forest method was slightly better than geostatistics. The use of other covariates such as land use and soil series maps can increase accuracy of the maps based on random forest methods, thereby improving decision-making. Using other environmental covariates that were not used in this study such as land unit and soil series map can also improve the accuracy of the map.

Keywords: Flat land, Digital soil mapping, Geostatistics

[1] Corresponding author: Soil and Water Research Department, Sistan Agricultural and Natural Resources Research and Education Center, Zabol, Iran.