

## ارزیابی روش‌های رگرسیون درختی و خطی چندگانه در برآورد ظرفیت تبادل کاتیونی

یاسر استواری<sup>۱\*</sup> - کامران عسگری<sup>۲</sup> - حمیدرضا متقیان<sup>۳</sup>

تاریخ دریافت: ۱۳۹۲/۱۱/۰۵

تاریخ پذیرش: ۱۳۹۳/۱۰/۰۹

### چکیده

با برآورد ظرفیت تبادل کاتیونی با استفاده از ویژگی‌های پایه‌ای و زودیافت خاک می‌توان در وقت و هزینه صرفه‌جویی کرد. هدف این بررسی، مقایسه دو روش رگرسیون درختی و رگرسیون خطی چندگانه در برآورد ظرفیت تبادل کاتیونی (CEC) با استفاده از ویژگی‌های خاک است. برای این منظور از داده‌های ۱۰۶ نمونه خاک UNSODA استفاده شد. جهت برآورد CEC با استفاده از روش‌های رگرسیون درختی و رگرسیون خطی از ویژگی‌های اجزای بافت خاک، pH، ماده آلی و چگالی ظاهری استفاده شد. کارایی روش رگرسیون درختی در برابر روش رگرسیون خطی چندگانه در برآورد CEC مقایسه شدند. نتایج نشان داد که در برآورد CEC با استفاده از روش‌های رگرسیونی تنها ضرایب ماده آلی (۳/۱۸۳) و درصد رس (۰/۲۷۴) که بیشترین همبستگی را با CEC دارند معنی‌دار شده و مدل رگرسیونی بر اساس این دو پارامتر توسعه یافت. همچنین از میان پارامترهای ورودی در روش رگرسیون درختی تنها پارامترهای ماده آلی و درصد رس در درخت رگرسیون ظاهر شد. روش رگرسیون درختی در دو مجموعه داده صحت‌سنجی و اعتبارسنجی بر اساس آماره‌های ارزیابی  $R^2$ ، RMSE، ME و GMER کارایی بالاتری نسبت به روش‌های رگرسیونی خطی در برآورد CEC داشت. از میان روش‌های رگرسیونی خطی، مدل پیشنهادی کارایی بالاتری نسبت به مدل‌های بل و ونکولن و بروسما و همکاران داشت.

واژه‌های کلیدی: توابع انتقالی، درخت رگرسیون، CEC، UNSODA

### مقدمه

(۱۲). رگرسیون درختی از روش‌های تجزیه و تحلیل در علوم محیطی است (۲۰ و ۲۱). این روش یکی از تکنیک‌های داده‌کاوی بر مبنای آشکار کردن ساختار داده‌ها است که از دو گروه متغیر کیفی و کمی به عنوان ورودی استفاده می‌کند. روش رگرسیون درختی نقطه بهینه جداسازی متغیرهای مستقل را تعیین و آن‌ها را به گروه‌هایی که تا حد ممکن از نظر متغیر همگن باشد تفکیک می‌کند (۲۰). برخی از مزایای روش رگرسیون درختی عبارتند از:

- **سادگی در نتایج:** در بیشتر موارد تفسیر نتایج خلاصه شده در درخت رگرسیون آسان است. سادگی نه تنها برای طبقه‌بندی سریع مشاهدات جدید، بلکه اغلب مدلی بسیار ساده در توصیف رفتار یک متغیر است. همچنین، برای برخی از اشخاص درک نتایج به صورت تصویری راحت‌تر از استفاده از معادلات رگرسیونی است (۲۳).

- **عدم نیاز به نرمال بودن توزیع داده‌ها:** روش رگرسیون درختی یک روش غیرپارامتری و غیرخطی است. نتایج نهایی روش رگرسیون درختی می‌تواند در قالب یک رشته شرایط منطقی اگر-سپس<sup>۶</sup> خلاصه و بیان کرد (۲۳).

ظرفیت تبدیلی کاتیونی (CEC<sup>۴</sup>) از مهم‌ترین خصوصیات خاک در ارتباط با توانایی خاک در نگهداری آب و مواد غذایی و همچنین شاخص خوبی برای کیفیت و بهره‌وری و مدیریت آلودگی خاک است که به عنوان ورودی در مدل‌های خاک و زیست محیطی مورد استفاده می‌باشد (۳ و ۱۴). برای نخستین بار، بوما (۱۰) با استفاده از روابط رگرسیونی متغیرهای دیرپافت خاک (متغیرهایی که اندازه‌گیری آن‌ها زمان و یا هزینه‌بر است) را با استفاده از متغیرهای زودپافت خاک (متغیرهایی که اندازه‌گیری آن‌ها به راحتی انجام می‌شود) بدست آورد و این روابط رگرسیونی را توابع انتقالی خاک (PTFs<sup>۵</sup>) نامید (۲۱). روش رگرسیونی خطی چندگانه روش رایجی است که هدف از آن، بیان متغیر وابسته به شکل تابعی ریاضی از متغیر(های) مستقل است

۱ و ۳- دانشجوی دکتری و استادیار گروه خاکشناسی، دانشکده کشاورزی، دانشگاه شهرکرد

\*- نویسنده مسئول: (Email: yaser.ostovary@gmail.com)

۲- دانشگاه آزاداسلامی، واحد خوراسگان (اصفهان)، باشگاه پژوهشگران جوان، اصفهان، ایران

4 - Cation Exchange Capacity

5 - Pedotransfser Functions

• **جلوگیری آسان از بیش‌برازشی:** این روش با امکان هرس شاخه‌ها با قرار دادن حداقل اندازه‌ی نمونه در گرهو اعتبارسنجی متقاطع از تقسیم شدن زیادی گره‌ها و ایجاد شاخه‌های زیاد جلوگیری می‌کند (۲۳).

در روش رگرسیون درختی ابتدا متغیرهای مستقل (پیوسته و کیفی) برای برآورد به نرم‌افزار ترسیم رگرسیون درختی معرفی می‌شوند. سپس نرم‌افزار متغیرهای موثر را انتخاب و آن‌ها را بر اساس اولویت به طور متوالی به دو گره تقسیم می‌کند. بر اساس این منطق، یک ساختار شبیه درخت ایجاد می‌شود. بنابراین، در نخستین گره، ویژگی دیربافت مد نظر به کمک متغیری که بیشترین سهم را در توجیه تغییرات آن دارد به دو گره فرزندی که در واقع میانگین کلی از متغیر دیربافت است تقسیم می‌شود. بهترین طبقه‌بندی زمانی انجام می‌شود که بهترین متغیر برای ایجاد بیشترین یکنواختی در گره‌های جدید انتخاب شود. مقدار متغیر تخمین زده شده در هر گره میانگینی از کل گره است (۲۴). با حرکت به سمت گره‌های پایین‌تر، همگن‌تر گره افزایش و پراش<sup>۱</sup> کاهش یافته و آن گره برآوردی دقیق‌تر از ویژگی دیربافت مورد نظر را ارائه می‌دهد (۲۰).

در سال‌های اخیر، توابع انتقالی متعددی در تخمین CEC از خصوصیات فیزیک و شیمیایی اساس خاک توسعه داده شده است (۵، ۹، ۱۰ و ۱۷). در اغلب این مدل‌ها CEC به عنوان تابع خطی از ماده‌آلی و مقدار رس در نظر گرفته شده است (۹، ۱۰ و ۱۷). در پژوهش‌های انجام شده به طور معمول از روش‌های رگرسیون چندخطی چندگانه و شبکه‌های عصبی-مصنوعی در پیدا کردن ضرایب مدل استفاده شده است، ولی تاکنون از روش رگرسیون درختی در تخمین ظرفیت تبادل کاتیونی استفاده نشده است. مکزی و جکویر (۱۹) و دهقانی‌بانیانی و همکاران (۱) روش‌های رگرسیون درختی و رگرسیون خطی چندگانه را در تخمین هدایت آبی اشیاع ارزیابی و مقایسه کردند. نتایج آن‌ها نشان داد که روش رگرسیون درختی تخمین‌های واقعی و دقیق‌تر یارائه می‌کند.

کروگ و همکاران (۱۶) از توابع انتقالی برای پیش‌بینی ظرفیت تبادل کاتیونی خاک‌های دانمارک به وسیله‌ی روش رگرسیون چندگانه برای ۱۶۴۳ نمونه‌ی خاک استفاده کردند. نتایج آن‌ها نشان داد که ۹۰ درصد از تغییرات در مقادیر CEC را می‌توان با استفاده از مقدار رس و موادآلی به عنوان متغیرهای وابسته توجیه کرد. میزان خطای پیش‌بینی این تخمین  $1/99 \text{ cmol}^+/\text{Kg}$  بود. فولادمند (۳) ظرفیت تبادل کاتیونی خاک را با استفاده از برخی از خصوصیات خاک تخمین زد. در این پژوهش ۲۰ نمونه خاک از مناطق اطراف شهرستان‌های مرودشت و فسا تهیه و ظرفیت تبادل کاتیونی و برخی دیگر از خصوصیات فیزیکی‌شیمیایی آن‌ها اندازه‌گیری شد. از روابط

همبستگی خطی و توانی جهت برآورد ظرفیت تبادل کاتیونی خاک استفاده شد. نتایج نشان داد که همبستگی خطی با درصد شن و همبستگی‌های توانی با میانگین هندسی قطر ذرات، درصد‌های رس و سیلت به ترتیب بهترین حالت‌های برآورد ظرفیت تبادل کاتیونی خاک می‌باشند.

مهربانیان و همکاران (۵) به بررسی کارایی توابع انتقالی جهت تخمین ظرفیت تبادل کاتیونی خاک‌های آهکی و گچی استان یزد پرداختند. در این مطالعه از اجزای بافت خاک و ماده‌آلی در تخمین ظرفیت تبادل کاتیونی استفاده گردید. نتایج این پژوهش نشان داد که شبکه عصبی-مصنوعی با میلی‌اکی والان برصد گرم  $\text{RMSE} = 1/04$  و ضریب تبیین  $0/97$  دارای کارایی بالاتری نسبت به رگرسیون خطی دارد که احتمالاً به علت وجود روابط غیرخطی میان متغیرهای وابسته و پیش‌بینی شونده می‌باشند. کشاورزی و همکاران (۸) به مقایسه روش فازی و شبکه عصبی مصنوعی در تخمین CEC با استفاده از پارامترهای درصد رس و ماده‌آلی پرداخت. نتایج نشان داد که CEC همبستگی زیادی با درصد رس و ماده‌آلی دارد و روش فازی در مقایسه با روش شبکه عصبی-مصنوعی کارایی بالاتری در برآورد CEC دارد.

به نظر می‌رسد تاکنون پژوهشی مرتبط با رگرسیون درختی و ظرفیت تبادل کاتیونی انجام نشده باشد. پس، بدیهی است که هنوز جنبه‌های مختلف موضوع بررسی شده نباشد. از جمله تاکنون کارایی رگرسیون درختی با رگرسیون خطی در حالتی که ورودی‌ها یکسان باشند مقایسه نشده است. لذا هدف از این مطالعه مقایسه کارایی دو روش رگرسیون خطی چندگانه (شامل مدل پیشنهادی بر اساس خصوصیات زود یافت خاک و نیز مدل‌های بروسما و همکاران (۱۱) و بل و ونکولن (۹) و روش رگرسیون درختی در برآورد ظرفیت تبادل کاتیونی خاک‌های UNSODA می‌باشد.

## مواد و روش‌ها

در این مطالعه از داده‌های ۱۰۶ نمونه خاک بانک اطلاعاتی<sup>۲</sup> UNSODA استفاده شد. این بانک اطلاعاتی شامل داده‌های هیدرولیکی (هدایت هیدرولیکی و منحنی رطوبتی) و کمکی (توزیع اندازه ذرات، ماده آلی و CEC) ۸۰۰ نمونه خاک دنیا است. داده‌ها به دو دسته ۸۱ تایی و ۲۵ تایی تقسیم شدند. دسته ۸۱ تایی که در اینجا از آن با نام داده‌های صحت‌سنجی یاد می‌شود برای توسعه مدل رگرسیونی خطی چندگانه و رگرسیون درختی مورد استفاده قرار گرفت. از دسته ۲۵ تایی برای سنجش اعتبار روش‌های رگرسیون درختی و رگرسیون خطی چندگانه استفاده شد. داده‌های دسته دوم

۱:۱، از معیارهای ضریب تعیین ( $R^2$ )، میانگین خطا (ME)، میانگین مربعات خطا (RMSE) و نسبت میانگین هندسی خطا (GMER) جهت آزمون صحت و اعتبار سنجی روش رگرسیون درختی و مقایسه‌ی آن با توابع رگرسیون چندگانه استفاده گردید. ضریب تعیین مستقیماً از نرم‌افزار استخراج شد ولی آماره‌های دیگر از روابط زیر به دست آمدند:

$$ME = \sum[(P_i - O_i)/n] \quad (1)$$

$$RMSE = \sqrt{(\sum[(P_i - O_i)^2]/n)} \quad (2)$$

$$GMER = \exp [\sum \ln (P_i / O_i)] \quad (3)$$

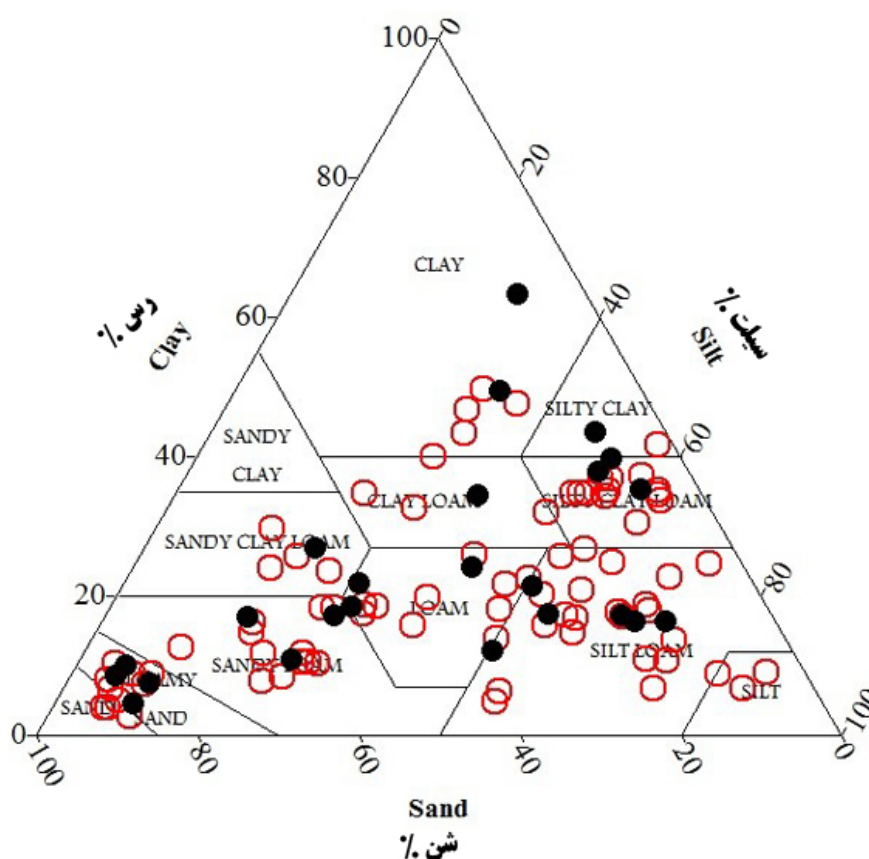
که در آن‌ها  $P_i$ : مقادیر برآورد شده CEC،  $O_i$ : مقادیر اندازه‌گیری شده CEC،  $n$ : تعداد نمونه‌ها است (۸). مقایسه میانگین پارامترها در مجموعه داده‌های صحت و اعتبارسنجی با آزمون  $t$  و در سطح ۵ درصد انجام شد. کلیه محاسبات آماری در محیط نرم‌افزار STATISTICA 8.0 انجام گرفت.

طوری انتخاب شدند که دارای توزیع بافتی یکسانی با داده‌های دسته‌ی اول باشند.

از همان ویژگی‌های زودیافت به کارگرفته شده در روش رگرسیون چندگانه به عنوان ورودی در روش رگرسیون درختی در برآورد CEC استفاده شد. چگالی ظاهری، pH، درصد رس و ماده‌آلی به عنوان متغیرهای پیوسته ورودی در نظر گرفته شدند. برای ایجاد رگرسیون درختی از مدول رگرسیون درختی استاندارد نرم‌افزار STATISTICA 8.0 و با شرایط زیر استفاده شد: برای جلوگیری از بزرگ‌شدن بیش از اندازه درخت رگرسیون، حداقل اندازه نمونه در هر گره ۲۸ و حداکثر تعداد گره ۳۰ تا و شرط شکستن گره‌ها بر مبنای واریانس گذاشته شد.

### ارزیابی کارایی مدل‌ها

جهت ارزیابی و مقایسه روش رگرسیون درختی و مدل‌های رگرسیونی خطی چندگانه از خطوط ۱:۱ استفاده شد. علاوه بر خطوط



شکل ۱- توزیع بافت خاک. دایره‌های توخالی و توپر به ترتیب داده‌های صحت‌سنجی توپر داده‌های اعتبارسنجی هستند  
Figure 1- Distribution of soil texture. Full and empty circles are development and validation data sets, respectively

جدول ۱- خلاصه آماری CEC و برخی از خصوصیات دیگر خاک (A) و (B) به تریب داده‌های صحت‌سنجی و اعتبار سنجی خاک UNSOD  
Table 1- statistical descriptive of CEC and some soil properties. (A) and (B) are development and validation data sets, respectively

ویژگی Properties	واحد Unit	Mean میانگین		Median میانه		Minimum کمینه		Maximum بیشینه		انحراف معیار Standard deviation	
		A	B	A	B	A	B	A	B	A	B
		رس Clay	%	18.3 <sup>a</sup>	15.8a	15.0	15.9	1.2	2.0	58.3	63.1
شن Sand	%	45.6 <sup>a</sup>	47.2a	41.3	49.0	4.0	4.0	89.8	86.0	29.6	28.9
سیلت Silt	%	40.5	38.4a	38.2	39.5	4.3	3.9	85.5	70.1	21.2	23.3
pH	—	5.7	5.5	5.8	5.6	3.3	3.2	8.3	8.1	2.2	2.3
BD	g m <sup>-3</sup>	1.40a	1.46a	1.45	1.49	1.02	1.07	7.76	1.69	0.18	0.14
OM	%	0.70a	0.81a	0.75	0.78	0.08	0.08	7.70	7.50	1.5	2.1
CEC	meq/100g	16.5a	20.4a	16.6	23.3	0.7	0.7	40.3	37.6	9.4	9.9

## نتایج و بحث

شکل ۱ توزیع بافت خاک‌های به کار رفته شده را نشان می‌دهد. گروه‌های خاک استفاده شده در ۱۱ کلاس از ۱۲ بافت کلاس USDA قرار دارند (شکل ۱). طبق این نمودار نمونه‌ای در کلاس رسی شنی (SaC) قرار نداشت. توزیع دو دسته داده‌های صحت و اعتبارسنجی در کلاس‌های بافتی مشابه بود (شکل ۱). خلاصه آماری بافت، ماده‌آلی، چگالی ظاهری و CEC نمونه‌های خاک در جدول ۱ آمده است.

میانگین اجزای بافت، چگالی ظاهری، ماده‌آلی، pH و CEC در دو مجموعه داده‌های صحت و اعتبارسنجی تفاوت معنی‌داری باهم ندارند ( $p > 0.05$ ). مقدار میانگین و میانه متغیرها نزدیک به هم بوده حاکی از نرمال بودن توزیع داده‌ها است (جدول ۱). در داده‌های صحت‌سنجی شن با میانگین ۴۵/۶ درصد بیشترین و رس با میانگین ۱۸/۳ درصد کمترین مقدار را در اجزای بافت خاک دارد (شکل ۱، جدول ۱). همین‌طور، بیشینه رس ۵۸ درصد است درحالی‌که بیشینه شن به ۸۹ درصد هم می‌رسد (جدول ۱). بنابراین تعدادی از نمونه‌های خاک در کلاس سبک: شنی (S) و شنی‌لومی (LS) قرار داشتند.

در داده‌های صحت‌سنجی مقدار چگالی ظاهری خاک‌ها از ۱/۰۲ تا ۱/۷۶ گرم بر سانتی‌متر مکعب متغیر است. میانگین مقدار ماده‌آلی ۰/۷۰ درصد و از ۰/۰۸ تا ۷/۷۰ درصد متغیر است. میانگین ظرفیت تبادل کاتیونی ۱۶/۵ میلی‌اکی والان بر صدگرم و از حداقل ۰/۷ تا حداکثر ۴۰/۳ میلی‌اکی والان بر صدگرم در مجموعه داده‌های صحت‌سنجی متغیر است (جدول ۱). اگرچه سعی شده است که خلاصه آماری دو مجموعه داده به هم نزدیک باشند وجود اندکی تفاوت، حتی اگر معنی‌دار نباشد، برای هر متغیر در بین دو مجموعه داده طبیعی است.

همبستگی ظرفیت تبادل کاتیونی با پارمترهای اجزای بافت خاک،

چگالی ظاهری، ماده‌آلی و pH خاک‌های UNSODA در جدول ۲ آمده است. نتایج نشان داد که CEC با ماده‌آلی ( $r = 0.70$ )، درصد رس ( $r = 0.59$ ) همبستگی مستقیم و با درصد شن ( $r = -0.35$ ) و چگالی ظاهری ( $r = -0.41$ ) همبستگی معکوس معنی‌داری دارد (جدول ۲). CEC با دیگر پارامترها (pH و درصد سیلت) رابطه معنی‌داری نداشت. کشاورزی و همکاران (۱۵) در پژوهش خود همبستگی بالاتر درصد رس ( $r = 0.92$ ) با CEC را نسبت به همبستگی CEC با ماده‌آلی ( $r = 0.56$ ) گزارش کردند. در پژوهش شیرانی و رفیع‌نژاد (۲) نیز همبستگی قوی CEC با درصد رس نسبت به ماده‌آلی گزارش شد. همبستگی منفی بین درصد شن و CEC به این دلیل است که با افزایش شن ذراتی که دارای توانایی ایجاد بار منفی هستند کاهش می‌یابد (۴). افزایش چگالی ظاهری در نتیجه افزایش مقدار اجزای شن است، لذا چنانچه گفته شد با افزایش شن توانایی ایجاد و نگهداشت بار منفی و در نتیجه CEC کاهش می‌یابد. در پژوهش معماریان‌فر و بیگی‌هرچگانی (۴) نیز بین CEC با ماده‌آلی و درصد رس رابطه‌ی مستقیم و با درصد شن رابطه‌ی معکوس معنی‌داری وجود داشت. میرخانی و همکاران (۶) نیز بین CEC با درصد رس و ماده‌آلی همبستگی خوبی مشاهده کردند. بین درصد شن با دیگر اجزای بافت خاک (درصدهای سیلت و رس) رابطه منفی معنی‌داری وجود داشت که با پژوهش معماریان‌فر و بیگی‌هرچگانی (۴) همخوانی دارد.

## توسعه تابع رگرسیونی خطی برآورد CEC

جهت ایجاد تابع رگرسیونی خطی برآورد CEC، بر اساس جدول ۲ پارامترهایی که همبستگی مثبتی با CEC داشتند وارد مدل شدند. در رگرسیون چندگانه خطی اگر بین متغیرهای مستقل وابستگی خطی قوی مشاهده شود، هم‌راستایی چندگانه به وجود می‌آید.

جدول ۲- ضرایب همبستگی CEC با برخی از خصوصیات خاک

Table 2- Correlation coefficient of CEC with some soil properties

پارامتر	CEC	Bd	OM	pH	رس Clay	سیلت Silt	شن Sand
CEC	1.00						
Bd	-0.41*	1.00					
OM	0.70*	-0.57*	1.00				
pH	0.13	0.15	0.09	1.00			
رس Clay	0.59*	-0.12	0.45*	0.20	1.00		
سیلت Silt	0.14	0.08	0.02	0.03	0.19	1.00	
شن Sand	-0.35*	-0.03	-0.21	0.15	-0.49*	-0.88*	1.00

\* معنی‌داری در سطح ۵٪ را نشان می‌دهد.

\* shows significant in 5% level.

عامل آنیونی به مقدار زیاد، مکان مناسبی برای جذب و تبادل کاتیون‌ها می‌باشد (۲).

هورن و همکاران (۱۳) نیز از یک تابع انتقالی با متغیرهای ورودی درصد رس و ماده‌آلی به صورت  $CEC = a_0 + a_1 OC + a_2 \text{ Clay}$  در برآورد CEC استفاده کردند. در مطالعه کروگ و همکاران (۱۶) در پیش‌بینی ظرفیت تبادل کاتیونی خاک‌های دانمارک به وسیله روش رگرسیون چند متغیره، ۹۰ درصد از تغییرات در مقادیر CEC با استفاده از مقدار رس و مواد آلی به عنوان متغیرهای وابسته توجیه شد. در پژوهش شیرانی و رفیع‌نژاد (۲) نیز تنها ضرایب متغیرهای درصد رس و ماده‌آلی معنی‌دار شد، اگرچه اثر درصد شن بر CEC به صورت کاهشی بود با این حال در مدل رگرسیونی این اثرات معنی‌دار نشد. تابع رگرسیونی توسعه یافته با ۲ تابع انتقالی رگرسیونی کلاسیک زیر در برآورد CEC مقایسه می‌شود:

$$CEC = 0.301 \text{ Clay} + 1.09 \text{ OM} \quad (۳) \quad \text{بروسما و همکاران (۱۱)}$$

$$CEC = 6.49 + 0.1 \text{ Clay} + 0.34 \text{ OM} \quad (۴) \quad \text{بل و ونکولن (۹)}$$

که در این معادلات، OM: درصد ماده‌آلی و Clay: درصد رس است. چنانچه از تابع توسعه یافته (معادله ۲) مشخص است مقدار ضریب عرض از مبدأ معنی‌دار شده و مقدار آن برابر ۶/۳۵ میلی‌اکی والان برصد گرم است که این ضریب به مقدار ضرایب عرض از مبدأ (۶/۴۹ میلی‌اکی والان برصد گرم) در توابع مانریکو و همکاران (۱۷) و بل و ونکولن (۹) نزدیک است.

### توسعه رگرسیون درختی برآورد CEC

از همان متغیرهای ورودی در ایجاد تابع رگرسیون خطی چندگانه (ماده‌آلی، درصد رس، درصد شن و چگالی ظاهری) در برآورد CEC

از میان متغیرهای مستقل بین درصدهای شن و سیلت همبستگی قوی ( $r = 0.88$ ) وجود داشت که احتمال هم‌راستایی چندگانه را ایجاد می‌کرد ولی چون درصد سیلت با CEC رابطه‌ی معنی‌داری نداشت، این متغیر در ایجاد تابع انتقالی وارد مدل نشد، لذا احتمال هم‌راستایی چندگانه در ایجاد تابع رگرسیونی وجود نداشت. شکل تابع رگرسیونی به صورت زیر است:

$$CEC = a_0 + a_1 \text{ Sand} + a_2 \text{ Bd} + a_3 \text{ Clay} + a_4 \text{ OM} \quad (۴)$$

که در این معادله،  $a_0$ ،  $a_1$ ،  $a_2$ ،  $a_3$  و  $a_4$  ضرایب پارامترهای ورودی مدل هستند. چنانچه از این معادله مشخص است تنها pH و درصد سیلت به دلیل عدم همبستگی با CEC وارد مدل نشدند. در پژوهش میکراتنی و همکاران (۱۸) در برآورد CEC نیز pH و درصد سیلت به دلیل عدم همبستگی با CEC به عنوان متغیر وارد مدل رگرسیونی نشدند.

از میان ضرایب پارامترهای ورودی مدل رگرسیونی، تنها ضرایب  $a_0$  (عرض از مبدأ)،  $a_3$  (ضریب درصد رس) و  $a_4$  (ضریب ماده‌آلی) در سطح ۵٪ معنی‌دار شدند، لذا مدل نهایی برآورد CEC به شکل معادله ۲ توسعه یافت:

$$CEC = 6.35 + 0.2735 \text{ Clay} + 3.2618 \text{ OM} \quad r=0.77, F=55.4, p<0.0001 \quad (۵)$$

که در این معادله رس و ماده‌آلی بر حسب درصد هستند. در بین ذرات معدنی خاک، ذرات رس بیشترین ظرفیت تبادل کاتیونی را دارا هستند. همچنین مواد نیز دارای ظرفیت تبادل کاتیونی بالایی می‌باشند. ذرات رس به دلیل داشتن سطح ویژه بالا، دارای مکان‌های بیشتری برای تبادل کاتیون‌ها می‌باشند. بنابراین بسته به نوع رس که دارای بار منفی متفاوت هستند، ظرفیت تبادل کاتیونی رس‌های مختلف، متفاوت است (۱۶). ماده‌آلی نیز به دلیل سطح ویژه بالا و داشتن گروه‌های

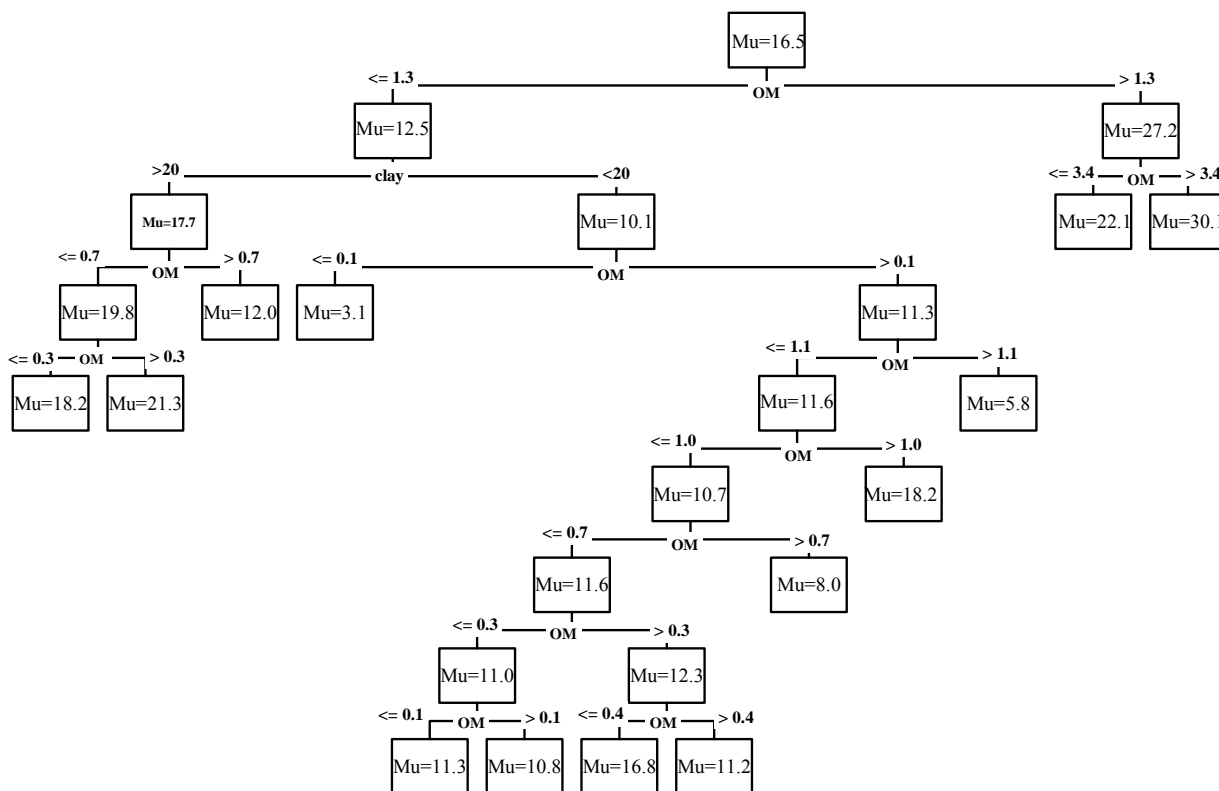
### صحت سنجی روش‌های رگرسیونی خطی و درختی

شکل ۳ نمودار خطوط ۱:۱ مقادیر اندازه‌گیری شده در برابر مقادیر تخمین زده شده CEC توسط روش‌های رگرسیونی خطی (مدل رگرسیونی پیشنهادی، مدل بروسما (۱۱) و همکاران و مدل بل و ونکولن (۹) و روش رگرسیون درختی در داده‌های صحت‌سنجی خاک‌های UNSODA را نشان می‌دهد.

روش رگرسیون درختی در تخمین مقدار CEC در داده‌های صحت سنجی از روش‌های رگرسیونی خطی کارایی بیشتری دارد. مقدار پراکندگی داده‌های حول خط ۱:۱ در روش رگرسیون درختی نسبت به روش‌های رگرسیونی خطی بیشتر است (شکل ۳).

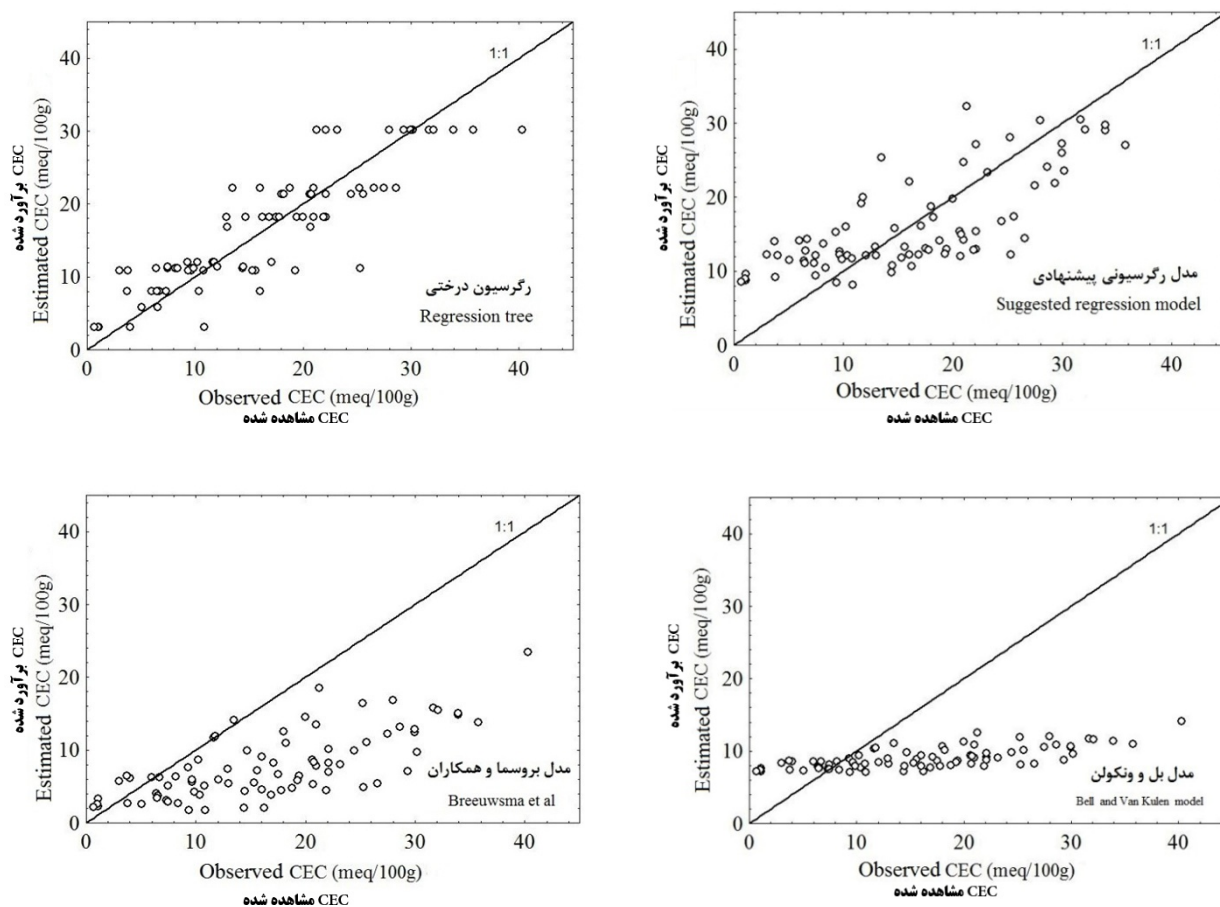
به عنوان ورودی جهت ایجاد درخت رگرسیون استفاده شد. شکل ۲ درخت رگرسیونی برآورد CEC خاک‌های UNSODA را نشان می‌دهد. همانند تابع رگرسیون خطی، از میان متغیرهای ورودی در روش رگرسیون درختی تنها ماده‌آلی و درصد رس در درخت رگرسیون ظاهر شد (شکل ۲). می‌توان این چنین بیان داشت که به دلیل همبستگی قوی پارامترهای ماده‌آلی و درصد رس با CEC، در درخت رگرسیون این دو پارامتر ظاهر می‌شود ولی پارامترهای درصد شن و چگالی ظاهری در درخت رگرسیون نمایان نمی‌شود.

اولین تقسیم‌بندی درخت رگرسیون بر اساس ماده‌آلی صورت گرفت و گره مادری به ۲ گره فرزندی تقسیم شد (شکل ۲). بیشتر تقسیم‌بندی‌ها در درخت رگرسیون بر اساس ماده‌آلی صورت گرفت که اهمیت ماده آلی و همبستگی قوی آن با CEC را نشان می‌دهد. درصد رس تنها در تقسیم‌بندی گره دوم ظاهر شد (شکل ۲).



شکل ۲- درخت رگرسیونی برآورد CEC با استفاده از متغیرهای درصد رس و ماده‌آلی. Mu: مقدار CEC در هر گره، OM: ماده‌آلی، Clay: درصد رس

Figure 2- tree structure of CEC estimation using clay and organic matter. Mu: CEC value in each node, OM: Organic matter, Clay:% Clay



شکل ۳- خطوط ۱:۱ مقادیر اندازه‌گیری شده در برابر مقادیر تخمین زده شده CEC توسط روش‌های رگرسیونی خطی (مدل پیشنهادی، مدل بروسما و همکاران و بل و ونکولن) و روش رگرسیون درختی در داده‌های صحت‌سنجی

Figure 3- 1:1 line observed versus estimated CEC using linear regression methods (suggested model, Bell and VanKolen and Breeuwsma et al) and tree regression in developing data set.

ونکولن (۹) نسبت به مدل پیشنهادی این پژوهش مربوط دانست. آماره‌های ارزیابی  $R^2$ ، RMSE، ME و GMER روش رگرسیون درختی و روش‌های رگرسیونی خطی (مدل پیشنهادی، مدل بروسما و همکاران و بل و ونکولن) در برآورد CEC در داده‌های صحت‌سنجی خاک‌های UNSODA در جدول ۳ آمده است.

همچنین بر اساس شکل ۳، تابع انتقالی پیشنهادی داده شده نسبت به توابع بروسما و همکاران (۱۱) و بل و ونکولن (۹) کارایی بیشتری دارد. توابع بروسما و همکاران (۱۱) و بل و ونکولن (۹) کم-برآوردی زیادی را نشان می‌دهد (پراکنش بیشتر داده‌ها در زیر خط ۱:۱ قرار دارد؛ شکل ۳). دلیل این کم‌برآوردی را می‌توان به ضریب بسیار کمتر ماده‌آلی در مدل‌های بروسما و همکاران (۱۱) و بل و

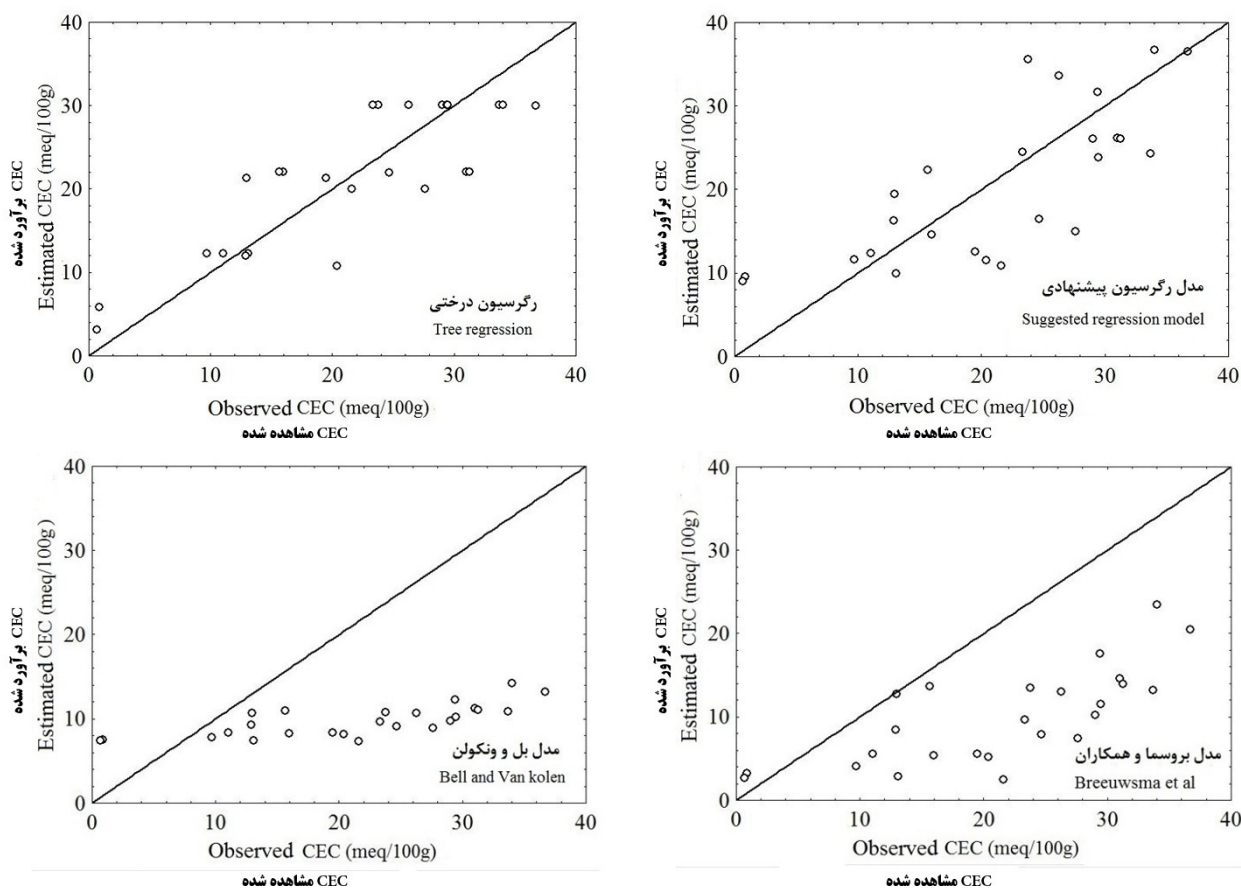
جدول ۳- آماره‌های ارزیابی  $R^2$ ، RMSE، ME و GMER روش رگرسیون درختی و روش‌های رگرسیونی خطی در داده‌های صحت‌سنجی

Table 3- evaluating indices  $R^2$ , RMSE, ME and GMER tree regression and linear regression methods in developing data set.

Regression model	مدل رگرسیونی	$R^2$	RMSE (meq/100g)	ME (meq/100g)	GMER
Tree regression	رگرسیون درختی	0.79	4.3	0.001	1.08
Suggested linear regression	مدل رگرسیونی خطی پیشنهادی	0.59	6.0	-0.005	1.17
Bell and VanKolen	مدل بل و ونکولن	0.51	11.2	-7.5	0.69
Breeuwsma et al	مدل بروسما و همکاران	0.52	11.1	-8.7	0.50

اندازه‌گیری شده نزدیک بوده و پراکندگی حول خط ۱:۱ بیشتر است).  
 آماره ME نشان‌دهنده وجود اریبی است (۲۲). مقادیر مثبت آن تخمین بیش از مقدار واقعی و مقادیر منفی آن تخمین کمتر از مقدار واقعی را نشان می‌دهد. هر اندازه آماره ME به صفر نزدیک‌تر باشد تخمین از اریبی کمتری برخوردار است (۲۲). بر اساس این آماره، روش رگرسیون درختی مقدار بیش‌برآوردی بسیار ناچیزی را نشان می‌دهد (جدول ۳). در حالی که روش‌های رگرسیونی خطی کم‌برآوردی در تخمین CEC را داشت. مدل بروسما و همکاران (۱۱) و مدل بل و ونکولن (۹) کم‌برآوردی شدیدی را نشان می‌داد (در شکل ۳ کم‌برآوردی این دو مدل رگرسیونی خطی در برآورد CEC به طور وضوح مشخص است). مقدار آماره GMER برابر ۱ نشان‌دهنده تطابق دقیق مقدار تخمین زده شده و مشاهده شده می‌باشد. مقدار GMER کمتر از ۱ دلالت بر کم‌برآوردی و GMER بیشتر از ۱ دلالت بر بیش‌برآوردی مدل دارد (۸).

مقدار ضریب تعیین ( $R^2=0.79$ ) روش رگرسیون درختی در داده‌های صحت‌سنجی در برآورد CEC به طور چشمگیری از روش‌های رگرسیونی خطی بیشتر است (جدول ۳). از بین سه روش رگرسیونی خطی، مدل پیشنهادی مقدار  $R^2$  بالاتری نسبت به مدل‌های بل و ونکولن (۹) و بروسما و همکاران (۱۱) دارد ولی تفاوت زیادی بین مدل بل و ونکولن (۹) و بروسما و همکاران (۱۱) وجود ندارد (جدول ۳). آماره RMSE خطا را نشان می‌دهد. مقدار RMSE در حالت بهینه یا حالتی که مقادیر برآورد شده و اندازه‌گیری شده مساوی باشند صفر است. هر اندازه مقدار RMSE کوچکتر باشد تخمین از صحت بیشتری برخوردار است. بر اساس آماره RMSE روش رگرسیون درختی کمترین خطا (میلی‌اکی والان بر صد گرم  $=4/3$ ) را نسبت به سه روش رگرسیونی خطی دارد. شکل ۳ نیز مبین خطای کم‌تر روش رگرسیون درختی نسبت به روش‌های رگرسیونی خطی است (مقدار CEC تخمین زده شده به مقدار



شکل ۴- خطوط ۱:۱ مقادیر اندازه‌گیری شده در برابر مقادیر تخمین زده شده CEC توسط روش‌های رگرسیونی خطی (مدل پیشنهادی، مدل بروسما و همکاران و بل و ونکولن) و روش رگرسیون درختی در داده‌های اعتبار سنجی

Figure 4- 1:1 line observed versus estimated CEC using linear regression methods (suggested model, Bell and VanKolen and Breeuwsma et al) and tree regression in validating data set.



روش‌های رگرسیونی خطی (مدل توسعه یافته، مدل بروسما و همکاران و بل و ونکولن) در برآورد CEC در داده‌های اعتبارسنجی را نشان می‌دهد.

در داده‌های اعتبارسنجی نیز مشابه با داده‌های صحت‌سنجی روش رگرسیون درختی در برآورد CEC بر اساس آماره‌ها کارایی بالاتری نسبت به روش‌های رگرسیونی خطی دارد. بالاتر بودن ضریب تعیین ( $R^2 = 0.71$ )، کمترین مقدار مقدار خطا (میلی‌اکی والان برصدگرم  $RMSE = 5/3$ ) و ناریبی ناچیز این روش، کارایی بالاتر این روش نسبت به مدل‌های رگرسیونی خطی را تأکید می‌کند (جدول ۴). روش رگرسیون درختی در داده‌های اعتبارسنجی نسبت به داده‌های صحت‌سنجی مقدار  $R^2$  پائین‌تر و  $RMSE$  بیشتری دارد. همچنین بر اساس آماره‌های ME و GMER روش رگرسیون درختی در داده‌های اعتبارسنجی بر خلاف داده‌های صحت‌سنجی ارزیابی بیشتری (بیش‌برآوردی) دارد (جدول‌های ۳ و ۴).

بر اساس آماره‌ها، مدل رگرسیون خطی پیشنهادی نسبت به مدل‌های بل و ونکولن (۹) و بروسما و همکاران (۱۱) کارایی بالاتری دارد (جدول ۴). مقدار بالاتر ضریب تعیین ( $R^2 = 0.56$ )، خطای پائین‌تر (میلی‌اکی والان برصدگرم  $RMSE = 6/7$ ) کارایی بهتر و تخمین‌های صحیح‌تر مدل رگرسیونی پیشنهادی را نسبت به مدل‌های بل و ونکولن (۹) و بروسما و همکاران (۱۱) را نشان می‌دهد (جدول ۴). مدل‌های بل و ونکولن (۹) و بروسما و همکاران (۱۱) بر اساس آماره‌های ME و GMER در داده‌های اعتبارسنجی همانند داده‌های صحت‌سنجی کم‌برآوردی شدید این دو مدل را در برآورد CEC نشان می‌دهد (جدول ۴ و شکل ۴).

بر اساس این آماره در داده‌های صحت‌سنجی روش رگرسیون درختی بیش‌برآوردی و روش‌های رگرسیونی خطی کم‌برآوردی را در برآورد CEC نشان می‌دهد (جدول ۳). بر اساس آماره‌ی GMER نیز دو مدل بروسما و همکاران (۱۱) و مدل بل و ونکولن (۹) کم‌برآوردی بسیار زیادی را نشان می‌داد. به طور کلی بر اساس آماره‌ها روش رگرسیون درختی کارایی بسیار بالاتری نسبت به روش‌های رگرسیونی خطی در برآورد CEC دارد.

### اعتبارسنجی روش‌های رگرسیونی خطی و درختی

نمودار خطوط ۱:۱ مقادیر CEC تخمین‌زده شده در برابر مقادیر اندازه‌گیری شده با روش‌های رگرسیون درختی و خطی (مدل رگرسیونی پیشنهادی، مدل بل و ونکولن و مدل بروسما و همکاران) در داده‌های اعتبارسنجی در شکل ۴ آمده است.

در داده‌های اعتبارسنجی نیز همانند داده‌های صحت‌سنجی روش رگرسیون درختی در برآورد CEC کارایی بیشتری نسبت به روش‌های رگرسیونی خطی دارد (پراکنش حول خط ۱:۱ در روش درختی بیشتر از روش‌های رگرسیونی است؛ شکل ۳). کارایی مدل رگرسیونی پیشنهادی در داده‌های اعتبارسنجی همانند آن در داده‌های صحت‌سنجی از مدل‌های رگرسیونی بل و ونکولن (۹) و بروسما و همکاران (۱۱) بالاتر بوده و مقادیر CEC تخمین‌زده شده به مقادیر اندازه‌گیری شده نزدیک‌تر است (پراکنش داده‌ها حول خط ۱:۱ بیشتر است؛ شکل ۴). مدل‌های بروسما و همکاران (۱۱) و بل و همکاران (۹) در داده‌های اعتبارسنجی همانند داده‌های صحت‌سنجی کم‌برآوردی شدیدی را نشان می‌دهد (شکل ۴). جدول ۴ آماره‌های ارزیابی  $R^2$ ،  $RMSE$ ، ME و GMER روش رگرسیون درختی و

جدول ۴- آماره‌های ارزیابی  $R^2$ ،  $RMSE$ ، ME و GMER روش رگرسیون درختی و روش‌های رگرسیونی خطی در داده‌های صحت‌سنجی

Table 4- Evaluating indices  $R^2$ ,  $RMSE$ , ME and GMER tree regression and linear regression methods in validating data set.

Regression model	مدل رگرسیونی	$R^2$	RMSE (meq/100g)	ME (meq/100g)	GMER
Tree regression	رگرسیون درختی	0.71	5.3	-0.09	1.13
Suggested linear regression	مدل رگرسیونی خطی پیشنهادی	0.56	6.7	-0.74	1.13
Bell and VanKolen	مدل بل و ونکولن	0.53	14.2	-11.7	0.57
Breeuwsm a et al	مدل بروسما و همکاران	0.52	13.3	-11.5	0.50

CEC داشتند معنی‌دار شده و مدل رگرسیونی بر اساس این دو پارامتر توسعه یافت. همچنین از میان پارمترهای ورودی در روش رگرسیون درختی تنها پارامترهای ماده‌آلی و درصد رس در درخت رگرسیون ظاهر شد و بیشترین تقسیم‌بندی گره‌ها بر اساس ماده‌آلی صورت گرفت.

۲. بر اساس آماره‌های ارزیابی روش رگرسیون درختی CEC را در

### نتیجه‌گیری کلی

نتایج این پژوهش نشان داد:

۱. ظرفیت تبادل کاتیونی بیشترین همبستگی را با ماده‌آلی و درصد رس داشت. در برآورد CEC با استفاده از روش‌های رگرسیونی از میان متغیرهای ورودی به مدل رگرسیونی تنها ضرایب متغیرهای ماده‌آلی و درصد رس که بیشترین همبستگی را با

کار به راحتی انجام پذیر است. بنابراین می توان از نمودارهای درختی شکل ۲ برای برآورد CEC استفاده کرد.

۳. پیشنهاد می شود با جمع آوری نمونه های مختلف خاک از نواحی مختلف ایران و افزودن آن به مجموعه داده های UNSODA روندنماهای جدید و دقیق تری تهیه شود تا پیش بینی بهبود یابد.

مقایسه با توابع رگرسیون خطی چندگانه (مدل های رگرسیونی پیشنهادی، مدل بل و ونکولن و بروسما و همکاران) با دقت و صحت بیشتری پیش بینی می کند. بنابراین روش رگرسیونی درختی جایگزین مناسبی برای توابع رگرسیونی خطی در برآورد CEC است. ضمناً در رگرسیون درختی بایستی از بیش برآزشی (و به ویژه توسعه ی بیش از حد شاخه ها) جلوگیری شود و این

## منابع

- 1- Abbasi Y., Ghanbarian-Alavijeh B., Liaghat A.M., and Shorafa M. 2011. Evaluation of pedotransfer functions for estimating soil water retention curve of saline and saline-alkali soils of Iran. *Pedosphere*, 21(2): 230-237.
- 2- Bell M.A., and Van Kullen H. 1995. Soil pedotransfer function for four Mexican soils. *Soil Science Society of American Journal*, 59: 865- 871.
- 3- Bouma J. 1989. Using soil survey data for quantitative land evaluation. *Advance Soil Science*, 9:177-213
- 4- Breeuwisma A., Wosten J.H.M., Vleeshouwer J.J., Van Slobbe A.M., and Bouma J. 1986. Derivation of land qualities to assess environmental problems from soil surveys. *Soil Science Society of American Journal*, 50:186-190.
- 5- Dehghani Baniani s, Ghorbani Dashtaki .Sh, Mohamadi J., and Khodaverdilo H. 2011. Comparing the Performance of Multiple Linear Regression and Regression Tree to Predict Saturated Hydraulic Conductivity and the Inverse of Macroscopic Capillary Length. *Iranian water research*, 9:193-204. (In Persian with English abstract).
- 6- Foladmand H.R. 2008. Estimation of CEC using soil properties. *Jornal of agricultural science and environment*, 1:1-8. (In Persian).
- 7- Herbst M., and Diekkruger B. 2002. Statistical regression. Chapter 1 In: Vereecken H and Herbst M (Editors). *Developments in Soil Science. Volume 30. Development of pedotransfer functions in soil hydrology. Elsevier Science*, Pp: 3-18.
- 8- Hezarjaribi A., Nosrati Karizak F., Abdollahnezhad K., and Ghorbani Kh. 2013. The Prediction Possibility of Soil Cation Exchange Capacity by Using of Easily Accessible Soil Parameters. *Journal of Water and Soil*, 27(4):712-719. (in Persian with English abstract)
- 9- Horn R., Fleige F.H., Richter E. A., Czyz A. Dexter E., Diaz-Pereira Damitru R., Enarche, F., Mayol K., Rajkai D., Delarosa and Simota C. 2005. SIDASS project 5: Prediction of mechanical strength of arable soils and its effects on physical properties at various map scales. *Soil and Tillage Research*, 82:47-56.
- 10- Keller A., Von Steiger B., Van der Zee S.T., and Schuline R. 2001. A stochastic empirical model for regional heavy metal balances in agroecosystems. *Journal of Environmental Quality*, 30:1976-1989.
- 11- Keshavarzi A., Sarmadian F., and Labbafi R. 2011. Modeling of Soil Cation Exchange Capacity Based on Fuzzy Table Look-up Scheme and Artificial Neural Network Approach. *Modern Applied Science*, 5(1):153-165.
- 12- Krogh L., Madsen H.B., and Greve M.H. 2000. Cation exchange capacity pedotransfer functions for Danish soils. *Acta Agricultural Science and Sect. B, Soil and Plant Science*, 50:1-12.
- 13- Manrique L.A., Jones C.A., and Dyke P.T. 1991. Predicting cation exchange capacity from soil physical and chemical properties. *Soil Science Society of American Journal*, 55:787-794.
- 14- McBratney A.B., Minasny B., Cattle S.R., and Vervoort R.W. 2002. From pedotransfer function to soil inference systems. *Geoderma*, 93:225-253.
- 15- McKenzie N.J., Jacquier D.W. 1997. Improving the Field Estimation of Saturated Hydraulic Conductivity in Soil Survey. *Australian Journal of Soil Research*, 35:803-825.
- 16- Memarianfar M. and Beigi H. 2008. Comparison of artificial neural network and regression pedotransfer functions models for prediction of soil cation exchange capacity in Chaharmahal - Bakhtiari province. *Water and soil science*, 23(4):90-99. (In Persian with English abstract).
- 17- Mehrabian M., mehrjedi R., and Dehghani F. 2009. Assessing the Efficiency of Pedotransfer Functions for Estimating CEC in some Calcareous and Gypsiferous Soils of Yazd Province. *Jornal of Water and Soil Conservation*, 17(1):113-129. (in Persian with English abstract)
- 18- Mirkhani R., Shaban Poor Shahrestani M.V. and Saadat S. 2004. Prediction of soil retention curve using PTFs. *Journal of Agriculture science*, 15(3):151-162. (In Persian).
- 19- Pachepsky Y., and Schaap M. G., (Editors). 2004. Data mining and exploration techniques. Chapter 2 In: *Developments in Soil Science. Volume 30. Development of pedotransfer functions in soil hydrology. Elsevier Science*, Pp: 21-31.

- 20- Pachepskya Y.A., and Rawls W. J. 2006. Hydropedology and pedotransfer functions .Geoderma, 131:308– 316.
- 21- Schaap M. G. 2004. Accuracy and uncertainty in PTF predictions. Chapter 3 In:Pachepsky Y and Rawls W.J (Editors). Developments in Soil Science. Volume 30. Development of pedotransfer functions in soil hydrology. Elsevier Science, Pp: 33-46.
- 22- Shirani H., and Rafieinejad N. 2011. Prediction of Some Difficult-to-measure Soil properties in Kerman state soils using regression and artificial neural networks. Iranian Journal of Soil Research, 25(4):349-359. (In Persian).
- 23- StatSoft. 2011. STATISTICA (data analysis software system). Version 10. www.statsoft.com.
- 24- Wosten J.H.M., Pachepsky Y.A., and Rawls W.J. 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology, 251: 123-150.

## Assessment of Tree and Multiple Linear Regressions in Estimation of Cation Exchange Capacity

Y. Ostovari<sup>1\*</sup> - K. Asgari<sup>2</sup> - H. R. Motaghian<sup>3</sup>

Received:25-01-2014

Accepted:30-12-2014

**Introduction:** Estimation of cation exchange capacity (CEC) with reliable soil properties can save time and cost. Pedotransfer function (PTF) is a common method in estimating certain soil properties (e.g. CEC) that has been widely used for many years. One of the common techniques that have been used to develop PTFs is multiple linear regressions. In this method, all easily obtained soil properties are linearly related to certain soil properties. In addition to multiple linear regressions method, more complex techniques such as artificial neural networks and regression tree have been used to develop PTFs. The regression tree method is a well-known method for analyzing the environmental science which determines optimal separation point of independent variables. The purposes of this study were to evaluate and compare tree and multiple linear regressions in estimating cation exchange capacity with reliable soil properties.

**Materials and Methods:** For this work, 106 soil samples of Unsaturated Soil hydraulic database (UNSODA), which contain a wide range of soil texture classes, were used. The examples were divided into 2 sets including 81 and 25 soil samples for developing and validating multiple linear regression and tree regression, respectively. For estimating CEC with tree and multiple regressions, soil texture properties, organic matter, pH and bulk density were used. To develop multiple linear regressions and create the tree structure, at first, correlation between cation exchange capacity with other soil properties were evaluated; then, soil properties that had significant correlation were chosen to introduce software. As well, the suggested linear function and tree structure were compared with 2 famous pedotransfer functions including Bell and Van-kolen and Breeuwsma et al., which have been used for estimating CEC. For investigating the performance of multiple linear regression and tree regression to estimate CEC 1:1 lines, determination coefficient ( $R^2$ ), mean error (ME), root mean square error (RMSE), and geometric mean error (GMER) were used. Statistica 8.0 software that was developed by ESRI was used to develop multiple linear regressions and generate tree structure.

**Results and Discussion:** The results showed for developing multiple linear regression model to estimate CEC among all inputs parameters (sand, silt, clay, organic matter, pH and bulk density) only just two parameters including organic (with  $r=0.70$ ) and clay percentage (with  $r=0.59$ ) had a significant coefficient, so organic and clay percentage appeared, and suggested multiple linear regression models based on this two parameters, with coefficient of 3.183 and 0.274, respectively, were developed. Also, only organic matter and clay percentage from inputs parameter in tree were shown. In tree structure most nodes were divided into 2 Childs nodes based on organic matter and only in the left side of tree structure in the second level clay percentage was appeared. Regression tree in two data sets (validation and development) based on  $R^2$ , RMSE, ME and GMER had a high quality for CEC estimation than regression methods. Proposed linear regression model had high performance than Bell and Van-kolen and Breeuwsma et al. to estimate CEC.

**Conclusions:** The main aim of this study was to investigate the efficiency of multiple linear regression model and regression tree to predict cation exchange capacity (CEC) based on relationships between CEC and easily measurable soil properties. For this work, 106 soil samples of UNSODA data set were used. Results showed that just clay percentage and organic matter that had higher correlation with CEC appeared in suggested linear regression and tree structure. Based on 1:1 lines,  $R^2$ , RMSE, ME and GMER, tree regression model had higher performance than all linear regression models (suggested function, Bell and Van-kolen and Breeuwsma et al.) to estimate cation exchange capacity. As well, suggested function had more efficiency than Bell and Van-kolen and Breeuwsma to predict CEC.

**Keywords:** Cation Exchange Capacity, Regression Tree, Transfer Function, UNSODA

1, 3- PhD Student and Assistant Professor of Soil Science Department, Shahrekord University  
(\*-Corresponding Author Email: Yaser.ostovary@gmail.com)

2 -Young Research Club, Khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran