کاربرد مدل‌های برنامه‌ریزی ریاضی در برآوردهای محیطی

(مورد مطالعه: برآورد توابع انتقالی خاک با استفاده از رگرسیون حداقل مطلق خطا؛ مبتنی بر برنامه‌ریزی آرمانی)

عباس امینی[۱]، محسن باقری بداغ‌آبادی[۲]، رضا مهاجر[۳]، حمید قیومی محمدی[۴]
۱- گروه جغرافیا و برنامه‌ریزی روستایی، دانشکده علوم جغرافیایی و برنامه‌ریزی، دانشگاه اصفهان؛ ۲- عضو هیأت علمی مؤسسه تحقیقات خاک و آب، بخش خاکشناسی؛ ۳- عضو هیأت علمی خاک‌شناسی گروه کشاورزی، دانشگاه پیام نور شهرکرد و ۴- عضو هیئت علمی موسسه تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی اصفهان

چکیده

در این مقاله با استفاده از منطق روش‌های خطی بهینه‌یابی و رویکرد ریاضی برنامه‌ریزی آرمانی (GP)، بدیلی برای مدل‌های رگرسیون معمولی (LSM) ارائه شده که کمتر به توزیع داده‌ها حساس بوده و در مقابل توزیع‌های غیرنرمال و بویژه داده‌های پرت، نتایج معتبرتری ارائه می‌دهد. با استفاده از مدل مذکور با عنوان رگرسیون حداقل مقادیر مطلق خطا (LAV)، توابع انتقالی خاک برای ۱۲۰ نمونهٔ تصادفیِ از عمق صفر تا۲۰ سانتی‌متری از دو واحد همگون نقشه خاک با سری‌های شهرک و چهار محال در اطراف شهرکرد بدست آمد، و ظرفیت تبادل کاتیونی بر اساس مشخصه‌های بافت خاک (شن، لای و رس) و درصد مواد آلی برآورد شد. نتایج، حاکی از افزایش دقت برآوردهای مدل LAV در مقایسه با رگرسیون LSM بوده و خطای برآورد یا مانده‌های رگرسیونی کمتر در مدل LAV باعث شده که خط رگرسیون برازش داده شده بین مقادیر مشاهده شده و مقادیر برآورده شده در رویکرد LAV انطباق بیشتری با خط y=x داشته باشد.

واژه‌های کلیدی: توابع انتقالی خاک، ظرفیت تبادل کاتیونی، رگرسیون و برنامه‌ریزی آرمانی.

مقدمه

فنّاوری‌های آماری جزء جدا نشدنی پژوهش‌های علمی و توانایی بهره‌گیری از استنباط‌های آماری برای هر پژوهشگری ضروری است. این امر برای همهٔ مراحل انجام یک پژوهش مانند برآورد حجم نمونه، روش نمونه‌برداری، طرح آزمایشی و غیره دارای اهمیت بوده و بویژه انجام و اعتبار تفسیر و نتیجه‌گیری، بدون آمار تقریباً امکان‌پذیر نیست. الگوهای رگرسیونی از کاربردی‌ترین روش‌های آماری برای برآورد و پیش‌بینی متغیر یا متغیرهایی وابسته از روی مجموعه‌ای از متغیرهای مستقل است. هدف اصلیِ این مدل‌ها، برازاندن بهترین خط یا منحنی به مجموعه‌ای از نقاط مشاهداتی با کمترین انحراف است. رایج‌ترین و ساده‌ترین روش ریاضی برای یافتن خط رگرسیونی رهیافت کمترین مجموع مربعات (LSM: Least Square Method) است. از جمله موارد مشکل‌ساز در این رویکرد (LSM) وجود داده‌های دورافتاده و پرت در مجموعه داده‌های مشاهداتی است؛ چرا که مجذور کردن اختلاف این مقادیر از برآوردهای بدست آمده از رگرسیون، منجر به پدیدآمدن اعداد بزرگی می‌شود که با سنگین‌تر کردن وزن این نقاط سبب انحراف خط رگرسیون به سمت آنها می‌گردد و در نتیجه به اعتبار نتایج نهایی مدل خدشه وارد می‌سازد. از طرفی رویکرد LSM وابسته به فرض نرمال بودن توزیع خطاهای تصادفی است و بنابراین وقتی توزیع خطاها نرمال نیست این رویکرد ممکن است کارآیی خود را به شدت کاهش دهد (امینی فسخودی و دیگران، ۱۳۸۷). در عمل چنین داده‌هایی پیش ازتجزیه و تحلیل حذف می‌شوند. لیکن باید توجه داشت این داده‌ها همیشه ناشی از خطا نیستند و گاهی به دلایل مختلف چنین تغییرات شدیدی بین داده‌ها به طور طبیعی وجود دارد (مانند تغییرات شدید در ناپیوستگی‌های لایه‌های خاک، گسل‌ها و غیره). بنابرین حذف داده‌های پرت باید با احتیاط و توجیه‌پذیر باشد. با توجه به آنچه گفته شد استفاده از روش‌هایی که حتی با وجود داده‌های پرت تحلیل‌های معتبری را بدست بدهد امری ضروری به نظر می‌رسد.

مدل‌های رگرسیون در خاک: مدل‌های رگرسیون در خاک در بخش‌های گوناگونی بکار می‌رود ولی یکی از بخش‌های بسیار کاربردی که اساس آنرا مدل‌های رگرسیونی تشکیل می‌دهد توابع انتقالی خاک (Pedo Transfer Functions = PTFs) می‌باشد.

این نام را بوما (۱۹۸۹) برای روابطی که بین خصوصیات هیدرولیکی و خصوصیات زود یافت خاک مانند توزیع اندازه ذرات، کربن آلی و جرم مخصوص ظاهری بدست آورده بود، بکار برد.

در مدل‌سازی این توابع روش معمول همان رویکرد کمترین مجموع مربعات در مدل‌های رگرسیونی است؛ هر چند همگام با پیشرفت‌های رایانه‌ای و فنّاوری، روش‌های دیگری مانند شبکه‌های عصبی (مهاجر و همکاران، ۱۳۸۷)، فازی (تقی‌زاده و همکاران، ۱۳۹۴)، الگوریتم ژنتیک (پالیزوانزند و احمدی، ۱۳۹۴) و غیره نیز مورد استفاده قرار گرفته است، لیکن به دلیل سهولت انجام محاسبات در رویکرد رگرسیونی، این روش (LSM) همچنان بیشترین کاربرد را دارا است. بنابراین مشکل‌های گفته شده بالا (برای رویکرد LSM) در توابع انتقالی خاک نیز وجودخواهد داشت.

مدل‌های برنامه‌ریزی آرمانی (Goal Programming) در دهه ۱۹۶۰ توسط چارنز و کوپر معرفی شد و مونـوگرافی و مقالـه‌های علمـی زیادی را بخود اختصاص داده است (Aouni and Kettani, 2000). در این مدل‌ها تابع هدف می‌تواند به منظور کمینه‌سازی انحرافها از حالت بهینه تعریف و به جای بکارگیری مربع خطاها از قدر مطلق آنها استفاده گردد. مدل GP امکان در نظر گرفتن همزمان چندین هدف برای یافتن بهترین راه حل از بین مجموعه‌ای‌از راه حل‌های ممکن را فراهم می‌آورد. از مزایای GP می‌توان به انعطاف‌پذیری بیشتر آن برای مدل سازی فرآیند برآورد و ارزیابی اشاره کرد. بنابراین با چنین دیدگاه و رویکردی می‌توان به حل مدل‌ها رگرسیونی و تعین ضرایب مورد نظر پرداخت و به رهیافتی نوین و آسان در راستای بر طرف نمودن مشکل‌های گفته شده برای LSM دست یافت (امینی فسخودی و دیگران، ۱۳۸۷).

مقاله حاضر معرفی رویکردی نوین برای مدل‌های رگرسیونی و نشان دادن دقت و توانایی‌های کاربردی آن می‌باشد. برای این منظور به طور آزمایشی توابع انتقالی جهت برآورد ظرفیت تبادل کاتیونی خاک (CEC) خاک مورد برسی قرار گرفته است.

مواد و روش‌ها

منطقه مورد مطالعه، محدوده‌های مختلف از دو واحد همگون نقشه خاک با دو سری خاک شهرک و چهار محال را شامل می شود که در اطراف شهرکرد و حومه آن در استان چهار محال و بختیاری واقع شده است. رده‌بندی دو خاک غالب انتخاب شده به شرح زیر گزارش شده است:

خاک اول: سری شهرک با فامیل: Fine, Mixed, Mesic Typic Calcixerepts

خاک دوم: سری چهارمحال با فامیل: Fine, Carbonatic, Mesic Typic Calcixerepts

تعداد ۱۲۰ نمونه خاک ازعمق صفر تا۲۰ سانتی متری به طریق تصادفی برداشت شد. پس از آماده سازی نمونه‌ها، علاوه بر تعیین ظرفیت تبادل کاتیونی، مشخصه های بافت خاک (شن، لای و رس) و درصد مواد آلی با روش استاندارد تعیین شد. تابع انتقالی خاک به منظور برآورد CEC با استفاده از روش رگرسیون معمولی (LSM) و روش پیشنهادی (کمترین مجموع مقادیر مطلق LAV: Least Absolute Values) بدست آمد. چون هدف اصلی مطالعه مقایسه توانایی و دقت برآورد دو مدل LSM و LAV است، بنابراین یکبار بدون توجه به معنی‌دار بودن یا نبودن ضرایب B از نظر آماری، تمامی متغیرهای مستقل وارد تجزیه و تحلیل شدند و یکبار نیز با توجه به معنی‌دار بودن ضرایب در روش رگرسیون معمولی مقایسه دو روش گفته شده (LSM و LAV) انجام گرفت.

در رویکرد رگرسیون معمولی مدل مربوطه در محیط Excel به صورت زیر فرمول گردید:

$$y_i = b_0 + \sum_{j=1}^{m} x_{ij} b_j + \varepsilon_i \qquad i = 1,2,\ldots,n \qquad (1)$$

$$\min \sum_{i=1}^{n} \varepsilon_i^2$$

که در آن y متغیر وابسته (CEC) و اندیس i معرف iامین مشاهده و اندیس j برای متغیر مستقل (x) یا ویژگی jام است.

در GP می‌توان تابع یا تابع‌هایی را برای رسیدن به هدفی مشخص تعریف کرد. به این توابع، تابع هدف گفته می‌شود. معمولاً تابع هدف در GP به منظور کمینه‌سازی مجموع خطاها (انحراف از حالت بهینه) فرموله می‌گردد و در این راستا به جای مجذور

نمودن خطاها از مقادیر مطلق آنها (LAV) استفاده می‌شود. چون حل مدل‌های GP در بیشتر موارد مستلزم انجام عملیات تکراری برای رسیدن به بهترین جواب است، استفاده از رایانه بسیار ضروری می‌باشد و بدون آن بویژه برای حالات پیچیده تقریباً حل این مدل‌ها غیر ممکن است.

با توجه به رویکرد رگرسیونی کمترین مقادیر مطلق خطا (LAV) مدل برنامه‌ریزی آرمانی زیر (فرمول ۲) در قالب یک برنامه خطی فرموله گردید و در محیط EXCEL با انجام عملیات تکرار حل شد (امینی فسخودی و دیگران، ۱۳۸۷):

$$\sum_{i=1}^{n} | PVi - AVi | = \min \sum_{i=1}^{n} | \varepsilon_i | \ \min \qquad (۲)$$

St: $PV_1 + \varepsilon_1 = AV_1$

$PV_2 + \varepsilon_2 = AV_2$

$\vdots$

$PV_n + \varepsilon_n = AV_n$

که در این مدل :

PVi = برآورد بدست آمده برای مشاهده iام (= CEC برآورد شده iام)

AVi = مقدار واقعی برای مشاهده‌iام (= CEC واقعی iام)

|εi| = |PVi-AVi| = قدر مطلق خطا = قدر مطلق انحراف مقدار برآورد شده از مقدار واقعی

در پایان مقدار متغیر وابسته و به عبارتیCEC برآورد شده توسط توابع انتقالی خاک بر اساس مدل خطی معمول در رگرسیون به صورت فرمول ۳ بدست آمد:

$$CECi = B0 + B(S)*Xi(S) + B(Si)*Xi(Si) + B(C)*Xi(C) + B(OC)*Xi(OC) \qquad (۳)$$

که در آن:

B0 = ضریب ثابت، B = ضرایب مربوط به هر متغیر مستقل شن (S)، سیلت (Si)، رس (C) و کربن آلی(OC) و Xi = مقدار اندازه‌گیری شده برای هر متغیر مستقل است.

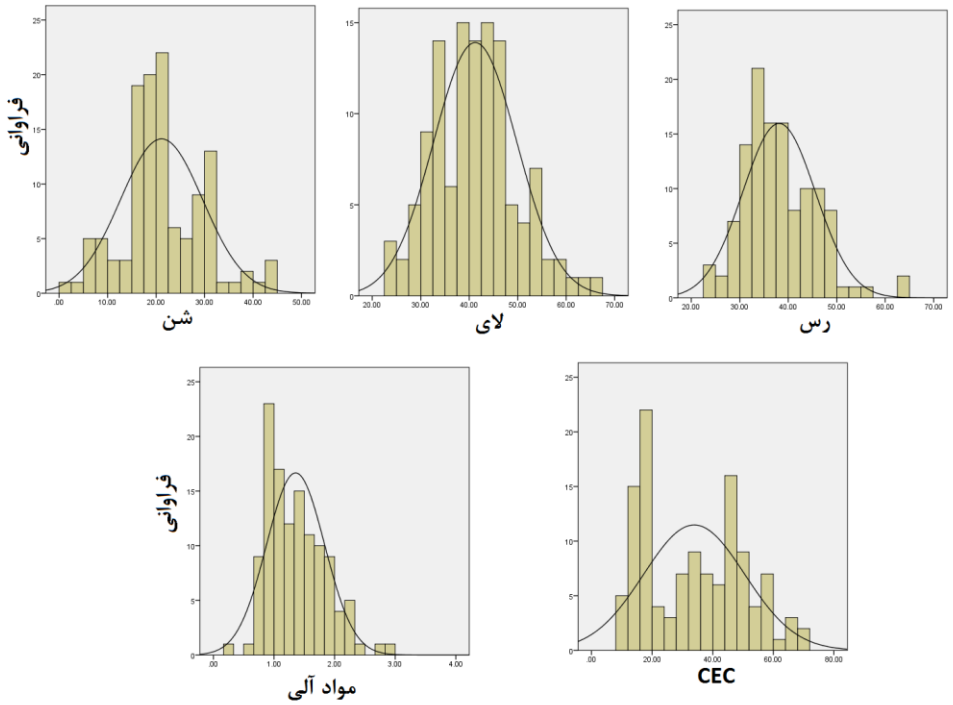
برای ارزیابی مدل‌ها از ضریب تبیین (R2)، ضریب x خط برازش برآوردها با مشاهدات، معیارهای میانگین خطا (Mean Error: ME)، مجذور میانگین مربعات خطا (Root Mean Square Error: RMSE) و مربع مطلق خطا (Mean Absolut Error: MAE) استفاده شد. در ارزیابی مدل هرچه معیارهای ME، RMSE و AME کوچکتر، ضریب x نزدیکتر به یک و R2 بزرگتر باشند نشان دهنده مدل بهتر و دقیقتر است.

نتایج و بحث

جدول ۱ مقادیر آماره‌های توصیفی داده‌ها را نشان می‌دهد. براساس یافته‌ها خاک‌های منطقه مورد مطالعه دارای دامنه تغییرات CEC زیاد برابر ۶۱/۷۱ می‌باشند. مقادیر چولگی و کشیدگی نشان دهنده نزدیک بودن داده‌ها به توزیع نرمال است. این موضوع با فراوانی داده‌ها در شکل ۲ نیز قابل مشاهده است. همان طور که دیده می‌شود داده‌ها تا حدودی از توزیع نرمال فاصله دارند لیکن چون انجام تبدیل داده (مانند لگاریتمی) نتایج را بهبود نبخشید از داده‌های اصلی در ادامه کار استفاده شد.

جدول ۱: مقادیر آماره‌های توصیفی داده‌ها

داده	دامنه	کمینه	بیشینه	میانگین		انحراف معیار	واریانس	چولگی		افراشتگی	
	آماره	آماره	آماره	آماره	خطای معیار	آماره	آماره	آماره	خطای معیار	آماره	خطای معیار
مواد آلی	۲/۵۷۰	۰/۳۲۰	۲/۸۹۰	۱/۳۵۶	۰/۰۴۴	۰/۴۷۹	۰/۲۲۹	۰/۶۴۷	۰/۲۲۱	۰/۱۴۲	۰/۴۳۸
شن	۴۲/۰۰۰	۲/۰۰۰	۴۴/۰۰۰	۲۱/۱۲۹	۰/۷۷۴	۸/۴۷۵	۷۱/۸۳۰	۰/۳۵۱	۰/۲۲۱	۰/۲۷۰	۰/۴۳۸
رس	۳۹/۱۰۰	۲۴/۶۰۰	۶۳/۷۰۰	۳۸/۱۰۵	۰/۶۸۵	۷/۵۰۰	۵۶/۲۴۴	۰/۸۰۴	۰/۲۲۱	۱/۰۰۱	۰/۴۳۸
لای	۴۲/۴۰۰	۲۳/۰۰۰	۶۵/۴۰۰	۴۱/۲۲۸	۰/۷۸۶	۸/۶۰۶	۷۴/۰۶۰	۰/۳۲۴	۰/۲۲۱	۰/۰۴۱	۰/۴۳۸
CEC	۶۱/۷۱	۹/۶۹	۷۱/۴۰	۳۳/۸۹	۱/۵۲	۱۶/۶۸	۲۷۸/۲۸	۰/۲۶۸	۰/۲۲۱	-۱/۱۲۹	۰/۴۳۸



شکل ۲: هیستوگرام داده‌ها به همراه منحنی نرمال

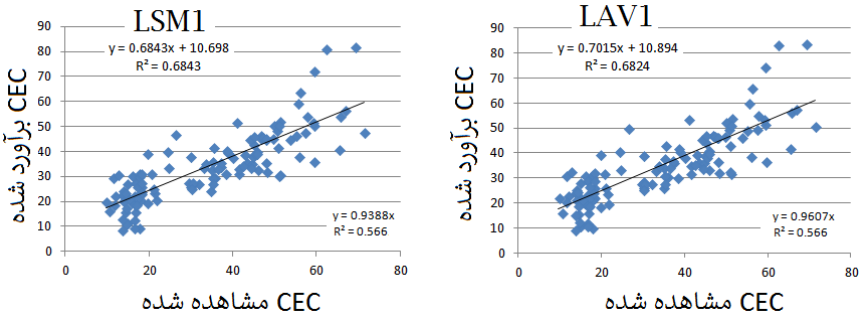
نتایج بدست آمده برای ضرایب مورد استفاده در توابع انتقالی خاک در روش‌های مختلف در جدول ۲ ارایه شده است. در حالت ۱ همه متغیرها و در حالت ۲ تنها متغیرهایی که از نظر آماری معنی دار هستند وارد مدل شده اند.

جدول ۲: مقادیر پارامترهای مربوط به واریوگرام‌های مدل‌ها و معیارهای ارزیابی برآوردها

متغیرها	حالت ۱		حالت ۲	
	LSM1	LAV1	LSM1	LAV1
رس (C)	۱/۹۲	۲/۶۵	۱/۴۲	۱/۳۹
سیلت (Si)	۰/۵۷	۱/۳۴	–	–
شن (S)	۰/۴۶	۱/۲۴	–	–
کربن آلي (OC)	۲۲/۸۸	۲۴/۰۸	۲۲/۸۱	۲۲/۶۶
ضریب ثابت	-۱۰۳/۲۵	-۱۸۰/۳۲	-۵۱/۲۶	-۴۸/۹۶
ME	۰/۰۰	-۰/۷۸	۰/۰۰	-۰/۸۹
RMSE	۹/۳۳	۹/۴۰	۹/۴۵	۹/۴۹
MAE	۷/۳۹	۷/۲۴	۷/۵۲	۷/۴۷
R2	۰/۶۸	۰/۶۸	۰/۶۸	۰/۶۸

نتایج حاصل از حالت ۲ با نتایج نوربخش و همکاران که رابطه CEC = ۴/۷۴ + ۷/۳۵OM + ۰/۲۷C را با ضریب تبیین R2 = ۰/۶۶۹ ارایه دادند همخوانی دارد. از نظر معیار ME مدل LAV برآوردهای بهتری داشته‌است (ME برابر صفر است) با در نظر گرفتن معیار AME نیز می‌توان دید مدل LAV برآوردهای بهتری داشته است اما RMSE برای مدل LSM کمتر است که این با توجه به منطق مربوط به آن که باید "مجموع مربعات خطا" کمینه شود، همخوانی دارد.

ضرایب تبیین (R2) در تمام حالات مقدار تقریباً برابر ۰/۶۸ بود اما در مدل LAV ضریب X به یک نزدیکتر بود. شکل ۲ پراکنش نقاط برآورد شده را در دو روش LAV و LSM برای حالت ۱ نسبت به مقادیر واقعی نشان می‌دهد. لازم به یادآوری است، هرچه پراکنش نقاط حول خط y=x کمتر باشد نشانگر برآورد دقیق‌تر است و در حالتی که کاملاً نقاط روی این خط قرار گیرند یعنی مقادیر برآورد شده و واقعی کاملاً برابرند. با توجه به خط رگرسیون برازش داده شده بین مقادیر واقعی و مقادیر برآورده شده نیز مشخص می‌شود که در روش LAV ضریب X به یک نزدیکتر است به عبارتی این خط انطباق بیشتری با خط y=x دارد. برای مقایسه واقعی‌تر نسبت به خط y=x می‌توان خط رگرسیون را از مبدا مختصات عبور داد و اختلاف ضریب x را از خط y =x بدست آورد؛ مقدار این اختلاف برای روش LSM برابر با۰/۰۶۱۲ و برای روش LAV برابر۰/۰۳۹۳ است. این موارد نیز برآورده‌های بهتر روش LAV را نشان می‌دهد.



شکل ۲: مقادیر برآوردشده CEC در برابر مقادیر مشاهده شده (واقعی)

نتیجه‌گیری

با توجه به وجود فناوری‌های رایانه‌ای و امکان انجام محاسبات پیچیده و طولانی و بویژه عملیات تکرار، بکارگیری روش‌هایی نوین که در عین سادگی کارآیی بالایی هم داشته باشند، برای حل مسایل نه تنها جایگاه ویژه‌ای را به خود اختصاص می‌دهد، بلکه امری ضروری به نظر می‌رسد. از جمله این روش‌ها مدل‌های برنامه‌ریزی آرمانی می‌باشد. این مدل‌ها علاوه بر پشتیبانی علمی و منطقی بالایی که در تصمیم‌گیری‌ها دارند، از انعطاف‌پذیری قابل قبولی نیز برخوردارند و بر این اساس می‌توان با بکارگیری آنها در زمینه‌های گوناگون به حل مسایل پرداخت. از جمله بخش‌های کاربردی و بسیار مهم در آمار، مدل‌های رگرسیونی است که همواره با مشکلات و محدودیت‌هایی مانند نرمال نبودن توزیع داده‌ها، وجود داده‌های پرت، تعداد و اندازه داده‌ها و غیره روبرو است. چنین مشکلاتی می‌تواند از عوامل ترغیب کننده پژوهشگران در بکارگیری مدل‌های برنامه‌ریزی آرمانی در حیطه روش‌های آماری باشد. نتایج مقاله حاضر تاییدی بر توانایی‌ها، دقت و سادگی کاربرد این مدل‌ها برای حل مسایل است.

منابع

امینی فسخودی ع.، خرسندی ج. و حمدحیدری ش. ۱۳۸۸. اصلاح الگوهای رگرسیونی به‌کمک مدل‌های برنامه‌ریزی آرمانی و کاربرد آن در مطالعات برنامه‌ریزی و توسعه روستایی. جغرافیا و برنامه‌ریزی محیطی. ۲۰ (۱)، پیاپی ۳۳: ۱۰۰-۹۱.

پالیزوانزند، ح. و احمدی، ع. ۱۳۹۴. مقایسه مدلهای رگرسیون خطی، فازی و فازی-ژنتیک در برآورد گنجایش تبادل کاتیونی خاک تحقیقات کاربردی خاک ۶۸ (۳۹۳): صفحه ۸۶ تا ۱۰۰.

تقی‌زاده مهرجردی ر.، سرمدیان، ف.، ذوالفقاری ع. و جعفری، ا. ۱۳۹۴. پیش‌بینی ظرفیت تبادل کاتیونی خاک‌های ایران با استفاده از روش‌های گوناگون . مهندسی زراعی (مجله علمی کشاورزی) ۱ (۳۸): صفحه ۵۹ تا ۷۷.

مهاجر ر.، صالحی م. و بیگی هرچگانی ح. ۱۳۸۸. تخمین ظرفیت تبادل کاتیونی خاک با استفاده از رگرسیون و شبکه عصبی و اثر تفکیک داده‌ها بر دقت و صحت توابع . مجله علوم آب و خاک. ۱۳ (۴۹۹) : صفحه ۸۳تا۹۷.

نوربخش، ف.، جلالیان، ا. و شریعتمداری، ح.۱۳۸۲. تخمین گنجایش تبادل کاتیونی خاک با استفاده از برخی ویژگی‌های فیزیک و شیمیایی خاک. مجله علوم آب و خاک. ۷ (۳): صفحه ۱۰۷تا ۱۱۸.

Bouma, J. 1989. Using soil survey data for quantitative land evaluation. Advanced Soil Science, 9, 177-213.

Aouni, B. and Ossama, K. (2000), Goal programming model: A glorious history and a promising future, Euro. J. of Op. Res. Vol. 133, pp 225 - 231

Application of mathematical programming models in environmental projections; case of pedo transfer function estimation using LAV regression based on goal programming (GP) approach

A. Amini[1], M. Bagheri Bodaghabadi[2], R. Mohajer[3], H. Ghaiumi Mohammadi[4]

1- Dept. of Geographical Sciences and Planning, University of Isfahan; 2- Dept. of Soil Science, Soil and Water Research Institute; 3- Dept. of Agriculture, PNU University of Shahr-e-Kord; 4- Dept. of soil & Water, Soil & Water Research Institute (SWRI**)**

Abstract

The aim of this article is introducing an alternative for the least Square regression (LSM) models referring to basis the linear optimization methods and goal programming approach. The procedure is less sensitive against of non-normal distributions and especially more robust for outlier data. Using the method namely least absolute values (LAV) regression, the pedo transfer functions for 120 soil random samples of 0-20 cm depths obtained from two homogenous soil map units of Shahr-e-Kord with series of Shahrak and Chahar Mahal. The cation exchange capacity (CEC) was estimated flowingly, based on the soil texture (sand, silt, clay) and OC as the predictors. Results revealed the more precise estimations of the LAV approach compared to the LSM one. Because of the less residuals in LAV regression, the fitted line between the observed and estimated values of ECE was more coincidence with the y=x line, leads to the better and more accurate estimations.

Keywords: PTFs, CEC, Regression, Goal Programming.