مدل های خطی برگشت فازی در خاکشناسی

جهانگرد محمدی
دانشیار گروه خاکشناسی، دانشکده کشاورزی، دانشگاه شهرکرد

مقدمه

یکی از جنبه های مهم مطالعه خاک اطلاع داشتن از روابط و همبستگی بین خصوصیات مختلف خاک و بیان کمی آنها در قـــــالب مدل های آماری است. این مدل ها که اصطلاحاً توابع انتقالی خاک (Pedotransfer Functions) نامیده می شوند، عمدتاً شامل مدل های رگرسیون آماری هستند. به وسیله این مدل ها خصوصیات مهم خاک که اندازه گیری آنها پرهزینه و وقت گیر است به مجموعه ای از ویژگی های مبنایی خاک که به سادگی و با هزینه کمتری قابل اندازه گیری هستند، مرتبط می شوند (۱). اخیراً تلاش هایی جهت بکارگیری روش ها و فنون جدید در مدل بندی چنین توابعی صورت گرفته است که می توان برای نمونه به استفاده از شبکه های عصبی مصنوعی (Artificial Neural Networks) اشاره کرد (۳). معهذا تمامی این روش ها بر پایه فرض دقیق بودن متغیرهای مورد مطالعه و مشاهدات مربوط به آنها استوار بوده و همچنین روابـــط بین متغیرها نیز دقیق فرض می گردد. حال آنکه در سیستم های طبیعی مانند خاک، عموماً با مشاهدات یا روابط نادقیق بین متغیرها مطرح است. در چنین شرایطی باید از مدل هایی بهره گرفت که قادر به ارائه الگوهای مناسب تر باشند، به گونه ای که این مدل ها بتوانند انطباق بیشتری با دنیای واقعی داشته باشند. بدین ترتیب با توجه به ماهیت ابهامی و مشکوک پدیده های مرتبط با خاک و یا نادقیق بودن مقادیر اندازه گیری شده خصوصیات مختلف خاک، به نظر می رسد لزوم بهره گیری از روش های پردازش مبتنی بر نظریه مجموعه ها و منطق فازی در توابع پدوترانسفر بیش از پیش ضروری باشد (۴). هدف از مقاله حاضر معرفی رگرسیون فازی و نشان دادن قابلیت های کاربردی آن در بسط و توسعه مدل های خطی (رگرسیون) بمنظور برآورد توابع انتقالی خاک است.

مواد و روش ها

رگرسیون فازی

در مدل رگرسیون با ضرایب فازی که در تحقیق حاضر از آن استفاده شده است، فرض بر آن است که اختلاف بین مقادیر مشاهده شده مربوط به متغیر وابسته و مقادیر برآورد شده به وسیله مدل، ناشی از ابهام در ساختار مدل می باشد. این ابهام، در ضرایب مدل که اعداد فازی منظور می شوند، نمود می یابد (۵ و۶). بدین ترتیب در این نوع مدل رگرسیون خطی، هدف آن است که بر پایه مجموعه مشاهدات، $(y_1,\underline{x}_1)$ و ... و $(y_m,\underline{x}_m)$، ضرایب (اعداد) فازی $\tilde{A}_0$ و$\tilde{A}_1$ و... و$\tilde{A}_n$ را طوری بیابیم که مدل:

$$\tilde{y} = f(\underline{x}) = \tilde{A}_0 + \tilde{A}_1 x_1 + ... + \tilde{A}_n x_n$$

یک $\underline{x}_i=(x_{i1},...,x_{in});i=1,...,m$ و x_{ij}مدل بهینه باشد، که ام است. jام از متغیر iمقدار مشاهده

تعیین ضرایب فازی

در رگرسیون با ضرایب فازی، هدف آن است کـه بـر پایـه مشـاهدات غیرفازی، ضرایب $\tilde{A}_i$، $i=1,...,n$، به گونه ای تعیین شوند که: اولاً خروجی فازی، $\hat{y}$، برای تمـامی مقـادیر y_j، $j=1,...,m$، حـداقل دارای درجه عضویتی به بزرگی h باشد: $Y(y_i)\geq h$ $\forall j=1,...,m$

ثانیاً ابهام یا فازی بودن خروجی مدل حداقل ممکن باشد. شرط اول تضمین می کند که در مدل نهایی، مقدار عضویت y_i یعنی i امین مقدار مشاهده شده متغیر وابسته در برآورد فازی آن به وسیله مدل، یعنی $\hat{y}(y_i)$، حداقل به اندازه h باشد. مقدار h به وسیله کاربر انتخاب مـی گردد و می توان آن را به عنوان درجه برازش مدل تعبیر کرد. شرط دوم نیز بیانگر این نکته است که ابهام در مقدار پیش بینی متغیر وابسته باید حداقل باشد. چون بر پایه هر مشاهده، یک خروجی از مدل وجود دارد، پس باید مجموع ابهام های خروجی ها را حداقل شویم. به بیان دیگر مایل هستیم که مقدار زیر حداقل شود:

$$Z = \left(S_0^{\,L} + S_0^{\,R}\right) + \sum_{i=1}^{n}\left[\left(S_i^{\,L} + S_i^{\,R}\right)\sum_{j=1}^{m} x_{ji}\right]$$

پس مساله یافتن ضرایب فازی مدل، معادل با حداقل سازی تابع هدف Z با توجه به $2\times m$ محدودیت تولید شده توسط m مشاهده است. بدین ترتیب با یک مساله برنامه ریزی خطی مطرح است که به روش‌های مختلف از جمله روش سیمپلکس قابل حل است (۲).

داده های مورد استفاده در تحقیق حاضر، بخشی از مجموعه اطلاعات حاصل از مطالعه خاکهای منطقه دشت سیلاخور در استان لرستان است. نمونه برداری در محدوده ۱۰۰ هکتاری و به صورت شبکه های منظم با ابعاد ۲۰۰ متر در ۲۰۰ متر و از لایه سطحی ۳۰-۰ سانتیمتری واقع در مرکز هر شبکه انجام شد. خصوصیات خاک شامل درصد رس، سیلت و شن، اندازه گیری به روش پیپت، درصد رطوبت اشباع (SP)، درصد کربن آلی (OC)، درصد آهک و ظرفیت تبادل کاتیونی (CEC) است. دو تابع انتقالی شیمیایی و فیزیکی خاک با استفاده از روش رگرسیون معمولی و رگرسیون فازی جهت تخمین ظرفیت تبادل کـاتیونی (CEC) و درصد رطوبت اشباع (SP) برازش داده شد.

نتایج و بحث

نتایج حاصل از برازش توابع مورد نظر بـه وسـیله دو روش رگرسـیون آمـاری و رگرسـیون فـازی در جـداول (۱) و(۲) ارائـه شـده اسـت. همانگونه که در جدول (۲) ملاحظه می شـود، پارامترهـای محاسـبه شده در رگرسیون فازی به صورت اعداد فازی هستند که در برگیرنـده مقدار مرکزی آنها و همچنین درجه ابهام و یا عدم قطعیت امکانی آنها مـی باشـد. بـا اعمـال تصـحیح پیشـنهادی بـر روی روش تاناکـا و همکاران(۱۹۸۲)، مدل هایی حاصل گردید که ابهام مـدل در ضـرایب مربوط به متغیرهای توضیحی پخش می شود. این ابهام ها می تواننـد متقارن و یا نامتقارن باشند. به طور مثـال در تـابع انتقـالی مربـوط بـه ظرفیت تبادل کاتیونی، ضریب متغیر کربن آلی دارای ابهام بـا پهنـای نامتقارن است. بدین ترتیب که پهنای راسـت ابهـام، کـه در برگیرنـده مقادیر بزرگتر کربن آلی است، بزرگتـر و کشـیده تـر از پهنـای چـپ، منطبق با مقادیر کوچکتر کربن می باشد.

جدول(۱) نتایج حاصل از برازش توابع پدوترانسفر به روش رگرسیون معمولی

توابع پدوترانسفر	ضرایب (سطح معنی دار) a	b	c	d	R^2
CEC=a+b(Sand)+c(OC)	۲۱/۹۸۹ (۰/۰۰۰)	-۰/۲۲۲ (۰/۰۰۰)	۲/۴۶۹ (۰/۰۰۰)	-	۰/۸۵۰
SP=a+b(Silt)+c(Sand)+d(OC)	۷۰/۸۴۰ (۰/۰۰۰)	-۰/۴۱۲ (۰/۰۰۳)	-۰/۵۷۵ (۰/۰۰۰)	۷/۲۴۹ (۰/۰۰۰)	۰/۹۳۳

تجزیه و تحلیل حساسیت

به منظور ارزیابی مدل های حاصل از رگرسیون فازی، اقدام به تجزیه و تحلیل حساسیت با استفاده از سطوح اعتبار h و مقدار کل ابهام Z شد. نتایج مربوط به تابع پدوترانسفر ظرفیت تبادل کاتیونی در جدول (۳) ارائه شده است. بطورکلی بزرگ شدن مقدار h دو تاثیر دارد: اولاً سطح اعتبار مدل بالا می رود. زیرا مدل حاصله با h های بزرگ، با امکان بیشتری نقاطی را که فرآیند مدل سازی مبتنی بر آنها بوده است را شامل می شود. ثانیاً ابهام در مقادیر پیش بینی و در نتیجه مقدار کل ابهام مدل، Z، بزرگ تر می شود. نتایج نشان می دهد که این تاثیر تا h های حدود ۰/۷ تقریباً خطی است و از آن پس شتاب بیشتری می گیرد. طبیعی است که کوچک انتخاب کردن h تاثیری معکوس داشته باشد. در این حالات، از یک طرف ابهام کل مدل، Z، کوچک می شود و لذا یک مدل متقن تر خواهیم داشت، و از طرف دیگر سطح اعتبار مدل کاهش می یابد.

جدول(۲) نتایج حاصل از برازش توابع پدوترانسفر به روش رگرسیون فازی

توابع پدوترانسفر	ضرایب (مقدار ابهام) $\tilde{A}_3$	$\tilde{A}_1$	$\tilde{A}_0$ $\tilde{A}_2$
$CEC= \tilde{A}_0-\tilde{A}_1(Sand)+\tilde{A}_2(OC)$	۱۹/۰۱ (۰/۰۰-۰/۰۰)	-۰/۱۷ (۰/۱۹-۰/۱۱)	۳/۳۰ (۰/۵۲-۰/۹۹)	-
$SP=\tilde{A}_0+\tilde{A}_1(Silt)+\tilde{A}_2(Sand)+\tilde{A}_3(OC)$	۶۸/۱۱ (۰/۰۰-۰/۰۰)	-۰/۴۲ (۰/۰۴-۰/۰۴)	-۰/۵۴ (۰/۲۲-۰/۲۲)	۸/۵۲ (۰/۰۰-۰/۰۰)

همانگونه که از جدول فوق مشاهده می شود، تغییر مقـادیر h تـاثیری بر روی مراکز (مقدار میانی) اعداد فازی (ضرایب مـدل) نـدارد و تنهـا پهناهای اعداد فازی تغییر می کننـد. پهناهـای اعـداد فـازی بایـد بـه گونه ای انتخاب گردد تا منجر به تخمین های غیر واقعی نشود. بطـور مثـال، مـدل بنـدی ظرفیـت تبـادل کـاتیونی بـا سـطح اعتبـار ۰/۸ مـــنجر به تخمین CEC با بـازه ۱/۶۳- تـا ۲۷/۹۷ مـی شـود. بدیهی است که تخمین منفی برای ظرفیت تبادل کاتیونی غیر واقعی و نادرست است. بطور کلی انتخاب سطح اعتبار مدل و در نتیجه مقدار ابهام آن، در اختیار کاربر است کـه براسـاس شـناخت و آگـاهی وی از پدیده و سیستم مورد مطالعه و توازن مورد نظر بین اعتبار و ابهام مدل انتخاب می شود. البته می توان سطح اعتبار h =۰/۵ را به عنوان یک سطح اعتبار موجه و معقول در نظرگرفت.

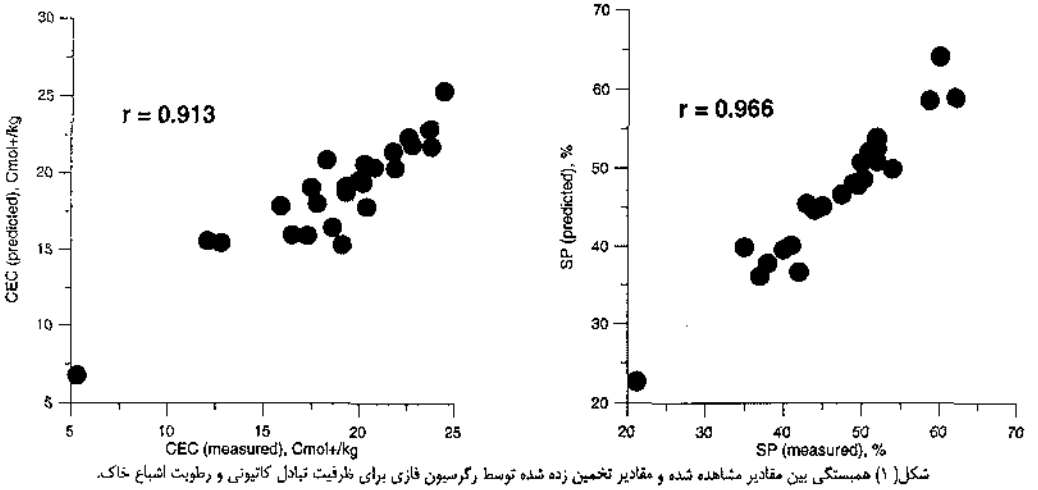
جدول (۳) نتایج تجزیه و تحلیل حساسیت برای تابع پدوترانسفر ظرفیت تبادل کاتیونی

سطح اعتبار(h)	(پهنای ابهام راست، پهنای ابهام چپ، پارامترمدل) پارامترهای مدل بهمراه پهنای ابهام چپ و راست	ابهام کل مدل (Z)
۰/۱	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۱۰ و ۰/۰۶) x_1 + (۳/۳۰ و ۰/۲۹ و ۰/۵۵) x_2$	۱۵۷/۵۳
۰/۲	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۱۲ و ۰/۰۷) x_1 + (۳/۳۰ و ۰/۳۳ و ۰/۶۳) x_2$	۱۷۷/۲۳
۰/۳	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۱۳ و ۰/۰۸) x_1 + (۳/۳۰ و ۰/۳۷ و ۰/۷۰) x_2$	۲۰۲/۵۴
۰/۴	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۱۶ و ۰/۰۹) x_1 + (۳/۳۰ و ۰/۴۳ و ۰/۸۲) x_2$	۲۳۶/۳۰
۰/۵	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۱۹ و ۰/۱۱) x_1 + (۳/۳۰ و ۰/۵۲ و ۰/۹۹) x_2$	۲۸۳/۵۶
۰/۶	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۲۳ و ۰/۱۴) x_1 + (۳/۳۰ و ۱/۲۴ و ۰/۶۵) x_2$	۳۵۴/۴۵
۰/۷	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۳۱ و ۰/۱۹) x_1 + (۳/۳۰ و ۱/۷۰ و ۰/۸۷) x_2$	۴۷۲/۶۰
۰/۸	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۴۷ و ۰/۲۸) x_1 + (۳/۳۰ و ۲/۴۷ و ۱/۳۰) x_2$	۷۰۸/۹۰
۰/۹	$y = (۱۹/۰۱ و ۰/۰ و ۰/۰) + (-۰/۱۷ و ۰/۹۳ و ۰/۵۶) x_1 + (۳/۳۰ و ۴/۹۴ و ۲/۶۰) x_2$	۱۴۱۷/۸۰

ارزیابی توابع پدوترانسفر

به منظور بررسی دقت تخمین های حاصل از رگرسیون فازی، همبستگی بین مقادیر مشاهده شده و مقادیر تخمینی برای دو تابع برازش داده شده در شکل (۱) ترسیم شد. همان گونه که ملاحظه می شود، مدل های رگرسیون فازی منجر به تخمین های قابل قبولی شده است. بدین صورت که ضرایب همبستگی خطی بین مشاهدات و تخمین ها در رگرسیون فازی برای توابع CEC و SP به ترتیب برابر ۰/۹۱۳ و ۰/۹۶۶ است. لازم به ذکر است که ضرایب همبستگی بین مشاهدات و تخمین ها توسط رگرسیون معمولی برای توابع CEC و SP به ترتیب برابر ۰/۹۲۷ و ۰/۹۶۵ بود.

در مطالعه سیستم هایی که روابط بین متغیرهای سیستم، نادقیق و مبهم است، در نظر گرفتن ساختار فازی برای مدل بندی آن سیستم ضروری است. در مورد توابع انتقالی خاک، این ساختار را می توان به صورت یک تابع خطی فازی که پارامترهای آن اعداد فازی تعیین می شوند، در نظرگرفت. در بسیاری از موارد به دلیل عدم تقارن در نحوه تاثیر متغیرها، باید از ضرایب نامتقارن در مدل بندی ها استفاده کرد. مشکلات و محدودیت های موجود در رابطه با کیفیت و اندازه (تعداد) داده های مورد استفاده در برازش توابع پدوترانسفر و همچنین ماهیت ابهامی مدلها در علوم خاک را می توان از عوامل ترغیب کننده خاکشناسان در بکارگیری روش رگرسیون فازی قلمداد کرد. معهذا حجم نسبتاً زیاد محاسبات، بکارگیری اعداد فازی غیرمثلثی مانند اعداد فازی نرمال و تاثیر آنها در نتایج حاصل، بسط الگوریتم های دیگر رگرسیون فازی خارج از چارچوب برنامه ریزی خطی و همچنین استفاده از روشهای رگرسیون فازی برپایه داده های فازی، از چالش های تحقیقاتی پیش روی متخصصان خاکشناسی کشور خواهد بود.



شکل(۱) همبستگی بین مقادیر مشاهده شده و مقادیر تخمین زده شده توسط رگرسیون فازی برای ظرفیت تبادل کاتیونی و رطوبت اشباع خاک.

منابع مورد استفاده

1- Bouma, J. 1989. Using soil survey data for quantitative land evaluation. Adv. Soil Sci., 9: 177-213.
2- Gass, S.I. 1975. Linear programming. McGraw-Hill, USA.
3- Minasny, B. and A.B. McBratney. 2002. The Neuro-m method for fitting neural network parametric pedotranfer functions. J. Soil Sci. Soc. Am, 66: 352-361.
4- Mohammadi, J. 2003. Variation and uncertainty in environmental sciences. Proc. of the Fourth Seminar on Fuzzy Sets and Its Applications, University of Mazandaran, Babolsar, pp: 136-141.
5- Tanaka, H., S. Uejima and K. Asai. 1982. Linear regression analysis with fuzzy model. IEEE Trans. Systems Man Cybernet. 12: 903-907.
6- Yen, K.K., S. Ghoshray and G. Roig. 1999. Linear regression model using triangular fuzzy number. Fuzzy Sets Syst, 106: 166-177.