

## تعیین ترکیب بهینه متغیرهای ورودی با استفاده از آزمون گاما برای مدل‌سازی پتاسیم قابل جذب در سیستم عصبی-فازی (مطالعه موردی: منطقه میانکنگی؛ زابل)

امین دلارامی<sup>۱</sup>، احمد غلامعلی زاده<sup>۲\*</sup>، اسما شعبانی<sup>۲</sup>

۱- دانشجوی کارشناسی ارشد، گروه علوم مهندسی خاک، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران

۲- گروه علوم مهندسی خاک، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران

\* مسئول مکاتبه: Ahangar@uoz.ac.ir

DOI: 10.22034/csrar.2020.119092

تاریخ پذیرش: ۱۳۹۷/۰۹/۱۸

تاریخ دریافت: ۱۳۹۶/۰۷/۳۰

### چکیده

یکی از مراحل مهم و پیچیده برای مدل‌سازی غیرخطی، پیش‌پردازش داده‌های ورودی به‌منظور انتخاب ترکیبی مناسب از آن‌ها در مدل می‌باشد. در این مطالعه آزمون گاما برای انتخاب ترکیب بهینه متغیرهای ورودی در مدل‌سازی پتاسیم قابل جذب استفاده شد. برای تعیین بهینه تعداد داده‌های مورد نیاز برای مدل‌سازی از آزمون M استفاده شد. به‌منظور مدل‌سازی تعداد هشت متغیر ورودی استفاده گردید. مدل‌سازی پتاسیم قابل جذب با استفاده از تعداد نقاط بهینه، متغیرهای منتخب با خوشه‌بندی کاهشی در سیستم عصبی فازی انجام شد. نتایج نشان داد که شش متغیر شامل (درصد رس، سیلت، ماده آلی، هدایت الکتریکی، رطوبت اشباع و pH) ترکیب بهینه متغیرها در مدل‌سازی پتاسیم قابل جذب در منطقه میانکنگی می‌باشند. همچنین با استفاده از خروجی آزمون M تعداد ۱۱۲ داده (۶۰ درصد داده‌ها) برای بخش آموزش مدل‌سازی مناسب تشخیص داده شد. نتایج حاکی از این واقعیت است که روش M در قسمت آموزش از دقت و سرعت مناسبی نسبت به روش آزمون و خطا در یافتن تعداد مناسب داده‌های ورودی، برخوردار می‌باشد. نتایج حاصل از مدل‌سازی نیز بیانگر آن بود که روش عصبی فازی توانایی و عملکرد بالایی در برآورد مقدار پتاسیم قابل جذب در خاک‌های منطقه میانکنگی را داشته است ( $RMSE=4.27$  و  $R^2=0.90$ ). همچنین، در این تحقیق، در راستای مدل‌سازی و پیش‌بینی پتاسیم قابل جذب، درصد کربن آلی مهم‌ترین ورودی شناخته شد.

کلمات کلیدی: آزمون M، آزمون گاما، پتاسیم قابل جذب، عصبی فازی، ویژگی‌های زودیافت

### مقدمه

2007). از آنجا که، بررسی تغییرات عناصر غذایی در خاک به‌ویژه در مناطقی با وسعت زیاد، پرهزینه و زمان‌بر است، از این‌رو نیاز به روش‌هایی احساس می‌شود که بتوانند به‌راحتی سطح گسترده‌ای از مناطق را پایش کرده و روند تغییرات را مورد بررسی قرار دهند. از جمله این روش‌ها، روش‌های مبتنی بر هوش محاسباتی‌اند که به دنبال کشف دانش نهفته در داده‌ها و استخراج روابط درونی بین آن‌ها و تعمیم آن به موقعیت‌های دیگر هستند.

اخیراً در زمینه مسائل خاکشناسی توابع انتقالی توسعه‌یافته با استفاده از روش‌های هوش محاسباتی توسط محققان بسیاری مورد استفاده قرار گرفته است و تقریباً نتیجه تمامی پژوهش‌های پیشین نشان می‌دهد که این مدل‌ها حداقل به اندازه بقیه روش‌های مورد استفاده، خوب عمل می‌کنند و به‌خوبی بر مفروضات آماری درگیر با توابع انتقالی فائق می‌آیند (Merdun *et al.*, 2006). پژوهشگران علوم کشاورزی از روش‌های مبتنی بر هوش مصنوعی در شبیه‌سازی پارامترهایی چون میزان تبخیر - تعرق هفتگی (Landeras *et al.*, 2009)، تبخیر روزانه

پتاسیم عنصری ضروری و پرمصرف برای همه موجودات زنده است و در فیزیولوژی و متابولیسم گیاه، نه تنها از نظر مقدار آن در بافت‌های گیاهی، بلکه از نظر وظایف فیزیولوژیکی و شیمیایی یکی از مهم‌ترین کاتیون‌ها است. غلظت پتاسیم در یاخته‌ها چنان زیاد است که به‌عنوان یکی از مهم‌ترین عوامل کنترل‌کننده پتانسیل اسمزی یاخته‌ها و در نتیجه فشار آماس شناخته می‌شود. غلظت زیاد پتاسیم در سیتوپلاسم و کلروپلاست، آنیون‌های آلی و معدنی محلول و غیر محلول را خنثی کرده و pH این بخش‌های سلول را بین ۷ تا ۸ ثابت حفظ می‌کند که برای اکثر واکنش‌های آنزیمی الزامی است (Bohra and Doerffling, 1993). حاصلخیزی خاک و فراهمی عناصر غذایی از جمله پتاسیم به‌منظور تأمین نیاز گیاهان از دیرباز مورد توجه کارشناسان علوم خاک بوده است. وجود تغییرات مکانی در ویژگی‌های خاک امری معمول است ولی شناخت این تغییرات به‌ویژه در اراضی کشاورزی جهت برنامه‌ریزی دقیق و مدیریت، امری ضروری است (Ayoubi *et al.*

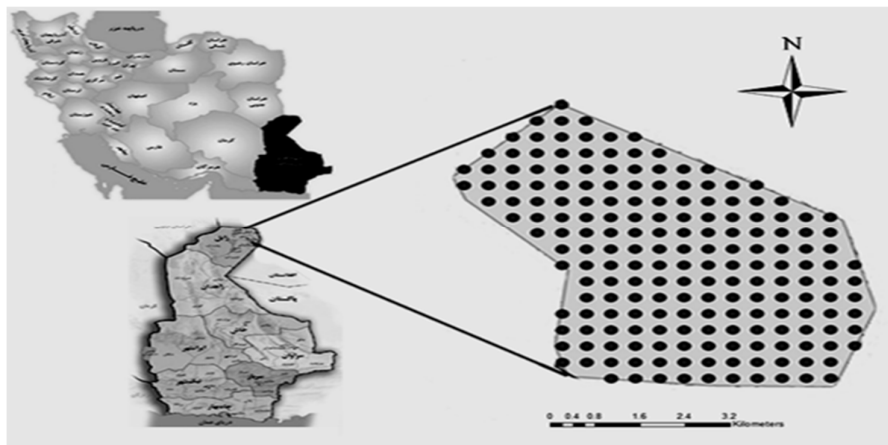
گردید. آزمون گاما اولین بار توسط کانکار (Koncar, 1997) و سپس توسط استفان و همکاران (Stefansson *et al.*, 1997) معرفی شد. ایوانز (Evans, 2001) و ایوانز و جونز (Evans and Jones, 2002) از این روش برای یافتن بهترین ترکیب متغیرهای ورودی به مدل استفاده کردند. مطالعات دیگری در موضوعات مختلف با استفاده از آزمون گاما، صورت گرفته است که مقدم‌نیا و همکاران (Moghaddamnia *et al.*, 2009) و پیری و همکاران (Piri *et al.*, 2009) از این جمله می‌باشند. سیستم استنتاج فازی بر مبنای تطبیق شبکه برای اولین بار توسط جانگ در سال ۱۹۹۳ ارائه گردید (Jang, 1993) و تاکنون کاربرد موفقیت آمیزی در زمینه حل مسائل مختلف داشته است (Chen and Chung, 2006; Shu and Ouarda, 2008). سیستم عصبی فازی از الگوریتم یادگیری شبکه عصبی جفت شده با استدلال فازی برای برآورد توزیع خروجی به کمک توزیع ورودی بهره می‌گیرد. این سیستم معمولاً با الگوریتم یادگیری هیبرید دگرآمورخته (نظارت شده) برای بهینه‌سازی پارامترهای خطی و غیرخطی آموزش داده می‌شود (Chung *et al.*, 2012). با این حال، این روش در مطالعات مرتبط با عناصر ضروری خاک کمتر مورد استفاده قرار گرفته است؛ بنابراین هدف از انجام این مطالعه یافتن ترکیب بهینه متغیرهای ورودی (شامل ویژگی‌های زودیافت خاک) در مدل‌سازی پتاسیم، تشخیص تعداد بهینه داده‌های مربوط به آموزش مدل به منظور پیش‌بینی پتاسیم قابل جذب در خاک-های منطقه میانکنگی می‌باشد.

### مواد و روش‌ها

#### منطقه مورد مطالعه

(Piri *et al.*, 2009)، پیش‌بینی درصد سدیم تبدلی (Mukerji *et al.*, 2016)، پیش‌بینی سیلاب (Hashemi *et al.*, 2009)، تخمین میزان فرسایش خاک (Kim and Gilley, 2008)، پیش‌بینی تغییرات مکانی فسفر خاک (Mir *et al.*, 2016) و مدل‌سازی کربن آلی خاک (Gholamalizadeh *et al.*, 2015) بهره برده‌اند. با این حال، با اینکه مطالعات زیادی برای مدل‌سازی و پیش‌بینی ویژگی‌های خاک با روش‌های مختلف انجام شده است. ولی، هنوز مسائل حل نشده فراوانی در این زمینه وجود دارد که انتخاب بهترین ترکیب متغیرهای ورودی مدل برای پیش‌بینی دقیق در علوم خاک یکی از آن‌ها می‌باشد.

در زمینه مدل‌سازی مسائل مربوط به خاک این سؤال مطرح است که اولاً کدام متغیرها به‌عنوان ورودی برای مدل-سازی مناسب می‌باشند و ثانیاً چند درصد از داده‌های مشاهداتی لازم است برای آموزش مدل در نظر گرفته شود. یکی از مراحل مهم و پیچیده برای مدل‌سازی غیرخطی، پیش‌پردازش داده‌های ورودی به‌منظور انتخاب ترکیبی مناسب از آن‌ها در مدل می‌باشد. این کار سبب کاهش مراحل سعی و خطا و شناخت مهم‌ترین متغیرهای مؤثر بر پدیده موردنظر در مدل‌سازی می‌گردد. برای این منظور تابه‌حال از روش‌های مختلف مانند روش تجزیه به مؤلفه‌های اصلی (Zhang *et al.*, 2006; Zhang, 2007; Noori *et al.*, 2009)، پروکراتس (Dinpashoh *et al.*, 2004) و آزمون گاما (Corcoran *et al.*, 2008; Moghaddamnia *et al.*, 2003) استفاده شده است. در مطالعه حاضر برای یافتن بهترین ترکیب از متغیرهای ورودی به‌منظور پیش‌بینی پتاسیم قابل جذب روش آزمون گاما انتخاب و مورد استفاده قرار گرفت. همچنین با استفاده از آزمون M تعداد داده‌های موردنیاز برای بخش آموزش مدل هم تعیین



شکل ۱- نقشه جغرافیایی منطقه نمونه‌برداری؛ میانکنگی، زابل، سیستان

Figure 1- Geographic map of the sampling area; Miankangi, Zabol, Sistan

تا با تجزیه و تحلیل منطقی این داده‌ها به‌عنوان نمونه، بتوان روابط غیرخطی یا نامشخص بین آن‌ها را استخراج و کار شبیه‌سازی را برای موارد احتمالی مشابه انجام داد. وارد کردن داده‌ها به‌صورت خام به شبکه باعث کاهش سرعت و دقت آن می‌شود. برای اجتناب از چنین شرایطی و همچنین یکسان نمودن ارزش داده‌ها، قبل از آموزش شبکه عصبی، داده‌های ورودی به آن بایستی استاندارد شوند (Kumar et al., 2002). همچنین، از آنجایی که داده‌های استفاده شده برای مدل‌سازی پتاسیم قابل جذب ابعاد متفاوتی دارند و مقادیر آن‌ها اختلاف زیادی با هم داشتند مبادرت به نرمال‌سازی داده‌ها شد. رابطه (۱) برای نرمال‌سازی استفاده گردید.

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

که  $x_n$  بیانگر مقدار نرمال شده،  $x$  نشان‌دهنده مقدار واقعی،  $x_{min}$  نشان‌دهنده مقدار حداقل و  $x_{max}$  معرف مقدار حداکثر پارامتر است.

در این مطالعه نرم‌افزار win-Gamma (version 1.98) برای یافتن بهینه ترکیب متغیرها و تعداد نقاط کافی برای بخش آموزش مدل به کار گرفته شد. همچنین برای ارزیابی عملکرد مدل‌ها از آماره‌های ضریب تبیین ( $R^2$ ) و RMSE استفاده شد.

هر چه مقدار RMSE به صفر نزدیک‌تر باشد، نمایانگر دقت بالاتر تخمین خواهند بود. همچنین هر چه مقدار  $R^2$  به یک نزدیک‌تر، دقت مدل بیشتر است (Rezace Pazhand, 2001).

### نتایج و بحث

جدول (۱) توصیفات آماری داده‌های مورد استفاده در این مطالعه را نشان می‌دهد. ایوبی و همکاران (Ayoubi et al., 2014) کمترین ضریب تغییرات ویژگی‌های خاک را به ترتیب در خاک‌های غرب و مرکز ایران برای pH گزارش کردند.

در تحقیق حاضر به‌منظور تعیین ترکیب بهینه متغیرهای ورودی در مدل عصبی فازی ابتدا با استفاده از آزمون گاما، مقدار گاما برای حالت استفاده از تمام متغیرهای ورودی به دست آمد. مدل‌های استفاده شده در این تحقیق در جدول (۲) آورده شده است. مدل ANFIS-1 مدلی است شامل همه‌ی متغیرهایی که به‌منظور مدل‌سازی محتوای پتاسیم قابل جذب استفاده شده است.

منطقه میانکنگی واقع در دشت سیستان در شرق ایران و در شمال استان سیستان و بلوچستان (شکل ۱) در طول جغرافیایی ۶۱ درجه و ۴۴ دقیقه تا ۶۱ درجه و ۴۹ دقیقه شرقی و عرض جغرافیایی ۳۱ درجه و ۴ دقیقه تا ۳۱ درجه و ۸ دقیقه شمالی واقع شده است. متوسط بارندگی سالیانه منطقه ۵۵ میلی‌متر، ارتفاع از سطح دریا ۴۷۸ متر و دمای آن از ۹/۵ تا ۴۹ درجه سانتی‌گراد متغیر است (Piri and Ansari, 2013). نمونه‌برداری خاک با فواصل ۵۰۰ متر از عمق ۰ تا ۳۰ سانتیمتر انجام و تعداد ۱۸۶ نمونه خاک برداشت شد. برای این منظور بر روی نقشه‌های توپوگرافی منطقه، شبکه‌هایی با فواصل تقریبی ۵۰۰ در ۵۰۰ متر پیاده و مختصات آن‌ها بر اساس سیستم مختصات متریک تعیین و پس از ورود مختصات نقاط به GPS<sup>1</sup>، برداشت نمونه‌ها انجام گردید. سپس نمونه‌های تهیه شده در مجاورت هوای آزاد، خشک و به نرمی کوبیده شده و از الک ۲ میلی‌متر (۱۰ mesh) عبور داده شدند. پس از آن، pH (Rhoades and Oster, 1986) و هدایت الکتریکی (Page, 1982) نمونه‌ها در نسبت ۱ به ۱ آب به خاک اندازه‌گیری شد. کربن آلی به روش تیتراسیون با آمونیوم فرسولفات نیم‌نرمال، پتاسیم قابل جذب خاک با استفاده از استات آمونیوم یک نرمال (pH=7) (Jafari Haghghi, 2003)، آهک به روش تیتراسیون اسید اضافی با سود (Nelson, 1982) و بافت نمونه‌های خاک به روش هیدرومتری تعیین گردید (Bouyoucos, 1962).

### آزمون گاما

آزمون گاما ابزاری مناسب برای یافتن بهترین ترکیب از متغیرهای ورودی به مدل غیرخطی می‌باشد. در واقع آزمون گاما تخمینی از حداقل میانگین مربعات خطای آن بخش از داده‌ها است که با مدل خطی نمی‌توان آن‌ها را تعیین کرد. در شرایطی که تعداد متغیرهای ورودی زیاد باشند با استفاده از آزمون گاما می‌توان بهترین ترکیب از میان تمام حالات ممکن را تشخیص داد. همچنین با استفاده از آزمون M تعداد بهینه داده‌های ورودی در بخش آموزش مدل‌سازی تشخیص داده می‌شود. آزمون M در واقع رسم نمودار تغییرات مقدار آماره G و یا خطای استاندارد در مقابل تعداد داده‌ها می‌باشد. در نقطه‌ای که این دو نمودار تقریباً به حالت افقی و پایدار می‌رسند آن نقطه تعیین‌کننده تعداد نقاط کافی جهت استفاده در بخش آموزش مدل‌سازی می‌باشد (Jones, 2004).

### مدل‌سازی

سیستم‌های عصبی - فازی (ANFIS) جهت طراحی و آموزش، نیازمند یک سری داده‌های ورودی و خروجی می‌باشند

<sup>1</sup>Global positioning system

جدول ۱- توصیف آماری خصوصیات فیزیکی و شیمیایی نمونه‌های خاک

Table 1- Statistical description of physical and chemical properties of soil samples

ویژگی‌های خاک (Soil Characteristics)	واحدها (Units)	حداقل (Min)	حداکثر (Max)	میانگین (Mean)	انحراف معیار (S.D.)	ضریب تغییرات (C.V.)
کربنات کلسیم معادل (CCE)	(%)	7	23	18.66	2.52	13.5
هدایت الکتریکی (EC)	(dS.m <sup>-1</sup> )	0.13	21.33	5.68	5.85	102.99
اسیدیته خاک (pH)	-	6.9	9	8.13	0.45	5.53
رس (Clay)	(%)	1	53	12.44	8.98	72.18
شن (Sand)	(%)	0	96.6	45.37	18.28	40.29
سیلت (Silt)	(%)	2	78	42.17	16.03	38.01
رطوبت اشباع (SP)	(%)	15.3	80	46.67	8.50	18.21
کربن آلی (OC)	(%)	0.03	0.70	0.26	0.14	53.84
پتاسیم (K)	(ppm)	2	78	14.85	14.37	96.8

جدول ۲- ترکیب ورودی‌ها در مدل‌های استفاده شده برای برآورد پتاسیم قابل جذب

Table 2- Combine input in the models used to estimate available potassium

اسم مدل (Model Name)	ورودی‌های انتخاب شده (Selected inputs)	متغیر انتخاب نشده (Not selected variable)
ANFIS-1	CCE, SP, pH, Sand, Silt, EC, OC, Clay	-
ANFIS-2	SP, pH, Sand, Silt, EC, OC, Clay	CCE
ANFIS-3	CCE, pH, Sand, Silt, EC, OC, Clay	SP
ANFIS-4	CCE, SP, Sand, Silt, EC, OC, Clay	pH
ANFIS-5	CCE, SP, pH, Silt, EC, OC, Clay	Sand
ANFIS-6	CCE, SP, pH, Sand, EC, OC, Clay	Silt
ANFIS-7	CCE, SP, pH, Sand, Silt, OC, Clay	EC
ANFIS-8	CCE, SP, pH, Sand, Silt, EC, Clay	OC
ANFIS-9	CCE, SP, pH, Sand, Silt, EC, OC	Clay

به‌دست‌آمده در جدول (۳) آورده شده است. جدول (۳) نتایج آزمون گاما را برای مدل‌سازی پتاسیم قابل جذب با استفاده از ویژگی‌های زودیافت (CCE, SP, pH, Sand, Silt, EC, OC, Clay) در منطقه میانکنگی نشان می‌دهد.

سپس، در ادامه یکی از متغیرهای ورودی حذف و مقدار گاما برای ترکیب مفروض محاسبه شد. سپس متغیر مذکور دوباره به مدل وارد و متغیر دیگری حذف شد و مقدار گاما برای ترکیب جدید به دست آمد. این روال برای همه متغیرها تکرار گردید و نتایج

جدول ۳- نتایج آزمون گاما به منظور تعیین ترکیب مناسب ورودی‌ها در مدل‌سازی پتاسیم قابل جذب

Table 3- Gamma test results to determine the right combination of inputs in the modeling of available potassium

مدل (Model)	متغیر انتخاب نشده (Not selected variable)	گاما (Gamma) (Γ)	شیب (Gradient) (A)	خطای استاندارد (Standard Error)	نسبت V (V Ratio)
ANFIS-1	-	0.013844	0.053414	0.0024098	0.35356
ANFIS-2	CCE	0.010719	0.079656	0.0027897	0.27376
ANFIS-3	SP	0.014069	0.057966	0.0015931	0.35931
ANFIS-4	pH	0.017661	0.061487	0.0032562	0.45107
ANFIS-5	Sand	0.013833	0.062509	0.002995	0.35329
ANFIS-6	Silt	0.01613	0.053692	0.0031149	0.41194
ANFIS-7	EC	0.016911	0.040931	0.0030753	0.4319
ANFIS-8	OC	0.025619	0.024583	0.004242	0.65429
ANFIS-9	Clay	0.01638	0.05416	0.0019652	0.41835

همسایگی انتخاب شد. در این مطالعه تعداد نزدیک‌ترین همسایگی برابر ۲۴ در نظر گرفته شد. بعد از اینکه مقدار گاما برای تمامی ترکیب‌های جدول (۲) به دست آمد مقدار آماره گامای هر یک از مدل‌های ANFIS-2 تا ANFIS-9 با مقدار آماره گامای مربوط به مدل ANFIS-1 مقایسه شد. مدلهایی که در آن‌ها مقدار گامای مدل با ترکیبات

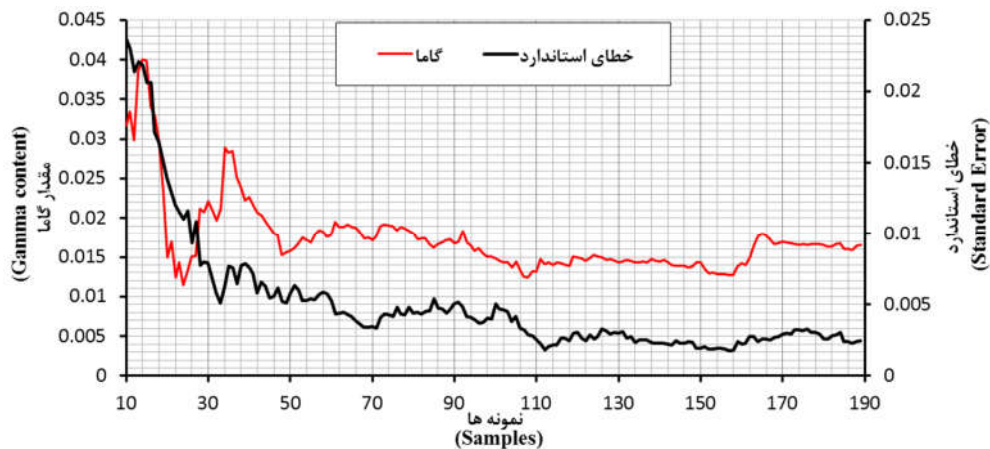
برای به‌دست‌آوردن مقدار بهینه حداکثر مقدار نزدیک‌ترین همسایگی، ابتدا آماره گاما برای مقادیر مختلف نزدیک‌ترین همسایگی محاسبه شد و نمودار مقدار آماره گاما تابعی از تعداد نزدیک‌ترین همسایگی رسم شد. کمترین مقدار گاما، در نمودار مذکور به‌عنوان حداکثر مقدار نزدیک‌ترین

مفروض بیشتر از مقدار گامای نظیر مدل ANFIS-1 بود، در نظر گرفته شدند. متغیرهای غایب در مدل‌های مذکور به‌عنوان متغیر مهم در تخمین پتاسیم قابل جذب تشخیص داده شدند و از آن‌ها برای مدل‌سازی پتاسیم قابل جذب در منطقه میانکنگی استفاده گردید. شریفی و همکاران (Sharifi *et al.*, 2013) نیز این روش را برای مدل‌سازی به کار برده‌اند.

با توجه به جدول (۳) می‌توان استنباط کرد، ۶ مدل وجود دارد که مقدار گامای آن‌ها بیش از گامای نظیر مدل ANFIS-1 می‌باشد؛ بنابراین متغیرهای غایب این مدل‌ها شامل ویژگی‌های SP, pH, Silt, EC, OC, Clay به‌عنوان ورودی‌های مؤثر در مدل‌سازی پتاسیم قابل جذب تشخیص داده شدند. همچنین، از آنجایی‌که کربن آلی (OC) از ترکیب ورودی‌های مدل حذف می‌شد مقدار گامای مدل ANFIS-8 نسبت به مدل ANFIS-1 (یعنی زمانی که همه متغیرها در ترکیب وجود دارند) بیشترین مقدار را به خود اختصاص می‌داد به‌عنوان مهم‌ترین متغیر در این مدل شناخته شد. هاشمی و همکاران (Hashemi *et al.*, 2016) بیان داشتند، در راستای تخمین درصد سدیم تبدیلی با استفاده از کمترین تعداد ورودی (هدایت الکتریکی و اسیدیته خاک)، سیستم عصبی فازی با کمترین میزان خطا ( $RMSE=2/34$ ) و بیشترین مقدار ضریب تبیین ( $R^2=0/81$ ) عملکرد بهتری داشت. تاکنون در بیشتر مطالعات خاکشناسی تعداد داده‌های مورد استفاده در بخش آموزش مدل در زمینه‌های مختلف به طور معمول ۷۰ یا ۸۰ درصد کل داده‌ها انتخاب می‌شد (Gholamalizadeh *et al.*, 2015; Mir *et al.*, 2016; )

(Hashemi *et al.*, 2016). باین‌حال ما در تحقیق حاضر برای تشخیص اینکه چه تعداد داده برای مدل‌سازی در بخش آموزش مدل کافی است از آزمون M استفاده نمودیم. شکل (۲) نتایج آزمون M را با استفاده از متغیرهای استفاده شده در مدل منتخب، نشان می‌دهد. در شکل (۲) نمودار آماره گاما (خط قرمز رنگ) و خطای استاندارد (خط مشکی) هر کدام از نقطه‌ای به بعد تقریباً هموار می‌شوند. با استفاده از این دو نمودار تعداد ۱۱۲ داده برای بخش آموزش مدل در نظر گرفته شد. این تعداد داده کمتر از ۶۰ درصد کل داده‌ها (۱۸۹ نمونه) می‌باشد و استفاده از این تعداد پیچیدگی مدل را به طور قابل ملاحظه‌ای نسبت به زمانی که از ۷۰ درصد داده‌ها برای آموزش مدل استفاده می‌شود (مرسوم در روش‌های هوشمند)، کاهش می‌دهد. در زمینه‌ها و مطالعات دیگری از روش آزمون M استفاده شده است از جمله؛ رمضان و همکاران (Remesan *et al.*, 2009)، در مطالعات خود تعداد ۱۰۵۶ نقطه از ۲۲۳۶ نقطه را (۴۷/۳ درصد) با استفاده از آزمون M جهت مدل‌سازی رواناب مناسب تشخیص دادند.

از معایب استفاده از مدل‌های هوش محاسباتی (از جمله سیستم عصبی - فازی) تعیین ترکیب مناسب ورودی‌ها با استفاده از سعی و خطا می‌باشد که نیاز به صرف وقت زیادی دارد. در این مطالعه با استفاده از روش آزمون گاما قبل از بسط هرگونه مدل خاصی، ترکیب بهینه متغیرهای ورودی انتخاب شد. استفاده از این روش منجر به کاهش مدت‌زمان لازم برای مدل‌سازی می‌شود. یکی دیگر از معایب مدل‌های هوش محاسباتی تعیین بخشی از داده‌ها جهت استفاده در بخش مدل‌سازی است.



شکل ۲- نمودار مقایسه تغییرات مقدار گاما در مقابل خطای استاندارد (آزمون M)  
Figure 2- Comparison gamma value changes chart versus standard error (M test)

آموزش داده‌شده شامل (درصد رس، سیلت، ماده آلی، هدایت الکتریکی، رطوبت اشباع و pH) و در لایه خروجی پتاسیم قابل جذب به‌عنوان هدف مدل‌سازی قرار گرفت. باتوجه‌به مدل توسعه‌یافته عصبی - فازی (خوشه‌بندی کاهشی) برای پیش‌بینی

### نتایج مدل‌سازی با روش خوشه‌بندی (Subtractive)

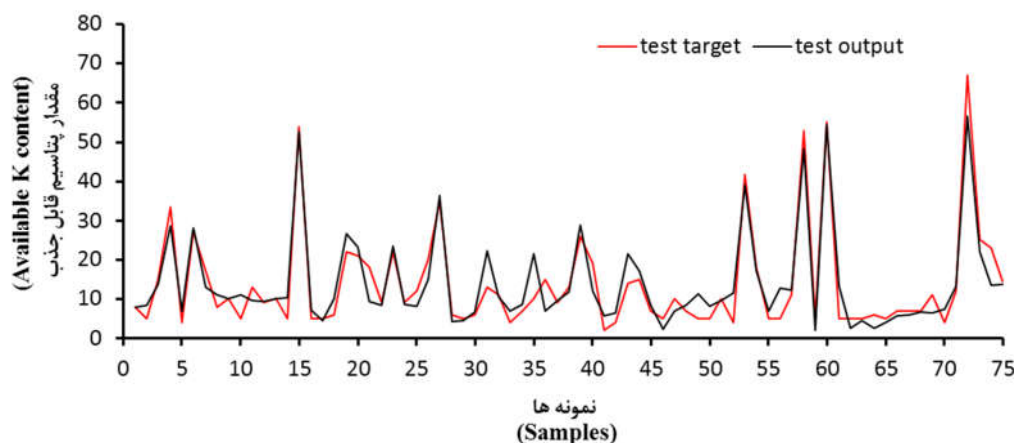
نتایج بهترین ساختار سیستم عصبی فازی که برای پیش‌بینی پتاسیم قابل جذب در منطقه مورد مطالعه مورد استفاده قرار گرفت، در جدول (۴) و شکل (۳) آمده است. ورودی‌ها در مدل

قابل جذب در منطقه مورد مطالعه توانستند بیش از ۹۰ درصد از تغییرپذیری فسفر قابل جذب در منطقه دشت سیستان را توجیه کنند.

پتاسیم قابل جذب در منطقه مورد مطالعه توانست بیش از ۹۰ درصد از تغییرپذیری پتاسیم قابل جذب در منطقه را توجیه کند (جدول ۴). میر و همکاران (Mir et al., 2016) با استفاده از مدل توسعه یافته شبکه عصبی مصنوعی برای پیش‌بینی فسفر

جدول ۴- نتایج بدست آمده از مدل‌سازی پتاسیم قابل جذب با استفاده از روش خوشه‌بندی کاهشی نروفازی  
Table 4- Results of available potassium modeling by using subtractive clustering neuro-fuzzy

آزمون (Test)		آموزش (Training)		مدل
R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE	
0.90	4.27	0.95	3.58	انفیس (خوشه‌بندی کاهشی) ANFIS (Subtractive clustering)



شکل ۳- نمودار مقادیر برآورد شده و مقادیر اندازه‌گیری شده‌ی پتاسیم قابل جذب در منطقه میانکنگی  
Figure 3- Graph of the estimated and measured values of K in Miankangi region

بالایی در برآورد مقدار پتاسیم قابل جذب در خاک‌های منطقه میانکنگی را داشته است. همچنین، قابل توجه است که در این تحقیق، در راستای مدل‌سازی و پیش‌بینی پتاسیم قابل جذب، درصد کربن آلی مهم‌ترین ورودی شناخته شد. در نهایت، پیشنهاد می‌شود روش آزمون گاما برای مدل‌سازی مسائل دیگر مربوط به حوزه علوم خاک نیز بکار رود. انتخاب ترکیب بهینه متغیرهای ورودی به مدل با روش آزمون گاما انجام و نتایج با دیگر روش‌های شناخته شده مرسوم مثل رگرسیون خطی چندگانه (با گزینه‌ی گام‌به‌گام) و یا روش شبکه‌های عصبی مصنوعی مقایسه گردد.

### نتیجه‌گیری

با استفاده از آزمون گاما دیگر نیازی به مدل‌سازی با استفاده از سعی و خطا در ترکیب ورودی برای یافتن بهترین مدل نیست. با استفاده از گاما سعی شد تا بهترین ترکیب برای مدل‌سازی انتخاب شود. هدف بعدی یافتن تعداد داده کافی برای بخش آموزش مدل-سازی بود که در این مطالعه با استفاده از آزمون M تعداد ۱۱۲ داده برای آموزش مدل مناسب تشخیص داده شد. نتایج حاصل از مدل-سازی نیز بیانگر آن بود که روش عصبی فازی توانایی و عملکرد

### References

- Ayoubi, Sh., Mohammad Zamani, S. and Khormali, F. 2007. Prediction total N by organic matter content using some geostatistic approaches in part of farm land of Sorhankalateh, Golestan Province. *Journal of Agricultural Science and Natural Resources*, 14(4): 215-225. (In Persian)
- Ayoubi, S., Mehnatkesh, A., Jalalian, A., Sahrawat, K.L. and Gheysari, M. 2014. Relationships between grain protein, Zn, Cu, Fe and Mn contents in wheat and soil and topographic attributes. *Archives of Agronomy and Soil Science*, 60(5): 625-638.
- Bohra, J.S. and Doerffling, K. 1993. Potassium nutrition of rice (*Oryza Sativa* L.) Varieties under NaCl salinity. *Plant and Soil*, 152: 299-303.

- Bouyoucos, G.J.** 1962. Hydrometer method improved for making particle size analyses of soils. *Agronomy Journal*, 54(5): 464-465.
- Corcoran, J., Wilson, I. and Ware, J.** 2003. Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, 19: 623-634.
- Chen, S.M. and Chung, N.Y.** 2006. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of Intelligent Systems*, 21: 485-501.
- Chung, C.H., Chiang, Y.M. and Chang, F.J.** 2012. A spatial neural fuzzy network for estimating pan evaporation at ungauged sites. *Hydrology and Earth Systems Sciences*, 16: 255-266.
- Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. and Mirnia, M.** 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology*, 297:109-123.
- Evans, D.** 2001. Data derived estimations of noise using near neighbor distance distributions. Ph.D. Thesis, Cardiff University, Wales, U.K.
- Evans, D. and Jones, A.J.** 2002. A proof of the gamma test. *Proceedings of Royal Society, Series A*, 458(20-27): 2759-2799.
- Gholamalizadeh Ahangar, A., Sarani, F., Hashemi, M. and Shabani, A.** 2015. Comparison of linear regression methods, geostatistical and artificial neural network modeling of organic carbon in dry land of Sistan plain. *Journal of Water and Soil*, 28(6):1250-1260. (In Persian)
- Hashemi, M., Gholamalizadeh Ahangar, A. and Shabani, A.** 2016. Evaluating pedotransfer functions for estimating ESP in the soils of Sistan plain. *Journal of Agricultural Engineering*, 38(2): 77-93. (In Persian)
- Jafari Haghighi, M.** 2003. Methods of Soil Analysis. First Edition. Publication of nedaye zoha. 129 p. (In Persian)
- Jang, J.S.R.** 1993. ANFIS Adaptive-Network-Based Fuzzy inference system. *IEEE Transactions on Systems Man and Cybernetics*, 23: 665-658.
- Jones, A.J.** 2004. New tools in non-linear modeling and prediction. *Computational Management Science*, 1: 109-149.
- Kim, M. and Gilley, J.E.** 2008. Artificial neural network estimation of soil erosion and nutrient concentrations in runoff from land application areas. *Computers and Electronics in Agriculture*, 64(2): 268-275.
- Koncar, N.** 1997. Optimization methodologies for direct inverse neuro control. Ph.D. Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London, London.
- Kumar, M., Raghuvanshi, N.S., Singh, R., Wallender, W.W. and Pruitt, W.O.** 2002. Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering-ASCE*, 128: 224-233.
- Landeras, G., Ortiz-Barredo, A. and López, J.J.** 2009. Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. *Journal of Irrigation and Drainage Engineering*, 135(3): 323-334.
- Merdun, H., Çınar, Ö., Meral, R. and Apan, M.** 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil and Tillage Research*, 90(1): 108-116.
- Mir, H., Gholamalizadeh Ahangar, A. and Shabani, A.** 2016. Determination of the most important soil parameters affecting the availability of phosphorus in Sistan plain, using connection weight method in neural networks. *Journal of Water and Soil*, 29(6):1674-1687. (In Persian)
- Moghaddamia, A., Ghafari Gousheh, M., Piri, J. and Han, D.** 2008. Evaporation estimation using support vector machines technique. *World Academy of Science, Engineering and Technology*, 43: 14-22.
- Moghaddamia, A., Remesan, R., Hassanpour Kashani, M., Mohammadi, M., Han, D. and Piri, J.** 2009. Comparison of LLR, MLP, Elman, NNARX and ANFIS Models- with a case study in solar radiation estimation. *Journal of Atmosphere and Solar-Terrestrial Physics*, 71: 975-982.
- Mukerji, A., Chatterjee, C. and Raghuvanshi, N.S.** 2009. Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. *Journal of Hydrologic Engineering*, 14(6): 647-652.
- Nelson, R.E.** 1982. Carbonate and gypsum. In: Page, A. L. (eds). *Methods of Soil Analysis-Part II*. Soil Science Society of America and American Society of Agronomy, Madison, Wisconsin, USA. PP: 181-197.
- Noori, R., Karbassi, A. and Sabahi, M.S.** 2009. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*, 91: 767-771.
- Page, A.L.** 1982. *Methods of Soil Analysis- Part 2: Chemical and Microbiological Properties*. Soil Science Society of America and American Society of Agronomy, Madison, Wisconsin, USA. 1572 p.
- Piri, J., Amin, S., Moghaddamia, A., Keshavarz, A., Han, D. and Remesan, R.** 2009. Daily pan evaporation modeling in a hot and dry climate. *Journal of Hydrologic Engineering*, 14(8): 803-811.
- Piri, H. and Ansari, H.** 2013. Study of drought in Sistan Plain and its impact on Hamoun international. *Journal of Wetland Ecobiology- Islamic Azad University, Ahvaz Branch*, 5(1): 63-74. (In Persian)
- Remesan, R., Shamim, M.A., Han, D. and Mathew, J.** 2009. Runoff prediction using an integrated hybrid modeling scheme. *Journal of Hydrology*, 372: 48-60.

- Rezaee Pazhand, H.** 2001. Application of Probability and Statistics in Water Resources. 1st Edition. Sokhan Gostar, Mashhad. 468 p. (in Persian)
- Rhoades, J.D. and Oster, J.D.** 1986. Solute Content. Methods of Soil Analysis-Part 1. Physical and Mineralogical Methods. PP: 985-1006.
- Sharifi, AR., Dinpashoh, Y., Fakheri-Fard, A. and Moghaddamnia, AR.** 2013. Optimum combination of variables for runoff simulation in Amameh watershed using Gamma test. *Journal of Soil and Water*, 23(4):59-72. (In Persian)
- Shu, C. and Ouarda, T.B.M.J.** 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology*, 349: 31-43
- Stefansson, A., Koncar, N. and Jones, A.J.** 1997. A note on the gamma test. *Neural Computing and Applications*, 5(3): 131-133.
- Zhang, Y.X.** 2007. Artificial neural networks based principal component analysis input selection for clinical pattern recognition analysis. *Talanta*, 73: 68-75.
- Zhang, Y.X., Li, H., Hou, A. and Haval, J.** 2006. Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemo Metrics and Intelligent Laboratory Systems*, 82: 165-175.



## **Determining optimal combination of input variables using gamma test for absorbable potassium modeling in the fuzzy-neural system (case study: Mian-Kangi region Zabol)**

**Amin Delarami<sup>1</sup>, Ahmad Gholamalizadeh<sup>\*2</sup>, Asma Shabani<sup>2</sup>**

<sup>1</sup>MSc Student, Department of soil sciences, Faculty of soil and water, University of Zabol, Zabol, Iran

<sup>2</sup>Department of soil sciences, Faculty of soil and water, University of Zabol, Zabol, Iran

\*Corresponding Author: [Ahangar@uoz.ac.ir](mailto:Ahangar@uoz.ac.ir)

Received: 22 October 2017

Accepted: 9 December 2018

DOI: 10.22034/csrar.2020.119092

### **Abstract**

One of the important and complex steps for nonlinear modeling is pre-processing of input data in order to select the appropriate combination of them in the model. The gamma test was used to select the optimal combination of input variables for available potassium modeling in this study. The M test was used for determining the optimal number of data needed for modeling. Eight input variables were used for modeling. Modeling the available potassium was done by the number of optimum points and selected variables with subtractive clustering in the fuzzy neural system. The results showed that six variables (clay percentage, silt, organic matter, electrical conductivity, saturation moisture and pH) are the optimal combination of variables in modeling the available potassium in Mian-Kangi region. Also, 112 of measured data (60%) were considered as suitable data for the modeling training section using the M test results. The results indicated that the M method has better accuracy and speed than the trial and error method for finding the appropriate number of input data in training section. The results of modeling also indicated that the fuzzy neural method has high capability and performance in estimating the amount of available potassium in the soil of Mian-Kangi region ( $R^2 = 0.90$  and  $RMSE = 4.27$ ). Also, organic carbon percentage was the most important input for modeling and predicting the amount of available potassium.

**Key words:** Available potassium, Fast measured properties, Fuzzy neural, Gamma test, M test

