مقایسه‌ی شبکه‌ی عصبی فازی و رگرسیون چند متغیره در پیش‌بینی فسفر خاک با استفاده از خصوصیات خاک

احمد بازوبندی [۱*]، صمد امامقلی زاده[۲] ، هادی قربانی [۲]

۱- دانشجویی دکتری دانشکده‌ی کشاورزی، دانشگاه فردوسی مشهد، ۲- دانشیاران دانشکده‌ی کشاورزی دانشگاه شاهرود

چکیده

در سال‌های اخیر استفاده از روش‌های غیرمستقیم مانند شبکه عصبی فازی برای برآورد پارامترهای خاک موردتوجه قرارگرفته است. در این پژوهش به‌منظور تخمین فسفر خاک با استفاده از شبکه عصبی فازی اقدام به نمونه‌برداری به تعداد ۲۵۰ نمونه سطحی از استان گیلان گردید. پارامترهای فیزیکی و شیمیایی خاک شامل فراوانی نسبی ذرات، PH، کربن آلی، مقادیر روی، کادمیوم به‌عنوان ورودی و میزان فسفر کل خاک به‌عنوان پارامتر خروجی در نظر گرفته شد. داده‌ها جمع‌آوری‌شده به‌صورت تصادفی به دودسته، آموزش (۸۰ درصد) و صحت سنجی (۲۰ درصد) تقسیم شدند و از آن‌ها برای آموزش و ارزیابی مدل‌های عصبی فازی و همچنین رگرسیون خطی و غیرخطی استفاده گردید. دقت پیش‌بینی به‌وسیله آماره‌های (MAE،RMSE و R2) بین فسفر اندازه‌گیری و پیش‌بینی‌شده، ارزیابی شد. نتایج کارایی بالاتر مدل شبکه عصبی فازی را با مقادیر MAE(۲/۰۶)،RMSE(۲/۷۲) و R^2(۰/۷۵)نشان داد.

واژه های کلیدی: مدل عصبی فازی، گیلان، رگرسیون چند متغیره، پیش بینی فسفر

مقدمه

فسفر یکی از سه عنصر اصلی و ضروری برای رشد گیاهان است که در بین عناصر غذایی بعد از نیتروژن مهم‌ترین عامل محدودکننده تولیدات کشاورزی در بیش‌تر مناطق جهان و ایران می‌باشد. روش‌های متعددی برای تعیین فسفر قابل‌استفاده گیاه در خاک پیشنهادشده است. اولسن و سومرس (۱۹۸۲) و تورکر و هاگین (۱۹۸۲) ولی هیچ‌یک از آن‌ها تمام عوامل مؤثر بر فسفر خاک را به‌حساب نیاورده‌اند. مناسب‌ترین روش تعیین فسفر قابل‌استفاده در خاک‌های آهکی روش اولسن می‌باشد. در بسیاری از کشورها همبستگی خوبی بین عملکرد و مقدار فسفری که با کاربرد روش السن اندازه‌گیری شده به‌دست‌آمده است. امروزه مدل شبکه عصبی مصنوعی و عصبی فازی(ANFIS) در بخش‌های مختلف علوم به‌منظور مدل‌سازی روابط پیچیده غیرخطی به کار گرفته می‌شود و تا حدودی جایگزین مدل‌های آماری شده است زیرا شبکه‌های عصبی مصنوعی و عصبی فازی(ANFIS) بدون نیاز به حل معادلات دیفرانسیل جزئی، غیرخطی بودن فرایند موردنظر را شبیه‌سازی می‌نمایند و حتی زمانی که مجموع داده‌های آموزشی حاوی داده‌های خطادار باشد، عملکرد مناسبی را نشان می‌دهند و برخلاف مدل‌های بر اساس چند رگرسیونی (مدل‌های آماری) نیاز به انتخاب اولیه شکل تابع ریاضی مرتبط کننده ورودی و خروجی سیستم وجود ندارد (دامنگیر،۲۰۰۱). تعیین مقدار فسفر به روش‌های آزمایشی در یک منطقه معمولاً بسیار پرهزینه و زمان‌بر است؛ اما روش‌های مدل‌سازی مانند شبکه‌های عصبی مصنوعی و عصبی مصنوعی برای جایگزین این روش معرفی شده‌اند (اینگورسن و اشترک۲۰۰۶ و التفندر۲۰۰۷). کاربرد مدل شبکه‌ی عصبی مصنوعی برای پیش‌بینی موارد مختلف در سال‌های اخیر توسط محققان زیادی گزارش شده است(قربانی و همکاران ۲۰۱۵). بررسی منابع نشان می‌دهد تابه‌حال تحقیقی در زمینه بررسی و پیش‌بینی میزان فسفر در مناطق شمال ایران انجام نشده است. بنابراین در تحقیق حاضر با استفاده از خصوصیات فیزیکی و شیمیایی زود یافت خاک به پیش‌بینی آن با مقایسه ی روش های شبکه ی عصبی فازی(ANFIS) و رگرسیون چند متغیره خطی و غیر خطی انجام شد.

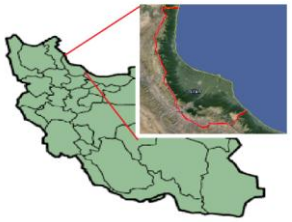
مواد و روش‌ها

منطقه‌ی موردمطالعه

استان گیلان با مساحت ۱۴۷۱۱ کیلومترمربع در میان رشته‌کوه‌های البرز و تالش در شمال ایران جای گرفته است.

تجزیه‌ی فیزیکی و شیمیایی

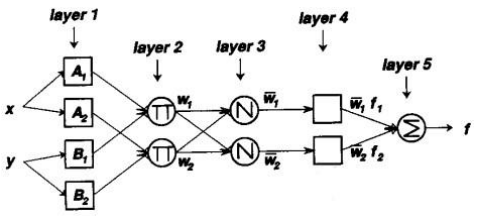
در این آزمایش نمونه‌برداری از مناطق مختلف استان گیلان از عمق (۰-۳۰ سانتیمتری) سطح خاک انتخاب شد پس از آماده‌سازی اولیه بافت خاک به روش هیدرومتری، فسفر خاک به روش اولسن و همکاران (۱۹۸۲) عناصر میکرو با اتمیک ابزوربشن (با عصاره گیر (DTPA، pH در خمیر اشباع و درصد کربن و ماده آلی به روش اکسایش با دی کرومات اندازه‌گیری شدند (همه در آزمایشگاه بخش خاکشناسی مرکز تحقیقات استان گیلان)



شکل ۱- موقعیت منطقه‌ی موردمطالعه

مدل فازی عصبی (ANFIS)

این ساختار حاصل تلفیق شبکه تطبیقی با قوانین فازی است. در این ساختار هیچ محدودیتی بر روی توابع گرهای وجود ندارد به‌جز اینکه باید به‌طور قطعه‌های مشتق‌پذیر باشند. قوانین فازی به‌صورت تاکاگی-سوگینو در نظر گرفته می‌شوند و اگر x و y دو متغیر بخش مقدمه قانون فازی i ام باشند، مؤخره این قانون فازی به‌صورت ترکیب خطی x و y به همراه یک مقدار ثابت خواهد بود (جی شینگ و همکاران ۱۹۹۳) ساختار شبکه ANFIS با پنج لایه، در شکل ۲ نشان داده‌شده است. لایه اول دارای گره‌های قابل تنظیم است که معمولاً توابع تعلق مربوط به آن‌ها به‌صورت گاوسی یا زنگوله شکل، با مقدار بیشینه‌ای برابر یک و کمیت‌های برابر صفر می‌باشند. پارامترهای توابع تعلق که همان پارامترهای مقدمه قوانین فازی هستند، بر اساس توصیف زبانی مربوط به متغیرها و زیر فضاهای فازی و بر اساس روش هایبرید تنظیم می‌شوند. گره‌های لایه دوم ثابت در نظر گرفته می‌شوند. این گره‌ها دو سیگنال ورودی را در هم ضرب کرده و حاصل را به‌عنوان خروجی تحویل شبکه می‌دهند. سیگنال‌های ورودی این گره‌ها درواقع میزان سازگاری ورودی با هریک از توابع تعلق و خروجی آن‌ها، وزن مربوط به هریک از قوانین است. گره‌های لایه سوم نیز ثابت بوده و وظیفه آن‌ها محاسبه وزن نرمالیزه هریک از قوانین است. گره‌های لایه چهارم وزن نرمالیزه هر یک از قوانین فازی را در خروجی بخش مؤخره آن قانون ضرب کرده و تنها گره لایه پنجم نیز کلیه سیگنال‌های خروجی از گره‌های لایه چهارم را جمع کرده و به‌عنوان خروجی شبکه تحویل می‌دهد (جی شینگ و همکاران ۱۹۹۳). الگوریتم هایبرید به‌صورت مستقیم برای شناسایی و تخمین پارامترهای شبکه قابل‌اعمال است. پارامترهای مؤخره قوانین همان پارامترهای خطی شبکه هستند که به روش LSE تخمین زده می‌شوند. پارامترهای مقدمه نیز به روش کاهش گرادیان تنظیم می‌شود (جی شینگ و همکاران ۱۹۹۳).



شکل ۲: ساختار شبکه ANFIS

رگرسیون چند متغیره

مدل رگرسیونی به شکل ماتریسی را می‌توان به‌صورت معادله زیر نشان داد:

$$Y\ X = \beta + e \quad (۱)$$

در فرمول (۱)،β ماتریس ضرایب رگرسیون، e ماتریس خطای برازش و Y نیز ماتریس پاسخ می‌باشد. با حل معادله بالا برحسب β خواهیم داشت:

$$\beta = (X')-1(X\ Y') \quad (۲)$$

که در رابطه (۲) X′ ترانهاده ماتریس X است. برای محاسبه معکوس (X′)، لازم است متغیرهای مستقل همبستگی زیادی نداشته باشند، زیرا در این صورت ماتریس (X′) را نمی‌توان معکوس کرد و باعث افزایش خطا در اثر گرد کردن داده‌ها و محاسبات می‌شود. برای رفع این مشکل باید قبل از ساخت مدل رگرسیونی، همبستگی بین متغیرهای مستقل را از بین برد. در این خصوص روش مناسب، استفاده از تحلیل مؤلفه‌های اصلی روی متغیرهای مستقل ورودی به مدل است. معیار قضاوت برای رفع این مشکل با اجرای تحلیل مؤلفه‌های اصلی روی متغیرهای ورودی، فاکتور تورم واریانس است. عدد ایدئال برای فاکتور تورم واریانس یک است و مقادیر بزرگ‌تر از ۱۰ برای تورم واریانس نشانه ناپایداری مدل رگرسیونی است (هاکینگ، ۲۰۰۳). در این تحقیق پس از رفع مشکل همبستگی در متغیرهای مستقل، مدلی مناسب با استفاده از فن رگرسیون خطی چندمتغیره برای پیش‌بینی فسفر کل خاک توسعه‌یافته و در محاسبات رگرسیونی از الگوریتم گام‌به‌گام (stepwise) استفاده‌شده است. در این روش ورود متغیرها به مدل رگرسیون به‌صورت مرحله‌ای، از مهم‌ترین متغیر تا کم‌اهمیت‌ترین آن‌ها، صورت می‌گیرد. معیار میزان اهمیت متغیر در مدل، مقدار سطح معنی‌داری یا آماره F متناظر با آن در جدول آزمون معنی‌داری متغیرها است.

ارزیابی مدل‌ها

به‌منظور مقایسه عملکرد الگوریتم‌های شبکه ی عصبی فازی و رگرسیون چند متغیره از پارامترهای آماری مانند ضریب تبیین (R2) (رابطه ۳)، میانگین خطای مطلق (MAE) (رابطه ۴)، مجذور میانگین مربعات خطا (RMSE) (رابطه ۵) که از روابط زیر قابل‌محاسبه می‌باشند، استفاده گردید.

$$R^2 = \frac{\left[\sum_{i=1}^{n}(\hat{y}_i - \hat{y})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(\hat{y}_i - \hat{y})^2 \sum_{i=1}^{n}{}_i(y_i - \bar{y})^2} \quad (۳)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_j - y_j\right| \quad (۴)$$

$$RMSE = \left[\frac{\sum_{i=1}^{n}(\hat{y}_j - y_j)^2}{n}\right]^{\frac{1}{2}} \quad (۵)$$

کـه در آن‌ها: می‌باشد همچنین y و به ترتیب مقدار مطلق و متوسط مقادیر خروجی شبکه عصبی و x و به ترتیب برابر مقادیر مطلق و متوسط فسفر کل اندازه‌گیری شده است.

نتایج و بحث

به‌منظور مدل‌سازی داده‌ها را به دودسته تقسیم کرده به‌طوری‌که ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای صحت سنجی اختصاص داده شد (امینی و همکاران ۲۰۰۵، منهاج ۲۰۰۹) برای این منظور دودسته داده به‌طور تصادفی تعیین گردید به‌طوری‌که ازنظر معیارهای آماری از قبیل میانگین و انحراف معیار تا حد ممکن شبیه یکدیگر باشند همچنین دو گروه داده تفاوت معنی‌داری در سطح ۵ درصد نداشتند. (توماسلاوهمکاران ۲۰۰۰ و سرمدیان و همکاران ۲۰۰۹) برای انتخاب داده‌های

آزمون و تست خود از تی تست استفاده کردند و بیان داشتند اگر بین داده‌های تست و آموزش تفاوت معنی‌داری وجود نداشته باشد می‌توان انتظار نتایج بهتری را از مدل داشت در جدول (۱) معیارهای آماری مربوط به داده‌های تست و آموزش ارائه‌شده است.

نتایج رگرسیون چند متغیره

برای تعیین رگرسیون چند متغیره (خطی و غیرخطی) پارامترهای موردمطالعه روابط رگرسیونی مربوطه که به‌صورت روابط (۶) و (۷) می‌باشد در رگرسیون چندگانه‌ی خطی اگر بین متغیرهای مستقل وابستگی خطی قوی مشاهده شود هم‌راستایی چندگانه به وجود می‌آید (نوابیان و همکاران ۲۰۰۳) ازآنجاکه فراوانی نسبی ذرات شامل رس، سیلت و شن هم‌راستا بوده برای جلوگیری از هم‌راستایی از نسبت رس به سیلت استفاده‌شده است.

رابطه (۶)

$$P = 33/85 - 0/580(Clay/silt) - 5/874(O.C) - 1/69(pH) + 0/042(Sand) - 0/023(Zn) - 1/93(Cadm)$$

(P<0/05)

رابطه (۷)

$$P = 74.706(clay/silt)^{-0.032}(O.C)^{-0.598}(pH)^{0.844}(Sand)^{0.056}(Zn)^{-0.028}(Cadm)^{-0.065}$$

(P<0/05)

نتایج مدل شبکه عصبی فازی (ANFIS)

در این ساختار مجموعه‌های فازی را می‌توان به‌عنوان وزن، ارتباطات و متغیرهای ورودی و خروجی را به‌عنوان نورون‌ها تفسیر کرد و الگوریتم یادگیری ساختارها، پارامترها و یا هردوی آن‌ها را اصلاح می‌کند. نتایج و خصوصیات مدل به‌کاررفته برای هر یک از پارامترها نیز در جدول ۲ ارائه‌شده است. بهترین ساختار با توجه به معیارهای آماری اعم از MAE و RMSE تعیین گردید.

جدول ۲-ویژگی‌های مدل ANFIS انتخابی برای پیش‌بینی فسفر کل خاک

پارامتر پیش‌بینی‌شده	روش یادگیری	نوع تابع عضویت	تکرار	تعداد توابع عضویت
فسفر کل خاک	هایبرید	گاوسی ۲	۱۰۰۰	۷

نتایج ارزیابی مدل‌ها

نتایج مربوط به ANFIS و رگرسیون چند متغیره مربوط به پارامترهای موردانداز‌ه‌گیری در جدول ۳ آورده شده است با توجه به این جدول ملاحظه می‌شود که در کل بهترین عملکرد مربوط به شبکه‌ی عصبی فازی (ANFIS) با ضریب تبیین ۰/۷۵ و RMSE و MAE به ترتیب ۲/۷۲و ۲/۰۶می‌باشد.

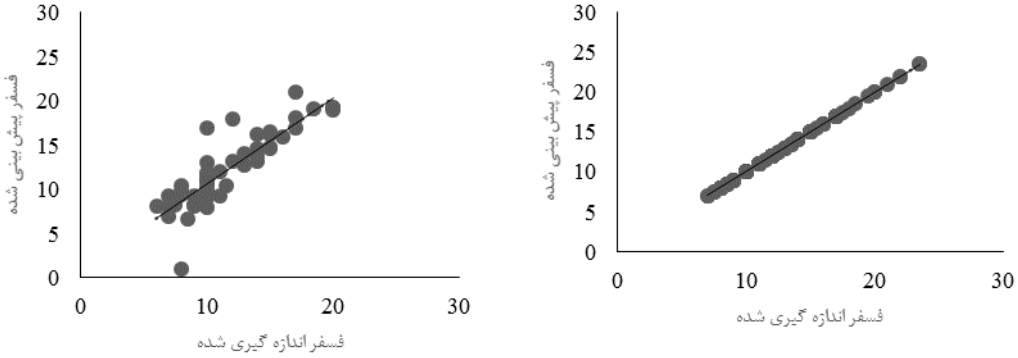
جدول ۳- نتایج به‌دست‌آمده از مدل‌های نروفازی، رگرسیون خطی و رگرسیون غیرخطی

خصوصیت خاک	مدل	مرحله‌ی محاسباتی	R^2	RMSE	MAE
	نروفازی (ANFIS)	آموزش	۰/۹۹	۱/۰۴	۰/۶۸
		صحت سنجی	۰/۷۵	۲/۷۲	۲/۰۶
	رگرسیون خطی		۰/۴۷	۳/۷۶	۳/۳۲
	رگرسیون غیرخطی		۰/۴۹	۲/۶۲	۲/۳۵

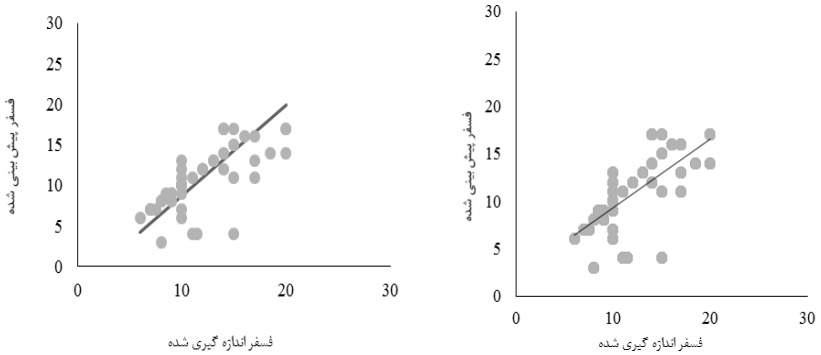
نتایج حاصل از مرحله‌ی آموزش نشان‌دهنده‌ی این است که هرچند مدل عصبی فازی در این مرحله به‌خوبی پیش‌بینی را انجام داده است اما در مرحله‌ی صحت سنجی توانایی این مدل کاهش قابل ملاحضه ای داشته است که می تواند بر اثر پراکندگی بالای داده ها باشد، مشخص‌شده است همان‌گونه که در جدول ۳ نشان داده‌شده است در شکل های ۳و۴ پراکنش

داده‌های اندازه‌گیری شده و پیش‌بینی‌شده‌ی شبکه‌ی عصبی فازی(ANFIS)، رگرسیون چند متغیره‌ی خطی و رگرسیون متغیره‌ی غیرخطی آورده شده است. همان‌طور که در این شکل‌ها مشاهده می‌شود مدل ANFIS با کمترین پراکندگی نسبت به برازش داده‌شده بالاترین ضریب تبیین را نشان می‌دهد؛ و کمترین دقت در پیش‌بینی مربوط به مدل رگرسیون خطی است که ضریب تبیین آن ۰/۴۷ است. در شکل ۵ همان‌طور که مشخص‌شده است ضرایب تبیین برای هر یک از مدل‌ها محاسبه‌شده است که نشان از کارایی بهتر مدل شبکه‌ی عصبی فازی (ANFIS) است.

الف ب



شکل۳- نمودار همبستگی متقابل داده‌های پیش‌بینی شده و اندازه‌گیری شده سرب در مرحله‌ی آموزش(الف) و صحت سنجی(ب)



شکل ۴-نمودار همبستگی متقابل داده‌های پیش‌بینی شده و اندازه‌گیری شده سرب در مدل رگرسیون خطی (الف) و رگرسیون غیر خطی(ب)

شکل ۵- نمودار ضریب تبیین مدل‌های مختلف در پیش‌بینی فسفر

نتیجه‌گیری

با توجه به معیارهای مورداستفاده در مطالعه‌ی حاضر (MAE،RMSE و R^2) جهت ارزیابی دقت مدل‌ها مورداستفاده در تخمین میزان فسفر خاک نتایج بررسی نشان داد که مدل شبکه عصبی فازی در مقایسه با مدل رگرسیون خطی چند متغیر در تخمین فسفر خاک از دقت و توانایی بالاتری برخوردار بوده است و همچنین بین دو شبکه نرو فازی و MLPشبکه چندلایه (MLP)

در تخمین فسفر خاک قابلیت بالاتری را دارا بوده است .در پایان باید اذعان داشت که یکی از دلایل کارایی بالای مدل‌های شبکه عصبی مصنوعی، ماهیت ابهامی پدیده‌های مرتبط به خاک و یا تقریبی بودن مقادیر اندازه‌گیری شده خصوصیات مختلف در محیط خاک است(کاشی و همکاران ۲۰۱۳)

منابع

سرمدیان ف.، تقی زاده ر.ا.، عسگری ح.م.، و اکبرزاده ع. ۱۳۸۸ . مقایسه روشهای نروفازی، شبکه عصبی و رگرسیون چند متغیره در پیش بینی. برخی خصوصیات خاک مطالعه موردی استان گلستان. مجله تحقیقات آب و خاک ایران۴۱:۲۲۰-۲۱۱

Amini, M., Afyuni, M., Fathianpourb, N., Khademi, H. and Fluchler, H. (2005). Continuous soil pollution mapping using fuzzy logic and spatial interpolation. Geoderma, 124, 223–233

Abraham, A, (2005), Artificial Neural Networks, Oklahama state university,Stillwater, usa

Altfelder S., Duijnisveld W. H. M., Streck T., Meyenburg G. and Utermann J. (2007). Quantifying the influence of uncertainty and variability on groundwater risk assessment for trace elements. Vandose Zone J., 6, 668-67.

Damangir, S., Jafarijashemi, G., Mamduhi, M., & Zohoor, H. (2006, September). Optimum Synthesis of Mechanisms for Path Generation Using a New Curvature Based—Deflection Based Objective Function. In The 6th WSEAS International Conference on Simulation, Modelling and Optimization, Lisbon, Portugal (pp. 672-676.(

Haykin S (1999) Neural networks, a comprehensive foundation,2nd edn. Prentice Hall, U

Jyh-Shing, Roger Jang, ”ANFIS: Adaptive Network-Based Fuzzy Inference Systems” IEEE Trans. On Systems, Man and Cybernetics, Vol. 23, No. 3, pp. 665-685, 1993.

Ingwersen J. (2001). The environmental fate of cadmium in the soils of the waste water irrigation area of Braunschweig: Measurement, Modeling and Assessment. PhD thesis, technische, University Braunschweig(Germany(

Ghorbani H., Kashi H., Hafezi Moghadas, N. and Emamgholizadeh S. (2015). Estimation of soil cation exchange capacity using multiple regression, artificial neural networks, and adaptive neuro-fuzzy inference system models in Golestan province, Iran. Commu in Soil Sci and Plant Anal., 46(6), 763-780.

Menhaj, M. (2009). Fundamental of Artificial neural networks, Amirkabir Press, 245p.

Olsen, S.R., and Sommers, J.F. 1982. Phosphorus. P 403-430, In: A.L. Page (ed.), Methods of soil Analysis. Agron. No. 9, part 2: Chemical and microbiological properties, 2nd edition, Am. Soc. Agron., Madison, WI, USA

Tomasella, J., Hodnett, M. G. and Rossato, L. (2000). Pedotransfer functions for the estimation of soil water retention in Brazilian soils. Journal of Soil Science Society of America, 49, 1100- 1105.

Comparison of Adaptive neuro fuzzy inference system (ANFIS) and multiple regression to predict soil phosphorus using soil properties

A. Bazoobandi[1], H. Ghorbani[2], S. Emamgholizadeh[2]
1. PhD Student, Faculty of Agriculture, Ferdowsi University of Mashhad
2. Associate Professor Faculty of Agriculture, Shahrrood University

Abstract

In recent years the use of indirect methods such as Adaptive Neuro-Fuzzy Inference System (ANFIS) to estimate soil parameters has been addressed by researchers. In this study, for estimation of the soil phosphorus 250 soil samples from Gilan province were used. The physical and chemical parameters of soil such as soil particle size, pH, organic carbon, zinc, cadmium used as input data to estimate the phosphor. The collected data were classified into two categories: training (80%) and test (20%) and they were used for training and testing of ANN ANFIS, linear and nonlinear regression models. The accuracy of these models to prediction of measured and predicted phosphor evaluated by statistics parameters such as mean-absolute-error (MAE), root-mean-square-error (RMSE) and determination coefficient (R2). The results showed the high efficiency of ANFIS model with values of MAE (2.06), RMSE (2.72) and R2 (0.75).

Keywords:MLP, Gilan province, ANFIS,Soil phosphorus