

کاربرد الگوریتم ژنتیک در برآورد توزیع مکانی پهاش خاک با ترکیب زمین آمار و برنامه‌ریزی بیان ژن

لاله پرویز^{۱*}، رضا احدی^۲، فاطمه سید میلانی^۲

(تاریخ دریافت: ۱۴۰۰/۰۲/۰۶ تاریخ پذیرش: ۱۴۰۰/۰۷/۰۹)

چکیده

آگاهی در مورد توزیع مکانی ویژگی‌های خاک نقش مهمی در سیستم رشد و عملکرد محصول در بخش کشاورزی دارد. در این راستا جهت افزایش دقت در توزیع مکانی pH خاک، الگوریتم ژنتیک در ترکیب برآوردهای رگرسیون خطی، برنامه‌ریزی بیان ژن (GEP) و زمین‌آمار (درون‌یابی کریجینگ) در مناطقی از استان آذربایجان شرقی (شهرستان‌های بناب، عجب‌شیر و مراغه) مورد استفاده قرار گرفت. مدل نیم‌تغییرنمای خطی در درون‌یابی کریجینگ دارای کمترین مقدار خطا است. برنامه‌ریزی بیان ژن و رگرسیون خطی به ترتیب در برآورد توزیع مکانی pH خاک، کمترین و بیشترین مقدار خطا را دارند، میزان کاهش میانگین مربع خطا، خطای جذر میانگین مربعات، خطای جذر میانگین مربعات نسبت به رگرسیون خطی به زمین‌آمار بهترتیب ۲۶/۴۷ و ۲۷/۴۷، ۵۸/۶۷ و ۳۷/۰۳ و ۲۳/۸ درصد است. روش ترکیب استفاده از الگوریتم ژنتیک نسبت به سه نوع مدل‌سازی، خطای برآورد توزیع مکانی pH خاک را کاهش می‌دهد، به‌عنوان نمونه میزان کاهش میانگین مربع خطا، خطای جذر میانگین مربعات، خطای جذر میانگین مربعات نسبت به رگرسیون خطی به ترتیب برابر با ۱۱/۲۳، ۷۶/۳۳ و ۱۰ درصد است. برآوردهای pH با روش ترکیبی در محدوده قلیایی می‌باشد که با مقادیر اندازه‌گیری شده هم‌خوانی دارد. کمینه و بیشینه مقدار قدر مطلق اختلاف بین مقادیر اندازه‌گیری و تخمینی در نقاط مورد بررسی به ترتیب در بناب (۰/۰۹) و عجب‌شیر (۰/۲۵) می‌باشد. معیار میانگین درصد خطای مطلق در روش ترکیبی در محدوده قابل قبول از نظر عملکرد است و این مسئله کارایی روش مورد استفاده را در برآورد توزیع مکانی پهاش خاک نشان می‌دهد.

واژه‌های کلیدی: ویژگی‌های خاک، درون‌یابی، خطا، ترکیبی

پرویز ل. ، احدی ر.، سید میلانی ف. ۱۴۰۱. کاربرد الگوریتم ژنتیک در برآورد توزیع مکانی پهاش خاک با ترکیب زمین‌آمار و برنامه‌ریزی بیان ژن. تحقیقات کاربردی خاک. جلد ۱۰، شماره ۳. صفحه: ۱۱۷-۱۲۹.

1- دانشیار دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

2- دانش‌آموخته کارشناسی دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

*پست الکترونیک: laleh_parviz@yahoo.com

مقدمه

و خاک و با به‌کارگیری زمین‌آمار و تحلیل مؤلفه‌های اصلی و رگرسیون خطی برخی از خصوصیات خاک درون‌یابی شدند. درون‌یابی کریجینگ بیزین^۱ نتایج قابل قبولی برای منیزیم، پتاسیم، فسفر، pH و نیتروژن کل داشت (John *et al.*, 2021). از کریجینگ معمولی و رگرسیون کریجینگ در تهیه نقشه‌های pH، کربن آلی خاک و نیتروژن کل در منطقه‌ای در ویتنام استفاده شد. دقت دو روش بکار گرفته‌شده در برآورد خصوصیات خاک متفاوت بود (Gia Pham *et al.*, 2019). خطای مربوط به روش‌های زمین‌آمار با خطا گزارش شده است که دلیل آن مربوط به تابع برازش داده شده است که شامل کلیه نقاط تجربی محاسبه شده نیست (Zamani Ahmad Mahmoodi *et al.*, 2014). در این میان از گزینه‌های مطرح‌شده در برآورد توزیع مکانی استفاده از روش‌های هوش مصنوعی است. مدل عددی، روش هوش مصنوعی (شبکه عصبی مصنوعی، شبکه عصبی - فازی و شبکه عصبی - موجک) و زمین‌آمار در برآورد سطح آب زیرزمینی در آبخوان دشت همدان - بهار استفاده شدند. روش شبکه عصبی - موجک و مدل عددی به ترتیب کمترین و بیشترین مقدار خطا را داشتند (Bayat Varkeshi & Fasihi, 2018). از روش‌های زمین‌آمار (کریجینگ و کوکریجینگ) و شبکه عصبی مصنوعی در تخمین سرعت آستانه فرسایش بادی در منطقه جزینک دشت سیستان استفاده شد. براساس آماره‌های ارزیابی، مدل پرسپترون چند لایه با سه لایه مخفی و سه نرون دارای دقت بالایی بوده است (Arbab *et al.*, 2018). روش‌های زمین‌آمار (کریجینگ) و شبکه عصبی مصنوعی پرسپترون چند لایه در دشت شبستر - صوفیان در برآورد سطح تراز آب زیرزمینی به‌کار گرفته شدند. عملکرد شبکه عصبی با میزان همبستگی ۰/۹۶ نسبت به زمین‌آمار بهتر بود (Ebadi *et al.*, 2019). با توجه به سابقه روش‌های هوش مصنوعی و زمین‌آمار در برآورد توزیع مکانی داده‌ها، جهت افزایش کارایی در دقت برآورد مکانی داده‌ها، روش‌های ترکیب‌گزینه مناسبی به‌نظر می‌رسند. روش ترکیبی شبکه عصبی مصنوعی (بهینه‌شده با الگوریتم ژنتیک) و زمین‌آمار در برآورد سطح استابی در دشت رامهرمز استفاده شد.

خاک در نتیجه فرآیندهای طبیعی در جهت افقی و عمودی دارای تغییرات است (Zhang & Hartemink, 2021). در این راستا نیاز زیادی به اطلاعات خاکمی باشد، زیرا آگاهی در مورد خاک و نحوه توزیع آن در مدیریت و حفاظت از آن ضروری است. اطلاعات خاک را می‌توان با استفاده از نقشه‌های خاک تهیه کرد که در واقع نمایش گرافیکی جهت بیان اطلاعات در مورد توزیع مکانی خصوصیات خاک است (Calzolari *et al.*, 2021). در این میان ارزیابی تغییرات مکانی پ‌ه‌اش و بافت خاک به‌علت تأثیر بر ساختمان خاک، وضعیت حاصلخیزی، قابلیت‌هدایت هیدرولیکی، نفوذ و نقش آن در مدیریت پایدار خاک مهم است (Pahlavan-Rad & Akbarrimoghaddam, 2018). از روش‌های پرکاربرد در سال‌های اخیر جهت بررسی توزیع مکانی داده‌ها، استفاده از روش‌های زمین‌آمار است. با استفاده از روش‌های زمین‌آمار توزیع مکانی برخی از ویژگی‌های فیزیکی و شیمیایی خاک در منطقه میانکنگی سیستان مورد بررسی قرار گرفت. تهیه نقشه‌های pH و قابلیت‌هدایت الکتریکی خاک با استفاده از متغیر کمکی درصد رس، درصد رطوبت اشباع و محتوای درصد ذرات رس با متغیر کمکی‌هدایت الکتریکی و کربنات کلسیم معادل با متغیر کمکی‌درصد سیلت انجام گرفت. برای تهیه نقشه درصد سیلت و درصد شن از روش کریجینگ گسسته استفاده شد. تغییرات مکانی مربوط به قابلیت‌هدایت الکتریکی خاک منطقه بسیار شدید بود (Hashemi *et al.*, 2016). با استفاده از نمونه‌برداری از ۵۴ نقطه در بخشی از شرق دشت فراهان در استان مرکزی تغییرات مکانی بافت خاک (درصد اندازه ذرات خاک) پیش‌بینی شد. شاخص‌های ارزیابی به‌دست آمده حاکی از کارایی مدل کریجینگ در این تحقیق بود (Afzali *et al.*, 2018). از درون‌یابی کریجینگ معمولی در تخمین توزیع مکانی برخی خصوصیات خاک (pH، قابلیت‌هدایت الکتریکی، فسفر، پتاسیم و کربن آلی) استفاده شد. در این تحقیق تعداد نمونه‌ها ۳۲ و نمونه‌برداری به‌صورت تصادفی بود. نیم‌تغییرنما به‌صورت نمایی در نظر گرفته شد (Shit *et al.*, 2016). با استفاده از متغیرهای محیطی

شده با دستگاه سیستم موقعیت‌یاب جهانی ثبت شد. در این تحقیق جهت برآورد دقیق توزیع مکانی پهاش خاک از رگرسیون خطی چند متغیره، برنامه‌ریزی بیان ژنو زمین‌آمار (درون‌یابی به روش کریجینگ) استفاده شد. سپس ترکیب تخمین‌های مدل‌سازی‌های مختلف با به-کارگیری الگوریتم ژنتیک انجام گرفت. در واقع هدف از استفاده از الگوریتم ژنتیک تعیین صحیح وزن‌های مربوط به تخمین روش‌های مختلف است.

درون‌یابی کریجینگ

کریجینگ یک روش تخمینی جهت برآورد توزیع مکانی داده‌های مورد بررسی است و پایه روش براساس میانگین متحرک وزنی است. تعریف تخمینگر براساس رابطه ۱ می‌باشد.

$$Z^*(x) = \sum_{i=1}^N \lambda_i Z(x_i) \quad (1)$$

$Z^*(x)$: مقدار اندازه‌گیری‌شده در موقعیت i ، λ_i : مقدار وزن نقاط معلوم در موقعیت i ، x : موقعیت مجهول (Mirzaee et al., 2017 - Nadiri et al., 2017). بافت نمونه‌های جمع‌آوری‌شده در محدوده سبکاست. خاک منطقه در بازه قلیایی قرار دارد.

برنامه‌ریزی بیان ژن

برنامه‌ریزی بیان ژن یک روش ناپارامتری است (Eleyedath et al., 2021). برنامه‌ریزی بیان ژن یک الگوریتم ژنتیک بر پایه جمعیت کروموزوم‌های انتخاب شده است که تنوع ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی معرفی می‌کند (Shi et al., 2015). در واقع برنامه‌ریزی بیان ژن توسعه الگوریتم‌های ژنتیک (GA^2) و برنامه‌ریزی ژنتیک (GP^3) است. برنامه‌ریزی بیان ژن یک الگوریتم یادگیری است و این یادگیری مربوط به روابط بین متغیرها در مجموعه داده‌ها است که مدل‌هایی برای بیان این روابط ایجاد می‌شود. همانند طبیعت به فرآیند رمزگشایی اطلاعات، ترجمه گفته می‌شود. این ترجمه به‌طور آشکار حاکی از نوعی کد و مجموعه‌ای از قوانین است.

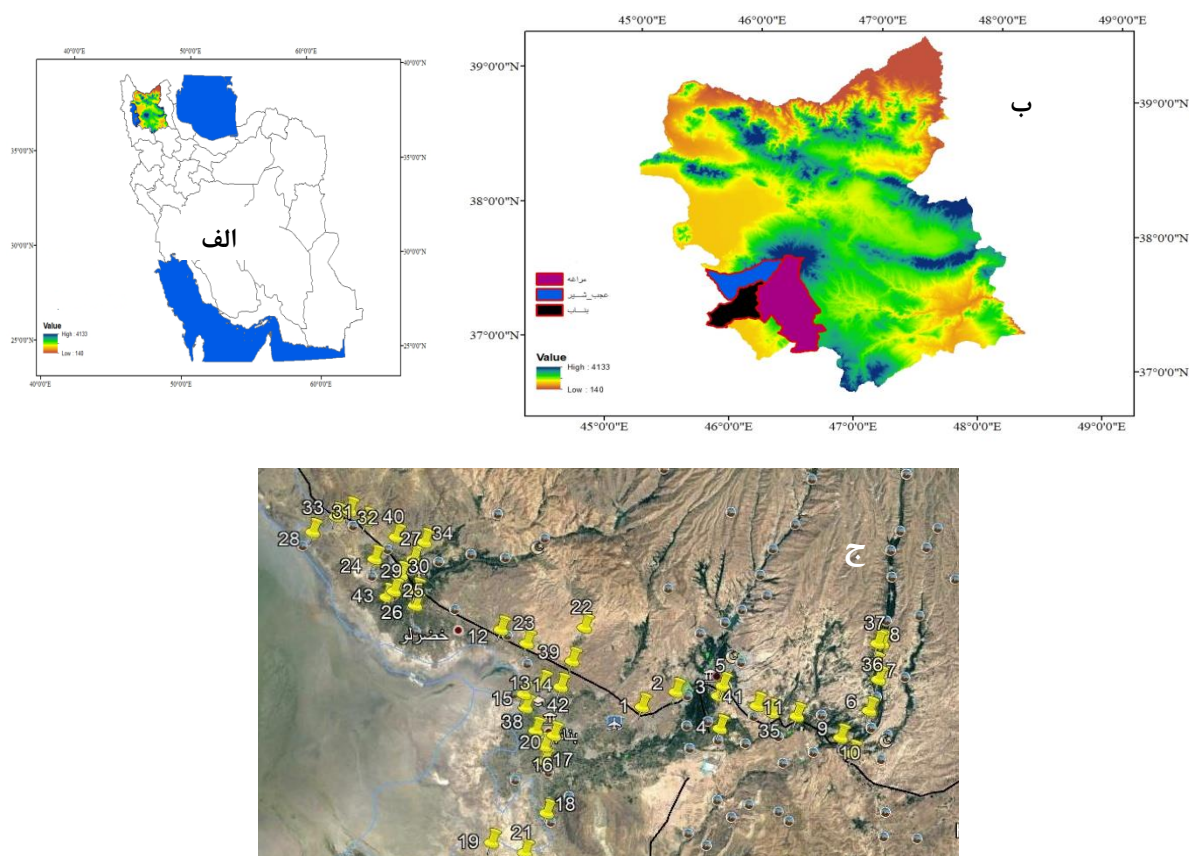
روش ترکیبی نسبت به مدل زمین‌آمار دارای دقت بالایی بود (Zamani Ahmad Mahmoodi et al., 2014). به-گونه‌ای دیگر هم در تخمین سطح آب زیرزمینی، ترکیب روش هوش مصنوعی و زمین‌آمار در دشت هادی شهر مورد بررسی قرار گرفت. در این حالت از پیرومتر مرکزی برای مدل‌سازی شبکه عصبی مصنوعی استفاده شد. سپس با روش خوشه‌بندی سلسله‌مراتبی پیرومترهای انتخاب‌شده، دسته‌بندی شدند و سپس با استفاده از مدل شبکه عصبی و فازی ساگنو هر گروه از پیرومترها مدل‌سازی شدند (Habibi et al., 2016). از ترکیب شبکه عصبی مصنوعی و سامانه اطلاعات جغرافیایی (GIS) در ارزیابی هوشمند جهت تعیین زمین مناسب برای کشت محصولات استفاده شد. شبکه عصبی طراحی‌شده با استفاده از نمونه‌های استخراج‌شده از سامانه اطلاعات جغرافیایی آموزش داده شد (Ahmadi & Layegh, 2015). بنابراین ترکیب مناسب روش‌های برآورد توزیع مکانی داده‌ها در افزایش دقت توزیع مکانی آن‌ها دارای اهمیت بسزایی است و استفاده از روشی کارآمد در این زمینه بر نقایص مدل‌سازی انفرادی غلبه می‌کند.

هدف این تحقیق به‌کارگیری روش ترکیبی جهت برآورد توزیع مکانی pH خاک است. برای این منظور از رگرسیون خطی چند متغیره، برنامه‌ریزی بیان ژن (GEP^1) و زمین‌آمار (درون‌یابی کریجینگ معمولی) استفاده شد. اساس ترکیب انتخاب وزن مناسب به مدل‌های مورد-استفاده براساس الگوریتم ژنتیک است. در بخش زمین‌آمار مدل‌های نیم‌تغییرنمای مناسب‌درون‌یابی کریجینگ نیز مورد بررسی قرار گرفت.

مواد و روش‌ها

منطقه مورد مطالعه

نمونه‌های مورد مطالعه مربوط به مناطقی از استان آذربایجان شرقی شامل شهرستان‌های مراغه، بناب و عجب‌شیر است که نقاط نمونه‌برداری در شکل ۱ نشان داده شده است. تعداد ۴۳ نمونه به روش تصادفی جمع-آوری شدند و نمونه‌برداری از خاک سطحی (۰ - ۳۰ سانتی‌متری) انجام گرفت. موقعیت نقاط نمونه‌برداری



شکل ۱- موقعیت مکانی استان مورد مطالعه در کشور (الف)، موقعیت مناطق نمونه برداری در استان آذربایجان شرقی (ب)، موقعیت نقاط نمونه برداری شده (ج)

Figure 1. The location of studied province in the country(a), the location of sampling areas in East Azerbaijan province(b), the location of sampling points

$$\sqrt{(a-b)*(c+d)} \quad (3)$$

الگوریتم ژنتیک

الگوریتم ژنتیک برای بهینه‌سازی مسائل پیچیده بکار می‌رود. پایه این روش بر مبنای فرآیند تکامل است. این روش با تولید نسل‌های متعدد از جواب‌های ممکن سعی در حرکت به طرف جواب بهینه دارد. در نسل‌ها به جواب‌های مناسب‌تر و بهتر امکان بقا در تولید جواب‌ها داده می‌شود. در این نوع مدل بهینه‌سازی جهت ایجاد کروموزوم (تولید نسل براساس یکسری کروموزوم است) باید فضای تصمیم‌گیری کدگذاری شود؛ بنابراین در این روش یک جمعیت اولیه از جواب‌ها به صورت تصادفی انتخاب می‌شوند. با استفاده از تابع هدف و قیود هر کدام از کروموزوم‌ها ارزیابی می‌شوند. از مقادیر تولیدشده اولیه تعدادی از جواب‌های برتر (با توجه به تابع هدف) به عنوان

کد ژنتیکی برنامه‌ریزی بیان ژن بسیار ساده است: یک رابطه یک به یک بین ژن و گره‌ها که به صورت نمایش درختی است. قوانین بسیار ساده هستند که وظیفه آن‌ها سازمان‌دهی مکانی گره‌ها در بیان درختی (ET1) و نوع تعامل بین زیر درخت‌ها. هر کروموزوم در برنامه‌ریزی بیان ژن یک رشته‌خطی با طول ثابت است که از ژن‌ها تشکیل یافته است. سازمان‌دهی ساختار ژن‌های برنامه‌ریزی بیان ژن با توجه به چارچوب تفسیر باز (ORF²) بهتر درک می‌شود. رشته زیر یک چارچوب تفسیر باز از کروموزوم برنامه‌ریزی بیان ژن با طول هشت است (۲) و بیان درختی مربوطه در شکل ۲ آورده شده است که فرم ریاضی آن به صورت رابطه ۳ است.

$$\left\{ \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ Q & * & + & - & a & b & c & d \end{matrix} \right\} \quad (2)$$

1.Expression Tree, ET
2.Open Reading Frames, ORF

$$AMAPE = \frac{1}{N} \sum_{i=1}^N \frac{|O_i - F_i|}{\frac{1}{N} \sum_{i=1}^N O_i} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - F_i)^2} \quad (5)$$

$$RRMSE = \frac{RMSE}{O} \quad (6)$$

$$SI = \frac{\sqrt{RMSE}}{O} \quad (7)$$

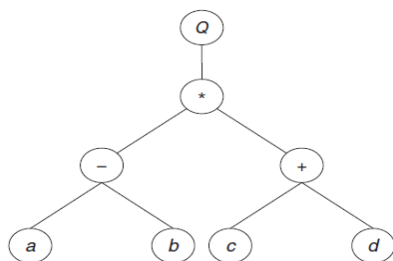
$$MSE = \frac{\sum_{i=1}^N (F_i - O_i)^2}{N} \quad (8)$$

$$UI = \frac{\left[\sum_{i=1}^N (O_i - F_i)^2 \right]^{0.5}}{\left[\sum_{i=1}^N (O_i)^2 \right]^{0.5} + \left[\sum_{i=1}^N (F_i)^2 \right]^{0.5}} \quad (9)$$

$$UII = \frac{\left[\sum_{i=1}^N (O_i - F_i)^2 \right]^{0.5}}{\left[\sum_{i=1}^N (O_i)^2 \right]^{0.5}} \quad (10)$$

F: داده‌های تخمینی، O: داده‌های اندازه‌گیری شده، N: تعداد داده‌ها جهت ارزیابی مدل.

در مقایسه عملکرد چندین مدل، مدلی با کمینه مقدار معیار خطا قابل قبول است (Del Carmen et al., 2017; Chen et al., 2021; Zaynoddin, 2018).



شکل ۲- نمونه‌ای از بیان درختی مربوط به برنامه‌ریزی بیان ژن (Luan et al., 2008)

Figure 2. The example for GEP expression tree (Luan et al., 2008)

کروموزوم‌های والد برای تولید نسل بعد انتخاب می‌شوند. سپس به صورت تصادفی کروموزوم‌ها در دسته‌های دوتایی قرار می‌گیرند و نسل بعد را تولید می‌کنند. در فرآیند بهینه‌سازی قسمت‌های از اطلاعات ژنتیکی دو کروموزوم با هم تعویض می‌شود که این مسئله با استفاده از عملگر ترویج انجام می‌گیرد. برای ممنوعیت در رسیدن به جواب‌های بهینه موضعی عملگر جهش بکار گرفته می‌شود. با تکرار مراحل بیان شده نسل‌های جدید جایگزین نسل‌های قبل شده و روند محاسبات تا آنجایی ادامه پیدا می‌کند که با افزایش تعداد تکرارها در جواب نهایی بهبودی مشاهده نشود (Karamouz & Kerachian, 2003).

به‌طور کلی مراحل انجام تحقیق را می‌توان به صورت خلاصه بیان کرد: گام اول: جمع‌آوری داده‌های مربوط به ویژگی‌های خاک (pH) به همراه اطلاعات مکانی نقاط جمع‌آوری شده. گام دوم: استفاده از مدل‌سازی جهت تعیین توزیع مکانی ویژگی‌های خاک همراه با انجام تحلیل حساسیت هر نوع مدل‌سازی (۱-۲) - رگرسیون خطی ۲-۲ - زمین‌آمار: درون‌یابی کریجینگ معمولی ۲-۳ - برنامه‌ریزی بیان ژن). گام سوم: استفاده از روش ترکیبی جهت ادغام برآوردهای رگرسیون خطی، زمین‌آمار و برنامه‌ریزی بیان ژن (استفاده از اطلاعات گام دوم) با اختصاص وزن به هر نوع مدل‌سازی (۱-۳) - استفاده از الگوریتم ژنتیک در تعیین وزن‌های بهینه براساس تابع هدف و قیود تعریف شده). گام چهارم: مقایسه عملکرد روش ترکیبی با الگوریتم ژنتیک با نتایج بخش گام دوم با استفاده از معیارهای ارزیابی.

معیارهای ارزیابی

برای بررسی عملکرد روش ترکیبی و مقایسه آن با عملکرد رگرسیون خطی، برنامه‌ریزی بیان ژن و زمین‌آمار در برآورد توزیع مکانی pH خاک از برخی معیارهای خطا مانند RMSE (خطای جذر میانگین مربعات)، RRMSE (خطای جذر میانگین مربعات نسبی)، MSE (میانگین مربع خطا)، SI (شاخص پراکندگی)، AMAPE (میانگین درصد خطای مطلق تعدیل شده)، معیار دقت پیش‌بینی مدل (UI)، معیار کیفیت پیش‌بینی مدل (UII) استفاده شد که در ادامه معادلات معیارها آورده شده است.

نتایج و بحث

تا ۱۵ درصد) ۲- دسته بالا (۱۵ تا ۷۵ درصد) (Dahiya et al., 1984; Hashemi et al., 2016). ضریب تغییرات محاسبه شده داده‌های pH خاک در محدوده پایین قرار دارد. در تحقیق هاشمی و همکاران (Hashemi et al., 2016) نیز ضریب تغییرات پ‌هاش در محدوده پایین قرار داشتند. یکی از عوامل مؤثر در ضریب تغییرات کم در مورد پ‌هاش خاک می‌تواند ناشی از عوامل ذاتی مانند مواد مادری باشد (Foroughifar et al., 2010).

مقایسه عملکرد مدل‌سازی‌های مورد مطالعه در برآورد توزیع مکانی پ‌هاش خاک

در بخش زمین‌آمار، درون‌یابی با کریجینگ معمولی و با استفاده از نرم افزار ArcGIS 10.1 انجام گرفت. مسئله مورد نظر مربوط به تعیین مدل‌های نیم‌تغییرنمای خصوصیت مورد نظر بود که جهت انتخاب مدل‌ها، از محاسبه معیار خطا در ۹ نقطه انتخابی با مدل‌های مختلف استفاده شد که نتایج در شکل ۳ آورده شده است.

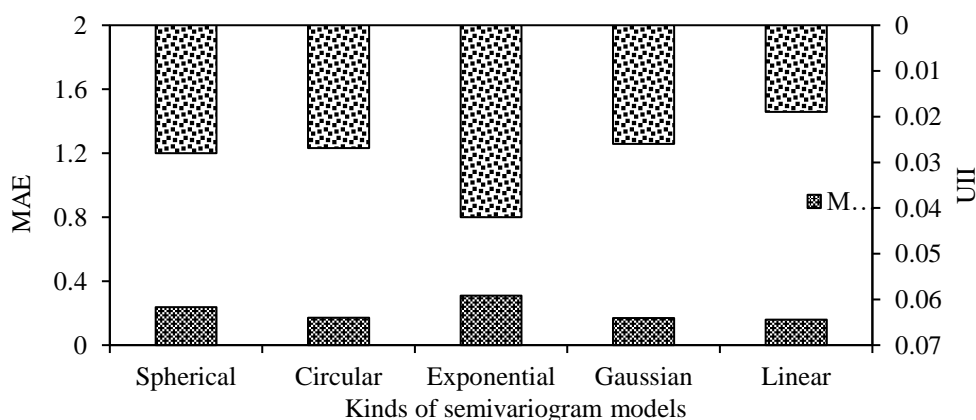
در این تحقیق جهت تخمین توزیع مکانی پ‌هاش خاک از روش ترکیبی برآوردهای مربوط به مدل‌سازی انفرادی استفاده شد. برای این منظور از اطلاعات ۴۳ نقطه استفاده شد که ۹ نقطه جهت ارزیابی روش ترکیبی انتخاب شدند. نقاط برداشت شده از مناطق عجب‌شیر، مراغه و بناب انجام گرفته است که در ۹ نقطه که برای صحت‌سنجی در نظر گرفته شده‌اند، سه منطقه در نظر گرفته شده‌اند. برخی از خصوصیات آماری مقادیر اندازه‌گیری شده و نتایج آزمون نرمال بودن داده‌ها در جدول ۱ آورده شده است.

با توجه به جدول ۱ سطح معنی‌داری در آزمون‌های بررسی نرمال بودن داده‌ها (کولموگروف-اسمیرنوف و شاپیرو-ویلک) بیشتر از ۵ درصد است، بنابراین می‌توان داده‌ها را با اطمینان بالایی نرمال فرض کرد. همچنین ترسیم نمودار احتمالاتی مربوط به توزیع نرمال نیز حاکی از نرمال بودن داده‌ها است ($P\text{-Value}=0.27$). تحقیقات نشان داده‌اند که محدوده طبقه‌بندی ضریب تغییرات ویژگی-های خاک به دو دسته تقسیم می‌شود ۱- دسته پایین (۰

جدول ۱- مقادیر مربوط به خصوصیات آماری داده‌های اندازه‌گیری شده پ‌هاش خاک و بررسی نرمال بودن داده‌ها

Table 1. The values of statistical properties of measured soil pH data and investigation the normality of data

Mean	8.55	Skewness coefficient	-0.2
Standard deviation	0.26	Kurtosis coefficient	1.69
Coefficient of variation	3.1	Kolmogorov-Smirnov test	0.2
Shapiro-Wilk test	0.35		



شکل ۳- مقادیر معیار خطا در انتخاب مدل‌های نیم‌تغییرنمای درون‌یابی کریجینگ مربوط به پ‌هاش خاک

Figure 3. Error values for selection the variogram models of kriging interpolation related to soil pH

روش زمین‌آمار (درون‌یابی کریجینگ) نسبت به نوع مدل نیم‌تغییرنا در تحقیقات شعبانی و همکاران (Shabani *et al.*, 2020) در برآورد برخی از خصوصیات خاک مانند کربن آلی، مقامی و همکاران (Maghami *et al.*, 2011) در پهنه‌بندی کیفیت آب، ندیری و همکاران (Nadiri *et al.*, 2014) در تخمین آلودگی نترات در منابع آب زیرزمینی، دهقانی و نورعلی (Dehghani & Noorali 2016) در برآورد سطح آب زیرزمینی و بلخیری و همکاران (Belkhiri *et al.*, 2020) در برآورد توزیع مکانی کیفیت آب زیرزمینی نیز مشاهده شد.

در مورد مدل‌سازی رگرسیون خطی و برنامه‌ریزی بیان ژن جهت برآورد توزیع مکانی پهاش خاک از موقعیت مکانی نقاط اندازه‌گیری شده به‌عنوان داده‌های ورودی استفاده شد. همچنین پارامترهای تعیین‌شده مدل‌سازی GEP بعد از تحلیل حساسیت در جدول ۲ آورده شده است.

با توجه به شکل ۳ کمینه مقدار MAE و UII مربوط به مدل خطی است. درصد کاهش UII از مدل نمایی با بیشترین خطا به مدل‌های کروی، دایره‌ای، گوسی و خطی به ترتیب ۳۳/۳۳، ۳۵/۳۹، ۳۸/۰۹ و ۵۴/۷۶ درصد است. که بیشترین میزان کاهش مربوط به مدل خطی است. میزان کاهش UII از مدل‌های کروی، دایره‌ای، نمایی، گوسی به خطی به ترتیب برابر با ۳۲/۱۴، ۲۹/۳۶، ۵۴/۷۶ و ۲۶/۹۲ درصد و میزان کاهش MAE به ترتیب برابر با ۳۲/۷۷، ۶/۴۳، ۴۸/۳۸ و ۵/۸۸ درصد است. در بررسی دیگر نیز قدر مطلق اختلاف بین متوسط مقادیر اندازه‌گیری و تخمینی در ۹ نقطه محاسبه شد که نتایج برای مدل‌های کروی، دایره‌ای، نمایی، گوسی و خطی به ترتیب ۰/۰۷۹/۰۶، ۰/۰۳۲/۱۸، ۰/۰۱۴، ۰/۰۱۴ بود. در این حالت نیز کمترین اختلاف مربوط به مدل خطی بود که با نتایج مربوط به محاسبات معیار خطا همخوانی دارد. حساسیت

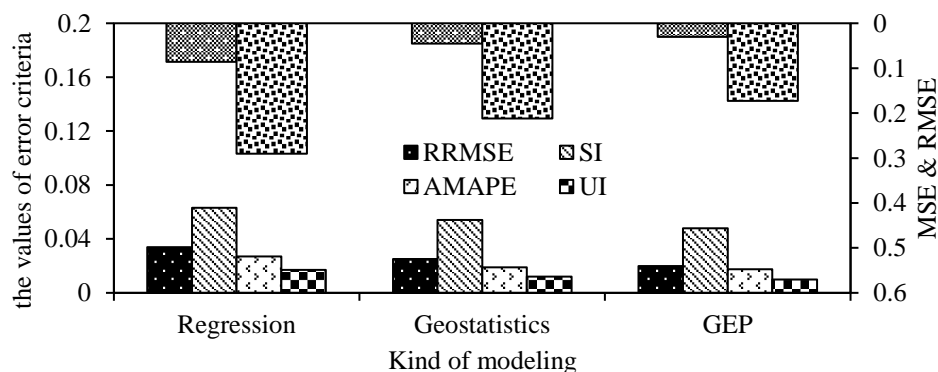
جدول ۲- مقادیر پارامترهای مدل برنامه‌ریزی بیان ژن

Table 2. The values of gene expression programming model parameters

One point recombination rate	0.2	Number of generation	1000
Two point recombination rate	0.3	Number of chromosome	15
Gene recombination rate	0.2	Head Size	8
Linking function	+	Number of genes	3
Inversion rate	0.1	Mutation rate	0.044

سر برابر با ۷، تعداد ژن برابر با ۳ و تعداد نسل ۱۰۰۰ به تعداد کروموزوم برابر با ۱۵، ۷/۹۷ درصد بود. در برآورد توزیع مکانی pH از سه نوع مدل‌سازی (رگرسیون خطی، برنامه‌ریزی بیان ژن و زمین‌آمار) استفاده شد که نتایج عملکرد سه نوع مدل‌سازی در شکل ۴ آورده شده است.

تحلیل حساسیت یکی از مراحل مهم در مدل‌سازی است، به‌طوری‌که میزان کاهش RMSE از حالت تعداد کروموزوم برابر با ۱۵، اندازه سر برابر با ۸، تعداد ژن برابر با ۳ و تعداد نسل برابر با ۱۰۰۰ به تعداد نسل برابر با ۱۰۰، ۹۱/۱۲ درصد و از حالت تعداد کروموزوم برابر با ۵۰، اندازه



شکل ۴- مقایسه عملکرد مدل‌سازی‌های مورد استفاده در تخمین پهاش خاک براساس معیارهای خطای مختلف

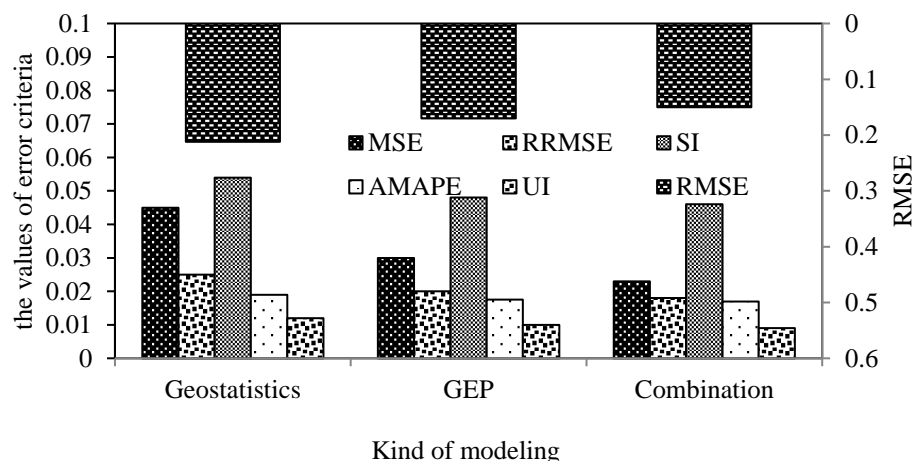
Figure 4. Comparison the performance of used modeling for soil pH estimation basis on different error criteria

(Arbab *et al.*, 2018) و همکاران (Varkeshi&Fasihi, 2018) نیز (Ebadi *et al.*, 2019) همکاران مشاهده شد.

استفاده از الگوریتم ژنتیک در ترکیب برآوردهای مدلسازی‌های توزیع مکانی پ-هاش خاک

در ادامه جهت افزایش دقت در برآورد توزیع مکانی pH خاک از ترکیب نتایج مدل‌سازی رگرسیونی، زمین‌آمار و برنامه‌ریزی بیان ژن استفاده شد که برای این منظور از الگوریتم ژنتیک جهت کمینه‌سازی خطای بین مقادیر اندازه‌گیری‌شده و تخمینی استفاده شد. پارامترهای مربوط به الگوریتم ژنتیک شامل احتمال تزویج برابر با ۰/۷۸، احتمال جهش یکنواخت برابر با ۰/۰۱، تعداد تکرار ۱۰۰۰ است (قابل ذکر است که بیشترین حساسیت الگوریتم ژنتیک به احتمال تزویج بود) و نتایج در شکل ۵ آورده شده است.

براساس شکل ۴، مقادیر معیارهای خطا در برنامه‌ریزی بیان ژن کمترین و در رگرسیون خطی بیشترین مقدار را داشتند، به‌عنوان نمونه میزان کاهش MSE، RMSE، RRMSE، SI، AMAPE و UI در برآورد پ-هاش از رگرسیون به زمین‌آمار به ترتیب برابر با ۴۷/۶۷، ۲۷/۵۸، ۲۶/۴۷، ۱۴/۲۸ و ۲۹/۶۲ و ۲۹/۴۱ درصد و از رگرسیون به برنامه‌ریزی بیان ژن به ترتیب برابر با ۶۵/۱۱، ۴۱/۳۷، ۲۳/۴۱، ۸/۱۷ و ۳۷/۰۳ و ۴۱/۱۷ درصد و از زمین‌آمار به برنامه‌ریزی بیان ژن به ترتیب برابر با ۳۳/۳۳، ۱۸/۳۹، ۲۰، ۱۱/۱۱، ۱۰/۵۲ و ۱۶/۶ بود. بیشترین مقدار درصد کاهش مربوط به رگرسیون به برنامه‌ریزی بیان ژن و سپس رگرسیون به زمین‌آمار است. در مقایسه عملکرد بین روش‌های هوش مصنوعی و زمین‌آمار، روش‌های هوش مصنوعی دارای قابلیت بیشتری بودند که این مسئله در تحقیقات بیات ورکشی و فصیحی (Bayat



شکل ۵- مقایسه عملکرد مدلسازی برنامه‌ریزی بیان ژن و زمین‌آمار با روش ترکیبی با استفاده از الگوریتم ژنتیک

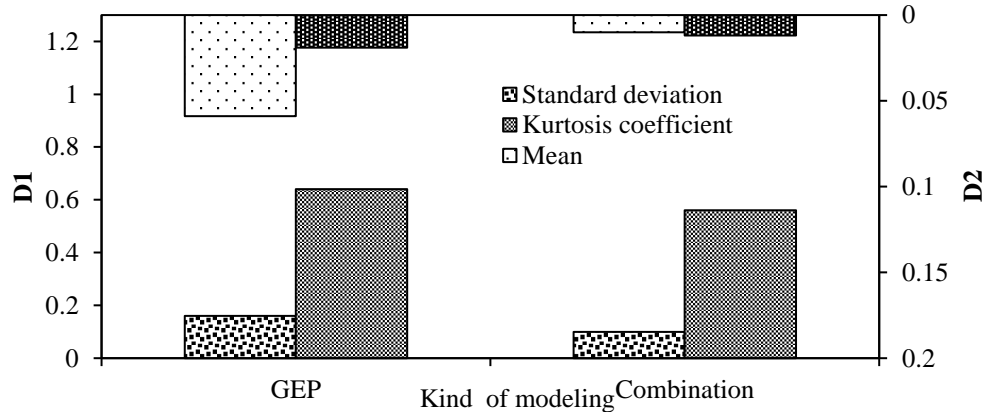
Figure 5. Comparison the performance of gene expression programming and geostatistics with combination method using genetic algorithm

حالت ترکیبی بیشتر از برنامه‌ریزی بیان ژن به حالت ترکیبی است. در مورد بهینه‌سازی وزن‌های روش ترکیبی با الگوریتم ژنتیک، مقادیر وزن‌های حاصل از الگوریتم برای ضعیف‌ترین نوع مدل‌سازی یعنی رگرسیون خطی، کمترین مقدار و برای بهترین نوع مدل‌سازی تحقیق یعنی برنامه‌ریزی بیان ژن بیشترین مقدار را داشت. استفاده از سایر روش‌های ترکیبی در تحقیقات تحبیبی و همکاران (Habib *et al.*, 2016) و زمانی احمد محمودی و همکاران (Zamani Ahmad Mahmoodi *et al.*)

براساس شکل ۵، تمامی معیارهای خطا در حالت ترکیبی با استفاده از الگوریتم ژنتیک نسبت به برنامه‌ریزی بیان ژن و زمین‌آمار کاهش داشتند، به‌عنوان نمونه میزان کاهش MSE، RMSE، RRMSE، SI، AMAPE و UI در برآورد pH از برنامه‌ریزی بیان ژن به حالت ترکیبی به ترتیب برابر با ۲۳/۳۳، ۱۱، ۱۰/۷۶، ۴/۱۶، ۲/۸۵ و ۱۰ درصد و از زمین‌آمار به حالت ترکیبی برابر با ۴۸/۸۸، ۲۹/۲۴، ۲۸، ۱۴/۸۱، ۸/۹۴ و ۲۵ درصد است. همان‌طور که مشاهده می‌شود میزان درصد کاهش از زمین‌آمار به

ترکیبی) به دست آمد که نتایج در شکل ۶ آورده شده است (در شکل ۶، منظور از D1 اختلاف‌های انحراف معیار و ضریب کشیدگی و D2 اختلاف‌های میانگین و ضریب تغییرات است).

(al., 2014) منجر به افزایش دقت در برآورد توزیع مکانی داده‌ها شد. در بررسی دیگر جهت بررسی عملکرد روش ترکیبی با الگوریتم ژنتیک، مقادیر برخی ویژگی‌های آماری pH خاک در ۹ نقطه محاسبه و قدرمطلق اختلاف بین مقادیر اندازه‌گیری و تخمینی (با برنامه‌ریزی بیان ژن و

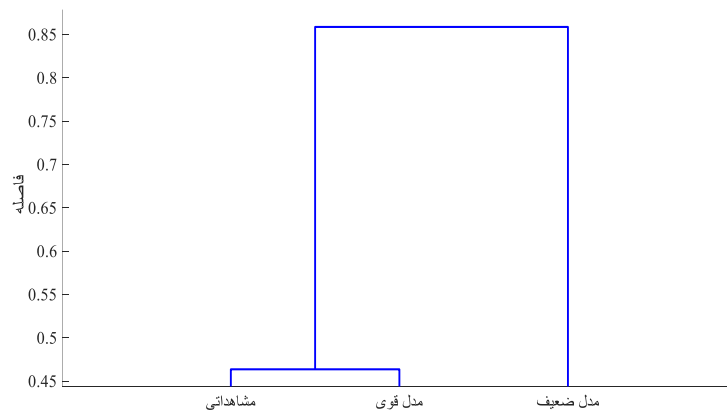


شکل ۶- مقایسه خصوصیات آماری داده‌های اندازه‌گیری شده با داده‌های حاصل از برنامه‌ریزی بیان ژن و ترکیبی

Figure 6. Comparison the statistical properties of measured data with data of gene expression programming

ترکیبی محاسبه شود، در بازه ۰ تا ۱۰ درصد قرار دارد که حاکی از پیش‌بینی دقیق روش ترکیبی است. مقادیر ضریب تغییرات داده‌های پهاش خاک با روش ترکیبی در محدوده پایین قرار گرفته‌اند که با محدوده مربوط به داده‌های اندازه‌گیری شده همخوانی دارد. در حالت دیگر جهت بررسی عملکرد روش ترکیبی از ترسیم نمودار درختی خوشه‌بندی استفاده شد به طوری که در این نمودار، داده‌های اندازه‌گیری شده (۱) با بهترین (۳) و بدترین (۲) نوع مدل‌سازی این تحقیق مقایسه شدند که نتایج در شکل ۷ آورده شده است.

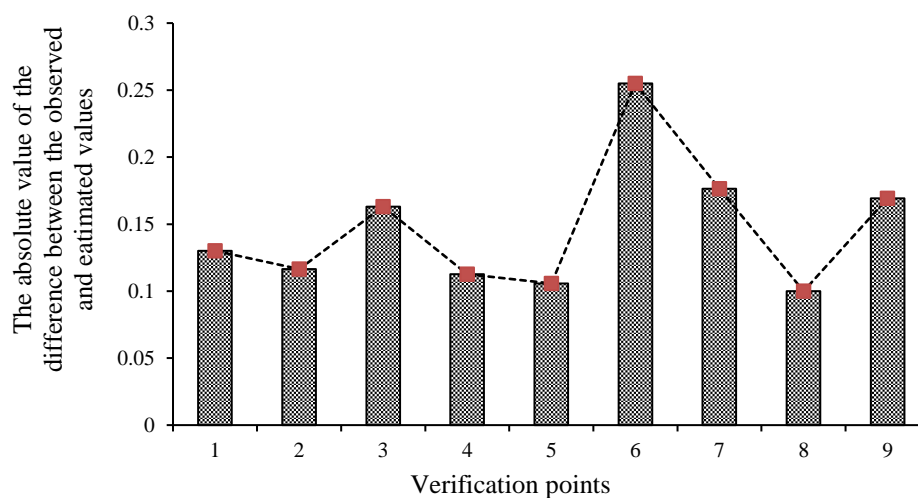
بر اساس شکل ۶، کمینه مقدار اختلاف بین خصوصیات آماری اندازه‌گیری و تخمینی در pH خاک مربوط به حالت ترکیبی بوده است. میزان درصد کاهش خطای میانگین، انحراف معیار، ضریب تغییرات و ضریب کشیدگی از برنامه‌ریزی بیان ژن به حالت ترکیبی برابر با ۸۳/۰۵، ۳۷/۵، ۳۶/۸۴ و ۱۲/۵ درصد محاسبه شد. همچنین برآوردهای مربوط به پهاش خاک با روش ترکیبی در محدوده قلیایی قرار دارد که در مورد داده‌های اندازه‌گیری شده نیز به این صورت است. اگر معیار میانگین درصد خطای مطلق برای داده‌های پهاش خاک حاصل از روش



شکل ۷- نمودار درختی خوشه‌بندی داده‌های اندازه‌گیری شده و تخمینی پهاش خاک

Figure 7. The dendrogram of measured and estimated soil pH data

قابل قبول روش ترکیبی است. در حالت دیگر قدرمطلق اختلاف بین مقادیر اندازه‌گیری و تخمینی خاک در ۹ نقطه صحت‌سنجی محاسبه شد و نتایج در شکل ۸ آورده شده است.



شکل ۸- بررسی عملکرد روش ترکیبی در برآورد توزیع مکانی پ‌هایش خاک در هر کدام از نقاط صحت‌سنجی
Figure 8. Investigation the performance of the combination method for soil pH spatial distribution estimation at each validation points

است که در این تحقیق با استفاده از الگوریتم ژنتیک سعی در ترکیب برآورد مدل‌سازی‌های مختلف است. در بررسی توزیع مکانی پ‌هایش خاک از مدل‌سازی رگرسیون خطی، برنامه‌ریزی بیان ژن و زمین‌آمار (کریجینگ) استفاده شد. براساس معیارهای خطای محاسبه‌شده، برنامه‌ریزی بیان ژن دارای کمترین میزان خطا بود. در برنامه‌ریزی بیان ژن نمایش کروموزوم‌ها به صورت بیان درختی در افزایش کارایی مدل‌سازی مفید است. به کارگیری الگوریتم ژنتیک جهت تعیین وزن‌های بهینه اختصاصی به مدل‌سازی‌های مورد استفاده، دقت توزیع مکانی pH خاک را افزایش داد. نتایج بهینه‌سازی حاکی از بیشترین مقدار وزن اختصاصی به مدل برنامه‌ریزی بیان ژن بود. در واقع الگوریتم ژنتیک با تعیین وزن‌های بهینه، اطلاعات مربوط به برآوردهای توزیع مکانی داده را از مدل‌سازی‌های موجود استخراج می‌کند و هرچه چقدر برآورد وزن‌ها دقیق‌تر باشد، استخراج اطلاعات با قدرت کافی خواهد بود. مقادیر ضریب تغییرات pH خاک در محدوده پایینی قرار دارد که این مسئله می‌تواند از مواد مادری تأثیرپذیر باشد. در بررسی توزیع مکانی پ‌هایش خاک، بیشترین و کمترین مقدار خطا به ترتیب مربوط به عجب‌شیر و بناب بود.

براساس شکل ۷ دو خوشه مجزا دیده می‌شود که در خوشه اول داده‌های اندازه‌گیری شده و تخمینی حاصل از روش ترکیبی قرار دارد و در خوشه دوم داده‌های تخمینی مدل‌سازی رگرسیونی قرار دارد. قرارگیری داده‌های اندازه‌گیری شده و ترکیبی در یک خوشه حاکی از عملکرد

براساس شکل ۸، کمینه و بیشینه مقدار قدر مطلق اختلاف بین مقادیر اندازه‌گیری شده و تخمینی به ترتیب در نقطه ۸ (دریناب) و ۶ (در عجب‌شیر) بود. کمترین مقدار پ‌هایش خاک در داده‌های اندازه‌گیری شده مربوط به نقطه ۷ بود که این مسئله در داده‌های حاصل از روش ترکیبی نیز حفظ شده است. نقاط ۲، ۶، ۷، ۹ نسبت به مقادیر اندازه‌گیری شده دارای بیش برآورد بودند. اگر مقایسه عملکرد روش ترکیبی نقاط براساس منطقه انجام گیرد، بیشترین مقدار RMSE و RRMSE مربوط به عجب‌شیر و کمترین مربوط به بناب بوده است. ضریب تغییرات مربوط به داده‌های اندازه‌گیری مراغه نسبت به دو منطقه دیگر بیشتر بود که این روند در داده‌های مربوط به روش ترکیبی هم حفظ شده است.

نتیجه‌گیری کلی

ارزیابی دقیق تغییرات مکانی خصوصیات خاک تأثیر چشمگیری بر اکوسیستم کشاورزی، مدل‌سازی محیطی و مدیریت پایدار دارد. در این راستا سعی در افزایش کارایی در زمینه مدل‌سازی توزیع مکانی ویژگی‌های خاک

افزایش دقت در زمینه مدل‌سازی مدل‌ها در این قسمت باید به دو نکته اشاره کرد ۱-۳- نوع مدل مورد استفاده مانند استفاده از سایر رویکردهای زمین‌آماری یا مدل‌های دیگر مانند رگرسیون بردار پشتیبان ۲-۳- تحلیل حساسیت دقیق مدل‌های مورد استفاده ۴- استفاده از روشی مناسب در روند بهینه‌سازی جهت تعیین دقیق وزن‌های هر مدل. به‌طور کلی روش ترکیبی با الگوریتم ژنتیک از کارایی قابل قبولی در برآورد توزیع مکانی پ-هاش خاک برخوردار بود.

توجه به این‌که روند مدل‌سازی یکسان است، این مسئله تأثیر سایر عوامل را در این تغییرات نشان می‌دهد. مقادیر معیار میانگین درصد خطای مطلق در روش ترکیبی در محدوده قابل قبول بود که قابلیت روش ترکیبی را در برآورد توزیع مکانی pH خاک نشان می‌دهد. برخی از مسائل می‌توانند روند مربوط به روش ترکیبی را تحت تأثیر قرار دهد که عبارتند از ۱- افزایش تعداد مدل‌های مورد استفاده در برآورد توزیع مکانی داده‌ها ۲- استفاده از تعداد زیادی متغیر مستقل در مدل‌های بکار رفته ۳-

Reference

- Adhikari R., and Agrawal R.K. 2014. Performance evaluation of weights selection schemes for linear combination of multiple forecasts. *Artificial Intelligence Review*, 42(4):529-48.
- Afzali A., Varvani J., and Jafania R. 2018. Application of Geostatistical analysis for prediction of Spatial Variations in Soil Texture (Case study: Farahan Plain, Markazi province). *Sarzamin*, 15(58):1-16. (In Persian)
- Ahmadi F.F., and Layegh N.F. 2015. Integration of artificial neural network and geographical information system for intelligent assessment of land suitability for the cultivation of a selected crop. *Neural Computing and Applications*, 26(6):1311-20.
- Arbab M., Pahlavanravi A., Piri Sahragard H., and Amiri M. 2018. Accuracy Assessment of Geo-Statistics and Artificial Neural Networks Methods to Estimate Threshold Wind Velocity: A Case of Jazinak Region, Sistan Plain. *Quarterly Journal of Environmental Erosion Research*, 31(3):87-105. (In Persian)
- Bayat Varkeshi M., and Fasihi R. 2018. Comparison of Numerical Model, Neural intelligent and geostatistical in Estimating Groundwater Table. *Journal of Geographical Sciences*, 18(48):165-182. (In Persian)
- Belkhir L., Tiri A., and Mouni L. 2020. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. *Groundwater for Sustainable Development*, 11:100473.
- Calzolari C., Ungaro F., and Vacca A. 2021. Effectiveness of a soil mapping geomatic approach to predict the spatial distribution of soil types and their properties. *Catena*, 196:104818.
- Chen W., Xu H., Chen Z., and Jiang M. 2021. A novel method for time series prediction based on error decomposition and nonlinear combination of forecasters. *Neurocomputing*, 426:85-103.
- Dahiya I.S., Richter J., and Malik R.S. 1984. Soil spatial variability: a review, *International Journal of Tropical Agricultural*, 11:1-102.
- Dehghani R., and Noorali A. 2016. Comparison of geo-statistical methods and artificial neural network in estimating groundwater level (Case Study: Nourabad Plain, Lorestan). *Journal of Environmental Science and Technology*, 18(1):31-43. (In Persian)
- Del Carmen Bas M., Ortiz J., Ballesteros L., and Martorell S. 2017. Forecasting 7BE concentrations in surface air using time series analysis. *Atmospheric Environment*, 155:154-61.
- Ebadi Y., Javdan J., and Rezaei Moghaddam M.H. 2019. Estimating accuracy of Artificial Neural Networks (ANN) and Geo Statistical Methods in interpolating Groundwater Levels Case study: Shabestar-Sufian Plain. *Scientific-Research Quarterly of Geographical Data*, 28(110):133-145. (In Persian)
- Eleyedath A., Kar S.S., and Swamy A.K. 2021. Modelling of expansion ratio and half-life of foamed bitumen using gene expression programming. *International Journal of Pavement Engineering*, 22(3):369-81.
- Foroughifar H., Jafarzadah A.A., Torabi Gelsefidi H., Aliasgharzadah N., Toomanian N., and Davatgar N. 2010. Spatial variations of surface soil pH physical and chemical properties on different landforms of Tabriz plain. *Journal of Soil and Water Science*, 21(3):1-21. (In Persian)

- Gia Pham T., Kappas M., Van Huynh C., and Nguyen L. 2019. Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of Central Vietnam. *ISPRS International Journal of Geo-Information*, 8(3):147.
- Habibi M.H., Nadiri A.A., Asghari Moghaddam A., and Naderi K. 2016. Combination of geostatistical and artificial intelligence methods for predicting spatiotemporal water level in the Hadishahr plain. *Iran-Watershed Management Science & Engineering*, 10(32):27-33. (In Persian)
- Hashemi M., Gholamalizadeh A., Bameri A., Sarani F., and Hejazizadeh A. 2016. Survey and zoning of soil physical and chemical properties using geostatistical methods in GIS (Case Study: Miankangi Region in Sistan). *Journal of Water and Soil*, 30(2):443-458. (In Persian)
- John K., Afu S.M., Isong I.A., Aki E.E., Kebonye N.M., Ayito E.O., Chapman P.A., Eyong M.O., and Penížek V. 2021. Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistics. *International Journal of Environmental Science and Technology*, 4:1-6.
- Karamouz M., and Kerachian R. 2003. Water Quality Planning and Management. *Amirkabir University of Technology*.
- Luan F., Si H.Z., Liu H.T., Wen Y.Y., and Zhang X.Y. 2008. Prediction of atmospheric degradation data for POPs by gene expression programming. *SAR and QSAR in Environmental Research*, 19(5-6):465-79.
- Maghami Y., Ghazavi R., Vali A.A., and Sharafi S. 2011. Evaluation of spatial interpolation methods for water quality zoning using GIS Case study, Abadeh Township. *Geography and Environmental Planning Journal*, 42(2):171-182. (In Persian)
- Mirzaee S., Asadzadeh F., and Zolphaghari A.A. 2017. Suitable geostatistical methods and incorporation of them with pedotransfer functions for mapping cation exchange capacity. *Applied Soil Research Journal*. 4(2): 1-13. (In Persian)
- Nadiri A., Naderi K., Asghari Moghaddam A., and Habibi M.H. 2017. Spatiotemporal Predicting of Groundwater Level using Artificial Intelligence Models and Geostatistics Model (Case Study: Duzduzan Plain). *Journal of Geography and Planning*, 20(58):218-301. (In Persian)
- Nadiri A., SHakor S., Asghari Moghaddam A., and Vadie M. 2014. Investigation of Groundwater Nitrate Pollution with Different Interpolation Methods (Case Study: East Azarbayjan, Bilverdy Plain). *HydroGemorphology*, 1(1):75-92. (In Persian)
- Pahlavan-Rad M.R., and Akbarimoghaddam A. 2018. Spatial variability of soil texture fractions and ϕ_{in} a flood plain (case study from eastern Iran). *Catena*, 160:275-81.
- Shabani H., Delavar M.A., Safari Y., and Alamdari P. 2020. Spatial Variability of Some Soil Characteristics in Lands of Zanjan University. *Applied Soil Research Journal*. 7(4): 164-178.
- Shi W., Liu Y., Kong W., and Shen Q. 2015. Tea classification by near infrared spectroscopy with projection discriminant analysis and gene expression programming. *Analytical Letters*, 48(18):2833-42.
- Shit P.K., Bhunia G.S., and Maiti R. 2016. Spatial analysis of soil properties using GIS based geostatistics models. *Modeling Earth Systems and Environment*, 2(2):1-6.
- Zamani Ahmad Mahmoodi R., Akhondali A.M., Zarei H., and Radmanesh F. 2014. Estimation of the groundwater level by using a combined optimized method with Genetic Algorithms in Ramhormoz plain. *Journal of Irrigation and Water Engineering*, 4(15):26-38. (In Persian)
- Zaynoddin M., Bonakdari H., Azari A., Ebtehaj I., Gharabaghi B., and Riahi Madavar H. 2018. Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. *Journal Environmental and Management*, 222:190-206.
- Zhang Y., and Hartemink AE. 2021. Quantifying short-range variation of soil texture and total carbon of a 330-ha farm. *Catena*, 201:105200.

Genetic Algorithm Application for Soil pH Spatial Distribution Estimation with Geostatistics and Gene Expression Programming Combination

Laleh Parviz^{1*}, Reza Ahadi², Fatemeh Seyed Milani³

(Received: August 2021 Accepted: October 2021)

Abstract

Awareness about spatial distribution of soil properties plays an important role in the system of crop growth and yield in the agricultural field. In this regard, genetic algorithm was used to increase the accuracy in spatial distribution of soil pH which combines the estimates of linear regression, gene expression programming (GEP) and geostatistics (kriging interpolation) with data related to some parts of East Azerbaijan province (Bonab, Maragheh and Ajabshir). The linear model of kriging interpolation had the minimum error. Gene expression programming and linear regression had the lowest and highest error for soil pH spatial distribution estimation, respectively, for example the mean square error, root mean square error, relative root mean square error decreasing from linear regression to geostatistics was 47.67%, 27.58%, 26.47% and scatter index and adopted mean absolute percentage error decreasing from linear regression to GEP was 23.8% and 37.03%, respectively. The use of genetic algorithm in combination method reduced the error of spatial distribution compared to the tree types of models, for example mean square error, root mean square error, relative root mean square error decreasing from GEP to combination method was 23.33%, 11.76%, 10%, respectively. The estimates of soil pH with combination method are in the alkaline range, which is consistent with the obtained data. The minimum and maximum value of the absolute difference between measured and estimated data were at point in Bonab (0.09) and Ajabshir (0.25), respectively. The mean absolute percentage error of combination method was in acceptable range and this shows the efficiency of combination method for soil pH spatial distribution estimation.

Key words: Soil properties, Interpolation, Error, Combination

Parviz L., Ahadi R., and Seyed Milani F. 2022. Genetic algorithm application for soil pH spatial distribution estimation with geostatistics and gene expression programming combination. *Applied Soil Research*, 10(3):.117-129.

1. Associate Professor, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran

2. Former BSc Student, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran

* Corresponding Author Email: laleh_parviz@yahoo.com