



## Modelling distribution of available phosphorous contents in surface soils of northern Khuzestan Province using linear and random forest models

Saeid Hojati<sup>1✉</sup>, Mojtaba Norouzi Masir<sup>2</sup>, Asim Biswas<sup>3</sup>

1. Corresponding Author, Department of Soil Science, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Khuzestan, Iran. Email: [s.hojati@scu.ac.ir](mailto:s.hojati@scu.ac.ir)

2. Department of Soil Science, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Khuzestan, Iran. Email: [mnm.scu@gmail.com](mailto:mnm.scu@gmail.com)

3. School of Environmental Science, University of Guelph, Ontario, Canada. Email: [biswas@uoguelph.ca](mailto:biswas@uoguelph.ca)

---

---

### Article Info

**Article type:** Research Article

#### Article history:

**Received:** May. 9, 2022

**Revised:** July. 24, 2022

**Accepted:** July. 30, 2022

**Published online:** Aug. 23, 2022

#### Keywords:

Conditioned Latin Hypercube,  
Prediction,  
Validation,  
Organic Carbon Content.

### ABSTRACT

There is little information about the spatial distribution of elements involved in assessing the fertility of soils in Khuzestan province, especially the available phosphorus contents of the soils. Therefore, this study conducted to determine the most effective soil properties controlling the concentration of available phosphorus contents of soils in the north of Khuzestan province and to introduce the most appropriate method for modeling the spatial distribution of available phosphorus contents of the soils analyzed using linear regression and random forest algorithm. For this purpose, 250 composite soil samples (0-10 cm depth) were randomly collected using the Conditional Latin Hypercube sampling approach from December 2020 to February 2021. Then, the physical and chemical properties of the samples were determined using standard laboratory methods. The experimental data were then analyzed for descriptive statistics using SPSS software. To model the spatial variability of available phosphorus contents of the soils, the experimental data were modeled using linear regression and random forest models in RStudio software. The results showed that according to the measured amounts of absorbable phosphorus in the soil samples in 32.4% of the samples, the concentration of available phosphorus is less than 5 mg/kg. Evaluation of multiple linear regression and random forest models based on model evaluation metrics including mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination ( $R^2$ ) using training, test and the whole dataset, showed that the random forest model provides better and more accurate estimates due to higher coefficients of determination as well as lower error values. The results also illustrated that the organic carbon content of the soils has the greatest contribution in the study area to predict available contents of soil phosphorus. In conclusion, models that include non-linear relationships between variables seem to be more suitable in predicting soil properties.

---

Cite this article: Hojati, S., Norouzi Masir, & M., Biswas, A. (2022) Modelling distribution of available phosphorous contents in surface soils of northern Khuzestan Province using linear and random forest models. *Iranian Journal of Soil and Water Research*, 53 (6), 1331-1344.

© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/ijswr.2022.342792.669262>





## مدل سازی پراکنش فسفر قابل جذب در خاک‌های سطحی شمال استان خوزستان با استفاده از مدل رگرسیون خطی چندگانه و الگوریتم جنگل تصادفی

سعید حجتی<sup>۱\*</sup>، مجتبی نوروزی مصیر<sup>۱</sup>، آسیم بیسواس<sup>۲</sup>

۱. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، خوزستان، ایران. ایمیل: [s.hojati@scu.ac.ir](mailto:s.hojati@scu.ac.ir)

۲. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، خوزستان، ایران. ایمیل: [nmm.scu@gmail.com](mailto:nmm.scu@gmail.com)

۳. گروه محیط زیست، دانشگاه گوئلف، آنتاریو، کانادا. ایمیل: [biswas@uoguelph.ca](mailto:biswas@uoguelph.ca)

### چکیده

### اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۱/۲/۱۹

تاریخ بازنگری: ۱۴۰۱/۵/۲

تاریخ پذیرش: ۱۴۰۱/۵/۸

تاریخ انتشار: ۱۴۰۱/۶/۱

### واژه‌های کلیدی:

مکعب لاتین شرطی،

پیش‌بینی،

اعتبارسنجی،

محتوای کربن آلی.

اطلاعات اندکی در رابطه با پراکنش مکانی عناصر دخیل در ارزیابی وضعیت حاصلخیزی خاک‌های استان خوزستان و به‌ویژه فسفر قابل جذب خاک‌ها وجود دارد. از این رو، مطالعه حاضر با هدف تعیین مؤثرترین ویژگی‌های خاکی کنترل‌کننده غلظت فسفر قابل جذب در خاک‌های شمال استان خوزستان و معرفی مناسب‌ترین روش مدل‌سازی تغییرات مکانی فسفر قابل جذب با استفاده از روش‌های رگرسیون خطی چندگانه و الگوریتم جنگل تصادفی انجام شد. بدین منظور تعداد ۲۵۰ نمونه مرکب صورت تصادفی به روش ابرمکعب لاتین مشروط در دی‌ماه تا بهمن‌ماه ۱۳۹۹ از خاک‌های سطحی (۱۰-۰ سانتی‌متری) اراضی شمال استان خوزستان تهیه شد. پس از آماده‌سازی‌های اولیه نمونه‌ها، ویژگی‌های فیزیکی و شیمیایی آن‌ها با استفاده از روش‌های استاندارد آزمایشگاهی اندازه‌گیری شد. سپس به‌منظور مدل‌سازی تغییرات مکانی فسفر قابل جذب خاک‌ها، داده‌های آزمایشی پس از آماده‌سازی‌های اولیه با استفاده از مدل‌های رگرسیون خطی چندگانه و جنگل تصادفی در محیط نرم‌افزار RStudio بررسی شدند. نتایج نشان داد که در ۳۲/۴ درصد از نمونه‌های مورد بررسی غلظت فسفر قابل جذب کمتر از ۵ میلی‌گرم بر کیلوگرم است. نتایج ارزیابی مدل‌های رگرسیون خطی چندگانه و جنگل تصادفی براساس آماره‌های ارزیابی مدل شامل میانگین خطای مطلق (MAE)، جذر میانگین مربعات خطا (RMSE) و ضریب تبیین ( $R^2$ ) در هر سه مرحله‌ی آموزش، آزمون و کل داده‌ها نشان داد که مدل جنگل تصادفی با توجه به ضرایب تبیین بالاتر و همچنین مقادیر خطای کمتر، تخمین‌های بهتر و دقیق‌تری را ارائه می‌دهد. بررسی اهمیت متغیرهای خاکی در مدل‌سازی پراکنش فسفر قابل جذب در هر دو روش رگرسیون خطی چندگانه و جنگل تصادفی نشان داد که محتوای کربن آلی خاک‌ها بیشترین نقش را در توزیع فسفر قابل جذب در منطقه مطالعاتی داراست. در کل به نظر می‌رسد استفاده از مدل‌هایی که روابط غیرخطی بین متغیرها را نیز لحاظ می‌نمایند در پیش‌بینی خصوصیات خاک‌ها بهتر است.

استناد: حجتی، سعید؛ نوروزی مصیر، مجتبی؛ بیسواس، آسیم (۱۴۰۱). مدل‌سازی پراکنش فسفر قابل جذب در خاک‌های سطحی شمال استان خوزستان با استفاده از مدل

رگرسیون خطی چندگانه و الگوریتم جنگل تصادفی. مجله تحقیقات آب و خاک ایران، ۵۳ (۶)، ۱۳۴۴-۱۳۳۱.

DOI: <http://doi.org/10.22059/ijswr.2022.342792.669262>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

## مقدمه

فسفر یکی از عناصر پر مصرف مورد نیاز گیاه بوده و از نظر کمی پس از نیتروژن رتبه‌ی دوم را به خود اختصاص می‌دهد. غلظت فسفر کل خاک‌های زراعی از ۰/۱ تا ۳ گرم در کیلوگرم متغیر است (Souri, 2018). فسفر محلول مورد نیاز برای اکثر محصولات کشاورزی در خاک بین ۰/۲ تا ۰/۳ میلی‌گرم در کیلوگرم می‌باشد که در این غلظت گیاهان قادر هستند بیش‌ترین محصول را تولید نمایند (Riemersma et al., 2006). با این حال، فراهمی فسفر اغلب به‌عنوان یک عامل محدود کننده برای تولید فرآورده‌های کشاورزی در سراسر جهان مطرح است. به‌گونه‌ای که برآوردها حاکی از آن است که کمبود فسفر در ۴۲ درصد از زمین‌های کشت شده در سراسر جهان بسیار شایع می‌باشد. این در حالی است که مطالعات پیشین در شمال استان خوزستان نشان داده‌اند که حدود ۷۷ درصد از این اراضی به کود فسفره نیاز دارند (Mirzashahi et al., 2014).

مدل‌سازی فرآیند ایجاد نمای انتزاعی یک موضوع است. به بیان دیگر، مدل نمادی از واقعیت است که مهم‌ترین ویژگی‌های دنیای واقعی را به‌صورت ساده و کلی بیان می‌کند. مدل‌سازی با قابلیت پیش‌بینی (که معمولاً از آن به‌عنوان مدل‌سازی آماری یا یادگیری ماشین یاد می‌شود) برای نقشه‌برداری رقومی خاک، بیانگر فرآیند تهیه یک مدل ریاضی است که رابطه واقعی بین ویژگی‌ها و/یا کلاس‌های خاک را به‌منظور ارائه‌ی یک پیش‌بینی دقیق تخمین می‌زند. این فرآیند شامل دو مرحله است:

الف- انتخاب داده‌هایی که به‌عنوان پیش‌بینی کننده‌ی مدل بتوانند نماینده‌ی خوبی از متغیر هدف باشند،

ب- انتخاب یک مدل یا الگوریتم مناسب برای پیش‌بینی.

یکی از مدل‌های رایج در نقشه‌برداری رقومی خاک‌ها، مدل جنگل تصادفی<sup>۱</sup> است. مدل جنگل تصادفی در واقع یک الگوریتم یادگیری ترکیبی<sup>۲</sup> نظارت شده است که از آن هم برای طبقه‌بندی (متغیرهای کیفی مانند کلاس‌های خاک) و هم رگرسیون (متغیرهای کمی مانند شوری) استفاده می‌شود. همان‌طور که از نام آن مشهود است، این الگوریتم جنگلی را به‌طور تصادفی می‌سازد. جنگل ساخته شده، در واقع گروهی از درخت‌های تصمیم (برای مثال ۱۰۰ درخت تصمیم) است. کار ساخت جنگل با استفاده از درخت‌ها اغلب اوقات به روش کیسه‌گذاری<sup>۳</sup> انجام می‌شود. ایده اصلی روش کیسه‌گذاری آن است که ترکیبی از مدل‌های یادگیری، نتایج کلی مدل را افزایش می‌دهند. به بیان ساده، جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند. در الگوریتم جنگل تصادفی به هر کدام از درخت‌ها، زیرمجموعه‌ای از داده‌ها تزیق می‌شود. برای مثال اگر مجموعه‌ی داده‌ی مورد استفاده دارای ۱۰۰۰ سطر (یعنی ۱۰۰۰ نمونه) و ۵۰ ستون (یعنی ۵۰ ویژگی) بود، الگوریتم جنگل تصادفی به هر کدام از درخت‌ها، مثلاً ۱۰۰ سطر و ۲۰ ستون اختصاص می‌دهد، که به‌صورت تصادفی انتخاب شده‌اند. الگوریتم جنگل تصادفی، به‌عنوان یک مدل نظارت شده، یاد می‌گیرد که در فاز آموزش<sup>۴</sup> یا برازش مدل، داده‌ها را (مانند مقادیر اندازه‌گیری شده‌ی قابلیت هدایت الکتریکی خاک، pH، بافت و ...) به خروجی(های) مورد نظر مرتبط کند تا مدل آموزش ببیند و بتواند مقدار صحیحی را برای متغیر مورد نظر پیش‌بینی کند (Breiman, 2001).

در رابطه با استفاده از این مدل در مطالعات خاک‌شناسی پژوهش‌های نسبتاً زیادی صورت گرفته است. نتایج بسیاری از این پژوهش‌ها نشان داده‌اند که مدل رگرسیونی جنگل تصادفی از پتانسیل زیادی برای بهینه‌سازی روابط غیرخطی بین خاک و زمین نما<sup>۵</sup> برخوردار است. (Jamshidi et al., 2019) نیز کارایی روش‌های رگرسیون خطی چندگانه، کوپست و جنگل تصادفی را برای پیش‌بینی تغییرات کربن آلی در سطح و عمق خاک اراضی دشت سعادت شهر مورد ارزیابی قرار دادند. موقعیت نقاط مطالعاتی بر اساس روش مربعات لاتین تعیین و مقدار کربن آلی نمونه‌های خاک در افق‌های ژنتیکی اندازه‌گیری شد. نتایج مطالعه‌ی آن‌ها نشان داده است که برخلاف اختلاف اندک این مدل‌ها، مدل جنگل تصادفی با توجه به قابلیتش در نشان دادن توزیع مکانی مقادیر کربن آلی خاک، برای تخمین کربن آلی در تمامی عمق‌ها انتخاب شده و نقشه‌های پراکنش کربن آلی در اعماق مختلف خاک دشت سعادت شهر در محیط سامانه اطلاعات جغرافیایی (GIS) تهیه شده است. (Mousavi et al., 2020) در مطالعه‌ای با نمونه‌برداری سطحی (۱۰-۰ سانتی‌متری) از ۱۳۱ نقطه در کرانه‌ی شرقی دریاچه

1- Random Forest  
2- Ensemble  
3- Bagging  
4- Training  
5- Landscape



ارومیه و بهره‌گیری از دو مدل جنگل تصادفی و کوبیست، نقشه پراکنش کربن آلی را تهیه نموده‌اند. نتایج نشان داده است که مدل جنگل تصادفی با مقادیر ضریب تبیین  $0/89$  و جذر میانگین مربعات خطای  $0/16$  در مقایسه با مدل کوبیست ( $R^2 = 0/85$  و  $RMSE = 0/21$ ) کارایی و صحت بیشتری در پیش‌بینی کربن آلی خاک در منطقه دارد. (Pahlevan-Rad et al., 2016) به‌منظور نقشه‌برداری رقومی خاک ۸۵۰۰۰ هکتار از اراضی استان گلستان از دو مدل جنگل تصادفی و رگرسیون لجستیک چندگانه استفاده کرده‌اند. بر اساس نتایج به دست آمده مدل جنگل تصادفی نسبت به مدل رگرسیون لجستیک چند متغیره توانایی بالاتری در پیش‌بینی سری‌های خاک منطقه داشته است. Camera et al., (2017) با استفاده از دو مدل رگرسیون لجستیک چندگانه و جنگل تصادفی اقدام به تهیه نقشه رقومی کلاس‌های خاک در کشور قبرس نموده و گزارش دادند که مدل جنگل تصادفی نسبت به مدل رگرسیون لجستیک چندگانه از توانایی بالاتری برای پیش‌بینی کلاس‌های خاک منطقه برخوردار بوده و همچنین این پیش‌بینی با خطای کمتری نسبت به مدل دیگر همراه بوده است.

اگرچه مطالعات زیادی در رابطه با استفاده از مدل‌های مختلف در پیش‌بینی ویژگی‌های خاک‌ها در مناطق مختلف دنیا و البته ایران صورت گرفته است، لیکن بررسی‌های انجام شده توسط نویسندگان مقاله‌ی حاضر، حاکی از آن است که در رابطه با پیش‌بینی ویژگی‌های خاک به‌ویژه متغیرهای مرتبط با ارزیابی توان حاصلخیزی خاک‌ها در استان خوزستان که به‌دلیل وجود منابع آبی و خاک حاصلخیز پتانسیل زیادی برای تولید محصولات متنوع کشاورزی دارد، اطلاعات محدودی در دست است. بر این اساس، مطالعه‌ی حاضر با هدف تشخیص مؤثرترین ویژگی‌های خاکی کنترل‌کننده غلظت فسفر قابل جذب در خاک‌های استان و معرفی مناسب‌ترین روش مدل‌سازی تغییرات مکانی فسفر قابل جذب با استفاده از برخی روش‌های داده‌کاوی شامل رگرسیون خطی چندگانه و الگوریتم جنگل تصادفی انجام شد.

## مواد و روش‌ها

### معرفی منطقه‌ی مطالعاتی

محدوده‌ی مطالعاتی به وسعت  $437476$  کیومتر مربع در شمال استان خوزستان شامل شهرستان‌های شوش، دزفول، اندیمشک، ایذه، باغملک، رامهرمز، اندیکا، لالی، گتوند، مسجدسلیمان، شوشتر، باوی، اهواز و دشت آزادگان قرار دارد. در این منطقه انواع کاربری اراضی اعم از مرتع، جنگل و کشاورزی دیده می‌شود. اما عمده کاربری اراضی در استان خوزستان کشاورزی است که عمدتاً نیز بخش‌هایی از مرکز و غرب منطقه را به خود اختصاص می‌دهند. کشاورزی در این مناطق عمدتاً از نوع آبی است؛ حال آن‌که اراضی دیم بیشتر بخش‌های شرقی استان و عمدتاً دامنه‌های کوهستانی را به خود اختصاص می‌دهند (Statistical Center of Iran, 2013). استان خوزستان از نظر پستی و بلندی به دو منطقه‌ی کوهستانی و جلگه‌ای تقسیم می‌شود. منطقه جلگه‌ای خوزستان از جنوب دزفول، مسجدسلیمان، رامهرمز و بهبهان آغاز شده تا کرانه‌های خلیج فارس و اروند ادامه دارد. این جلگه تقریباً دارای شیب کم است و در برخی از نقاط آن گنبد‌های نمکی مربوط به دوران کامبرین وجود دارد که در شور کردن اراضی و آب‌های منطقه تأثیر عمده دارد. وسعت این منطقه تقریباً ۶۰ درصد کل استان می‌باشد. وجود رسوبات ریزدانه، حاکمیت فرسایش بیابانی، امکان تشکیل پهنه‌های ماسه‌ای را نیز در جلگه خوزستان فراهم کرده است. تراکم عمده این عوارض در غرب اهواز، شمال حمیدیه و سوسنگرد می‌باشد. بادهای غالب غربی و جنوب غربی که از سوی بین‌النهرین و عربستان می‌وزند را مسئول ایجاد تپه‌های ماسه‌ای اطراف می‌توان معرفی نمود. منطقه کوهستانی به‌طور عمده در شمال و شرق استان (رشته کوه زاگرس با روند شمال غربی - جنوب شرقی) حدود دو پنجم مساحت کل استان را شامل می‌شود. مهم‌ترین کوه‌های استان عبارتند از: کوه چو، کوه دلا، کوه منار، کوه آب‌بندان، کوه سیاه، کوه منگشت و کوه سلطان. ارتفاع در نقاط مختلف استان بسیار متغیر است؛ اما به‌طور متوسط از جنوب به شمال و از غرب به شرق افزایش می‌یابد. به‌گونه‌ای که تغییرات ارتفاعی در استان از صفر در سواحل خلیج فارس تا بیش از  $3400$  متر در ارتفاعات شمال و شمال شرقی می‌باشد (Statistical Center of Iran, 2013).

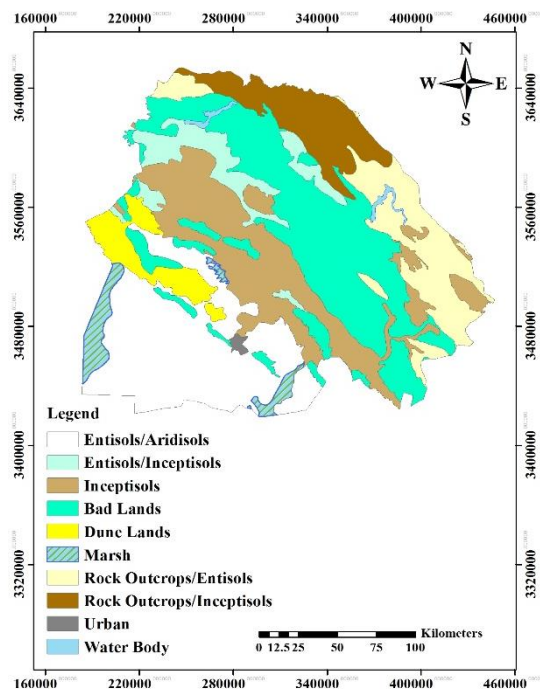
از لحاظ اقلیمی، میانگین بارندگی سالانه استان طی دوره آماری بلندمدت  $248/3$  میلی‌متر است. بیشترین میزان بارش سالانه مربوط به ایستگاه ایذه با  $614/8$  میلی‌متر و کمترین میزان بارش سالانه در آبادان با  $149/7$  میلی‌متر ثبت شده است. نتایج حاصل از بررسی ویژگی‌های دمايي طی دوره آماری بلند مدت در سطح استان خوزستان نشان می‌دهد، متوسط دمای سالیانه‌ی استان  $25/3$  درجه سانتیگراد است و بالاترین میانگین سالانه دما مربوط به ایستگاه‌های آغاچاری، شوشتر و رامهرمز به ترتیب  $27/6$ ،  $26/8$  و  $26/7$  درجه سانتی‌گراد و کمترین میانگین سالانه دما به ایستگاه دهدز با  $18/9$  درجه سانتی‌گراد تعلق دارد (Statistical Center of Iran, 2013). خاک‌های استان خوزستان بر اساس سامانه رده‌بندی خاک آمریکایی (۲۰۱۴) عمدتاً شامل رده‌های اریدی‌سولز، اینسپتی‌سولز و انتی‌سولز می‌باشند که با حرکت از بخش‌های غربی منطقه به سمت بخش‌های شرقی‌تر به‌واسطه تغییر ویژگی‌های اقلیمی مانند بارندگی بیشتر (رژیم‌های رطوبتی

زریک و یوستیک) و کاهش نسبی درجه حرارت (رژیم‌های حرارتی ترمیک و مزیک در مرز ترمیک) تکامل خاک‌ها افزایش یافته و رده‌بندی آن‌ها از انتی‌سولز به سمت اینسپتی‌سولز تغییر می‌یابد. همچنین، قسمت‌های غربی استان متأثر از فرآیندهای فرسایش بادی بوده و مقادیر نسبتاً قابل توجهی تپه‌های شنی در نزدیکی مرزهای ایران با عراق در حوالی شهرهای شوش و بستان به چشم می‌خورند (شکل ۱).

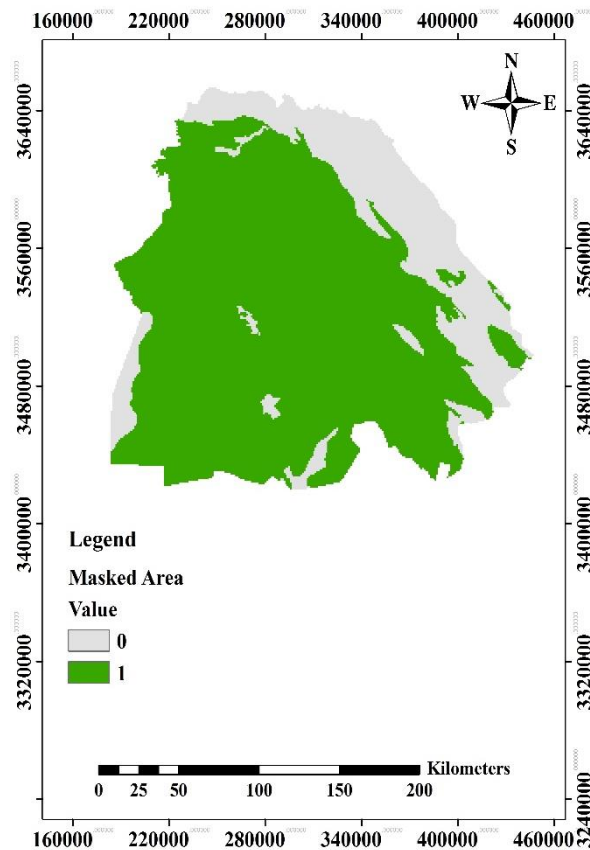
### فرآیند نمونه‌برداری

در این مطالعه تصاویر ماهواره‌ی لندست ۸، از پایگاه داده‌های مکانی اداره زمین‌شناسی آمریکا (earthexplorer.usgs.gov) تهیه شد. باندهای طیفی این ماهواره شامل ۹ باند OLI و ۲ باند TIRS می‌باشد. زمان برداشت تصاویر سنجنده به نحوی انتخاب شدند که زمان تصویر برداری ماهواره با زمان نمونه‌برداری خاک (دی ماه تا بهمن ماه ۱۳۹۹) نزدیک باشد. بعد از تهیه تصاویر، تمام تصحیحات مربوط به تصاویر ماهواره‌ای شامل تصحیح اتمسفری، رادیومتری و هندسی در محیط نرم‌افزار ENVI نسخه ۵/۳ بر روی کلیه باندها انجام شد. لازم به توضیح است که برای انجام تصحیحات هندسی، از نقشه‌های توپوگرافی با مقیاس ۱:۵۰۰۰۰ تهیه شده از سازمان جغرافیایی کشور استفاده گردید. تصحیح هندسی تصاویر با روی هم‌گذاری لایه‌های خطی (مانند رودخانه‌ها) بر روی تصویر تطابق یافته انجام شد و بدین ترتیب دقت بالای عمل تطابق هندسی مورد تأیید قرار گرفت. برای تصحیح رادیومتری و به منظور کاهش اثرات پخش اتمسفری از روش تصحیح رادیومتریکی نسبی (Savari et al., 2021) استفاده گردید.

تعداد ۲۵۰ نمونه از عمق صفر تا ۱۰ سانتیمتری (به منظور ایجاد ارتباط بهتر با لایه‌های اطلاعاتی تهیه شده از تصاویر ماهواره‌ی لندست ۸) خاک اراضی زراعی و مرتعی در شمال استان خوزستان با استفاده از روش نمونه‌برداری مکعب لاتین شرطی<sup>۱</sup> (cLHS) (Minasny and McBratney, 2006) و به صورت مرکب انجام شد؛ به گونه‌ای که در هر نقطه یک مربع یک در یک متری ایجاد و از رئوس و مرکز هر مربع نمونه‌هایی برداشت شد و سپس نمونه‌های خاک با یکدیگر مخلوط گردید. بدین منظور ابتدا در محیط نرم‌افزار ArcGIS اراضی بایر، مناطق شهری، تالاب‌ها و اراضی دارای رخنمون سنگی (فاقد خاک) ماسک شدند. به دیگر سخن، مناطق مزبور در تعیین نقاط نمونه‌برداری لحاظ نگردیدند (شکل ۲).



شکل ۱- نقشه تنوع خاک‌ها و پوشش‌های اراضی در مناطق شمالی استان خوزستان (Hojati, 2018)

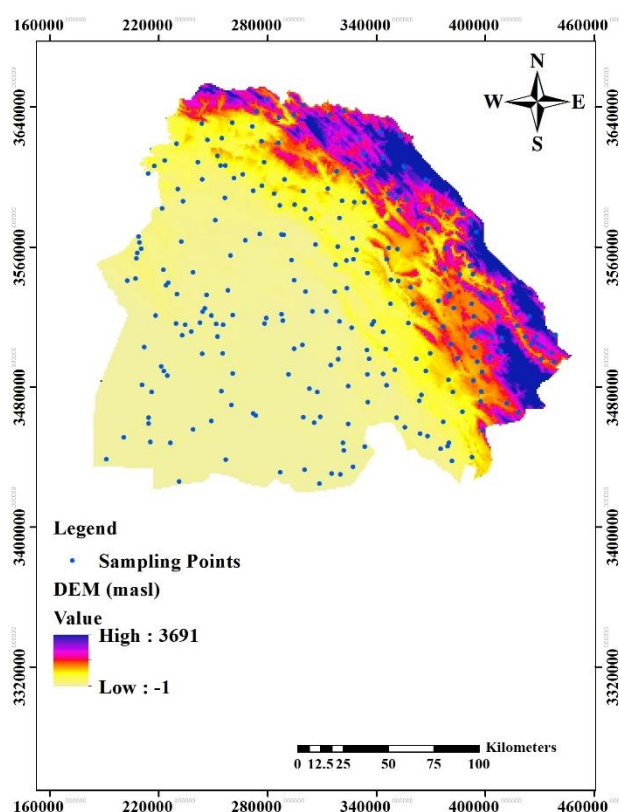


شکل ۲- نقشه مناطق ماسک شده (مناطق با رنگ روشن) در نیمه‌ی شمالی استان خوزستان. عدد صفر در راهنمای نقشه نشان دهنده مناطقی است که در نمونه‌برداری لحاظ نشده‌اند.

در روش نمونه‌برداری cLHS، تعداد نمونه توسط کاربر تصمیم‌گیری می‌شود؛ لیکن محل نقاط نمونه‌برداری توسط نرم‌افزار مشخص می‌گردد. این تکنیک، داده‌های متغیرهای محیطی را براساس تعداد نمونه‌های خواسته شده به زیرخوشه‌هایی تقسیم‌بندی می‌کند و سپس با استفاده از داده‌های متغیرهای محیطی در هر زیرخوشه، یک نمونه تصادفی را به‌نحوی انتخاب می‌کند که در هر بعد a و b فقط یک نمونه وجود داشته باشد (Minasny and McBratney, 2006).

در مطالعه حاضر از مدل رقومی ارتفاع (DEM)<sup>۱</sup> منطقه با قدرت تفکیک مکانی ۹۰ متر، شاخص پوشش گیاهی نرمال شده (NDVI)<sup>۲</sup> و شاخص شوری (SI)<sup>۳</sup>، نقشه خاک و نقشه زمین‌شناسی منطقه با مقیاس ۱:۲۰۰۰۰۰۰ به‌عنوان متغیرهای محیطی ورودی برای تعیین موقعیت نقاط نمونه‌برداری در روش cLHS استفاده شد. بدین منظور، از افزونه‌ی TEUI نسخه ۴۶۰ در محیط نرم‌افزار ArcGIS نسخه ۱۰/۳ استفاده شد. لازم به توضیح است برای استفاه از افزونه‌ی فوق در طراحی نقاط پیشنهادی برای نمونه‌برداری از خاک‌ها تمامی لایه‌های ورودی باید دارای ساختار شطرنجی<sup>۴</sup> بوده و اندازه پیکسل‌های مشابهی داشته باشند. در مرحله بعد، نقاط پیشنهادی به دستگاه موقعیت یاب جهانی (GPS)<sup>۵</sup> منتقل و در آن ذخیره گردید و طی پیمایش‌های صحرائی از خاک نقاط مورد نظر نمونه‌برداری صورت گرفت. لازم به توضیح است که در مواردی که نقاط پیشنهادی برای نمونه‌برداری مناسب تشخیص داده نشد (۳۲ نقطه) از نزدیک‌ترین نقطه به نقطه پیشنهادی نمونه‌برداری و موقعیت جغرافیایی نقاط جدید توسط دستگاه GPS ثبت گردید. شکل ۳ موقعیت نهایی نقاط نمونه‌برداری پیشنهادی توسط الگوریتم cLHS را نشان می‌دهد.

1- Digital Elevation Model  
 2- Normalized Difference Vegetation Index  
 3- Salinity Index  
 4- Raster format  
 5- Global Positioning System



شکل ۳- موقعیت نقاط نمونه‌برداری پیشنهادی توسط الگوریتم cLHS در شمال استان خوزستان

### اندازه‌گیری ویژگی‌های فیزیکی و شیمیایی خاک‌ها

پس از هوا خشک کردن نمونه‌های خاک برداشت شده و عبور از الک دو میلی‌متری، تجزیه‌های آزمایشگاهی لازم بر روی آن‌ها انجام گرفت. برای این منظور، فسفر قابل جذب به روش اولسن (Olsen *et al.*, 1954) و اکشن خاک در خمیر اشباع با دستگاه pH متر، قابلیت هدایت الکتریکی عصاره اشباع خاک با استفاده از دستگاه هدایت‌سنج (Rhoades, 1986)، بافت خاک به روش هیدرومتری (Gee and Bauder, 1986)، کربنات کلسیم معادل به روش تیتراسیون برگشتی (Loeppert and Suarez, 1996)، گچ به روش ترسیب با استون (Nelson, 1982) و کربن آلی به روش اکسیداسیون تر در مجاورت دی‌کرومات پتاسیم (Walkly and Black, 1934) اندازه‌گیری شدند.

### مدل‌سازی پراکنش فسفر قابل جذب در منطقه مطالعاتی

ابتدا داده‌های فسفر قابل جذب و سایر ویژگی‌های فیزیکی و شیمیایی خاک‌ها در مواردی که داده‌های آزمایشی فاقد توزیع نرمال بودند، با استفاده از فرآیند تبدیل داده‌ها نرمال شده و سپس تحلیل رگرسیون چند متغیره خطی بین فسفر قابل جذب با ویژگی‌ها خاک به روش رگرسیون خطی چندگانه از نوع گام به گام در محیط نرم‌افزار Rstudio نسخه ۱/۴/۱۱۰۶ صورت گرفت (R Core Team, 2021). لازم به توضیح است که برای بررسی نرمال بودن داده‌ها از آزمون آماری شاپیرو-ویلک استفاده شد. در این ارتباط، نوع تبدیل داده‌ها در رابطه با متغیرهای با توزیع غیرنرمال متفاوت بود. بر این اساس، در رابطه با کربنات کلسیم معادل و قابلیت هدایت الکتریکی خاک‌ها از لگاریتم‌گیری و در رابطه با واکنش خاک از روش نرمال‌سازی مینیمم-ماکزیمم استفاده شد. برای تعیین رابطه‌ی رگرسیونی ابتدا معنی‌داری ضرایب برای عرض از مبدأ و متغیرهای مستقل با استفاده از آزمون t تأییدگردید. لازم به توضیح است که از تمامی داده‌های آزمایشی، ۶۷ نمونه (حدوداً ۲۷ درصد کل نمونه‌ها) به طور تصادفی وارد مدل نشده و به منظور اعتبار سنجی مدل‌ها مورد استفاده قرار گرفت.

به‌منظور مدل‌سازی پراکنش فسفر قابل جذب خاک‌ها در منطقه‌ی مطالعاتی با استفاده از الگوریتم جنگل تصادفی، داده‌های به‌دست آمده از مطالعات آزمایشگاهی در محیط نرم‌افزار RStudio و پس از نصب و فراخوانی بسته‌ی RandomForest اجرا گردید (R Core Team, 2021). لازم به ذکر است که در الگوریتم جنگل تصادفی تعداد درختان جنگل "ntree" و تعداد متغیرهای کمکی در گره هر درخت "mtry" توسط کاربر مشخص می‌شود و مناسب‌ترین مقدار این دو پارامتر با روش سعی و خطا برای دست‌یابی به کم‌ترین مقدار خطا به دست می‌آید.

(Pahlevan-Rad *et al.*, 2015). در این پژوهش، دو مشخصه "mtry" و "ntree" با تکرار کردن مقادیر "mtry" از ۱ تا ۹ برابر با تعداد کل متغیرهای پیش‌بینی کننده و همچنین مقادیر "ntree" از ۱۰۰ تا ۱۰۰۰۰ با افزایش ۱۰۰ مرتبه‌ای بهینه سازی شدند (Hengl *et al.*, 2015; Mousavi *et al.*, 2020).

در استفاده از الگوریتم جنگل تصادفی نیز مانند مدل رگرسیون خطی ۶۷ نمونه در ایجاد هر درخت استفاده نگردید که تحت عنوان نمونه‌های بیرون از سبد<sup>۱</sup> (OOB) نام‌گذاری شدند. از آنجایی که این نمونه‌ها در آموزش درخت استفاده نشده‌اند، در فرآیند آزمون مدل مورد استفاده قرار می‌گیرند و بنابراین در این روش نیازی به اعتبارسنجی مستقل نیست.

#### اعتبارسنجی نتایج مدل‌های مورد استفاده

برای ارزیابی میزان توانایی مدل‌ها در پیش‌بینی مقدار فسفر قابل جذب از معیارهایی مانند میانگین مربعات خطا (MSE)، جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب تبیین ( $R^2$ ) استفاده گردید.

$$MAE = \frac{1}{n} \sum_{i=1}^n [Z^*(x_i) - Z(x_i)] \quad \text{رابطه (۱)}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_i))^2 \quad \text{رابطه (۲)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_i))^2} \quad \text{رابطه (۳)}$$

$$R^2 = \left[ \frac{\sum_{i=1}^n (Z(x_i) - Z(\text{ave})) (Z^*(x_i) - Z^*(\text{ave}))}{\sqrt{\sum_{i=1}^n (Z(x_i) - Z(\text{ave}))^2 \sum_{i=1}^n (Z^*(x_i) - Z^*(\text{ave}))^2}} \right]^2 \quad \text{رابطه (۴)}$$

در معادلات بالا،  $n$  تعداد نقاط مشاهده‌ای،  $Z^*(x_i)$  مقدار برآوردی بر نقطه  $x_i$  و  $Z(x_i)$  مقدار مشاهده‌ای برای نقطه  $x_i$  می‌باشند. مقادیر  $Z(\text{ave})$  و  $Z^*(\text{ave})$  نیز به ترتیب میانگین مقادیر مشاهده‌ای و مقادیر پیش‌بینی شده را نشان می‌دهند.

## نتایج و بحث

### نتایج توصیف آماری داده‌های آزمایشی

جدول ۱ توصیف آماری فسفر قابل جذب و ویژگی‌های فیزیکی و شیمیایی اندازه‌گیری شده در نمونه‌های خاک مورد بررسی را نشان می‌دهد. همان‌گونه که مشاهده می‌گردد با توجه به مقادیر pH و کربنات کلسیم معادل، خاک‌های مورد بررسی در زمره خاک‌های قلیایی و آهکی هستند. به علاوه، با توجه به مقادیر قابلیت هدایت الکتریکی اندازه‌گیری شده در عصاره اشباع خاک‌های مورد بررسی، این خاک‌ها در هر دو کلاس غیر شور (اراضی واقع در بخش‌های شرقی و شمال شرقی) و شور (مناطق مرکزی) طبقه‌بندی می‌شوند (Rhoades *et al.*, 1999). بر این اساس، ۳۸/۴ درصد از نمونه‌های مورد بررسی (۹۶ نمونه) در کلاس غیر شور (قابلیت هدایت الکتریکی کمتر از ۴ دسی زیمنس بر متر) و سایر نمونه‌های جمع‌آوری شده (۱۵۴ نمونه) در کلاس شور قرار دارند. همچنین، با توجه به فراوانی ذرات شن، سیلت و رس، خاک‌های مورد بررسی در کلاس‌های بافتی لومی (۴۰ درصد)، سیلتی رسی (۲۵ درصد)، رسی لومی سیلتی لوم (۱۵ درصد)، رسی (۱۰ درصد)، سیلتی رسی لومی (۵ درصد)، شنی لومی (۵ درصد) قرار دارند. با توجه به مقادیر اندازه‌گیری شده‌ی فسفر قابل جذب در نمونه خاک‌های مورد بررسی در ۳۲/۴ درصد (۸۱ نمونه) از نمونه‌های مورد بررسی غلظت فسفر قابل جذب کمتر از ۵ میلی‌گرم بر کیلوگرم است و از لحاظ میزان فسفر قابل جذب، این خاک‌ها فقیر محسوب می‌شوند که نیاز به تامین فسفر مورد نیاز محصولات زراعی از طریق کوددهی در آن‌ها ضروری به‌نظر می‌رسد. بر این اساس، غلظت فسفر قابل جذب در ۴۵/۲ درصد (۱۱۳ نمونه) از نمونه‌های مورد بررسی در محدوده‌ی غلظتی ۵ تا ۱۵ میلی‌گرم بر کیلوگرم است و در ۲۲/۴ درصد (۵۶ نمونه) از نمونه‌های مورد بررسی غلظت فسفر قابل جذب از ۱۵ میلی‌گرم



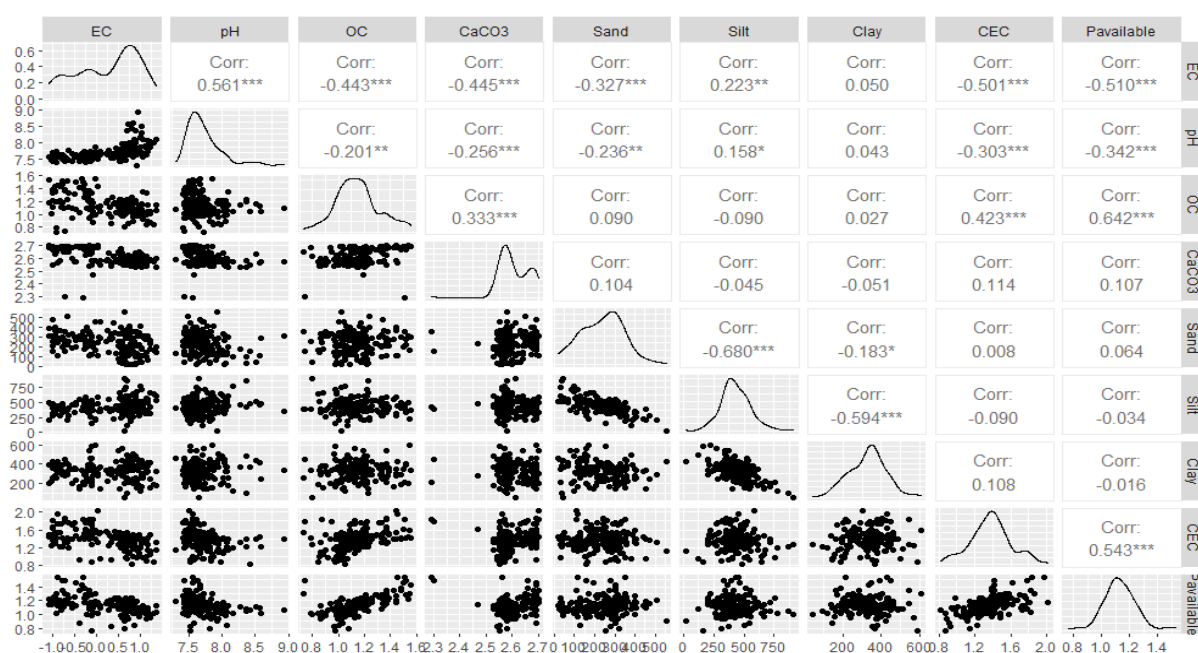
بر کیلوگرم بیشتر است و به نظر می‌رسد در این نمونه‌ها غلظت فسفر به گونه‌ای است که قادر به تامین نیاز گیاهان زراعی است.

جدول ۱- توصیف آماری ویژگی‌های فیزیکوشیمیایی اندازه‌گیری شده در خاک‌های سطحی جمع‌آوری شده از شمال استان خوزستان

Properties	Min	Max	Mean	Median	Skewness	Kurtosis	CV (%) <sup>a</sup>
Sand (g/Kg)	۲۰/۰	۵۵۵/۰	۲۳۷/۱	۲۵۰/۰	۰/۴۸	۱/۵۵	۴۵/۲
Silt (g/Kg)	۲۰/۰	۹۱۰/۰	۴۳۴/۲	۴۱۵/۰	-۰/۰۶	۰/۸۲	۳۰/۲
Clay (g/Kg)	۴۵/۰	۶۰۵/۰	۳۲۸/۱	۳۳۵/۰	-۰/۰۳	۰/۲۱	۲۹/۹
pH	۷/۲	۹/۰	۷/۹	۸/۰	-۱/۰۵	۲/۲۳	۵/۰
<sup>b</sup> EC (dS/m)	-/۱	۲۴/۰	۴/۴	۳/۰	۱/۵۸	۲/۷۹	۱۱۱/۴
Organic C (g/Kg)	۳/۰	۳۷/۰	۱۵/۱	۱۳/۰	۱/۳۵	۱/۶۹	۳۱/۹
<sup>c</sup> CCE (g/Kg)	۱۹۵/۰	۴۹۸/۰	۴۰۵/۴	۳۹۸/۰	۰/۲۳	۱/۰۷	۱۳/۴
<sup>d</sup> CEC (cmole/Kg)	۷/۰	۱۰۴/۰	۲۷/۵	۲۵/۰	۱/۸۴	۴/۵۱	۵۷/۹
Available P (mg/kg)	۱/۲	۳۴/۱	۱۴/۲	۱۳/۳	۱/۷۱	۴/۵۵	۴۹/۳

<sup>a</sup> Coefficient of variation; <sup>b</sup> Electrical conductivity; <sup>c</sup> Calcium carbonate equivalent; <sup>d</sup> Cation exchange capacity

بررسی رابطه‌ی بین فسفر قابل جذب خاک‌ها با ویژگی‌های فیزیکی و شیمیایی اندازه‌گیری شده در این مطالعه (شکل ۴) حاکی از رابطه معنی‌دار و مثبت بین فسفر قابل جذب خاک‌ها با کربن آلی و ظرفیت تبادل کاتیونی و رابطه منفی و معنی‌دار با قابلیت هدایت الکتریکی و واکنش خاک (pH) در سطح احتمال ۰/۱ درصد است. مطالعات مشابهی به ارتباط مثبت بین فسفر قابل جذب و ماده آلی (Mir *et al.*, 2014; Mohebbi, 2016) و ارتباط منفی آن با واکنش خاک (Mir *et al.*, 2016) اشاره نموده‌اند. به طور کلی می‌توان گفت مواد آلی از طریق مکانیسم‌هایی مانند پوشش سزکویی اکسیدها توسط هوموس، جایگزینی آنیون فسفات به وسیله یون هومات، تشکیل کمپلکس-های پایدار بین آنیون‌های آلی حاصل از تجزیه مواد آلی خاک با آهن و آلومینیوم که از واکنش آنها با فسفر ممانعت می‌کند، باعث کاهش تثبیت فسفر در خاک و افزایش قابلیت دسترسی آن برای گیاه می‌شوند (هاولین ۲۰۰۵). معاضد و همکاران (۲۰۱۰) نیز دریافتند که اختلاف در ظرفیت جذب فسفر در خاک‌های مختلف در امیدیه خوزستان به دلیل اختلاف در مقدار ماده آلی، کلسیم و آلومینیوم قابل تبادل و pH خاک می‌باشد. اما نکته قابل توجه عدم وجود ارتباط معنی‌دار بین فسفر قابل جذب با کربنات کلسیم خاک‌ها است. (2003) Bertrand *et al.* نیز در مطالعه خود بین فسفر قابل جذب و کربنات کلسیم رابطه معنی‌داری گزارش نکردند. این محققین اظهار داشته‌اند که تأثیر کربنات کلسیم افزون بر مقدار، به نوع و اندازه ذرات آن نیز بستگی دارد.



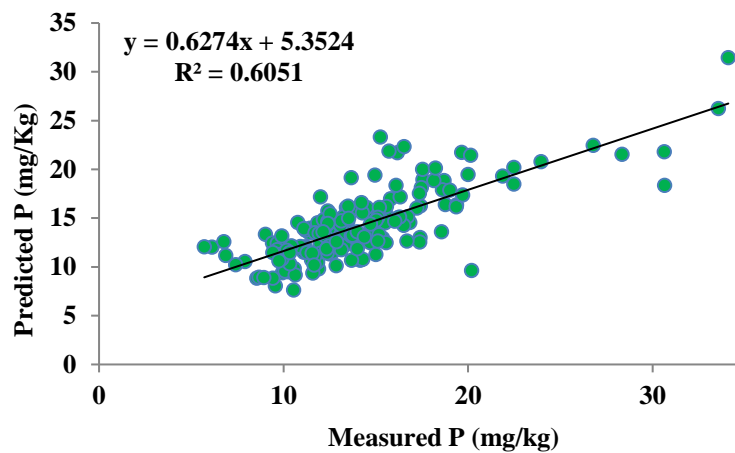
شکل ۴- ضرایب همبستگی بین فسفر قابل جذب و ویژگی‌های فیزیکوشیمیایی مطالعه شده‌ی خاک‌ها در شمال استان خوزستان

### نتایج پیش‌بینی فسفر قابل جذب خاک‌ها

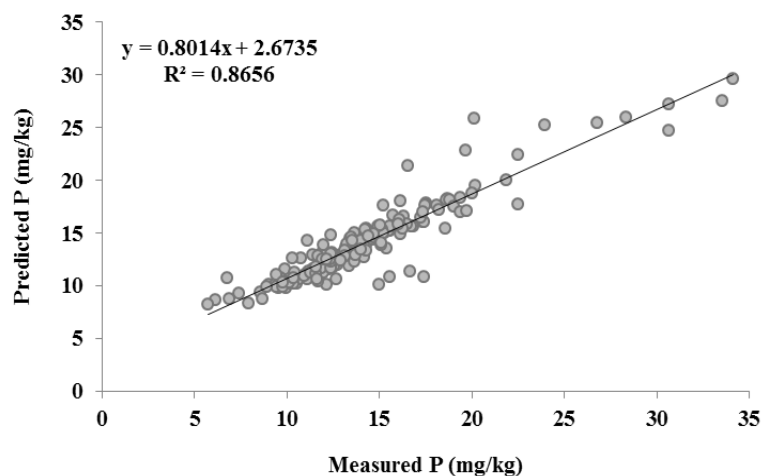
نتایج این مطالعه نشان داد که در مدل‌سازی پیش‌بینی فسفر قابل جذب خاک‌ها با استفاده از سایر ویژگی‌های خاکی اندازه‌گیری شده در این مطالعه (شکل ۵) کیفیت پیش‌بینی‌ها حدوداً ۶۰/۵ درصد است. *Pilevar Shahri et al. (2011)* نیز در مطالعه‌ای با استفاده از آنالیز داده‌های سطح زمین و بهره‌گیری از مدل رگرسیون خطی، کربن آلی را در خاک‌های سطحی (۰-۱۰ سانتی‌متری) منطقه ضرغام آباد سمیرم پیش‌بینی و گزارش نمودند که مدل رگرسیونی قادر به برآورد ۶۰ درصد از تغییرات مکانی کربن آلی در منطقه است. *Mir et al. (2016)* نیز در مطالعه‌ای که در دشت سیستان انجام دادند، گزارش نمودند که استفاده از رابطه‌ی رگرسیونی تنها قادر به پیش‌بینی ۴۳ درصد از تغییرات فسفر قابل جذب خاک می‌باشد.

در مقابل در مدل‌سازی با استفاده از الگوریتم جنگل تصادفی، کم‌ترین مقدار خطای تخمین نمونه‌های بیرون از سبد (OOB) که توسط الگوریتم جنگل تصادفی در مدل‌سازی استفاده نشده‌اند و برای اعتبارسنجی مستقل به کار می‌روند، ۵۶/۴ درصد بود که با تعداد ۵۰۰ درخت و ۸ متغیر در هر درخت به دست آمد. *Grimm et al. (2008)* نیز مقدار خطای تخمین بیرون از سبد (OOB) را برای تخمین مقدار ذخایر کربن آلی خاک با استفاده از الگوریتم جنگل تصادفی ۹۵-۷۴ درصد گزارش نمودند. *Pahlevan-Rad et al. (2015)* نیز مقدار خطای تخمین OOB در رابطه با مدل‌سازی پیش‌بینی کلاس‌های خاک در استان گلستان با استفاده از الگوریتم جنگل تصادفی را ۵۲/۳ درصد گزارش نمودند که از بررسی تعداد ۳۰۰ درخت و ۱۷ متغیر در هر درخت به دست آمده بود. مقایسه نتایج این پژوهش با نتایج سایر پژوهشگران که در بالا به برخی از آن‌ها اشاره شد حاکی از آن است که خطای به دست آمده در مدل‌سازی تغییرات مکانی فسفر قابل جذب به روش جنگل تصادفی تقریباً مشابه نتایج گزارش شده در این مطالعات بوده است.

مدل خطی (Linear Model)



الگوریتم جنگل تصادفی (Random Forest Algorithm)



شکل ۵- مقایسه مقادیر واقعی و پیش‌بینی شده‌ی فسفر قابل جذب توسط مدل‌های رگرسیون خطی و جنگل تصادفی

### تعیین مناسب‌ترین روش مدل‌سازی در پیش‌بینی فسفر قابل جذب خاک‌های منطقه‌ی مطالعاتی

به‌منظور ارزیابی نتایج مدل‌سازی تغییرات مکانی فسفر قابل جذب در خاک‌های سطحی شمال استان خوزستان با استفاده از روش‌های مختلف، از شاخص‌های آماری پیشنهاد شده در مطالعات قبلی شامل ضریب تبیین ( $R^2$ )، جذر میانگین مربعات خطا (RMSE)، میانگین مربعات خطا (MSE) و میانگین خطای مطلق (MAE) استفاده شد که نتایج آن در جدول ۲ ارائه شده است. همانگونه که مشاهده می‌گردد از میان مدل‌های بررسی شده، پیش‌بینی‌های انجام شده توسط مدل جنگل تصادفی در مقایسه با مدل خطی در رابطه با داده‌های مرحله‌ی آموزش، مرحله اعتبارسنجی و کل داده‌های آزمایشی به مقادیر اندازه‌گیری شده (واقعی) فسفر قابل جذب نزدیک‌تر بوده و خطای تخمین این مدل در مقایسه با مدل رگرسیون خطی کمتر است. این موضوع احتمالاً به دلیل وجود روابط غیرخطی بین پدیده‌ها است که در مدل خطی در نظر گرفته نمی‌شود. نتایج مشابه در سایر مطالعات نیز بر برتری مدل جنگل تصادفی در پیش‌بینی ویژگی‌ها و کلاس‌های خاک دلالت دارد. به‌عنوان نمونه، Khordehbin *et al.* (2020) در مطالعه‌ای با مقایسه‌ی روش‌های مختلف داده‌کاوی در پیش‌بینی ذخیره‌ی کربن آلی خاک در شهرستان بهبهان گزارش نمودند که الگوریتم جنگل تصادفی در مقایسه با مدل‌های شبکه‌ی عصبی مصنوعی و ماشین بردار پشتیبان تخمین‌های نزدیک به واقعیت‌تری را ارائه می‌دهد. Hengl *et al.* (2015) نیز در بخشی از آفریقا به تهیه نقشه رقوم‌ی ویژگی‌های خاک با دو مدل رگرسیون خطی چندگانه و جنگل تصادفی پرداخته و بر اساس نتایج اعتبارسنجی مدل‌ها گزارش نمودند که الگوریتم جنگل تصادفی نسبت به مدل رگرسیون خطی چندگانه توانسته است که به‌طور معنی‌داری ویژگی‌های خاک را دقیق‌تر پیش‌بینی نماید.

جدول ۲- مقایسه‌ی مدل‌های خطی و جنگل تصادفی در پیش‌بینی فسفر قابل جذب خاک‌های شمال استان خوزستان

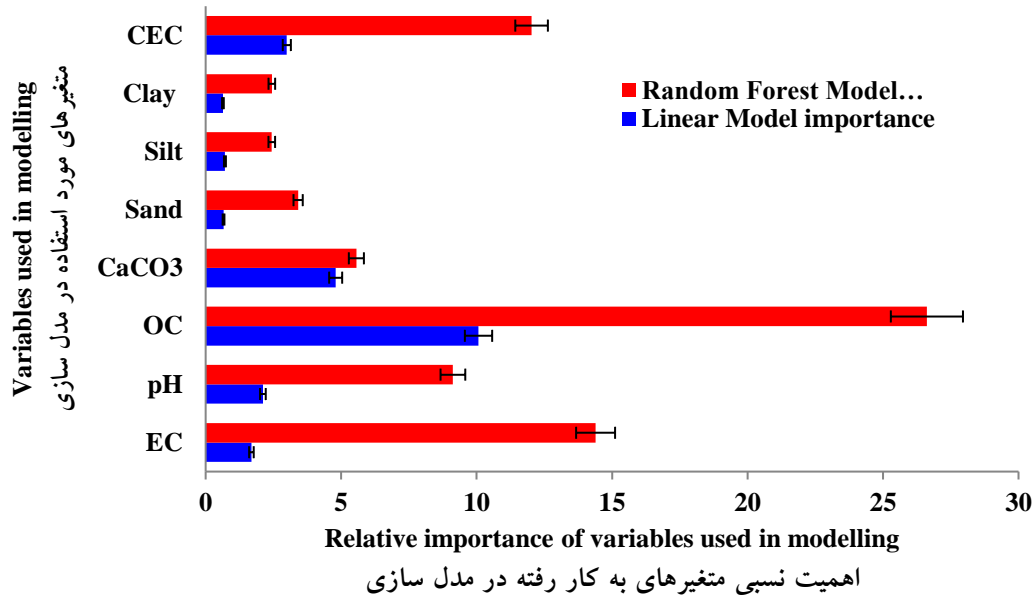
Types of Data	N	R <sup>2</sup>	Linear Model		
			RMSE	MSE	MAE
Training	۱۸۳	۰/۶۴۲	۲/۹۱۲	۸/۴۸۲	۲/۰۵۶
Test	۶۷	۰/۳۶۳	۲/۷۸۷	۷/۷۷۲	۲/۳۳۶
Total	۲۵۰	۰/۶۰۵	۲/۸۸۶	۸/۳۳۱	۲/۱۱۶
Random Forest Algorithm					
Training	۱۸۳	۰/۹۴۵	۱/۴۲۹	۲/۰۴۳	۰/۹۶۸
Test	۶۷	۰/۴۶۴	۲/۵۳۱	۶/۴۰۷	۱/۸۴۹
Total	۲۵۰	۰/۸۸۳	۱/۷۱۷	۲/۹۴۹	۱/۱۰۳

### تعیین اهمیت متغیرهای مورد بررسی در تخمین فسفر قابل جذب خاک‌ها

اهمیت متغیرهای خاکی استفاده شده در مدل‌سازی خطی در شکل ۶ نشان داده شده است. در الگوریتم جنگل تصادفی دو روش میانگین کاهش دقت و میانگین کاهش ضریب جینی اهمیت متغیرها را نشان می‌دهند. در روش میانگین کاهش دقت که در بیش‌تر مطالعات استفاده شده، مقادیر درست متغیرها با مقادیری که به‌طور تصادفی برای هر درخت تولید شده است، جایگزین می‌گردد و اثر این تغییر را روی طبقه‌بندی اندازه‌گیری می‌کند؛ اگر این جایگزینی اثری روی خطای اندازه‌گیری نداشته باشد اهمیت آن متغیر کم است و اگر مقدار خطای اندازه‌گیری افزایش یابد آن متغیر در زمره‌ی متغیرهای مهم قرار می‌گیرد (Breiman and Cutler, 2004; Pahlevan-Rad *et al.*, 2015). در این پژوهش نیز از روش میانگین کاهش دقت برای تعیین اهمیت متغیرهای به‌کار رفته در فرآیند مدل‌سازی فسفر قابل جذب استفاده گردید.

همان‌گونه که مشاهده می‌گردد در مدل‌سازی فسفر قابل جذب خاک‌ها با استفاده از الگوریتم جنگل تصادفی، از میان ویژگی‌های فیزیکی و شیمیایی اندازه‌گیری شده در این پژوهش، محتوای کربن آلی خاک‌ها بیشترین اهمیت را در پیش‌بینی فسفر قابل جذب خاک‌ها داشته است (شکل ۶). پس از آن ویژگی‌هایی مانند سوری، ظرفیت تبادل کاتیونی و محتوای کربنات کلسیم معادل خاک‌ها مؤثرترین ویژگی‌های فیزیکی و شیمیایی خاک‌ها برای پیش‌بینی فسفر قابل جذب خاک‌ها محسوب می‌گردند. این در حالی است که از میان متغیرهای مورد بررسی در مدل‌سازی خطی، ویژگی‌های کربن آلی خاک، کربنات کلسیم معادل و ظرفیت تبادل کاتیونی خاک‌ها به‌ترتیب بیش‌ترین تأثیر را در مدل‌سازی تغییرات مکانی فسفر قابل جذب خاک‌ها نشان داده‌اند. مقایسه نتایج به‌دست آمده از دو روش رگرسیون خطی چندگانه و الگوریتم جنگل تصادفی حاکی از آن است که هر دو مدل بررسی شده در رابطه با اهمیت ویژگی‌هایی مانند کربن آلی، ظرفیت تبادل کاتیونی و محتوای کربنات کلسیم معادل خاک‌ها در کنترل محتوای فسفر قابل جذب در خاک‌های منطقه توافق دارند. با این حال، از آنجا

که در الگوریتم جنگل تصادفی، روابط غیرخطی بین ویژگی‌ها نیز مورد بررسی قرار می‌گیرد، شوری خاک نیز به‌عنوان یکی از مهم‌ترین شاخص‌های تأثیرگذار بر محتوای قابل جذب فسفر خاک‌ها لحاظ شده است که به‌نظر می‌رسد با واقعیت‌های منطقه مطالعاتی تطابق بیشتری را نشان می‌دهد.



شکل ۶- اهمیت نسبی ویژگی‌های خاکی به کار برده شده در مدل‌سازی فسفر قابل جذب خاک‌ها با استفاده از مدل خطی و الگوریتم جنگل تصادفی

## نتیجه‌گیری کلی

در این پژوهش، پیش‌بینی پراکنش مکانی فسفر قابل جذب خاک‌ها در شمال استان خوزستان با استفاده از دو روش مدل‌سازی رگرسیون چندگانه خطی و الگوریتم جنگل تصادفی مورد بررسی قرار گرفت. یافته‌های این پژوهش حاکی از آن است که ۳۲/۴ درصد از نمونه‌های مورد بررسی از لحاظ محتوای فسفر قابل جذب در زمره‌ی خاک‌های فقیر (کمتر از ۵ میلی‌گرم بر کیلوگرم) محسوب می‌گردند که نیازمند توجه جدی بوده و کوددهی به ویژه استفاده از کودهای با منشاء آلی در آن‌ها ضروری به نظر می‌رسد. همچنین، نتایج این پژوهش یافته‌های مطالعات پیشین دال بر برتری الگوریتم جنگل تصادفی در پیش‌بینی ویژگی‌های خاک‌ها را تأیید می‌نماید؛ به‌گونه‌ای که این الگوریتم می‌تواند بیش از ۸۷ درصد از تغییرات فسفر قابل جذب خاک‌ها را در منطقه‌ی مطالعاتی توجیه نماید. لذا، به‌نظر می‌رسد در پیش‌بینی فسفر قابل جذب خاک‌ها علاوه بر لحاظ نمودن روابط خطی باید به روابط غیرخطی بین متغیر هدف (فسفر قابل جذب) و سایر متغیرهای پیش‌بینی‌کننده نیز توجه نمود و در این ارتباط ویژگی‌هایی مانند محتوای کربن آلی، قابلیت هدایت الکتریکی خاک‌ها، ظرفیت تبادل کاتیونی و کربنات کلسیم معادل خاک‌ها به ترتیب مؤثرترین ویژگی‌های پیش‌بینی‌کننده‌ی فسفر قابل جذب خاک‌ها در شمال استان خوزستان تشخیص داده شد.

## سپاس‌گزاری

هزینه‌های اجرای این پژوهش توسط صندوق حمایت از پژوهشگران و فناوران کشور (طرح پژوهشی شماره ۹۸۰۲۲۷۸۳) و دانشگاه شهید چمران اهواز پرداخت شده است.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

## REFERENCES

- Bertrand, I., Holloway, R.E., Armstrong, R.D., McLaughlin, M.J. 2003. Chemical characteristics of phosphorus in alkaline soils from southern Australia. *Australian Journal of Soil Research*, 41: 61-76.
- Breiman, L. 2001. Random forests. *Machine Learning*, 45: 5-32.
- Breiman, L., Cutler, A. 2004. Random Forests. Department of Statistics, University of Berkeley. [http://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm).
- Camera, C., Zomeni, Z., Noller, J.S., Zissimos, A.M., Christiforou, I.C., Bruggeman, A. 2017. A high

- resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma*, 285: 35-49.
- Gee, G.W., Bauder, J.W., 1986. Particle size analysis. In: Klute, A. (Ed.), *Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods*, second ed. ASA and SSSA, Madison, WI, pp. 383-411.
- Grimm, R., Behrens, T., Marker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using random forests analysis. *Geoderma*, 146 (1-2): 102-113.
- Havlin, J.L., S.L. Tisdale, J.D. Beaton, Nelson, W.L. 2005. *Soil fertility and fertilizers: An introduction to nutrient management*. Seventh edition, Pearson Education, Upper Saddle River, NJ.
- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E. 2015. Mapping soil properties of Africa at 250 m resolution: Random Forests significantly improve current predictions. *PLoS ONE* 10 (6): e0125814.
- Hojati, S. 2018. Spatial variability of selected heavy metals in surface soils of northern Khuzestan Province. Project No. 93021034, Iran National Science Foundation, Tehran, Iran, 133 pages.
- Jamshidi, M., Delavar, M.A., Taghizadeh-Mehrjerdi, R., Brungard, C. 2019. Evaluating digital soil mapping approaches for 3D mapping of soil organic carbon. *Iranian Journal of Soil Research*, 33(2): 227-239.
- Khordehbin, S., Hojati, S., Landi, A., Ahmadianfar, I. 2020. Comparison of different data mining methods in predicting soil organic carbon storage in some lands of Behbahan city. *Iranian Journal of Soil and Water Research*, 51(4): 1041-1054.
- Loeppert, R.H., Suarez, D.L. 1996. Carbonate and Gypsum. In: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), pp. 437-474. *Methods of Soil Analysis. Part 3. Chemical Methods*, ASA and SSSA, Madison, USA.
- Minasny, B., McBratney, A.B., 2006. A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32(9): 1378-1388.
- Mir, H., Gholamalizadeh Ahangar, A., Shabani, A. 2015. 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.
- Mirzashahi, K., Salimpour, S., Paknejad, A.R. 2014. Fertilizer recommendation for crops and orchards in the north of Khuzestan province. Journal No. 536, Soil and Water Research Institute, Tehran, Iran, 56 pages.
- Moazed, H., Y, Hoseini, A.A. Naseri, Abbasi, F. 2010. Determining phosphorus adsorption isotherm in soil and its relation to soil characteristic. *Journal of Food, Agriculture & Environment*. 8(2): 1153-1157.
- Mohebbi, S.M.J. 2014. Investigation of relationships between available phosphorus, potassium and some soil properties in agricultural lands of Varamin- Iran. *International Journal of Agriculture and Biosciences*, 3(1): 7-12.
- Mousavi, A., Shahbazi, F., Oustan, Sh., Jafarzadeh, A.A., Minasny, B. 2020. Application of two data mining techniques for mapping the spatial distribution of soil organic carbon (case study: East shore of Urmia Lake). *Journal of Water and Soil*, 34(3): 689-705.
- Nelson, D.W., Sommers, L., 1982. Total carbon, organic carbon, and organic matter. In: Page, A.L., (Ed.), *Methods of soil analysis. Part 2: Chemical and microbiological properties*, pp. 539-579. American Society of Agronomy and Soil Science Society of America, Madison, WI, USA.
- Olsen, S.R., Cole, C.V., Watanabe, F.S., Dean, L.A., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. U.S. Dep. Of Agric. Circ, 939 pp.
- Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Brungard, C.W., Kiani, F., Komaki, C.B., Bogaert, P. 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma*, 279: 141-148.
- Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Kiani, F., Komaki, C.B. 2015. Digital soil mapping using Random Forest model in Golestan province. *Journal of Soil and Water Conservation*, 21(6): 73-93.
- Pilevar Shahri, A.J., Ayoubi, Sh., Khademi, H. 2011. Comparison of artificial neural network (ANN) and multivariate linear regression (MLR) models to predict soil organic carbon using digital terrain analysis (case study: Zargham Abad Semirom, Isfahan province). *Journal of Water and Soil*, 24(6): 1151-1163.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rhoades, J.D., 1986. Soluble salts. In: Page, A.L., Miller, R.H., Keeny, D.R. (Eds.), *Methods of Soil Analysis*,



- Part 2. Chemical and Microbiological Properties, second ed. ASA and SSSA, Madison, WI, pp. 167–180.
- Rhoades, J. D., Chanduvi, F., Lesch, S., 1999. Food and agricultural organization of the UN. Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO Irrigation and Drainage, Paper No. 57.
- Riemersma, S., Little, J., Ontkean, G., Moskal-Hébert, T., 2006. Phosphorus sources and sinks in watersheds: A review. Alberta soil phosphorus limits project, 82 p.
- Souri, M.K. 1397. Phosphorus in agriculture and environment with emphasis on management and measurement methods. Agricultural Education and Extension Publications, 348 pages.
- Statistical Center of Iran. 2013. Iran statistical yearbook. Tehran: Management and Planning Organization.
- Walkley, A., Black, I.A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 37(1): 29-38.