

Research Article

Agricultural Engineering., 47(3) (2024) 411-427

ISSN (P): 2588-526X

DOI: 10.22055/agen.2024.47331.1735

ISSN (E): 2588-5944

Incorporation of spatial autocorrelation in soil salinity distribution modeling in Qazvin area

A. Jafari^{1*}, F. Sarmadian^{2*}, Ahmad Heidari³ and Zahra Rasaie⁴

1. Associate Professor, Department of Soil Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran. * Corresponding author, E-mail: a.jafari@uk.ac.ir
2. Professor, Soil Science Department, Faculty of Agricultural, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. * Co-corresponding author, E-mail: fsarmad@ut.ac.ir
3. Professor, Soil Science Department, Faculty of Agricultural, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: ahaidari@ut.ac.ir
4. Researcher, Soil Science Department, Faculty of Agricultural, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: zahra.rasaie@ut.ac.ir

Received: 8 July 2024

Accepted: 26 August 2024

*Corresponding Author: a.jafari@uk.ac.ir

Abstract

Introduction: Machine learning algorithms usually do not consider spatial autocorrelation in soil data, unless it is perspicuity specified. Machine learning algorithms that compute autocorrelated observations have been recently formulated, such as geographic random forest (Georganos et al., 2019), or spatial ensemble techniques (Jiang et al., 2017). In theory, if we include all relevant environmental variables to model a soil property or class, there should be no spatial autocorrelation in the residuals of the fitted models. If this happens, some important predictors are likely to be missed. Despite the availability of the data set and the care taken during modeling, residual autocorrelation is still likely to occur. Several researchers have suggested the use of spatial alternative covariates as an indicator of spatial location in the SCORPAN model. The most common alternative is to use geographic coordinates (east and north) as covariates in the model, which leads to synthetic maps, especially when used in combination with tree-based algorithms. On the other hand, distance maps from observation locations are proposed by Hengl et al. (2018). Distance maps to observation locations usually do not have a clear meaning in terms of soil processes in an area (e.g., distance from a river). In the field of digital soil mapping, the current use of distance maps is not satisfactory for several reasons. The presence of pseudo-covariates with a set of covariates related to pedology is not very useful because it prevents the analysis of residuals and the creation of new hypotheses from these residuals. It also hinders the interpretation of the most important key predictors. Finally, pseudo-distance covariates may be well integrated into multiple pedology-related covariates, making them better predictors or masking the effect of pedology-related covariates. In spatial ecology, spatial eigen-vector maps, spatial filters or trend level regression replace distance maps in reducing or eliminating spatial autocorrelation (Kuhn et al., 2009). The purpose of this study is first to detect and calculate the spatial autocorrelation in the soil data. In the second step, it is going to develop a non-spatial model without considering the spatial autocorrelation, then to extract the spatial eigenvectors as an index of the spatial autocorrelation, and finally to use them as independent variables in spatial modeling.

Methods and Materials: In this study, the soil salinity data utilized of 297 soil samples from a section of the Qazvin plain. The first and second derivatives of a digital elevation model as topography factors, remote sensing



indices, parent material map, geofom map, and annual average temperature and rainfall maps were used to select the most important auxiliary variables. Finally, in order to select the best and most relevant environmental variables for modeling, the correlation between these variables and the dependent variable i.e. soil salinity in 297 study points was used using FSelector package of R software. Moran's I and Jerry's C indices were used to evaluate the spatial autocorrelation of soil data. First, the non-spatial ordinary least square (OLS) model was fitted to predict the spatial distribution of soil salinity. At this stage, spatial autocorrelation was not considered. Then spatial regression was fitted by calculating spatial filters through spatial eigenvectors as independent variables. Finally, the comparison of the outputs of the non-spatial OLS model and the spatial regression model was done with criteria such as R^2 , Akaike information criteria (AIC), autocorrelation of residuals and root mean of square error (RMSE).

Results and Discussion: Statistical analysis indicated the high variability of soil salinity in the study area (coefficient of variation or CV more than 35%). Also, soil salinity shows high skewness and kurtosis, indicating its abnormal distribution. The high variability of this soil characteristic emphasizes the interaction of complex and numerous factors, including soil forming processes and different management strategies. The most important variables selected based on the correlation analysis include elevation, Multi-resolution Valley Bottom Flatness (MrVBF), wetness index, drainage basin, greenness index, normalized differential vegetation index (NDVI) and the corrected and transformed vegetation index (CTVI). A total of 7 variables were selected, which include four topography variables and three remote sensing variables. Among the topographical variables, the MrVBF had the most importance (correlation: 0.70). The spatial distribution map of soil salinity shows that the soil salinity is low in the northern, northeastern and northwestern parts towards the center of the studied area. The highest amount of salinity is found in the southern and southeastern regions. Moran's I and Jerry's C indices were 0.57 and 0.4, respectively. Based on both indices, soil salinity in the study area exhibits spatial autocorrelation. In the spatial regression model, by considering spatial autocorrelation, compared to the non-spatial model, the results are improved. By considering the spatial autocorrelation, the value of R^2 increased, while the values of AIC, spatial autocorrelation of the residuals and RMSE decreased. The distribution maps of residuals from the non-spatial OLS model and the spatial regression model differ in terms of the spatial sign of the residuals and the spatial autocorrelation distribution that can be recognized in the form of clusters. Clusters (red or blue) indicate the presence of spatial autocorrelation in the residuals. In the distribution map of the residuals of the non-spatial model, more and larger clusters (marked with green ovals) are identified, indicating the existence of spatial autocorrelation in the residuals of the model. The presence of spatial autocorrelation in the residuals of a model shows that the model is not able to remove the spatial dependence, which may be due to not considering an important auxiliary variable in the modeling.

Conclusion: This study was conducted in order to investigate the effect of spatial autocorrelation on the results of soil salinity modeling. Soil salinity prediction was done by non-spatial OLS model (without considering spatial autocorrelation) and spatial regression model (with spatial autocorrelation considered). The results indicated the improvement of the performance of the spatial regression model compared to the non-spatial ordinary least squares model. In the spatial model, considering the spatial autocorrelation as a covariate, the value of R^2 increased, while the values of AIC, spatial autocorrelation of the residuals, and RMSE decreased. From a statistical point of view, spatial autocorrelation may be viewed as a problem, but in fact, it is a fundamental and pervasive feature of almost all soil data that must be taken into account in the modeling in order to avoid poor model fit and bias of predictions.

Keywords: *covariates, machine learning, spatial and non-spatial modeling, spatial eigen-vector mapping*

تلفیق خودهمبستگی مکانی در مدل سازی توزیع شوری خاک در منطقه قزوین

اعظم جعفری^{۱*}، فریدون سرمدیان^{۲*}، احمد حیدری^۳، زهرا رسائی^۴

- ۱- دانشیار گروه علوم خاک، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران.
 ۲- استاد گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکده گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.
 ۳- استاد گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکده گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.
 ۴- پژوهشگر گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکده گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

تاریخچه مقاله

چکیده

دریافت: ۱۴۰۳/۰۴/۱۸

پذیرش نهایی: ۱۴۰۳/۰۶/۰۵

کلمات کلیدی:

متغیرهای محیطی،

ماشین یادگیری،

مدل سازی مکانی و

غیرمکانی،

نقشه های بردار ویژه مکانی

داده های خاک مانند هر داده مکانی ممکن است دارای خودهمبستگی مکانی باشند. اگر این وابستگی مکانی در باقیمانده های یک مدل آماری مشاهده شود، یکی از فرضیه های کلیدی آنالیز آماری که شامل استقلال و توزیع یکنواخت باقیمانده ها است، نقض می شود. معمولا الگوریتم های یادگیری ماشین، خودهمبستگی مکانی در داده های خاک را در نظر نمی گیرند. مطالعه حاضر سعی دارد خودهمبستگی مکانی را به عنوان یک متغیر مستقل در مدلسازی تخمین شوری خاک لحاظ کند و نتایج تخمین را بررسی کند. بدین منظور، شوری خاک سطحی در ۲۹۷ نقطه در منطقه آبیگ قزوین اندازه گیری شد و متغیرهای محیطی مهم انتخاب شدند. سپس یک مدل غیرمکانی حداقل مربعات معمولی و یک مدل رگرسیون مکانی بر داده های شوری خاک برازش داده شد. از دو شاخص موران و جری برای تشخیص خودهمبستگی مکانی استفاده گردید. نقشه توزیع مکانی شوری خاک در منطقه آبیگ قزوین نشان می دهد در بخش های شمالی، شمال شرقی و شمال غرب به سمت مرکز منطقه مورد مطالعه میزان شوری خاک کم بوده و بیشترین مقدار و محدودیت شوری در مناطق جنوب و جنوب شرقی دیده می شود. مقدار شاخص موران ۰/۵۷ و مقدار شاخص جری ۰/۴ به دست آمد که براساس هر دو شاخص، ویژگی شوری خاک در منطقه مورد نظر دارای خودهمبستگی مکانی است. با وارد کردن خودهمبستگی مکانی در مدل رگرسیون مکانی در مقایسه با مدل غیرمکانی نتایج پیش بینی بهبود یافت. با لحاظ کردن خودهمبستگی مکانی، مقدار R^2 افزایش، درحالیکه مقادیر AIC، خودهمبستگی مکانی باقیمانده ها و RMSE کاهش یافت. به نظر می رسد ادغام خودهمبستگی مکانی در مدل سازی خصوصیات خاک امری ضروری است و باید در نظر گرفته شود.

* عهده دار مکاتبات:

Email: a.jafari@uk.ac.ir,
fsarmad@ut.ac.ir

مدل فرعی برازش دادند که هم همبستگی غیرایستایی و هم خودهمبستگی مکانی را در نظر بگیرد.

اعمال یک مدل غیر مکانی برای نقشه برداری رقومی خاک^۷ (DSM) به خودی خود مشکلی نیست. تعریف DSM ارائه شده توسط لاگچری و مک برتنی^۸ (۲۹) این موضوع را تأیید می کند، چرا که پیش بینی هایی را برای نقشه برداری با استفاده از سیستم های استنتاج غیر مکانی خاک ارائه می دهد. از لحاظ تئوری، اگر همه متغیرهای محیطی مربوط را برای مدل سازی ویژگی یا کلاس خاک در نظر بگیریم، نباید هیچ خودهمبستگی مکانی در باقیمانده های مدل های برازش شده وجود داشته باشد. منتهی اگر خودهمبستگی مکانی اتفاق بیفتد، احتمالاً برخی از پیش بینی کننده های مهم وارد نشده اند. اما مورد مهمتر این است که، ممکن است پیش بینی های انجام شده توسط الگوریتم یادگیری ماشین مغرضانه (بایاس) باشد یا مدل به درستی برازش داده نشده باشد، چرا که وجود خودهمبستگی مکانی، نقض فرضیه استقلال بین نقاط داده است. کوهن و دورمن^۹ (۲۶) تهیه نقشه توزیع مکانی خودهمبستگی باقیمانده ها را برای تسهیل شناسایی یک فرآیند مکانی نامشخص توصیه می کنند. حتی در برخی موارد، نقشه باقیمانده ها یک الگوی واضح را نشان می دهد (به عنوان مثال افزایش باقیمانده ها با فاصله از رودخانه) و یا ممکن است به ایجاد یک فرضیه جدید یا بازسازی مدل موجود کمک کند.

با وجود در دسترس بودن مجموعه داده ها و مراقبت های انجام شده در طول مدل سازی، همچنان خودهمبستگی باقیمانده ها به احتمال زیاد رخ می دهد. چندین محقق استفاده از متغیرهای کمکی جایگزین مکانی^{۱۰} را به عنوان شاخصی از موقعیت مکانی در مدل اسکورپن پیشنهاد کرده اند یا اینکه آنها را برای خودهمبستگی مکانی موجود در داده ها در نظر بگیرند (۲۰ و ۲۸). رایج ترین جایگزین،

مقدمه

داده های خاک مانند هر داده مکانی دیگر، اغلب مستقل از یکدیگر نیستند و وابستگی آنها ناشی از موقعیت و مکان-شان در فضای مورد مطالعه است. به عبارت دیگر داده های خاک ممکن است خودهمبستگی مکانی^۱ نشان دهند. خودهمبستگی مکانی یعنی مشاهدات به دست آمده از موقعیت های نزدیک به یکدیگر وابستگی بیشتری دارند و مقادیر آنها به هم شبیه تر است و با افزایش فاصله، این وابستگی کمتر می شود. اگر این وابستگی در باقیمانده های یک مدل آماری مشاهده شود، یکی از فرضیه های کلیدی آنالیز آماری، که بیان می کند باقیمانده ها مستقل و دارای توزیع یکنواخت هستند، نقض می شود. الگوریتم های یادگیری ماشین، خودهمبستگی مکانی در داده های خاک را معمولاً در نظر نمی گیرند. سینها و همکاران^۲ (۴۷) مدل جنگل تصادفی را برای سناریوهای مختلف خودهمبستگی مکانی در مشاهدات مورد بررسی قرار دادند و به این نتیجه رسیدند که وجود خودهمبستگی مکانی منجر به واریانس بالایی در باقیمانده ها می شود. اخیراً الگوریتم های یادگیری ماشین که خودهمبستگی مکانی داده ها را در نظر می گیرند، ارائه شده اند، مانند جنگل تصادفی جغرافیایی^۳ (۱۲) یا تکنیک های مکانی گروهی^۴ (۲۱). این دو روش با برازش مدل های محلی مکانی با استفاده از تنها مشاهدات همسایه به رگرسیون وزن دار جغرافیایی خلاصه می شوند. جیانگ و همکاران^۵ (۲۱) منطقه جغرافیایی در مینه سوتا آمریکا را به زیرمنطقه های مجزای جغرافیایی تجزیه کردند و یک مدل محلی را در هر زیر ناحیه ارائه کردند. یافته های آنها، بهبود عملکرد طبقه بندی در نتیجه گنجاندن وابستگی مکانی در ماشین یادگیری را نشان داد. جورجانوس و همکاران^۶ (۱۲) برای هر مشاهده با استفاده از الگوریتم جنگل تصادفی، یک

1- Spatial autocorrelation

2- Sinha *et al.*

3- geographical random forest

4- spatial ensemble techniques

5- Jiang *et al.*

6 - Georganos *et al.*

7- Digital Soil Mapping

8- Lagacherie & McBratney

9- Kuhn & Dormann

10- spatial surrogate covariates

گرفتند (۷ و ۱۶). نقشه‌های بردار ویژه مکانی براساس وزن‌دهی فاصله بین مشاهدات کار می‌کند و یک ماتریسی از وزن‌ها براساس مقدار خودهمبستگی مکانی متغیر در فواصل مختلف به عنوان متغیر کمکی ارائه می‌کند و قابلیت استفاده در مدل‌های یادگیری ماشین را دارد.

امروزه، گسترش روزافزون روش‌های پیش‌بینی، خاکشناسان را قادر می‌سازد تا توزیع مکانی منابع خاک را با توجه به شرایط توپوگرافی محیط و درک مکانیسم‌هایی که چنین روابط خاک-لندفرم را ایجاد می‌کنند، بهتر مدل‌سازی کنند. در این مطالعه، ما قصد داشتیم به یک موضوع اساسی اما تا حدودی کمتر ارائه شده بپردازیم که در واقع تأثیر خودهمبستگی مکانی بر مدل‌سازی ویژگی خاک است. هدف از مطالعه حاضر، در مرحله اول تشخیص و محاسبه خودهمبستگی مکانی در داده‌های خاک است و در مرحله دوم توسعه یک مدل غیرمکانی بدون لحاظ خودهمبستگی مکانی و سپس استخراج بردارهای ویژه مکانی^۷ به عنوان شاخصی از خودهمبستگی مکانی و استفاده از آنها به عنوان متغیر مستقل در مدل‌سازی مکانی است. به طور کلی هدف مطالعه حاضر، بهبود مدل‌سازی با در نظر گرفتن خودهمبستگی مکانی است.

مواد و روش‌ها

تشریح منطقه مورد مطالعه

این مطالعه در بخشی از دشت قزوین در مرز استان‌های قزوین و البرز به مساحت تقریبی ۶۰۰۰۰ هکتار انجام شد (شکل ۱). منطقه مورد نظر بین عرض‌های جغرافیایی ۳۵ درجه و ۵۵ دقیقه تا ۳۶ درجه و ۱۰ دقیقه شمالی و طول‌های جغرافیایی ۵۰ درجه و ۱۵ دقیقه تا ۵۰ درجه و ۳۵ دقیقه شرقی قرار دارد. ارتفاع از ۱۱۴۱ متر تا ۱۷۵۲ متر بالای سطح دریا متغیر است. با حرکت از قسمت‌های شمالی به جنوبی منطقه میانگین دمای سالانه از ۱۳/۳ تا ۱۵/۵ درجه سانتی‌گراد افزایش و بارندگی از ۳۱۴/۲ به ۲۵۳/۵ میلی‌متر

استفاده از مختصات جغرافیایی (شرق و شمال) به عنوان متغیر کمکی در مدل است. از طرف دیگر، نقشه‌های فاصله از مکان‌های مشاهده‌ای به عنوان متغیر جایگزین مکانی توسط هنگل و همکاران^۱ (۱۸) پیشنهاد شده است. نقشه‌های فاصله تا مکان‌های مشاهده معمولاً معنا و مفهوم مشخصی از نظر فرآیندهای خاک‌سازی در یک منطقه (مثلاً فاصله از رودخانه) ندارند. بهرنز و همکاران^۲ (۳) پیشنهاد می‌کنند از فاصله اقلیدسی استفاده شود که نقشه‌های فاصله از موقعیت‌های مرجع در منطقه مورد مطالعه مانند گوشه یا مرکز، مشخص هستند. در چندین مطالعه موردی، استفاده از نقشه‌های فاصله به عنوان متغیرهای کمکی در مقایسه با یک مدل بدون نقشه‌های فاصله در مجموعه متغیرهای کمکی، کاهش قابل ملاحظه‌ای از خودهمبستگی باقیمانده را نشان داده‌اند (۴ و ۱۷).

در زمینه نقشه‌برداری رقومی خاک، استفاده فعلی از نقشه‌های فاصله به دلایل متعدد رضایت بخش نیست. حضور متغیرهای شبه-کمکی^۳ (متغیر فاصله) با مجموعه متغیرهای کمکی مرتبط با پدولوژی چندان مفید نیست، زیرا از تجزیه و تحلیل باقیمانده‌ها و ایجاد فرضیه‌های جدید از این باقیمانده‌ها جلوگیری می‌کند (۱۶)، همچنین تفسیر مهم‌ترین پیش‌بینی‌کننده‌های کلیدی را مختل می‌کند. در نهایت، متغیرهای شبه کمکی فاصله ممکن است به خوبی در چندین متغیر کمکی مرتبط با پدولوژی ادغام شوند، و آنها را به پیش‌بینی‌کننده‌های بهتری تبدیل کند یا اثر متغیرهای کمکی مرتبط با پدولوژی را پنهان کند. در حال حاضر، نقشه‌های بردار ویژه مکانی^۴، فیلترهای مکانی^۵ یا آنالیز سطح روند^۶ جایگزین نقشه‌های فاصله‌ای در کاهش یا حذف خودهمبستگی مکانی شدند (۲۷). معمولاً فیلترهای مکانی در مطالعات سنجش از دور و تکنیک آنالیز سطح روند در مطالعات درون‌یابی و زمین‌آمار مورد استفاده قرار

- 1- Hengl *et al.*
- 2- Behrens *et al.*
- 3- pseudo-covariates
- 4- spatial eigenvector maps
- 5- spatial filters
- 6- trend-surface regression

7- spatial eigen vectors

محیطی در نرم‌افزار SAGA نسخه 8.0.1 استخراج گردید. همچنین، میانگین ۶ ساله تصاویر ماهواره‌ای لندست ۸ مربوط به فصل‌های بهار و تابستان سال‌های ۲۰۱۳ تا ۲۰۲۳ که با بازه تاریخی نمونه‌برداری از خاک‌ها منطبق می‌باشند، برای استخراج شاخص‌های تصاویر ماهواره‌ای استفاده شد که در نهایت، ۲۷ شاخص مختلف تهیه شد.

نقشه مواد مادری با رقوم‌سازی نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ موجود در منطقه (سازمان زمین‌شناسی و اکتشاف معادن ایران) و نقشه ژئوفرم از روی هم انداختن نقشه‌های زمین‌شناسی و شکل زمین در نرم‌افزار ArcGIS نسخه 10.3 تهیه شدند.

همچنین، نقشه‌های میانگین سالانه دما و بارندگی با استفاده از داده‌های اقلیمی بلندمدت (۴۹ ساله - ایستگاه هواشناسی قزوین ۱۳۴۹-۱۳۹۸) ۲۴ ایستگاه هواشناسی در منطقه مورد مطالعه با درون‌یابی بر اساس روش درون‌یابی تجربی کریجینگ بیزی با قدرت تفکیک مکانی ۲۵۰ متر در نرم‌افزار ArcGIS تهیه شدند (۳۳، ۳۴، ۳۵، ۳۶، ۴۰).

آماده‌سازی پایگاه داده

در نهایت تمامی این نقشه‌ها با استفاده از روش نمونه‌برداری مجدد به قدرت تفکیک مکانی ۱۵ متر در ۱۵ متر تغییر داده شدند و برای تمامی نقشه‌ها سیستم مختصات مرجع متریک (WGS 1984, UTM 39N) تعریف شد. مقادیر متغیرهای محیطی متناظر با موقعیت جغرافیایی خاک‌رخ‌ها استخراج شدند. در نهایت، به منظور انتخاب بهترین و مربوط‌ترین متغیرهای محیطی برای مدل‌سازی (جدول ۱)، از همبستگی بین این متغیرها و متغیر وابسته یعنی شوری خاک در ۲۹۷ نقطه مطالعاتی با استفاده از بسته FSelector (۴۴) در نرم‌افزار R استفاده شد.

کاهش می‌یابد. بر اساس محاسبات انجام شده در نرم‌افزار نیوهای (۴۹) خاک‌های منطقه در سه رژیم رطوبتی زیریک خشک، اریدیک ضعیف و آکوئیک و رژیم حرارتی ترمیک قرار دارند (۴۸). مواد مادری تشکیل دهنده خاک‌های بخش‌های شمالی و مرکزی منطقه رسوبات آبرفتی و واریزه‌ای شامل سنگ‌های آذرین بازالتی آهکی مربوط به دوره کواترنر می‌باشند که در بخش‌های جنوبی منطقه به کفه‌های گلی و نمکی تغییر می‌یابد. منطقه مورد مطالعه نسبتاً مسطح است و از نظر توپوگرافی از شمال به جنوب واحدهای فیزیوگرافی مختلف شامل تپه، فلات، دشت دامنه‌ای و اراضی پست در منطقه دیده می‌شوند (۳۵، ۳۶، ۴۰، ۴۲).

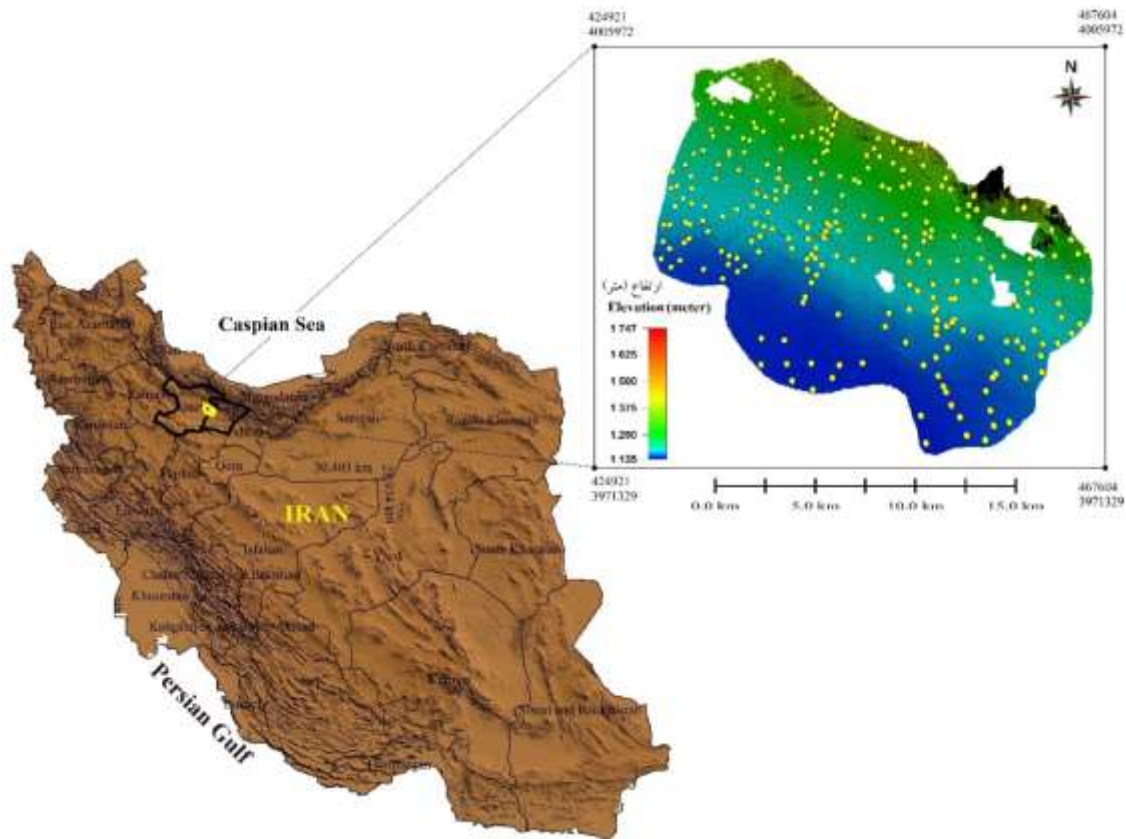
داده‌های خاک

مطالعات میدانی در مقیاس نیمه‌تفصیلی طی یک دوره ۶ ساله (۲۰۱۶ تا ۲۰۲۲) انجام شده است (۳۵، ۳۶، ۴۰، ۴۲). نمونه برداری بصورت ترکیبی از روش‌های معمول نمونه‌برداری از جمله تصادفی، شبکه‌ای و فرا مکعب لاتین مشروط (cLHS) انجام شد که شامل خاک‌رخ و مته با میانگین فاصله ۴۰۰ تا ۶۰۰ متر می‌باشد (شکل ۱) (۴۵). در نهایت، ۲۹۷ نمونه خاک سطحی (۰-۳۰ سانتی‌متری) برای اندازه‌گیری هدایت الکتریکی برداشت گردید.

این داده‌ها از گستره وسیعی از مطالعات پیشین انجام شده در منطقه (۲۲، ۲۳، ۳۰، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۲) در دسترس می‌باشند.

متغیرهای کمکی

مشتقات درجه اول و دوم مدل رقومی ارتفاع به عنوان فاکتورهای توپوگرافی در نظر گرفته شدند. بدین منظور مدل رقومی ارتفاعی آلوس پالسار (<https://asf.alaska.edu/>) با دقت تفکیک مکانی ۱۲/۵ متر مورد استفاده قرار گرفت. بر این اساس ۴۵ متغیر



شکل (۱) منطقه مطالعاتی در مرز استان‌های قزوین و البرز، نقاط نمونه‌برداری (دایره‌های زرد رنگ) به همراه تغییرات ارتفاع (DEM) در منطقه مورد مطالعه

Figure (1) Location of the study area in boundary of Alborz and Ghazvin provinces, and the sampling points (yellow circles) on DEM

برای ارزیابی خودهمبستگی مکانی، باید مفهوم فاصله بین مشاهدات مشخص شود. این فواصل در ماتریس وزن که رابطه بین موقعیت‌های اندازه‌گیری را تعریف می‌کند، نشان داده می‌شوند. اگر داده‌ها در n موقعیت جمع‌آوری شوند، ماتریس وزن دارای ابعاد $n \times n$ با مقادیر صفر در قطر می‌باشد (۱۴). از دو شاخص موران و جری می‌توان برای تشخیص خودهمبستگی مکانی استفاده کرد. براساس این دو شاخص، میزان شباهت نقاط داده (مثلاً i و j) به عنوان تابعی از فاصله (d_{ij}) اندازه‌گیری می‌شود.

شاخص موران^۱ (۳۱) براساس میزان انحراف از میانگین است و برای n مشاهده روی متغیر X در موقعیت‌های i و j به شکل معادله ۱ محاسبه می‌شود:

تشخیص خودهمبستگی مکانی در مشاهدات

مشاهدات در موقعیت‌های مختلف ممکن است مستقل نباشند. به عنوان مثال، مقادیر اندازه‌گیری‌ها در موقعیت‌های نزدیک ممکن است شبیه هم باشند تا در موقعیت‌های دور. این پدیده خودهمبستگی مکانی نامیده می‌شود. خودهمبستگی مکانی، همبستگی یک متغیر با خودش در مکان را اندازه‌گیری می‌کند. خودهمبستگی مکانی ممکن است مثبت یا منفی باشد. خودهمبستگی مکانی مثبت زمانی که مقادیر مشابه نزدیک یکدیگر باشند، اتفاق می‌افتد. در حالیکه خودهمبستگی مکانی منفی زمانی رخ می‌دهد که مقادیر غیرمشابه نزدیک یکدیگر باشند. قبل از مدل‌سازی، لازم است که وجود خودهمبستگی مکانی در متغیر وابسته بررسی شود.

جعفری و همکاران: تلفیق خودهمبستگی مکانی در...

. شاخص موران با مقادیر بزرگتر از ، خودهمبستگی مثبت و شاخص موران با مقادیر کمتر از خودهمبستگی منفی را نشان می‌دهد. به طور معمول، شاخص موران با افزایش فاصله روند کاهشی نشان می‌دهد تا اینکه به مقدار صفر یا کمتر از صفر یعنی عدم وجود خودهمبستگی مکانی می‌رسد (۹).

$$I = \frac{n}{S_0} * \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (1)$$

در معادله ۱، \bar{x} میانگین متغیر X ، عناصر ماتریس وزن و $S_0 = \sum_i \sum_j w_{ij}$ مجموع عناصر ماتریس وزن: مقادیر شاخص موران از -۱ تا +۱ تغییر می‌کند. در غیاب خودهمبستگی و بدون توجه به ماتریس وزن، مقدار مورد انتظار شاخص موران $-1/(n-1)$ است که به صفر تمایل دارد.

جدول (۱) متغیرهای کمکی انتخاب شده به ترتیب اهمیت

Table (1) List of the selected ancillary covariates, sorted by their importance.

Reference مرجع	Source منبع	SCORPAN factors فاکتورهای اسکورپن	Parameter پارامتر	Abbreviation علامت اختصاری
https://asf.alaska.edu/		r	DEM مدل رقومی ارتفاع	Elevation
(۱۰)	DEM مدل رقومی ارتفاع	r, c	Multiresolution index of valley bottom flatness شاخص تفکیک همواری کف دره	MrVBF
(۱)		r	Valley depth عمق دره	ValD
(۶)		r	Saga wetness index شاخص خیسگی ساگا	SWI
(۴۳)		o	Perpendicular vegetation index شاخص پوشش گیاهی عمودی	PVI
(۲)	RS تصویر ماهواره-ای	s, p	Salinity index = B4/B8A شاخص شوری	SI
(۳۹)		s, p	The first principle component of the Sentinel-2 bands اولین مولفه اصلی باندهای تصویر سنتینل ۲	PC1
(۲۵)	Climate data داده اقلیمی	c	Mean annual precipitation میانگین سالانه بارندگی	PreSum
(۲۵)		c	Average annual temperature میانگین سالانه درجه حرارت	Tave

(مواد مادری p: parent material، پستی و بلندی r: relief، موجودات زنده o: organism، اقلیم c: climate، خاک s: soil)

ترسیم این بردارهای ویژه، الگوی مکانی خودهمبستگی مکانی را نشان می‌دهد. این روش مبتنی بر تجزیه تابع ویژه^۴ ماتریس‌های اتصال مکانی است، روشی نسبتاً جدید و هنوز ناآشنا برای توصیف الگوهای مکانی در داده‌های پیچیده است. یک رویکرد بسیار مشابه، به نام فیلتر بردار ویژه^۵، توسط دینیز فیلهو و بینی^۶ (۸) برای توضیح عدم استقلال فیلوژنتیک در داده‌های یولوژیکی ارائه شد. با توجه به ماهیت غیر اقلیدسی ماتریس‌های مکانی، مقادیر ویژه منفی و مثبت تولید می‌شود. بردارهای ویژه با مقادیر ویژه مثبت، خودهمبستگی مکانی مثبت و بردارهای ویژه با مقادیر ویژه منفی، خودهمبستگی مکانی منفی را نشان می‌دهد.

در این مطالعه، یک رویکرد بردار ویژه مبتنی بر فاصله بکار گرفته شد که در چهار مرحله به طور خلاصه آورده می‌شود: (۱) محاسبه یک ماتریس فاصله اقلیدسی (جغرافیایی) در بین واحدهای نمونه برداری: $D = [d_{ij}]$

(۲) انتخاب مقدار آستانه t و ساخت یک ماتریس اتصال با استفاده از معادله ۳:

$$W = [w_{ij}] = \begin{cases} 0 & \text{if } d_{ij} = t \\ 0 & \text{if } d_{ij} > t \\ [1 - (d_{ij}/4t)^2] & \text{if } d_{ij} \leq t \end{cases} \quad (3)$$

که در آن، t به عنوان حداکثر فاصله‌ای انتخاب می‌شود که اتصالات را بین تمام واحدهای نمونه برداری حفظ می‌کند. باید توجه کرد که در طرح‌های نمونه برداری شبکه‌ای منظم و نمونه برداری‌های با شکل هندسی نامشخص مقدار t تغییر می‌کند. اگر نمونه برداری منظم باشد، مقدار آن یک می‌شود.

(۳) محاسبه بردارهای ویژه ماتریس شباهت

(۴) انتخاب بردارهای ویژه به عنوان پیش‌بینی کننده‌های مکانی در یک مدل خطی یا خطی تعمیم یافته. در اینجا، یک روش انتخاب مدل که میزان خودهمبستگی مکانی را در باقیمانده‌های مدل به حداقل می‌رساند، استفاده شد. در این رویکرد، بردارهای ویژه به یک مدل اضافه می‌شوند تا زمانی که خودهمبستگی مکانی در باقیمانده‌ها، یا همان شاخص موران، معنی دار نباشد (۹).

شاخص جری^۱ (۱۱) براساس انحراف در پاسخ هر مشاهده با یکدیگر است:

$$C = \frac{n-1}{2S_0} * \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

شاخص جری از صفر (ماکزیم خودهمبستگی مثبت) تا مقدار نامشخصی بزرگتر از یک تغییر می‌کند. مقدار یک نشان دهنده عدم خودهمبستگی مکانی است. مقادیر شاخص جری کمتر از ۱ نشان دهنده افزایش خودهمبستگی مکانی مثبت است، در حالی که مقادیر شاخص جری بالاتر از ۱ نشان دهنده افزایش خودهمبستگی مکانی منفی است (۹).

آنالیز رگرسیون و نقشه‌های بردار ویژه مکانی^۲ (SEVM)

ابتدا مدل رگرسیون غیر مکانی حداقل مربعات معمولی^۳ (OLS) برای پیش‌بینی توزیع مکانی متغیر وابسته شوری خاک با کمک متغیرهای مستقل مهم شامل متغیرهای حاصل از مدل رقمی ارتفاع، تصاویر ماهواره‌ای و اقلیم برازش داده شد. در این مرحله خودهمبستگی مکانی لحاظ نمی‌شود. سپس همان مدل تحت عنوان رگرسیون مکانی علاوه بر متغیرهای مهم محیطی، با محاسبه فیلترهای مکانی از طریق بردارهای ویژه مکانی به عنوان متغیرهای مستقل برازش داده می‌شود.

نقشه‌های بردار ویژه مکانی مبتنی بر این ایده است که آرایش مکانی نقاط داده را می‌توان به عنوان متغیرهای توصیفی در نظر گرفت، چرا که اثرات مکانی را در مقیاس‌های مکانی مختلف ثبت می‌کند. به عنوان مثال، در مقیاس‌های وسیع، زمین شناسی زیرین و اقلیم منطقه‌ای بر فرآیندهای ژئودولوژیکی تأثیر می‌گذارد و ساختار مکانی توزیع خاک را باعث می‌شود. در مقیاس‌های کوچک‌تر، فرایندهای بهم‌ریختگی (به ویژه، بهم‌ریختگی‌های جانوری و گیاهی)، می‌توانند عوامل محلی حاکم باشند. در طول تجزیه و تحلیل، آن دسته از بردارهای ویژه‌ای که خودهمبستگی مکانی در باقیمانده‌ها را به بهترین نحو کاهش می‌دهند، به عنوان پیش‌بینی کننده‌های مکانی انتخاب می‌شوند.

4- eigenfunction
5- eigenvector filtering
6- Diniz-Filho and Bini

1- Geary's C
2- Spatial Eigen Vector Mapping (SEVM)
3- Ordinary least square (OLS)

داده‌های شوری خاک منطقه میزان چولگی و کشیدگی بالایی را نشان می‌دهد که بیانگر توزیع غیرنرمال آن است (۱۵). تغییرپذیری بالای این ویژگی خاک تأکیدی بر برهمکنش عوامل پیچیده و متعدد، از جمله فرآیندهای تشکیل خاک و شیوه‌های مدیریتی مختلف خاک‌ها می‌باشد (۴۶).

انتخاب متغیرهای محیطی مهم برای مدل‌سازی

متغیرهای مهم انتخاب شده بر اساس آنالیز همبستگی انجام شده شامل می‌باشد. همانطور که ملاحظه می‌گردد، در مجموع ۷ متغیر انتخاب شده که شامل چهار متغیر توپوگرافی و سه متغیر سنجش از دور می‌باشد. از میان متغیرهای توپوگرافی (استخراج شده از مدل رقومی ارتفاع)، شاخص همواری کف دره با درجه تفکیک بالا (MrVBF) بیشترین میزان اهمیت را دارد (همبستگی: ۰/۷۰). این ویژگی توپوگرافی، نقشی اساسی در شکل‌دهی خاک و تعیین الگوهای مکانی ویژگی‌های خاک از طریق کنترل آن‌ها بر حرکت آب و رسوب‌گذاری را دارد (۱۹)، (۴۱).

توزیع مکانی متغیر شوری خاک در منطقه مورد

مطالعه

نقشه توزیع مکانی شوری خاک در منطقه آبیگ قزوین نشان می‌دهد که در بخش‌های شمالی، شمال شرق و شمال غرب به سمت مرکز منطقه مورد مطالعه میزان شوری خاک کم بوده و بیشترین مقدار و محدودیت شوری در مناطق جنوب و جنوب شرقی دیده می‌شود (شکل ۳). در قسمت‌های شمالی، مقدار شوری حتی ممکن است کمتر از یک (dS/m) باشد و به سمت جنوب، مقدار آن افزایش پیدا می‌کند. شوری خاک از جمله خصوصیات است که به راحتی تحت تاثیر فاکتورها و فرایندهای خاکسازي دچار تغییرات زیادی می‌شود.

در منطقه مورد مطالعه، قسمت‌های شمالی با واحد فیزیوگرافی تپه و فلات با ارتفاع بیشتر همراه است و به سمت جنوب ارتفاع کم می‌شود که در این قسمت واحد فیزیوگرافی اراضی پست با آب زیرزمینی بالا مشخص شده است. به نظر می‌رسد تغییرات شوری خاک با تغییرات توپوگرافی همخوانی دارد چرا که از ارتفاعات شیبدار منطقه (قسمت‌های شمالی) به سمت اراضی پست (قسمت‌های جنوبی)، شوری خاک افزایش یافته است.

مزیت اصلی بردارهای ویژه مکانی برای محاسبه خودهمبستگی مکانی توانایی آنها در برآورد خودهمبستگی مکانی در مقیاس‌های چندگانه به طور همزمان است. چراکه ممکن است منابع موثر در تغییرات مکانی خاک در مقیاس‌های مختلف، متفاوت باشد (۵ و ۱۳). به عنوان مثال، در مقیاس‌های بزرگ، زمین‌شناسی و اقلیم عامل تاثیرگذار در ساختار مکانی توزیع خاک هستند. در حالیکه در مقیاس کوچک‌تر، فعالیت‌های متابولیکی پوشش گیاهی از فاکتورهای محلی موثر می‌باشد. به طور کلی، ترکیبی از عوامل مختلف ناشی از مقیاس‌های مکانی مختلف می‌تواند الگوی پیچیده‌ای از خودهمبستگی مکانی در پراکنندگی خصوصیات خاک ایجاد کنند (۲۴). در بردارهای ویژه مکانی (SEVM)، این فاکتورهای چندمقیاسی توسط مجموعه‌ای از بردارهای ویژه حاصل از تابع ویژه نمایش داده می‌شوند که شامل ماتریسی از فاصله اقلیدسی بین مشاهدات است. یعنی بردارهای ویژه با مقادیر ویژه بزرگ یا کوچک به ترتیب به عنوان معیاری برای تغییرات در مقیاس وسیع یا ریز داده‌ها عمل می‌کنند. در مطالعه حاضر، مقادیر ویژه که مثبت بودند و همچنین منجر به حداقل شاخص موران باقیمانده‌ها شدند، انتخاب گردید. اجرای مدل رگرسیون غیرمکانی حداقل مربعات معمولی، بردارهای ویژه مکانی و رگرسیون مکانی در نرم‌افزار R با کمک بسته‌های *spdep* و *spatialreg* انجام گردید.

مقایسه نتایج مدل

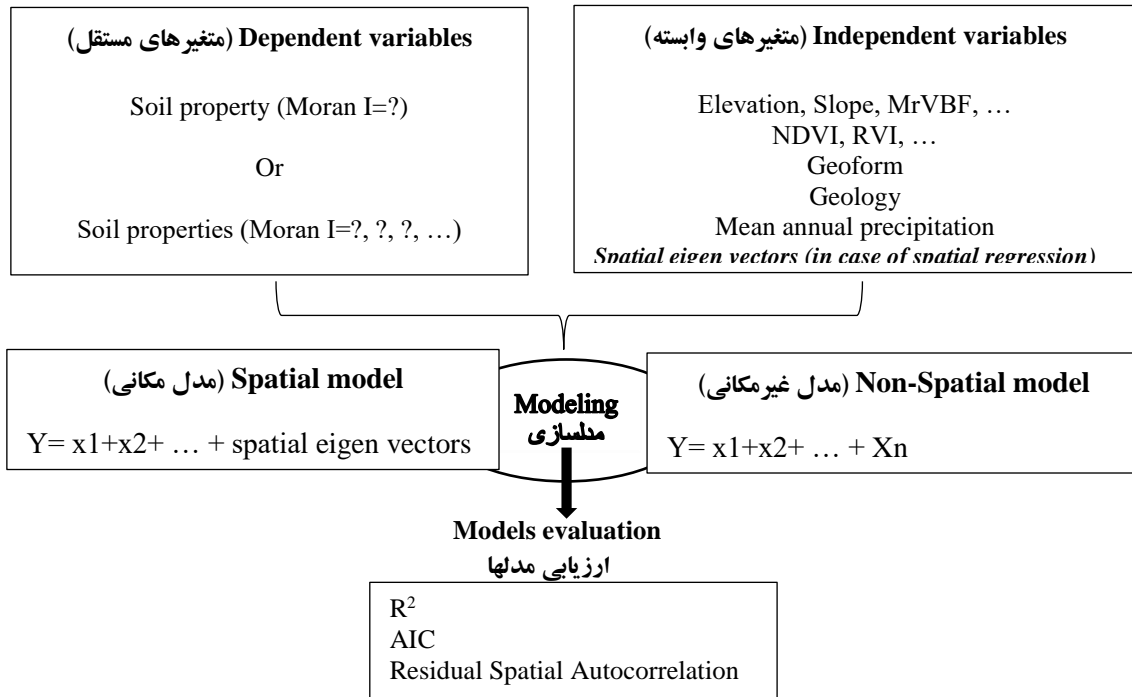
باید توجه داشت که هدف مقاله حاضر، مقایسه عملکرد مدل‌های تخمین نیست، بلکه هدف بهبود مدل‌سازی با در نظر گرفتن خودهمبستگی مکانی است. بنابراین مقایسه خروجی مدل غیرمکانی حداقل مربعات معمولی و مدل رگرسیون مکانی با معیارهایی همچون R^2 ، AIC^1 ، خودهمبستگی باقیمانده‌ها و $RMSE^2$ انجام گرفت. مراحل انجام تحقیق در شکل ۲ نشان داده شده است.

جدول ۲ توصیف آماری داده‌های شوری خاک سطحی را نشان می‌دهد که حاکی از تغییرپذیری بالایی در شوری خاک منطقه مورد مطالعه است (CV بیش‌تر از ۳۵٪). براین اساس،

1- Akaike Information Criteria

2- Root Mean Square Error

نتایج و بحث



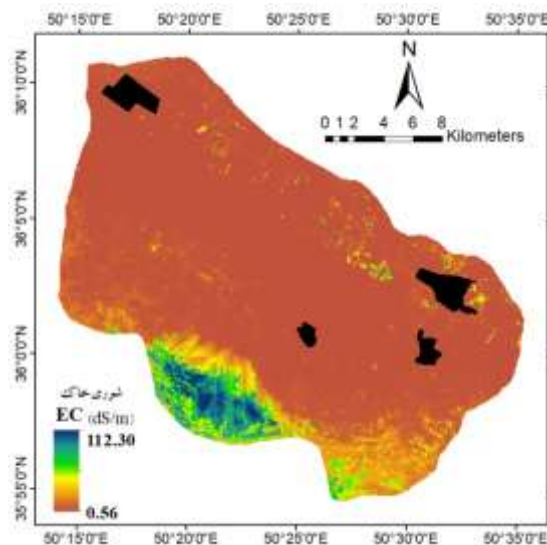
شکل (۲) روندنمای مراحل انجام پژوهش

Figure (2) Flowchart of methodology used in this study

جدول (۲) توصیف آماری شوری خاک سطحی

Table (1) Statistical description of surface soil salinity

کشیدگی	چولگی	ضریب تغییرات (%)	انحراف معیار	میانگین	حداکثر	حداقل	واحد	ویژگی خاک
Kurt.	Skew.	CV	SD	Mean	Max	Min	Unit	Soil property
40.5	5.9	326.4	17.3	5.3	163.7	0.1	dS/m	Soil salinity شوری خاک



شکل (۳) نقشه توزیع مکانی شوری خاک (قسمتهایی که با رنگ سیاه مشخص شدند، منطقه مسکونی است)

Figure (3) spatial distribution map of soil salinity (The parts marked in black are the residential areas)

مشخص است که در نقشه توزیع باقیمانده‌های مدل غیرمکانی، خوشه‌های (که با بیضی به رنگ سبز مشخص شده است) بیشتر و بزرگتری قابل شناسایی است که نشان از وجود خودهمبستگی مکانی در باقیمانده‌های مدل است. دورمن و همکاران^۲ (۹) حضور خودهمبستگی مکانی به صورت خوشه‌هایی در نقشه توزیع باقیمانده‌های مدل حداقل مربعات معمولی را گزارش کردند. حضور خودهمبستگی مکانی در باقیمانده‌های یک مدل نشان می‌دهد مدل قادر به حذف وابستگی مکانی نیست که ممکن است به دلیل در نظر نگرفتن یک متغیر کمکی مهم در مدلسازی باشد.

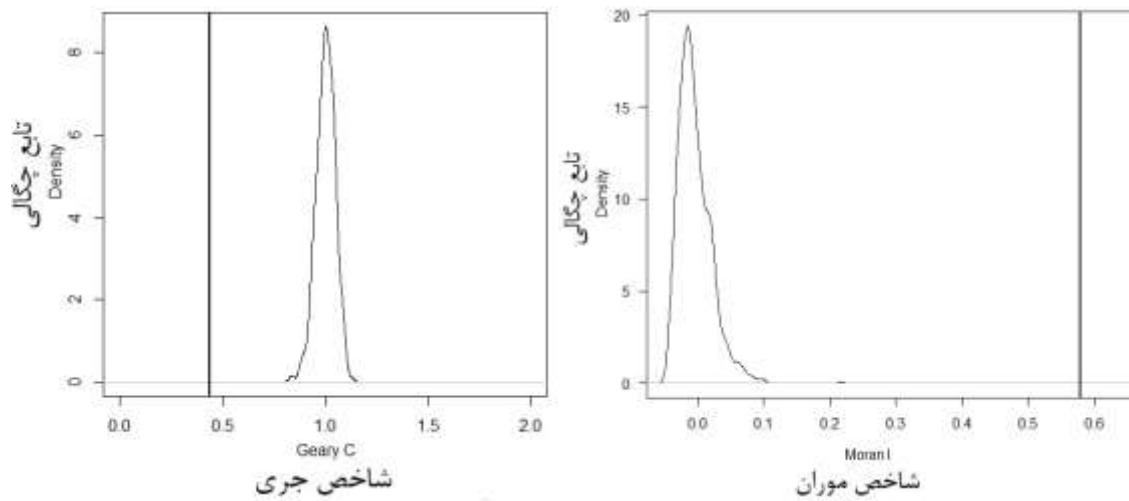
تشخیص خودهمبستگی مکانی در داده‌ها

شکل ۴، نمودار شاخص موران و جری محاسبه شده برای داده‌های شوری خاک را نشان می‌دهد. خطوط عمودی در شکل ۴، نشان‌دهنده مقادیر شاخص موران و جری به دست آمده در این مطالعه است که برای شاخص موران ۰/۵۷ و شاخص جری ۰/۴ به دست آمد. براساس هر دو شاخص، ویژگی شوری خاک در منطقه مورد نظر دارای خودهمبستگی مکانی است. منحنی‌های رسم شده در این شکل، محدوده بدون خودهمبستگی مکانی را نشان می‌دهد. اگر مقادیر شاخص‌های موران و جری در این محدوده قرار بگیرد، داده‌ها دارای خودهمبستگی مکانی نیستند و تغییرات آنها مشتق از مکان است.

عملکرد مدل

معیارهای R^2 ، AIC، خودهمبستگی باقیمانده‌ها و RMSE برای مقایسه خروجی مدل غیرمکانی حداقل مربعات معمولی و رگرسیون مکانی در جدول‌های ۳ و ۴ نشان داده شده است. در مدل رگرسیون مکانی با وارد کردن خودهمبستگی مکانی در مقایسه با مدل غیرمکانی نتایج بهبود یافت. با لحاظ کردن خودهمبستگی مکانی، مقدار R^2 افزایش، درحالی‌که مقادیر AIC، خودهمبستگی مکانی باقیمانده‌ها و RMSE کاهش یافت. نتایج مطالعه حاضر با نتایج کیم و همکاران^۱ (۲۴) همخوانی دارد. نتایج آنها عملکرد بهتر مدل‌های مکانی در مقایسه با مدل‌های غیرمکانی در پیش‌بینی خصوصیات از جمله pH، CEC و برخی کاتیون‌ها نشان داد که با کاهش مقدار خطا و خودهمبستگی مکانی باقیمانده‌ها در مدل مکانی همراه بود.

نقشه‌های توزیع باقیمانده‌های مدل غیرمکانی حداقل مربعات معمولی و مدل رگرسیون مکانی در شکل ۵ نشان داده شده است. این نقشه‌ها از نظر علامت مکانی باقیمانده‌ها و توزیع خودهمبستگی مکانی که به شکل خوشه‌هایی قابل تشخیص است، تفاوت دارند. هرچه خوشه‌ها و کلاسترها (رنگ قرمز یا آبی) بزرگتر باشد، نشان‌دهنده حضور خودهمبستگی مکانی در باقیمانده‌ها است. با مقایسه دو شکل مربوط به رگرسیون مکانی و حداقل مربعات معمولی



شکل (۴) شاخص موران و شاخص جری محاسبه شده برای داده‌های شوری خاک
Figure (4) Moran's I and Geary's C for soil salinity data

جدول (۳) نتایج مدل غیرمکانی حداقل مربعات معمولی و رگرسیون مکانی برای شوری خاک

Table (2) Results of the non-spatial model of ordinary least squares and the spatial regression for soil salinity

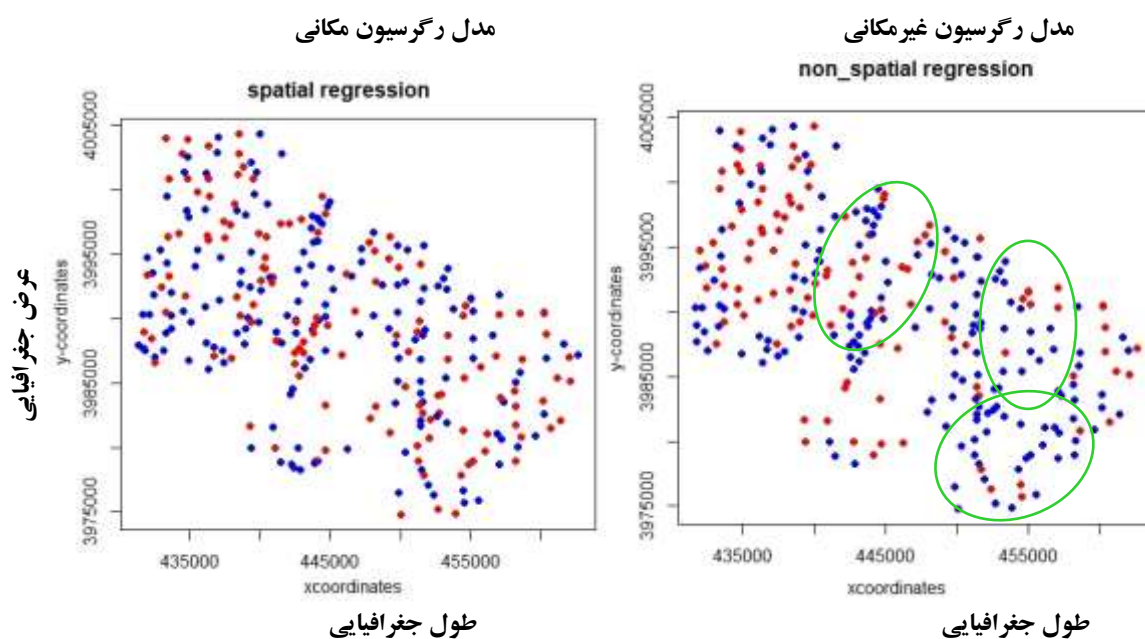
Resid. I ^a	AIC	R ²	مدل Model
0.31	1491.1	0.33	رگرسیون غیرمکانی حداقل مربعات معمولی Non-spatial ordinary least squares
-0.01	1360.1	0.59	رگرسیون مکانی spatial regression

a: مقدار خودهمبستگی مکانی در باقیمانده‌های مدل (spatial autocorrelation in the residuals of model)

جدول (۴) کاهش RMSE پس از ترکیب خودهمبستگی مکانی در تخمین شوری خاک

Table (3) Decreases in RMSE after incorporating spatial autocorrelation in prediction of soil salinity

درصد کاهش خطا %reduction	رگرسیون مکانی spatial regression	رگرسیون غیرمکانی حداقل مربعات معمولی Non-spatial ordinary least squares
21.9	0.775	0.994



شکل (۵) توزیع باقیمانده‌های مدل غیر مکانی حداقل مربعات معمولی و رگرسیون مکانی (رنگ قرمز و آبی: به ترتیب مقادیر منفی و مثبت خودهمبستگی مکانی را نشان می‌دهد).

Figure (5) Residual maps of the non-spatial model of ordinary least squares and the spatial regression (Blue and red indicate positive and negative values, respectively)

کردن خودهمبستگی مکانی به عنوان متغیر کمکی، مقدار R^2 افزایش، درحالی‌که مقادیر AIC، خودهمبستگی مکانی باقیمانده‌ها و RMSE کاهش یافت. از نقطه نظر آماری، خودهمبستگی مکانی، ممکن است به عنوان مشکل تلفی شود، اما در واقع، این یک ویژگی اساسی و فراگیر در تقریباً تمام داده‌های خاک است که به منظور جلوگیری از برازش نامناسب مدل و پیش‌بینی‌های مغرضانه باید تعیین شود و در مدلسازی لحاظ گردد.

نتیجه‌گیری

این مطالعه به منظور بررسی تاثیر خودهمبستگی مکانی روی نتایج حاصل از مدلسازی تخمین شوری خاک انجام گرفت. پیش‌بینی شوری خاک توسط مدل غیر مکانی حداقل مربعات معمولی (بدون در نظر گرفتن خودهمبستگی مکانی) و مدل رگرسیون مکانی (با در نظر گرفتن خودهمبستگی مکانی) صورت گرفت. نتایج حاکی از بهبود عملکرد مدل رگرسیون مکانی در مقایسه با مدل غیر مکانی حداقل مربعات معمولی می‌باشد. در مدل مکانی با لحاظ

References

1. Abdel-Kader, F.H. 2011. Digital soil mapping at pilot sites in the northwest coast of Egypt: a multinomial logistic regression approach. *Egypt. Journal of Remote Sensing, Space Science* 29–40.
2. Allbed, A., and Kumar, L. 2013. Soil Salinity Mapping and Monitoring in Arid and Semi-Arid Regions Using Remote Sensing Technology: A Review. *Advances in Remote Sensing*, 2, 373–385.
3. Behrens, T., Schmidt, K., Viscarra-Rossel, R. A., Gries, P., Scholten, T., and MacMillan, R. A. 2018b. Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, 69, 757-770.
4. Bjørn Møller A., Beucher A.M., Pouladi N., and Humlekrog Greve M. 2020. Oblique geographic coordinates as covariates for digital soil mapping. *SOIL*, 6, 269–289.
5. Borcard, D., and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modeling*, 153, 51–68.
6. Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991-2007.
7. Davies B.E, and Gamm S.A. 1970. Trend surface analysis applied to soil reaction values from Kent, England. *Geoderma* 3(3), 223-231.
8. Diniz-Filho, J.A., and Bini, L.M. 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14, 177-185.
9. Dormann, C., F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609-628.
10. Gallant, J.C., and Dowling, T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resource Research*. 39, 1347–1359.
11. Geary, R.C. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*. 5 (3), 115–145.
12. Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuyse, S., Mboga, N.O., Wol, E., and Kalogirou, S. 2019. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 1, 1-12.
13. Griffith, D.A. 2013. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Springer Science and Business Media.
14. Haining, R.P. 2001. Spatial Autocorrelation. *International Encyclopedia of the Social and Behavioral Sciences*. 14763-14768.
15. Hair, J.F., Hult, G.T.M., Ringle, C.M., and Sarstedt, M. 2022. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3rd ed.). Thousand Oaks, CA: Sage.
16. Harnett, P.R., Mountain, G.D., Barnett, M. E. 1978. Spatial filtering applied to remote sensing imagery. *Optica Acta*, 25, 801-809.
17. Hawkins, B.A. 2012. Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography*, 39:1-9.

18. Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., and Graler, B. 2018. Random forest as a generic framework for predictive modeling of spatial and spatiotemporal variables. *PeerJ*, 6, e5518.
19. Hu, C., Wright, A.L., and Lian, G. 2019. Estimating the Spatial Distribution of Soil Properties Using Environmental Variables at a Catchment Scale in the Loess Hilly Area, China. *International Journal of Environmental Research and Public Health*, 16(3), 491.
20. Inakwu O.A.O., Crawford M., and McBratney A. 2006. Digital Mapping of Soil Attributes for Regional and Catchment Modelling, using Ancillary Covariates, Statistical and Geostatistical Techniques. *Developments in Soil Science*, Chapter 32. Volume 31, pp: 437-453.
21. Jiang, Z., Li, Y., Shekhar, S., Rampi, L., and Knight, J. 2017. Spatial ensemble learning for heterogeneous geographic data with class ambiguity: A summary of results. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 23* (pp. 1-10). ACM.
22. Khamoshi, S.E., Sarmadian, F., and Omid, M. 2023. Predicting and Mapping of Soil Organic Carbon Stock Using Machine Learning Algorithm, *Iranian Journal of Soil and Water Research*, 53 (11), 2671-2681. (in Persian with English abstract)
23. Khamoshi, S.E., Sarmadian, F., Keshavarzi, A. 2019. Digital soil mapping Using Random Forests and Land Suitability Evaluation for Abyek Region, Qazvin Province. *Journal of Range and Watershed Management*. 71, 885-899. (in Persian with English abstract)
24. Kim, D., Hirmas, D.R., McEwan, R.W., Mueller, T.G., Park, S.J., Šamonil, P., Thompson, J.A., and Wendroth, O. 2016. Predicting the influence of multi-scale spatial autocorrelation on soil–landform modeling. *Soil Science Society of American Journal*, 80: 409–419.
25. Krivoruchko, K., and Gribov, A. 2019. Evaluation of empirical Bayesian kriging. *Spatial Statistics*, 32,100368.
26. Kuhn, I., and Dormann, C.F. 2012. Less than eight (and a half) misconceptions of spatial analysis. *Journal of Biogeography*, 39: 995-998.
27. Kuhn, I., Nobis, M.P., and Durka, W. 2009. Combining spatial and phylogenetic eigenvector filtering in trait analysis. *Global Ecology and Biogeography*, 18, 745-758.
28. Kpade O.L.H., Stendahl J., Lundblad M., and Karlton E. 2021. Predicting the spatial distribution of soil organic carbon stock in Swedish forests using a group of covariates and site-specific data. *SOIL*, 7, 377–398.
29. Lagacherie, P., and McBratney, A. 2006. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. *Developments in Soil Science*, 31, 3-22.
30. Momtazi Burojeni, M., and Sarmadian, F. 2023. Spatial prediction of soil classes using C5.0 boosted decision tree model Abyek Area. *Journal of Range and Watershed Management*, 75 (4), 553 – 572. (in Persian with English abstract)
31. Moran, P.A.P. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika*. 37 (1), 17–23.
32. Mousavi, S.R., Sarmadian, F., Omid, M., and P. Bogaert. 2021. Digital Modeling of Three-Dimensional Soil Salinity Variation Using Machine Learning Algorithms in Arid and Semi-Arid lands of Qazvin Plain. *Iranian Journal of Soil Research*, 52, 1915-1929. (in Persian with English abstract)
33. Mousavi, S.R., Sarmadian, F., Omid, M., and P. Bogaert. 2021. Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Journal of Water and Soil*. 35, 719-734. (in Persian with English abstract)
34. Mousavi, S.R., Sarmadian, F., Omid, M., and P. Bogaert. 2022. Application of Machine Learning Models in Spatial Estimation of Soil Phosphorus and Potassium in Some Parts of Abyek Plain. *Iranian Journal of Soil Research*, 35, 397-411. (in Persian with English abstract)

35. Mousavi, S.R., Sarmadian, F., Angelini, M. E., Bogaert, P., and Omid, M. 2023. Cause-effect relationships using structural equation modeling for soil properties in arid and semi-arid regions. *Catena*, 232, 107392.
36. Mousavi, S.R., Sarmadian, F., Omid, M., and Bogaert, P. 2022. Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement*, 201, 111706.
37. Mousavi, S.R., Sarmadian, F., Rahmani, A. 2020. Modelling and Prediction of Soil Classes Using Boosting Regression Tree and Random Forests Machine Learning Algorithms in Some Part of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 50, 2525-2538. (in Persian with English abstract)
38. Neyestani, M., Sarmadian, F., Jafari, A., Keshavarzi, A., and Sharififar, A. 2021. Digital mapping of soil classes using spatial extrapolation with imbalanced data. *Geoderma Regional*, 26, e00422.
39. Nield, S.J., Boettner, J.L., and Ramsey, R.D. 2007. Digital mapping gypsic and nitric soil areas using Landsat ETM data. *Soil Science Society of America Journal* 71:245–252.
40. Rahmani, A., Sarmadian, F., and Arefi, H. 2023. Digital modeling and prediction of soil subgroup classes using deep learning approach in a part of arid and semi-arid lands of Qazvin Plain, *Iranian Journal of Soil and Water Research*, 53 (11), 2477-2499. (in Persian with English abstract)
41. Rasaei, Z., Rossiter, D.G., and Farshad, A. 2020. Rescue and renewal of legacy soil resource inventories in Iran as an input to digital soil mapping. *Geoderma regional*, 21, e00262.
42. Rezaie, G., Sarmadian, F., Mohammadi Torkashvand, A., Seyedmohammadi, J., and Marashi Aliabadi M. 2023. Digital Mapping of Surface and Subsurface Soil Organic Carbon and Soil Salinity Variation in a Part of Qazvin Plain (Case Study: Abyek and Nazarabad Regions). *Journal of Water and Soil*, 37, 315-331. (in Persian with English abstract)
43. Richardson, A. J., and Wiegand, C. L. 1977. Distinguishing vegetation from soil background information. *Photogrammetric engineering and remote sensing*, 43(12), 1541-1552.
44. Romanski, P., Kothhoff, L., and Schratz, P. 2023. FSelector R Package. CRAN.
45. Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., and Soil Survey Staff. 2021. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
46. Selmy, S., Abd El-Aziz, S., El-Desoky, A., and El-Sayed, M. 2022. Characterizing, predicting, and mapping of soil spatial variability in Gharb El-Mawhoub area of Dakhla Oasis using geostatistics and GIS approaches. *Journal of the Saudi Society of Agricultural Sciences*, 21(6), 383-3967
47. Sinha, P., Gaughan, A.E., Stevens, F.R., Nieves, J.J., Sorichetta, A., and Tatem, A.J. 2019. Assessing the spatial sensitivity of a random forest model: Application in gridded population modeling. *Computers, Environment and Urban Systems*, 75, 132-145.
48. Soil Survey Staff. 2022. Keys to Soil Taxonomy, 13th ed. USDA-Natural Resources Conservation Service.
49. Van Wambeke, A.R. 2000. The Newhall Simulation Model for estimating soil moisture & temperature regimes. Conservation Service: Department of Crop and Soil Sciences Cornell University, Ithaca, NY USA.