

Random Forest Algorithm Training and Test for Monitoring Soil Salinity Changes in Pistachio Orchards

Abstract

A total of 720 soil samples collected from 240 sampling locations within pistachio orchards in the provinces of Yazd and Khorasan Razavi were used for training and testing the Random Forest algorithm. The auxiliary variables used in this modeling included the median values of 32 surface reflectance variables and spectral indices derived from Sentinel-2 satellite images, covering the period from March 1 to October 1 of the sampling year. These variables were extracted for the sampling points using the Google Earth Engine platform. The Random Forest model was developed and optimized through coding in the R environment. The Leave-One-Out Cross-Validation (LOOCV) method was used to validate the model. After identifying and removing outliers, 191 points remained for retraining and retesting the Random Forest algorithm. The RMSE was 1.1 dS/m in the training set and 2.6 dS/m in the test set. The R^2 value was 93% in both datasets. This algorithm was then used to predict changes in soil salinity in the studied areas and years. Based on the results, in the Tanour Lahour drainage project, Class 3 of salinity is decreasing while Class 2 land area is increasing. In the Mortazieh farm, Class 3 salinity has decreased but shifted to Class 4. In the Rezaei farm, Class 4 land area is significantly increasing. Likewise, in the Dadyar farm in Khorasan Razavi Province, Class 4 salinity area has relatively increased. The results of this study demonstrated that the Random Forest algorithm is capable of successfully predicting soil salinity changes within the study areas and similar regions.

Keywords: Google Earth Engine, Modeling, Machine Learning, Sentinel-2, Spectral Indices

Introduction

Given the limitations of traditional soil salinity assessment methods, alternative techniques such as remote sensing (RS) have been used to predict soil salinity and sodicity in unsampled areas. Vegetation indices (VIs), as indirect indicators of salinity, have been used to assess soil salinity through their negative effects on crop growth and plant stress. On the other hand, salinity indices, which serve as direct indicators of salinity, highlight the spectral reflectance of salt crusts on the soil surface, especially in the visible and near-infrared (NIR) regions of the electromagnetic spectrum. The scientific literature includes many successful estimations of soil properties using various soil sensing technologies. Despite significant advances in developing soil salinity prediction models, such models have not been widely developed for pistachio orchards, which typically lack full vegetation cover. Therefore, this study aimed to develop a Random Forest model to predict soil salinity using auxiliary variables derived from satellite imagery.

Materials and Methods

For this study, 669 soil samples collected from 223 sampling locations within four large pistachio-growing areas in Yazd and Khorasan Razavi provinces were used to train and test the Random Forest algorithm. At each sampling location, soil was collected down to a depth of 90 cm in 30 cm increments. The ground-truth data corresponded to the years 2018, 2022, and 2023. To train the Random Forest algorithm, spectral reflectance values and vegetation indices derived from Sentinel-2 satellite images were used. The surface reflectance values and spectral indices for the sampling points were extracted using the Google Earth Engine (GEE) platform. The Random Forest algorithm was trained, tested, and validated using code written in the R programming environment. The spectral indices used in this study were important variables commonly used as auxiliary inputs for soil salinity modeling.

Results and Discussion

According to the correlation matrix, although there was no strong correlation between the main variable (soil salinity) and individual auxiliary variables, some of the auxiliary variables showed positive or negative correlations with one another. Due to multicollinearity among the auxiliary variables, stepwise regression models could not be used to determine the relationship between soil salinity and these variables. The Random Forest model was trained using 500 decision trees. The RMSE of the model was 1.10 dS/m for the training set and 2.59 dS/m for the testing set. The R^2 value was 0.932 for both sets. The spectral indices used in this study are related to vegetation stress conditions and thus can also reflect plant performance under stress. The algorithm was used to predict changes in soil salinity across the study regions and years. The results showed that in the Tanour Lahour drainage project, the area of Class

3 salinity land is decreasing while Class 2 land is increasing. In Mortazieh farm, Class 3 salinity has declined but shifted toward Class 4. In Rezaei farm, Class 4 salinity has increased significantly. Similarly, in Dadyar farm in Khorasan Razavi province, Class 4 salinity area has relatively increased. These findings indicate that the Random Forest algorithm is capable of successfully predicting soil salinity changes in the study areas and similar regions.

Conclusion

The results of this study demonstrated that the Random Forest model, utilizing variables extracted from Sentinel-2 satellite imagery, has a high capability for predicting soil salinity changes in pistachio orchards in Yazd and Khorasan Razavi provinces. The spatial analysis of salinity changes also showed that in certain areas, such as the Tanour Lahour drainage project, salinity has decreased, indicating the positive effects of management practices. In contrast, in areas like Rezaei farm, the increasing trend of salinity is alarming. These results highlight the importance of continuous monitoring and the use of advanced tools such as the integration of remote sensing data and machine learning algorithms for effective soil salinity management. This approach can also be applied in other saline agricultural regions to prevent further soil degradation and yield reduction.

Author Contributions

Conceptualization, Y. H. and F. D.; methodology, Y.H. and H.B.; software, Y.H.; validation, H. B., M. S. and M.M.; formal analysis, Y.H.; investigation, Y.H., F.D. and H.B.; resources, F.D. and M.M.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H., F.D. and H.B.; visualization, Y.H.; supervision, Y.H.; project administration, Y.H.. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Not applicable.

Acknowledgements

The authors would like to thank the National Salinity Research Center for support of the present study.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

آموزش و آزمون الگوریتم جنگل تصادفی برای پایش تغییرات شوری خاک باغات پسته

چکیده

تعداد ۷۲۰ نمونه خاک که از ۲۴۰ موقعیت نمونه‌برداری در داخل باغات پسته از دو استان یزد و خراسان رضوی جمع‌آوری شده بود برای آموزش و آزمون الگوریتم جنگل تصادفی مورد استفاده قرار گرفتند. متغیرهای کمکی که برای این مدل‌سازی مورد استفاده قرار گرفتند عبارت بودند از مقادیر میانه ۳۲ متغیر بازتاب سطحی و شاخص‌های طیفی مستخرج از تصاویر ماهواره‌ای سنتینل ۲ از اول مارس تا اول اکتبر سال نمونه‌برداری بود که توسط پلتفرم گوگل ارث انجین برای نقاط نمونه‌برداری استخراج شده بود. مدل جنگل تصادفی از طریق کد نویسی در محیط R توسعه و بهینه‌سازی شد. برای اعتبارسنجی مدل از روش کنار گذاشتن یک نقطه استفاده شد. پس از شناسایی و حذف نقاط پرت، تعداد ۱۹۱ نقطه برای آموزش و آزمون مجدد الگوریتم جنگل تصادفی باقی ماند. مقدار RMSE در دسته آموزش ۱/۱ و در دسته آزمون ۲/۶ دسی‌زیمنس بر متر بود. مقدار R^2 در هر دو دسته ۹۳ درصد بود. این الگوریتم برای پیش‌بینی تغییرات شوری خاک در مناطق و سال‌های مورد مطالعه استفاده شد. بر اساس این نتایج در پروژه زهکشی تنور لاهور، مساحت اراضی کلاس ۳ در حال کاهش و اراضی کلاس ۲ در حال افزایش است. در مزرعه مرتاضیه شوری کلاس ۳ کاهش یافته ولی در عوض به کلاس ۴ افزوده شده است. در مزرعه رضایی نیز به شدت اراضی کلاس ۴ در حال افزایش است. همچنین در مزرعه دادیار در استان خراسان رضوی نیز بطور نسبی اراضی کلاس ۴ شوری افزایش یافته است. نتایج این تحقیق نشان داد که الگوریتم جنگل تصادفی بطور موفقیت‌آمیزی قادر به پیش‌بینی تغییرات شوری خاک در محدوده‌های مورد مطالعه و مناطق مشابه می‌باشد.

کلمات کلیدی: سنتینل ۲، شاخص‌های طیفی، گوگل ارث انجین، مدل‌سازی، یادگیری ماشین

مقدمه

تخمین زده می‌شود که ۴۲۴ میلیون هکتار از خاک رویی و ۸۳۳ میلیون هکتار از خاک زیرین اراضی تحت کشت دنیا تحت تاثیر نمک باشند. از خاک‌های رویی ۸۵ درصد شور، ۱۰ درصد سدیمی و ۵ درصد شورسدیمی هستند و از خاک‌های زیرین ۶۲ درصد شور، ۲۴ درصد سدیمی و ۱۴ درصد شورسدیمی هستند (FAO, 2024). درواقع این تخمین‌ها در بهترین حالت برآوردی تجربی هستند چرا که به دلیل ذات پیچیده و پویای زمانی و مکانی شوری خاک (Scudiero et al., 2014) مقادیر مطلق آن در حال تغییر است.

شوری خاک از دو منبع سرچشمه می‌گیرد: حمل و نقل توسط آب یا باد، یا فرآیندهای طبیعی مانند هوازدگی (که به عنوان شوری/سدیمی اولیه شناخته می‌شود). علاوه بر این، عواملی مانند توپوگرافی، آبیاری و شوری زمین‌های خشک که ناشی از جنگل‌زدایی، چرای بیش از حد و قطع درختان است نیز به شوری خاک کمک می‌کنند (Hassani et al., 2020). این فرآیندهای پویا تأثیرات منفی بر کیفیت خاک، محصولات کشاورزی، هیدرولوژی، ژئوشیمی و جنبه‌های مختلف اقتصادی و اجتماعی دارند (Albed and Kumar, 2013). شوری خاک همچنین در برخی شرایط رابطه خاک و آب را بدتر کرده و باعث کاهش نفوذ، هدایت هیدرولیکی، نگهداشت، زهکشی و اثرات یونی خاص بر گیاهان می‌شود (Lin and Banuelos, 2015).

با توجه به محدودیت‌های روش‌های سنتی، تکنیک‌های جایگزینی مانند ژئواستاتیک (Brunner et al., 2007) و سنجش از دور (RS) (Taghizadeh-Mehrjardi et al., 2021; Dong et al., 2019; Ren et al., 2019; Fourati et al., 2017;) برای

پیش‌بینی شوری و سدیمی خاک در مناطق نمونه‌برداری نشده استفاده شده‌اند. در میان حسگرهای ماهواره‌ای مختلف که برای نقشه‌برداری شوری و تولید نقشه‌های هدایت الکتریکی خاک استفاده می‌شوند، تصاویر ماهواره‌ای چندطیفی به‌طور گسترده‌ای مورد مطالعه قرار گرفته و به‌عنوان ابزاری بسیار امیدوارکننده برای این کار شناخته شده‌اند (Taghadosi et al., 2019).

شاخص‌های پوشش گیاهی (VI) به عنوان شاخص‌های غیرمستقیم شوری نیز برای ارزیابی شوری خاک از طریق اثرات منفی آن بر رشد محصولات کشاورزی و استرس گیاه استفاده شده‌اند. از سوی دیگر، شاخص‌های شوری که به‌عنوان شاخص‌های مستقیم شوری عمل می‌کنند، بازتاب طیفی پوسته‌های نمکی سطح خاک، به‌ویژه در محدوده مرئی و مادون قرمز نزدیک (NIR) طیف الکترومغناطیسی را برجسته می‌کنند (Gorji et al., 2017).

ادغام تکنیک‌های نقشه‌برداری دیجیتال خاک (DSM) و یادگیری ماشین برای تخمین ویژگی‌های خاک نیز باعث گسترش کاربرد این روش‌ها در پیش‌بینی شوری خاک شده است (Peng et al., 2019). در مورد استفاده از ادغام متغیرهای محیطی با یادگیری ماشین برای پیش‌بینی شوری خاک، Haq et al., (2023) مطالعه‌ای در پاکستان انجام دادند که از پنج مدل یادگیری ماشین همراه با داده‌های RS استفاده کردند. یافته‌های آن‌ها نشان داد که مدل جنگل تصادفی (RF) از نظر دقت پیش‌بینی بهتر از سایر مدل‌ها عمل کرده است. به طور مشابه، Xiao et al., (2023) اثربخشی سه مدل یادگیری ماشین جنگل تصادفی، ماشین بردار پشتیبان و بوستینگ گرادینل شدید را در پیش‌بینی شوری خاک در منطقه خشک شمال غربی چین ارزیابی کردند. نتایج آن‌ها نشان داد که مدل بوستینگ گرادینل شدید، به دنبال آن RF، نسبت به رگرسیون بردار پشتیبان، به‌ویژه زمانی که متغیرهای ورودی خاصی استفاده می‌شد، عملکرد بهتری داشتند.

Zhao et al., (2024) در مطالعه‌ای به بررسی کاربرد متغیرهای نهفته‌ی مشتق شده از روش حداقل مربعات جزئی (PLS) از طیف‌سنجی مرئی و مادون قرمز نزدیک (Vis-NIR)، در ترکیب با داده‌های سنجنش از دور (RS) و متغیرهای کمکی، برای پیش‌بینی هدایت الکتریکی (EC) و نسبت جذب سدیم (SAR) در منطقه شمالی سین‌کیانگ، چین پرداختند. نتایج مدل‌سازی نشان داد که سنجنش از دور و کاربری زمین به تنهایی پیش‌بینی‌کننده‌های قابل اعتمادی نبودند، اما افزودن ویژگی‌های توپوگرافی به طور قابل توجهی دقت پیش‌بینی هر دو شاخص EC و SAR را بهبود بخشید. Zare et al., (2021) در یک مطالعه امکان‌سنجی استفاده از سه روش مختلف برای پیش‌بینی شوری خاک سطحی را ارزیابی کردند: طیف‌سنجی بازتاب پراکنده مرئی و نزدیک مادون قرمز (VisNIR DRS)، طیف‌سنجی فلورسانس پرتوی ایکس قابل حمل (PXRF)، و سنجنش از دور (RS) دو منطقه شورزار در غرب تگزاس، ایالات متحده مورد ارزیابی قرار گرفتند. ترکیب سه تکنیک بالاترین قابلیت پیش‌بینی را داشت.

برخی تحقیقات اخیر نشان داده‌اند که افزایش تراکم نمونه‌برداری فضایی خاک به‌طور قابل توجهی عملکرد پیش‌بینی مدل‌های نقشه‌برداری دیجیتال خاک را افزایش می‌دهد (Lagacherie et al., 2020). سنجنش خاک یک جایگزین امیدبخش برای اندازه‌گیری مستقیم خاک است که تحت شرایط خاص اندازه‌گیری می‌تواند نمونه‌برداری‌های فضایی متراکم‌تری ارائه دهد. می‌توان سنجنش از راه دور (Mulder et al., 2011) و سنجنش نزدیک به خاک (Rossel et al., 2011) را متمایز کرد.

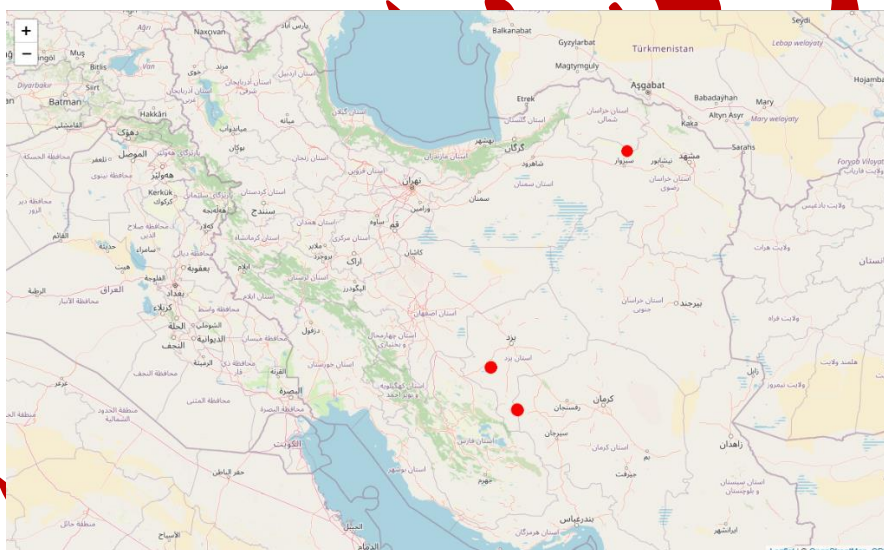
ادبیات علمی شامل تخمین‌های موفقیت‌آمیز زیادی از ویژگی‌های خاک با استفاده از فناوری‌های مختلف سنجنش خاک است، مانند طیف‌سنجی VIS-NIR-SWIR (Gomez and Coulouma, 2018)، طیف‌سنجی گاما (Zare et al., 2018)، رادار نفوذی زمین (Afshar et al., 2016)، تصاویر هوابرد هایپراسپکترال (Gholizadeh et al., 2018) و بازتاب‌سنج دامنه‌زمانی (Arsoy et al., 2013).

تحقیقات جاری درباره شوری خاک جهانی به دلیل دقت پایین مکانی داده‌های موجود (بیش از ۲۵۰ متر) و قوانین طبقه‌بندی نیمه کیفی (پنج طبقه‌ای) محدود است. برای رفع این محدودیت‌ها، Wang et al., (2024) چارچوبی جدید پیشنهاد کردند تا به‌طور کمی محتوای نمک خاک را در مقیاس جهانی در پنج منطقه اقلیمی با دقت ۱۰ متر تخمین بزنند.

بر اساس بررسی منابع انجام شده، هر چند پیشرفت‌های چشم‌گیری در زمینه توسعه مدل‌های پیش‌بینی شوری خاک صورت گرفته است، اما در زمینه باغات پسته که فاقد پوشش کامل هستند، این چنین مدل‌هایی توسعه نیافته‌اند. به همین دلیل این تحقیق با هدف توسعه مدل جنگل تصادفی برای پیش‌بینی شوری خاک با استفاده از متغیرهای کمکی مستخرج از تصاویر ماهواره‌ای انجام گرفت.

مواد و روش‌ها

برای انجام این پژوهش تعداد ۶۶۹ نمونه خاک که از ۲۲۳ موقعیت نمونه‌برداری در داخل ۴ مجموعه پسته‌کاری وسیع از دو استان یزد و خراسان رضوی جمع‌آوری شده بود برای آموزش و آزمون الگوریتم جنگل تصادفی مورد استفاده قرار گرفتند. شکل ۱ موقعیت نقاط نمونه‌برداری را در استان‌های یزد و خراسان رضوی نشان می‌دهد. موقعیت دو باغ پسته در استان یزد بسیار نزدیک به یکدیگر بود و در نتیجه روی نقشه بصورت یک نقطه نمایش داده شده است.



شکل ۱- موقعیت نقاط نمونه‌برداری مورد استفاده در پژوهش

هر موقعیت نمونه‌برداری تا عمق ۹۰ سانتی‌متری به فواصل ۳۰ سانتی‌متری نمونه‌برداری شده بود. داده‌های اندازه‌گیری زمینی مربوط به سال‌های ۲۰۱۸، ۲۰۲۲ و ۲۰۲۳ میلادی بودند. جدول ۱ بطور خلاصه اطلاعات زمانی- مکانی نقاط نمونه‌برداری را نشان می‌دهد.

جدول ۱- اطلاعات مکانی- زمانی نمونه‌های خاک جمع‌آوری شده

نام	شهرستان	استان	عرض جغرافیایی	طول جغرافیایی	سال	مساحت (هکتار)	تعداد نمونه
دادپار	جوین	خراسان رضوی	۳۶/۵۷	۵۷/۷۶	۲۰۲۲	۸۸	۷۵
مرتاضیه	تفت	یزد	۳۱/۲۶	۵۳/۷۲	۲۰۲۳	۴۸۳	۲۲۲
رضایی	تفت	یزد	۳۱/۲۶	۵۳/۷۴	۲۰۲۱	۷۵	۲۹۷
تنور لاهور	خاتم	یزد	۳۰/۱۷	۵۴/۵۲	۲۰۲۱	۵۷۴	۱۲۶

نمونه‌های خاک با اهداف متفاوتی جمع‌آوری شده بودند و در نتیجه تراکم نمونه‌برداری‌ها متفاوت بود.

برای آموزش الگوریتم جنگل تصادفی از مقادیر بازتاب طیفی و شاخص‌های پوشش گیاهی مستخرج از تصاویر ماهواره سنتینل ۲ استفاده شد.

دومین سری از ماهواره‌های سنتینل در ۲۳ ژوئن ۲۰۱۵ مأموریت خود را آغاز کرد. سنتینل ۲ ماهواره مدار قطبی است که مأموریت آن اخذ تصاویر با وضوح بالا برای پایش پدیده‌هایی همانند آب، خاک، پوشش گیاهی و... می‌باشد. این ماهواره از ۱۳ باند طیفی در محدوده‌های مرئی، مادون قرمز نزدیک و مادون قرمز طول موج کوتاه برخوردار می‌باشد. سنتینل-۲ از توان تفکیک زمانی ۵ روز در استوا و ۳ روز در عرض‌های میانی برخوردار می‌باشد. باندهای ماهواره‌های مختلف با همدیگر متفاوت‌اند. در این پژوهش شاخص‌های پوشش گیاهی و شوری خاکی که از تصاویر ماهواره سنتینل ۲ استخراج شده است محاسبه و مورد استفاده قرار گرفته‌اند. شاخص‌ها هم طیف بسیار گسترده‌ای را شامل می‌شوند.

برای استخراج مقادیر بازتاب سطحی و شاخص‌های طیفی برای نقاط نمونه‌برداری از پلتفرم Google Earth Engine (GEE) استفاده شد. مراحل کار به شرح زیر بودند:

۱. بارگیری داده‌های مکانی نقاط نمونه‌برداری: یک مجموعه ویژگی (FeatureCollection) از دارایی (Asset) های GEE که قبلاً بارگیری شده بود فراخوانی شده و مختصات جغرافیایی (عرض و طول جغرافیایی) آن به هندسه نقطه‌ای تبدیل شد.
۲. انتخاب نقاط داخل محدوده مورد مطالعه: نقاطی که داخل منطقه مورد مطالعه قرار داشتند فیلتر شده و روی نقشه نمایش داده شدند.
۳. بارگیری تصاویر سنتینل ۲: مجموعه تصاویر سنتینل ۲ از تاریخ ۱ مارس ۲۰۲۳ تا ۱۰ اکتبر ۲۰۲۳ انتخاب شده و بر اساس درصد پوشش ابری کمتر از ۶٪ فیلتر شدند. سپس، مقدار میانه باندهای منتخب محاسبه و داده‌ها به فرمت Float تبدیل شدند.
۴. محاسبه شاخص‌های طیفی: چندین شاخص طیفی مانند NDVI، EVI، NDMI، NDSI و سایر شاخص‌های مورد نظر بر اساس جدول ۲ محاسبه و به تصویر چندباندی اضافه شدند.
۵. نمونه‌گیری از تصاویر در نقاط انتخاب‌شده: مقادیر باندهای سنتینل ۲ و شاخص‌های طیفی برای نقاط داخل منطقه استخراج شدند. همچنین، مختصات نقاط (عرض و طول جغرافیایی) به ویژگی‌های خروجی اضافه شد.
۶. چاپ و ذخیره داده‌ها:

- داده‌های نمونه‌گیری شده به‌صورت فایل جدولی در Google Drive ذخیره شدند.

- تصویر چندباندی همراه با شاخص‌های طیفی به صورت (tif) GeoTIFF در Google Drive ذخیره شد.
- برای آموزش، آزمون و اعتبارسنجی الگوریتم جنگل تصادفی از کدنویسی در محیط R استفاده شد. پکیج‌های مورد استفاده به این منظور عبارت بودند از: tidyverse، caret، randomforest و ggplot2. برای رستری کردن نتایج پیش‌بینی‌ها از پکیج‌های terra و sf و برای طبقه‌بندی رسترهای ایجاد شده و محاسبه سطح هر کلاس علاوه بر پکیج‌های قبلی از پکیج dplyr نیز استفاده شد. جدول ۲ بطور خلاصه کارکرد هر کدام از این پکیج‌ها در محیط R را نشان می‌دهد.
- جدول ۳ شاخص‌های طیفی مورد استفاده در این پژوهش و نحوه محاسبه آن‌ها را نشان می‌دهد. این شاخص‌ها شاخص‌های مهمی هستند که بعنوان متغیرهای کمکی برای مدل‌سازی شوری خاک مورد استفاده قرار می‌گیرند.

جدول ۲- خلاصه کارکرد پکیج‌های اصلی مورد استفاده در پژوهش

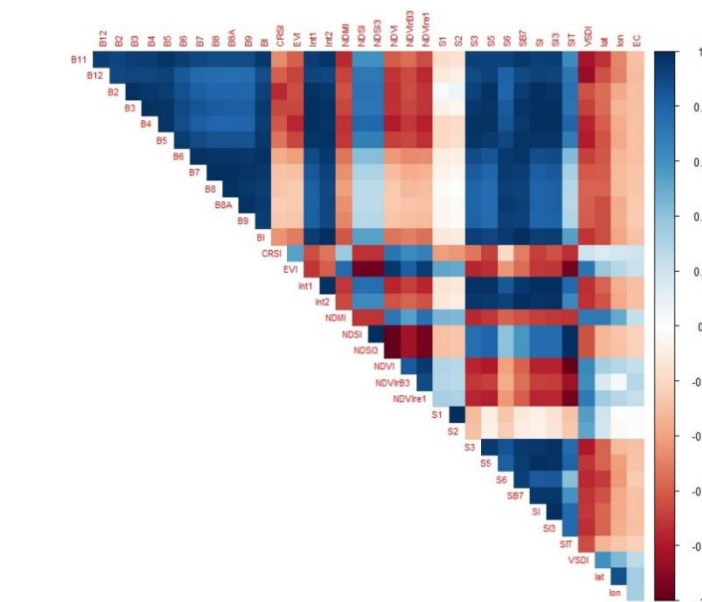
نام پکیج	کارکرد	رفرنس
tidyverse	این پکیج مجموعه‌ای از پکیج‌های محبوب R برای تحلیل و تجسم داده‌ها است. این پکیج‌ها (مانند readr، tidyr، ggplot2، dplyr) بر پایه دستورهای ساده و خوانا طراحی شده‌اند. این پکیج کار با داده‌ها را به صورت ساختارمند، سریع و قابل فهم آسان می‌سازد.	Wickham et al., 2019
caret	این پکیج ابزاری قدرتمند برای ساخت، آموزش و ارزیابی مدل‌های یادگیری ماشین است. این پکیج فرآیندهایی مانند پیش‌پردازش داده، تنظیم پارامترها (tuning) و اعتبارسنجی متقابل (cross-validation) را ساده‌سازی می‌کند. این پکیج از صدها الگوریتم مختلف پشتیبانی کرده و محیطی یکپارچه برای مقایسه عملکرد مدل‌ها فراهم می‌سازد.	Kuhn, 2008
randomforest	برای ساخت مدل‌های جنگل تصادفی به کار می‌رود. این پکیج برای طبقه‌بندی و رگرسیون بسیار دقیق، مقاوم در برابر بیش‌برازش، و قابل استفاده برای ارزیابی اهمیت متغیرها است. عملکرد آن برای داده‌های بزرگ و پیچیده بسیار مناسب است.	Liaw and Winer, 2002
ggplot2	برای ترسیم نمودارهای آماری زیبا و قابل تنظیم استفاده می‌شود. این پکیج امکان ساخت نمودارهای پیچیده با لایه‌بندی داده‌ها را فراهم می‌کند. ابزاری قدرتمند برای تحلیل و نمایش بصری داده‌هاست.	Wickham, 2016
terra	برای تحلیل و پردازش داده‌های مکانی و رستری به کار می‌رود. این پکیج قابلیت‌های متنوعی برای خواندن، نوشتن، پردازش و تجزیه و تحلیل داده‌های جغرافیایی دارد.	Hijmans, 2020
sf	خواندن، نوشتن و تحلیل داده‌های مکانی. این پکیج انواع هندسه‌های مکانی مانند نقطه، خط و چندضلعی را پشتیبانی می‌کند و با توابعی مثل st_read()، st_intersection() و st_crs() کار می‌کند. همچنین با tidyverse و ggplot2 به خوبی سازگار است و امکان تحلیل و ترسیم داده‌های مکانی را فراهم می‌کند.	Pebesma and Bivand, 2023
dplyr	برای دستکاری و تحلیل داده‌های جدولی طراحی شده است. این پکیج توابع ساده و قدرتمندی برای فیلتر کردن، گروه‌بندی، خلاصه‌سازی و ترکیب داده‌ها فراهم می‌کند. این پکیج بخشی از مجموعه tidyverse است و برای کدنویسی خوانا و مؤثر بسیار مناسب است.	Wickham, 2015

جدول ۳- شاخص‌های طیفی استخراج شده از تصاویر ماهواره سنتینل ۲

شاخص طیفی	رابطه	رفرنس
شاخص شوری (Salinity Index- SI)	$\sqrt{B2 \times B4}$	
شاخص شوری ۲ (Salinity Index- SI2)	$\sqrt{B3^2 + B4^2 + B8A^2}$	
شاخص شوری ۳ (Salinity Index- SI3)	$\sqrt{B3^2 + B4^2}$	
شاخص شوری (Salinity Index- S1)	$B2/B4$	
شاخص شوری (Salinity Index- S2)	$(B2 - B4)/(B2 + B4)$	
شاخص شوری (Salinity Index- S3)	$(B3 \times B4)/B2$	(Allbed et al., 2014)
شاخص شوری (Salinity Index- S5)	$(B2 \times B4)/B3$	
شاخص شوری (Salinity Index- S6)	$(B4 \times B8A)/B3$	
شاخص شوری T (Salinity Index T- SI-T)	$(B4/B8A)/100$	
شاخص تفاوت شوری نرمال شده (Normalized)	$(B4 - B8A)/(B4 + B8A)$	
شاخص شوری تفاوت (Difference Salinity Index- NDSI)	$\sqrt{B4^2 + B8A^2}$	
شاخص شفافیت (Brightness Index- BI)	$\frac{B8A \times B4 - B3 \times B2}{\sqrt{B8A \times B4 + B3 \times B2}}$	(Scudiero et al., 2014)
شاخص واکنش به شوری تاج پوشش (Canopy Response)	$(B3 + B4)/2$	
شاخص شوری (Salinity Index- CRSI)	$(B3 + B4 + B8A)/2$	(Fourati et al., 2015)
شاخص شدت ۱ (Intensity Index 1- Int1)	$(B8A - B4)/(B8A + B4)$	(Cho et al., 2018)
شاخص تفاوت پوشش گیاهی نرمال شده (Normalized)	$\frac{2.5 \times (B8A - B4)}{B8A + 6 \times B4 - 7.5 \times B2 + 1}$	(Qiu et al., 2019)
شاخص پوشش گیاهی ارتباط یافته (Enhanced)	$(B8A - B5)/(B8A + B5)$	
شاخص تفاوت پوشش گیاهی نرمال شده حاشیه سرخ (Normalized Difference Vegetation Index)۱	$(B8A - B6)/(B5 + B6)$	
شاخص تفاوت پوشش گیاهی نرمال شده حاشیه سرخ (Normalized Difference Vegetation Index)۲	$(B6 - B5)/(B6 + B5)$	
تفاوت نرمال شده حاشیه سرخ (Normalized)	$(B7 - B5)/(B7 + B5)$	(Chen et al., 2020)
شاخص مثلثی کلروفیل حاشیه سرخ ۱ (Triangular)	$1.2 \times (B5 - B3) - 1.5 \times (B4 - B3) \times \sqrt{B5/B4}$	
شاخص پوشش گیاهی تغییر یافته (Modified Soil)	$\frac{2 \times B4 + 1 - \sqrt{(2 \times B8A + 1)^2 - 8 \times (B8A - B4)}}{2}$	
شاخص تفاوت نرمال شده خشکی (Normalized)	$(NDVI - NDWI)/(NDVI + NDWI)$	
شاخص پایش رطوبت خاک (Soil Moisture)	$\sqrt{B8A^2 + B11^2}/\sqrt{2}$	
شاخص پایش (Monitoring Index- SMMI)	$(B11 - B12)/(B11 + B12)$	(Yue et al., 2019)
شاخص خشکی نرمال شده چند باندی (Normalized)	$\frac{B8A - (B11 - B12)}{B8A + (B11 + B12)}$	
شاخص تفاوت نرمال شده آب (Normalized Difference)	$(B8A - B11)/(B8A + B11)$	Hazaymeh & Hassan,) (2017
شاخص خشکی موج مرئی و کوتاه (Visible and)	$1 - (B12 + B4 - 2 \times B2)$	
شاخص خشکی موج مرئی و کوتاه (Shortwave Drought Index- VSDI)		

برای پیش‌بینی مقادیر شوری در سال‌های مختلف، یک شبکه منظم ۱۰ در ۱۰ متری بر روی هر کدام از باغات مورد بررسی ایجاد شد، مقادیر متغیرهای کمکی برای هر نقطه توسط GEE استخراج شد. برای مقادیر استخراج شده الگوریتم جنگل تصادفی در R اجرا شد و مقدار شوری خاک پیش‌بینی شد. مقادیر پیش‌بینی شده توسط R تبدیل به رستر و نقشه شدند.

مدل جنگل تصادفی مورد استفاده با ۵۰۰ درخت تصمیم آموزش داده شد. شکل ۳ مقادیر شوری متوسط خاک اندازه‌گیری شده در مقابل مقادیر پیش‌بینی شده توسط مدل جنگل تصادفی را نشان می‌دهد. مقدار $RMSE$ این مدل $1/1.03$ دسی‌زیمنس بر متر و مقدار R^2 در دسته آموزش ۰/۹۳۲ بود.

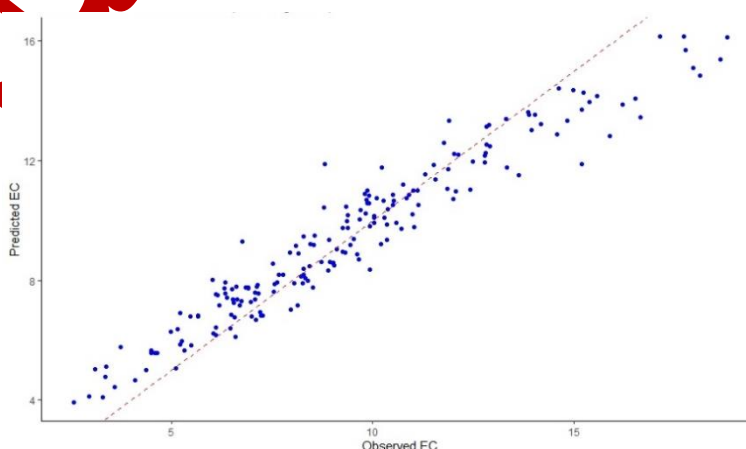


بین جفت متغیرهای کمکی و اصلی (EC)

آموزش داده شد. شکل ۳ نقادیر شوری متوسط خاک اندازه گیری

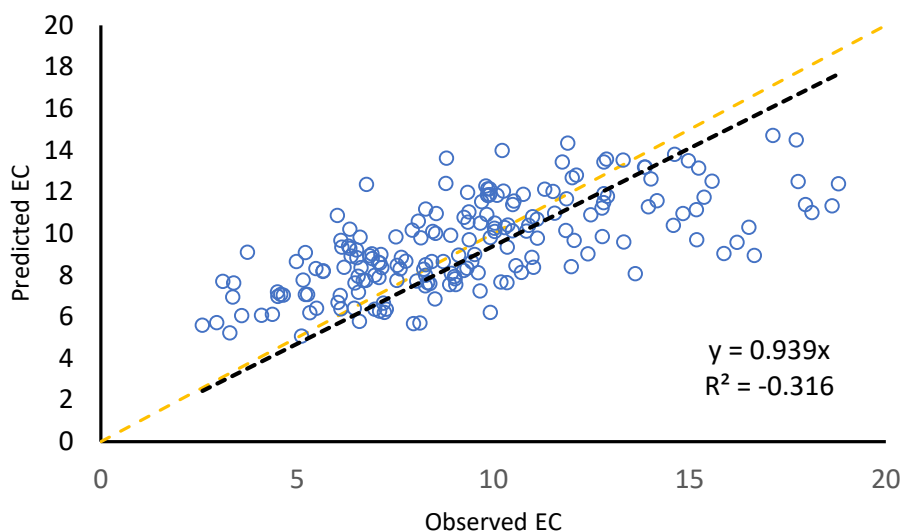
فی را نشان می دهد. مقدار RMSE این مدل $1/1.03$ دسی زیمنس

در مقابل مقادیر پیش‌بینی شده توسط مدل جنگل تصادفی را نشان می‌دهد. مقدار RMSE این مدل ۱/۱۰۳ دسی‌زیمنس بر متر و مقدار R^2 در دسته آموزش ۰/۹۳۲ بود.



شکل ۳- همبستگی بین مقادیر اندازه‌گیری و پیش‌بینی شده توسط جنگل تصادفی در دسته آموزش

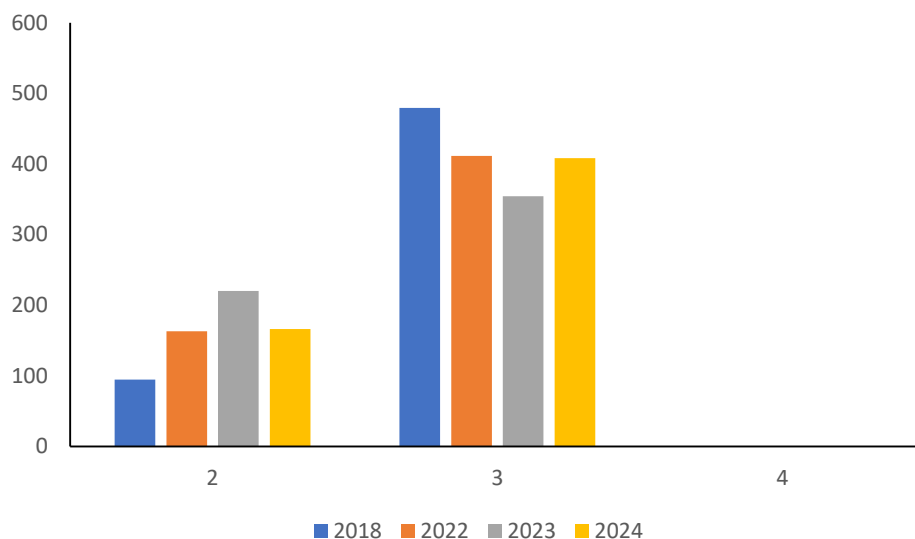
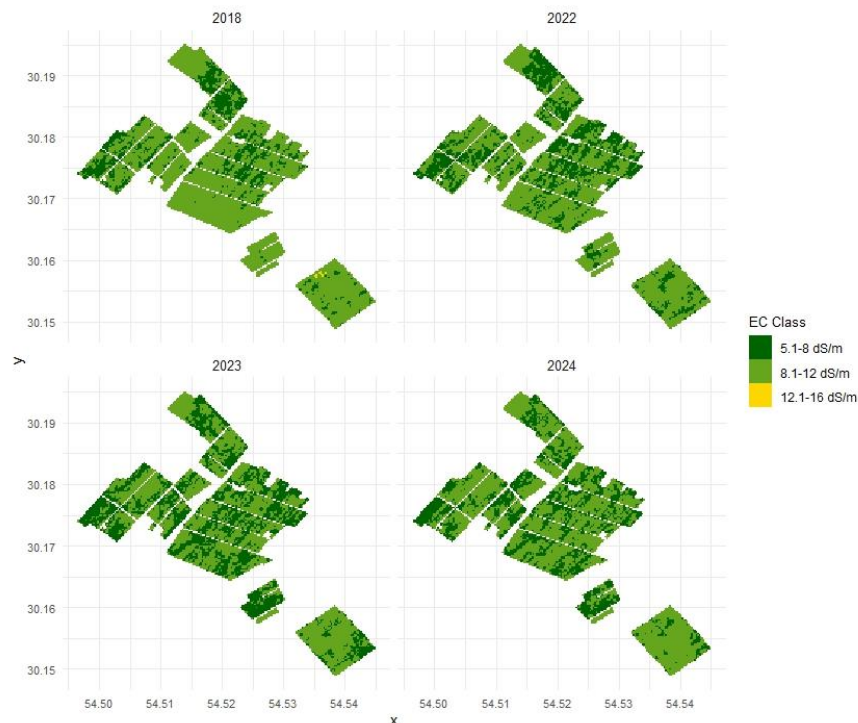
برای اعتبارسنجی از روش کنار گذاشتن یک نقطه (Leave One Out Cross Validation LOOCV) استفاده شد. رابطه رگرسیون بین مقادیر اندازه‌گیری شده و پیش‌بینی شده در این دسته به همراه خط برازش داده شده و خط یک به یک در شکل ۴ نشان داده شده است. مقدار RMSE در این دسته ۲/۵۹ دسی‌زیمنس بر متر بود.



شکل ۴- همبستگی بین مقادیر اندازه‌گیری و پیش‌بینی شده توسط جنگل تصادفی در دسته آزمون

شاخص‌های طیفی استفاده شده در این تحقیق، شاخص‌هایی هستند که با وضعیت تنش در پوشش گیاهی رابطه دارند. به همین دلیل می‌توانند منعکس کننده عملکرد گیاه نیز باشند. سیدمحمدی و همکاران (۲۰۲۴) نشان دادند که با استفاده از الگوریتم جنگل تصادفی می‌توان ۹۶ درصد تغییرپذیری عملکرد باغات پسته در استان‌های مختلف کشور را توضیح داد.

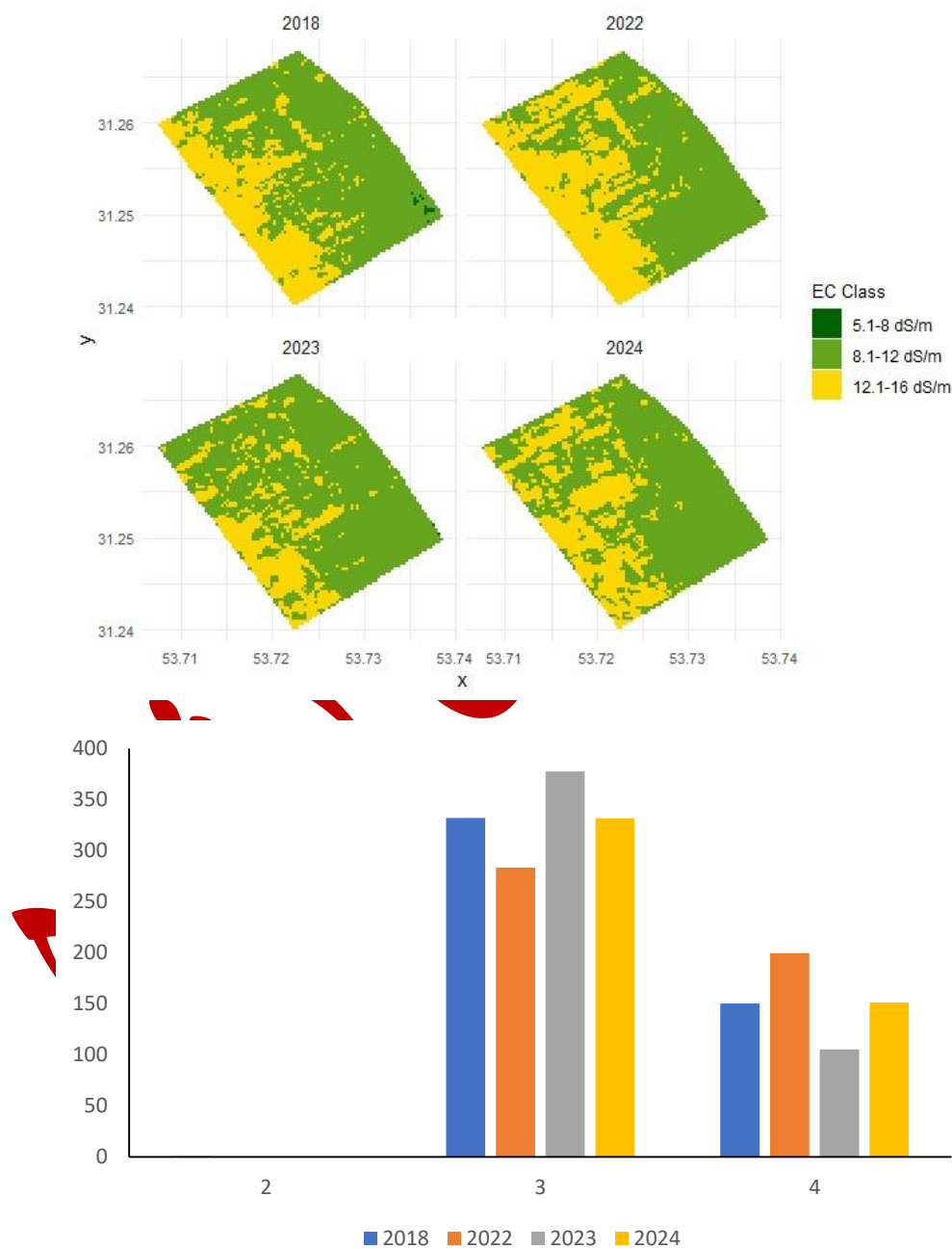
نشانه



شکل ۵- نقشه و نمودار تغییرات میانگین شوری خاک پروژه زهکشی و بهسازی خاک تنور لاهور در استان یزد

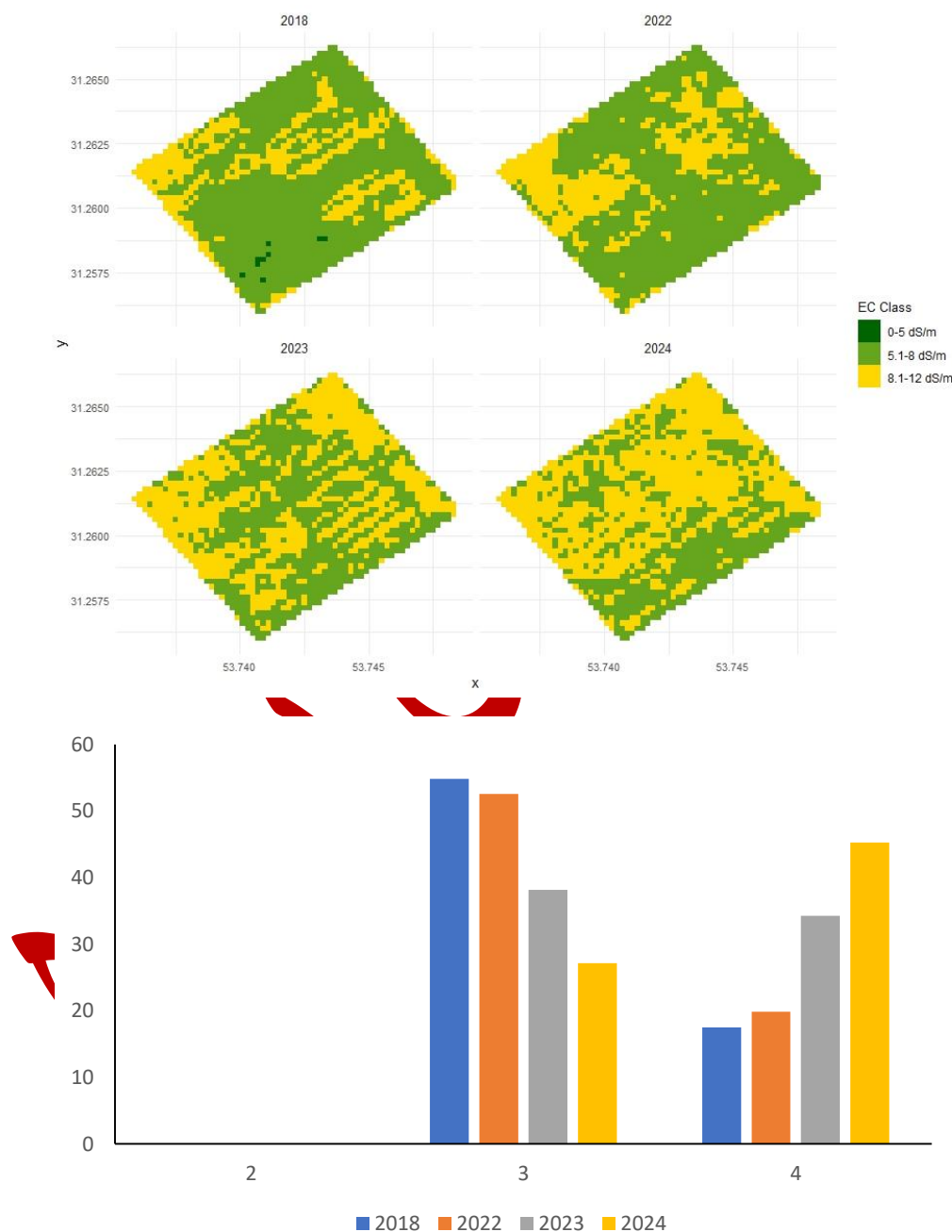
پس از تعمیم نتایج به سال‌های مختلف مشاهده شد که مقادیر شوری خاک پیش‌بینی شده در سال‌های ۲۰۱۹، ۲۰۲۰ و ۲۰۲۱ برای تمام مناطق مورد بررسی بطور چشمگیری بیشتر از سال‌های قبل و بعد از آن بود. به همین دلیل نتایج این سال‌ها حذف شد. برای سال‌های ۲۰۱۵ تا ۲۰۱۷ نیز تصاویر سنتینل فاقد تصحیح اتمسفری بودند. به همین دلیل برای تمام نقاط مورد بررسی نقشه تغییرات شوری خاک در سال‌های ۲۰۱۸، ۲۰۲۲، ۲۰۲۳ و ۲۰۲۴ نشان داده شده است. در تمامی این شکل‌ها کلاس ۱ نشان دهنده شوری ۰ تا ۵ دسی‌زیمنس بر متر، کلاس ۲ نشان‌دهنده شوری ۵ تا ۸ دسی‌زیمنس بر متر، کلاس ۳ نشان‌دهنده شوری ۸ تا ۱۲ دسی‌زیمنس بر متر، کلاس ۴ نشان‌دهنده شوری ۱۲ تا ۱۶ دسی‌زیمنس بر متر و کلاس ۵ نشان‌دهنده شوری ۱۶ تا ۲۰ دسی‌زیمنس بر متر است.

دسی‌زیمنس بر متر می‌باشد. شکل ۵ نقشه و نمودار تغییرات میانگین شوری خاک پروژه زهکشی و بهسازی خاک تنور لاهور در استان یزد را نشان می‌دهد. بر این اساس مساحت اراضی کلاس ۳ کاهش و کلاس ۲ افزایش یافته است. بدین ترتیب می‌توان نتیجه‌گیری کرد که شوری خاک در محدوده این پروژه رو به کاهش است. با توجه به زمان احداث این پروژه زهکشی (سال ۱۳۹۶) کاهش شوری خاک تا زمان رسیدن به یک سطح تعادلی مورد انتظار است. زمان رسیدن به سطح تعادلی جدید در آبیاری با آب‌های با شوری‌های مختلف، متفاوت است (Mehdi Tounsi et al., 2017) و بسته به سطح شوری اولیه خاک و شوری آب آبیاری، این روند می‌تواند افزایشی یا کاهش‌ی باشد.



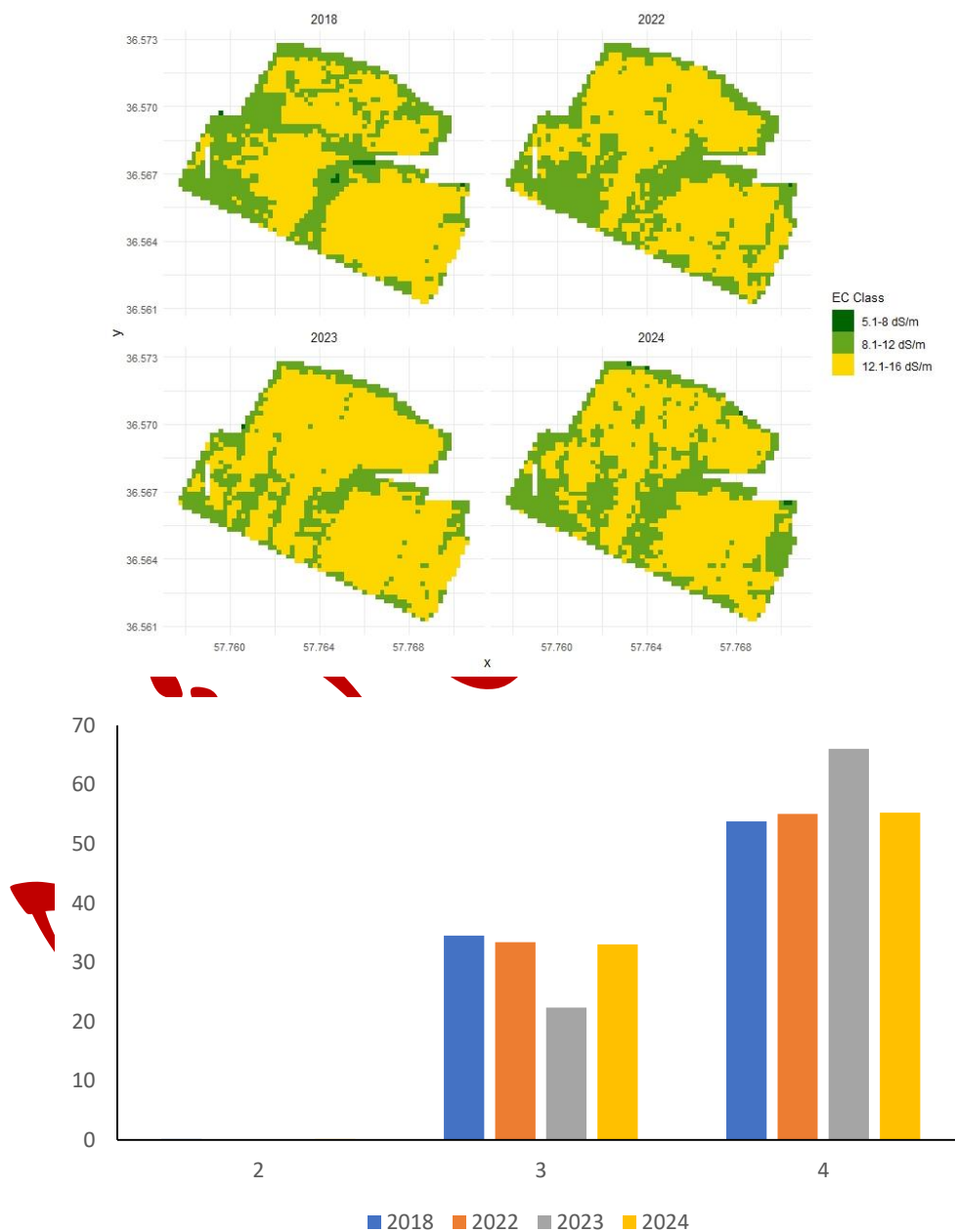
شکل ۶- نقشه و نمودار تغییرات میانگین شوری خاک باغ مرتاضیه در استان یزد

شکل ۶ نقشه و نمودار تغییرات میانگین شوری خاک باغ مرتاضیه در استان یزد را نشان می‌دهد. بر این اساس شوری خاک در سال ۲۰۲۲ نسبت به قبل از آن روند افزایشی و در سال ۲۰۲۳ روند کاهشی نشان می‌دهد. این مجموعه در طی سال‌های متمادی آبیاری با آب‌هایی با شوری‌های نسبتاً ثابت به سطحی از تعادل رسیده است و نوسانات جزئی شوری خاک احتمالاً مربوط به تغییرات مدیریت آبیاری بوده است (Shamsi et al., 2022).



شکل ۷- نقشه و نمودار تغییرات میانگین شوری خاک باغ رضایی در استان یزد

شکل ۷ نقشه و نمودار تغییرات میانگین شوری خاک باغ رضایی در استان یزد را نشان می‌دهد. بر این اساس مساحت کلاس ۳ کاهش یافته و به همان نسبت مساحت کلاس ۴ افزایش یافته است که نشان دهنده روند شور شدن خاک در طی سال‌های مورد بررسی است. در سال‌های اخیر سیاست قطع برق در طی فصل گرم سال منجر به افزایش فاصله بین آبیاری‌ها شده است. از طرفی چون حجم آبیاری در هر دور ثابت است، مقدار آبیاری در بسیاری از موارد کفاف تامین نیاز آبی و آبشویی را نمی‌دهد و به همین دلیل شور شدن تدریجی خاک اتفاق می‌افتد. در سال‌های اخیر هر چند افزایش جزئی در شوری آب زیرزمینی به‌عنوان منبع آب آبیاری اتفاق افتاده است (داده‌های خارج از پژوهش)، اما این تغییرات کسر آبشویی است که منجر به تغییر شوری خاک در این سطح شده است (Hasheminejad, 2011).



شکل ۸- نقشه و نمودار تغییرات میانگین شوری خاک باغ دادیار در استان خراسان رضوی

شکل ۸ نقشه و نمودار تغییرات میانگین شوری خاک باغ دادیار در استان خراسان رضوی را نشان می‌دهد. بر این اساس مساحت کلاس ۳ نسبتاً کاهش یافته و به همان نسبت مساحت کلاس ۴ افزایش یافته است که نشان دهنده روند شور شدن نسبی خاک در طی سال‌های مورد بررسی است. علت این روند افزایشی در بین باغداران کوچک مقیاس‌تر را می‌توان مانند مورد قبل به قطع برق نسبت داد. در این مورد نیز داده‌های خارج از پژوهش نشان دهنده تغییرات نامحسوس شوری آب آبیاری در طی سال‌های پژوهش است. از طرفی در سال‌های اخیر و پس از نصب کنتورهای هوشمند، مدیریت زمان برداشت آب در اختیار بهره‌برداران است و معمولاً طوری برنامه‌ریزی می‌شود که امکان آبیاری زمستانه وجود داشته باشد. کشاورزان انتخاب شده در پژوهش همگی جزو کشاورزان پیش‌رو بوده و فعالیت‌های اصلاح خاک، آبیاری زمستانه و کوددهی را انجام می‌دهند و به همین دلیل این تغییرات را می‌توان به کاهش حجم آبیاری نسبت داد.

تجمع نمک در خاک‌های تحت آبیاری که فاقد مشکل زهکشی می‌باشند در اثر ناترازی آبیاری و زهکشی اتفاق می‌افتد (Ritzema, 2016). پسته‌کاران ایرانی به‌طور معمول در طی فصل رشد آبیاری به مقدار کمتر از نیاز انجام می‌دهند که منجر به تجمع نمک در منطقه ریشه می‌شود و در عوض در فصل زمستان عملیات آبیاری و آبشویی زمستانه را انجام می‌دهند (Dehghani et al., 2023). این در حالی است که شوری خاک در همه سطح باغ بطور یکنواخت نیست. آبشویی یکنواخت زمستانه باعث هدررفت آب در برخی مناطق و باقی ماندن نمک در مناطق دیگر می‌شود. تعیین این غیریکنواختی در سطح مزرعه نیازمند نمونه‌برداری گسترده است که زمان و هزینه زیادی دربردارد و عملاً تکرار آن در طی سال‌های مختلف را غیرممکن می‌سازد. روش پیشنهاد شده در این پژوهش امکان ارائه نقشه تغییرات شوری خاک را با استفاده از شاخص‌های سنجش از دور و الگوریتم جنگل تصادفی فراهم می‌آورد که با استفاده از آن کشاورزان می‌توانند تشخیص دهند کدام مناطق از مزرعه نیازمند آبشویی زمستانه بیشتری است.

نتیجه‌گیری

نتایج این تحقیق نشان داد که مدل جنگل تصادفی با بهره‌گیری از منفرجهای استخراج‌شده از تصاویر ماهواره‌ای سنتینل-۲، توان بالایی در پیش‌بینی تغییرات شوری خاک در باغات پسته استان‌های یزد و خراسان رضوی دارد. دقت بالای مدل داده‌های آموزش و آزمون، همراه با مقدار خطای پایین، عملکرد قابل‌قبولی را در برآورد شوری خاک نشان می‌دهد. بررسی تغییرات فضایی شوری نیز نشان داد که در برخی مناطق مانند پروژه زهکشی تنور لاهور، کاهش شوری مشاهده شده و اثر مثبت اقدامات مدیریتی قابل مشاهده است، در حالی که در مناطقی مانند مزرعه رضایی، روند افزایش شوری هشداردهنده است. این نتایج بر اهمیت پایش مستمر و استفاده از ابزارهای پیشرفته مانند ترکیب داده‌های سنجش از دور و الگوریتم‌های یادگیری ماشین برای مدیریت مؤثر شوری خاک تأکید دارد و می‌تواند در سایر مناطق کشاورزی شور نیز به کار گرفته شود تا از شور شدن خاک و کاهش عملکرد محصولات جلوگیری شود.

منابع

- Afshar, F. A., Ayoubi, S., Besalatpour, A. A., Khademi, H., and Castrignano, A. (2016). Integrating auxiliary data and geophysical techniques for the estimation of soil clay content using CHAID algorithm. *Journal of Applied Geophysics*, 126, 87-97.
- 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*, 2013.
- Allbed, A., Kumar, L., & Aldakheel, Y. Y. (2014). Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma*, 230, 1-8.

- Arsoy, S., Ozgur, M., Keskin, E., and Yilmaz, C. (2013). Enhancing TDR based water content measurements by ANN in sandy soils. *Geoderma*, 195, 133-144.
- Brunner, P. H. T. L., Li, H. T., Kinzelbach, W., and Li, W. P. (2007). Generating soil electrical conductivity maps at regional level by integrating measurements on the ground and remote sensing data. *International Journal of Remote Sensing*, 28(15), 3341-3361.
- Chen, Y., Qiu, Y., Zhang, Z., Zhang, J., Chen, C., Han, J., & Liu, D. (2020). Estimating salt content of vegetated soil at different depths with Sentinel-2 data. *PeerJ*, 8, e10585.
- Cho, K. H., Beon, M.-S., & Jeong, J.-C. (2018). Dynamics of soil salinity and vegetation in a reclaimed area in Saemangeum, Republic of Korea. *Geoderma*, 321, 42-51.
- Dehghani, F., Rahnemaie, R., Dalir, N., Saadat, S., & Zarebanadkouki, M. (2023). Interactive effect of salinity and Ca to Mg ratio of irrigation water on pistachio growth parameters and its ionic composition in a calcareous soil. *New Zealand Journal of Crop and Horticultural Science*, 51(3), 432-450.
- Dong, W., Wu, T., Luo, J., Sun, Y., and Xia, L. (2019). Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. *Geoderma*, 340, 234-248.
- FAO. (2024). Global status of salt-affected soils—Main report. Rome. <https://doi.org/10.4060/cd3044en>.
- Fourati, H. T., Bouaziz, M., Benzina, M., & Bouaziz, S. (2015). Modeling of soil salinity within a semi-arid region using spectral analysis. *Arabian Journal of Geosciences*, 8(12), 11175-11182.
- Fourati, T.H., Bouaziz, M., Benzina, M., and Bouaziz, S. (2017). Detection of terrain indices related to soil salinity and mapping salt-affected soils using remote sensing and geostatistical techniques. *Environmental monitoring and assessment*, 189, 1-11.
- Ghassemi, F., Jakeman, A. J., and Nix, H. A. (1995). Salinisation of land and water resources: human causes, extent, management and case studies: *CAB international*.
- Gholizadeh, A., Žižala, D., Saberioon, M., and Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218, 89-103.
- Gomez, C., and Coulouma, G. (2018). Importance of the spatial extent for using soil properties estimated by laboratory VNIR/SWIR spectroscopy: Examples of the clay and calcium carbonate content. *Geoderma*, 330, 244-253.
- Gorji, T., Sertel, E., & Tanik, A. (2017). Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecological indicators*, 74, 384-391.
- Haq, Y. U., Shahbaz, M., Asif, S., Ouahada, K., & Hamam, H. (2023). Identification of soil types and salinity using MODIS terra data and machine learning techniques in multiple regions of Pakistan. *Sensors*, 23(19), 8121.
- Hasheminejad, Y. (2011). Irrigation Management under Saline Conditions Using Wetting Front Detector (WFD). *Iranian Journal of Soil Research*, 24(3), 265-272. doi: 10.22092/ijsr.2011.126640
- Hassani, A., Azapagic, A., and Shokri, N. (2020). Predicting long-term dynamics of soil salinity and sodicity on a global scale. *Proceedings of the National Academy of Sciences*, 117(52), 33017-33027.
- Hazaymeh, K., & Hassan, Q. K. (2017). A remote sensing-based agricultural drought indicator and its implementation over a semi-arid region, Jordan. *Journal of Arid Land*, 9(3), 319-330.
- Hijmans, R. J. (2020). terra: Spatial data analysis. *CRAN: Contributed Packages*.

- Kuhn, M. (2008). "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software*, 28(5), 1–26. doi:10.18637/jss.v028.i05, <https://www.jstatsoft.org/index.php/jss/article/view/v028i05>.
- Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., and Nkuba-Kasanda, L. (2020). Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. *Geoderma*, 375, 114503.
- Liaw A. and Wiener M. (2002). "Classification and Regression by randomForest." *R News*, 2(3), 18–22. <https://CRAN.R-project.org/doc/Rnews/>.
- Lin, Z. Q., and Banuelos, G. S. (2015). Soil salination indicators. *Environmental indicators*, 319–330.
- Mehdi-Tounsi, H., Chelli-Chaabouni, A., Mahjoub-Boujnah, D., & Boukhris, M. (2017). Long-term field response of pistachio to irrigation water salinity. *Agricultural Water Management*, 185, 1–12.
- Mulder, V. L., De Bruin, S., Schaepman, M. E., and Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1–2), 1–19.
- Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>
- Peng, J., Biswas, A., Jiang, Q., Zhao, R., Hu, J., Hu, B., and Shi, Z. (2019). Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma*, 337, 1309–1319.
- Qiu, Y., Chen, C., Han, J., Wang, X., Wei, S., & Zhang, Z. (2019). Satellite remote sensing estimation model of soil salinity in Jiefangzha irrigation under vegetation coverage. *Water Saving Irrigation*, 44, 108–112.
- Ren, D., Wei, B., Xu, X., Engel, B., Li, G., Huang, Q., Xiong, Y. and Huang, G. (2019). Analyzing spatiotemporal characteristics of soil salinity in arid irrigated agro-ecosystems using integrated approaches. *Geoderma*, 356, 113935.
- Ritzema, H. P. (2016). Drain for Gain: Managing salinity in irrigated lands—A review. *Agricultural Water Management*, 176, 18–28.
- Rossel, R. V., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., and Lobsey, C. (2011). Proximal soil sensing: An effective approach for soil measurements in space and time. *Advances in agronomy*, 113, 243–291.
- Scudiero, E., Skaggs, T. H., and Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional*, 2, 82–90.
- Seyedmohammadi, J., Navidi, M. N., Zeinadini, A., & McDowell, R. W. (2024). Random forest, an efficient smart technique for analyzing the influence of soil properties on pistachio yield. *Environment, Development and Sustainability*, 26(1), 2615–2636.
- Shamsi, S., Kamali, A., & Hasheminejhad, Y. (2022). An approach to predict soil salinity changes in irrigated pistachio orchards (Ardakan, Yazd Province): A case study. *Dry Land Soil Research (DLSR)*, 1(1), 1–10.
- Taghadosi, M. M., Hasanlou, M., & Eftekhari, K. (2019). Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. *International journal of remote sensing*, 40(1), 237–252.

- Taghizadeh-Mehrjardi, R., Schmidt, K., Toomanian, N., Heung, B., Behrens, T., Mosavi, A., Band, S.S., Amirian-Chakan, A., Fathabadi, A. and Scholten, T. (2021). Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. *Geoderma*, 383, 114793.
- Wang, N., Chen, S., Huang, J., Frappart, F., Taghizadeh, R., Zhang, X., Wigneron, J.P., Xue, J., Xiao, Y., Peng, J. and Shi, Z. (2024). Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing. *Journal of Remote Sensing*, 4, 0130.
- Wickham, H. (2015). dplyr: A grammar of data manipulation. *R package version 04.*, 3, p156.
- Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M. (2019). Welcome to the tidyverse. *Journal of open source software*, 4(43), 1686.doi:10.21105/joss.01686.
- Xi, W. F., Jiang, Q. W., & Yang, A. M. (2024). Using stepwise regression to address multicollinearity is not appropriate. *International Journal of Surgery*, 110(5), 3122-3123.
- Xiao, C., Ji, Q., Chen, J., Zhang, F., Li, Y., Fan, J., Hou, X., Yan, F. and Wang, H. (2023). Prediction of soil salinity parameters using machine learning models in an arid region of northwest China. *Computers and Electronics in Agriculture*, 204, 107512.
- Yue, J., Tian, J., Tian, Q., Xu, K., & Xu, N. (2019). Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *Isprs Journal of Photogrammetry and Remote Sensing*, 154, 216-230.
- Zare, E., Ahmed, M. F., Malik, R. S., Subasinghe, R., Huang, J., and Triantafilis, J. (2018). Comparing traditional and digital soil mapping at a district scale using residual maximum likelihood analysis. *Soil Research*, 56(5), 535-547.
- Zare, S., Abtahi, A., Shamsi, S. R. F., and Lagacherie, P. (2021). Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches. *Catena*, 207, 105702.
- Zhao, S., Ayoubi, S., Mousavi, S.R., Mireei, S.A., Shahpouri, F., Wu, S.X., Chen, C.B., Zhao, Z.Y. and Tian, C.Y. (2024). Integrating proximal soil sensing data and environmental variables to enhance the prediction accuracy for soil salinity and sodicity in a region of Xinjiang Province, China. *Journal of Environmental Management*, 364, 121311.