Essay

Impact of Land Cover Changes on Soil Mapping in Plain Areas: Evidence from Tongzhou District of Beijing,China

Xiangyuan Wu , Kening Wu Wu * , Huafu Zhao , Shiheng Hao , Long Kang , Zhenyu Zhou

Posted Date: 1 August 2023

doi: 10.20944/preprints202308.0080.v1

Keywords: and cover changes; soil mapping; random forest; plain areas

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

Essay

Impact of Land Cover Changes on Soil Mapping in Plain Areas: Evidence from Tongzhou District of Beijing, China

Xiangyuan Wu [1], Kening Wu [1,2,*], Huafu Zhao [1,2], Shiheng Hao [1], Long Kang [1] and Zhenyu Zhou [1]

[1] School of Land Science and Technology, China University of Geosciences, Beijing 100083, China
[2] Key Laboratory of Land Consolidation and Rehabilitation, Ministry of Natural Resources, Beijing 100035, China
* Correspondence: wukening@cugb.edu.cn

Abstract: The flat terrain in plain areas makes the land easily accessible for cultivation and farming, providing vast opportunities for agricultural development. Additionally, these areas are crucial for urban construction and economic growth. Soil mapping plays a crucial role in understanding soil characteristics and guiding land management practices. However, accurately mapping soils in plain regions can be challenging due to their low spatial variability and diverse land use types. This study focuses on the impact of land cover changes on the accuracy of soil mapping in plain areas, aiming to provide effective assistance in soil mapping through the analysis of their coupling relationship. Starting with a 20-year land cover change analysis, this study utilizes a unified approach that combines expert knowledge, mixed sampling methods, and random forest mapping techniques. The study incorporates environmental covariates that have minimal period influence and synergistically use NDVI (Normalized Difference Vegetation Index) and land cover data from the same year. The analysis is based on transition matrices, confusion matrices, and their derived indicators. The research findings indicate that Tongzhou District has experienced rapid development over the past 20 years, with the area of construction land nearly doubling. 29% of arable land has been converted into construction land, resulting in an increase in the accuracy of the soil map from 58.99% to 66.91% over the 20-year period. The soil change area during this period accounts for 16.5% of the total area, with 51.9% of the changed areas overlapping with land cover change areas. These overlapping regions are predominantly influenced by human activities. In terms of cultivated land types in the study area, the quantity of arable land has decreased by approximately 29% over the 20 years, while the proportion of sandy loam calcareous fluvo-aquic soil and light loam calcareous fluvo-aquic soil, which constitute nearly half of the soil types, has increased. These data demonstrate the coupling relationship between land cover changes and soil type variations, particularly the significant influence of human activities on soil structure. It is evident that on one hand, improving the extent of land use in plain areas enhances the credibility of soil mapping. On the other hand, human activities impact land cover, which in turn affects and reflects changes in the soil.

Keywords: land cover changes; soil mapping; random forest; plain areas

1. Introduction

Soil mapping plays a significant role in agriculture, environmental protection, and land planning (Kristensen et al., 2019). By providing detailed information about soil characteristics and spatial distribution, soil mapping can guide farmers and land managers in developing rational land cover and crop cultivation strategies to maximize crop yield and quality (Dewage et al., 2019). In plain areas, the low spatial variability of soils and the subtle differences in soil types and properties pose challenges for accurate classification and mapping (De-Cai et al., 2019; Liu et al., 2012; Zhao et al.,

2013). Human activities such as crop cultivation, irrigation, and fertilization can have complex impacts on soils, resulting in more intricate and irregular spatial distribution of soil properties (Rostaminia et al., 2021; Murray et al., 2017). Different land cover types have significant influences on soil formation, evolution, and properties (Adugna et al., 2016). For instance, different land cover types like cropland, grassland, forest, and wetland can lead to variations in soil texture, organic matter content, and water-holding capacity, which in turn affect the chemical, physical, and biological properties of soils (Hounkpatin et al., 2020; Papadopoulou-Vrynioti et al., 2014; Yang et al., 2021). These differences are reflected in soil mapping, with different land cover types often exhibiting distinct spatial distribution patterns in the soil (Chen et al., 2020). Studying the impact of land cover changes on soil mapping can help explore ways to improve mapping accuracy in plain areas and provide valuable guidance for land protection and management measures (Gan-Lin et al., 2017). Additionally, it contributes to optimizing land cover structure, enhancing the sustainability of agricultural production, and offering scientifically informed decision-making support for policymakers to promote sustainable development in plain regions (Wang et al., 2020).

Many scholars have dedicated their efforts to exploring the impact and correlations between land cover changes and soil properties. For example, Zhang et al. (2021) proposed a method to identify agricultural land cover history by overlaying land cover data from 1980 to 2018, and their model, which incorporates agricultural land cover history, improved the mapping accuracy of soil organic carbon compared to models that only utilized natural variables, providing more spatial details attributed to land development. Taveira et al. (2018) sampled and classified soil profiles from regional soil surveys and compared their land cover capabilities with different management levels displayed on maps. The results revealed that higher land cover capability in Minas Gerais state corresponded to higher accuracy in soil mapping. Due to the complex and diverse relationships between land cover changes and soil properties, as well as the interactions among different land cover types, geographic environments, and management practices, there are still research gaps and challenges in this field (Zhang et al., 2022). In this study, expert knowledge was used to extract virtual points and combined with other sampled points as mixed points. In addition to using consistent environmental covariates such as parent material, texture, groundwater depth, distance to water bodies, and elevation as auxiliary variables, random forest was employed for soil mapping from 2000 to 2020 at 5-year intervals, ensuring the use of corresponding land cover types and NDVI as environmental covariates. The coupling relationship between land cover changes in plain areas and soil mapping was revealed through confusion matrices, derived evaluation indicators, and transition matrices. This study will provides new insights to enhance the accuracy of soil mapping in plain areas and address challenges such as rich land use types and low spatial variability, offering a reliable theoretical foundation for agricultural development, urban construction, and economic growth in plain regions.

2. Material and Methods

2.1. Overview of the Study Area

Tongzhou District is located in the southeastern part of Beijing, China, with geographic coordinates ranging from 39°36' to 40°02' and 116°32' to 116°56'. The district covers an area of 903 km², predominantly consisting of alluvial fans and floodplains formed by the Yongding River, Chaobai River, and Wenyu River (Cui et al., 2020). The land surface in Tongzhou District is covered by deep deposits of Quaternary sediments, forming modern alluvial fan plains and impact low plains. The land cover in Tongzhou District exhibits diversity, encompassing various land use types including agriculture, forestry, urbanization, industrialization, transportation, and water resources (Dai et al., 2022). According to available data, the land cover distribution in Tongzhou District is as follows: arable land covers 37.02%, orchards cover 3.81%, forests cover 8.62%, grassland covers 0.13%, urban and industrial land covers 33.43%, transportation land covers 5.31%, water resources and facilities cover 9.47%, and other land covers 2.21%. These surface characteristics and land use types provide essential baseline data and background information for soil mapping in Tongzhou District.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

The district consists of 3 soil classes, 8 sub-classes, 13 soil orders, and 42 soil species, including alluvial soil, brown soil, and sandy soil, with alluvial soil being the dominant soil type in Tongzhou District. (The soil types mentioned in the text are named according to the Chinese soil classification system).

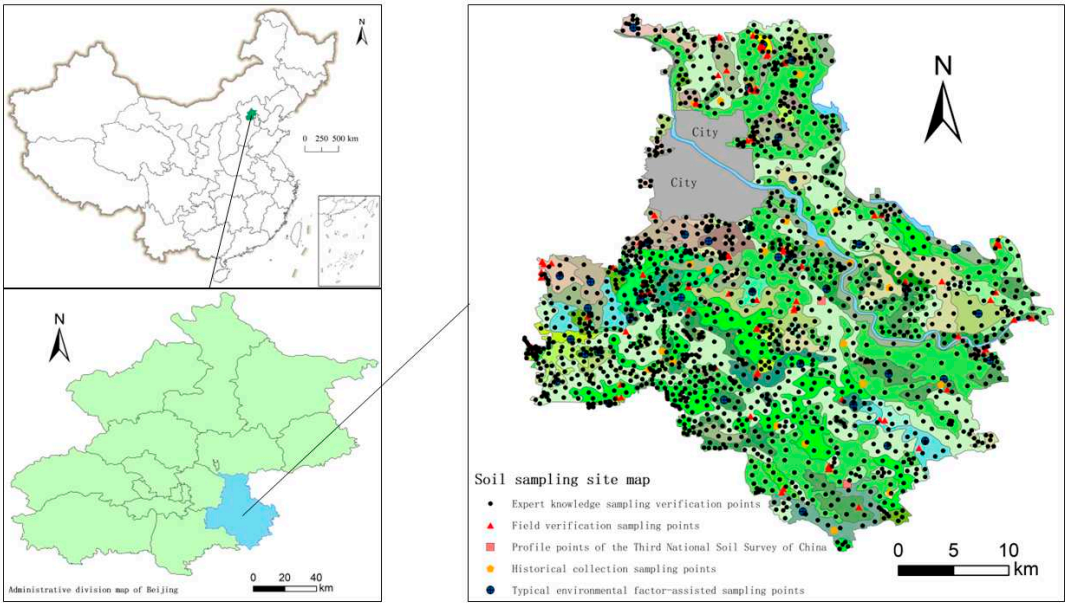

Figure 1. Study Area and Soil Samples (Sampling Point Locations with Calibration Soil Map as the Base Map).

2.2. Data Sources and Processing

2.2.1. Handling Environmental Covariates

Environmental covariates can help explain the spatial distribution of soil properties (Sun et al., 2022). By collecting and analyzing environmental factors related to soil properties, such as topography, climate, vegetation, and land cover, the associations between soil properties and these environmental factors can be revealed (Poggio et al., 2021; Dong et al., 2014; Dunkl et al., 2020). As shown in Figure 2, to ensure data consistency, this study selects factors such as parent material, groundwater depth, land cover type, distance to water bodies, texture, and elevation that have not undergone significant changes over the past 20 years as environmental covariates. Based on this, when conducting soil mapping in different years in the study area, the corresponding NDVI and current land cover data are used for mapping, ensuring that the NDVI and land cover data are from the same data source to ensure the reliability and accuracy of the research results (Yang et al., 2021). Among them, the original soil map, texture, elevation, groundwater depth, and distance to water bodies are derived from the dataset of the National Earth System Science Center, and the parent material information is obtained from 1:25,000 geological maps of the region. NDVI data are sourced from Landsat series data in Google Earth Engine (GEE) after a series of processing steps. The current land cover data are collected from stable samples extracted from the Chinese Land Use/Land Cover Dataset (CLUD) using GEE, as well as visual interpretation samples from satellite time series data, Google Earth, and Google Maps. Multiple temporal indicators were constructed using all available Landsat data and fed into a random forest classifier to obtain the classification results. All the above data are processed with a resolution of 30m to ensure the uniformity of the data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

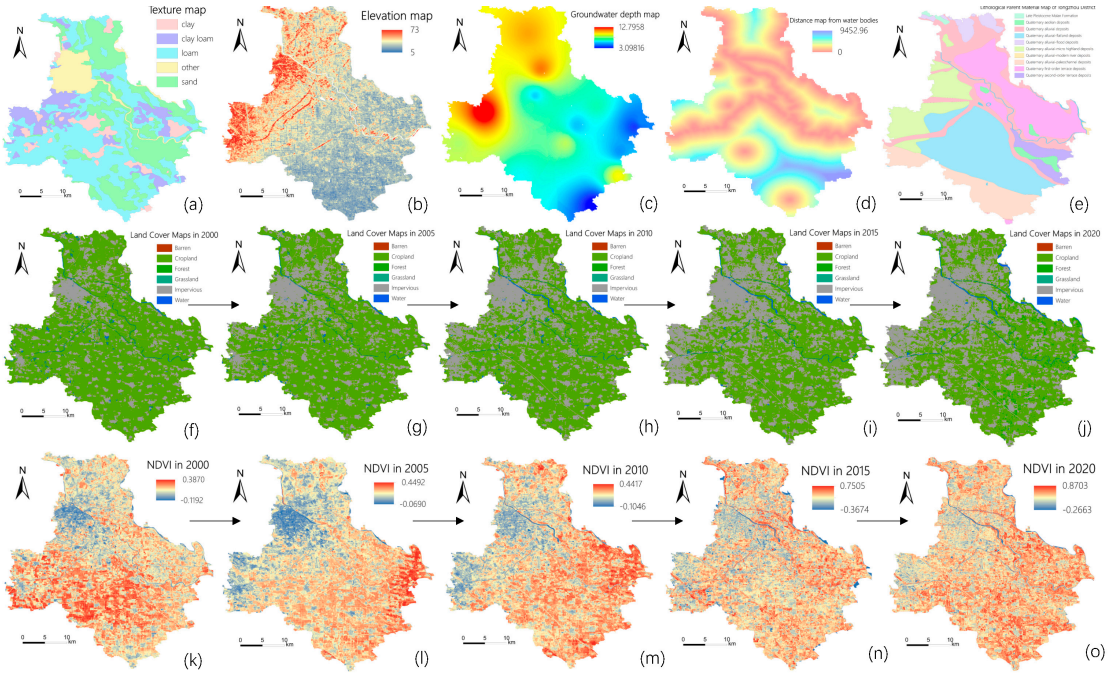

Figure 2. Set of Environmental Covariates (Images a-e represent the texture map, elevation map, groundwater depth map, distance from water bodies map, and parent material map, respectively. Images f-j depict land cover maps for every 5-year interval from 2000 to 2020, while images k-o represent NDVI maps for the same time period.).

2.2.2. Source of Sampling Points

Due to the use of conventional environmental covariate-assisted sampling methods that require real-time remote sensing data, the positions of sampling points may vary across different years (Yang et al., 2020; Liu et al., 2020; Miller et al., 2015). To ensure data consistency, this study primarily relied on a labor-intensive targeted sampling approach based on expert knowledge. Expert knowledge can provide prior information about the distribution of soil properties and, guided by expert experience and domain knowledge, enable the targeted selection of representative and critical sampling points, thus improving sampling efficiency and cost-effectiveness (Zhang et al., 2021; Ng et al., 2020).

The calibration soil map was obtained by conducting indoor verification on the original soil map to address issues related to inconsistent names for the same soil and identical names for different soils. After overlaying the validated soil map with the land cover map and image, the boundary divisions were further refined. In this study, a targeted sampling approach based on expert knowledge was employed, where the validated soil map, current land cover map, and image map were overlaid. Soil experts, using their knowledge in soil science, made judgments based on the overlaid maps to identify virtual points that potentially represent regional environmental factors in the vicinity of the plot center. Leaf vein structures were used for point selection in large plots, ensuring that each plot obtained a representative virtual point for each soil type while guaranteeing a minimum of five virtual points per plot. Ultimately, 1,564 virtual points representing plot characteristics were obtained after manual screening. In addition to these virtual points, other sources of sampling points included field validation points, third national soil survey profile points, historical data collection points, and typical environmental factor-assisted points. Typical environmental factor-assisted points were obtained by overlaying the input environmental covariates and mapping the distribution of typical ranges of each environmental covariate in the same geographic space. Then, the frequency distribution analysis of the environmental covariates was conducted after overlaying them with the validated indoor soil map. The typical value ranges for each key environmental covariate were derived. After outputting the distribution areas or multiple patches of typical environmental conditions for each soil type, the center point of each patch was extracted as a typical point for that soil type, resulting in a total of 41 typical environmental factor-assisted points. The third

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

national soil survey profile points referred to the profile points collected during the unified third national soil survey in China, totaling two points. Historical data collection points referred to 21 points obtained through query and processing of soil data from books such as the Soil of Tong County, Soil Atlas of China, Soil of Beijing, and Soil Series of China (Beijing and Tianjin Volume). Field validation points were obtained by conducting field sampling in key areas of doubt to update the indoor validated soil map, totaling 95 sampling points. As shown in Figure 1, the soil information from the validated soil map was extracted based on the information of the sampling points to obtain the mapping points for this study.

2.3. Modeling and Mapping Methods

This study adopts the Random Forest model for soil mapping. Random Forest is a machine learning algorithm based on ensemble learning, composed of multiple decision trees. Each decision tree is independently generated and increases the diversity of the model through random sampling with replacement and random feature selection on the input samples (Speiser et al., 2023; Moller A B et al., 2019). Random Forest can handle missing data and effectively capture the complex non-linear relationships between soil properties and environmental factors. Its robustness allows it to handle outliers and noise, improving the stability and reliability of soil mapping (Rostaminia M et al., 2021; Abowarda A S et al., 2021). The training process of the Random Forest model includes the following steps:

Data Preparation: Collecting and preparing soil sample data, including measured values of soil properties and associated environmental variable data.

Feature Selection: Selecting a subset of available environmental variables as input features for the decision trees. The feature selection in Random Forest is random, with each decision tree considering only a subset of features.

Random Sampling: Conducting random sampling with replacement on the training data to build training datasets for each decision tree. This allows each decision tree to be trained on different subsets of samples.

Decision Tree Construction: Constructing multiple decision trees using the training datasets. Each decision tree progressively splits the data based on feature selection and splitting criteria, forming a tree structure.

Prediction Ensemble: For new input samples, each decision tree independently predicts the outcome, and the final prediction result is obtained through voting or averaging the predictions from the decision trees. This ensemble approach reduces the risk of random errors and overfitting.

The Random Forest modeling in this study was implemented by calling the randomForest package in the R programming language.

2.4. Evaluation Methods for Mapping Results

The Confusion Matrix is a tool used in machine learning and statistics to evaluate the performance of classification models (Sun H W D et al., 2021). It visualizes and summarizes the relationship between the predicted results of a classification model and the true labels (Wang et al., 2021). The Confusion Matrix and its derived evaluation metrics provide a comprehensive way to assess the performance of classification models (Zeng G et al., 2019). In soil mapping, classification models are used to classify soil types in different regions, making the evaluation of classification model accuracy crucial for determining the correctness of soil types (Fromm S F V et al., 2020). The following are derived evaluation metrics:

Precision is the most commonly used performance metric for classification. It represents the accuracy of the model, i.e., the number of correctly identified instances divided by the total number of samples. Generally, higher precision indicates better model performance (B S L A et al., 2019).

$$\text{Accuracy} = (TP+TN)/(TP+FN+FP+TN) \quad (1)$$

True Positive (TP): The model correctly predicts positive instances as positive.

False Negative (FN): The model incorrectly predicts positive instances as negative.

False Positive (FP): The model incorrectly predicts negative instances as positive.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

True Negative (TN): The model correctly predicts negative instances as negative.

In statistics, a Confidence Interval (CI) is used to estimate the range of a population parameter, and the 95% CI is a commonly used interval (Wang M et al., 2023). A 95% CI indicates a 95% confidence level for the estimated parameter, meaning that in repeated sampling, about 95% of the confidence intervals will contain the true parameter value. It is expressed as follows:

$$95\%\ \text{CI}=\left[\alpha - 1.96 * \frac{\sigma}{\sqrt{n}},\ \alpha + 1.96 * \frac{\sigma}{\sqrt{n}}\right] \tag{2}$$

where α is the mean, σ is the standard deviation, and n is the sample size of a single experiment.

No Information Rate (NIR), also known as the baseline accuracy or random accuracy, refers to the accuracy achieved when making predictions based solely on the majority class in a classification task (Asa et al., 2018). NIR is useful in evaluating classification models because it provides a baseline reference for comparing whether the model's accuracy has substantially improved. If the model's accuracy is significantly lower than the NIR, further model optimization or considering improved data processing methods may be necessary to enhance the model's performance.

P-Value [Acc > NIR] refers to the probability of the observed accuracy being significantly higher than the NIR when conducting a hypothesis test. P-Value is a statistical measure used to assess the significance of the difference between the observed sample data or statistics and the null hypothesis under the assumption that the null hypothesis is true (Zhang J et al., 2021). If the P-Value is smaller than a pre-defined significance level (usually 0.05), the null hypothesis can be rejected, and it can be concluded that the accuracy is significantly higher than the NIR. Conversely, if the P-Value is greater than the significance level, the null hypothesis cannot be rejected, indicating that the difference between the accuracy and the NIR may be due to random factors rather than a true improvement in model performance (Feyisa G L et al., 2013).

Kappa (Cohen's Kappa) is a statistical measure used to assess the performance of a classification model, particularly for evaluating the consistency between observers or the consistency between a model and observers in a classification task (Zou D et al., 2017). The formula is as follows:

$$K = \frac{p_o - p_e}{1 - p_e} \tag{3}$$

where p_o is the sum of the number of correctly classified samples for each class divided by the total number of samples, representing the overall classification accuracy.

$$p_o = \frac{\sum_{i=1}^{C} T_i}{n} \tag{4}$$

$$p_e = \frac{\sum_{i=1}^{C} a_i * b_i}{n^2} \tag{5}$$

where C is the total number of categories, T_i represents the number of samples correctly classified for each category, a_i represents the number of true samples for each category, b_i represents the number of predicted samples for each category, and n is the total number of samples.

3. Results and Analysis

3.1. Analysis of Land Cover Changes

Since the reform and opening up, China has undergone rapid development, and the highly accessible plain areas have experienced significant land cover changes under human influence. As one of the key sub-centers of Beijing, Tongzhou District has expanded rapidly over the past 20 years. With the increase in population and limited land resources in the central urban areas of Beijing, Tongzhou District has become an important area for urbanization in Beijing (Song et al., 2020). The government has increased investment in urban planning and infrastructure construction in Tongzhou District, promoting its urbanization process. As shown in Figure 3, a large area of cropland in the western part of Tongzhou District has been transformed into urban areas during the 20-year urbanization process, and the new urban areas have become connected to the main urban areas. Although there has been some expansion of built-up areas in the eastern part, the overall change is not significant.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

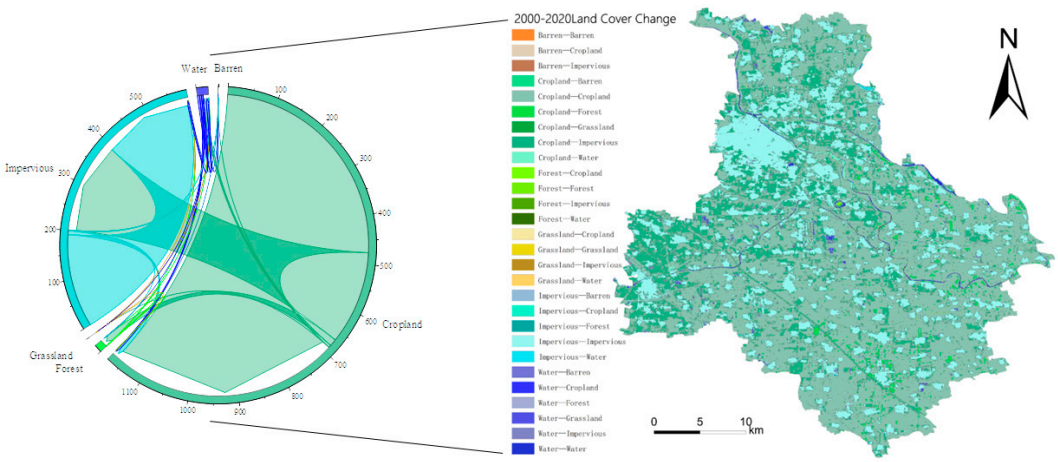

Figure 3. Trend of Land Cover Changes in Tongzhou District from 2000 to 2020.

As indicated in Table 1, cropland is the largest land type in Tongzhou District and maintains a leading position in terms of overall quantity. Although the cropland area has decreased by about 29% in the past 20 years, it remains the primary land type in Tongzhou District. The area of impervious surfaces has significantly increased over the past 20 years, from approximately 201.12 km² in 2000 to around 387.33 km² in 2020, showing a growth of 92.59%. This increase is mainly due to the conversion of cropland, reflecting the rapid development of urbanization in Tongzhou District. The forest area has increased from approximately 1.09 km² in 2000 to around 14.98 km² in 2020. This indicates that Tongzhou District has been working on the protection and increase of green spaces. The overall changes in other land types are not significant, showing no clear trend of growth or decline. The areas of these land types have remained relatively stable over the past 20 years. In summary, the land cover changes in Tongzhou District over the past 20 years have shown a trend of decreasing cropland, increasing impervious surfaces, and increasing forest area. This reflects the impact of urbanization on land use and indicates that Tongzhou District has made efforts and achievements in pursuing sustainable development and building a green and ecological city.

Table 1. 2000-2020 Land Cover Transfer Matrix（km^2）.

2000 \ 2020	Barren	Cropland	Forest	Grassland	Impervious	Water	Total
Barren	0.0055	0.0023	0.0000	0.0000	0.0759	0.0000	0.0837
Cropland	0.0045	476.4026	14.4315	0.0138	192.1693	5.3429	688.3646
Forest	0.0000	0.4103	0.3608	0.0000	0.0768	0.2408	1.0887
Grassland	0.0000	0.0095	0.0000	0.0005	0.0419	0.0019	0.0538
Impervious	0.0027	8.0304	0.0776	0.0000	191.6857	1.3196	201.1160
Water	0.0079	3.8467	0.1113	0.0001	3.2763	4.6823	11.9247
Total	0.0207	488.7018	14.9812	0.0144	387.3259	11.5876	902.6316

3.2. Soil Prediction Mapping and Accuracy Analysis in Different Years

According to Table 2, the accuracy varies among different years but remains at a relatively high level. NIR (Normalized Improvement Ratio) represents the accuracy of predicting classifications without any classification information. In this case, the NIR is 0.1871, indicating that the proportion of accurately predicted classifications based solely on prior probabilities is 0.1871. This suggests that the inclusion of environmental covariates effectively improves the predictive ability of the model, demonstrating a significant impact of the selected environmental covariates on soil formation. The p-value [Acc > NIR] < 2.2e-16 indicates that the actual accuracy is significantly higher than the accuracy based solely on prior probabilities, and the difference is highly significant. This confirms that the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

difference between accuracy and NIR is not due to random factors but rather reflects a genuine improvement in model performance.

Table 2. Index table of soil map in different years.

Year \ Index	Accuracy	95% CI	No Information Rate	P-Value [Acc > NIR]	Kappa
2000	0.5899	(0.5296, 0.6483)	0.1871	< 2.2e-16	0.5451
2005	0.6265	(0.5789, 0.6723)	0.1871	< 2.2e-16	0.5876
2010	0.6043	(0.5442, 0.6622)	0.1871	< 2.2e-16	0.5656
2015	0.6403	(0.5808, 0.6967)	0.1871	< 2.2e-16	0.6046
2020	0.6691	(0.6104, 0.7241)	0.1871	< 2.2e-16	0.6342

In terms of trends, Tongzhou District has experienced an increasing level of land use intensity during the urbanization process. The impervious area has nearly doubled, and the Kappa value of the soil map has also increased, indicating an improvement in the consistency of soil map classification during this period. The accuracy has increased from 58.99% in 2000 to 66.91% in 2020. The 95% CI represents the range of uncertainty for estimating accuracy. Each confidence interval is a range, with the lower and upper limits representing the estimated lower and upper bounds of accuracy. To avoid the influence of chance and randomness, the observation of the 95% CI intervals reveals that the overall accuracy of the soil map is also on an upward trend. This suggests a positive correlation between land use intensity and soil mapping accuracy. As the land use intensity in the study area strengthens, the accuracy of soil mapping in the study area will also increase. This is because human activities tend to develop land in a way that is most suitable for human use, and the utilization patterns indirectly reflect the properties of the soil, which are ultimately manifested in land cover.Table 2 Index table of soil map in different years.

According to Figure 4, over the past 20 years, there have been some changes in the soil of Tongzhou District, but most areas have maintained a relatively stable soil type. The most significant and drastic changes are mainly concentrated in the urban areas of Tongzhou and its surroundings, indicating that human activities have had a certain impact on soil type changes. Overall, the epicenter of the changing areas is located in the western part of Tongzhou District. This is because the western region is closer to the center of Beijing and was the first to be affected by urban expansion. In the past 20 years, the government has continuously enhanced the utilization capacity of land (Xue Y, 2020). For suitable arable land, the government designates the areas as cropland, while for unsuitable soils, such as sandy soils and marshy soils, reasonable transformation measures have been taken to better serve urban development (Lu Q et al., 2017). In the process of historical evolution, the sandy soil areas have been largely afforested and transformed into forest land, which helps prevent sand erosion and improve land quality. The former marshy areas have been completely integrated into the urban area through land leveling and other engineering measures. These examples demonstrate that soil changes in Tongzhou District are influenced to some extent by human activities, especially urban expansion and adjustments in land use policies. By transforming and utilizing different types of soil, the government adapts to the needs of urban development and improves land utilization efficiency. These efforts contribute to the protection of land resources, improvement of the ecological environment, and support for sustainable urban development. It also reflects the trend of soil cover changes influencing soil changes under human influence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

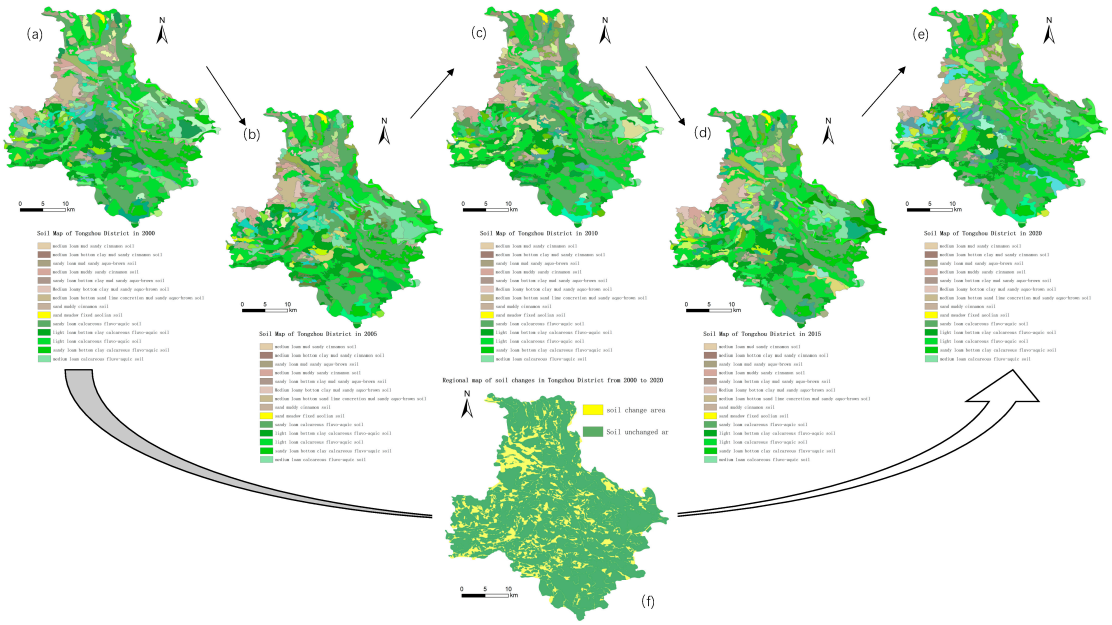

Figure 4. Soil Change Trend in Tongzhou District (a, b, c, e, f represent soil prediction maps for every 5 years from 2000 to 2020, while d shows the soil change regions in Tongzhou District from 2000 to 2020. Due to the large number of soil types (48) in Tongzhou District, it is not feasible to display all of them in the legend. Therefore, the legend focuses on the five largest soil types and includes loam soil and sandy soil.).

3.3. Coupling Analysis of Land Cover Changes and Soil Changes in the Same Space

According to Figure 5, analyzing from the same spatial location, the soil change area in Tongzhou District from 2000 to 2020 accounts for 16.5% of the total area. Overall, there have not been significant changes in soil types, and most of the soil has remained relatively stable. However, among the 16.5% of soil areas that experienced changes, 51.9% of the areas overlap with land cover change areas, indicating a certain degree of synergy between land cover change and soil change, suggesting that changes in land cover have influenced soil types. Further observation of the overlapping areas reveals that 48% of the changes are related to the changes in Impervious areas. Impervious areas are typically the focus of human activities and show significant soil modifications. Therefore, nearly half of the overlapping areas are associated with changes in built-up areas, demonstrating the significant impact of construction and redevelopment on soil structure. Additionally, it is worth noting that among all the land cover types transitioning to Impervious, cropland accounts for 47.6%. This means that a portion of land previously used for agricultural cultivation has been converted into built-up areas, and the remaining 52.4% mostly consists of pre-existing Impervious areas, indicating that Tongzhou District has primarily chosen agricultural land parcels for urban expansion during its development. These findings indicate the coupling between land cover change and soil type change, particularly the significant impact of built-up area change on soil structure. The conversion of cropland to impermeable surfaces represents a considerable proportion. These findings emphasize the significant influence of human activities on soil and highlight the importance of considering soil conservation and sustainable utilization in urban development and land use planning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

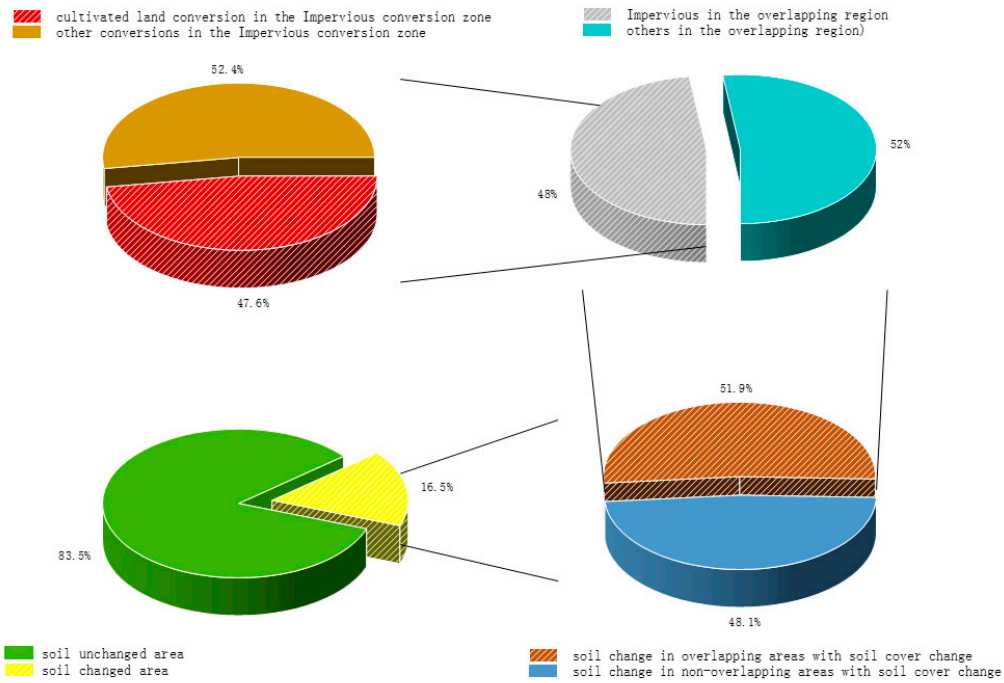

Figure 5. Soil Change and Land Cover Change Coupling Analysis Percentage Map.

3.4. Analysis of Proportional Changes in Soil Types within Cropland over a 20-Year Period

Cropland is the foundation of agricultural production and directly relates to food security and human survival and development. It is also an important component of ecosystems, contributing to maintaining healthy soil and sustainable utilization of water resources (Zhou Y et al., 2021; Liu Y et al., 2021). As shown in Figure 6, analyzing from the perspective of cropland, it can be observed that the quantity of cropland decreased by approximately 29% over the 20-year period, but the main soil types of cropland remained almost unchanged. Sandy loam calcareous fluvo-aquic soil and light loam calcareous fluvo-aquic soil accounted for almost half of the cropland's soil types, and their proportion increased during the evolution of the total cropland quantity reduction over the 20-year period. This is because sandy loam calcareous fluvo-aquic soil is widely distributed in Tongzhou District and is more likely to be utilized by farmers. Light loam calcareous fluvo-aquic soil has moderate permeability, strong capillary force, and certain drought resistance (Han Y et al., 2021). It is rich in soil nutrients and has good soil structure, allowing for coordinated development of water, nutrients, gases, and heat factors within the soil, thus exhibiting a high level of fertility suitable for high-yield crop growth (Ding S et al., 2022). Under the influence of human activities, the soil types of cropland gradually approach suitable types for cultivation. On one hand, farmers choose suitable land for cultivation, and on the other hand, land engineering transforms cropland soil into a suitable soil structure for cultivation (Hasan S S et al., 2020; Shen X et al., 2021). Human activities and environmental changes are the direct factors leading to land cover changes, and land cover changes can directly and indirectly impact soil changes. Taking cropland as an example, human activities act on cropland, causing its soil types to transition towards suitable types for cultivation, and over time, more soil types suitable for cultivation are used for cropland. This direct factor leads to changes in land cover, reflecting the coupling between land cover and soil changes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 August 2023 doi:10.20944/preprints202308.0080.v1

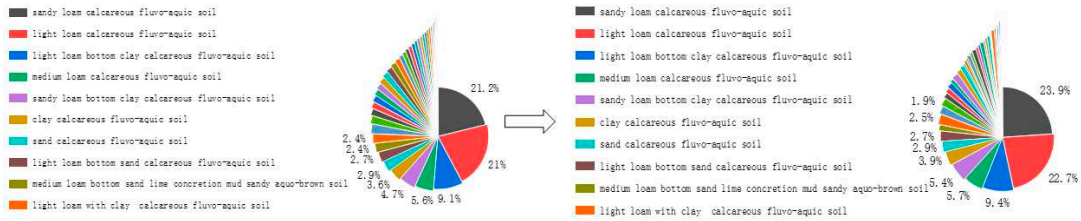

Figure 6. Percentage of Top 10 Soil Types in Cropland for the 20-year Period.

4. Discussion and Conclusion

As a representative study area in plain regions, Tongzhou District exhibits characteristics such as diverse land use types, conflicts between agriculture and urbanization, the impact of human activities on land cover, and soil type diversity, making it representative of land cover and soil changes in plain regions (Li H et al., 2020). This study selected a 20-year time span to comprehensively reveal the land cover changes and soil map evolution in plain regions, aiming to obtain more comprehensive and accurate results. Under the condition of using the same treatment methods for corresponding years and consistent environmental covariates of the same type, a unified expert knowledge-assisted sampling method and random forest mapping method were adopted to enhance the uniformity and credibility of the soil mapping results. By inheriting the expert knowledge contained in the original soil map, the coupling and correlation between soil type changes and land cover changes were further analyzed.

Based on the analysis and results of this study, the following conclusions can be drawn for the plain regions:

1.There is a coupling relationship between land cover changes and soil type changes. Nearly half of the soil change areas overlap with land cover change areas, and human activities play a dominant role in the process of land cover changes. The urbanization process has a significant impact on land use and soil type changes.

2.While human activities influence land cover changes, land cover also affects and reveals soil changes. Human activities tend to lead land use towards suitable soil types, and soil types guide human activities towards maximizing the value of land use.

3.The accuracy of soil mapping in Tongzhou District has improved over the past 20 years, increasing from 58.99% to 66.91%. This indicates that the improvement in land use intensity in plain regions can enhance the credibility of soil mapping.

In the future, further exploration of the driving factors of land cover's influence on soil changes can be conducted. To address the issue of low credibility in soil mapping in plain regions, it is recommended to consider using land cover-related environmental covariates in areas with high land use intensity. Additionally, the integration of high-resolution remote sensing data, ground-based observations, and advanced machine learning algorithms can be considered to improve the accuracy and detail of soil mapping.

Funding: This study was supported by the National Natural Science Foundation of China (42171261).

Conflicts of Interest: The contact author has declared that none of the authors has any competing interests.

References

1. Abowarda A S, Bai L, Zhang C, et al. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale[J].Remote Sensing of Environment, 2021, 255:112301.DOI:10.1016/j.rse.2021.112301.
2. Adugna A, Abegaz A. Effects of land use changes on the dynamics of selected soil properties in northeast Wellega, Ethiopia[J].The Soil, 2016.DOI:10.5194/soil-2-63-2016.
3. Asa, Gholizadeh, Daniel, et al. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging[J].Remote Sensing of Environment, 2018.DOI:10.1016/j.rse.2018.09.015.

4. B, A, Miller, et al. Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks[J].SOIL, 2015.DOI:10.5194/soil-1-217-2015.
5. B S L A, A L K, A B W. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review-ScienceDirect[J].Geoderma, 2019, 352:395-413.DOI:10.1016/j.geoderma.2019.05.031.
6. Chen X, Yu L, Du Z, et al. Distribution of ecological restoration projects associated with land use and land cover change in China and their ecological impacts[J].Science of The Total Environment, 2022, 825:153938-.DOI:10.1016/j.scitotenv.2022.153938.
7. Cui L, Wang J, Sun L, et al. Construction and optimization of green space ecological networks in urban fringe areas: A case study with the urban fringe area of Tongzhou district in Beijing[J].Journal of Cleaner Production, 2020, 276(Pt 2):124266.DOI:10.1016/j.jclepro.2020.124266.
8. C Y Y A B, D Y D, D B W, et al. Increasing contribution from microbial residues to soil organic carbon in grassland restoration chrono-sequence[J].Soil Biology and Biochemistry, 2022.
9. Dai L, Zhan Z, Shu Y, et al. Land Use Change in the Cross-Boundary Regions of a Metropolitan Area: A Case Study of Tongzhou-Wuqing-Langfang[J].Land, 2022, 11.DOI:10.3390/land11020153.
10. Dewage S N S P, Minasny B, Malone B. Disaggregating a Regional Extent Digital Soil Map using Bayesian Area-to-Point Regression Kriging for Farm-Scale Soil Carbon Assessment[J]. 2020.DOI:10.5194/soil-2019-75.
11. De-Cai W, Gan-Lin Z, Xian-Zhang P, et al. Mapping Soil Texture of a Plain Area Using Fuzzy-c-Means Clustering Method Based on Land Surface Diurnal Temperature Difference[J].Pedosphere, 2012, 22(003):394-403.DOI:10.1016/S1002-0160(12)60025-3.
12. Ding S, Xin X, Yang W, et al. Transformation of fertilizer nitrogen in fluvo-aquic soils with different textures and its influencing factors[J].Plant and Soil, 2022.DOI:10.1007/s11104-021-05255-7.
13. Dong Y X L. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China[J].Geoderma: An International Journal of Soil Science, 2019, 340.
14. Dong Z, Wang Z, Liu D, et al. Mapping wetland areas using Landsat-derived NDVI and LSWI: a case study of West Songnen Plain, Northeast China[J].Journal of the Indian Society of Remote Sensing, 2014, 42(3).DOI:10.1007/s12524-013-0357-1.
15. Dunkl I, Liess M. On the benefits of clustering approaches in digital soil mapping: an application example concerning soil texture regionalization[J].2021.DOI:10.5194/soil-2020-102.
16. Feyisa G L, Meilby H, Fensholt R, et al. Automated Water Extraction Index: a new technique for surface water mapping using Landsat imagery.[J].Remote Sensing of Environment, 2014(140).DOI:10.1016/j.rse.2013.08.029.
17. Fromm S F V, Hoyt A M, Acquah G E, et al. Continental-scale controls on soil organic carbon across sub-Saharan Africa[J]. 2020.DOI:10.5194/soil-2020-69.
18. Gan-Lin Z, Feng L, Xiao-Dong S, et al. Recent progress and future prospect of digital soil mapping: A review[J].Journal of Integrative Agriculture, 2017.DOI:CNKI:SUN:ZGNX.0.2017-12-020.
19. Kristensen J A, Balstrm T, Jones A, et al. Development of a harmonised soil profile analytical database for Europe: a resource for supporting regional soil management[J].2019.DOI:10.5194/soil-5-289-2019.
20. Hasan S S, Zhen L, Miah M G, et al. Impact of land use change on ecosystem services: A review[J].Environmental Development, 2020, 34:100527.DOI:10.1016/j.envdev.2020.100527.
21. Han Y, Teng Y, Wang X, et al. Soil Type Driven Change in Microbial Community Affects Poly(butylene adipate- co -terephthalate) Degradation Potential[J].Environmental Science and Technology, 2021.DOI:10.1021/acs.est.0c04850.
22. Hounkpatin K O L, Stendahl J, Lundblad M, et al. Predicting the spatial distribution of soil organic carbon stock in Swedish forests using remotely sensed and site-specific variables[J]. 2020.DOI:10.5194/soil-2020-75.
23. Liu F, Geng X, Zhu A X, et al. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS[J]. Geoderma, 2012, 171(none):44-52.DOI:10.1016/j.geoderma.2011.05.007.
24. Li H, Song W. Evolution of rural settlements in the Tongzhou District of Beijing under the new-type urbanization policies[J].Habitat International, 2020, 101:102198.DOI:10.1016/j.habitatint.2020.102198.

25. Liu, XueqiZhu, A-XingYang, LinPei, TaoLiu, JunzhiZeng, CanyingWang, Desheng. A graded proportion method of training sample selection for updating conventional soil maps[J].Geoderma: An International Journal of Soil Science, 2020, 357.
26. Liu Y, Zhou Y. Reflections on China's food security and land use policy under rapid urbanization[J].Land Use Policy, 2021(109-):109.
27. Luís Renato Silva Taveira, Menezes M D D, Teixeira A F D S, et al. Mapping land use capability in tropical conditions adapting criteria to different levels of agricultural management[J].Ciência e Agrotecnologia, 2018, 42(6):631-642.DOI:10.1590/1413-70542018426026518.
28. Lu Q, Chang N B, Joyce J. Predicting long-term urban growth in Beijing (China) with new factors and constraints of environmental change under integrated stochastic and fuzzy uncertainties[J].Stochastic Environmental Research and Risk Assessment, 2017.DOI:10.1007/s00477-017-1493-x.
29. Moller A B, Beucher A, Pouladi N, et al. Oblique geographic coordinates as covariates for digital soil mapping[J].2019.DOI:10.5194/soil-2019-83.
30. Murray L R, Hamilton E M, Belinda K, et al. Planning spatial sampling of the soil from an uncertain reconnaissance variogram[J]. SOIL Discussions, 2017:1-20.DOI:10.5194/soil-2017-21.
31. Ng W, Minasny B, Mendes W D S, et al. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data[J].SOIL, 2020, 6(2):565-578.DOI:10.5194/soil-6-565-2020.
32. Nussbaum M, Spiess K, Baltensweiler A, et al. Evaluation of digital soil mapping approaches with large sets of environmental covariates[J].SOIL, 2018, 4(1):1-22.DOI:10.5194/soil-2017-1410.5194/soil-2017-14-supplement.
33. Poggio L, De Sousa L M, Batjes N H, et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty[J].SOIL, 2021, 7.DOI:10.5194/soil-7-217-2021.
34. Papadopoulou-Vrynioti K, Alexakis D, Bathrellos G D, et al. Environmental research and evaluation of agricultural soil of the Arta plain, western Hellas[J]. Journal of Geochemical Exploration, 2014, 136:84-92.DOI:10.1016/j.gexplo.2013.10.007.
35. Rostaminia M, Rahmani A, Mousavi S R, et al. Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms[J]. Environmental Monitoring and Assessment, 2021, 193(12):1-17.DOI:10.1007/s10661-021-09543-8.
36. Shen X ,Wang L ,Yang Q, et al. Dynamics of Soil Organic Carbon and Labile Carbon Fractions in Soil Aggregates Affected by Different Tillage Managements[J].Sustainability, 2021, 13.DOI:10.3390/su13031541.
37. Speiser, Lynn J, Miller, et al. A comparison of random forest variable selection methods for classification prediction modeling[J].[2023-07-09].
38. Song W, Li H. Spatial pattern evolution of rural settlements from 1961 to 2030 in Tongzhou District, China[J].Land Use Policy, 2020, 99.DOI:10.1016/j.landusepol.2020.105044.
39. Sun W, Liu S, Zhang X, et al. Estimation of soil organic matter content using selected spectral subset of hyperspectral data[J].Geoderma, 2022, 409:115653-.DOI:10.1016/j.geoderma.2021.115653.
40. Sun H W D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest[J].Engineering Geology, 2021, 281(1).
41. Wang, uHuang, ShengzhiHuang, QiangLeng, GuoyongHan, ZhimingZhao, JingGuo, Yi. Vegetation vulnerability and resistance to hydrometeorological stresses in water- and energy-limited watersheds based on a Bayesian framework[J].Catena: An Interdisciplinary Journal of Soil Science Hydrology-Geomorphology Focusing on Geoecology and Landscape Evolution, 2021, 196(1).
42. Wang M, Guo X, Zhang S, et al. Global soil profiles indicate depth-dependent soil carbon losses under a warmer climate[J].Nature Communications[2023-07-09].DOI:10.1038/s41467-022-33278-w.
43. Wang X , Wu J , Liu Y ,et al.Driving factors of ecosystem services and their spatiotemporal change assessment based on land use types in the Loess Plateau[J].Journal of Environmental Management, 2022, 311:114835-.DOI:10.1016/j.jenvman.2022.114835.
44. Xue Y. Spatial accessibility between commercial and ecological spaces: A case study in Beijing, China[J].[2023-07-09].DOI:10.1515/geo-2020-0333.
45. Yang H, Huang K, Zhang K, et al. Predicting Heavy Metal Adsorption on Soil with Machine Learning and Mapping Global Distribution of Soil Adsorption Capacities[J]. 2021.DOI:10.1021/acs.est.1c02479.

46. Yang J, Huang X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019[J].Earth System Science Data, 2021, 13(8):3907-3925.DOI:10.5194/essd-13-3907-2021.
47. Yang L, Li X, Shi J, et al. Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method[J].Geoderma, 2020, 369:114337.DOI:10.1016/j.geoderma.2020.114337.
48. Zeng G. On the confusion matrix in credit scoring and its analytical properties[J].Communication in Statistics-Theory and Methods, 2019:1-14.DOI:10.1080/03610926.2019.1568485.
49. Zhang J, Feng Q, Zhu M, et al. Effects of land use and land cover change on soil organic carbon storage in the Hexi regions, Northwest China[J].Journal of Environmental Management, 2022(312-).DOI:10.1016/j.jenvman.2022.114911.
50. Zhang J, Zhang M, Tao H Y, et al. A QSAR–ICE–SSD Model Prediction of the PNECs for Per- and Polyfluoroalkyl Substances and Their Ecological Risks in an Area of Electroplating Factories[J].Molecules (Basel, Switzerland), 2021.DOI:10.3390/molecules26216574.
51. Zhang, LeiYang, LinMa, TianwuShen, FeixueCai, YanyanZhou, Chenghu. A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data[J].Geoderma: An International Journal of Soil Science, 2021, 384(1).
52. Zhang Z, Zhang H, Xu E. Enhancing the digital mapping accuracy of farmland soil organic carbon in arid areas using agricultural land use history[J].Journal of Cleaner Production, 2022, 334:130232-.DOI:10.1016/j.jclepro.2021.130232.
53. Zhao M S, Rossiter D G, Li D C, et al. Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index[J]. Ecological Indicators, 2014, 39(apr.):120-133.DOI:10.1016/j.ecolind.2013.12.015.
54. Zou D, Zhao L, Sheng Y, et al. A new map of permafrost distribution on the Tibetan Plateau[J].The Cryosphere, 2017, 11(6):2527-2542.DOI:10.5194/tc-11-2527-2017.
55. Zhou Y, Li X, Liu Y. Cultivated land protection and rational use in China[J].Land Use Policy, 2021, 106(1).DOI:10.1016/j.landusepol.2021.105454.