agronomy

MDPI

Article

Integrated Exploitation of Sentinel-1 Backscatter, Interferometric Coherence, and Texture Features for Digital Mapping of Soil Total Nitrogen Across the Iberian Peninsula

Dongxu Dai [1,2,†], Hongmin Zhang [3,†], Yajun Geng [3,*], Tao Zhou [1,*], Huijie Li [1], Junming Liu [1], Tingting Liu [4], Angela Lausch [5,6,7] and Bingcheng Si [1,8]

1 College of Hydraulic and Civil Engineering, Ludong University, Yantai 264025, China; daidongxumtkx@163.com (D.D.); lihuijie2015@gmail.com (H.L.); liujunming0225@163.com (J.L.); bing.si@usask.ca (B.S.)
2 College of Resources and Environment, Xinjiang Agricultural University, Urumqi 830091, China
3 College of Resources and Environmental Engineering, Ludong University, Yantai 264025, China; 2024110296@m.ldu.edu.cn
4 Institute of Applied Plant Nutrition, University of Göttingen, 37075 Göttingen, Germany; tingting.liu@agr.uni-goettingen.de
5 Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research—UFZ, Permoserstraße 15, 04318 Leipzig, Germany; angela.lausch@ufz.de
6 Department of Architecture, Facility Management and Geoinformation, Institute for Geo-Information and Land Surveying, Anhalt University of Applied Sciences; Seminarplatz 2a, 06846 Dessau, Germany
7 Landscape Ecology Lab, Department of Geography, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
8 Department of Soil Science, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada
* Correspondence: yajungeng@126.com (Y.G.); tao.zhou@ldu.edu.cn (T.Z.)
† These authors contributed equally to this work.

Abstract

Accurate mapping of soil total nitrogen (STN) is fundamental for advancing sustainable and precision soil management. While digital soil mapping (DSM) has increasingly relied on Earth observation (EO) data, the potential of various synthetic aperture radar (SAR) features, particularly interferometric coherence and texture, remains underexplored for large-scale STN prediction. This study aimed to systematically evaluate the potential of multiple Sentinel-1 SAR-derived features, including backscatter coefficients, interferometric coherence, and texture metrics, for modeling and mapping STN across the Iberian Peninsula. We integrated 4296 soil samples from the 2018 LUCAS dataset with multi-source environmental covariates processed via the Google Earth Engine (GEE) platform. Nine modeling scenarios were designed to compare individual and combined contributions of Sentinel-1, Sentinel-2, topographic, and climatic variables using random forest (RF) and extreme gradient boosting (XGBoost) algorithms. The results indicated that the selection of SAR-derived features significantly influences prediction accuracy. Among individual Sentinel-1 feature groups, texture metrics and interferometric coherence outperformed the traditionally used backscatter coefficients, emphasizing their effectiveness in STN mapping. Specifically, texture-based and coherence-based models achieved R^2 values of 0.34 to 0.35 and 0.33, respectively, whereas backscatter-only models yielded the lowest accuracy (R^2 = 0.29 to 0.30). The integration of all three radar categories substantially improved performance (R^2 = 0.39 to 0.42), surpassing the performance of models based solely on Sentinel-2 optical data (R^2 = 0.33 to 0.34). The most comprehensive model, which combined multi-source EO data with topographic and climatic variables, achieved the highest overall accuracy with R^2 values of 0.51 for RF and 0.52 for XGBoost. Variable importance analysis confirmed that satellite-derived variables were the most influential group. Spatial

Check for updates

Academic Editor: Paul Kwan

Received: 29 January 2026
Revised: 23 March 2026
Accepted: 30 March 2026
Published: 1 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

predictions successfully captured the heterogeneity of STN across the peninsula, with higher concentrations in humid, mountainous regions and lower values in arid central plateaus and southern regions. This study demonstrates that integrating diverse Sentinel-1 radar information, particularly coherence and texture, provides a robust alternative or complement to optical data, offering a powerful tool for large-scale soil property mapping.

Keywords: earth observation; soil total nitrogen; Sentinel-1; Sentinel-2; soil mapping

1. Introduction

Soil total nitrogen (STN) plays a fundamental role in sustaining soil fertility and regulating terrestrial nutrient cycles [1]. As a key element of soil nutrient dynamics, STN directly affects plant productivity, microbial processes, and ecosystem functioning [2,3]. Accurate knowledge of STN distribution is crucial for optimizing fertilizer use, minimizing environmental impacts, and promoting sustainable agricultural practices. However, traditional laboratory-based soil surveys are costly and time-consuming, limiting their applicability for large-area assessments [4]. In this context, satellite-based remote sensing offers a powerful and cost-effective alternative for characterizing the spatio-temporal variability of soil properties at large scales. Therefore, developing accurate satellite-based large-scale STN mapping methods has become a pressing priority for advancing soil monitoring and precision agriculture.

The emergence of digital soil mapping (DSM) has provided a valuable framework for predicting soil properties by linking field measurements with environmental covariates [5]. This approach relies on the quantitative relationships between observed soil attributes and auxiliary variables such as climate, topography, land use, and remotely sensed data [6]. Among these, Earth observation (EO) data from satellite missions have become particularly indispensable due to their wide spatial coverage, temporal continuity, and capacity to represent surface conditions relevant to soil processes [7]. Optical remote sensing, in particular, has dominated DSM studies owing to its extensive spectral information and long historical record [8]. However, the reliance on cloud-free imagery restricts its applicability in regions characterized by persistent cloud cover or rainy seasons. In contrast, synthetic aperture radar (SAR) systems operate independently of daylight and weather conditions, offering a robust alternative source of information [9]. As a result, SAR observations have attracted increasing attention as a promising data source for DSM [10,11], although several methodological challenges remain unresolved.

One of the major challenges in SAR-based DSM is the effective selection and utilization of radar-derived features. Diverse SAR features can be extracted through different processing strategies, such as backscatter coefficient, interferometric coherence, and textural metrics [12]. Among these, the backscatter coefficient represents the most direct and widely used SAR observable, capturing the interaction between microwave signals and land surface characteristics [13]. Its relative ease of acquisition has led to its extensive use in radar-based DSM studies [8,11]. Early investigations demonstrated that multi-temporal SAR backscatter contains useful information for predicting soil properties, including soil organic carbon and total nitrogen, although radar-only models often lagged behind optical-based counterparts in terms of accuracy [14,15]. The emergence of cloud-computing platforms such as Google Earth Engine (GEE) has substantially reduced computational barriers, enabling the efficient processing and integration of long-term SAR time series [16]. Recent studies have shown that when multi-frequency, multi-polarization, or multi-orbit backscatter datasets are systematically exploited, SAR-based models can achieve perfor-

mance comparable to, or even exceeding, that of optical-only approaches [17,18]. Despite this progress, most existing studies remain heavily focused on backscatter intensity, while other SAR-derived features relevant to STN mapping have received far less attention.

Interferometric coherence constitutes another important yet underexplored SAR feature for DSM applications. Coherence measures the degree of similarity between repeated SAR acquisitions and has been widely used to monitor surface dynamics [19,20]. Time-series coherence has been successfully applied in agricultural and vegetation-related studies, including crop monitoring, vegetation parameter prediction, and soil moisture retrieval [21–23]. However, its value for predicting soil properties, including STN, is still largely unknown. In addition, SAR backscatter images can be further analyzed using grey-level co-occurrence matrix (GLCM) techniques to extract texture features, which help characterize spatial heterogeneity and reduce speckle-related noise [24,25]. Texture features have shown promise in a wide range of radar-based applications [26–28], yet their contribution to DSM tasks, particularly for STN estimation, has received limited attention. Backscatter, coherence, and texture features provide complementary information that captures different aspects of the land surface. Their combination has been emphasized in many radar-based studies as an effective strategy to enhance information richness and improve model robustness [29,30]. Nevertheless, the joint exploitation of backscatter, coherence, and texture for STN mapping has rarely been evaluated in a systematic manner, leaving a significant research gap regarding their relative and synergistic contributions to model performance. To further illustrate these research gaps, Table 1 summarizes the specific SAR feature types employed in recent representative satellite-based DSM studies, highlighting the limited focus on SAR coherence and texture compared to backscatter.

Table 1. Comparison of satellite data sources and SAR-derived feature types employed in recent representative satellite-based DSM studies.

Study	Year	Study Region	Optical	SAR Backscatter	SAR Coherence	SAR Texture
[3]	2022	Yellow River Basin, China	Yes	No	No	No
[8]	2026	Europe	Yes	Yes	No	No
[10]	2024	Xinjiang, China	Yes	Yes	No	No
[11]	2026	Danjiangkou, China	Yes	Yes	No	No
[14]	2020	Spain	Yes	Yes	No	No
[15]	2019	Heihe River Basin, China	Yes	Yes	No	No
[17]	2025	Europe	Yes	Yes	No	No
[18]	2023	Austria	Yes	Yes	No	No
[28]	2022	Western Australia	Yes	Yes	No	Yes
This study	2026	Iberian Peninsula	Yes	Yes	Yes	Yes

Against this background, the objective of this study was to systematically evaluate the capability of multiple Sentinel-1 SAR-derived features—including backscatter, interferometric coherence, and texture metrics—for modeling and mapping STN across the Iberian Peninsula. Specifically, the objectives are to: (1) evaluate the prediction value of radar backscatter, interferometric coherence, and texture features individually and in combination; (2) unlock the potential of Sentinel-1 radar observations for STN mapping by systematically exploiting complementary SAR information; and (3) compare the performance of SAR-based models with those developed using Sentinel-2 optical imagery, which is commonly employed in DSM. By systematically comparing these feature sets and their synergies, this work seeks to clarify the role of multi-source SAR information in improving large-scale STN mapping and to provide guidance for future integration of radar data into operational DSM frameworks.

2. Materials and Methods

2.1. Study Area

This study was conducted on the Iberian Peninsula, which forms the southwestern extremity of Europe and is bounded by the Atlantic Ocean to the west and north and the Mediterranean Sea to the east and southeast (Figure 1). Its physiography is highly heterogeneous, characterized by an extensive interior plateau surrounded by multiple mountain systems [31]. Topographic variation is considerable, with elevations ranging from sea level to over 3400 m in mountainous areas, and an average altitude close to 600 m [31,32]. The climate is predominantly Mediterranean; however, the northwestern and northern sectors are more strongly influenced by a temperate oceanic climate, and high-elevation areas show localized alpine climatic effects [33]. The pronounced climatic gradients across the peninsula support a wide variety of ecosystems, ranging from sparsely vegetated semi-arid landscapes to dense deciduous broadleaf forests and evergreen coniferous woodlands [34]. Agricultural landscapes are extensive, accounting for nearly half of the total land surface, and include a variety of crops such as olives, cereals, and grapes [35]. STN plays a pivotal role in maintaining crop yields and regulating soil fertility, making it a key determinant of agricultural sustainability and soil health in the Iberian Peninsula. Therefore, accurate spatial prediction of STN is essential for optimizing fertilization strategies and supporting sustainable agricultural development in the region. The prevailing soil groups are Cambisols, Regosols, and Leptosols, with Luvisols, Fluvisols, and Calcisols also widely distributed [36].

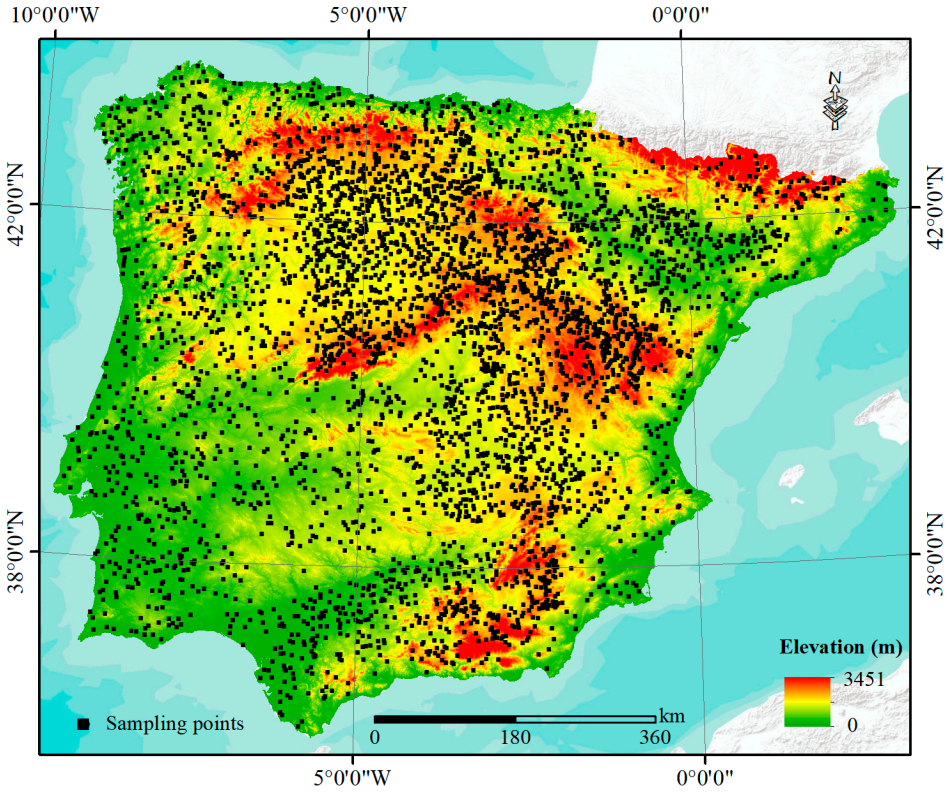


Figure 1. Map of the study area showing soil sampling locations from the LUCAS 2018 Soil Survey and elevation across the Iberian Peninsula.

2.2. Soil Dataset

The soil observations used in this study were obtained from the LUCAS Soil Survey, initiated by Eurostat as part of a European Union (EU)-wide effort to systematically monitor soil conditions [37,38]. The LUCAS Soil Survey represents the only harmonized and

regularly implemented soil sampling campaign across the EU. The first sampling campaign was conducted in 2009, followed by subsequent surveys in 2015, 2018, and most recently in 2022. For the LUCAS Soil Survey, soil sampling locations were distributed systematically across the survey domain at an approximate density of one site per ~200 km^2 [39]. A total of 18,984 georeferenced sampling locations are available in the quality-controlled 2018 LUCAS Soil dataset. At each site, surveyors collected a composite topsoil sample by pooling five subsamples: one taken at the reference point and four taken 2 m away in the cardinal directions (N, E, S, W), with a target composite mass of 500 g [40]. The composite samples were air-dried and sieved to 2 mm before laboratory analysis to determine a range of physical and chemical soil properties. All analytical procedures were performed in an ISO-certified laboratory following standardized protocols [8]. Detailed descriptions of the field sampling procedures, laboratory workflows, and data quality controls are provided in [40]. For this study, we extracted a subset of 4296 LUCAS 2018 samples located within the Iberian Peninsula to support the spatial modeling of STN across the region (Figure 1). The STN values in the study area ranged from 0.20 to 20.00 g/kg, with a mean of 2.01 g/kg and a standard deviation of 1.64 g/kg; additional descriptive statistics are provided in Table S1. The LUCAS 2018 soil dataset was selected for this study as it is the most recent publicly available LUCAS soil dataset, with the LUCAS 2022 soil dataset not yet publicly accessible. Additionally, the operational status of Sentinel-1/2 satellites by 2018 facilitated the effective integration of satellite imagery with the soil observations. The dataset was accessed via the European Soil Data Centre (ESDAC) [38].

2.3. Environmental Data

The environmental covariates used in this study encompassed Sentinel-1 SAR and Sentinel-2 optical imagery, as well as topographic and climatic data. Since the original spatial resolutions of these covariates varied across data sources, all raster layers were resampled to a common spatial resolution of 100 m to ensure consistency in subsequent analyses. Attribute values of each predictor were extracted at the locations of the sampling points. The extracted values, together with the soil data and all raster layers, were integrated within a geographic information system using the UTM/WGS84 projection for subsequent modeling. Details regarding the sources and preprocessing of the environmental data are provided below.

2.3.1. Topographic and Climatic Variables

Topographic variables were derived from the SRTM DEM available on the GEE platform. The SRTM dataset provides near-global elevation coverage at a spatial resolution of approximately 30 m. Based on this DEM, nine terrain indices were calculated using SAGA GIS, including elevation, slope, plan curvature, profile curvature, and topographic wetness index (TWI). Climatic variables were obtained from the WorldClim dataset accessible in GEE, which provides long-term climatic summaries derived from monthly temperature and precipitation records. These layers are interpolated from global meteorological station observations and have been widely used in regional-scale environmental and soil modeling [41,42]. In this study, mean annual temperature (MAT) and mean annual precipitation (MAP) were extracted from the WorldClim dataset and used as climatic inputs for subsequent modelling. The Pearson correlation coefficients between STN and the selected topographic and climatic variables are illustrated in Figure 2. The results indicated that STN was significantly correlated with most variables, particularly showing strong positive correlations with MAP (correlation coefficient of 0.53) and slope (0.37), while exhibiting a negative correlation with TWI (−0.38).

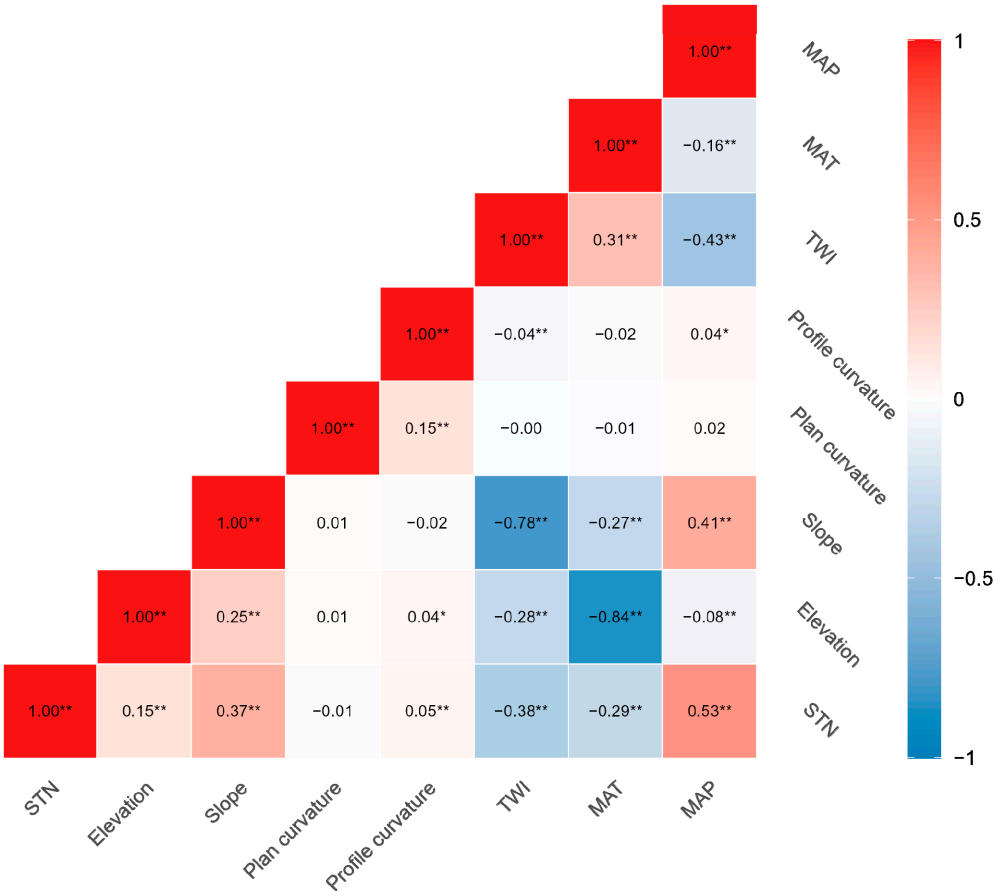


Figure 2. Correlogram showing Pearson correlation coefficients between STN and environmental factors. Values in the cells represent correlation coefficients (*: significant at p <0.05; **: significant at $p < 0.01$).

2.3.2. Sentinel-1

The Sentinel-1 constellation, operated by the European Space Agency (ESA), comprises two C-band SAR satellites that provide all-weather, day-and-night observations [43]. Each satellite has a 12-day revisit cycle, which is reduced to approximately 6 days when both satellites operate concurrently [44]. Sentinel-1 offers multiple imaging modes with varying spatial resolutions and swath widths. Over terrestrial areas, the interferometric wide swath (IW) mode is most commonly employed, offering a spatial resolution of 5 m × 20 m and a swath width of up to 250 km [18]. In this study, Sentinel-1 IW ground range detected (GRD) imagery available on the GEE platform was used. The GRD archive in GEE is provided after standard preprocessing using the Sentinel-1 Toolbox workflow within the Sentinel Application Platform (SNAP) [45], an open-source architecture for the exploitation and analysis of Sentinel data products. We assembled all Sentinel-1 GRD acquisitions covering the study region up to 2018 and then filtered the archive based on image metadata, including acquisition date, orbit direction, and polarization configuration. Monthly composites are widely adopted in previous DSM studies using Sentinel-1 data and have demonstrated strong prediction performance [17,46]. Accordingly, this compositing strategy was implemented in the present study. Specifically, all available Sentinel-1 GRD images within each month were composited using the median function to produce monthly composite datasets. We generated monthly median composites for each polarization and orbit direction, resulting in 12 temporal intervals and a total of 48 Sentinel-1-derived backscatter features used for model development (Table 2).

Table 2. Detailed description of Sentinel-1 derived features used in this study.

Feature Category	Derived Variables	Polarizations	Data Source
Backscatter	Monthly median composites	VV, VH	Sentinel-1 GRD
Texture	Mean, Variance, Homogeneity, Contrast, Dissimilarity, Entropy, Second Moment, Correlation	VV, VH	GLCM-derived texture features from backscatter composites
InSAR Coherence	Seasonal median coherence	VV	Global seasonal Sentinel-1 repeat-pass InSAR coherence dataset [47]

Additionally, we extracted textural information from the VV and VH polarized backscatter composites using the GLCM method. For each backscatter image, eight GLCM texture metrics were calculated: mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation. These metrics were calculated using a 5×5 moving window, averaging across four directions (0°, 45°, 90°, and 135°) [25]. All texture calculations were implemented using the "glcm" package (version 1.6.5) in the R software (version 4.2.3).

Furthermore, we incorporated the global seasonal Sentinel-1 repeat-pass InSAR coherence dataset published by [47]. This product was generated from ~205,000 Sentinel-1 IW acquisitions collected from 1 December 2019 to 30 November 2020 and provides median seasonal coherence estimates for multiple temporal baselines (6/12/18/24/36/48 days). Because HH-polarized coherence is not spatially complete over our study region, only the VV coherence images were retained; specifically, 24 VV coherence images (4 seasons $\times$ 6 baselines) were used in the subsequent analysis (Table 2).

2.3.3. Sentinel-2

The Sentinel-2 constellation operated by ESA consists of two multispectral imaging satellites with a combined revisit interval of 5 days [48]. It forms part of the Copernicus programme and carries the MultiSpectral Instrument (MSI), which acquires data in 13 spectral bands at spatial resolutions of 10, 20, and 60 m across a swath width of about 290 km [49]. Owing to its high spatial resolution, wide spectral coverage, and reliable data quality, Sentinel-2 has become one of the most widely used optical data sources in DSM applications [3,50]. In this study, all Sentinel-2 surface reflectance images (Level–2A) with less than 10% cloud cover available on the GEE platform up to 2018 were collected. The Sentinel-2 surface reflectance dataset in GEE was geometrically, radiometrically and atmospherically corrected using Sen2Cor [51], a specialized atmospheric correction processor designed to generate Sentinel-2 Level-2A surface reflectance data. During preprocessing, cloud and cirrus contamination were removed using the QA60 quality assessment band. Subsequently, all available images were composited using the median reducer function to generate a Sentinel-2 composite. Finally, ten spectral bands (B2–B8A, B11, and B12) were extracted from the composite imagery and used as optical variables for the modeling analyses.

2.4. Predictor Selection

Selecting an appropriate subset of predictors is a crucial step in machine learning–based DSM, as high-dimensional predictor spaces often contain redundant or weakly informative variables that can degrade model performance and increase computational cost [52,53]. To address this issue, recursive feature elimination (RFE) was employed to identify the most relevant predictors without compromising model performance. RFE is a wrapper-based backward elimination technique that iteratively removes the least important predictors based on their contribution to model performance [54]. The procedure starts by fitting a model with all available predictors and ranking them according to

variable importance measures. In each subsequent iteration, the least influential predictor is removed, the model is refitted using the reduced predictor subset, and performance metrics are recalculated [55]. The subset of predictors that yields the lowest prediction error is then retained as the optimal feature set. This approach effectively reduces redundancy and multicollinearity among environmental variables while maintaining prediction accuracy, as widely demonstrated in DSM applications [52,56].

2.5. Prediction Models

In this study, random forest (RF) and extreme gradient boosting (XGBoost) were selected to construct the DSM models, as they are currently the most widely adopted machine learning algorithms in the field, consistently demonstrating superior prediction performance and stability in capturing complex soil–environment relationships. Compared to traditional geostatistical methods or other machine learning models, these ensemble tree-based approaches have become the preferred choice in numerous DSM studies [57,58]. In this study, the "caret" package (version 6.0-94) in the R software (version 4.2.3) was employed to perform RFE for predictor selection and to optimize user-defined parameters for both RF and XGBoost via a grid-search approach [18]. The final RF and XGBoost models were constructed using the "randomForest" (version 4.7-1.1) and "xgboost" (version 1.7.6.1) packages, respectively.

2.5.1. Random Forest

RF is an ensemble tree-based machine learning algorithm widely used for classification and regression tasks. It operates by aggregating the predictions of multiple decision trees trained on different subsets of the data to produce a final output [8]. In RF, a large number of decision trees are constructed independently, with randomness introduced during the training process. This randomization not only improves model generalization but also mitigates the risk of overfitting. The final results from individual trees are then aggregated to produce the final output, typically by averaging for regression tasks or majority voting for classification tasks [59]. Owing to its strong prediction performance, robustness to overfitting, and relatively simple implementation, RF has been widely adopted in DSM applications [8,60]. In RF modeling, two main parameters require tuning: (i) the number of trees in the forest (ntree); and (ii) the number of predictor variables randomly selected at each node (mtry) [61].

2.5.2. Extreme Gradient Boosting

XGBoost, introduced by [62], is an efficient and scalable gradient-boosting algorithm that has been widely used due to its strong prediction performance and computational efficiency. XGBoost is essentially a tree-based ensemble learning algorithm that constructs a series of decision trees in sequence, where each subsequent tree is trained to fit and correct the residuals of the previous one, thereby progressively improving overall prediction accuracy. A distinctive feature of XGBoost is the inclusion of regularization terms in the objective function, which serve to constrain model complexity and reduce the risk of overfitting, thus improving model robustness and generalization [63]. With its strong prediction capability, effective overfitting control, and excellent scalability, XGBoost has been increasingly applied in DSM studies [53,58]. The model includes several key parameters that require tuning, which significantly influence model performance. Detailed descriptions of these parameters are available in [62].

2.6. Evaluation of Model Performance

To systematically assess the prediction capability of different data sources and their combinations, we designed nine prediction scenarios representing distinct groups of input

variables (Table 3). The scenarios included different configurations of Sentinel-1 radar data, Sentinel-2 optical features, and topographic and climatic covariates. Specifically, the first three scenarios were based on individual categories of Sentinel-1 data, including coherence, backscatter, and texture features, while the fourth combined all three radar-derived variables. Scenario 5 was built using Sentinel-2 optical features, followed by Scenario 6, which integrated all variables derived from both Sentinel-1 and Sentinel-2 imagery. Scenario 7 consisted solely of topographic and climatic covariates. Scenario 8 combined Sentinel-2 optical features with topographic and climatic variables, while Scenario 9 combined all Sentinel-1/2-derived features with topographic and climatic variables to form the most comprehensive dataset. These nine scenarios were intentionally structured to (i) compare the modeling performance of different Sentinel-1 feature types (coherence, backscatter, and texture), (ii) assess the added value of integrating Sentinel-1 and Sentinel-2 imagery, and (iii) compare the performance of DSM models constructed using Sentinel-1/2 data either independently or in combination with common topographic and climatic covariates, thereby evaluating their respective predictive power for STN. For each of the nine scenarios, RF and XGBoost algorithms were applied to develop prediction models. Model performance was evaluated using a 10-fold cross-validation approach, in which the dataset was randomly divided into ten equal subsets [64]. During each iteration, nine subsets were used for training and the remaining one for validation. This procedure was repeated ten times so that every subset served as a validation set once, and the overall performance metrics were computed as the mean of the ten iterations. Three statistical metrics were used to assess model accuracy: the coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE).

Table 3. Description of the experimental scenarios and the composition of input variable groups.

Scenario	Coherence	Backscatter	Texture	Sentinel-2	Terrain and Climate	Description
Scenario 1	√					Individual SAR coherence
Scenario 2		√				Individual SAR backscatter
Scenario 3			√			Individual SAR texture
Scenario 4	√	√	√			All SAR-derived features
Scenario 5				√		Sentinel-2 optical features
Scenario 6	√	√	√	√		Sentinel-1/2 imagery
Scenario 7					√	Terrain and climate
Scenario 8				√	√	Sentinel-2, terrain, and climate data
Scenario 9	√	√	√	√	√	Full feature set

3. Results

3.1. Modeling Performance Under Different Prediction Scenarios

The prediction accuracy of STN under the nine modeling scenarios is summarized in Table 4. The results show clear differences in model performance depending on the input variable configuration. Among the scenarios based on individual Sentinel-1 categories (Scenarios 1–3), texture features (Scenario 3) provided the most accurate predictions, achieving R^2 values of 0.34 and 0.35 for RF and XGBoost, respectively. The coherence-based model (Scenario 1) showed intermediate performance within the single-source radar group, yielding an R^2 of 0.33 and an RMSE of 1.37 g/kg for both algorithms. In contrast, backscatter (Scenario 2) exhibited the lowest prediction capability among all tested scenarios, with R^2 values ranging from 0.29 to 0.30 and the highest RMSE (1.45 g/kg). The combination of all radar-derived variables in Scenario 4 significantly improved performance compared to the individual radar scenarios, increasing the R^2 to 0.39 for RF and 0.42 for XGBoost. Sentinel-2 optical features (Scenario 5) provided a prediction accuracy (R^2 = 0.33–0.34) comparable to the individual radar coherence and texture scenarios. However, this perfor-

mance was inferior to the model integrating all three radar categories (Scenario 4), which yielded higher R^2 values (R^2 = 0.39–0.42). Despite this, the integration of all Sentinel-1 and Sentinel-2 features (Scenario 6) led to a marked improvement, with R^2 values increasing from 0.33–0.34 to 0.44–0.45. Topographic and climatic covariates alone (Scenario 7) yielded slightly lower accuracy (R^2 = 0.42–0.43) compared to the combined remote sensing data (Scenario 6). When Sentinel-2 optical features were integrated with topographic and climatic variables (Scenario 8), model performance further improved, reaching R^2 values of 0.47 for RF and 0.48 for XGBoost. Finally, the most comprehensive configuration, Scenario 9 (Figures S1 and S2), which integrated Sentinel-1/2-derived features with topographic and climatic variables, attained the highest overall accuracy, achieving an MAE of 0.71 g/kg, an RMSE of 1.24 g/kg, and an R^2 of 0.51 for RF, while XGBoost performed slightly better with an MAE of 0.70 g/kg, an RMSE of 1.22 g/kg, and an R^2 of 0.52.

Table 4. Comparison of the mean and standard deviation (SD) of soil total nitrogen prediction performance for RF and XGBoost across nine scenarios based on ten-fold cross-validation.

Scenario	RF						XGBoost					
	MAE (g/kg)		RMSE (g/kg)		R^2		MAE (g/kg)		RMSE (g/kg)		R^2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Scenario 1	0.81	0.02	1.37	0.17	0.33	0.04	0.81	0.02	1.37	0.17	0.33	0.05
Scenario 2	0.84	0.02	1.45	0.17	0.30	0.04	0.84	0.03	1.45	0.17	0.29	0.04
Scenario 3	0.81	0.02	1.42	0.17	0.34	0.04	0.80	0.03	1.40	0.17	0.35	0.04
Scenario 4	0.78	0.02	1.37	0.18	0.39	0.05	0.77	0.02	1.34	0.18	0.42	0.05
Scenario 5	0.81	0.03	1.38	0.17	0.33	0.03	0.81	0.03	1.37	0.17	0.34	0.04
Scenario 6	0.76	0.03	1.32	0.18	0.44	0.05	0.75	0.03	1.29	0.18	0.45	0.04
Scenario 7	0.75	0.03	1.28	0.17	0.43	0.04	0.76	0.03	1.29	0.17	0.42	0.04
Scenario 8	0.74	0.03	1.27	0.18	0.47	0.05	0.73	0.03	1.25	0.18	0.48	0.05
Scenario 9	0.71	0.03	1.24	0.19	0.51	0.06	0.70	0.04	1.22	0.19	0.52	0.06

Note: R^2: the squared Pearson correlation coefficient between observed and predicted values.

3.2. Variable Importance

Figure 3 illustrates the relative importance of the top 20 predictors for STN mapping using the RF algorithm under Scenarios 4 and 9. The relative importance of variables was quantified using the increase in mean squared error (%IncMSE) derived from the RF algorithm. In the model based solely on radar coherence, backscatter, and texture features (Scenario 4), the cumulative importance of these categories was 69%, 17%, and 14%, respectively. This ranking order of collective importance remained consistent in the model incorporating all environmental variables (Scenario 9), where these categories accounted for 14%, 6%, and 2% of the total importance, respectively. In Scenario 4, the top five most important variables were all derived from coherence features, with the two highest-ranked variables corresponding to coherence features with the shortest temporal baselines. In the comprehensive model (Scenario 9), satellite-derived variables exhibited the highest cumulative importance (52%), followed by climatic (34%) and topographic (14%) variables. Specifically, the top five variables in this model comprised two climatic variables (ranked 1st and 2nd), two satellite-derived variables, and one DEM derivative.

3.3. Spatial Prediction of STN

The spatial distributions of STN across the Iberian Peninsula, predicted by the RF and XGBoost models under six comprehensive scenarios (Scenarios 4–9), are illustrated in Figures 4 and 5, respectively, while the results for individual SAR-derived features (Scenarios 1–3) are provided in Figures S3 and S4. Overall, the RF and XGBoost models exhibited consistent spatial patterns across the study area. A distinct spatial heterogeneity

in STN content was observed in all predicted maps, which aligns well with the observed gradient in the original LUCAS samples (Figure S5). High STN concentrations were predominantly distributed in the northern and northwestern regions, corresponding to the mountainous and humid areas. In contrast, the central plateau and the southern regions generally exhibited lower STN values, forming a clear north–south gradient. Predictions based solely on remote sensing data (Scenario 4 and Scenario 5) captured fine-scale spatial variability. Scenario 7, derived exclusively from topographic and climatic covariates, produced smoother spatial transitions and clearly delineated the broader regional trends of STN distribution. Scenario 9, which integrated Sentinel-1/2 features with environmental covariates, combined the advantages of the other scenarios, preserving both the broad spatial trends and the detailed local variations.

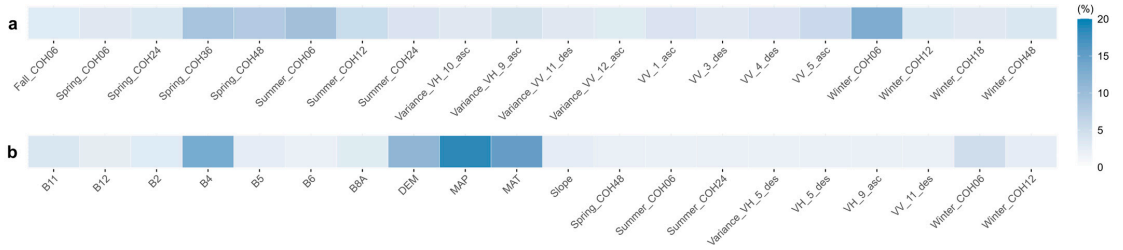


Figure 3. Ranking of the top 20 most important variables identified by the RF model. (**a**) Scenario 4: model constructed using only Sentinel-1 radar features (coherence, backscatter, and texture); (**b**) Scenario 9: comprehensive model combining Sentinel-1/2 features with climatic and topographic covariates.

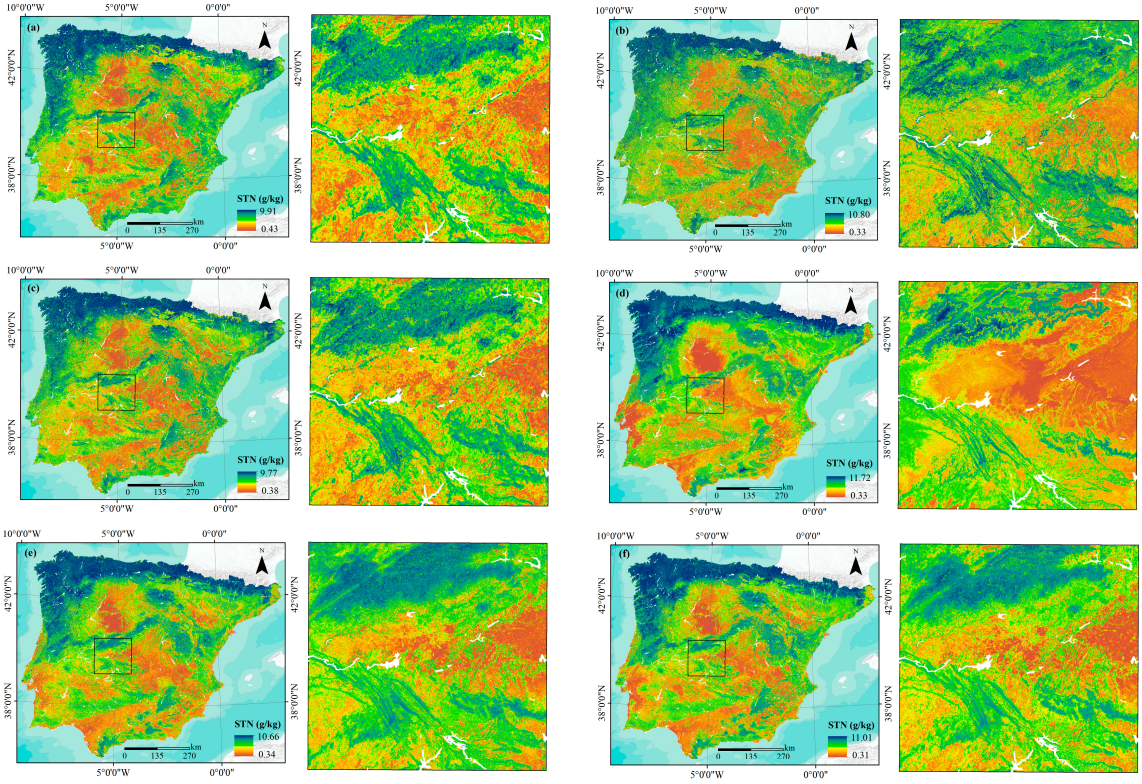


Figure 4. Spatial distribution of STN predicted by the RF model across the Iberian Peninsula under six modeling scenarios (Scenarios 4–9). The subplots correspond to predictions based on different variable combinations: (**a**) Scenario 4 (all SAR-derived features); (**b**) Scenario 5 (Sentinel-2 optical features); (**c**) Scenario 6 (Sentinel-1/2 imagery); (**d**) Scenario 7 (terrain and climate data); (**e**) Scenario 8 (Sentinel-2, terrain, and climate data); and (**f**) Scenario 9 (full feature set). The black rectangles and adjacent panels illustrate zoomed-in details of the prediction maps.

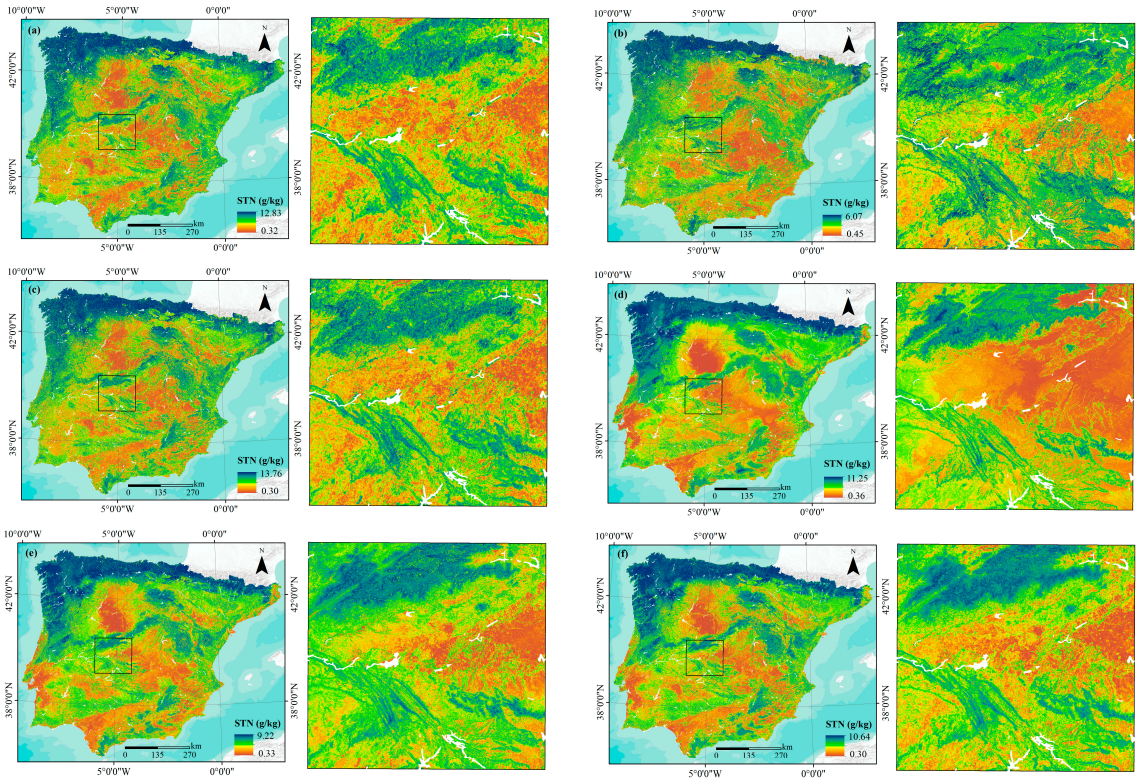


Figure 5. Spatial distribution of STN predicted by the XGBoost model across the Iberian Peninsula under six modeling scenarios (Scenarios 4–9). The subplots correspond to predictions based on different variable combinations: (**a**) Scenario 4 (all SAR-derived features); (**b**) Scenario 5 (Sentinel-2 optical features); (**c**) Scenario 6 (Sentinel-1/2 imagery); (**d**) Scenario 7 (terrain and climate data); (**e**) Scenario 8 (Sentinel-2, terrain, and climate data); and (**f**) Scenario 9 (full feature set). The black rectangles and adjacent panels illustrate zoomed-in details of the prediction maps.

4. Discussion

4.1. Comparison of Prediction Capabilities Among Different Scenarios

Our results demonstrate that the selection of input variables and the integration of multi-source data significantly influence the prediction accuracy of STN. The comparative analysis of different modeling scenarios highlights the varying capacities and mutual complementarity of Sentinel-1 radar features, Sentinel-2 optical data, and traditional environmental covariates (topography and climate) in capturing the spatial variability of STN. Among the individual Sentinel-1 feature groups, the superior performance of texture features and interferometric coherence compared to backscatter provides important insights into the utility of SAR data for STN mapping. Remote sensing-based texture features capture the spatial heterogeneity and structural patterns of the landscape [65], which often correlate with the spatial distribution of soil properties. Coherence also provided competitive results, outperforming backscatter. This suggests that coherence observation provides a unique and valuable radar-derived feature for STN mapping that is frequently overlooked in traditional SAR-based DSM [46,66]. This finding is supported by the results of [22,67], who reported that coherence provides valuable information regarding surface characteristics that is not captured by backscatter features.

Furthermore, the integration of all radar-derived variables significantly outperformed scenarios relying on a single radar-derived feature group. This improvement is consistent with reports in the existing literature indicating that different SAR features provide complementary information regarding surface characteristics [19,67]. The combined use of these features effectively mitigates the limitations of individual SAR parameters, providing a

more comprehensive characterization of the land surface as it relates to STN variability. Our results indicate that Sentinel-2 optical features provided a prediction accuracy comparable to individual SAR features (coherence and texture) but were inferior to the model incorporating all radar features. This differs from previous DSM research comparing Sentinel-1 and Sentinel-2, which reported that Sentinel-2 achieved higher prediction accuracy than Sentinel-1 [68]. This discrepancy may be due to our integration of multiple radar-derived features from the Sentinel-1 sensor, which further mined the potential of Sentinel-1 for STN mapping. While Sentinel-2 is the most commonly used optical sensor in DSM due to its high prediction accuracy, resolution, and data quality [7], the superior performance of Sentinel-1 features in this study underscores the reliability of SAR as a robust alternative or supplement, particularly in regions where cloud cover limits optical data availability. However, the joint use of Sentinel-1 radar and Sentinel-2 optical data significantly improved prediction accuracy. This result indicates that these two sensors provide highly complementary information, consistent with previous DSM studies that emphasized the advantages of integrating optical and SAR sensors to enhance model performance [11,69]. Nevertheless, those previous efforts primarily relied on the backscatter coefficients of radar sensors, while ignoring the great potential of coherence and texture features in mapping soil properties.

Our results show that while topographic and climatic variables alone demonstrated favorable prediction performance, their performance was lower than the combined use of Sentinel-1 radar and Sentinel-2 optical data. This result further emphasizes the indispensable role of Sentinel-1/2 features. This differs from previous DSM research that reported higher prediction power for topographic and climatic variables [68]. These differences may stem from the different strategies utilized for Sentinel-1 and Sentinel-2 data; specifically, we utilized three categories of SAR-derived features from long-term Sentinel-1 observations to fully exploit the potential of the sensor. The highest overall accuracy was achieved in Scenario 9, which combined all remote sensing features with topographic and climatic data (R^2 = 0.51 for RF and 0.52 for XGBoost), which is consistent with previous DSM studies integrating EO and traditional environmental variables [15,70]. However, the moderate R^2 values suggest that there is still room for further enhancement in model performance. Unlike many previous DSM studies that integrated a large number of variables representing soil-forming factors, this study primarily focused on evaluating the potential of different radar-derived metrics for STN mapping [71,72]. Future research could focus on identifying new explanatory variables that better capture historical soil-forming processes and fine-scale soil heterogeneity to further improve modeling performance. This study focused on nine scenarios specifically designed to address the research objectives; however, the impact of alternative data configurations, such as different Sentinel-1 polarization modes and acquisition dates, warrants further investigation in future research. Notably, all EO and traditional environmental variables in this study were obtained and processed via the GEE cloud platform. The use of GEE allowed for the efficient harmonization of massive multi-source datasets and provided high-performance parallel computing capabilities, which greatly enhanced the efficiency of large-scale environmental covariate extraction and modeling. Our findings suggest that for future large-scale STN mapping, the integration of multi-source EO data with traditional environmental variables within the GEE framework remains a highly robust strategy for improving model prediction accuracy and reliability.

4.2. Analysis of Feature Importance

In the model based exclusively on radar-derived features (Scenario 4), InSAR coherence observations exhibited a dominant influence, accounting for 69% of the total variable importance, which significantly outperformed backscatter (17%) and texture features (14%).

While previous DSM studies utilizing SAR data have predominantly focused on backscatter features [69,73], our analysis suggests that backscatter alone may not fully exploit the potential of Sentinel-1 data for STN mapping. Instead, our findings demonstrate that coherence provides more critical information for STN modeling than backscatter features alone. Notably, the two highest-ranked variables in Scenario 4 were coherence features with the shortest temporal baselines. This result is supported by many studies that have emphasized the advantages of short temporal baseline coherence in surface monitoring [74,75]. Consequently, utilizing multi-temporal coherence, particularly from short temporal baselines, can provide information that is complementary to backscatter and texture features, thereby significantly enhancing the capability of SAR data for STN mapping. Overall, this underscores the untapped potential of using InSAR coherence derived from the Sentinel-1 time series as a key predictor in DSM frameworks.

When incorporating all environmental covariates (Scenario 9), satellite-derived variables remained the most influential group in terms of cumulative importance (52%), reaffirming the essential role of Sentinel radar and optical data for STN mapping. Climatic variables played a pivotal role in this comprehensive model, occupying the top two positions in the individual variable ranking and accounting for 34% of the total importance. Temperature and precipitation are key factors regulating microbial activity, organic matter decomposition, and nitrogen mineralization rates, which determine STN levels [76]. These results are consistent with large-scale DSM studies where climatic factors often emerge as primary predictors for soil chemical properties [77,78]. In addition to satellite and climate data, topographic variables also played an important role in modeling STN. DEM derivatives, especially elevation, were among the most influential predictors, paralleling previous research that identifies topography as a critical factor in soil formation and nutrient distribution [68]. Terrain attributes, such as elevation and slope, influence local hydrothermal conditions and the redistribution of water and sediments, thereby affecting the spatial heterogeneity of STN [79].

4.3. Analysis of Spatial Patterns

The spatial distributions of STN predicted by our models show strong consistency with established pedological knowledge of the Iberian Peninsula and align with spatial trends observed in continental-scale STN mapping products [80]. Our models, which integrated Sentinel-1/2 derived features with environmental factors, preserved both regional trends and fine-scale variability in STN, providing a comprehensive view of spatial distribution across the Iberian Peninsula. These results align with the growing body of literature that emphasizes the value of combining remote sensing and environmental covariates for accurate large-scale soil property predictions [17,81]. Geographically, the highest STN concentrations were clearly identified in the northern and northwestern territories and the high-altitude mountain ranges. This distribution pattern aligns with findings from large-scale soil surveys in Europe, which attribute elevated nitrogen levels to the cool, humid conditions in these regions [40]. In these regions, the cool and humid conditions favor organic matter accumulation, which results in high contents of STN in the topsoil. Furthermore, these high-value areas predominantly correspond to forest and grassland ecosystems. The dense vegetation cover in these mountainous and humid zones contributes significantly to nitrogen inputs through litterfall and root turnover. Conversely, lower STN values were prevalent in the arid inland plateaus, the southern regions, and the Mediterranean coastal zones. These areas generally experience higher temperatures and limited rainfall, conditions that accelerate the decomposition of organic matter. Additionally, these low-altitude regions are extensively utilized for intensive agriculture [33]. The lower STN content in these croplands can be attributed to management practices such as tillage and

harvest removal, which deplete soil nutrient reservoirs. Despite the potential application of fertilizers, croplands in these Mediterranean zones typically exhibit lower STN levels compared to natural vegetation [82].

5. Conclusions

This study systematically evaluated the capability of diverse Sentinel-1 SAR information, including backscatter coefficient, interferometric coherence, and texture features, for modeling and mapping STN across the Iberian Peninsula. By integrating diverse SAR-derived features with Sentinel-2 optical imagery and traditional environmental covariates within the GEE platform, we clarified the synergistic potential of diverse EO data in DSM frameworks. Overall, the results showed that the selection of radar-derived features significantly influences STN prediction accuracy. Among the individual Sentinel-1 feature groups, texture and interferometric coherence features outperformed the backscatter features commonly used in DSM. Moreover, the integration of multiple radar features effectively exploits the complementary information of SAR sensors, and the combined use of backscatter, coherence, and texture improved performance compared to models based on a single radar feature group. Although Sentinel-2 optical features achieved performance levels comparable to a single radar feature group, the model incorporating all radar features outperformed the optical-only model, indicating that Sentinel-1 radar information provides a robust alternative or supplement to optical data for STN mapping. Furthermore, the synergistic use of Sentinel-1 and Sentinel-2 further enhanced prediction accuracy, demonstrating a substantial complementarity between microwave and multispectral observations, while variable importance analysis confirmed that EO data was the most influential group in the modeling process. Under this integrated framework, the comprehensive model integrating multi-source EO data with topographic and climatic variables achieved the highest overall accuracy, with R^2 values reaching 0.51 for RF and 0.52 for XGBoost. The GEE platform also proved indispensable for the efficient harmonization and processing of multi-source geospatial data. Finally, the predicted maps successfully captured the spatial heterogeneity of STN across the Iberian Peninsula, with high STN concentrations concentrated in the humid, mountainous northern regions, whereas lower values were observed in the arid central plateaus and southern regions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy16070750/s1, Table S1: Descriptive statistics of soil total nitrogen (g/kg) for the LUCAS 2018 soil dataset used in this study; Figure S1: Predicted versus observed soil total nitrogen based on 10-fold cross-validation for Scenarios 4 to 9 using the RF algorithm. The red line represents the linear fit of observed and predicted values, and the black line is the 1:1 line; Figure S2: Predicted versus observed soil total nitrogen based on 10-fold cross-validation for Scenarios 4 to 9 using the XGBoost algorithm. The red line represents the linear fit of observed and predicted values, and the black line is the 1:1 line; Figure S3: Spatial distribution of STN predicted by the RF model across the Iberian Peninsula under scenarios based on individual SAR features (Scenarios 1–3). The subplots correspond to predictions using single categories of radar information: (a) Scenario 1 (SAR coherence); (b) Scenario 2 (SAR backscatter); and (c) Scenario 3 (SAR texture). The black rectangles and adjacent panels illustrate zoomed-in details of the prediction maps; Figure S4: Spatial distribution of STN predicted by the XGBoost model across the Iberian Peninsula under scenarios based on individual SAR features (Scenarios 1–3). The subplots correspond to predictions using single categories of radar information: (a) Scenario 1 (SAR coherence); (b) Scenario 2 (SAR backscatter); and (c) Scenario 3 (SAR texture). The black rectangles and adjacent panels illustrate zoomed-in details of the prediction maps; Figure S5: Map of the original LUCAS soil sampling points showing the spatial variability in measured STN content across the study area.

Author Contributions: Conceptualization, D.D., H.Z., T.Z., Y.G., H.L., J.L., T.L., A.L. and B.S.; Data curation, D.D., H.Z., T.Z. and Y.G.; writing—original draft preparation, D.D., H.Z., T.Z. and Y.G.; supervision, B.S., T.Z. and Y.G.; writing—review & editing, D.D., H.Z., T.Z., Y.G., H.L., J.L., T.L., A.L. and B.S.; methodology, D.D., H.Z., T.Z., Y.G., H.L., J.L., T.L., A.L. and B.S.; formal analysis, D.D., H.Z., T.Z. and Y.G.; investigation, D.D., H.Z., T.Z., Y.G., H.L., J.L., T.L., A.L. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this research was provided by the National Natural Science Foundation of China (No. 42501059, No. 42577364), the Natural Science Foundation of Shandong Province (No. ZR2023QD100), the Yantai City Science and Technology Innovation Development Plan (No. 2023JCYJ092), and the Shandong Province Higher Education Institutions' Youth Entrepreneurship Team Program (No. 2022KJ289).

Data Availability Statement: The dataset used in our study is derived from the LUCAS 2018 Soil Database, which is provided by the European Soil Data Centre. Any researcher can request and download the original LUCAS 2018 dataset directly from the European Soil Data Centre website at: https://esdac.jrc.ec.europa.eu/euso (accessed on 22 July 2025).

Acknowledgments: The authors acknowledge the use of AI tools for linguistic editing and grammar refinement during the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huang, T.; Yang, N.; Lu, C.; Qin, X.; Siddique, K.H.M. Soil organic carbon, total nitrogen, available nutrients, and yield under different straw returning methods. *Soil Tillage Res.* **2021**, *214*, 105171. [CrossRef]
2. Huang, B.; Sun, W.; Zhao, Y.; Zhu, J.; Yang, R.; Zou, Z.; Ding, F.; Su, J. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. *Geoderma* **2007**, *139*, 336–345. [CrossRef]
3. Xu, Y.; Li, B.; Shen, X.; Li, K.; Cao, X.; Cui, G.; Yao, Z. Digital soil mapping of soil total nitrogen based on Landsat 8, Sentinel 2, and WorldView-2 images in smallholder farms in Yellow River Basin, China. *Environ. Monit. Assess.* **2022**, *194*, 282. [CrossRef] [PubMed]
4. Santra, P.; Kumar, M.; Panwar, N.R.; Yadav, R.S. Digital Soil Mapping: The Future Need of Sustainable Soil Management. In *Geospatial Technologies for Crops and Soils*; Mitran, T., Meena, R.S., Chakraborty, A., Eds.; Springer: Singapore, 2021; pp. 319–355.
5. Fathololoumi, S.; Biswas, A. Incorporating digital soil mapping-derived soil properties for enhanced soil moisture prediction. *Soil Tillage Res.* **2026**, *256*, 106901. [CrossRef]
6. Mahmoudabadi, E.; Karimi, A.; Haghnia, G.H.; Sepehr, A. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. *Environ. Monit. Assess.* **2017**, *189*, 500. [CrossRef] [PubMed]
7. Radočaj, D.; Gašparović, M.; Jurišić, M. Open Remote Sensing Data in Digital Soil Organic Carbon Mapping: A Review. *Agriculture* **2024**, *14*, 1005. [CrossRef]
8. Zhou, T.; Geng, Y.; Li, H.; Zhang, H.; Li, H.; Liu, J.; Li, S.; Liu, T.; Pan, J.; Si, B.; et al. Multi-frequency SAR and optical data integration for continental-scale digital mapping of soil chemical properties across Europe. *Soil Tillage Res.* **2026**, *258*, 107034. [CrossRef]
9. Ottinger, M.; Kuenzer, C. Spaceborne L-Band Synthetic Aperture Radar Data for Geoscientific Analyses in Coastal Land Applications: A Review. *Remote Sens.* **2020**, *12*, 2228. [CrossRef]
10. Liang, H.; Japaer, G.; Yu, T.; Zhang, L.; Chen, B.; Lin, K.; Ju, T.; Zhao, Y.; Pei, T.; Aizizi, Y. Comparison of global and zonal modeling strategies—A case study of soil organic matter and C:N ratio mapping in Altay, Xinjiang, China. *Ecol. Inform.* **2024**, *84*, 102882. [CrossRef]
11. Ye, L.; Lin, X.; Huang, L.; Ji, K.; Drost, T.A.; Wan, W.; Zheng, Y.; Huang, C.; Zuo, J. Digital mapping of soil fauna abundance and diversity using Sentinel-1/2 and environmental attributes. *Soil Tillage Res.* **2026**, *255*, 106796. [CrossRef]
12. Santos, A.R.; Barbosa, M.A.G.A.; Anjinho, P.S.; Parizotto, D.; Mauad, F.F. Integrated Use of Synthetic Aperture Radar and Optical Data in Mapping Native Vegetation: A Study in a Transitional Brazilian Cerrado–Atlantic Forest Interface. *Remote Sens.* **2024**, *16*, 2559. [CrossRef]
13. Tian, Y.; Ackermann, K.; McCarthy, C.; Sternberg, T.; Purevtseren, M.; Limuge, C.; Hagiwara, K.; Ogawa, K.; Hobara, S.; Hoshino, B. Seasonal Dynamics of the Land-Surface Characteristics in Arid Regions Retrieved by Optical and Microwave Satellite Data. *Remote Sens.* **2024**, *16*, 3143. [CrossRef]

14. Wang, X.; Zhang, Y.; Atkinson, P.M.; Yao, H. Predicting soil organic carbon content in Spain by combining Landsat TM and ALOS PALSAR images. *IJAEO* **2020**, *92*, 102182. [CrossRef]
15. Zhou, T.; Geng, Y.; Chen, J.; Sun, C.; Haase, D.; Lausch, A. Mapping of Soil Total Nitrogen Content in the Middle Reaches of the Heihe River Basin in China Using Multi-Source Remote Sensing-Derived Variables. *Remote Sens.* **2019**, *11*, 2934. [CrossRef]
16. Mayer, T.; Poortinga, A.; Bhandari, B.; Nicolau, A.P.; Markert, K.; Thwal, N.S.; Markert, A.; Haag, A.; Kilbride, J.; Chishtie, F.; et al. Deep learning approach for Sentinel-1 surface water mapping leveraging Google Earth Engine. *ISPRS Open J. Photogramm. Remote Sens.* **2021**, *2*, 100005. [CrossRef]
17. Wang, X.; Geng, Y.; Zhou, T.; Zhao, Y.; Li, H.; Liu, Y.; Li, H.; Ren, R.; Zhang, Y.; Xu, X.; et al. Mapping the soil C:N ratio at the European scale by combining multi-year Sentinel radar and optical data via cloud computing. *Soil Tillage Res.* **2025**, *245*, 106311. [CrossRef]
18. Zhou, T.; Lv, W.; Geng, Y.; Xiao, S.; Chen, J.; Xu, X.; Pan, J.; Si, B.; Lausch, A. National-scale spatial prediction of soil organic carbon and total nitrogen using long-term optical and microwave satellite observations in Google Earth Engine. *Comput. Electron. Agric.* **2023**, *210*, 107928. [CrossRef]
19. Jacob, A.W.; Vicente-Guijalba, F.; Lopez-Martinez, C.; Lopez-Sanchez, J.M.; Litzinger, M.; Kristen, H.; Mestre-Quereda, A.; Ziółkowski, D.; Lavalle, M.; Notarnicola, C.; et al. Sentinel-1 InSAR Coherence for Land Cover Mapping: A Comparison of Multiple Feature-Based Classifiers. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 535–552. [CrossRef]
20. Wu, W.; Guo, S.; Shao, Z.; Li, D. Urban Impervious Surface Extraction Based on Deep Convolutional Networks Using Intensity, Polarimetric Scattering and Interferometric Coherence Information from Sentinel-1 SAR Images. *Remote Sens.* **2023**, *15*, 1431. [CrossRef]
21. Ezzahar, J.; Chehbouni, A.; Ouaadi, N.; Madiafi, M.; Said, K.; Er-Raki, S.; Laamrani, A.; Chakir, A.; Chabaane, Z.L.; Zribi, M. Sentinel-1 Backscatter and Interferometric Coherence for Soil Moisture Retrieval in Winter Wheat Fields Within a Semiarid South-Mediterranean Climate: Machine Learning Versus Semiempirical Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 2256–2271. [CrossRef]
22. Zhao, Q.; Xie, Q.; Peng, X.; Lai, K.; Wang, J.; Fu, H.; Zhu, J.; Song, Y. Understanding the Temporal Dynamics of Coherence and Backscattering Using Sentinel-1 Imagery for Crop-Type Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6875–6893. [CrossRef]
23. Ghosh, S.M.; Behera, M.D. Aboveground biomass estimates of tropical mangrove forest using Sentinel-1 SAR coherence data—The superiority of deep learning over a semi-empirical model. *Comput. Geosci.* **2021**, *150*, 104737. [CrossRef]
24. Fakhri, F.; Gkanatsios, I. Quantitative evaluation of flood extent detection using attention U-Net case studies from Eastern South Wales Australia in March 2021 and July 2022. *Sci. Rep.* **2025**, *15*, 12377. [CrossRef] [PubMed]
25. Numbisi, F.N.; Van Coillie, F.M.B.; De Wulf, R. Delineation of Cocoa Agroforests Using Multiseason Sentinel-1 SAR Images: A Low Grey Level Range Reduces Uncertainties in GLCM Texture-Based Mapping. *ISPRS Int. J. Geo.-Inf.* **2019**, *8*, 179. [CrossRef]
26. Attarchi, S. Extracting impervious surfaces from full polarimetric SAR images in different urban areas. *IJRS* **2020**, *41*, 4644–4663. [CrossRef]
27. Iyyappan, M.; Ramakrishnan, S.S. Enhancing land cover classification for multispectral images using hybrid polarimetry SAR data. *IJRS* **2020**, *41*, 6718–6754. [CrossRef]
28. Nguyen, T.T.; Pham, T.D.; Nguyen, C.T.; Delfos, J.; Archibald, R.; Dang, K.B.; Hoang, N.B.; Guo, W.; Ngo, H.H. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *ScTEn* **2022**, *804*, 150187. [CrossRef]
29. Pulella, A.; Aragão Santos, R.; Sica, F.; Posovszky, P.; Rizzoli, P. Multi-Temporal Sentinel-1 Backscatter and Coherence for Rainforest Mapping. *Remote Sens.* **2020**, *12*, 847. [CrossRef]
30. Bian, S.; Xie, C.; Tian, B.; Guo, Y.; Zhu, Y.; Yang, Y.; Zhang, M.; Yang, Y.; Ruan, Y. Classification of plants based on time-series SAR coherence and intensity data in Yancheng coastal wetland. *IJRS* **2025**, *46*, 859–881. [CrossRef]
31. Oliva, M.; Serrano, E.; Fernández-Fernández, J.M.; Palacios, D.; Fernandes, M.; García-Ruiz, J.M.; López-Moreno, J.I.; Pérez-Alberti, A.; Antoniades, D. The Iberian Peninsula. In *Periglacial Landscapes of Europe*; Oliva, M., Nývlt, D., Fernández-Fernández, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 43–68.
32. Serrano, E.; Oliva, M.; González-García, M.; López-Moreno, J.I.; González-Trueba, J.; Martín-Moreno, R.; Gómez-Lende, M.; Martín-Díaz, J.; Nofre, J.; Palma, P. Post-little ice age paraglacial processes and landforms in the high Iberian mountains: A review. *Land Degrad. Dev.* **2018**, *29*, 4186–4208. [CrossRef]
33. Zhou, T.; Geng, Y.; Lv, W.; Xiao, S.; Zhang, P.; Xu, X.; Chen, J.; Wu, Z.; Pan, J.; Si, B.; et al. Effects of optical and radar satellite observations within Google Earth Engine on soil organic carbon prediction models in Spain. *J. Environ. Manag.* **2023**, *338*, 117810. [CrossRef]
34. Caparros-Santiago, J.A.; Rodriguez-Galiano, V. Analysing long-term spatiotemporal land surface phenology patterns over the Iberian Peninsula using 250 m MODIS EVI2 data. *ScTEn* **2024**, *954*, 176453. [CrossRef]

35. Pérez-Jordà, G.; Peña-Chocarro, L.; Pardo-Gordó, S. Fruits arriving to the west. Introduction of cultivated fruits in the Iberian Peninsula. *J. Archaeol. Sci. Rep.* **2021**, *35*, 102683. [CrossRef]
36. Fernández-Getino, A.P.; Duarte, A.C. Soil management guidelines in Spain and Portugal related to EU Soil Protection Strategy based on analysis of soil databases. *Catena* **2015**, *126*, 146–154. [CrossRef]
37. Panagos, P.; Van Liedekerke, M.; Jones, A.; Montanarella, L. European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy* **2012**, *29*, 329–338. [CrossRef]
38. Panagos, P.; Van Liedekerke, M.; Borrelli, P.; Köninger, J.; Ballabio, C.; Orgiazzi, A.; Lugato, E.; Liakos, L.; Hervas, J.; Jones, A.; et al. European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *Eur. J. Soil Sci.* **2022**, *73*, e13315. [CrossRef]
39. Ungaro, F.; Tarocco, P.; Aprea, A. Digital Soil Mapping of Soil Macronutrients (N, P, K) in Emilia-Romagna (NE Italy): A Regional Baseline for the EU Soil Monitoring Law. *Land* **2025**, *14*, 2142. [CrossRef]
40. Fernandez-Ugalde, O.; Scarpa, S.; Orgiazzi, A.; Panagos, P.; Van Liedekerke, M.; Marechal, A.; Jones, A. *LUCAS 2018 Soil Module*; Presentation of dataset and results; Publications Office of the European Union: Luxembourg, 2022; p. 31144.
41. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *IJCli* **2017**, *37*, 4302–4315. [CrossRef]
42. Poggio, L.; Simonetti, E.; Gimona, A. Enhancing the WorldClim data set for national and regional applications. *ScTEn* **2018**, *625*, 1628–1643. [CrossRef] [PubMed]
43. Jiao, M.; Li, X.; Sun, X.; Wu, J.; Zhao, T.; Tang, R.; Bai, Y. Multi-Layer Soil Moisture Profiling Based on BKA-CNN by Integrating Sentinel-1/2 SAR and Multispectral Data. *Agronomy* **2025**, *15*, 2542. [CrossRef]
44. Shang, J.; Liu, J.; Poncos, V.; Geng, X.; Qian, B.; Chen, Q.; Dong, T.; Macdonald, D.; Martin, T.; Kovacs, J.; et al. Detection of Crop Seeding and Harvest through Analysis of Time-Series Sentinel-1 Interferometric SAR Data. *Remote Sens.* **2020**, *12*, 1551. [CrossRef]
45. Paluba, D.; Papale, L.G.; Laštovička, J.; Perivolioti, T.M.; Kalaitzis, P.; Mouratidis, A.; Karadimou, G.; Štych, P. Tracking Burned Area Progression in an Unsupervised Manner Using Sentinel-1 SAR Data in Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15612–15634. [CrossRef]
46. Luo, C.; Zhang, W.; Zhang, X.; Liu, H. Mapping the soil organic matter content in a typical black-soil area using optical data, radar data and environmental covariates. *Soil Tillage Res.* **2024**, *235*, 105912. [CrossRef]
47. Kellndorfer, J.; Cartus, O.; Lavalle, M.; Magnard, C.; Milillo, P.; Oveisgharan, S.; Osmanoglu, B.; Rosen, P.A.; Wegmüller, U. Global seasonal Sentinel-1 interferometric coherence and backscatter data set. *Sci. Data* **2022**, *9*, 73. [CrossRef]
48. Dou, W.; Zhang, W.; He, S.; Li, X.; Luo, C. Research on Spatiotemporal Combination Optimization of Remote Sensing Mapping of Farmland Soil Organic Matter Considering Annual Variability. *Agronomy* **2025**, *15*, 2714. [CrossRef]
49. Trevisiol, F.; Mattivi, P.; Mandanici, E.; Bitelli, G. Cross-Sensors Comparison of Popular Vegetation Indexes From Landsat TM, ETM=, OLI, and Sentinel MSI for Time-Series Analysis Over Europe. *ITGRS* **2024**, *62*, 1–16. [CrossRef]
50. Castaldi, F.; Chabrillat, S.; Don, A.; van Wesemael, B. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. *Remote Sens.* **2019**, *11*, 2121. [CrossRef]
51. Sandamali, J.; Narine, L.L. A data-driven, cloud-based approach for forest aboveground biomass mapping using GEDI and other earth observation data: An ecoregion-specific Investigation across the state of Alabama, USA. *GeoIn* **2025**, *40*, 2465446. [CrossRef]
52. Zhang, X.; Chen, S.; Xue, J.; Wang, N.; Xiao, Y.; Chen, Q.; Hong, Y.; Zhou, Y.; Teng, H.; Hu, B.; et al. Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping. *Geoderma* **2023**, *432*, 116383. [CrossRef]
53. Chen, Y.; Ma, L.; Yu, D.; Zhang, H.; Feng, K.; Wang, X.; Song, J. Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests. *Ecol. Indic.* **2022**, *135*, 108545. [CrossRef]
54. Bulut, O.; Tan, B.; Mazzullo, E.; Syed, A. Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. *Information* **2025**, *16*, 476. [CrossRef]
55. Srujana, B.; Verma, D.; Naqvi, S. Machine Learning vs. Survival Analysis Models: A study on right censored heart failure data. *Commun. Stat.-Simul. Comput.* **2024**, *53*, 1899–1916. [CrossRef]
56. Rostaminia, M.; Rahmani, A.; Mousavi, S.R.; Taghizadeh-Mehrjardi, R.; Maghsodi, Z. Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms. *Environ. Monit. Assess.* **2021**, *193*, 815. [CrossRef] [PubMed]
57. Liu, F.; Rossiter, D.G.; Zhang, G.-L.; Li, D.-C. A soil colour map of China. *Geoderma* **2020**, *379*, 114556. [CrossRef]
58. Geng, Y.; Zhou, T.; Zhang, Z.; Cui, B.; Sun, J.; Zeng, L.; Yang, R.; Wu, N.; Liu, T.; Pan, J.; et al. Continental-scale mapping of soil pH with SAR-optical fusion based on long-term earth observation data in google earth engine. *Ecol. Indic.* **2024**, *165*, 112246. [CrossRef]
59. Mellor, A.; Haywood, A.; Stone, C.; Jones, S. The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification. *Remote Sens.* **2013**, *5*, 2838–2856. [CrossRef]
60. Dharumarajan, S.; Hegde, R. Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use Manag.* **2022**, *38*, 135–149. [CrossRef]

61. Fox, E.W.; Hill, R.A.; Leibowitz, S.G.; Olsen, A.R.; Thornbrugh, D.J.; Weber, M.H. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environ. Monit. Assess.* **2017**, *189*, 316. [CrossRef] [PubMed]
62. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. Xgboost: Extreme Gradient Boosting, R Package Version 0.4-2. 2015. Available online: http://r.meteo.uni.wroc.pl/web/packages/xgboost/vignettes/xgboost.pdf (accessed on 22 July 2025).
63. Mathew, A.; Yadalam, P.K.; Radeideh, A.; Hady, S.; Swed, R.; Cheema, R.; Mousa AL-Mohammad, M.; Alsaegh, M.; Shetty, S. A hybrid approach to predicting and classifying dental impaction: Integrating regularized regression and XG boost methods. *Front. Oral Health* **2025**, *6*, 1524206. [CrossRef]
64. Amirian-Chakan, A.; Minasny, B.; Taghizadeh-Mehrjardi, R.; Akbarifazli, R.; Darvishpasand, Z.; Khordehbin, S. Some practical aspects of predicting texture data in digital soil mapping. *Soil Tillage Res.* **2019**, *194*, 104289. [CrossRef]
65. Hofmann, S.; Everaars, J.; Schweiger, O.; Frenzel, M.; Bannehr, L.; Cord, A.F. Modelling patterns of pollinator species richness and diversity using satellite image texture. *PLoS ONE* **2017**, *12*, e0185591. [CrossRef] [PubMed]
66. Peng, L.; Wu, X.; Feng, C.; Gao, L.; Li, Q.; Xu, J.; Li, B. Assessing the potential of multi-source remote sensing data for cropland soil organic matter mapping in hilly and mountainous areas. *Catena* **2024**, *245*, 108312. [CrossRef]
67. Ding, Y.; Ding, Q.; Yang, J.; Shao, Z.; Huang, X. PRISI: A novel piecewise radar impervious surface index for urban monitoring using Sentinel-1 data. *IJAEO* **2022**, *114*, 103033. [CrossRef]
68. Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *ScTEn* **2020**, *729*, 138244. [CrossRef]
69. Azizi, K.; Garosi, Y.; Ayoubi, S.; Tajik, S. Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. *Soil Tillage Res.* **2023**, *229*, 105681. [CrossRef]
70. Wang, H.; Zhang, X.; Wu, W.; Liu, H. Prediction of Soil Organic Carbon under Different Land Use Types Using Sentinel-1/-2 Data in a Small Watershed. *Remote Sens.* **2021**, *13*, 1229. [CrossRef]
71. Liu, F.; Wu, H.; Zhao, Y.; Li, D.; Yang, J.-L.; Song, X.; Shi, Z.; Zhu, A.X.; Zhang, G.-L. Mapping high resolution National Soil Information Grids of China. *Sci. Bull.* **2022**, *67*, 328–340. [CrossRef]
72. Veloso, G.V.; Mello, D.C.d.; Silva, L.V.; Fernandes-Filho, E.I.; Rosas, J.T.F.; Mello, F.A.d.O.; de Souza, J.J.L.L.; Francelino, M.R.; dos Santos, S.R.; Firmino, F.H.T.; et al. Strategies for predictive digital soil mapping by geophysical, remote sensing and machine learning approaches. *Catena* **2026**, *264*, 109822. [CrossRef]
73. Shafizadeh-Moghadam, H.; Minaei, F.; Talebi-khiavi, H.; Xu, T.; Homaee, M. Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. *Catena* **2022**, *212*, 106077. [CrossRef]
74. Voormansik, K.; Zalite, K.; Sünter, I.; Tamm, T.; Koppel, K.; Verro, T.; Brauns, A.; Jakovels, D.; Praks, J. Separability of Mowing and Ploughing Events on Short Temporal Baseline Sentinel-1 Coherence Time Series. *Remote Sens.* **2020**, *12*, 3784. [CrossRef]
75. Wang, S.; Zhang, G.; Chen, Z.; Cui, H.; Zheng, Y.; Xu, Z.; Li, Q. Surface deformation extraction from small baseline subset synthetic aperture radar interferometry (SBAS-InSAR) using coherence-optimized baseline combinations. *GIScience Remote Sens.* **2022**, *59*, 295–309. [CrossRef]
76. Zhou, W.; Li, S.; Sun, X.; Zou, R.; He, L.; Yu, J.; Zhao, G.; Chen, Z.; Bai, X.; Zhang, J. Soil Organic Carbon and Total Nitrogen Stocks and Interactions with Soil Metal Oxides in Different Climatic Zones. *Forests* **2023**, *14*, 1572. [CrossRef]
77. Žížala, D.; Minařík, R.; Skála, J.; Beitlerová, H.; Juřicová, A.; Reyes Rojas, J.; Penížek, V.; Zádorová, T. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. *Catena* **2022**, *212*, 106024. [CrossRef]
78. Yang, R.-M.; Liu, L.-A.; Zhang, X.; He, R.-X.; Zhu, C.-M.; Zhang, Z.-Q.; Li, J.-G. The effectiveness of digital soil mapping with temporal variables in modeling soil organic carbon changes. *Geoderma* **2022**, *405*, 115407. [CrossRef]
79. Liu, Y.; Gao, P.; Zhang, L.; Niu, X.; Wang, B. Spatial heterogeneity distribution of soil total nitrogen and total phosphorus in the Yaoxiang watershed in a hilly area of northern China based on geographic information system and geostatistics. *Ecol. Evol.* **2016**, *6*, 6807–6816. [CrossRef]
80. Ballabio, C.; Lugato, E.; Fernández-Ugalde, O.; Orgiazzi, A.; Jones, A.; Borrelli, P.; Montanarella, L.; Panagos, P. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* **2019**, *355*, 113912. [CrossRef]
81. Ma, G.; Ding, J.; Han, L.; Zhang, Z.; Ran, S. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. *Reg. Sustain.* **2021**, *2*, 177–188. [CrossRef]
82. Jones, A.; Fernandez-Ugalde, O.; Scarpa, S. Lucas 2015 topsoil survey. *Present. Dataset* **2020**, *30332*, 616084.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.