

Article

Machine Learning Prediction of Soil Nutrients and pH in Tibetan Alpine Croplands: A Comparison of Nine Algorithms

Chenjun Zhao ^{1,2} , Shaowei Li ¹, Yuan Tian ¹ , Chengqun Yu ¹, Gang Fu ¹ , Zhiming Zhong ¹ and Wei Sun ^{1,*}¹ Lhasa Alpine Ecosystem Research Station, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wsun@igsnr.ac.cn

Abstract

Soils underpin terrestrial ecosystem functioning and exert strong control over farmland productivity, nutrient turnover, environmental quality, and agricultural sustainability. Reliable prediction of soil nutrients and pH is therefore important for precision fertilization, soil-quality assessment, and the sustainability of farmland management, particularly in fragile alpine agroecosystems. However, the comparative performance of candidate algorithms for estimating soil properties in alpine croplands of the Qinghai-Tibetan Plateau remains poorly characterized. Here, we assessed nine approaches, namely random forest (RF), generalized boosting regression (GBR), multiple linear regression (MLR), artificial neural network (ANN), generalized linear regression (GLR), conditional inference tree (CIT), extreme gradient boosting (eXGB), support vector machine (SVM), and recursive partitioning regression tree (RRT), for predicting soil carbon, nitrogen, phosphorus, potassium, and pH from annual mean temperature, annual precipitation, annual solar radiation, and maximum normalized difference vegetation index (NDVImax). Field observations were obtained from 103 sampling plots, with four individual soil samples collected per plot and depth, resulting in 412 observations for each depth layer. Clear contrasts in model performance emerged under alpine agricultural conditions. Across all soil variables and depths, eXGB and RF showed the best overall performance, with the lowest mean relative bias values of 1.35% and 1.55%, respectively, and the lowest mean RMSE values of 0.43 and 0.47. These results indicate that nonlinear ensemble algorithms better captured the heterogeneous relationships between environmental covariates and soil properties than linear models or single-tree approaches. Overall, RF and eXGB provide a practical methodological basis for precision fertilization, soil-quality assessment, and long-term sustainability in fragile alpine agroecosystems.



Academic Editor: Anna De Marco

Received: 8 August 2026

Revised: 1 September 2026

Accepted: 3 September 2026

Published: 27 September 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: data mining; extreme gradient boosting; random forest; global change; Tibetan Plateau; alpine cropland

1. Introduction

Farmland soils provide the material foundation for crop production, nutrient cycling, terrestrial carbon storage, and agricultural sustainability. Variation in soil carbon and nutrient pools directly affects agricultural output, nutrient-use efficiency, environmental quality, and the sustainability of cultivated land management [1]. Soil also constitutes a core component of terrestrial ecosystems [2], and the spatial distribution of soil properties

such as carbon (C), nitrogen (N), phosphorus (P), potassium (K), and pH strongly shapes ecosystem processes. Reliable estimation of these properties is therefore important not only for precision agriculture but also for soil-quality assessment, long-term land stewardship, and regional sustainability [3]. This challenge is particularly acute in Tibetan farmlands, where high environmental heterogeneity complicates soil monitoring and sustainable agricultural management [4,5].

Agricultural soils in Tibet develop under the joint influence of a high-elevation climate, complex terrain, and region-specific management practices, resulting in distinctive pedogenic pathways and spatial heterogeneity [6]. Conventional soil surveys are costly, laborious, and spatially restrictive [7], which limits their ability to resolve fine-scale variability in cold agroecosystems and to support timely decisions for sustainable fertilization and soil conservation. For this reason, digital soil mapping has become an increasingly valuable alternative, especially in regions characterized by sparse observations and high topographic complexity [8,9]. Within that framework, machine learning methods are attractive because they can capture nonlinear responses and accommodate complex combinations of environmental covariates [10]. Random forest (RF), generalized boosting regression (GBR), and support vector machine (SVM) have all been widely used in soil prediction, while ensemble strategies often perform especially well when predictors are high-dimensional and highly collinear [11]. These approaches can therefore contribute to more efficient soil monitoring and provide technical support for the sustainability of cropland management.

Despite these advances, important uncertainties remain in the prediction of soil properties in croplands of the Tibetan Plateau. First, the relative suitability of different machine learning algorithms under high-altitude agricultural conditions has not been systematically evaluated [12]. Second, the sources of prediction uncertainty and the mechanisms behind model error remain unclear, particularly when sample sizes are modest and data heterogeneity is pronounced [13]. Third, the practical limitations of conventional linear models and simple recursive tree approaches in complex soil–environment systems still require more rigorous empirical evaluation [14,15]. These knowledge gaps restrict the application of digital soil prediction to precision fertilization, soil-quality assessment, and the sustainability of agricultural development in fragile alpine regions. To address these issues, we compared nine algorithms using field observations from 103 sampling plots in Tibetan croplands, with four individual soil samples collected per plot and depth. Our objectives were to evaluate predictive accuracy and robustness under alpine farmland conditions, examine model responses to data heterogeneity and noise, and identify suitable approaches for soil nutrient management and farmland sustainability in Tibet.

2. Materials and Methods

2.1. Study Area and Soil Sampling

The investigation was carried out in alpine croplands of Xizang, China, a region with a plateau monsoon climate characterized by relatively warm, humid summers and cold, dry winters (Figure S1) [16]. Mean annual precipitation and air temperature are approximately 400 mm and 8 °C, respectively. The dominant soil type is alpine fluvo-aquic soil, corresponding to Fluvents in USDA Soil Taxonomy, and the main crops include *Medicago sativa* L., *Elymus nutans* Griseb., and *Hordeum vulgare* L. [17]. During the 2023 growing season, soil was sampled on 15 August from two depth intervals (0–10 and 10–20 cm) within 2 m × 2 m quadrats using a 4 cm diameter auger. Four soil cores were collected randomly at each depth within each quadrat and analyzed as individual samples. Thus, 412 observations were obtained for each soil depth layer and were used separately for model construction. To ensure regional representativeness, sampling was implemented using a stratified random design based on major agroecological zones, elevation gradi-

ents, and dominant soil types derived from existing maps. Samples were allocated in proportion to stratum area so that the dataset captured the principal environmental and agricultural gradients of Tibetan Plateau croplands. The distribution of sampling sites is shown in Figure S1. Soil physicochemical properties were analyzed following standard soil analytical procedures used in previous studies of alpine agricultural ecosystems in Tibet [3]. Soil organic carbon (SOC) was measured using the potassium dichromate oxidation method, total nitrogen (TN) using the Kjeldahl procedure, total phosphorus (TP) using the molybdenum–antimony colorimetric method, and soil pH using a calibrated pH meter. Total potassium (TK) was determined by flame atomic absorption spectrophotometry. The density variables, including soil organic carbon density (SOCD), total nitrogen density (TND), total phosphorus density (TPD), and total potassium density (TKD), represent nutrient stocks per unit ground area within a given soil layer. They were calculated from nutrient concentration, soil bulk density, and soil layer thickness and are expressed as kg/m^2 . The calculation was as follows: nutrient density (kg/m^2) = nutrient concentration (g/kg) $\times$ bulk density (g/cm^3) $\times$ soil thickness (cm)/100.

2.2. NDVImax and Climate Covariates

NDVImax for the growing season (May–September) at 30 m spatial resolution was obtained from the National Ecosystem Science Data Center of China [18] (<http://www.nesdc.org.cn>; accessed on 10 July 2024). This dataset represents the maximum NDVI extracted from all cloud-free Landsat observations within the seasonal window for each year. Climatic predictors, including annual mean temperature, annual precipitation, and annual solar radiation, were downloaded from the National Tibetan Plateau Data Center [19]. Both the NDVImax product and the climate data used in this study correspond to 2023. These variables were used as predictive covariates to characterize spatial variation in vegetation activity and climatic background during the sampling year, rather than to infer direct short-term causal effects on soil carbon, nutrient pools, or pH. Because these soil properties are shaped by long-term pedogenic processes, nutrient cycling, and agricultural management, the modeled relationships should be interpreted as predictive associations between soil observations and environmental covariates.

2.3. Model Development

Training and validation subsets were generated using an 80:20 split. The validation subset ($n = 82$) was designed to preserve stratification while reducing spatial dependence through two constraints: validation quadrats were spaced at least 500 m apart (Figure S1), and sampling also respected stratification by elevation, slope, and NDVI. This design reduced the risk of data leakage while maintaining representativeness across soil nutrient variables and pH. To further evaluate whether residual spatial dependence affected model validation, we conducted supplementary Moran’s I diagnostics for both observed validation values and model residuals. All statistical analyses were conducted in R 4.4.3 with widely used and validated packages [20]. Multiple linear regression (MLR) was fitted using the base ‘stats’ package because of its interpretability in describing linear relationships [21]. Generalized linear regression (GLR) extended this framework to accommodate broader response distributions. To capture nonlinear structure and interactions, we employed recursive regression trees (RRT; ‘rpart’) [21] and conditional inference trees (CITs), both of which partition predictor space recursively. Random forest (RF; ‘randomForest’) [22] was used to mitigate overfitting risk and quantify variable importance through an ensemble of decorrelated trees. Generalized boosted regression (GBR; ‘gbm’) [23] iteratively fits trees to residual error, whereas extreme gradient boosting (eXGB; ‘xgboost’) [24] extends boosting with regularization to enhance computational efficiency and predictive accuracy. Support

vector machine (SVM; 'e1071') [25] identifies an optimal separating hyperplane to improve generalization, and artificial neural network (ANN; 'rminer') [24] was used as a flexible nonlinear learner inspired by biological neural architectures.

Specifically, the nine regression algorithms were fitted using model-specific parameter settings. For GBR, Gaussian loss was used, and the number of trees was selected by 2-fold cross-validation repeated 100 times from an initial maximum of 1000 trees. The eXGB, ANN, GLR, and CIT models were implemented using 'rminer::fit()' with the 'heuristic5' search strategy, with 'mpar = NULL', 'feature = "none"', 'scale = "none"', and 'transform = "none"'. SVM was implemented using the default settings of 'e1071::svm()'; CART was fitted using 'rpart::rpart()' with 'method = "anova"' and 'cp = 0.01'; and LM was fitted as ordinary least-squares regression. For RF, the final 'ntree' and 'mtry' values were derived from the saved final RF model parameter files. Detailed parameter settings and reproducible code are provided in Supplementary Table S1 and Code S1.

Taken together, these algorithms represent a broad set of modeling strategies, including ensemble learning (RF, GBR, and eXGB), statistical regression (MLR and GLR), recursive partitioning (RRT and CIT), margin-based learning (SVM), and adaptive nonlinear fitting (ANN). Evaluating such a diverse model set improves reproducibility and provides a clearer basis for judging which approaches are best suited to heterogeneous soil datasets.

2.4. Model Accuracy Evaluation

Because the candidate models rely on different optimization schemes, training error was quantified using the criteria reported by each package. RF minimized mean squared error (MSE) and returned R^2 values (Table S1), whereas GBR used mean training error together with cross-validation thresholds to determine stopping points (Table S2). SVM performance was characterized from residual structure and margin-related diagnostics (Table S3), while MLR and RRT primarily relied on R^2 -based optimization (Tables S4 and S5). GLR, ANN, CIT, and eXGB likewise followed model-specific error-reduction procedures (Table S6). The source data and R code used to document the model parameter-setting procedures are provided in Table S7 and Code S1.

To compare all models on a common basis, we used four standardized validation statistics: relative bias, root mean squared error (RMSE), the slope of the regression of predicted values against observed values, and R^2 . This combination captures systematic over- or underestimation, overall prediction error magnitude, proportional agreement, and explained variance, thereby providing a balanced evaluation framework for ecological modeling.

2.5. Spatial Autocorrelation Complementarity Analysis

To further assess whether spatial autocorrelation affected model validation, we conducted an additional Moran's I diagnostic analysis for the validation dataset. Because four soil samples were collected at each sampling location, validation observations were first aggregated to the sampling-location level based on geographic coordinates. Moran's I was then calculated for both observed soil-property values and model residuals, defined as observed minus predicted values. Spatial weights were constructed using a 4-nearest-neighbor approach, and statistical significance was assessed using 999 permutation tests. This analysis was used as a supplementary diagnostic to evaluate residual spatial dependence in model predictions. The detailed Moran's I results are provided in Figure S7 and Tables S8 and S9.

3. Results

3.1. Model Construction

Model development was based on 330 quadrat-level observations, which provided a sufficiently broad basis for comparing algorithm performance. RF models used between 591 and 970 trees (Table S1), placing them between the larger GBR ensembles (1000 trees; Table S2) and the SVM models defined by 265–285 support vectors. RF also exhibited the highest training performance, with R^2 values from 0.67 to 0.91 (Table S1), clearly exceeding the corresponding values for multiple linear regression (0.05–0.27; Table S4) and recursive regression (0.38–0.64; Table S5). Substantial variation in training error was also evident among ANN, GLR, CIT, and eXGB (Table S6).

3.2. Overall Model Performance

When all 18 combinations of soil depth and soil property were considered together, eXGB and RF emerged as the two best-performing algorithms. Mean relative bias across variables and depths was lowest for eXGB (1.35%), followed by RF (1.55%) and GBR (1.85%), whereas GLR, SVM, ANN, MLR, and CIT showed substantially larger mean absolute relative bias, ranging from 3.00% to 3.49% (Figure S2).

A similar ranking was observed for RMSE. eXGB and RF yielded the lowest average RMSE values (0.43 and 0.47, respectively), while MLR, GLR, and ANN each exceeded 1.20 on average (Figure S3). Their advantage was particularly evident for nutrient concentrations and density variables, which may involve stronger nonlinear relationships with environmental covariates. In the 0–10 cm layer, for example, RF and eXGB showed low RMSE values for SOC, TN, TP, and TK, and the same pair of models also ranked lowest or near-lowest for SOCD, TKD, SOC, TN, TP, TK, and pH in the 10–20 cm layer. These patterns indicate that tree-based ensemble models generalized more effectively than both linear approaches and several flexible alternatives.

3.3. Prediction of Soil Nutrient Densities

For nutrient-density variables in the 0–10 cm layer, the validation results showed close agreement between observed and predicted values for RF and eXGB (Figures 1, 2 and S2). The strongest agreement occurred for TND, TPD, and TKD, whereas SOCD displayed slightly greater dispersion. eXGB produced the lowest RMSE for TPD (0.01 kg P/m²), TKD (0.36 kg K/m²), and TND (0.02 kg N/m²), while RF achieved the lowest or a very similar error for SOCD (0.24 kg C/m²). Bias for the ensemble models (RF, GBR, and eXGB) generally remained within $\pm 5\%$ (Table 1).

Table 1. Relative bias (%) between predicted and observed values for soil organic carbon density (SOCD, kg/m²), total nitrogen density (TND, kg/m²), total phosphorus density (TPD, kg/m²), total potassium density (TKD, kg/m²), soil organic carbon (SOC, g/kg), total nitrogen (TN, g/kg), total phosphorus (TP, g/kg), total potassium (TK, g/kg), and pH in the 0–10 cm and 10–20 cm layers (n = 82).

Soil Depth	Variable	RF	GBR	SVM	MLR	RRT	GLR	ANN	CIT	eXGB
0–10 cm	SOCD	4.40	4.16	1.92	1.29	6.88	3.49	1.34	4.28	3.08
	TND	2.77	3.87	−1.70	2.70	4.12	3.73	4.59	4.06	1.81
	TPD	−1.75	−3.41	−8.53	−4.31	−4.44	−3.68	−3.28	−0.76	−1.61
	TKD	1.20	2.71	2.34	0.88	3.05	1.64	0.95	3.14	1.07
	SOC	−2.45	−0.46	−7.89	−4.22	−0.88	−3.45	−4.23	−5.63	−2.48
	TN	0.96	2.36	2.46	7.01	8.79	6.58	7.10	6.89	−0.39
	TP	0.18	3.35	−0.81	5.35	2.92	8.41	5.82	−0.79	−1.20
	TK	−0.65	−0.87	−1.30	−1.57	−0.85	−1.65	−1.58	−1.11	−0.41
	pH	−0.16	0.14	0.21	0.38	0.44	0.18	0.42	0.01	−0.22

Table 1. Cont.

Soil Depth	Variable	RF	GBR	SVM	MLR	RRT	GLR	ANN	CIT	eXGB
10–20 cm	SOCD	1.47	3.77	5.47	10.89	6.62	10.84	11.13	9.05	−0.15
	TND	2.22	0.39	−4.22	1.49	4.56	−1.85	0.48	1.26	1.84
	TPD	1.46	1.27	−4.25	2.48	−1.15	2.09	2.47	2.04	1.57
	TKD	−1.74	−0.56	−0.29	−0.78	−0.62	1.29	1.76	1.22	−2.19
	SOC	3.23	−1.40	−3.75	3.18	2.90	3.28	3.18	5.86	1.86
	TN	1.95	−2.65	−7.26	−4.47	1.49	−6.85	−4.38	−6.90	2.33
	TP	−0.73	−1.27	−2.96	2.46	−0.37	2.44	2.71	0.29	−1.61
	TK	−0.46	−0.40	2.86	1.13	0.67	1.35	1.13	−0.75	−0.48
	pH	0.04	0.34	0.29	−0.06	0.17	−0.01	−0.06	−0.03	−0.03

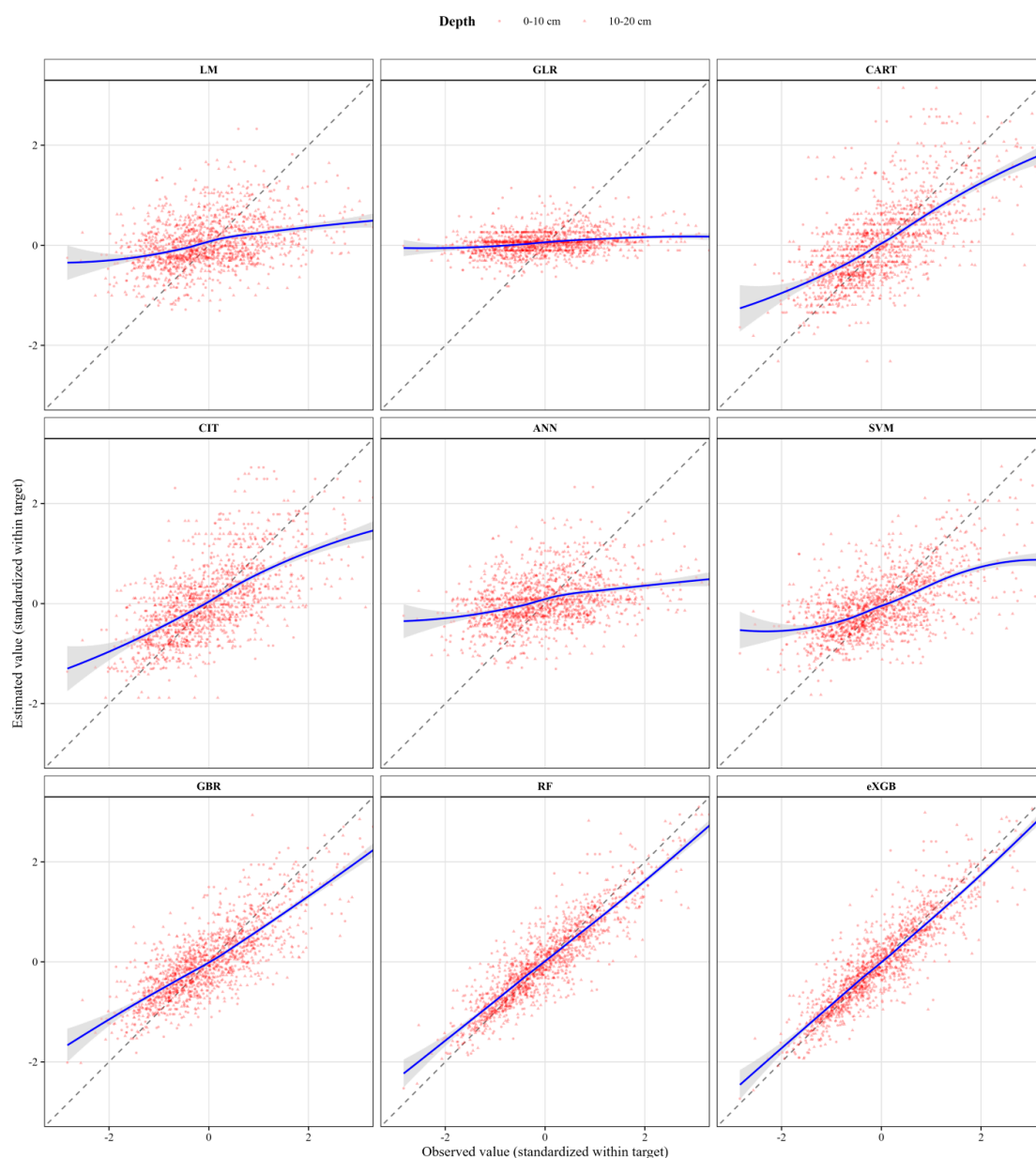


Figure 1. Standardized observed–versus–predicted relationships for all soil fertility indicators across the nine modeling algorithms. Panels represent LM, GLR, CART, CIT, ANN, SVM, GBR, RF, and eXGB models. Red points denote validation samples, grey dashed lines indicate 1:1 reference lines, and blue curves denote LOESS–smoothed fitted trends, and grey shaded bands indicate the 95% confidence intervals of the fitted trends.

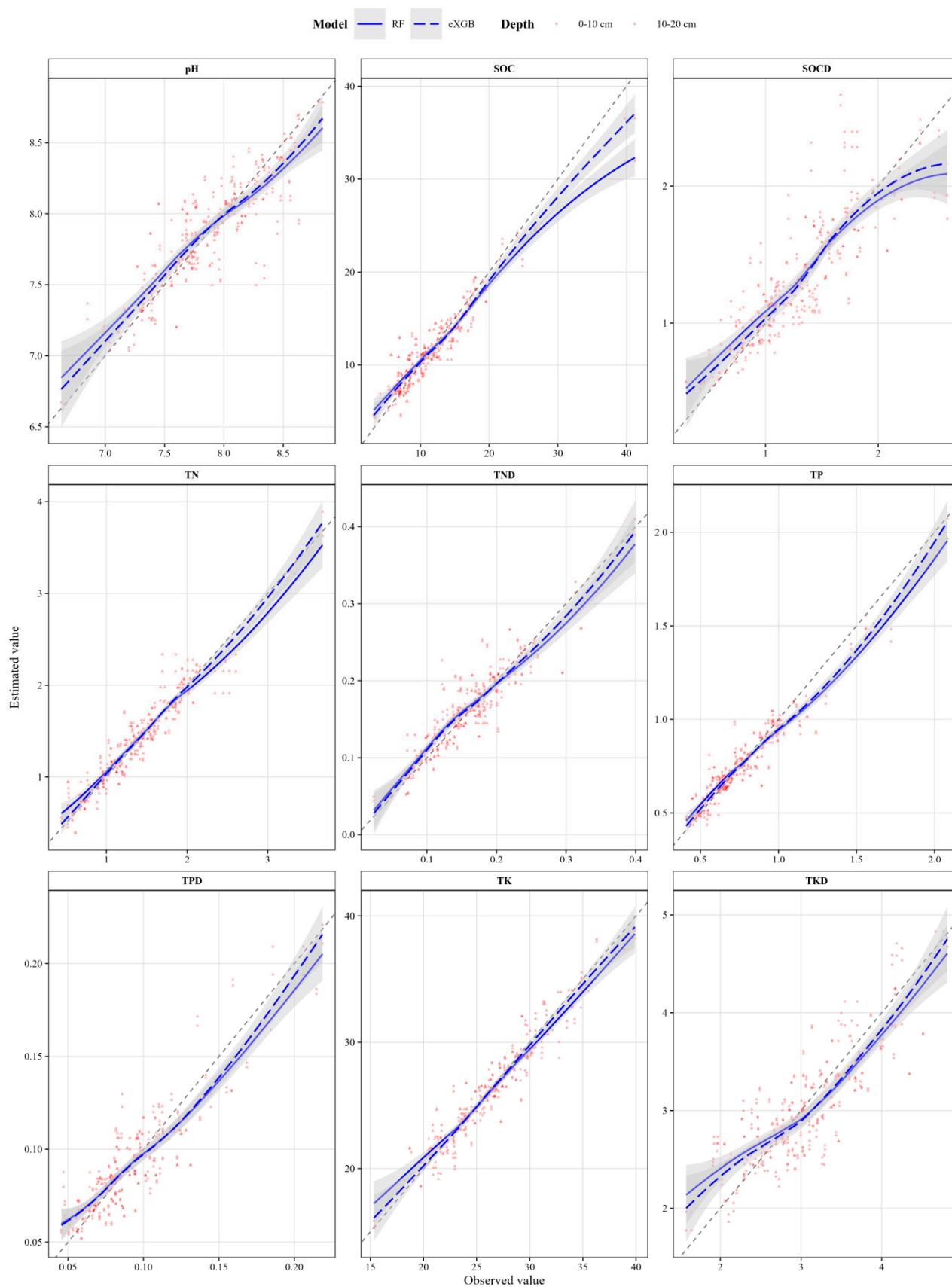


Figure 2. Observed-versus-predicted relationships for RF and eXGB models across soil fertility indicators. Panels represent pH, SOC, SOCD, TN, TND, TP, TPD, TK, and TKD. Red points denote validation samples, grey dashed lines indicate 1:1 reference lines, and blue curves denote LOESS-smoothed fitted trends, and grey shaded bands indicate the 95% confidence intervals of the fitted trends.

The 10–20 cm layer showed a similar pattern (Figures 1, 2 and S3; Table 1). RF and eXGB again matched the observations most closely and maintained low RMSE values. By contrast, linear models and ANN tended to over- or underestimate SOCD and produced larger errors for TKD. For deep-layer SOCD, the relative bias of eXGB was close to zero (-0.15%), with an RMSE of 0.25 kg C/m^2 , whereas MLR, GLR, and ANN all had bias values above 10% and RMSE values close to $0.44\text{--}0.45 \text{ kg C/m}^2$.

3.4. Prediction of Soil Nutrient Concentrations and pH

For SOC, TN, TP, and TK concentrations, RF and eXGB again showed the best overall performance at both depths (Figures 3 and 4). eXGB yielded the lowest RMSE values for surface SOC, TN, TP, and TK, while RF produced similarly low errors. The advantage of ensemble methods was especially evident for SOC and TN, two variables that are strongly controlled by vegetation productivity and microbial decomposition. Single-tree and linear models showed greater scatter and, in some cases, systematic flattening of the predicted range, suggesting that they could not adequately represent nonlinear gradients.

All models predicted pH more accurately than the nutrient variables, but RF and eXGB again ranked highest (Table 2; Figure S6). At 0–10 cm, both models had an RMSE of 0.21 pH units, whereas at 10–20 cm, eXGB performed best (RMSE = 0.19), closely followed by RF (RMSE = 0.20). Although several models generated high apparent agreement for pH, some flexible algorithms compressed the prediction range, indicating that RMSE and bias should be interpreted together rather than relying on correlation alone.

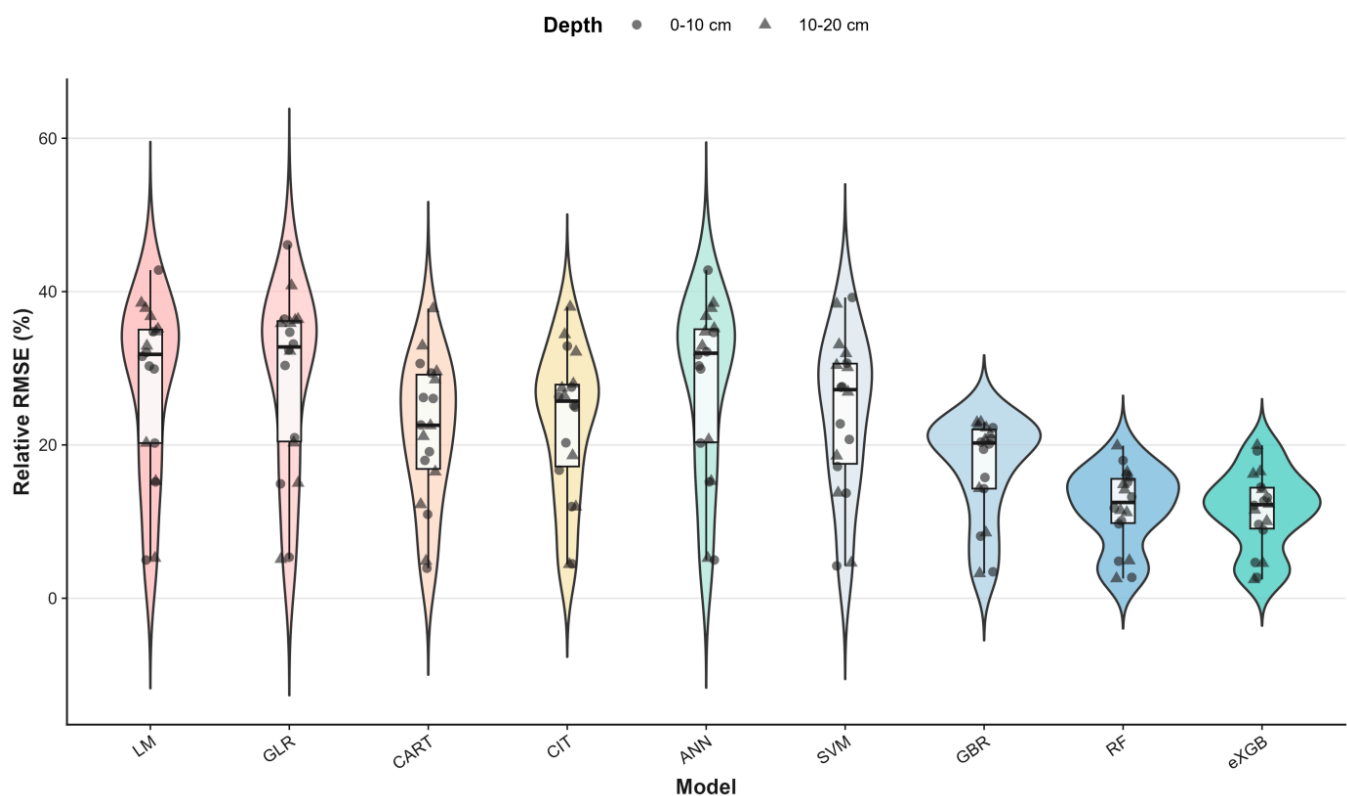


Figure 3. Distribution of relative RMSE across the nine modelling algorithms. Violin plots show the distribution of relative RMSE values across soil fertility indicators and soil depths. Boxplots indicate the median and interquartile range, and points represent individual target-variable/depth combinations. Circles and triangles denote the 0–10 cm and 10–20 cm soil layers, respectively.

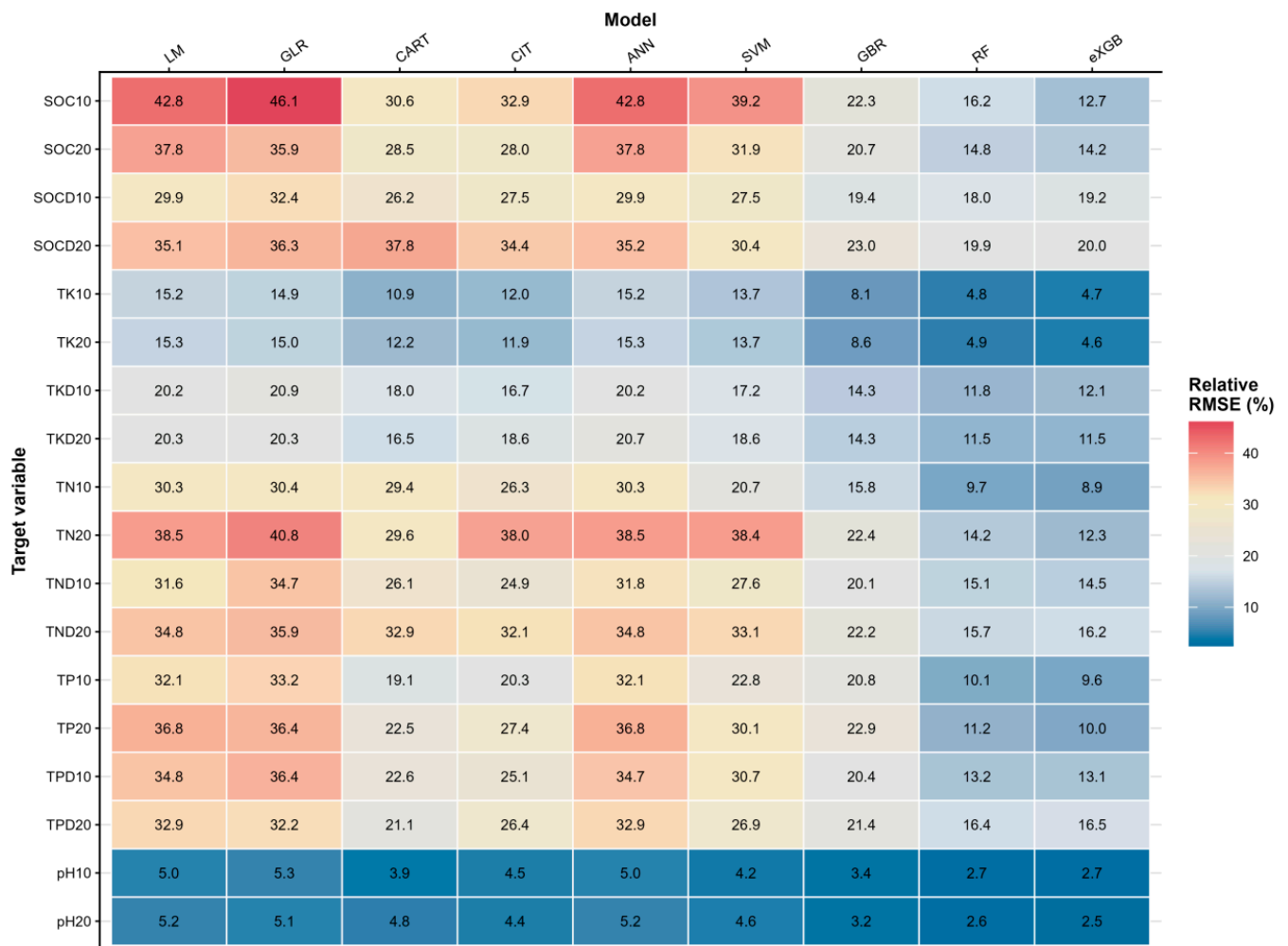


Figure 4. Target-specific relative RMSE across soil fertility indicators and modelling algorithms. The heatmap shows relative RMSE values for each soil fertility indicator and model combination. Cooler colors indicate lower prediction errors, whereas warmer colors indicate higher prediction errors. The values inside cells represent relative RMSE percentages.

Table 2. Root mean square error (RMSE) between predicted and observed values for soil organic carbon density (SOCD, kg/m²), total nitrogen density (TND, kg/m²), total phosphorus density (TPD, kg/m²), total potassium density (TKD, kg/m²), soil organic carbon (SOC, g/kg), total nitrogen (TN, g/kg), total phosphorus (TP, g/kg), total potassium (TK, g/kg), and pH in the 0–10 cm and 10–20 cm layers (n = 82).

Soil Depth	Variable	RF	GBR	SVM	MLR	RRT	GLR	ANN	CIT	eXGB
0–10 cm	SOCD	0.24	0.26	0.36	0.40	0.35	0.43	0.40	0.36	0.25
	TND	0.02	0.03	0.04	0.05	0.04	0.05	0.05	0.04	0.02
	TPD	0.01	0.02	0.03	0.03	0.02	0.03	0.03	0.02	0.01
	TKD	0.35	0.43	0.51	0.60	0.54	0.63	0.60	0.50	0.36
	SOC	2.04	2.81	4.95	5.40	2.80	5.81	5.40	4.15	1.60
	TN	0.13	0.21	0.28	0.41	0.39	0.41	0.41	0.35	0.12
	TP	0.08	0.16	0.17	0.24	0.14	0.25	0.24	0.15	0.07
	TK	1.30	2.17	3.68	4.07	2.94	4.01	4.07	3.21	1.25
	pH	0.21	0.27	0.33	0.39	0.31	0.42	0.39	0.35	0.21

Table 2. Cont.

Soil Depth	Variable	RF	GBR	SVM	MLR	RRT	GLR	ANN	CIT	eXGB
10–20 cm	SOCD	0.25	0.29	0.38	0.44	0.47	0.45	0.44	0.43	0.25
	TND	0.03	0.04	0.05	0.06	0.05	0.06	0.06	0.05	0.03
	TPD	0.01	0.02	0.02	0.03	0.02	0.03	0.03	0.02	0.02
	TKD	0.36	0.44	0.58	0.63	0.51	0.63	0.64	0.58	0.36
	SOC	1.67	2.32	3.59	4.26	3.21	4.03	4.26	3.15	1.60
	TN	0.20	0.32	0.55	0.55	0.42	0.58	0.55	0.54	0.18
	TP	0.09	0.18	0.23	0.28	0.17	0.28	0.28	0.21	0.08
	TK	1.28	2.24	3.59	3.99	3.20	3.93	3.99	3.12	1.20
	pH	0.20	0.25	0.36	0.41	0.38	0.40	0.41	0.35	0.19

3.5. The Impact of Spatial Autocorrelation on the Model

Moran's I diagnostic showed that several observed soil properties in the validation dataset exhibited significant positive spatial autocorrelation, confirming the presence of spatial structure in the measured soil attributes. However, residual spatial autocorrelation was generally weak for the two best-performing models. No significant residual Moran's I was detected for eXGB across the 18 target variables, and only one significant case was detected for RF. In contrast, stronger or more frequent residual spatial autocorrelation was observed for several lower-performing models, including LM, GLR, and ANN. These results suggest that the superior performance of RF and eXGB was not primarily driven by residual spatial dependence among nearby validation samples. The full results and code are provided in Figure S7, Tables S8 and S9, and Code S2.

4. Discussion

4.1. Tree-Based Ensembles More Accurately Capture Alpine Soil–Environment Relationships

The present study demonstrates that tree-based ensemble methods, particularly RF and eXGB, provided the most stable and accurate predictions across soil nutrient densities, nutrient concentrations, and pH. This finding is consistent with many digital soil mapping studies reporting that RF and boosting-based algorithms often outperform linear regression and single-tree models when soil properties are influenced by nonlinear relationships among climate, vegetation, terrain, and management factors [26–31]. Similar patterns have been reported in croplands, grasslands, forest soils, and regional soil-mapping products, where ensemble models are better able to capture threshold responses, variable interactions, and heterogeneous residual structures than purely linear approaches [28–33]. The strong performance of eXGB deserves special attention. Gradient boosting can fit complex response surfaces by iteratively correcting residual error, and XGBoost-style algorithms further incorporate shrinkage, tree-depth constraints, and regularization to reduce overfitting [29,34].

However, the relative performance of different algorithms is not universal. Previous studies have also shown that SVM, regression kriging, or simpler regression models can perform comparably to RF when environmental gradients are relatively smooth, strong terrain covariates are available, or sample size and data heterogeneity differ [27–31,35]. Therefore, the comparatively weaker performance of MLR, GLR, SVM, RRT, and CIT in the present study should not be interpreted as an inherent limitation of these algorithms under all soil-mapping conditions. Rather, it indicates that RF and eXGB were better suited to the current Tibetan alpine cropland dataset, where soil nutrients and pH were likely shaped by combined climatic, vegetation, and management-related variability.

4.2. Soil Depth Influences Predictive Accuracy

Prediction accuracy differed between the 0–10 and 10–20 cm layers, reflecting the vertical heterogeneity of soil carbon, nutrients, and pH [36,37]. Surface soils are more directly affected by crop residues, fine roots, fertilization, tillage, and seasonal changes in temperature and moisture. These factors may strengthen the association between NDVImax and surface soil nutrients, because peak vegetation greenness partly reflects crop growth, biomass production, and potential organic-matter return. This interpretation agrees with previous studies on the Tibetan Plateau showing that vegetation productivity is closely related to soil carbon and nutrient patterns [37,38].

The 10–20 cm layer showed a different but still predictable pattern. Deeper soil properties are less directly observable from remote sensing and are influenced by longer-term processes, such as organic matter stabilization, vertical nutrient movement, soil texture, hydrology, and freeze–thaw dynamics [39,40]. Even so, RF and eXGB maintained relatively low validation errors in this layer, suggesting that NDVImax and climate variables captured spatial gradients associated with subsurface soil conditions. This result supports the practical value of ensemble algorithms for estimating both surface and shallow subsurface soil fertility in alpine croplands.

4.3. Ecological Implications for Alpine Farmland on the Tibetan Plateau

The Qinghai–Tibetan Plateau harbors large soil carbon and nitrogen pools and is highly sensitive to climate warming, hydrological change, and intensified land use [38,39,41]. Previous studies have focused mainly on alpine grasslands, meadows, and permafrost regions, whereas croplands, despite their importance for food security and their combined exposure to climate and management, have received far less attention [41,42]. The present study shows that refined remote-sensing and climatic predictors can estimate several soil-fertility attributes in alpine farmlands with useful accuracy.

From a predictive perspective, aboveground vegetation dynamics can provide useful proxy information for estimating belowground soil fertility. This interpretation is consistent with earlier studies that used NDVI as an indicator of vegetation productivity and ecosystem response [43–45]. In this study, NDVImax reflected spatial variation in crop growth conditions during the sampling year and may have partly integrated the combined effects of climate, soil fertility, and management. Therefore, the strong performance of tree-based ensembles supports the use of NDVImax as a practical predictive covariate for monitoring soil nutrients in alpine agricultural regions.

However, the results do not justify overreliance on vegetation greenness alone. Soil nutrient variation is also strongly affected by fertilization, residue return, irrigation, crop rotation, grazing, and harvest removal, and these processes are not always mirrored by NDVI. Recent studies on the Tibetan Plateau show that management can alter or even reverse climate-driven soil carbon trends [46–48]. Thus, for croplands, management is not a secondary disturbance but a major driver of soil fertility. The shared signal between alpine grasslands and croplands is therefore only partial, and unexplained management effects likely remain embedded in the model residuals.

Compared with natural grasslands, cropland soils are governed not only by climate–vegetation–soil feedbacks but also by tillage, fertilization, irrigation, and residue removal. Accordingly, the consistency between our findings and grassland-based mechanisms should be understood as partial agreement. The successful application of NDVImax and climate variables indicates that alpine grasslands and croplands share certain ecological regulatory patterns, but model residuals probably still contain uncharacterized management signals. This reflects both the novelty and the limitations of applying digital soil mapping to plateau farmlands.

4.4. Limitations and Future Research

This study still has several limitations. First, model validation relied on an independent hold-out set of only 82 observations, which provides a preliminary external validation but cannot fully represent the geographic, climatic, and management heterogeneity of Tibetan croplands. It has been noted that random cross-validation may overestimate model performance when spatial proximity is present [49]. Therefore, future studies should further employ spatial block cross-validation, leave-region-out validation, or additional external validation with newly sampled sites to test the transferability of RF and eXGB models in unsampled areas.

Second, the predictor set was intentionally parsimonious and did not include several potentially important factors, such as detailed topographic variables, soil texture, parent material, irrigation, crop type, fertilization management, and freeze–thaw conditions, even though these factors are known to influence soil carbon, nitrogen, and fertility patterns on the Tibetan Plateau [36,38,41]. In addition, the use of single-year NDVImax and climate variables may not fully capture the long-term processes underlying soil formation and nutrient accumulation. In this study, the use of a consistent predictor set helped ensure a fair comparison among the nine modelling algorithms. To examine whether the omission of topographic information affected the robustness of the model comparison, we conducted a supplementary sensitivity analysis by adding elevation to the original predictor set. The inclusion of elevation caused only minor changes in validation R^2 and did not change the main conclusion that RF and eXGB were the two best-performing algorithms. Future studies should incorporate multi-year vegetation indices, long-term climate averages, and richer terrain, soil, and management covariates to improve both predictive performance and mechanistic interpretability (Tables S10 and S11; Figures S8 and S10).

Third, the present study focused on two shallow soil layers (0–10 and 10–20 cm) and did not extend to deeper profiles. Given that deep soil carbon and nutrients are crucial for long-term fertility and carbon sequestration, future sampling and modeling should cover deeper soil layers and explore vertical prediction strategies, such as depth functions coupled with machine learning.

5. Conclusions

Using 412 soil observations from alpine croplands on the Tibetan Plateau, we compared nine algorithms for predicting soil carbon density, nutrient density, nutrient concentration, and pH at 0–10 and 10–20 cm depths under a consistent predictor framework. Tree-based ensemble methods were the most reliable model class. eXGB and RF consistently produced the lowest mean relative bias and RMSE across properties and depths, whereas linear models, single-tree models, and ANN generally showed larger errors or narrower prediction ranges. Supplementary analyses of residual spatial autocorrelation and elevation sensitivity further supported the robustness of the main model-comparison conclusion. Overall, eXGB and RF provide a useful high-accuracy modelling framework that can support precision fertilization and soil-quality assessment in Tibetan alpine farmlands.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18199875/s1>, Figure S1: Sampling sites; Figure S2: Observed-versus-predicted relationships for soil organic carbon density (SOCD0–10), total nitrogen density (TND0–10), total phosphorus density (TPD0–10), and total potassium density (TKD0–10) in the 0–10 cm layer for (a–d) RF, (e–h) GBR, (i–l) SVM, (m–p) MLR, (q–t) RRT, (u–x) GLR, (y–bb) ANN, (cc–ff) CIT, and (gg–jj) eXGB. Blue lines denote fitted linear regressions; Figure S3: Observed-versus-predicted relationships for soil organic carbon density (SOCD10–20), total nitrogen density (TND10–20), total phosphorus density (TPD10–20), and total potassium density (TKD10–20) in the 10–20 cm layer for (a–d) RF, (e–h) GBR, (i–l) SVM, (m–p) MLR, (q–t) RRT, (u–x) GLR, (y–bb) ANN, (cc–ff) CIT, and (gg–jj) eXGB.

eXGB. Blue lines denote fitted linear regressions; Figure S4: Observed-versus-predicted relationships for soil organic carbon (SOC0–10), total nitrogen (TN0–10), total phosphorus (TP0–10), and total potassium (TK0–10) in the 0–10 cm layer for (a–d) RF, (e–h) GBR, (i–l) SVM, (m–p) MLR, (q–t) RRT, (u–x) GLR, (y–bb) ANN, (cc–ff) CIT, and (gg–jj) eXGB. Blue lines denote fitted linear regressions; Figure S5: Observed-versus-predicted relationships for soil organic carbon (SOC10–20), total nitrogen (TN10–20), total phosphorus (TP10–20), and total potassium (TK10–20) in the 10–20 cm layer for (a–d) RF, (e–h) GBR, (i–l) SVM, (m–p) MLR, (q–t) RRT, (u–x) GLR, (y–bb) ANN, (cc–ff) CIT, and (gg–jj) eXGB. Blue lines denote fitted linear regressions; Figure S6: Observed-versus-predicted relationships for soil pH in the 0–10 cm (pH0–10) and 10–20 cm (pH10–20) layers for (a–b) RF, (c–d) GBR, (e–f) SVM, (g–h) MLR, (i–j) RRT, (k–l) GLR, (m–n) ANN, (o–p) CIT, and (q–r) eXGB. Blue lines denote fitted linear regressions; Figure S7: Spatial autocorrelation diagnostics for observed soil properties and model validation residuals. (a) Moran’s I values of observed validation soil properties. (b) Distributions of residual Moran’s I values across the nine models. Triangles indicate significant Moran’s I values based on permutation tests ($p < 0.05$); Figure S8: Mean validation R^2 before and after adding elevation as an additional covariate; Figure S9: Changes in mean validation R^2 after adding elevation as an additional covariate; Figure S10: Soil-property-specific changes in validation R^2 after adding elevation as an additional covariate; Table S1: Random forest model parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S2: Generalized boosted regression parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S3: Support vector machine parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S4: Multiple linear regression parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S5: Recursive regression tree parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S6: Artificial neural network (ANN), generalized linear regression (GLR), conditional inference tree (CIT), and eXtreme gradient boosting (eXGB) parameters of soil organic carbon density (SOCD, kg m^{-2}), total nitrogen density (TND, kg m^{-2}), total phosphorus density (TPD, kg m^{-2}), total potassium density (TKD, kg m^{-2}), soil organic carbon content (SOC, g kg^{-1}), total nitrogen content (TN, g kg^{-1}), total phosphorus content (TP, g kg^{-1}), total potassium content (TK, g kg^{-1}), and pH at 0–10 cm and 10–20 cm ($n = 330$); Table S7: Source data used for model construction and supplementary reproducibility analysis; Table S8: Moran’s I diagnostics for observed validation soil properties and model residuals; Table S9: Model-level summary of residual Moran’s I values across the nine algorithms; Table S10: Target-level validation R^2 before and after adding elevation as an additional covariate; Table S11: Model-level validation R^2 before and after adding elevation as an additional covariate; Code S1: R code for documenting the original model parameter-setting procedures; Code S2: R code for Moran’s I spatial autocorrelation diagnostics of observed validation values and model residuals; Code S3a: R code for running the elevation-added sensitivity analysis; Code S3b: R code for recalculating validation R^2 using the same metric as the original manuscript and generating the elevation sensitivity figures.

Author Contributions: Conceptualization, C.Z.; Methodology, C.Z., S.L. and Y.T.; Software, Y.T.; Validation, S.L.; Formal analysis, Y.T.; Investigation, C.Z.; Data curation, S.L.; Writing—original draft, W.S.; Writing—review & editing, G.F., Z.Z. and W.S.; Visualization, C.Y. and G.F.; Supervision, C.Y. and G.F.; Project administration, Z.Z.; Funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Xizang Autonomous Region Science and Technology Project [XZ202501ZY0056], the Lhasa Science and Technology Plan Project [LSKJ202422], and the Xizang Autonomous Region Science and Technology Project [XZ202501ZY0086, and XZ202401JD0029].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Tully, K.; Ryals, R. Nutrient Cycling in Agroecosystems: Balancing Food and Environmental Objectives. *Agroecol. Sustain. Food Syst.* **2017**, *41*, 761–798. [\[CrossRef\]](#)
2. Leff, J.W. *Biogeography of Plant and Soil Microbial Communities*; ProQuest Dissertations Publishing: Ann Arbor, MI, USA, 2016.
3. Zhong, Z.; Fu, G. Response of Soil Fungal Species, Phylogenetic and Functional Diversity to Diurnal Asymmetric Warming in an Alpine Agricultural Ecosystem. *Agric. Ecosyst. Environ.* **2022**, *335*, 107993. [\[CrossRef\]](#)
4. Niu, B.; Zhang, X.; Piao, S.; Janssens, I.A.; Fu, G.; He, Y.; Zhang, Y.; Shi, P.; Dai, E.; Yu, C. Warming Homogenizes Apparent Temperature Sensitivity of Ecosystem Respiration. *Sci. Adv.* **2021**, *7*, eabc7358. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Wu, H.; Liu, E.; Jin, T.; Liu, B.; Gopalakrishnan, S.; Zhou, J.; Shao, G.; Mei, X.; Delaplace, P.; De Clerck, C. Crop Rotation Increases Tibetan Barley Yield and Soil Quality on the Tibetan Plateau. *Nat. Food* **2025**, *6*, 151–160. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Li, F.-F.; Lu, H.-L.; Wang, G.-Q.; Yao, Z.-Y.; Li, Q.; Qiu, J. Zoning of Precipitation Regimes on the Qinghai-Tibet Plateau and Its Surrounding Areas Responded by the Vegetation Distribution. *Sci. Total Environ.* **2022**, *838*, 155844. [\[CrossRef\]](#) [\[PubMed\]](#)
7. McBratney, A.B.; Odeh, I.O.; Bishop, T.F.; Dunbar, M.S.; Shatar, T.M. An Overview of Pedometric Techniques for Use in Soil Survey. *Geoderma* **2000**, *97*, 293–327. [\[CrossRef\]](#)
8. Kasraei, B.; Schmidt, M.G.; Saurette, D.D.; Bulmer, C.E.; Zhang, J.; Pennell, T.; John, K.; Heung, B. Advancing Digital Soil Mapping with Multi-Year Crop Cover Data: Impacts on Model Accuracy and Soil Interpretation. *Geoderma* **2025**, *461*, 117481. [\[CrossRef\]](#)
9. Xu, C.; Huang, J.; Hartemink, A.E.; Chaney, N.W. Pruned Hierarchical Random Forest Framework for Digital Soil Mapping: Evaluation Using NEON Soil Properties. *Geoderma* **2025**, *459*, 117392. [\[CrossRef\]](#)
10. O’Leary, D. Mapping Soil Using Neural Network Machine Learning and Remotely Sensed Geoscience Data; a Study in Peat. PhD Thesis, NUI Galway, Galway, Ireland, 2023.
11. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional Variable Importance for Random Forests. *BMC Bioinform.* **2008**, *9*, 307. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Chen, Y.; Zhang, X.; Ye, Q.; Zhang, X.; Cao, N.; Li, S.-Y.; Yu, J.; Zhao, S.-T.; Zhang, J.; Xu, X.-M. Machine Learning-Based Prediction Model for Myocardial Ischemia under High Altitude Exposure: A Cohort Study. *Sci. Rep.* **2024**, *14*, 686. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Wang, M.; Zhang, S.; Guo, X.; Wang, G.; Xia, J.; Xiao, L.; Luo, Z. Whole-profile Soil Carbon Responses to Concurrent Warming and Precipitation Changes across Global Biomes. *Glob. Change Biol.* **2025**, *31*, e70105. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Shalev, M. *Limits and Alternatives to Multiple Regression in Comparative Research*; Capitalisms Compared; Emerald Group Publishing Limited: Bingley, UK, 2007; pp. 261–308.
15. Kianifard, F.; Swallow, W.H. A Review of the Development and Application of Recursive Residuals in Linear Models. *J. Am. Stat. Assoc.* **1996**, *91*, 391–400. [\[CrossRef\]](#)
16. Ding, J.; Yang, T.; Zhao, Y.; Liu, D.; Wang, X.; Yao, Y.; Peng, S.; Wang, T.; Piao, S. Increasingly Important Role of Atmospheric Aridity on Tibetan Alpine Grasslands. *Geophys. Res. Lett.* **2018**, *45*, 2852–2859. [\[CrossRef\]](#)
17. Zhong, Z.M.; Qin, Y.; Zhang, G.Y.; Fu, G. Effects of Warming and No-Tillage on Soil Carbon, Nitrogen, Phosphorus and Potassium Contents and pH of an Alpine Farmland in Tibet. *Agronomy* **2024**, *14*, 1327. [\[CrossRef\]](#)
18. Khalil, M.I.; Gibson, D.J.; Baer, S.G.; Willand, J.E. Functional Diversity Is More Sensitive to Biotic Filters than Phylogenetic Diversity during Community Assembly. *Ecosphere* **2018**, *9*, e02164. [\[CrossRef\]](#)
19. Yang, K.; Jiang, Y.; Tang, W.; He, J.; Shao, C.; Zhou, X.; Lu, H.; Chen, Y.; Li, X.; Shi, J. *A High-Resolution near-Surface Meteorological Forcing Dataset for the Third Pole Region (TPMFD, 1979–2023)*; National Tibetan Plateau Data Center: Beijing, China, 2023.

20. R Core Team. R: A Language and Environment for Statistical Computing. 2022. Available online: <https://biostatistique.vetagro-sup.fr/guideRM1.pdf> (accessed on 2 September 2026).
21. Therneau, T.; Atkinson, B.; Ripley, B.; Ripley, M.B. Package “Rpart”. 2015. Available online: <https://cran.ms.unimelb.edu.au/web/packages/rpart/rpart.pdf> (accessed on 2 September 2026).
22. Garge, N.R.; Bobashev, G.; Eggleston, B. Random Forest Methodology for Model-Based Recursive Partitioning: The mobForest Package for R. *BMC Bioinform.* **2013**, *14*, 125. [[CrossRef](#)] [[PubMed](#)]
23. Greenwell, B.; Boehmke, B.; Cunningham, J.; Developers, G.; Greenwell, M.B. Package “Gbm”. 2019. Available online: <https://espejito.fder.edu.uy/cran/web/packages/ebm/ebm.pdf> (accessed on 5 May 2026).
24. Cortez, P.; Cortez, M.P. Package “Rminer”. 2016. Available online: <https://freebsd.yz.yamagata-u.ac.jp/pub/cran/web/packages/rminer/rminer.pdf> (accessed on 5 May 2026).
25. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Chang, C.-C.; Lin, C.-C.; Meyer, M.D. Package “E1071”. 2019. Available online: <http://www.maths.bristol.ac.uk/R/web/packages/e1071/e1071.pdf> (accessed on 7 May 2026).
26. Hengl, T.; Consoli, D.; Tian, X.; Nauman, T.W.; Nussbaum, M.; Isik, M.S.; Parente, L.; Ho, Y.F.; Simoes, R.; Gupta, S.; et al. OpenLandMap-Soildb: Global Soil Information at 30 M Spatial Resolution for 2000–2022+ Based on Spatiotemporal Machine Learning and Harmonized Legacy Soil Samples and Observations. *Earth Syst. Sci. Data* **2026**, *18*, 989–1036. [[CrossRef](#)]
27. Tian, Y.; Cao, S.; Sun, P.; Kang, Q.; Liu, S.; Zheng, X.; Wei, L.; Lu, Q. Digital Mapping of Soil pH Using Tree-Based Models Coupled with Residual Kriging. *Land* **2026**, *15*, 365. [[CrossRef](#)]
28. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE* **2017**, *12*, e0170478. [[CrossRef](#)] [[PubMed](#)]
29. Makungwe, M.; Chabala, L.; Chishala, B.; Lark, R. Performance of Linear Mixed Models and Random Forests for Spatial Prediction of Soil pH. *Geoderma* **2021**, *397*, 115079. [[CrossRef](#)]
30. Lin, L.; Wang, Q.; Zhang, Z.; He, J. Warming Enhances Soil Freezing and Thawing Circles in the Non-Growing Season in a Tibetan Alpine Grassland. *Acta Sci. Nat. Univ. Pekin.* **2017**, *53*, 171–178.
31. Wadoux, A.; Minasny, B.; McBratney, A. Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. [[CrossRef](#)]
32. Barkov, V.; Schmidinger, J.; Gebbers, R.; Atzmueller, M. Modern Neural Networks for Small Tabular Datasets: The New Default for Field-Scale Digital Soil Mapping? *Eur. J. Soil Sci.* **2026**, *77*, e70299. [[CrossRef](#)]
33. Bravo-García, J.; Camarillo-Naranjo, J.M.; Blanco-Velázquez, F.J.; Anaya-Romero, M. Soil Organic Carbon Mapping Through Remote Sensing and In Situ Data with Random Forest by Using Google Earth Engine: A Case Study in Southern Africa. *Land* **2025**, *14*, 1436. [[CrossRef](#)]
34. Nenkam, A.M.; Wadoux, A.M.J.C.; Minasny, B.; Silatsa, F.B.T.; Yemefack, M.; Ugbaje, S.U.; Akpa, S.; Zijl, G.V.; Bouasria, A.; Bouslihlim, Y.; et al. Applications and Challenges of Digital Soil Mapping in Africa. *Geoderma* **2024**, *449*, 117007. [[CrossRef](#)]
35. Carbajal-Llosa, C.; Barja, A.; Pizarro, S. Ensemble Machine Learning for Digital Mapping of Soil pH and Electrical Conductivity in the Andean Agroecosystem of Peru. *Front. Soil Sci.* **2025**, *5*, 1673628. [[CrossRef](#)]
36. Jobbágy, E.; Jackson, R. The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation. *Ecol. Appl.* **2000**, *10*, 423–436. [[CrossRef](#)]
37. Pries, C.; Ryals, R.; Zhu, B.; Min, K.; Cooper, A.; Goldsmith, S.; Pett-Ridge, J.; Torn, M.; Berhe, A. The Deep Soil Organic Carbon Response to Global Change. *Annu. Rev. Ecol. Evol. Syst.* **2023**, *54*, 375–401. [[CrossRef](#)]
38. Liu, X.; Zhou, T.; Shi, P.; Zhang, Y.; Luo, H.; Yu, P.; Xu, Y.; Zhou, P.; Zhang, J. Uncertainties of Soil Organic Carbon Stock Estimation Caused by Paleoclimate and Human Footprint on the Qinghai Plateau. *Carbon Balance Manag.* **2022**, *17*, 8. [[CrossRef](#)] [[PubMed](#)]
39. Chen, H.; Ju, P.; Zhu, Q.; Xu, X.; Wu, N.; Gao, Y.; Feng, X.; Tian, J.; Niu, S.; Zhang, Y.; et al. Carbon and Nitrogen Cycling on the Qinghai-Tibetan Plateau. *Nat. Rev. Earth Environ.* **2022**, *3*, 701–716. [[CrossRef](#)]
40. Zhao, L.; Wu, X.; Wang, Z.; Sheng, Y.; Fang, H.; Zhao, Y.; Hu, G.; Li, W.; Pang, Q.; Shi, J.; et al. Soil Organic Carbon and Total Nitrogen Pools in Permafrost Zones of the Qinghai-Tibetan Plateau. *Sci. Rep.* **2018**, *8*, 3656. [[CrossRef](#)] [[PubMed](#)]
41. Tian, J.; Zhang, B.; Wang, X.; He, C. In Situ Observations of Soil Hydraulic Properties and Soil Moisture in a High, Cold Mountainous Area of the Northeastern Qinghai-Tibet Plateau. *Sci. China-Earth Sci.* **2023**, *66*, 1757–1775. [[CrossRef](#)]
42. Gao, H.; Gong, J.; Liu, J.; Ye, T. Effects of Land Use/Cover Changes on Soil Organic Carbon Stocks in Qinghai-Tibet Plateau: A Comparative Analysis of Different Ecological Functional Areas Based on Machine Learning Methods and Soil Carbon Pool Data. *J. Clean. Prod.* **2024**, *434*, 139854. [[CrossRef](#)]
43. Zhang, S.; Tong, S.; Ren, J.; Bao, G.; Huang, X.; Bao, Y.; Altantuya, D. Bilateral Coupling Relationships between Vegetation NDVI and Multi-Depth Soil Moisture in the Mongolian Plateau. *Int. J. Digit. Earth* **2025**, *18*, 2532774. [[CrossRef](#)]
44. Ding, Z.; Liu, K.; Grunwald, S.; Smith, P.; Ciaia, P.; Wang, B.; Wadoux, A.; Ferreira, C.; Karunaratne, S.; Shurpali, N.; et al. Advancing Soil Organic Carbon Prediction: A Comprehensive Review of Technologies, AI, Process-Based and Hybrid Modelling Approaches. *Adv. Sci.* **2025**, *12*, e04152. [[CrossRef](#)] [[PubMed](#)]

45. Zhang, X.; Ramos, B.; Cladera, J. Research on Key Influencing Factors of Ecological Environment Quality in Barcelona Metropolitan Region Based on Remote Sensing. *Remote Sens.* **2024**, *16*, 4735. [[CrossRef](#)]
46. Stanley, P.; Wilson, C.; Patterson, E.; Machmuller, M.; Cotrufo, M. Ruminating on Soil Carbon: Applying Current Understanding to Inform Grazing Management. *Glob. Change Biol.* **2024**, *30*, e17223. [[CrossRef](#)] [[PubMed](#)]
47. Zhang, R.; Wei, Y.; Dong, C.; Wei, B.; Rillig, M.; Badgery, W.; Yang, G.; Liu, N.; Zhang, Y. Quantifying the Positive Effect of Ungulate Herbivory on Living Root-Derived Soil Organic Carbon Formation: Evidence From an Eight-Year Simulated Grazing Field Experiment With ¹³C Pulse Labeling. *Glob. Change Biol.* **2025**, *31*, e70336. [[CrossRef](#)] [[PubMed](#)]
48. Zhou, Y.; Xu, M.; Ren, S.; Du, Y.; Yue, Y.; Yu, H.; Zhang, Y.; Jiang, S.; Xu, T.; Wang, L. More Than a Decade of Moderate Grazing: No Impact on Soil Organic Carbon Stocks and Enhancement of Mineral-Associated Organic Carbon via Livestock Diversification. *Glob. Change Biol.* **2025**, *31*, e70466. [[CrossRef](#)] [[PubMed](#)]
49. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]

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.