Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial mismatch between total and extractable zinc and its geochemical drivers in coastal saline-alkaline soils

taolin zhou
University of Chinese Academy of Sciences
mengge zhou
Hainan Provincial Ecological Environmental Monitoring Center
yi yang
Institute of Geographic Sciences and Natural Resources Research
ziqiao li
University of Chinese Academy of Sciences
yonghua li
yhli@igsnrr.ac.cn

Institute of Geographic Sciences and Natural Resources Research

Research Article

Keywords: DTPA-extractable zinc, Saline-alkaline soils, Soil monitoring, Soil organic carbon, Soil pH, Spatial prediction

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11751518/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spatial mismatch between total and extractable zinc and its geochemical drivers in coastal saline-alkaline soils

Taolin Zhou 1, 3, Mengge Zhou 1, 2, Yi Yang 1, *, Ziqiao Li 1, 3, Yonghua Li 1, *

1 Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

2 Hainan Provincial Ecological Environmental Monitoring Center, Haikou 571126, China

3 University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding authors:

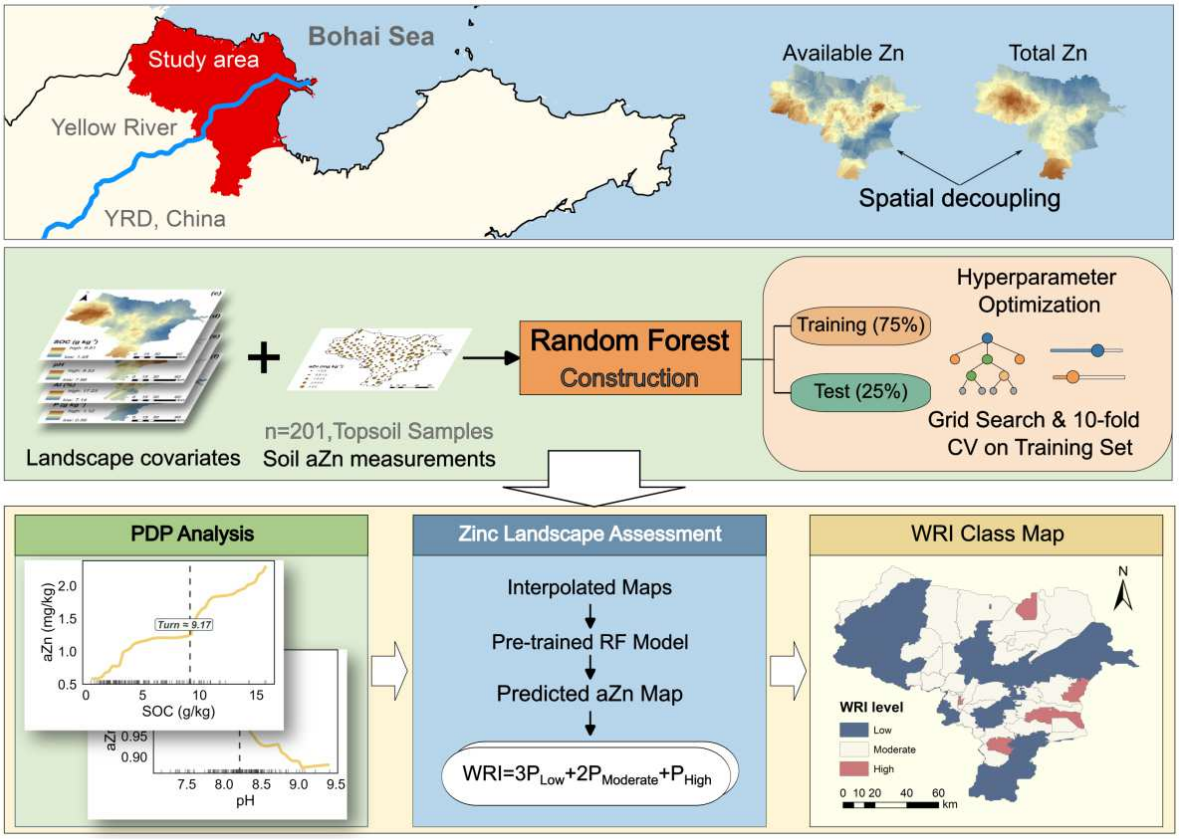
Yonghua Li: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. 11A Datun Road, Chaoyang District, Beijing 100101, China Email: yhli@igsnrr.ac.cn

Yi Yang: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. 11A Datun Road, Chaoyang District, Beijing 100101, China Email: yangyi@igsnrr.ac.cn

Highlights

- Total and extractable soil zinc showed contrasting spatial patterns.
- Extractable zinc was below 0.5 mg kg^{-1} in 22.4% of soil samples.
- Organic carbon and pH were major correlates of extractable zinc.
- Spatial assessment elucidates the geochemical behavior of zinc in salt-affected soils.

Graphical abstract



Abstract

Assessing soil zinc (Zn) status requires distinguishing total concentration from the fraction recovered by an availability test. We analyzed 201 agricultural topsoil samples from the Yellow River Delta, China, to compare total Zn with diethylenetriaminepentaacetic acid (DTPA)-extractable Zn and examine their associations with soil properties. Mean total Zn was 59.90 mg kg^{-1}, close to the regional geochemical background, whereas 45 samples (22.4%) contained less than 0.5 mg kg^{-1} extractable Zn. Their mapped distributions differed, and relatively high total Zn did not consistently coincide with high extractable Zn. Random forest analysis identified soil organic carbon (SOC), SOC-derived predictors, and pH as major contributors to predictions, with a test-set R^2 of 0.46 under a random data split. Structural equation modeling identified positive associations with SOC and negative associations with pH. Within the 45 samples with low extractable Zn, none of the examined environmental correlations was significant. Predictions on a 1-km grid were summarized for 106 townships to describe the spatial composition of extractable-Zn classes. The results support measuring extractable Zn alongside total Zn, SOC, and pH to select locations for confirmatory sampling and crop-response trials. This joint assessment provides a basis for interpreting trace element behavior and planning agricultural health monitoring in salt-affected landscapes.

Keywords: DTPA-extractable zinc; Saline-alkaline soils; Soil monitoring; Soil organic carbon; Soil pH; Spatial prediction

1. Introduction

Zinc (Zn) is an essential micronutrient for plant growth and human nutrition (Broadley et al., 2007). Soil is the main reservoir that supplies Zn to terrestrial ecosystems (Kabata-Pendias, 2010). Total Zn concentration describes element abundance, whereas an availability test measures an operationally defined extractable fraction. A soil with a substantial total Zn concentration can therefore have low extractable Zn (Alloway, 2008; Moreno-Lora and Delgado, 2020). Extractability depends on Zn partitioning among soil phases and on pH, carbonates, organic matter, and other nutrients (Duffner et al., 2014; Khoshgoftarmanesh et al., 2018; Peng et al., 2023). These factors are particularly relevant in coastal saline-alkaline soils, where several chemical gradients overlap. Comparing total and extractable Zn can identify locations that warrant further soil and crop assessment.

Landscape surveys show that total and available Zn can differ in their spatial distributions and relationships with soil properties (Peng et al., 2023; Taşpınar et al., 2025). Reliable soil information is important for managing salt-affected agricultural land, where degradation and restoration impose substantial costs (Qadir et al., 2014). Soil monitoring therefore needs to account for Zn availability as well as total concentration. Spatial soil information has been used to delineate management zones in salt-affected and micronutrient-deficient agricultural regions (Barman et al., 2021; Shukla et al., 2024). Its practical value depends on how the mapped patterns guide decisions, including the selection of monitoring sites and locations for field trials.

Interpreting these spatial patterns requires accounting for the covariation of soil properties and landscape position. Digital soil mapping uses environmental covariates to describe spatial variation, but correlations among predictors complicate attribution (McBratney et al., 2003; Minasny and McBratney, 2016). In coastal saline-alkaline landscapes, pH, salinity, organic carbon, and coastal position form overlapping gradients (Hanif et al., 2024). Dividing samples by Zn availability also changes the response range and sample size within each group. These differences affect the interpretation of subgroup correlations (Gelman and Stern, 2006).

Random forest and structural equation modeling (SEM) provide complementary descriptions of soil-property relationships. Random forest represents nonlinear predictive associations, while Shapley additive explanations (SHAP) and partial dependence plots (PDPs) describe how the fitted model uses its predictors (Breiman, 2001; Friedman, 2001; Lundberg et al., 2020). Correlated and

derived predictors require particular attention because they share information and can generate unsupported combinations during model interpretation (Apley and Zhu, 2020; Wadoux et al., 2020; Pichler and Hartig, 2023). SEM evaluates hypothesized direct and indirect associations within the observed covariance structure (Grace and Irvine, 2020).

We measured total and DTPA-extractable Zn in 201 agricultural topsoil samples from the Yellow River Delta, China, a coastal region affected by salinization and alkalinity (Wang et al., 2023; Xia et al., 2019). We examined three questions: (1) How closely do the mapped distributions of total and extractable Zn correspond? (2) Which soil properties are associated with extractable Zn after accounting for total Zn, and how do the observed associations vary between availability groups? (3) How can spatial predictions and township summaries guide further monitoring? We combined partial correlations, random forest interpretation, SEM, and spatial prediction to relate soil Zn status to the selection of locations for follow-up assessment.

2. Materials and methods

2.1. Study area

The study covered approximately 5,400 km² on the western coast of the Bohai Sea, China, between 117°31′ and 119°19′ E and 37°04′ and 38°16′ N (Fig. 1). The main crops are winter wheat, summer maize, and cotton. The predominant soils are Salic Fluvisols and Solonchaks under the World Reference Base for Soil Resources (WRB). The region has a warm-temperate continental monsoon climate, with dry, windy springs and hot, rainy summers. Seasonal wetting and drying, high evaporative demand, historical seawater intrusion, and alluvial deposition contribute to variation in salinity and soil chemistry (Wang et al., 2023). These conditions allow the relationships between soil properties and extractable Zn to be examined across a coastal saline-alkaline landscape (Hanif et al., 2024). Figure 1 shows the study extent and distribution of the 201 sampling sites.

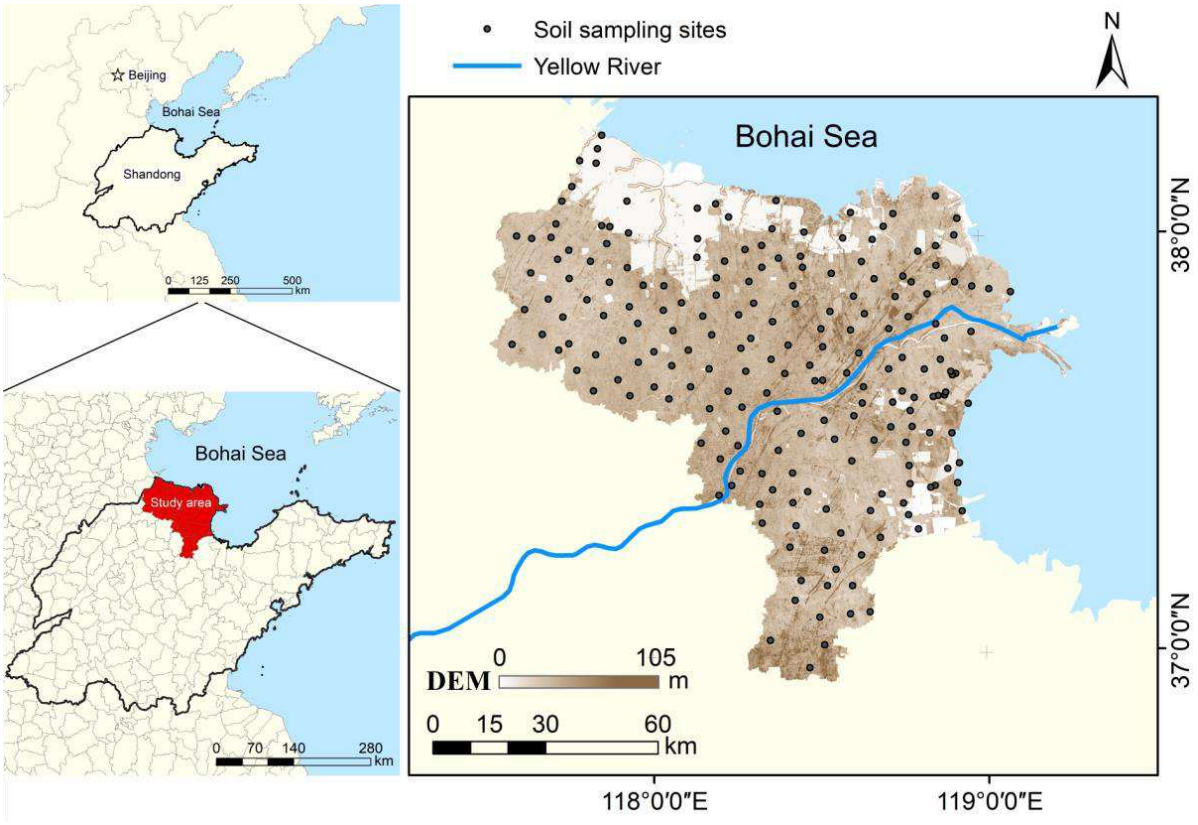


Fig. 1. Study area and distribution of agricultural topsoil sampling sites in the Yellow River Delta, China.

2.2. Field sampling and laboratory analysis

Soils were sampled in July 2021 using a 10 km × 10 km base grid with additional sites to improve spatial coverage. We collected 201 topsoil samples from 0–30 cm depth. At each site, five subsamples were taken with a wooden shovel within a 3 × 3 m quadrat and combined into approximately 1 kg of homogenized soil. Coordinates were recorded with a Trimble GeoXT GPS, together with land use, sampling time, and personnel. Sampling was confined to agricultural fields with relatively uniform management. Land-use records were used to verify sampling sites and were not included as predictors. Samples were air-dried, cleared of stones and roots, gently crushed with a wooden stick, and passed through 10- and 20-mesh nylon sieves. A subsample was further ground to pass through a 100-mesh nylon sieve.

Soil pH was measured in a 1:2.5 soil-to-water suspension (shaken for 2 min and left to stand for 30 min), and electrical conductivity (EC) was determined at a 1:5 soil-to-water ratio (shaken at 20 °C for 30 min, then centrifuged at 4000 rpm for 10 min). Soil organic carbon (SOC) was measured using the Walkley–Black method (Walkley and Black, 1934). For total elemental analysis, soil

samples (approximately 0.10 g) were digested on a hot plate (100 to 180 °C) using a mixed acid solution (HNO_3–HF–$HClO_4$ at a 3:3:1 ratio). Total P, K, Ca, Mg, Al, and Na were determined by inductively coupled plasma optical emission spectrometry (ICP-OES; PerkinElmer, USA), whereas total Zn was determined by inductively coupled plasma mass spectrometry (ICP-MS; PerkinElmer, Waltham, USA). The instrumental detection limits in the analytical solutions were 0.1 mg L^{-1} for K, Ca, Na, and Mg, and 0.1 ng L^{-1} for Zn. To maintain consistency with spatial modeling and visualization, major elements (Ca, Mg, Na, K) are reported as elemental concentrations rather than their oxide forms.

Available Zn (aZn) was extracted using a DTPA–$CaCl_2$–TEA buffer solution (0.005 M DTPA, 0.01 M $CaCl_2$, and 0.1 M TEA, adjusted to pH 7.30 ± 0.05). Briefly, 5.0 g of soil was extracted with 10 mL of extractant, shaken at 200 rpm for 2 h at 25 °C, centrifuged at 4000 rpm for 7 min, and filtered. Zinc in the extract was then measured by ICP-OES. Analytical quality was controlled using certified reference materials (GBW07986 and GBW07959), with triplicates analyzed for one in every ten samples to minimize error. The relative standard deviation of replicates was ≤5%. Hereafter, aZn denotes DTPA-extractable Zn, rather than a direct measurement of plant uptake or of all labile Zn pools. Available Zn (aZn) was measured by DTPA extraction (Lindsay and Norvell, 1978). The extractant contained 0.005 M DTPA, 0.01 M $CaCl_2$, and 0.1 M triethanolamine (TEA), adjusted to pH 7.30 ± 0.05. A 5.0-g soil sample was shaken with 10 mL of extractant at 200 rpm for 2 h at 25 °C, centrifuged at 4000 rpm for 7 min, and filtered. Zn in the extract was measured by ICP-OES. Throughout the manuscript, aZn denotes this DTPA-extractable fraction.

One in every 10 samples was analyzed in triplicate. Accuracy was assessed with certified soil reference materials GBW07986 and GBW07959. The relative standard deviation of replicate measurements was ≤5%.

2.3. Environmental covariates

Environmental covariates described soil properties, terrain, vegetation, and spatial position. Raster datasets were assembled on a common grid with a nominal spacing of 1 km. Clay content, total nitrogen (N), bulk density (BD), and porosity (POR) were obtained from the 1-km China Soil Map-Based Harmonized Data Base at the National Tibetan Plateau Data Center (Shangguan et al., 2013). POR was obtained directly from the dataset. Elevation from a digital elevation model (DEM) and the 2021 normalized difference vegetation index (NDVI; Xu, 2018) were obtained

from the Resource and Environmental Science and Data Center, Chinese Academy of Sciences. Distance to the nearest coastline (CoastD) was calculated from OpenStreetMap vector data (Haklay and Weber, 2008). Geographic coordinates recorded during sampling, including latitude (Lat), were also used as predictors.

For products with multiple depth intervals, including N, BD, and POR, the upper four layers were combined using a depth-weighted mean. Their thicknesses were 4.5, 4.6, 7.5, and 12.3 cm, giving a 0–28.9 cm profile that approximated the sampled 0–30 cm layer. Spatial covariates were aligned under the WGS 1984 geographic coordinate system. Gridded data maintained a nominal spatial resolution of approximately 1 km (30 arc-seconds) provided by the original databases, and were harmonized using bilinear interpolation. NDVI for May 2021 (Xu, 2018) represented vegetation conditions in the sampling year. Covariate values were assigned to sampling locations by nearest-neighbor extraction in R 4.5.1.

2.4. Statistical analysis and random forest modeling

Samples were divided into low-aZn (<0.5 mg kg^{-1}; n = 45) and higher-aZn (≥0.5 mg kg^{-1}; n = 156) groups using an operational concentration cutoff. Environmental variables were compared with Mann–Whitney U tests. Spearman rank correlations described pairwise associations, and partial Spearman correlations controlling for total Zn were calculated with ppcor in R 4.5.1. Missing values were handled by pairwise deletion. Subgroup comparisons were descriptive because sample sizes and response ranges differed. Differences between correlation coefficients were not formally tested (Gelman and Stern, 2006).

Random forest regression was implemented with ranger in R 4.5.1 (Breiman, 2001; Wright and Ziegler, 2017). The right-skewed aZn response was transformed using the natural logarithm. Predictors included soil properties, terrain, vegetation, geographic covariates, and three derived indices: the pH × SOC product (g kg^{-1}), the Ca/Mg mass ratio Ca/ (Mg + 0.001), and the SOC–DEM index SOC/log (DEM + 2). Variance inflation factors (VIFs) were calculated for the base covariates, and variables with VIF >10 were excluded. The three derived indices were retained outside this screening step.

The 201 samples were randomly divided into training (75%; n = 151) and test (25%; n = 50) subsets using a fixed random seed (seed = 123). Missing values were imputed using overall

medians calculated from the entire training subset prior to cross-validation. Hyperparameters were then selected by grid search with 10-fold cross-validation within this imputed training subset to maximize R^2. The search ranges were mtry = 2–4, min.node.size = 15–30, max.depth = 3–5, and sample.fraction = 0.55–0.75. The final model used 800 trees, mtry = 4, a minimum node size of 15, a maximum depth of 5, and a sampling fraction of 0.55. Performance was evaluated on the log scale using R^2, root mean square error (RMSE), and mean absolute error (MAE). Moran's I was calculated for full-sample residuals using inverse-distance weights (Moran, 1950). This diagnostic assessed residual spatial autocorrelation under the specified weighting scheme.

SHAP values were calculated with kernelshap and visualized with shapviz to describe contributions to model predictions (Lundberg et al., 2020). They were computed for all 201 samples using a background dataset of 100 randomly selected observations and a fixed seed of 123. One- and two-dimensional PDPs were generated with pdp (Friedman, 2001).

A custom prediction wrapper recalculated the derived indices whenever their component variables were varied across the PDP evaluation grid. This maintained algebraic consistency between the base and derived predictors. The resulting curves described the fitted model's responses to these changes.

2.5. Structural equation modeling

SEM was fitted with lavaan in R 4.5.1 to examine hypothesized pathways linking aZn to total Zn, SOC, P, Mg, Al, Ca, and pH (Rosseel, 2012). Variables with absolute skewness >1.5, specifically aZn, P, and total Zn, were natural-log transformed. Where the minimum value was ≤0, $\ln(x - \min(x) + 1)$ was used. The model was estimated by maximum likelihood with listwise deletion, using 201 samples. Covariances were specified a priori from established geochemical relationships (Kline, 2023). Nonsignificant paths were subsequently removed, yielding a final model with nine degrees of freedom. Model fit was evaluated using χ^2/df, the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) (Hu and Bentler, 1999). Indirect associations were estimated with 1,000 bootstrap resamples and percentile confidence intervals. Standardized and unstandardized estimates were checked against their corresponding intervals. Reported R^2 values refer to the modeled scale, including the log scale for aZn. The initial and final lavaan specifications are provided in Supplementary Text S1.

2.6. Landscape-scale prediction and spatial aggregation

The final random forest model was refitted to all 201 samples and used to predict aZn on a 1-km grid. DEM and BD were obtained from the gridded datasets, while pH, SOC, P, Al, Ca, and Mg were interpolated from the sampling points by ordinary kriging, with inverse distance weighting (IDW) used as a fallback. All layers were aligned in WGS 84 (EPSG:4326). Predictions covered the mapping extent without a mask for non-agricultural land or internal water bodies. This grid was used to summarize regional spatial patterns. Predictions on the log scale were back-transformed with the exponential function without a retransformation bias correction.

Predicted aZn was divided into three nonoverlapping classes: <0.5 mg kg^{-1}, 0.5 to <1.0 mg kg^{-1}, and ≥1.0 mg kg^{-1}. The boundaries were adapted from agronomic classifications (Liu, 1996; Moreno-Lora and Delgado, 2020) to summarize the model outputs.

For each of 106 townships, we calculated the percentage of mapped area in each aZn class. A weighted relative index (WRI) summarized class composition using weights of 3, 2, and 1 for the <0.5, 0.5 to <1.0, and ≥1.0 mg kg^{-1} classes, respectively:

$$WRI = 3P_{<0.5} + 2P_{0.5–<1.0} + P_{\geq 1.0}$$

Each P is the percentage of the total mapped administrative area within a township assigned to the corresponding class; the three percentages sum to 100. The theoretical WRI range is 100–300. The regional mean and the mean plus 1.5 standard deviations separated the low-, moderate-, and high-WRI categories.

The empirical 3:2:1 weight gave greater weight to lower predicted aZn classes. WRI was used to compare township class composition and identify areas for closer examination of soil Zn status.

3. Results

3.1. Total and extractable Zn concentrations and spatial patterns

Total Zn varied less than aZn across the study area. Total Zn ranged from 33.39 to 175.45 mg kg^{-1}, with a mean of 59.90 mg kg^{-1} and a coefficient of variation (CV) of 0.28. aZn ranged from 0.21 to 17.93 mg kg^{-1}, with a mean of 1.43 mg kg^{-1}, a median of 0.97 mg kg^{-1}, and a CV of 1.23. Of the 201 samples, 106 (52.7%) contained less than 1.0 mg kg^{-1} aZn, including 45 (22.4%) below 0.5 mg kg^{-1} (Liu, 1996; Moreno-Lora and Delgado, 2020).

Mean aZn ranged from 0.38 mg kg^{-1} in the <0.5 mg kg^{-1} class to 8.79 mg kg^{-1} in the >5.0 mg kg^{-1} class (Table 1).

The mapped distributions of total Zn and aZn differed (Fig. 2a, b). High total Zn formed a continuous area in the mid-western and southern inland parts of the study area, whereas high aZn occurred mainly in western and central-eastern patches. These patches generally coincided with lower pH, higher SOC and P, and moderate Al concentrations (Fig. 2c–f). Contiguous areas with aZn <0.5 mg kg^{-1} occurred mainly in strongly alkaline belts, including locations with relatively high total Zn and SOC. High total Zn therefore did not consistently coincide with high aZn.

Table 1. Descriptive statistics of total Zn and DTPA-extractable Zn (aZn) in 201 agricultural topsoil samples

aZn class	n	Min	Max	Mean	Median	SD	CV
Very Low: <0.5	45	0.21	0.49	0.38	0.38	0.09	0.23
Low: 0.5–1.0	61	0.50	0.99	0.76	0.78	0.15	0.19
Medium: 1.0–2.0	60	1.01	1.91	1.39	1.36	0.26	0.18
High: 2.0–5.0	29	2.02	4.86	3.00	2.96	0.76	0.25
Very High: >5.0	6	5.17	17.93	8.79	7.40	4.88	0.56
All aZn	201	0.21	17.93	1.43	0.97	1.75	1.23
Total Zn	201	33.39	175.45	59.90	57.04	16.58	0.28

Note: concentrations are in mg kg^{-1}. Zn, total zinc; aZn, DTPA-extractable zinc; SD, standard deviation; CV, coefficient of variation. Agronomic aZn classes follow Liu (1996). The cited soil Zn background for Shandong Province is 60.9 mg kg^{-1} (Wei, 1990).

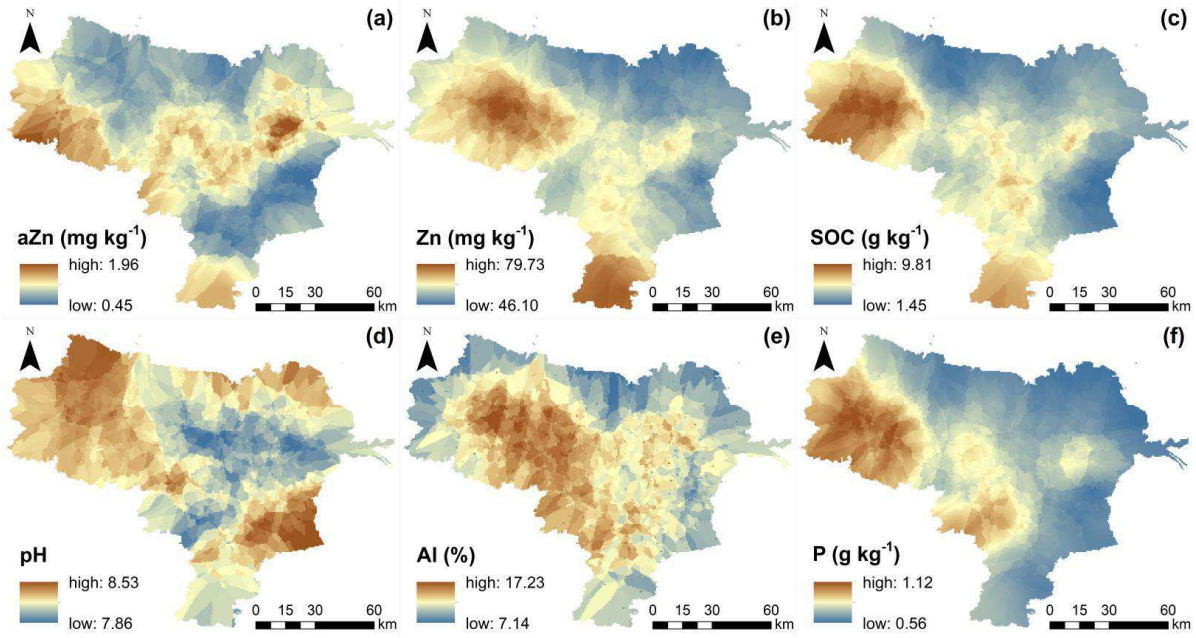


Fig. 2. Mapped distributions of aZn, total Zn, and selected soil properties: (a) aZn, (b) total Zn, (c) SOC, (d) pH, (e) total Al, and (f) total P.

3.2. Environmental associations in the full dataset and aZn groups

The low-aZn group had higher pH, EC, and Na and was closer to the coast, whereas the higher-aZn group had higher SOC and total P (Fig. 3). The distributions overlapped, indicating that the groups occupied partly shared environmental gradients. Mann–Whitney U-test results are reported in Supplementary Table S1.

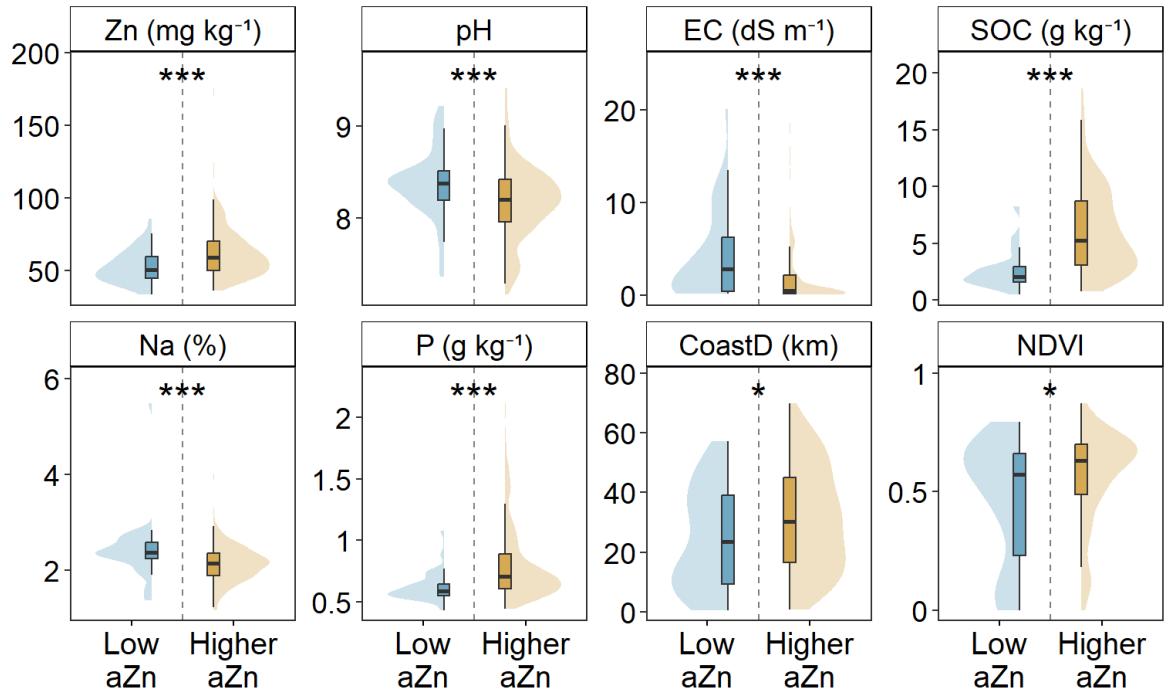


Fig. 3. Environmental characteristics of the low-aZn (<0.5 mg kg^{-1}) and higher-aZn (≥0.5 mg kg^{-1}) groups. Half-violin plots show distributions; box plots show medians and interquartile ranges (IQRs, 25th–75th percentiles), with whiskers extending to 1.5 × IQR. Asterisks indicate Mann–Whitney U-test results: *p <0.05, **p <0.01, and ***p <0.001. EC, electrical conductivity; SOC, soil organic carbon; Na, total sodium; P, total phosphorus; CoastD, distance to the coastline; NDVI, normalized difference vegetation index.

The overall Spearman correlation network showed several associations among the examined environmental variables and spatial proxies (Fig. 4a). Controlling for total Zn changed the magnitude or significance of some associations (Fig. 4b). CoastD and NDVI were no longer significant, while pH, SOC, and P remained significant and Al, Ca, and Mg were additional correlates.

No significant environmental correlations were detected within the low-aZn group. Significant associations with SOC, Al, and Ca were retained in the higher-aZn group (Fig. 4b). Supplementary Table S2 reports the correlation results.

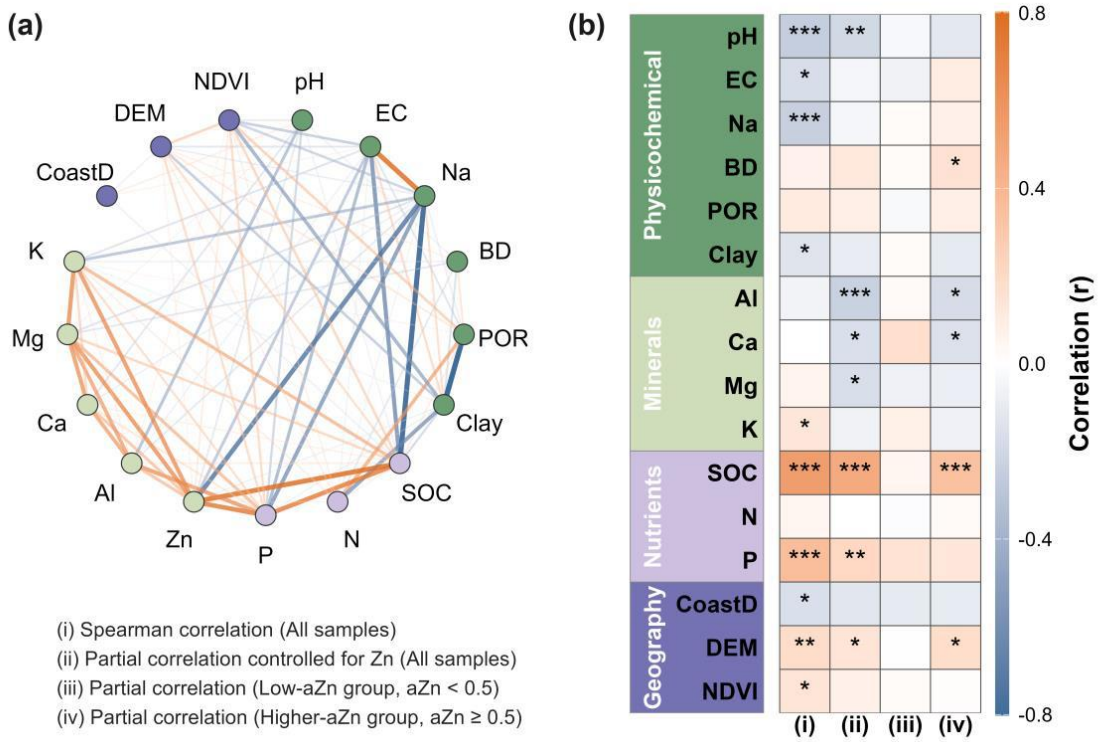


Fig. 4. Associations with aZn in the full dataset and the two aZn groups. (a) Spearman correlation network; edge width represents correlation magnitude, and orange and blue indicate positive and negative correlations. (b) Spearman and partial Spearman correlations controlling for total Zn for all samples (n = 201), the low-aZn group (n = 45), and the higher-aZn group (n = 156). Color intensity represents correlation magnitude; *p <0.05, **p <0.01, and ***p <0.001. Between-group differences in correlation coefficients were not formally tested.

3.3. Random forest performance and model interpretation

Training- and test-set R^2 values for ln(aZn) were 0.59 and 0.46, respectively (Fig. S1a, c). Training-set RMSE and MAE were 0.50 and 0.40; test-set values were 0.60 and 0.48, respectively (Fig. S1c). Full-sample residuals showed no detectable spatial autocorrelation under inverse-distance weights (Moran's I = −0.0038, p = 0.8762; Fig. S1b).

SOC accounted for 24.8% of the summed mean absolute SHAP values, followed by the pH × SOC product (20.0%), pH (12.9%), and the SOC–DEM index (9.9%; Fig. 5a). These percentages describe the distribution of model attribution among the inputs, including derived predictors that share components.

One-dimensional PDPs showed nonlinear responses to pH, SOC, and total P (Fig. 5b). Curve features near pH 8.2, SOC 9 g kg^{-1}, and total P 0.7 g kg^{-1} were identified from derivatives of the

smoothed curves and annotated as a drop, turn, and minimum, respectively. In the two-dimensional PDPs, predicted aZn remained low below approximately 9 g kg^{-1} SOC; above this value, it increased with SOC and decreased with pH (Fig. 5c). The Ca–Mg and SOC–DEM surfaces mainly followed the Ca and SOC gradients, respectively.

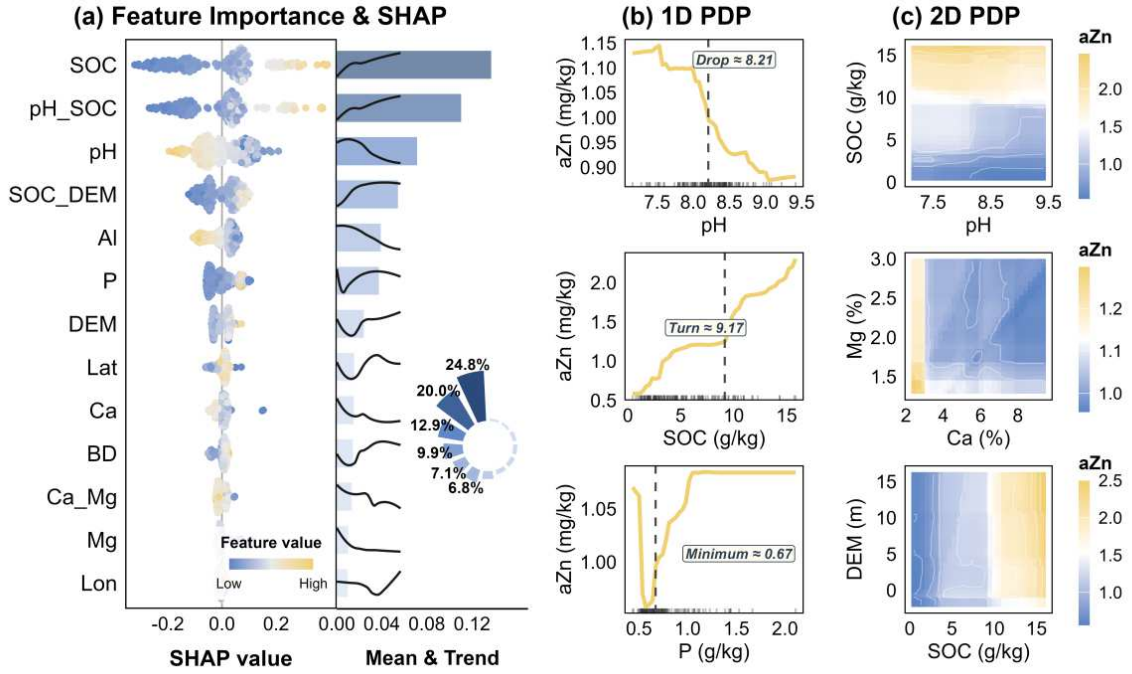


Fig. 5. Interpretation of the fitted random forest model. (a) SHAP distributions and relative contributions of the predictors. (b) One-dimensional PDPs for pH, SOC, and total P; annotations identify approximate curve features. (c) Two-dimensional PDPs for pH and SOC, Ca and Mg, and SOC and DEM. Derived predictors were recalculated as described in Section 2.4.

3.4. Structural relationships associated with aZn

The final SEM had $\chi^2/df = 2.078$, CFI = 0.985, RMSEA = 0.073, and SRMR = 0.081 (Fig. 6). SOC had the largest positive standardized direct association with aZn ($\beta = 0.561$, $p < 0.001$) and a positive indirect association through P and pH ($\beta = 0.038$, $p = 0.013$). pH and Al had negative direct associations with aZn ($\beta = -0.282$ and -0.279, respectively; both $p < 0.001$). Al had a positive indirect association through pH ($\beta = 0.116$, $p < 0.001$). The indirect associations of Mg and P through pH were −0.065 and 0.066, respectively ($p < 0.01$). The model explained 42.6% of the variance in aZn, 33.6% in P, and 21.1% in pH.

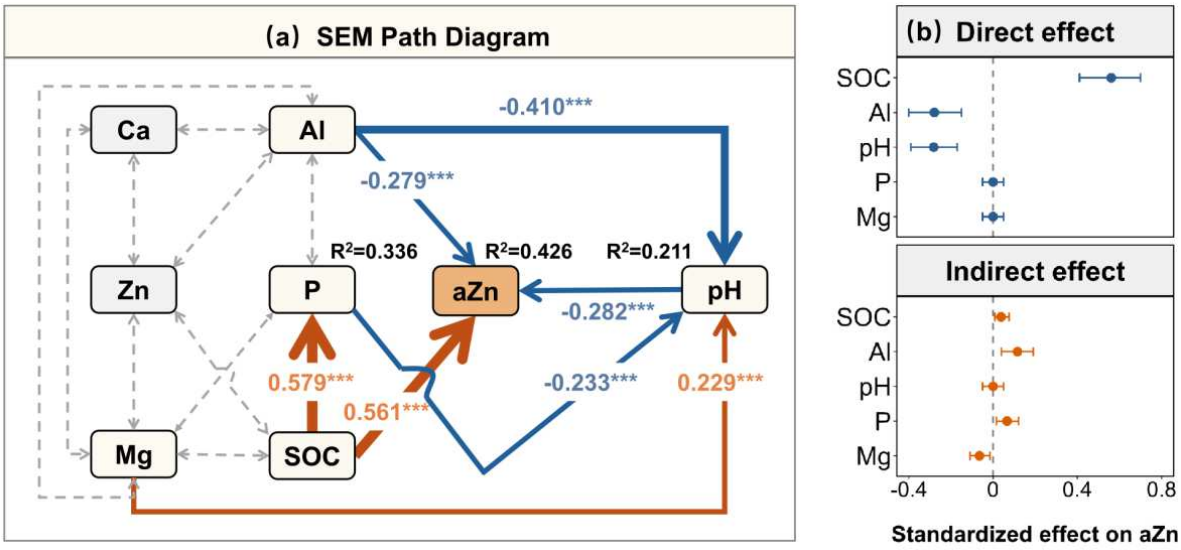


Fig. 6. SEM of soil-property associations with aZn. (a) Path diagram; orange and blue arrows indicate positive and negative paths, and values are standardized coefficients (β). Gray dashed double-headed arrows indicate specified covariances. R^2 is the proportion of variance explained for each endogenous variable on its modeled scale. (b) Standardized direct and indirect associations with aZn; points and error bars show estimates and confidence intervals.

3.5. Predicted concentration classes and township summaries

Predicted aZn varied across the landscape, with distinct patches of low and high values (Fig. 7a). The intermediate class (0.5 to < 1.0 mg kg^{-1}) occupied the largest mapped area, while the low class (<0.5 mg kg^{-1}) had a limited distribution.

WRI ranged from 100 to 200 across the 106 townships. Using the regional mean and mean + 1.5 SD as category boundaries, 57.5% of townships had low WRI, 34.9% had moderate WRI, and 7.5% had high WRI (Fig. 7b). Mean WRI in the high category was 189.97. Supplementary Table S3 lists township class proportions and WRI scores. These percentages refer to township counts. Higher WRI indicates a greater relative contribution of lower predicted aZn classes.

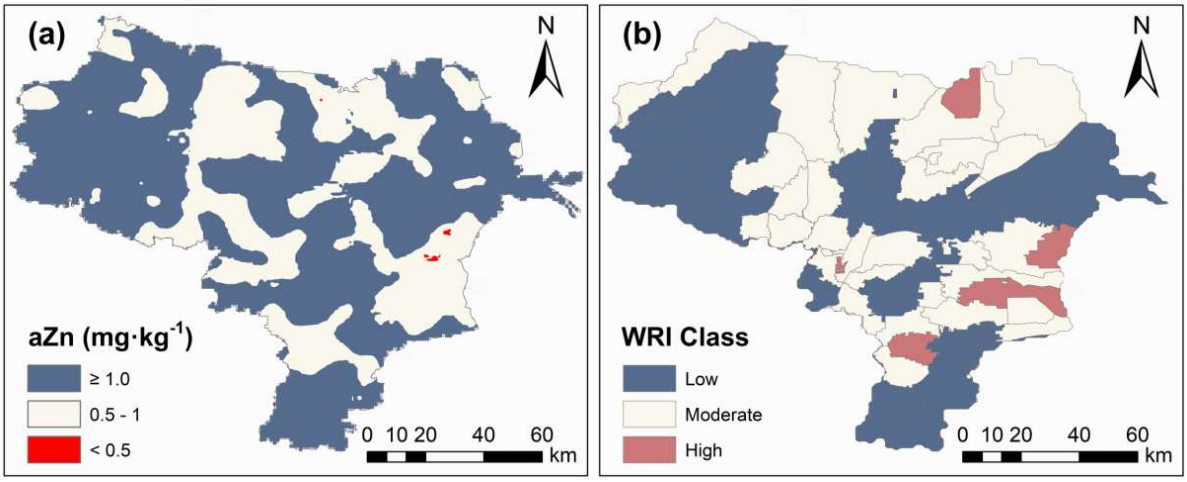


Fig. 7. Predicted aZn classes and township WRI categories. (a) Classes derived from the 1-km prediction grid: <0.5, 0.5 to <1.0, and ≥1.0 mg kg^{-1}. (b) WRI categories summarizing the proportions of the three predicted classes within each township.

4. Discussion

4.1. Interpreting spatial differences between total and extractable Zn

The spatial mismatch shows that total Zn and aZn describe different aspects of soil Zn status. Some soils had low aZn despite relatively high total Zn, while the regional mean total Zn was close to the cited background. Total concentration therefore gave an incomplete account of extractable Zn. This finding is consistent with studies of Zn partitioning and availability (Alloway, 2009; Liu et al., 2020) and with spatial contrasts reported in other agricultural regions (Taşpınar et al., 2025). The present survey relates these contrasts to the soil conditions in a coastal saline-alkaline landscape. Higher pH can reduce Zn solubility through sorption and precipitation (Chahal et al., 2023; de Francisco et al., 2024; Sparks et al., 1996). Here, pH remained associated with aZn after controlling for total Zn, whereas EC and NDVI did not. Partial correlations with Al, Ca, Mg, and P further place aZn variation within a broader soil chemical context. These associations are consistent with several possible retention processes; they do not identify individual reactions or reactive mineral surfaces.

Salinity can influence Zn speciation and adsorption through ionic strength and aqueous complexation, with effects that depend on pH and soil composition (Hanif et al., 2024). Covariation among soil properties complicates the separation of these effects in a field survey. The weak EC association after accounting for total Zn therefore needs to be interpreted alongside pH

and the other measured soil properties. For monitoring, the relevant combination is extractable Zn and the chemical conditions that influence its interpretation.

4.2. Environmental associations in the low-aZn group

The low-aZn group contained 45 samples spanning 0.21–0.49 mg kg^{-1}, compared with 156 samples spanning 0.50–17.93 mg kg^{-1} in the higher-aZn group. These differences in sample size and response range affect within-group correlations. Restricting the range of the response can attenuate associations, and a significant correlation in one group and a nonsignificant correlation in another do not establish a difference between their relationships (Gelman and Stern, 2006). Data exploration can help assess these patterns (Zuur et al., 2010), while a direct comparison requires a formal test of the correlations.

Geochemical retention offers a plausible explanation for low extractability. In alkaline soils, Zn can associate with sorbing surfaces or less labile solid phases (Xiang et al., 1995; Antoniadis et al., 2018; Sparks et al., 2022). The present survey measured bulk soil properties and DTPA-extractable Zn. Fractionation, solution-speciation measurements, or controlled pH experiments would be needed to distinguish the underlying retention processes.

The 0.5 mg kg^{-1} cutoff provides an operational basis for selecting soils for confirmatory sampling and crop diagnosis. Its agronomic interpretation depends on calibration for the crops and soil conditions of interest.

4.3. SOC and pH as joint correlates of extractable Zn

SOC and pH were prominent in both model descriptions. Because the pH × SOC and SOC–DEM predictors contain their component variables, their SHAP contributions share information. The percentages therefore describe attribution within this model. Recalculating the derived predictors maintains algebraic consistency in the PDPs, while their environmental interpretation also depends on the combinations of SOC, pH, and other variables represented in the observations (Apley and Zhu, 2020).

SOC contributed most to random forest predictions, and the modeled SOC–aZn relationship varied with pH. Organic matter can influence Zn through complexation, sorption, and dissolved organic ligands (Kaur et al., 2024; Violante et al., 2010). These effects depend on organic-matter composition and soil chemistry (Klinkert and Comans, 2020; Laurent et al., 2024). The positive

SOC–aZn association is therefore chemically plausible. Its interpretation also needs to account for covariation with texture, management history, and other soil characteristics.

The negative pH–aZn association is consistent with Zn retention in alkaline soils (Lakshmi et al., 2021). The PDP changes near pH 8.2 and SOC 9 g kg^{-1} describe features of the fitted model. The low modeled aZn below approximately 9 g kg^{-1} SOC and its pH dependence at higher SOC distinguish soil chemical settings for further comparison. These numerical features have not been calibrated as agronomic thresholds.

SEM also associated aZn with SOC and a network of pH-related pathways. It complements the random forest description by representing the covariance structure among the measured soil properties. Both analyses use the same observations, and the SEM includes data-driven path selection. Testing the response to carbon addition or pH adjustment would require controlled trials with crop measurements.

4.4. Selecting locations for further soil assessment

The predicted classes and township summaries provide a spatial overview for planning follow-up sampling. Previous studies have used soil information to delineate nutrient and salt-affected land management zones (Barman et al., 2021; Shukla et al., 2024). Here, sampling can focus on locations where total Zn and aZn disagree and where predicted aZn approaches a classification boundary. Confirming these locations would clarify the local extent of low extractability and the stability of the mapped classes.

WRI condenses the proportions of three concentration classes into one score. A higher score directs attention to townships with a greater contribution from lower-aZn classes, but different class combinations can produce the same score. Interpretation therefore requires the underlying class proportions and local soil evidence. Sensitivity to the concentration boundaries, empirical weights, and interpolation uncertainty should be assessed before the index is used to rank intervention priorities.

A monitoring program could first measure total Zn, aZn, pH, and SOC together, then combine the soil results with crop observations and locally calibrated tests of Zn response. Where low aZn is confirmed, contrasting SOC and pH conditions can guide site selection for field trials. This

sequence connects regional soil assessment to the evaluation of locally appropriate Zn management. Applying it in other saline-alkaline regions would help assess its wider relevance.

4.5. Scope of inference and validation needs

The survey covered one region at one sampling time, and the models did not include agricultural management history. DTPA extraction measures an operational soil fraction whose relationship with crop uptake varies with soil conditions, crop species, and analytical protocol. Crop uptake and yield measurements would allow the soil-test results to be calibrated for agronomic use.

Model performance was evaluated with a random data split. Spatially blocked validation and independent samples are needed to assess prediction at unsampled locations (Roberts et al., 2017). The 1-km grid specifies prediction spacing; it does not establish accuracy at that scale. Uncertainty in interpolated covariates and class assignment also affects township summaries (Wadoux et al., 2020).

Two features of the prediction workflow affect interpretation. The PDP wrapper preserves algebraic relationships among derived predictors, while joint data support remains relevant when predictors are correlated (Apley and Zhu, 2020). Exponentiating log-scale predictions returns values on the concentration scale without correcting retransformation bias (Duan, 1983). Independent assessment of class stability is therefore important when these predictions are compared with fixed concentration boundaries.

5. Conclusions

Total and DTPA-extractable Zn showed contrasting spatial distributions in the coastal saline-alkaline agricultural soils of the Yellow River Delta. Mean total Zn was close to the cited regional background, but 45 of 201 samples contained less than 0.5 mg kg^{-1} aZn. Total Zn alone did not characterize extractable Zn in these soils.

SOC and pH were major correlates of aZn. Random forest interpretation indicated nonlinear, conditional patterns, while SEM described related soil-property associations. Subgroup comparisons were descriptive because sample sizes and response ranges differed.

The results support measuring aZn alongside total Zn, SOC, and pH and using spatial summaries to select sites for confirmatory sampling and crop-response trials. Spatial validation, uncertainty

assessment, and crop calibration would strengthen the application of these spatial classifications in understanding trace element behavior and agricultural health.

CRediT authorship contribution statement

Taolin Zhou: Methodology, Data curation, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. Mengge Zhou: Investigation, Data curation, Methodology. Ziqiao Li: Investigation. Yi Yang: Conceptualization, Investigation, Funding acquisition. Yonghua Li: Supervision, Writing - review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant number 42507632].

Data availability

Data will be made available on request.

Declaration of generative AI and AI-assisted technologies in the writing process

During manuscript preparation, the authors used ChatGPT (OpenAI) to improve language, and readability. The authors reviewed and edited the output and take full responsibility for the content of the manuscript.

References

Alloway, B.J., 2008. Zinc in Soils and Crop Nutrition. International Zinc Association, Brussels.

Alloway, B.J., 2009. Soil factors associated with zinc deficiency in crops and humans. Environmental Geochemistry and Health 31, 537–548. https://doi.org/10.1007/s10653-009-9255-4

Antoniadis, V., Shaheen, S.M., Tsadilas, C.D., Selim, M.H., Rinklebe, J., 2018. Zinc sorption by different soils as affected by selective removal of carbonates and hydrous oxides. Applied Geochemistry 88, 49–58. https://doi.org/10.1016/j.apgeochem.2017.04.007

Apley, D.W., Zhu, J., 2020. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Journal of the Royal Statistical Society Series B: Statistical Methodology 82, 1059–1086. https://doi.org/10.1111/rssb.12377

Barman, A., Sheoran, P., Yadav, R.K., Abhishek, R., Sharma, R., Prajapat, K., Singh, R.K., Kumar, S., 2021. Soil spatial variability characterization: Delineating index-based management zones in salt-affected agroecosystem of India. Journal of Environmental Management 296, 113243. https://doi.org/10.1016/j.jenvman.2021.113243

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Broadley, M.R., White, P.J., Hammond, J.P., Zelko, I., Lux, A., 2007. Zinc in plants. New Phytologist 173, 677–702. https://doi.org/10.1111/j.1469-8137.2007.01996.x

Chahal, S.K., Hettiarachchi, G.M., Nelson, N.O., Guttieri, M.J., 2023. Fate and plant uptake of different zinc fertilizer sources upon their application to an alkaline calcareous soil. ACS Agricultural Science & Technology 3, 725–737. https://doi.org/10.1021/acsagscitech.2c00287

de Francisco, M., Romeiro, A., Durães, L., Álvarez-Torrellas, S., Ibañez, M.A., Almendros, P., 2024. Environmental behaviour of synthesized and commercial agricultural zinc products: Leaching, migration, and availability in soils. Journal of Soil Science and Plant Nutrition 24, 5293–5308. https://doi.org/10.1007/s42729-024-01908-5

Duan, N., 1983. Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association 78, 605–610. https://doi.org/10.1080/01621459.1983.10478017

Duffner, A., Weng, L., Hoffland, E., van der Zee, S.E.A.T.M., 2014. Multi-surface Modeling To Predict Free Zinc Ion Concentrations in Low-Zinc Soils. Environmental Science & Technology 48, 5700–5708. https://doi.org/10.1021/es500257e

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Gelman, A., Stern, H., 2006. The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. The American Statistician 60, 328–331. https://doi.org/10.1198/000313006X152649

Grace, J.B., Irvine, K.M., 2020. Scientist’s guide to developing explanatory statistical models using causal analysis principles. Ecology 101, e02962. https://doi.org/10.1002/ecy.2962

Haklay, M., Weber, P., 2008. OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing 7, 12–18. https://doi.org/10.1109/MPRV.2008.80

Hanif, Md., Bullen, J., Plancherel, Y., Kirby, M., Kirk, G., Weiss, D., 2024. Significant effect of salinity on zinc adsorption on tropical coastal and floodplain soils. European Journal of Soil Science 75, e13575. https://doi.org/10.1111/ejss.13575

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 6, 1–55. https://doi.org/10.1080/10705519909540118

Kabata-Pendias, A., 2010. Trace Elements in Soils and Plants, 4th ed. CRC Press, Boca Raton. https://doi.org/10.1201/b10158

Kaur, H., Srivastava, S., Goyal, N., Walia, S., 2024. Behavior of zinc in soils and recent advances on strategies for ameliorating zinc phyto-toxicity. Environmental and Experimental Botany 220, 105676. https://doi.org/10.1016/j.envexpbot.2024.105676

Khoshgoftarmanesh, A.H., Afyuni, M., Norouzi, M., Ghiasi, S., Schulin, R., 2018. Fractionation and bioavailability of zinc (Zn) in the rhizosphere of two wheat cultivars with different Zn deficiency tolerance. Geoderma 309, 1–6. https://doi.org/10.1016/j.geoderma.2017.08.019

Kline, R.B., 2023. Principles and Practice of Structural Equation Modeling, 5th ed. The Guilford Press, New York.

Klinkert, S., Comans, R.N.J., 2020. Geochemical Multisurface Modeling of Reactive Zinc Speciation in Compost as Influenced by Extraction Conditions. Environmental Science & Technology 54, 2467–2475. https://doi.org/10.1021/acs.est.9b04104

Lakshmi, P.V., Singh, S.K., Pramanick, B., Kumar, M., Laik, R., Kumari, A., Shukla, A.K., Abdel Latef, A.A.H., Ali, O.M., Hossain, A., 2021. Long-Term Zinc Fertilization in Calcareous Soils Improves Wheat (Triticum aestivum L.) Productivity and Soil Zinc Status in the Rice–Wheat Cropping System. Agronomy 11, 1306. https://doi.org/10.3390/agronomy11071306

Laurent, C., Bravin, M.N., Crouzet, O., Lamy, I., 2024. Does a decade of soil organic fertilization promote copper and zinc phytoavailability? Evidence from a laboratory biotest with field-collected soil samples. Science of The Total Environment 906, 167771. https://doi.org/10.1016/j.scitotenv.2023.167771

Lindsay, W.L., Norvell, W.A., 1978. Development of a DTPA Soil Test for Zinc, Iron, Manganese, and Copper. Soil Science Society of America Journal 42, 421–428. https://doi.org/10.2136/sssaj1978.03615995004200030009x

Liu, Y.-M., Liu, D.-Y., Zhao, Q.-Y., Zhang, W., Chen, X.-X., Xu, S.-J., Zou, C.-Q., 2020. Zinc fractions in soils and uptake in winter wheat as affected by repeated applications of zinc fertilizer. Soil and Tillage Research 200, 104612. https://doi.org/10.1016/j.still.2020.104612

Liu, Z., 1996. Microelements in soils of China. Jiangsu Science and Technology Publishing House, Nanjing.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Moreno-Lora, A., Delgado, A., 2020. Factors determining Zn availability and uptake by plants in soils developed under Mediterranean climate. Geoderma 376, 114509. https://doi.org/10.1016/j.geoderma.2020.114509

Peng, Z., Liu, Y., An, Y., Wang, Z., Li, X., Gao, H., Wei, G., Jiao, S., 2023. Contrasting patterns and drivers of soil micronutrient availability in paddy and maize fields of eastern China. Geoderma 431, 116342. https://doi.org/10.1016/j.geoderma.2023.116342

Pichler, M., Hartig, F., 2023. Machine learning and deep learning—A review for ecologists. Methods in Ecology and Evolution 14, 994–1016. https://doi.org/10.1111/2041-210X.14061

Qadir, M., Quillérou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R.J., Drechsel, P., Noble, A.D., 2014. Economics of salt-induced land degradation and restoration. Natural Resources Forum 38, 282–295. https://doi.org/10.1111/1477-8947.12054

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., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881

Rosseel, Y., 2012. lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software 48, 1–36. https://doi.org/10.18637/jss.v048.i02

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., Zhang, Y., 2013. A China data set of soil properties for land surface modeling. Journal of Advances in Modeling Earth Systems 5, 212–224. https://doi.org/10.1002/jame.20026

Shukla, A.K., Behera, S.K., Basumatary, A., Sarangthem, I., Mishra, R., Dutta, S., Sikaniya, Y., Sikarwar, A., Shukla, V., Datta, S.P., 2024. PCA and fuzzy clustering-based delineation of soil nutrient (S, B, Zn, Mn, Fe, and Cu) management zones of sub-tropical Northeastern India for precision nutrient management. Journal of Environmental Management 365, 121511. https://doi.org/10.1016/j.jenvman.2024.121511

Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), 1996. Methods of Soil Analysis: Part 3 Chemical Methods, 1st ed, SSSA Book Series. Wiley, Madison. https://doi.org/10.2136/sssabookser5.3

Sparks, D.L., Singh, B., Siebecker, M.G., 2022. Environmental Soil Chemistry, 3rd ed. Academic Press.

Taşpınar, K., Aytop, H., Ateş, Ö., Varol, M., Yalçın, G., Kızılaslan, F., Çakıllı, H., Toprak, S., 2025. Zinc bioavailability in semiarid agricultural regions: a soil property-based assessment. Environmental Geochemistry and Health 47, 238. https://doi.org/10.1007/s10653-025-02544-3

Violante, A., Cozzolino, V., Perelomov, L., Caporale, A.G., Pigna, M., 2010. Mobility and bioavailability of heavy metals and metalloids in soil environments. Journal of Soil Science and Plant Nutrition 10, 268–292. https://doi.org/10.4067/S0718-95162010000100005

Wadoux, A.M.J.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

568 Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil
569 organic matter, and a proposed modification of the chromic acid titration method. Soil Science
570 37, 29–38.
571 Wang, J., Yang, T., Zhu, K., Shao, C., Zhu, W., Hou, G., Sun, Z., 2023. A novel retrieval model
572 for soil salinity from CYGNSS: Algorithm and test in the Yellow River Delta. Geoderma 432,
573 116417. https://doi.org/10.1016/j.geoderma.2023.116417
574 Wei, F. (Ed.), 1990. Background Values of Soil Elements in China. China Environmental Science
575 Press, Beijing (in Chinese).
576 Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High
577 Dimensional Data in C++ and R. Journal of Statistical Software 77, 1–17.
578 https://doi.org/10.18637/jss.v077.i01
579 Xia, J., Ren, J., Zhang, S., Wang, Y., Fang, Y., 2019. Forest and grass composite patterns improve
580 the soil quality in the coastal saline-alkali land of the Yellow River Delta, China. Geoderma
581 349, 25–35. https://doi.org/10.1016/j.geoderma.2019.04.032
582 Xiang, H.F., Tang, H.A., Ying, Q.H., 1995. Transformation and distribution of forms of zinc in
583 acid, neutral and calcareous soils of China. Geoderma 66, 121–135.
584 https://doi.org/10.1016/0016-7061(94)00067-K
585 Xu, X., 2018. China's Annual 1 km NDVI and EVI Dataset. Resource and Environmental Science
586 and Data Center, Chinese Academy of Sciences. https://doi.org/10.12078/2018060601
587 Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common
588 statistical problems. Methods in Ecology and Evolution 1, 3–14.
589 https://doi.org/10.1111/j.2041-210X.2009.00001.x

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx