soil systems

MDPI

Article

Local Surrogate Relationships Between Soil Texture Fractions and Near-Surface Hydro-Structural Properties for Hydrological Parameterization in High-Andean Catchments

Christian Mera-Parra [1,2,*], Pablo Ochoa-Cueva [3], Jose Damian Ruiz Sinoga [4] and Paola Duque Sarango [1]

[1] Water Resources Research Group (GIRH-UPS), Universidad Politécnica Salesiana, Cuenca 010102, Ecuador; pduque@ups.edu.ec
[2] Doctoral Program in City, Territory and Sustainable Planning, University of Málaga, 29010 Malaga, Spain
[3] Department of Biological and Agricultural Sciences, Universidad Técnica Particular de Loja (UTPL), Loja 110107, Ecuador; paochoa@utpl.edu.ec
[4] Physical Geography and Territory Group, Department of Geography, University of Málaga, 29071 Malaga, Spain; sinoga@uma.es
* Correspondence: cmera@ups.edu.ec or cmera@uma.es; Tel.: +593-7-370-1342

Abstract

For hydrological parameterization in high-Andean catchments, it is necessary to understand whether near-surface hydro-structural soil properties can provide a surrogate signal of particle-size composition when direct texture information is sparse. This study evaluated the extent to which sand, silt, and clay fractions can be approximated from organic matter (OM), bulk density (ρ_b), and saturated hydraulic conductivity (K_{sat}) in the Zamora Huayco (ZH) and Irquis catchments, southern Ecuador. A harmonized dataset ($n = 44$) was analyzed through exploratory statistics, compositional assessment, correlation analysis, PCA, fraction-wise regression, ILR-based modeling, AIC/BIC term reduction, sensitivity analysis excluding OM, nested LOOCV, and bootstrap-based uncertainty intervals. Among LULC classes, samples classified as paramo occupied a distinct high-Andean hydro-edaphic domain, characterized by a differentiated relationship between soil physical properties and hydrological behavior. PCA showed that the dominant covariance structure involved OM, ρ_b, K_{sat}, and the redistribution between sand and silt. The BIC-reduced ILR model provided the most balanced formulation, with positive nested LOOCV performance for sand, silt, and clay ($R^2_{\mathrm{LOOCV}} = 0.147$, 0.704, and 0.124, respectively) and exact 100% compositional closure after inverse transformation. Silt was the most stable predicted fraction, whereas sand and clay retained larger residual uncertainty, stronger tail departures, and partial compression of the observed variability. The proposed equations provide local hydro-pedotransfer support, although their predictive signal remains dependent on further refinement, uncertainty assessment, and external validation before regional application.

Keywords: hydro-pedotransfer functions; soil physical properties; high-Andean catchments; data-scarce basins; predictive uncertainty

Check for updates

Academic Editor: Richard Wilkin

Received: 28 March 2026
Revised: 15 June 2026
Accepted: 18 June 2026
Published: 23 June 2026

1. Introduction

Distributed and semi-distributed hydrological models need spatial soil information to represent infiltration, runoff partitioning, subsurface flow, and soil water storage [1–3]. Among the most relevant inputs are particle-size fractions, ρ_b, OM, and K_{sat}, because these properties control water storage and the activation of flow paths [4]. Uncertainty in these variables, especially porosity-related properties and K_{sat}, can propagate nonlinearly into simulated hydrological responses, making soil parameterization a key source of error in basin-scale modeling [5,6].

At the field scale, infiltration regulates how rainfall is partitioned into groundwater recharge, drainage response, and near-surface soil water storage across Andean vegetation units [7]. Experimental evidence has shown that infiltration patterns are strongly influenced by soil physical properties and land-use conditions, making them relevant for sustainable soil and water management [8]. In mountain catchments, this hydrological response becomes particularly sensitive to soil-structure changes, organic matter accumulation, and land-cover transitions [9]. Although digital soil products and higher-resolution mapping can reduce part of the spatial uncertainty, locally representative hydraulic and edaphic measurements remain necessary, especially where steep topographic and ecological gradients occur over short distances and available soil observations are sparse [2].

Soil texture is commonly treated as a fundamental and relatively accessible descriptor in pedotransfer and hydro-pedological parameterization [10,11]. Conventional PTF development has therefore followed a direct direction, estimating soil water retention or hydraulic parameters from texture and related descriptors through empirical equations, van Genuchten-type formulations, hierarchical tools such as ROSETTA, and recent random-forest approaches [4,12–15]. However, this assumption is less straightforward in organic-rich Andean volcanic soils, where low bulk density, strong aggregation, high OM contents, and incomplete dispersion may affect particle-size determination [9,16,17]. In addition, OM quantification itself requires caution because loss-on-ignition can include not only organic matter but also volatilized mineral water and carbonates, particularly in soils with high organic and short-range-order mineral contents [18,19]. Therefore, K_{sat}, ρ_b, and OM are not proposed as simpler substitutes for direct texture analysis, but as hydro-structural descriptors that may retain a local surrogate signal of sand–silt–clay composition under Andean conditions.

This question must be addressed with approaches that remain interpretable under limited sample size. In data-scarce mountain basins, highly flexible predictive schemes may be difficult to generalize robustly when data volume, representativeness, and covariate transferability are limited [20,21], whereas parsimonious regression-based surrogate relationships preserve explicit equations and facilitate interpretation of the dominant links between structural descriptors and textural variability [21,22]. Under this perspective, the goal is not to replace direct particle-size analysis, but to evaluate how much of textural variability can be approximated from measurable structural proxies and where the main information losses occur.

High-Andean environments intensify this problem because steep topographic gradients, organic-rich Andosol/Histosol surface horizons developed on volcanic materials, and land-use pressures coexist over short distances [23,24]. In high-Andean soils, the imprint of LULC disturbance is expected to appear more clearly in hydro-structural properties than in sand–silt–clay composition, making OM, ρ_b, ϕ, water retention, and K_{sat} more sensitive indicators of short- to medium-term soil reorganization [25–27]. Recent studies in southern Ecuador have likewise shown that land-use dynamics and ecohydrological gradients modulate water regulation and ecosystem-service provision across Andean basins [24,28,29]. Accordingly, texture is treated here as a relatively stable descriptor of the mineral fraction, whereas OM, ρ_b, and K_{sat} are interpreted as more disturbance-sensitive variables that may reveal shifts in hydro-edaphic organization when direct textural information is limited [23,27,30,31].

Within this context, the key unresolved issue is not only whether sand, silt, and clay fractions are statistically associated with near-surface OM, ρ_b, and K_{sat}, but whether this association can be represented through parsimonious, physically interpretable, and com-

https://doi.org/10.3390/soilsystems10070068

positionally coherent surrogate relationships for hydrological parameterization. In high-Andean catchments, the sparse and discontinuous distribution of soil observations increases the relevance of this issue [32,33]. Under such conditions, the practical value of a surrogate approach depends less on predictive complexity than on interpretability, parsimony, and transferability [21,22], as well as on its capacity to preserve the closed nature of sand–silt–clay fractions while supporting subsequent hydrological model parameterization [22].

Organic-rich Andosols and associated Histosols in Quimsacocha and southern Ecuador show distinctive hydro-physical behaviour and texture-analysis constraints [9,34,35], while broader evidence indicates that OM removal, dispersion efficiency, and sedimentation assumptions can alter recovered particle-size fractions in organic-rich or volcanic soils [36–40]. Accordingly, this study evaluates the extent to which sand, silt, and clay fractions can be approximated from near-surface OM, ρ_b, and K_{sat} in two Andean catchments of southern Ecuador, Irquis and ZH, using a harmonized soil database. The analysis focuses on the covariance structure linking hydro-structural properties with textural composition, while LULC is used as a descriptive context for interpreting disturbance-related variability rather than as a direct driver of mineral texture. The aim is to quantify how far near-surface hydro-structural properties capture textural variability, identify the dominant hydro-edaphic gradients in the dataset, and evaluate the usefulness and limitations of this local, domain-specific approach for hydrological parameterization.

2. Materials and Methods

2.1. Study Area

The study was conducted in two headwater catchments of the southern Ecuadorian Andes: the Irquis Basin (Azuay Province, Cuenca canton) and ZH Basin (Loja Province). Figure 1 shows the geographic setting, sample points, and elevation gradients used to frame the field sampling and spatial analyses.

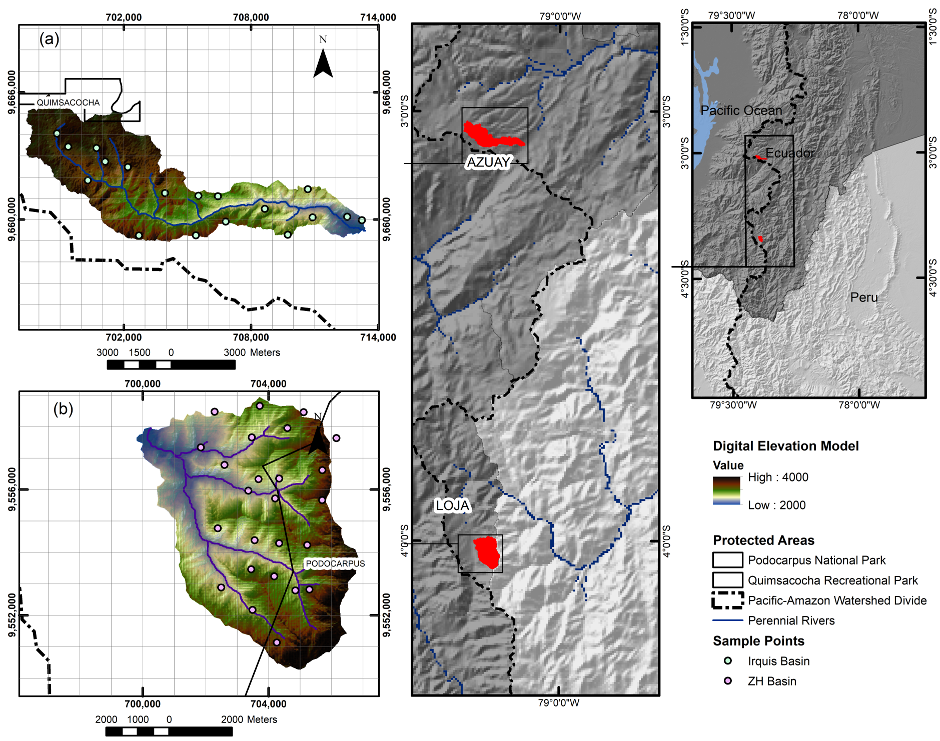

Figure 1. The location of the study basins in southern Ecuador. (**a**) Irquis Basin. (**b**) ZH Basin. The elevation, perennial rivers, sampling points, and nearby protected areas are shown for each basin. The black dashed line indicates the regional Andean drainage divide.

2.1.1. Irquis Basin

The Irquis Basin is located in the parish of Victoria del Portete, Cuenca canton, Azuay Province, in the southern Ecuadorian Andes (Figure 1a). From our study map, the basin spans approximately 702,000–714,000 m E and 9,660,000–9,666,000 m N (UTM Zone 17S), and lies in the vicinity of the Quimsacocha National Recreation Area. Based on the basin delineation adopted in this study, the drainage area is approximately 44.97 km^2, and elevation ranges from about 2640 to 3860 m a.s.l. The basin is representative of high-Andean headwaters where steep topographic gradients, paramo environments, and volcanic materials coexist over short distances. At the regional scale, the Quimsacocha paramo has been described as a humid tropical alpine environment with low temperatures and high precipitation, whereas southern Ecuadorian mountain catchments are characterized by strong hydroclimatic gradients and frequent data scarcity [41,42]. In comparable high-elevation catchments of southern Ecuador, Andosols and Histosols formed from volcanic ash are rich in organic matter and have high water-storage capacity [43]. For the Irquis natural-cover site at 2800 m a.s.l., Ref. [31] reported OM values ranging from 7.09% to 14.75%, while median K_{sat} decreased from 12.91 cm h^{-1} in the 0–10 cm layer to 2.98 cm h^{-1} in the 10–25 cm layer. In the Quimsacocha páramo, Ref. [35] reported higher surface OM values, ranging from 191.00 to 483.83 g kg^{-1}, equivalent to approximately 19.10–48.38%, and surface K_{sat} values between 0.50 and 10.84 cm h^{-1} across land-cover classes.

2.1.2. ZH Basin

The ZH Basin is an inter-Andean mountain catchment in Loja Province (southern Ecuador), with an upper sector influenced by the Podocarpus National Park (Figure 1b), and a downstream sector increasingly exposed to peri-urban and urban pressures from Loja city [24]. From our study map, the basin spans approximately 700,000–706,000 m E and 9,552,000–9,558,000 m N, with elevations of ~2060 to 3381 m.a.s.l. ZH has been described as steep (mean slope~0.65 m m^{-1}), where topography and seasonal rainfall promote runoff and sediment redistribution, with higher susceptibility under disturbed land cover [44]. The climate is reported as cold temperate mesothermal (~12–18°C; ~1047 mm), with a wet period centered on December–May and a dry period from June–November [1]. Land cover follows a conserved-to-disturbed gradient (more natural vegetation upslope; more agriculture and cattle raising downslope), while urban growth increases imperviousness, reducing infiltration and baseflow and increasing runoff [44].Available basin-scale soil information identifies loam as the dominant texture class and reports ranges of bulk density (~0.07–1.09 g cm^{-3}), organic carbon (~1.66–5.98%), and saturated hydraulic conductivity (~4.6–8.9 mm h^{-1}), suggesting a disturbance gradient in which less-intervened sectors tend to conserve soil structure and infiltration pathways, while downslope grazing, agriculture, and sealing are more consistent with compaction and lower effective infiltration capacity [24,44].

Figure 1 shows both basins as high-Andean headwater catchments near the Pacific–Amazon divide, where precipitation is partly conditioned by Amazonian moisture advection and windward orographic processes [45].

2.2. Soil Database, Sampling Design, and Variables

A harmonized soil database was assembled by merging two independent datasets: (i) ZH, provided by [44], and (ii) Irquis, specifically generated for this research by the Water Resources Research Group (GIRH-UPS), Universidad Politécnica Salesiana. The original ZH dataset comprised 43 field records distributed across the study microcatchments; however, only 26 observations were retained because they contained the complete set of variables required for the harmonized analysis. Because both datasets were generated independently, the analytical origin of OM was explicitly documented. In the Irquis Basin, OM was quantified by loss on ignition following AASHTO T 267-22, whereas OM values in the ZH database were determined using the Walkley–Black wet oxidation method [44]. OM was retained as a common edaphological variable, while the basin-specific analytical method was reported to preserve traceability and avoid treating LOI- and Walkley–Black-derived values as strictly interchangeable without method-specific calibration [18,46,47].

For Irquis, sampling followed a representativeness-oriented spatial framework that subsequently evolved into a sequential, interpolation-guided strategy [48–51]. Preliminary observations were used to generate interpolation surfaces of the main hydro-edaphic variables, which were then reclassified to identify spatial inconsistencies, transition zones, and areas requiring additional coverage [52,53]. Sampling was progressively intensified through stratified random selection over updated interpolation-based strata, allowing improved representation of hydro-edaphic heterogeneity across successive campaigns [54,55]. The complete Irquis spatial database comprised 79 georeferenced points with measurements of OM, K_{sat}, ρ_b, ρ_s, and ϕ. The sequential sampling procedure, including the intermediate interpolation maps, reclassification criteria, phase-specific point additions, sample-size calculations, and final spatial distribution of the Irquis and ZH records, is provided in the Supplementary Material.

Particle-size analysis was conducted on a subset of 18 Irquis points selected from the consolidated spatial database after surface hydro-structural heterogeneity had been characterized. Soil texture was determined using the Bouyoucos hydrometer method [39,56]. Prior to hydrometer readings, OM was removed with hydrogen peroxide, and sodium hexametaphosphate was used as a dispersing agent to promote particle separation [38,40,56]. Hydrometer readings were taken at 40 s and 2 h to estimate sand, silt, and clay fractions, and textural classes were assigned according to the USDA textural triangle [57]. Only these 18 Irquis points, containing both textural composition and structural soil properties, were included in the present comparative analysis. Together with the 26 complete ZH observations selected from the original 43 field records, the final harmonized database used in this study comprised $n = 44$ sampling points.

The laboratory and field procedures described below refer to near-surface soil samples and field measurements from the Irquis dataset generated in this study; soil samples were collected adjacent to the infiltration points from the 0–15 cm layer after removing the surface vegetation/litter cover and visible coarse roots. ρ_b, was determined from air-dried soil passed through a No. 40 sieve and packed into a graduated cylinder of known final volume. Particle density, ρ_s, was measured using a volumetric flask based on water displacement. Total porosity, ϕ, was then calculated from ρ_b and ρ_s as $\phi = \left(1 - \frac{\rho_b}{\rho_s}\right) \times 100$ [58,59]. OM was quantified by loss on ignition following AASHTO T 267-22 using 20 g subsamples ignited at $(455 \pm 10\ ^\circ\text{C})$ for 6 h, and reported as a percentage [60]. K_{sat} was estimated in situ from double-ring infiltrometer tests with rings embedded to approximately 150 mm according to ASTM D3385-03 [61]. In the project workflow, K_{sat} was approximated from the stabilized basic infiltration rate (I_b) obtained during the final quasi-steady stage of double-ring infiltrometer tests, when near-surface flow approaches field-saturated conditions [62,63]. This field-based estimate is better supported in organic-rich high-Andean surface horizons, where high OM, low ρ_b, high ϕ, and high water-storage capacity are commonly reported [9,35]. Because organic-rich Andosols in southern Ecuador can hold water contents of 80–90% during wet periods and remain above the wilting point during droughts [9], stabilized infiltration was used as an operational field-scale estimate of saturated or near-saturated hydraulic behavior.

2.3. Exploratory Data Analysis and Data-Driven Modeling Framework

A harmonized soil database was assembled by integrating the ZH and Irquis datasets, retaining OM, ρ_b, K_{sat}, and available textural information. After standardizing abbreviations, variable names, and units, the workflow combined EDA, correlation screening, dimensionality reduction, and data-driven regression modeling [64–67]. Pearson and Spearman correlations described linear and monotonic associations, while histograms, boxplots, scatterplots, and scatter matrices were used to inspect distributions, outliers, and bivariate structure [64,66,67]. Predictor–response blocks were defined using OM, ρ_b, and $\log_{10}(K_{\mathrm{sat}})$ as core predictors and textural fractions as responses when available, to evaluate whether hydro-structural soil properties retained an interpretable surrogate signal of textural variability for hydrological parameterization [22,68]. Group contrasts and multivariate structure were explored using stratified boxplots and dimensionality reduction [64–67,69].

2.4. Exploratory Data Analysis and Correlation-Based Screening

The merged ZH–Irquis database was first inspected through EDA to evaluate variable distributions, detect potential inconsistencies, and identify candidate dependencies prior to modeling [64]. Histograms and boxplots were used for OM, ρ_b, K_{sat}, and available textural fractions; because K_{sat} commonly spans several orders of magnitude and is often right-skewed, it was analyzed as $\log_{10}(K_{\mathrm{sat}})$ for correlation screening and regression [64,70,71].

Bivariate structure was assessed using scatterplots, scatter matrices, and Pearson–Spearman correlation matrices to describe linear, monotonic, and potential nonlinear associations [64]. Predictor–response blocks were defined by hydro-edaphic relevance and data availability: OM, ρ_b, and $\log_{10}(K_{\mathrm{sat}})$ were treated as predictors, sand–silt–clay fractions as texture-inference targets, and texture class as a descriptive variable [22,68]. Group contrasts were explored with boxplots stratified by land-use and/or intervention class to evaluate changes in central tendency, dispersion, and overlap across disturbance levels [64].

2.5. Multivariate Analysis and Dimensionality Reduction

To summarize joint variability among hydro-structural and textural soil properties, we applied Principal Component Analysis (PCA) to the standardized variables in the combined ZH–Irquis database [65,72]. Variables were centered and scaled using z-scores before decomposition to prevent differences in measurement units from dominating component extraction [65]. PCA was used to identify the dominant hydro-edaphic gradients linking OM, ρ_b, and $\log_{10}(K_{\mathrm{sat}})$ with textural fractions in the subset with complete texture information. The analysis was exploratory and was intended to evaluate the covariance structure among hydro-structural and textural properties, rather than to provide predictive validation [72]. The first components explaining the largest share of total variance were interpreted through their score distribution and loading patterns, allowing the main structural and textural axes of the dataset to be described [65,72].

2.6. Surrogate Modeling, Compositional Closure, and Internal Validation

To evaluate the robustness of surrogate relationships without selecting a single model structure a priori, three model families were compared under a common internal validation framework [73,74]. First, fraction-wise models were fitted separately for sand, clay, and silt using OM, ρ_b, and $\log_{10}(K_{\mathrm{sat}})$ as predictors. This family included linear, full quadratic, and AIC- and BIC-reduced quadratic formulations [75]. Term selection followed a hierarchical rule, retaining interaction or quadratic terms only when their corresponding main effects were included [76]. Second, a compositional family was implemented using ILR coordinates [77].

The closed sand–clay–silt composition was transformed into two orthogonal ILR coordinates before regression, and predicted ILR values were back-transformed to sand, clay, and silt fractions to satisfy the 100% closure condition. Given the different procedures used to obtain OM, a sensitivity analysis excluded OM. This sensitivity analysis used only ρ_b and $\log_{10}(K_{\text{sat}})$ within both fraction-wise and ILR frameworks [78]. All models were validated using LOOCV, with centering, scaling, fitting, and prediction performed within each training subset [79].

For AIC- and BIC-reduced models, term selection was also performed within each training subset, ensuring nested model selection within the validation procedure [80,81]. Prediction uncertainty was quantified using a non-parametric bootstrap nested within LOOCV [82]. For each left-out observation, $B = 1000$ bootstrap samples were drawn with replacement from the corresponding training subset. In each replicate, predictors were re-centered and re-scaled using only the resampled training data, following the same validation logic used to avoid overfitting during model assessment [79]. BIC-based term selection was repeated for each ilr coordinate, and the left-out observation was predicted in ilr space. Bootstrap predictions were then back-transformed to sand, clay, and silt fractions, consistent with compositional modeling of soil particle-size fractions [81]. The 2.5th and 97.5th percentiles of the inverse-transformed predictions were used as bootstrap-based 95% uncertainty intervals. Performance was assessed using cross-validated R^2, $RMSE$, MAE, and bias; raw, post hoc closed, and ILR-based closed predictions were evaluated [81].

2.7. Laboratory Materials, Field Instruments, and Software

Laboratory and field procedures for the Irquis site were conducted at the Civil Engineering Laboratory of Universidad Politécnica Salesiana, Cuenca, Ecuador. Particle-size analysis was carried out using a Digi-Sense ASTM 152H soil hydrometer, model 08297-42 (Cole-Parmer, Vernon Hills, IL, USA), supplied by LabSupply Cía. Ltda. (Guayaquil, Ecuador). A No. 40 ASTM E11 test sieve (425 µm; Impact Test Equipment Ltd., Stevenston, Ayrshire, Scotland, UK) was used for sample preparation. Analytical reagents used for particle-size analysis were supplied by LabSupply Cía. Ltda. (Guayaquil, Ecuador). Samples were weighed using a Mettler Toledo ME203 precision balance (Mettler-Toledo GmbH, Greifensee, Switzerland). Loss-on-ignition was determined using a digital muffle furnace equipped with a SAMO Thermal 3K temperature controller (SAMO Thermal S.A.S., Cuenca, Ecuador). Field infiltration was measured using a double-ring infiltrometer set, model EJK-0904 (Royal Eijkelkamp, Giesbeek, The Netherlands). Statistical analyses, model fitting, LOOCV, bootstrap resampling, and figure generation were performed in MATLAB R2024b (MathWorks, Natick, MA, USA) under the academic license of Universidad Politécnica Salesiana. Spatial processing and map preparation were conducted using QGIS 3.34.11 (QGIS Development Team/QGIS Association, Laax, Switzerland).

3. Results

3.1. Exploratory Statistical Patterns of Soil Variables

Figure 2 presents scatterplot matrices showing the pairwise relationships among the analyzed soil variables: OM, ρ_b, $\log_{10}(K_{sat})$, and the particle-size fractions, grouped according to anthropogenic activity intensity. Low anthropogenic activity corresponds to conserved covers, including native forest and paramo, whereas high anthropogenic activity corresponds to disturbed covers dominated by pasture and secondary vegetation. The figure shows the dispersion of the soil variables and the distribution of samples across the multivariate space defined by these properties.

The scatter matrices show that the local LULC signal is expressed through hydro-structural variables; conserved covers occupy the portion of the multivariate space with higher OM and lower ρ_b, whereas disturbed covers are displaced toward lower OM and higher ρ_b. Variability in $\log_{10}(K_{\text{sat}})$ follows this hydro-structural gradient only partially, with substantial overlap among samples, suggesting that infiltration response is controlled by local structural heterogeneity. Sand, clay, and silt fractions show broader overlap between activity levels, suggesting that, within this dataset, texture provides a background edaphic condition.

To further examine these patterns, univariate and bivariate exploratory plots were analyzed to visualize the distribution of the main structural soil variables and their relationships with particle-size fractions. Figure 3 presents histograms of OM, ρ_b, and saturated hydraulic conductivity expressed as $\log_{10}(K_{sat})$, together with scatterplots relating these variables to sand, clay, and silt contents across the sampled sites.

The exploratory patterns suggest partial associations between hydro-structural properties and particle-size fractions. Variability in $\log_{10}(K_{\text{sat}})$ aligns broadly with sand and clay content, whereas OM and ρ_b show more dispersed relationships. This supported a more detailed evaluation of whether hydro-structural variables retain a surrogate signal of textural variability across the study area.

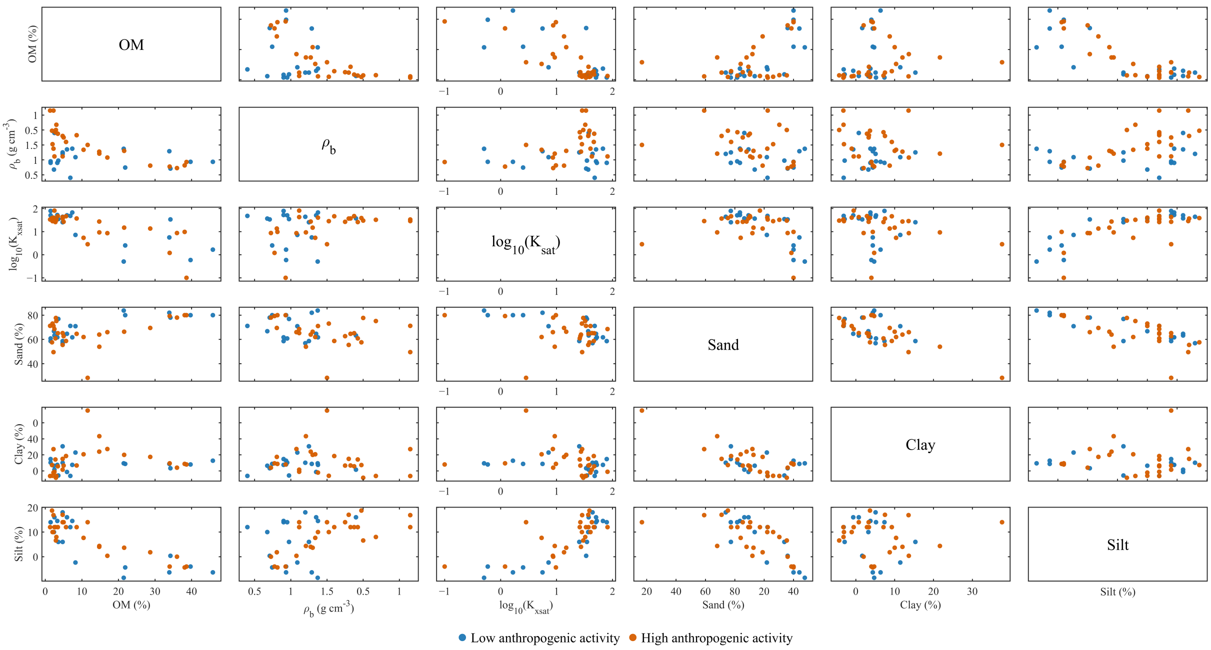

Figure 2. Multivariate scatter matrix of soil variables aggregated by anthropogenic activity level. Relationships between OM, ρ_b, $\log_{10}(K_{sat})$, sand, clay, and silt are shown for areas with low anthropogenic activity (blue) and high anthropogenic activity (orange).

Complementing that analysis, Figure 4 presents the boxplot distributions of OM, ρ_b, $\log_{10}(K_{\text{sat}})$, and the sand, silt, and clay fractions across the LULC classes considered in this study. While the previous analysis examined whether structural soil properties could help explain texture variability, this figure jointly evaluates whether hydro-structural properties and particle-size fractions vary systematically across LULC conditions.

The boxplots indicate systematic differences in structural soil properties across the land-use gradient. OM shows greater variability under pasture and paramo, with wider dispersion and higher maximum values, whereas native forest and shrubland exhibit lower and more compact distributions. ρ_b increases from native forest to shrubland and pasture, indicating denser soil conditions under disturbed covers. $\log_{10}(K_{\text{sat}})$, shows the highest median values under native forest, while shrubland and pasture display lower medians and greater dispersion. The particle-size fractions also vary among classes: paramo shows higher sand and lower silt contents with compact distributions, whereas pasture displays wider dispersion, particularly for sand, silt, and clay. These patterns indicate that conserved systems tend to maintain lower bulk density and higher hydraulic conductivity, whereas disturbed land uses show greater compaction and more variable structural and textural conditions.

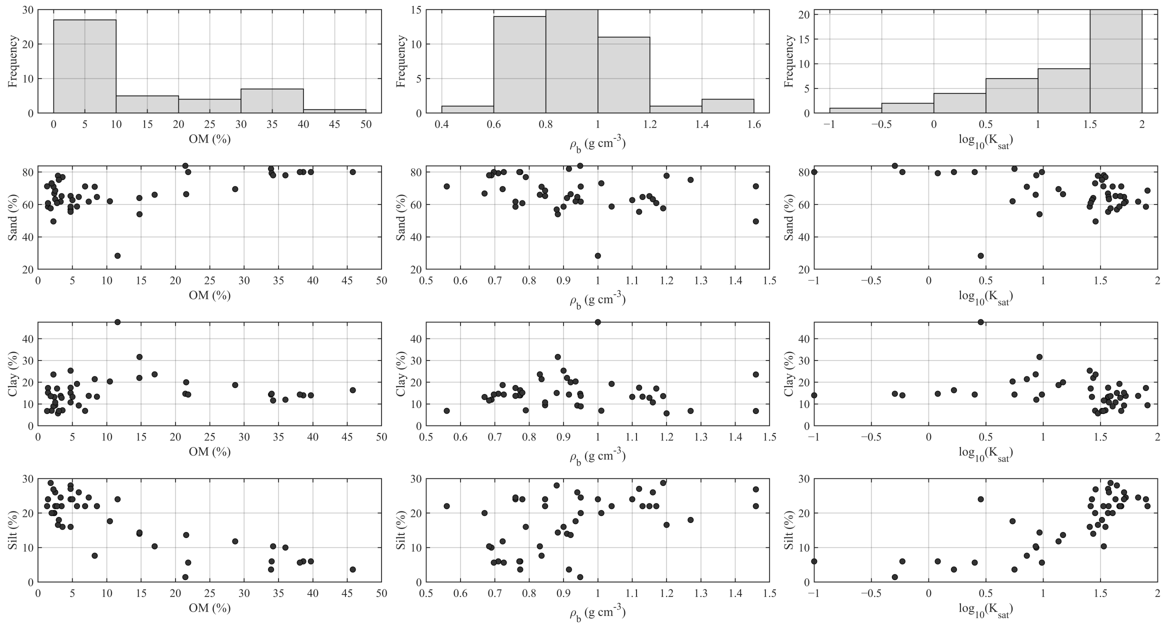

Figure 3. Exploratory histograms and bivariate scatterplots of soil predictors (OM, ρ_b, and $\log_{10}(K_{sat})$) against particle-size fractions (sand, clay, and silt). Histograms illustrate the distribution of predictor variables, while scatterplots highlight their relationships with soil texture components.

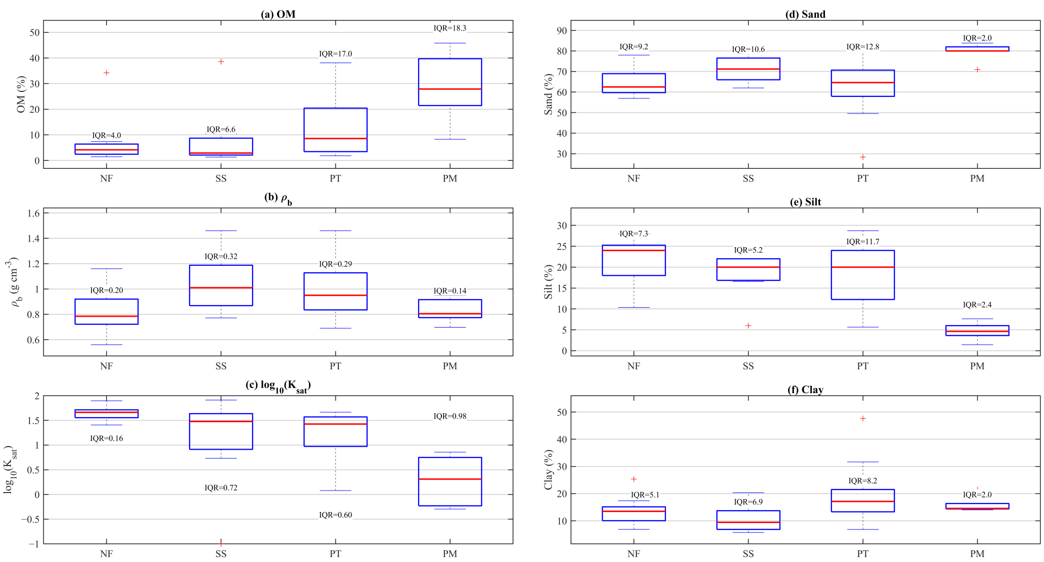

Figure 4. Boxplot distributions of hydro-structural variables and particle-size fractions across land-use classes. Panels show (**a**) OM, (**b**) ρ_b, (**c**) $\log_{10}(K_{\text{sat}})$, (**d**) sand, (**e**) silt, and (**f**) clay. Red lines indicate medians, blue boxes represent the interquartile range (IQR), whiskers indicate non-outlier ranges, and red crosses denote outliers. IQR values are shown above each box and are expressed in the units of the corresponding variable. NF = native forest; SS = shrubland/secondary forest; PT = pasture; PM = paramo.

Multivariate Compositional Association with Land-Use Class

Although exploratory analyses indicate that LULC is associated with systematic differences in soil structural properties, the particle-size fractions remain widely overlapped among LULC classes. Because sand, clay, and silt constitute a compositional system constrained to sum to a constant, their relationships cannot be properly evaluated through univariate comparisons [83]. Therefore, the association between soil texture and land-use class was examined within a compositional framework using PERMANOVA under Aitchison geometry, while the ternary distribution of sand, clay, and silt fractions together with the centroid of each land-use class was used to visualize their multivariate arrangement (Figure 5).

The ternary diagram shows a wide dispersion of samples within the sandy-loam and loamy-sand domains across all land-use classes. Native forest, paramo, shrubland, and pasture occupy overlapping regions of the ternary space, and their centroids remain close to each other.

PERMANOVA indicated a significant multivariate association between land-use class and soil texture composition (Pseudo-F = 11.36, R^2 = 0.46, p = 0.0001). This result indicates that LULC partitions part of the distance-based compositional variability within the sand–silt–clay system in this dataset, but it should not be interpreted as evidence of clear textural discrimination or predictive capacity at the land-use-class level (Table 1).

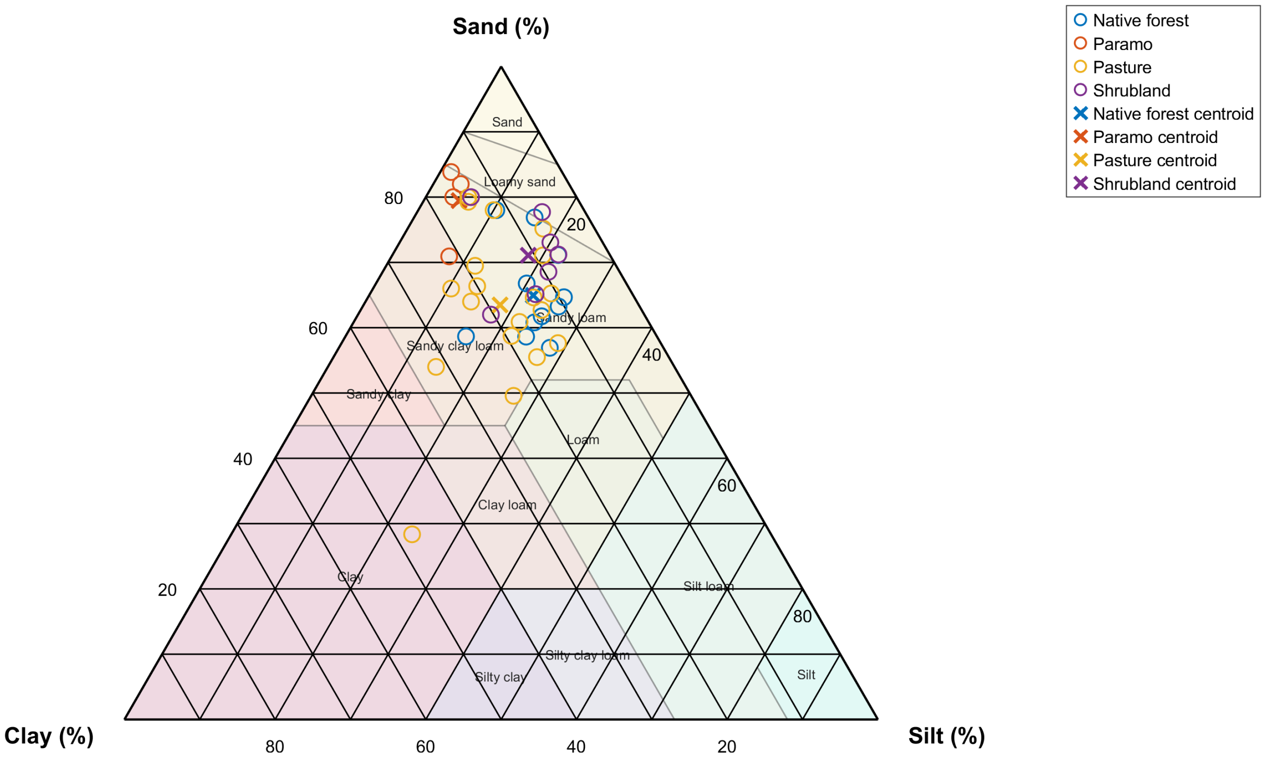

Figure 5. Ternary distribution of soil texture fractions (sand, clay, and silt) across LULC classes. Points represent individual samples and crosses indicate group centroids. Gray background polygons correspond to the main USDA textural classes, included as visual reference for the location of the observations within the soil texture triangle.

Table 1. Global PERMANOVA results for the compositional relationship between soil texture fractions and land-use class using Aitchison geometry.

Pseudo-F	R^2	p-Value	$df_{between}$	df_{within}	$SS_{between}$	SS_{within}
11.358	0.46	0.0001	3	40	11.095	13.025

Pairwise PERMANOVA comparisons revealed that the main compositional differences were associated with the samples classified as paramo. This class differed significantly from native forest, pasture, and shrubland after Holm correction (Table 2). In contrast, the comparisons among native forest, pasture, and shrubland were not statistically significant after adjustment, indicating substantial compositional overlap among these three LULC types.

Table 2. Pairwise PERMANOVA comparisons among land-use classes based on clr-transformed texture compositions. Holm correction was applied to control the family-wise error rate.

Group Comparison	N_1	N_2	Pseudo-F	R^2	p_{adj}
Native forest vs. paramo	12	6	44.614	0.736	0.0012
Native forest vs. Pasture	12	19	3.268	0.101	0.1206
Native forest vs. Shrubland	12	7	1.224	0.067	0.3168
paramo vs. Pasture	6	19	17.876	0.437	0.0020
paramo vs. Shrubland	6	7	19.972	0.645	0.0080
Pasture vs. Shrubland	19	7	3.055	0.113	0.1206

The ternary diagram (Figure 5) illustrates that most observations are concentrated within the sandy-loam and loamy-sand domains of the soil texture triangle. Native forest, pasture, and shrubland samples occupy a largely overlapping compositional region characterized by sand-dominated textures with moderate clay content. In contrast, samples classified as paramo are displaced toward the sand apex, consistent with the strong pairwise differences detected by PERMANOVA.

Group dispersion metrics (Table 3) indicate that pasture and shrubland exhibit greater internal heterogeneity in texture composition compared with native forest and paramo. Mean distances to centroid were highest in pasture (0.568) and shrubland (0.508), whereas native forest (0.354) and paramo (0.379) showed more compact compositional clusters.

Table 3. Within-group multivariate dispersion based on distances to group centroid in clr-transformed compositional space.

Group	N	Mean Distance	Median Distance
Native forest	12	0.354	0.303
paramo	6	0.379	0.328
Pasture	19	0.568	0.537
Shrubland	7	0.508	0.461

These results indicate that the significant LULC–texture association mainly reflects the distinct position of the paramo class, whereas native forest, pasture, and shrubland share an overlapping sandy to sandy-loam domain. Thus, texture provides a background edaphic condition in the studied catchments, supporting the subsequent focus on OM, ρ_b, and K_{sat} as hydro-structural indicators of soil variability observed among LULC groups.

The broad overlap of particle-size fractions among LULC classes, together with clearer contrasts in hydro-structural variables, motivated a correlation analysis between texture fractions and OM, ρ_b, and $\log_{10}(K_{\text{sat}})$. Pearson and Spearman coefficients were calculated using pairwise deletion to reduce data loss from missing values. Correlations were evaluated for the complete dataset and by anthropogenic activity level, with low activity including native forest and paramo, and high activity including pasture and shrubland.

Figure 6 presents the resulting Pearson and Spearman correlation matrices for the complete dataset and for each anthropogenic activity group. The figure summarizes the pairwise associations among textural fractions and structural soil variables, allowing comparison of the correlation structure under different disturbance conditions.

https://doi.org/10.3390/soilsystems10070068

Figure 6 shows that the overall correlation structure is broadly preserved across anthropogenic activity levels. The most stable pattern is the positive association of silt with $\log_{10}(K_{sat})$ and ρ_b, while sand remains positively related to OM. Stratification does not indicate a different correlation framework, but rather a shift in the intensity of specific linkages: under low anthropogenic activity, the sand–OM relationship is more evident, whereas under high anthropogenic activity, the associations between silt, $\log_{10}(K_{sat})$, and ρ_b are strengthened.

3.2. Multivariate Structure of Hydro-Edaphic Variables

Because the previous analyses showed broad overlap of textural fractions across LULC classes but clearer contrasts in hydro-structural variables, PCA was applied as an exploratory dimensionality reduction approach to isolate the intrinsic covariance structure among hydro-edaphic and textural variables and to identify the dominant multivariate gradients supporting subsequent predictive modeling.

The PCA was performed on the standardized hydro-edaphic dataset including OM, ρ_b, $\log_{10}(K_{sat})$, porosity and the particle-size fractions. In this framework, the first principal component (PC1) represents the linear combination of variables explaining the largest portion of total variance, while the second component (PC2) explains the largest share of the remaining variance under orthogonality constraints.

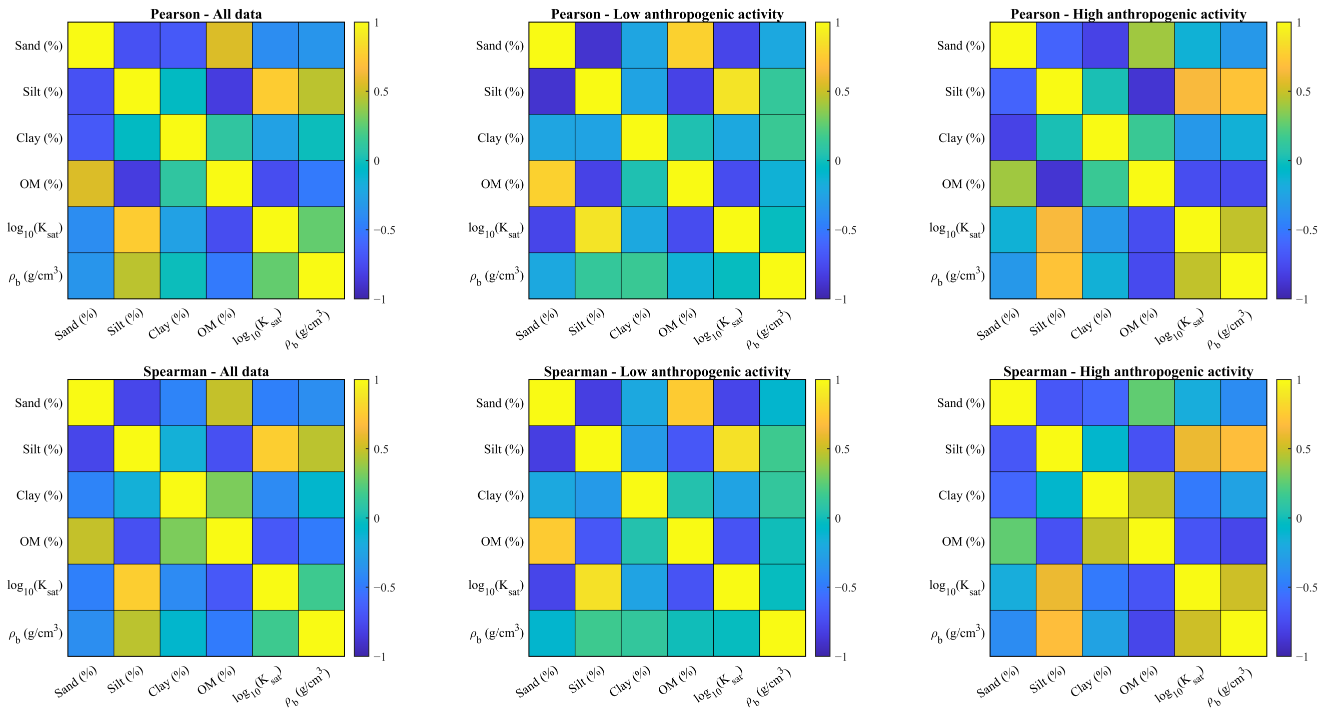

Figure 6. Correlation matrices between soil texture fractions and structural soil properties. Pearson and Spearman coefficients are shown for the complete dataset and for subsets representing low and high anthropogenic activity levels.

Figure 7 summarizes the PCA results through the scree plot, the explained variance diagram, the variable loading biplot, and the distribution of sample scores by LULC class.

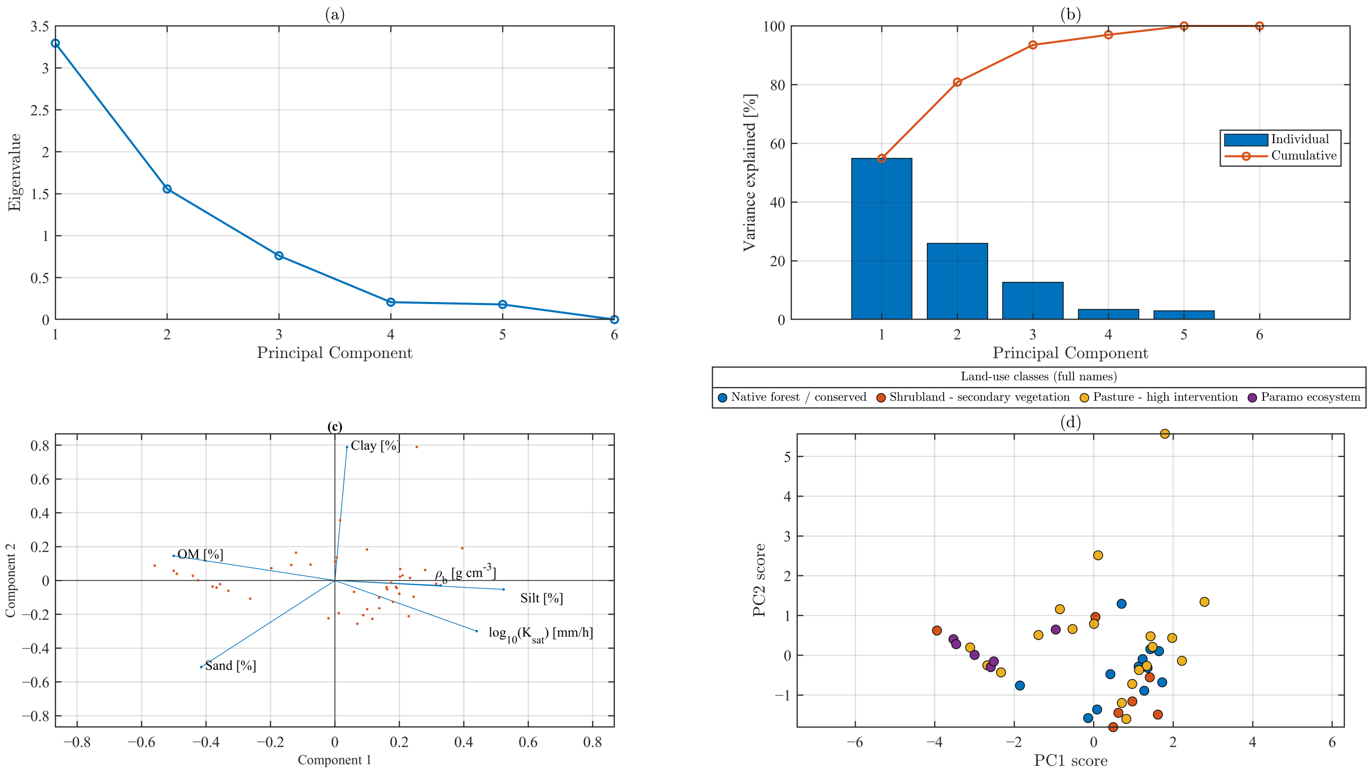

Figure 7. Principal component analysis of the hydro-edaphic dataset. Panel (**a**) shows the scree plot, panel (**b**) the individual and cumulative explained variance, panel (**c**) the PCA biplot for PC1 and PC2, and panel (**d**) the PCA scores by land-use class.

Figure 7 illustrates the multivariate structure of the hydro-edaphic dataset. The scree plot and the explained-variance diagram indicate that the first two components capture most of the variability of the system, allowing a substantial reduction of dimensionality. The biplot shows that PC1 is primarily driven by the joint influence of OM, ρ_b, $\log_{10}(K_{sat})$, and the redistribution between sand and silt fractions, whereas PC2 is mainly associated with variation in clay relative to the other texture fractions. The score distribution reveals considerable overlap among land-use classes in the reduced space, indicating that the dominant gradients of variability are governed by intrinsic soil properties rather than by categorical land-use separation. The exact standardized expressions defining the principal components are provided in Equations (1) and (2).

$$
\begin{aligned}
PC1_{\text{texture}} = & -0.508112\left(\frac{\text{OM}-16.264061}{20.448468}\right)+0.340710\left(\frac{\rho_b-0.930916}{0.209802}\right) \\
& +0.428000\left(\frac{\log_{10}(K_{\text{sat}})-1.259236}{0.602403}\right)-0.407728\left(\frac{\text{Sand}-66.582381}{10.346902}\right) \\
& +0.048265\left(\frac{\text{Clay}-15.443905}{7.607505}\right)+0.523434\left(\frac{\text{Silt}-17.973714}{7.358224}\right)
\end{aligned} \tag{1}
$$

For the structural dataset, the exact standardized expression for the first component is

$$
\begin{aligned}
PC1_{\text{structural}} = & +0.009904\left(\frac{\text{OM}-33.513073}{20.833354}\right)-0.563585\left(\frac{\rho_b-0.787419}{0.186373}\right) \\
& +0.587860\left(\frac{\log_{10}(K_{\text{sat}})-0.929308}{0.679974}\right)+0.580253\left(\frac{\phi-52.451849}{12.696891}\right)
\end{aligned} \tag{2}
$$

3.3. Cross-Validated Performance of Surrogate Texture Models

Because PCA showed that hydro-structural variables retained part of the covariance structure linked to particle-size redistribution, surrogate texture models were evaluated under LOOCV to test whether this multivariate signal translated into reproducible fraction-wise predictions. Figure 8 presents the cross-validated R^2_{LOOCV} obtained for direct fraction-wise models, ILR-based models, AIC- and BIC-reduced structures, and sensitivity models fitted without OM. Model performance was strongly fraction-dependent: silt was the most consistently predictable fraction, whereas sand and clay showed weaker and more structure-dependent validation responses.

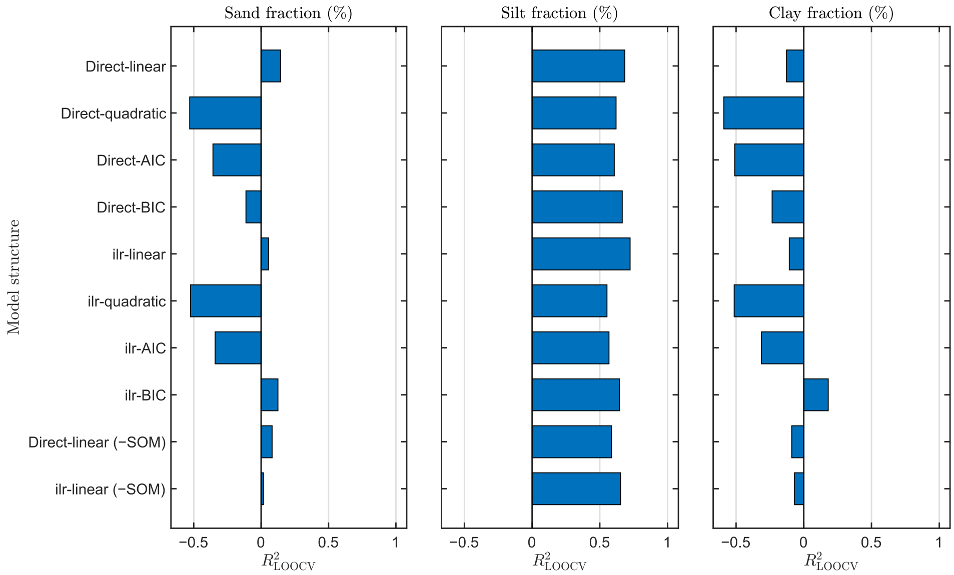

Figure 8. Cross-validated predictive performance of the surrogate texture models for sand, silt, and clay fractions. Bars represent R^2_{LOOCV} obtained under direct fraction-wise models, ILR-based models, AIC- and BIC-reduced structures, and sensitivity models excluding OM.

The BIC-reduced ILR model reached positive validation performance for sand, silt, and clay, with R^2_{LOOCV} values of 0.147, 0.704, and 0.124, respectively. Although silt reached slightly higher values under other model structures, the BIC-reduced ILR formulation provided the most balanced behaviour across the complete sand–silt–clay composition. For sand and clay, several models yielded negative R^2_{LOOCV} values. The sensitivity models excluding OM retained part of the predictive signal, especially for silt, where the decrease in R^2_{LOOCV} was moderate. The similar performance of the models with and without OM suggests that potential methodological differences in OM determination did not dominate the predictive structure.

3.4. Compositional Closure and Residual Diagnostics

After the cross-validated comparison, the BIC-reduced ILR structure was refitted with the full dataset to derive the final closure-preserving surrogate equations. The selected structure was parsimonious for ilr_1, retaining only OM, whereas ilr_2 retained OM, ρ_b, $\log_{10}(K_{\mathrm{sat}})$, and two interaction terms involving OM. Using standardized predictors, the fitted equations were

$$\mathrm{ilr}_1 = 1.250606 + 0.259304\,\mathrm{OM}_z \tag{3}$$

$$\begin{aligned}\mathrm{ilr}_2 = {} & 0.286104 + 0.557114\,\mathrm{OM}_z + 0.180562\,\rho_{b,z} - 0.424326\,\log_{10}(K_{\mathrm{sat}})_z \\ & + 0.368930\,(\mathrm{OM}_z\rho_{b,z}) + 0.199442\,(\mathrm{OM}_z\log_{10}(K_{\mathrm{sat}})_z)\end{aligned} \tag{4}$$

The predicted ILR coordinates were back-transformed to texture fractions through

$$A_{\text{sand}} = \exp\left(\sqrt{\frac{2}{3}}\,\text{ilr}_1\right) \tag{5a}$$

$$A_{\text{clay}} = \exp\left(-\sqrt{\frac{1}{6}}\,\text{ilr}_1 + \sqrt{\frac{1}{2}}\,\text{ilr}_2\right) \tag{5b}$$

$$A_{\text{silt}} = \exp\left(-\sqrt{\frac{1}{6}}\,\text{ilr}_1 - \sqrt{\frac{1}{2}}\,\text{ilr}_2\right) \tag{5c}$$

The final texture fractions were recovered by closure normalization:

$$\text{Sand} = 100\,\frac{A_{\text{sand}}}{A_{\text{sand}} + A_{\text{clay}} + A_{\text{silt}}} \tag{6a}$$

$$\text{Clay} = 100\,\frac{A_{\text{clay}}}{A_{\text{sand}} + A_{\text{clay}} + A_{\text{silt}}} \tag{6b}$$

$$\text{Silt} = 100\,\frac{A_{\text{silt}}}{A_{\text{sand}} + A_{\text{clay}} + A_{\text{silt}}} \tag{6c}$$

The standardized predictors were computed using the observed means and standard deviations of the analytical dataset: OM (18.067 ± 22.508), ρ_b (0.928 ± 0.206), and $\log_{10}(K_{\text{sat}})$ (1.200 ± 0.651). The signs and retained terms indicate that OM contributed to both ILR coordinates, while $\log_{10}(K_{\text{sat}})$ mainly affected ilr_2 through a negative main effect and a positive interaction with OM. Figure 9 compares observed and predicted sand, silt, and clay fractions from the BIC-reduced ILR surrogate model, including bootstrap-based 95% uncertainty intervals and Q–Q plots of standardized residuals.

The observed–predicted comparison showed that the model reproduced the central tendency of the texture fractions but with different levels of dispersion among components (Figure 9). Silt showed the clearest agreement with the 1:1 reference line, whereas sand and clay presented broader scatter, particularly at the extremes of their observed ranges.

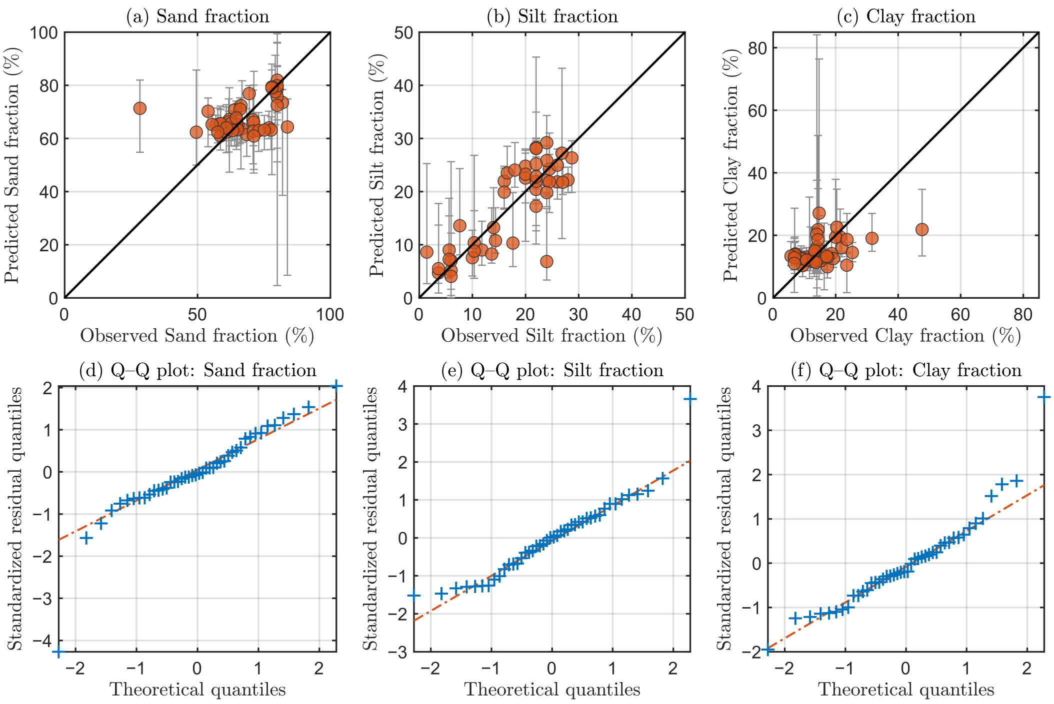

Figure 9. Observed and predicted texture fractions obtained with the BIC-reduced ILR surrogate model. The upper panels compare observed and predicted sand, silt, and clay fractions; vertical bars indicate bootstrap-based 95% uncertainty intervals after inverse ILR transformation. The lower panels show Q–Q plots of standardized residuals for each fraction.

The Q–Q plots showed that standardized residuals followed the theoretical quantiles more closely in the central range, especially for silt, whereas sand and clay showed stronger departures at the tails. The bootstrap-based uncertainty intervals followed the same fraction-dependent pattern: silt displayed narrower intervals and closer agreement with the 1:1 reference line, while sand and clay showed wider uncertainty bands near the extremes of their observed ranges. These diagnostics indicate that the BIC-reduced ILR model captured the central compositional tendency more reliably than extreme sand- and clay-rich observations. Tables 4 and 5 report the fitted and nested LOOCV performance metrics for each recovered texture fraction.

Table 4. Goodness-of-fit metrics for the final BIC-reduced ILR surrogate model fitted with the full dataset.

Fraction	R^2_{fit}	$RMSE_{\text{fit}}$	MAE_{fit}	Bias_{fit}
Sand	0.262	9.015	6.275	0.675
Silt	0.744	3.948	2.968	−0.076
Clay	0.264	6.300	4.468	−0.600

Note: $RMSE$, MAE, and bias are expressed in percentage points of the corresponding texture fraction.

Table 5. Nested LOOCV validation metrics for the BIC-reduced ILR surrogate model.

Fraction	R^2_{LOOCV}	$RMSE_{\text{LOOCV}}$	MAE_{LOOCV}	$\text{Bias}_{\text{LOOCV}}$
Sand	0.147	9.694	6.607	0.729
Silt	0.704	4.243	3.173	−0.113
Clay	0.124	6.874	4.904	−0.616

Note: LOOCV metrics were computed using nested BIC-based term selection within each training subset. $RMSE$, MAE, and bias are expressed in percentage points.

The goodness-of-fit metrics confirmed this fraction-dependent behaviour. For the model fitted with all data, R^2_{fit} reached 0.262 for sand, 0.744 for silt, and 0.264 for clay. The corresponding $RMSE$ values were 9.015, 3.948, and 6.300 percentage points, respectively, with low mean bias values for the three fractions. Under nested LOOCV, the same structure maintained positive validation performance for all fractions, with R^2_{LOOCV} values of 0.147 for sand, 0.704 for silt, and 0.124 for clay. The validation errors were higher than the fitted errors, as expected, but the model preserved the same hierarchy of predictive performance, with silt as the most stable fraction and sand and clay as the least constrained components.

The inverse ILR transformation produced closed sand–clay–silt compositions, with predicted sums equal to 100% and negligible closure residuals for both the fitted model and nested LOOCV predictions. Figure 10 places the observed and inverse-transformed ILR-predicted compositions in the USDA textural triangle to verify whether the surrogate model preserved physically valid sand–clay–silt combinations.

The predicted compositions remained within the USDA textural triangle and were mainly concentrated in the sandy-loam and loamy-sand domains. This pattern confirms that the inverse ILR transformation preserved the 100% closure condition and generated valid textural compositions. However, the observed samples covered a broader textural range, particularly toward sandy clay loam and clay-rich conditions.

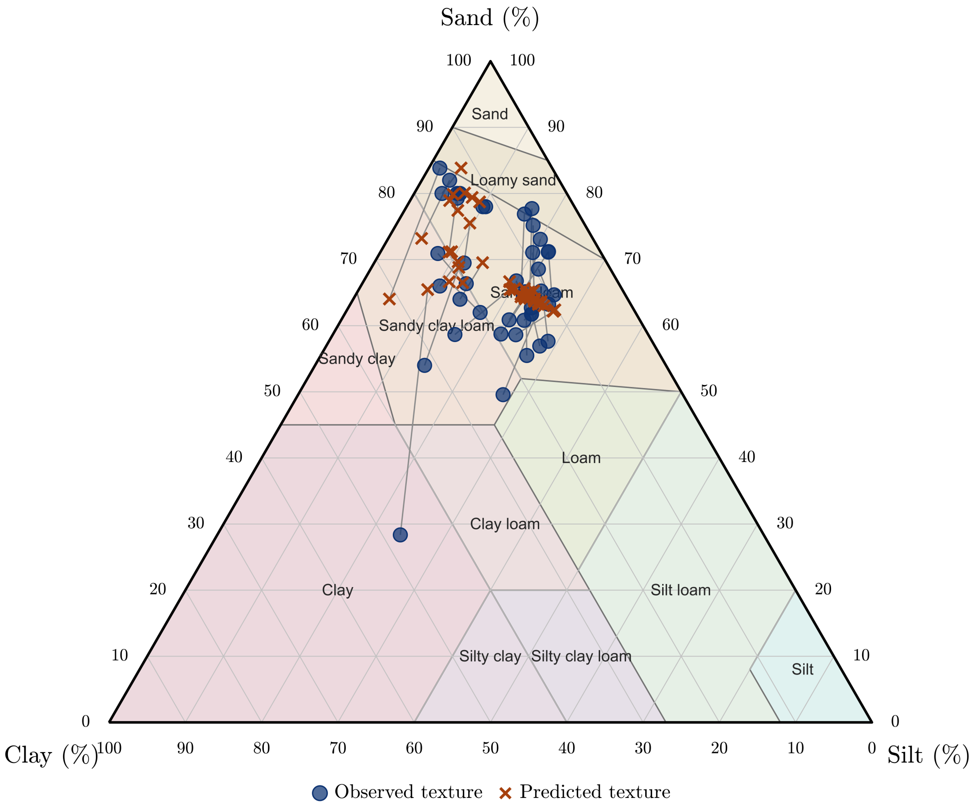

Figure 10. Observed and predicted soil texture fractions represented in the USDA textural triangle. Blue circles correspond to observed sand–clay–silt compositions, while orange crosses represent the fractions recovered from the BIC-reduced ILR surrogate model after inverse transformation.

4. Discussion

4.1. Land-Use Disturbance, Compositional Overlap, and the Edaphic Singularity of Paramo Soils

The hydro-structural expression of the soil system was altered by LULC disturbance, whereas particle-size composition remained broadly conserved across LULC classes. This overlap is consistent with the view that land-cover effects are often expressed more clearly through soil aggregation and hydro-structural attributes than through the sand–silt–clay triad alone [84]. Structurally, ρ_b increased under disturbed covers, whereas native forest maintained higher median $\log_{10}(K_{\mathrm{sat}})$, suggesting compaction, pore-network reorganization, and altered hydraulic functioning, as also reported by [23,31]. Therefore, this distinction is central to the surrogate modeling framework because LULC can modify the hydro-structural predictors used to approximate part of the sand–clay–silt signal.

The paramo class, in contrast, produced the strongest compositional contrast and behaved as a distinct hydro-edaphic domain within the dataset. This singular behavior is consistent with high-Andean volcanic and organic-rich soils, where a cold and wet climate, volcanic ash inputs, high accumulation of OM, open porous structure and strong water-retention capacity create soil conditions that differ structurally and functionally from lower-elevation Andean landscapes [43,85,86].

Studies in southern Ecuador indicate that the regional variability of volcanic ash soils is mainly related to parent material and climate, while clay mineralogy reflects local substrate conditions [30,86]. As reported by [23,31,87,88], grazing, afforestation, cultivation, and vegetation-cover change mainly affect disturbance-sensitive hydro-structural properties, including OM, ρ_b, ϕ, water retention, and K_{sat}. The predictive signal arises from the extent to which hydro-structural variables retain covariance with a soil compositional template controlled by parent material, pedogenesis, and geomorphic setting, as suggested by [89]. Similarly, refs. [90,91] showed that LULC mainly affects OM, ρ_b, and related soil structural properties, rather than directly modifying the mineral texture. Therefore, LULC may affect the surrogate signal indirectly through OM, ρ_b, and K_{sat}, but it should not be interpreted as a direct driver of sand–clay–silt composition.

For a local distributed hydrological model, this distinction is relevant because LULC can support the spatial stratification of hydro-structural parameters such as OM, ρ_b, and K_{sat}, whereas paramo should be treated as a differentiated hydro-edaphic unit rather than as a simple land-cover category.

4.2. Multivariate Hydro-Edaphic Covariance and Fraction-Dependent Predictability

The correlation matrices and PCA indicate that the response was structured by coupled changes among texture, soil structure, and hydraulic behavior rather than by isolated univariate trends. Under disturbed conditions, stronger associations emerged between silt, ρ_b, and $\log_{10}(K_{sat})$, while the main multivariate gradient combined OM, ρ_b, hydraulic response, and the redistribution between sand and silt. Similar patterns have been reported in hydro-pedological studies, where K_{sat} depends not only on texture, but also on pore continuity, aggregation, ϕ, and OM dynamics [22,92]. Structural disturbance has also been shown to modify the relationship between K_{sat} and basic soil descriptors, indicating that hydraulic behavior cannot be explained from particle-size composition alone [27,68]. In Andean environments, LULC change has been reported to alter ρ_b, K_{sat}, and soil water regulation more clearly than the mineral fraction itself [23,31].

Domain-Specific Predictability of the Silt Fraction

The better fit obtained for silt in this study should be interpreted as domain-specific rather than as a general rule. Recent reviews and comparative studies showed that the predictive performance of textural fractions depends mainly on scale, calibration domain, covariate type, and harmonization procedure, not on an intrinsic advantage of a given fraction [10,83,93]. In our dataset, the main PCA gradient combined OM, ρ_b, K_{sat}, and the redistribution between sand and silt, placing silt in the core hydro-edaphic covariance structure. Sand and clay, by contrast, likely retained a larger share of variability linked to parent material, local depositional heterogeneity, and compositional closure. Methodological effects must also be considered, since textural estimates are sensitive to analytical protocol, fraction definition, and class conversion; under a harmonized dataset within a relatively restricted domain, the apparent noise of the silt fraction may decrease substantially [10,18,93].

4.3. Model Selection, ILR Closure, and Residual Behaviour

Increasing algebraic complexity did not necessarily improve internal predictive robustness. As discussed by [94], validation estimates may become optimistic when model selection and performance assessment are not properly separated, while nested procedures provide a more reliable evaluation of the full modeling workflow. Similarly, ref. [21] showed that soil hydraulic models with fewer covariates can transfer better than more complex alternatives, because additional predictors may introduce uncertainty or unstable information. In this study, AIC- and BIC-reduced formulations helped control complexity and avoid retaining higher-order terms with apparent but inconsistent validation support.

The BIC-reduced ILR formulation provided the most defensible compromise between parsimony, validation performance, and compositional consistency. Although silt reached slightly higher R^2_{LOOCV} values under other structures, this was the only model with positive nested LOOCV performance for sand, silt, and clay simultaneously. This result is consistent with [95], who emphasized that soil particle-size fractions are compositional data constrained by a constant sum and require transformations, including ILR, when modeled jointly. Thus, inverse transformation avoided independent fraction-wise predictions and recovered closed compositions constrained to 100%.

The sensitivity analysis excluding OM showed that the predictive signal was not controlled by OM alone. Models without OM retained part of the validation structure, especially for silt, suggesting that methodological differences in OM determination did not dominate the surrogate relationships. This interpretation is consistent with the transferability analysis of [21], who showed that covariate number and measurement uncertainty can affect the generalization of PTFs. However, the final BIC-reduced ILR equations retained OM in both coordinates, indicating that it remained relevant within the hydro-edaphic covariance structure.

Residual behaviour confirmed the fraction-dependent limitations of the surrogate approach. The Q–Q plots showed closer agreement with theoretical quantiles in the central range, especially for silt, whereas sand and clay showed stronger tail departures, indicating errors associated with extreme textural observations. A comparable sensitivity to particle-size variation was modeled by [96], who showed that PTF-derived water-retention parameters and crop-model outputs can vary substantially with changes in soil texture. The USDA textural triangle supported the same interpretation: predictions remained compositionally valid, but were compressed toward the sandy-loam and loamy-sand domains, underestimating part of the observed spread in sand- and clay-rich samples.

The BIC-reduced ILR model should therefore be interpreted as a local, compositionally coherent surrogate formulation rather than as a substitute for direct particle-size analysis. Its strength lies in preserving closure, reducing overparameterization, and providing internally validated equations for exploratory hydrological parameterization. Its main limitation is the reduced ability to reproduce the full compositional variability, particularly at the tails of the sand and clay distributions.

4.4. Applicability Domain and Implications for Andean Hydrological Parameterization

Hydro-pedotransfer functions are relevant for hydrological modeling because they convert available soil information into hydraulic properties required by physically based and distributed models, particularly retention and conductivity functions controlling infiltration, storage, lateral flow, and runoff generation [22]. Their role becomes more critical in mountain regions, where direct hydraulic measurements are scarce and model performance is highly sensitive to soil parameterization [3,22,97]. As shown by [3,97,98], the choice of PTF is not neutral, since different formulations can modify runoff partitioning, water-balance components, infiltration response, and the spatial organization of simulated hydrological processes. In particular, [98] showed that infiltration modeling can be sensitive to PTF choice and initial moisture conditions.

The Irquis hydro-structural dataset shows the magnitude of this parameterization problem in the study area. Irquis soils retained higher OM and lower ρ_b than ZH soils, with mean values of $38.76 \pm 22.55\%$ and 0.818 ± 0.103 g cm^{-3}, respectively, compared with $3.74 \pm 1.95\%$ and 1.004 ± 0.226 g cm^{-3} in ZH. These contrasts indicate that the upper horizons analyzed here are not only textural units, but hydro-structural domains where OM accumulation, low bulk density, and variable hydraulic conductivity regulate subsurface response.

This interpretation is consistent with [9], who reported volumetric soil water contents of 0.80–0.90 cm^3 cm^{-3}, equivalent to 80–90% water by soil volume, during wet periods in southern Ecuadorian páramo Andosols, with wilting points close to 0.40 cm^3 cm^{-3}. During drought events, soil moisture decreased from approximately 0.84 to 0.60 cm^3 cm^{-3} and recovered within 2–3 months, whereas lower-altitude mineral soils decreased from 0.54 to 0.39 cm^3 cm^{-3}.

https://doi.org/10.3390/soilsystems10070068

Direct particle-size analysis is operationally difficult in these soils. As reported by [34] for Quimsacocha Andosols and Histosols, bulk density reached very low values, with averages of 0.44 g cm^{-3} in Andosol A horizons and 0.10 g cm^{-3} in Histosol H horizons, while OM reached 26.50% and 56.16%, respectively. The same study reported K_s values of 1.04 cm h^{-1} for Andosol A horizons and 0.39 cm h^{-1} for Histosol H horizons, together with high sand contents of 79.16% and 83.50%. These values illustrate why conventional texture determination can be time-consuming and analytically sensitive in organic-rich volcanic soils: mineral–organic associations, organo-metallic complexes, and pseudo-structures may affect the apparent sand fraction, while organic matter removal or oxidation increases laboratory effort and uncertainty. This problem is consistent with [38], who showed that organic matter removal can substantially alter clay and silt estimates, especially in soils with high organic carbon.

The proposed equations address this need by estimating sand–clay–silt composition from measured OM, ρ_b, and K_{sat}, while preserving the 100% compositional closure through the ILR framework. In the analytical dataset, the standardized predictors were computed from $\text{OM} = 18.067 \pm 22.508\%$, $\rho_b = 0.928 \pm 0.206$ g cm^{-3}, and $\log_{10}(K_{\text{sat}}) = 1.200 \pm 0.651$. This is relevant for distributed hydrological and sediment-export models because texture is commonly required to define infiltration, water-retention, storage, erodibility, detachment, and transport parameters, while reducing reliance on generic PTFs beyond their validated domain when spatially explicit textural information is sparse in high-Andean basins. As discussed by [99], recent hydro-pedotransfer developments increasingly rely on readily available predictors such as bulk density, organic matter, and texture to estimate soil hydraulic properties under uncertainty. In this study, the inverse direction was explored: hydro-structural properties were used to recover a local approximation of the textural composition needed for subsequent hydrological and sediment-related parameterization.

The results define a restricted applicability domain. The BIC-reduced ILR model retained positive nested LOOCV performance for the three fractions, with R^2_{LOOCV} values of 0.147 for sand, 0.704 for silt, and 0.124 for clay. The fitted model showed the same hierarchy, with R^2_{fit} values of 0.262, 0.744, and 0.264, respectively, and $RMSE_{\text{fit}}$ values of 9.015, 3.948, and 6.300 percentage points. Therefore, the model is useful for first-order, locally grounded parameterization, sensitivity analysis, and uncertainty stratification, but it should not be used as a substitute for direct particle-size analysis or as a universal PTF. Its main contribution is to provide interpretable hydro-pedotransfer support for upper soil horizons, where OM, ρ_b, and K_{sat} regulate subsurface hydrological response but do not fully recover the mineral compositional variability.

The Quimsacocha–Irquis area gives this issue additional relevance because the hydrological regulation provided by páramo soils overlaps with territorial water-resource concerns and the persistent water–mining conflict linked to the Loma Larga project [100]. Under these conditions, locally calibrated surrogate relationships can strengthen the technical basis for testing how different soil-parameter assumptions affect recharge, streamflow buffering, runoff generation, and sediment-export response under disturbance. Their application must remain conditional on independent validation, additional sampling, and transparent uncertainty analysis, since the current model preserved compositional closure but still compressed the observed variability toward sandy-loam and loamy-sand domains and showed larger uncertainty for sand- and clay-rich observations [22,97].

At the regional scale, the contrasting parent materials of the two basins help explain the limited transferability of texture-based prediction. ZH is influenced by the Paleozoic metamorphic basement of the Chigüinda Formation and by Miocene sedimentary infill of the Loja Basin, favouring acidic, weathered clay–silt matrices. In contrast, Irquis is developed on younger volcanic materials associated with the Tarqui Formation and the

https://doi.org/10.3390/soilsystems10070068

Quimsacocha volcanic domain, where pyroclastic deposits, volcanic ash, and short-range-order minerals promote Andisols with strong organo-mineral aggregation [30,34,101]. This contrast may explain why sand and clay were less robustly predicted than silt: incomplete dispersion can hide clay within microaggregates, while pseudo-sand structures can inflate the apparent sand fraction. Therefore, the stronger silt performance should be interpreted as a local response of highly weathered and structured Andean soils, not as a universal rule. This supports the need for Andean-specific PTFs that include aggregation-sensitive descriptors, rather than relying only on the classical sand–silt–clay triad [4].

4.5. Methodological Limitations

Several limitations should be considered when interpreting the proposed surrogate relationships. First, the analytical dataset was limited to the available harmonized observations from the ZH and Irquis basins, restricting the representation of extreme textural conditions and reducing the capacity to evaluate model behaviour beyond the observed hydro-edaphic domain [20,21]. Second, both basins were integrated from independent data sources, and OM was obtained using different analytical procedures: loss on ignition in Irquis and Walkley–Black wet oxidation in ZH. Although OM was retained as a common edaphological variable and the sensitivity analysis excluding OM suggested that this difference did not dominate the predictive structure, some method-related uncertainty may remain in the harmonized database [18,46,47]. Third, model performance was assessed through nested LOOCV, providing internal validation but not an independent external validation dataset, because no external dataset with the same variables, near-surface layer, and comparable analytical traceability was available. Consequently, the transferability of the equations to other Andean environments remains uncertain [21,73,74,79].

An additional limitation arises from the nature of the studied soils. In volcanic and organic-rich horizons, aggregation processes, organo-mineral associations, and incomplete particle dispersion may influence laboratory estimates of sand and clay fractions. Organic matter removal, dispersion efficiency, and sedimentation assumptions can also modify recovered particle-size fractions, especially in organic-rich or volcanic soils [36–40]. Such effects may contribute to part of the residual variability observed in the surrogate models, particularly at the tails of the sand and clay distributions. The stronger performance obtained for silt suggests that some hydro-structural signals are more stable than others within the analyzed domain, although this behaviour should not be generalized beyond the conditions represented in the present dataset [83,93,95].

Finally, the study was designed as an exploratory cross-domain assessment of hydro-structural and compositional relationships rather than as the development of a universally transferable pedotransfer function. Future work should expand the geographical coverage, incorporate additional parent materials and soil environments, and evaluate the proposed equations using independent datasets in order to better define their applicability domain and predictive uncertainty [21,22,97].

5. Conclusions

The results indicate that LULC disturbance mainly modified the hydro-structural expression of the soil system through OM, ρ_b, and K_{sat}, whereas sand–silt–clay composition remained broadly conserved across most LULC classes. The paramo class behaved as a distinct hydro-edaphic domain and should therefore be treated separately in Andean hydrological parameterization.

The correlation structure and PCA showed that the dominant response was not controlled by isolated bivariate trends, but by a coupled hydro-edaphic gradient involving OM, ρ_b, K_{sat}, and the redistribution between sand and silt. This confirms that hydro-structural predictors retain part of the textural signal, although this signal is indirect and conditioned by soil-forming controls.

Silt was the most stable and predictable fraction within the studied domain, with $R^2_{\mathrm{LOOCV}} = 0.704$ under the BIC-reduced ILR model. This behaviour should be interpreted as domain-specific, since sand and clay retained greater residual uncertainty associated with local parent material, aggregation, dispersion limitations, and compositional closure.

The BIC-reduced ILR formulation provided the most defensible surrogate model because it combined parsimony, positive nested LOOCV performance for the three fractions, and exact 100% compositional closure. The residual diagnostics and USDA triangle showed that the model reproduced the central compositional tendency but compressed the observed variability, especially for sand- and clay-rich samples.

The proposed equations are useful as local hydro-pedotransfer support for exploratory hydrological parameterization, sensitivity analysis, and uncertainty stratification in data-scarce Andean catchments. They should not be used as a substitute for direct particle-size analysis or as a universal PTF, and regional transfer requires external validation, additional sampling, and aggregation-sensitive descriptors.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/soilsystems10070068/s1, Figure S1: Evolution of the Irquis sampling network by phase; Figure S2: Diagnostic interpolation maps of K_{sat} used to guide sequential sampling expansion; Figure S3: Reclassified diagnostic surfaces of K_{sat} and OM used to identify spatial gaps and transition zones; Figure S4: Final distribution of the 79 Irquis georeferenced points used for spatial analysis of OM, K_{sat}, ρ_b, particle density (ρ_s), and porosity (ϕ); Figure S5: Distribution of the 43 original ZH field records and the 26 observations retained for the study; Table S1: Structure of the soil databases used in the harmonized analysis; Table S2: Sequential construction of the Irquis sampling network; Table S3: Phase-specific criteria used to define additional sampling points in the Irquis Basin; Table S4: Harmonized soil database used in the comparative analysis.

Author Contributions: Conceptualization, C.M.-P. and P.O.-C.; methodology, C.M.-P. and P.O.-C.; software, C.M.-P.; validation, C.M.-P. and P.O.-C.; formal analysis, C.M.-P.; investigation, C.M.-P. and P.O.-C.; resources, C.M.-P.; data curation, C.M.-P.; writing—original draft preparation, C.M.-P.; writing—review and editing, P.O.-C., J.D.R.S. and P.D.S.; visualization, C.M.-P.; supervision, J.D.R.S. and P.D.S.; project administration, C.M.-P.; funding acquisition, C.M.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad Politécnica Salesiana (Ecuador), which supported the field campaigns and data collection activities. The Article Processing Charge (APC) for publication was also funded by Universidad Politécnica Salesiana. The research activities and project management were coordinated by the Water Resources Research Group.

Informed Consent Statement: Not applicable.

Data Availability Statement: The soil dataset supporting the results of this study is provided in the Supplementary Material accompanying this article.

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

Abbreviations

The following abbreviations are used in this manuscript:

LULC	Land Use and Land Cover
K_{sat}	Saturated Hydraulic Conductivity
OM	Organic Matter
ρ_b	Bulk Density

ρ_s	Particle Density
ϕ	Porosity
I_b	Basic Infiltration Rate
PCA	Principal Component Analysis
PC1	Principal Component 1
PC2	Principal Component 2
ZH	Zamora Huayco
PERMANOVA	Permutational Multivariate Analysis of Variance
clr	centered log-ratio
ILR	Isometric Log-Ratio
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
LOOCV	Leave-One-Out Cross-Validation
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
EDA	Exploratory Data Analysis
PTFs	Pedotransfer Functions
USDA	United States Department of Agriculture
UTM	Universal Transverse Mercator
AASHTO	American Association of State Highway and Transportation Officials
ASTM	ASTM International
m.a.s.l.	meters above sea level

References

1. Oñate-Valdivieso, F.; Oñate-Paladines, A.; Collaguazo, M. Spatiotemporal Dynamics of Soil Impermeability and Its Impact on the Hydrology of an Urban Basin. *Land* **2022**, *11*, 250. [CrossRef]
2. López-Ballesteros, A.; Nielsen, A.; Castellanos-Osorio, G.; Trolle, D.; Senent-Aparicio, J. DSOLMap, a novel high-resolution global digital soil property map for the SWAT+ model: Development and hydrological evaluation. *Catena* **2023**, *231*, 107339. [CrossRef]
3. Mohajerani, H.; Teschemacher, S.; Casper, M.C. A Comparative Investigation of Various Pedotransfer Functions and Their Impact on Hydrological Simulations. *Water* **2021**, *13*, 1401. [CrossRef]
4. Vereecken, H.; Weynants, M.; Javaux, M.; Pachepsky, Y.; Schaap, M.G.; van Genuchten, M.T. Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review. *Vadose Zone J.* **2010**, *9*, 795–820. [CrossRef]
5. Lei, Q.; Zhang, T.; An, M.; Luo, J.; Qin, L.; Zhu, A.; Qiu, W.; Du, X.; Liu, H. Sensitivity analysis of SWAT streamflow and water quality to the uncertainty in soil properties generated by the SoLIM model. *J. Hydrol.* **2024**, *642*, 131879. [CrossRef]
6. Han, S.; Coulibaly, P. Bayesian flood forecasting methods: A review. *J. Hydrol.* **2017**, *551*, 340–351. [CrossRef]
7. Fries, A.; Silva, K.; Pucha-Cofrep, F.; Oñate-Valdivieso, F.; Ochoa-Cueva, P. Water Balance and Soil Moisture Deficit of Different Vegetation Units under Semiarid Conditions in the Andes of Southern Ecuador. *Climate* **2020**, *8*, 30. [CrossRef]
8. Kumar, V.; Chaplot, B.; Omar, P.J.; Mishra, S.; Azamathulla, H.M. Experimental study on infiltration pattern: Opportunities for sustainable management in the Northern region of India. *Water Sci. Technol.* **2021**, *84*, 2675–2685. [CrossRef] [PubMed]
9. Iñiguez, V.; Morales, O.; Cisneros, F.; Bauwens, W.; Wyseure, G. Analysis of the Drought Recovery of Andosols on Southern Ecuadorian Andean Páramos. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2421–2435. [CrossRef]
10. Mozaffari, H.; Moosavi, A.A.; Baghernejad, M.; Cornelis, W. Revisiting soil texture analysis: Introducing a rapid single-reading hydrometer approach. *Measurement* **2024**, *228*, 114330. [CrossRef]
11. Saxton, K.E.; Rawls, W.J.; Romberger, J.S.; Papendick, R.I. Estimating Generalized Soil-Water Characteristics from Texture. *Soil Sci. Soc. Am. J.* **1986**, *50*, 1031–1036. [CrossRef]
12. Vereecken, H.; Maes, J.; Feyen, J.; Darius, P. Estimating the Soil Moisture Retention Characteristic from Texture, Bulk Density, and Carbon Content. *Soil Sci.* **1989**, *148*, 389–403. [CrossRef]
13. van Genuchten, M.T. A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [CrossRef]
14. Schaap, M.G.; Leij, F.J.; van Genuchten, M.T. ROSETTA: A Computer Program for Estimating Soil Hydraulic Parameters with Hierarchical Pedotransfer Functions. *J. Hydrol.* **2001**, *251*, 163–176. [CrossRef]
15. Szab'o, B.; Szatm'ari, G.; Tak'acs, K.; Laborczi, A.; Mak'o, A.; Rajkai, K.; P'asztor, L. Mapping Soil Hydraulic Properties Using Random-Forest-Based Pedotransfer Functions and Geostatistics. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2615–2635. [CrossRef]

16. McDaniel, P.A.; Lowe, D.J.; Arnalds, O.; Ping, C.L. Andisols. In *Handbook of Soil Sciences: Properties and Processes*, 2nd ed.; Chapter 33; Huang, P.M., Li, Y., Sumner, M.E., Eds.; CRC Press: Boca Raton, FL, USA, 2012.
17. Takahashi, T.; Dahlgren, R.A. Nature, Properties and Function of Aluminum–Humus Complexes in Volcanic Soils. *Geoderma* **2016**, *263*, 110–121. [CrossRef]
18. Shamrikova, E.V.; Kondratenok, B.M.; Tumanova, E.A.; Vanchikova, E.V.; Lapteva, E.M.; Zonova, T.V.; Lu-Lyan-Min, E.I.; Davydova, A.P.; Libohova, Z.; Suvannang, N. Transferability between Soil Organic Matter Measurement Methods for Database Harmonization. *Geoderma* **2022**, *412*, 115547. [CrossRef]
19. Al-Shammary, A.A.G.; Kouzani, A.Z.; Kaynak, A.; Khoo, S.Y.; Norton, M.; Gates, W. Soil Bulk Density Estimation Methods: A Review. *Pedosphere* **2018**, *28*, 581–596. [CrossRef]
20. Ahmadisharaf, A.; Nematirad, R.; Sabouri, S.; Pachepsky, Y.; Ghanbarian, B. Representative Sample Size for Estimating Saturated Hydraulic Conductivity via Machine Learning: A Proof-of-Concept Study. *Water Resour. Res.* **2024**, *60*, e2023WR036783. [CrossRef]
21. Schoch, J.; Nussbaum, M.; Walthert, L.; Carminati, A.; Lehmann, P. Transferability of Pedotransfer Functions for Estimating Soil Hydraulic Properties: An Analysis of Controlling Factors for Forest Soils in Switzerland. *Geoderma* **2025**, *460*, 117397. [CrossRef]
22. Weber, T.K.D.; Weihermüller, L.; Nemes, A.; Bechtold, M.; Degré, A.; Diamantopoulos, E.; Fatichi, S.; Filipović, V.; Gupta, S.; Hohenbrink, T.L.; et al. Hydro-pedotransfer Functions: A Roadmap for Future Development. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 3391–3434. [CrossRef]
23. Patiño, S.; Hernández, Y.; Plata, C.; Domínguez, I.; Daza, M.; Oviedo-Ocaña, R.; Buytaert, W.; Ochoa-Tocachi, B.F. Influence of land use on hydro-physical soil properties of Andean páramos and its effect on streamflow buffering. *Catena* **2021**, *202*, 105227. [CrossRef]
24. Mejía-Veintimilla, D.; Ochoa-Cueva, P.; Arteaga-Marín, J. Evaluation of the Hydrological Response to Land Use Change Scenarios in Urban and Non-Urban Mountain Basins in Ecuador. *Land* **2024**, *13*, 1907. [CrossRef]
25. Sillero-Medina, J.A.; Hueso-González, P.; Ruiz-Sinoga, J.D. Differences in the Soil Quality Index for Two Contrasting Mediterranean Landscapes in Southern Spain. *Land* **2020**, *9*, 405. [CrossRef]
26. Rodrigo-Comino, J.; Senciales-González, J.M.; Ruiz-Sinoga, J.D. The Effect of Hydrology on Soil Erosion. *Water* **2020**, *12*, 839. [CrossRef]
27. Robinson, D.A.; Thomas, A.; Reinsch, S.; Lebron, I.; Feeney, C.J.; Maskell, L.C.; Wood, C.M.; Seaton, F.M.; Emmett, B.A.; Cosby, B.J. Analytical modelling of soil porosity and bulk density across the soil organic matter and land-use continuum. *Sci. Rep.* **2022**, *12*, 7758. [CrossRef] [PubMed]
28. Benavides-Muñoz, H.M.; Correa-Escudero, V.; Pucha-Cofrep, D.; Pucha-Cofrep, F. Analysis of Land Use Change and Hydrogeological Parameters in the Andean Semiarid Region of Ecuador. *Water* **2024**, *16*, 892. [CrossRef]
29. Oñate-Valdivieso, F.; Oñate-Paladines, A.; Díaz, R. Soil degradation in Andean watersheds: A case study using remote sensing. *Front. Earth Sci.* **2024**, *12*, 1325189. [CrossRef]
30. Buytaert, W.; Deckers, J.; Wyseure, G. Regional Variability of Volcanic Ash Soils in South Ecuador: The Relation with Parent Material, Climate and Land Use. *Catena* **2007**, *70*, 143–154. [CrossRef]
31. Marín, F.; Quiroz Dahik, C.; Mosquera, G.M.; Celleri, R.; Feyen, J.; Crespo, P. Changes in Soil Hydro-Physical Properties and SOM Due to Pine Afforestation and Grazing in Andean Environments Cannot Be Generalized. *Forests* **2019**, *10*, 17. [CrossRef]
32. Correa, A.; Breuer, L.; Crespo, P.; Célleri, R.; Feyen, J.; Birkel, C.; Silva, C.; Windhorst, D. Spatially distributed hydro-chemical data with temporally high-resolution is needed to adequately assess the hydrological functioning of headwater catchments. *Sci. Total Environ.* **2019**, *651*, 1613–1626. [CrossRef] [PubMed]
33. Senent-Aparicio, J.; Peñafiel, L.; Alcalá, F.J.; Jimeno-Sáez, P.; Pérez-Sánchez, J. Climate change impacts on renewable groundwater resources in the andosol-dominated Andean highlands, Ecuador. *Catena* **2024**, *234*, 107766. [CrossRef]
34. Borja Ramón, P.; Iñiguez, V.; Crespo, P.; Cisneros, P.; Cisneros, F. Características físico-químicas de Histosoles y Andosoles de los páramos de Quimsacocha, Ecuador. In Proceedings of the XI Congreso Ecuatoriano de la Ciencia del Suelo, Quito, Ecuador, 29–31 October 2008; pp. 1–11.
35. Quichimbo, P.; Tenorio, G.; Borja, P.; Cárdenas, I.; Crespo, P.; Célleri, R. Efectos sobre las propiedades físicas y químicas de los suelos por el cambio de la cobertura vegetal y uso del suelo: Páramo de Quimsacocha al sur del Ecuador. *Suelos Ecuat.* **2012**, *42*, 138–153.
36. Mikutta, R.; Kleber, M.; Kaiser, K.; Jahn, R. Review: Organic Matter Removal from Soils Using Hydrogen Peroxide, Sodium Hypochlorite, and Disodium Peroxodisulfate. *Soil Sci. Soc. Am. J.* **2005**, *69*, 120–135. [CrossRef]
37. Nanzyo, M. Unique Properties of Volcanic Ash Soils. *Glob. Environ. Res.* **2002**, *6*, 99–112.
38. Jensen, J.L.; Schjønning, P.; Watts, C.W.; Christensen, B.T.; Munkholm, L.J. Soil texture analysis revisited: Removal of organic matter matters more than ever. *PLoS ONE* **2017**, *12*, e0178039. [CrossRef] [PubMed]
39. Mwendwa, S. Revisiting soil texture analysis: Practices towards a more accurate Bouyoucos method. *Heliyon* **2022**, *8*, e09395. [CrossRef] [PubMed]

40. Papuga, K.; Kaszubkiewicz, J.; Kawałko, D.; Kreimeyer, M. Effect of Organic Matter Removal by Hydrogen Peroxide on the Determination of Soil Particle Size Distribution Using the Dynamometer Method. *Agriculture* **2022**, *12*, 226. [CrossRef]
41. Jantz, N.; Behling, H. A Holocene Environmental Record Reflecting Vegetation, Climate, and Fire Variability at the Páramo of Quimsacocha, Southwestern Ecuadorian Andes. *Veg. Hist. Archaeobotany* **2012**, *21*, 169–185. [CrossRef]
42. Crespo, P.J.; Feyen, J.; Buytaert, W.; Bücker, A.; Breuer, L.; Frede, H.G.; Ramírez, M. Identifying Controls of the Rainfall–Runoff Response of Small Catchments in the Tropical Andes (Ecuador). *J. Hydrol.* **2011**, *407*, 164–174. [CrossRef]
43. Lazo, P.X.; Mosquera, G.M.; McDonnell, J.J.; Crespo, P. The Role of Vegetation, Soils, and Precipitation on Water Storage and Hydrological Services in Andean Páramo Catchments. *J. Hydrol.* **2019**, *572*, 805–819. [CrossRef]
44. Ochoa-Cueva, P.; Fries, A.; Montesinos, P.; Rodríguez-Díaz, J.A.; Boll, J. Spatial Estimation of Soil Erosion Risk by Land-cover Change in the Andes of Southern Ecuador. *Land Degrad. Dev.* **2015**, *26*, 565–573. [CrossRef]
45. Oñate-Valdivieso, F.; Piedra Faicán, P.; Oñate-Paladines, A. Forecasting of Macroclimatic Phases Through Stochastic Modeling and Machine Learning: Implications for Regional Hydrological Analysis. *Water* **2026**, *18*, 1358. [CrossRef]
46. El Mouridi, Z.; Ziri, R.; Douaik, A.; Bennani, S.; Lembaid, I.; Bouharou, L.; Brhadda, N.; Moussadek, R. Comparison between Walkley-Black and Loss on Ignition Methods for Organic Matter Estimation in Different Moroccan Soils. *Ecol. Eng. Environ. Technol.* **2023**, *24*, 253–259. [CrossRef] [PubMed]
47. Ratefinjanahary, I.; MacKenzie, R.A.; Sharma, S.; Razakamanarivo, H.; Razafintsalama, V.; Ravelosamiarinirina, K.; Welti, A.; Ramifehiarivo, N. Development of soil organic carbon quantification model and comparison based on CHN analyser, Loss on Ignition, and Walkley-Black methods for mangrove soils in Madagascar. *Estuarine Coast. Shelf Sci.* **2025**, *317*, 109182. [CrossRef]
48. Brus, D.J.; Kempen, B.; Heuvelink, G.B.M. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* **2011**, *62*, 394–407. [CrossRef]
49. Biswas, A.; Zhang, Y. Sampling Designs for Validating Digital Soil Maps: A Review. *Pedosphere* **2018**, *28*, 1–15. [CrossRef]
50. Stumpf, F.; Schmidt, K.; Goebes, P.; Behrens, T.; Schönbrodt-Stitt, S.; Wadoux, A.; Xiang, W.; Scholten, T. Uncertainty-guided sampling to improve digital soil maps. *Catena* **2017**, *153*, 30–38. [CrossRef]
51. Blackford, C.; Heung, B.; Webster, K.L. Incorporating spatial uncertainty maps into soil sampling improves digital soil mapping classification accuracy in Ontario, Canada. *Geoderma Reg.* **2022**, *29*, e00495. [CrossRef]
52. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]
53. Skøien, J.O.; Blöschl, G.; Western, A.W. Scale Effects in Estimating the Variogram and Implications for Soil Hydrology. *Vadose Zone J.* **2006**, *5*, 153–167. [CrossRef]
54. Huang, J.; McBratney, A.B.; Triantafilis, J.; Malone, B.P.; Minasny, B. Evaluating an adaptive sampling algorithm to assist soil survey in New South Wales, Australia. *Geoderma Reg.* **2020**, *21*, e00284. [CrossRef]
55. Moradi, M.; Brown, J.A. Mapping using an adaptive sampling design. *Spat. Stat.* **2024**, *60*, 100820. [CrossRef]
56. Gee, G.W.; Or, D. Particle-Size Analysis. In *Methods of Soil Analysis. Part 4. Physical Methods*; Dane, J.H., Topp, G.C., Eds.; Soil Science Society of America: Madison, WI, USA, 2002; pp. 255–293. [CrossRef]
57. Soil Survey Staff. *Soil Survey Manual*; USDA Handbook No. 18, United States Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA, 2017.
58. Flint, A.L.; Flint, L.E. Particle Density. In *Methods of Soil Analysis. Part 4. Physical Methods*; Dane, J.H., Topp, G.C., Eds.; SSSA Book Series No. 5, Soil Science Society of America: Madison, WI, USA, 2002; pp. 229–240. [CrossRef]
59. Flint, L.E.; Flint, A.L. Porosity. In *Methods of Soil Analysis. Part 4. Physical Methods*; Dane, J.H., Topp, G.C., Eds.; SSSA Book Series No. 5, Soil Science Society of America: Madison, WI, USA, 2002; pp. 241–254. [CrossRef]
60. *AASHTO T 267-22*; Determination of Organic Content in Soils by Loss on Ignition. American Association of State Highway and Transportation Officials: Washington, DC, USA, 2022.
61. *ASTM D3385-03*; Standard Test Method for Infiltration Rate of Soils in Field Using Double-Ring Infiltrometer. ASTM International: West Conshohocken, PA, USA, 2003.
62. Fatehnia, M.; Tawfiq, K.; Ye, M. Estimation of saturated hydraulic conductivity from double-ring infiltrometer measurements. *Eur. J. Soil Sci.* **2016**, *67*, 135–147. [CrossRef]
63. Chowdary, V.M.; Rao, M.D.; Jaiswal, C.S. Study of infiltration process under different experimental conditions. *Agric. Water Manag.* **2006**, *83*, 69–78. [CrossRef]
64. Webster, R. Statistics to Support Soil Research and Their Presentation. *Eur. J. Soil Sci.* **2001**, *52*, 331–340. [CrossRef]
65. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef]
66. Wadoux, A.M.J.C.; Minasny, B.; McBratney, A.B. Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. [CrossRef]
67. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. [CrossRef]

68. Araya, S.N.; Ghezzehei, T.A. Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. *Water Resour. Res.* **2019**, *55*, 5715–5737. [CrossRef]
69. Rastgou, M.; Bayat, H.; Mansoorizadeh, M.; Gregory, A.S. Prediction of Soil Hydraulic Properties by Gaussian Process Regression Algorithm in Arid and Semiarid Zones in Iran. *Soil Tillage Res.* **2021**, *210*, 104980. [CrossRef]
70. Gupta, S.; Hengl, T.; Lehmann, P.; Bonetti, S.; Or, D. SoilKsatDB: Global Database of Soil Saturated Hydraulic Conductivity Measurements for Geoscience Applications. *Earth Syst. Sci. Data* **2021**, *13*, 1593–1612. [CrossRef]
71. Fewster, R.E.; Morris, P.J.; Swindles, G.T.; Baird, A.J.; Turner, T.E.; Ivanovic, R.F. Controls on Saturated Hydraulic Conductivity in a Degrading Permafrost Peatland Complex. *Water Resour. Res.* **2023**, *59*, e2023WR035398. [CrossRef]
72. Greenacre, M.; Groenen, P.J.F.; Hastie, T.; D'Enza, A.I.; Markos, A.; Tuzhilina, E. Principal Component Analysis. *Nat. Rev. Methods Prim.* **2022**, *2*, 100. [CrossRef]
73. Takada, T.; Nijman, S.; Denaxas, S.; Snell, K.I.E.; Uijl, A.; Nguyen, T.L.; Asselbergs, F.W.; Debray, T.P.A. Internal-External Cross-Validation Helped to Evaluate the Generalizability of Prediction Models in Large Clustered Datasets. *J. Clin. Epidemiol.* **2021**, *137*, 83–91. [CrossRef] [PubMed]
74. Collins, G.S.; Dhiman, P.; Ma, J.; Schlussel, M.M.; Archer, L.; Van Calster, B.; Harrell, F.E.; Martin, G.P.; Moons, K.G.M.; van Smeden, M.; et al. Evaluation of Clinical Prediction Models (Part 1): From Development to External Validation. *BMJ* **2024**, *384*, e074819. [CrossRef] [PubMed]
75. Zhang, J. Information Criteria for Model Selection. *WIREs Comput. Stat.* **2023**, *15*, e1607. [CrossRef]
76. Morris, T.P.; van Smeden, M.; Pham, T.M. The Marginality Principle Revisited: Should Higher-Order Terms Always Be Accompanied by Lower-Order Terms in Regression Analyses? *Biom. J.* **2023**, *65*, 2300069. [CrossRef] [PubMed]
77. Oh, J.; Kim, K.H.; Kim, H.R.; Park, S.; Yun, S.T. Using Isometric Log-Ratio in Compositional Data Analysis for Developing a Groundwater Pollution Index. *Sci. Rep.* **2024**, *14*, 12196. [CrossRef] [PubMed]
78. Cayuela-Sánchez, J.A.; López-Núñez, R. Compositional Data Methods and VISNIRS to Predict Soil Organic Carbon Contents. *Eur. J. Soil Sci.* **2025**, *76*, e70200. [CrossRef]
79. Lopez, E.; Gorla, G.; Etxebarria-Elezgarai, J.; Amigo, J.M.; Seifert, A. The Importance of Choosing a Proper Validation Strategy in Predictive Models. Part 2: Recipes for (Avoiding) Overfitting–A Tutorial. *Anal. Chim. Acta* **2026**, *1384*, 344838. [CrossRef] [PubMed]
80. de Mello, D.C.; Veloso, G.V.; de Mello, M.F.; de Lana, M.G.; de Oliveira, I.d.A.; de Oliveira Mello, F.A.; Siqueira, R.G.; Gomes, L.C.; Fernandes-Filho, E.I.; Schaefer, C.E.G.R.; et al. Integrating Proximal Geophysical Sensing and Machine Learning for Digital Soil Mapping: Spatial Prediction and Model Evaluation Using a Small Dataset. *Soil Adv.* **2025**, *3*, 100024. [CrossRef]
81. Zhang, M.; Shi, W.; Ma, Y.; Ge, Y. Mapping Soil Particle-Size Fractions Based on Compositional Balances. *Catena* **2024**, *234*, 107643. [CrossRef]
82. Dikta, G.; Scheer, M. *Bootstrap Methods: With Applications in R*; Springer: Cham, Switzerland, 2021. [CrossRef]
83. Mgohele, R.N.; Massawe, B.H.J.; Shitindi, M.J.; Sanga, H.G.; Omar, M.M. Prediction of Soil Texture Using Remote Sensing Data: A Systematic Review. *Front. Remote Sens.* **2024**, *5*, 1461537. [CrossRef]
84. Vrána, M.; Kubát, J.F.; Kavka, P.; Zumr, D. A laser diffractometry technique for determining the soil water stable aggregates index. *Geoderma* **2024**, *441*, 116756. [CrossRef]
85. Buytaert, W.; Deckers, J.; Wyseure, G. Description and Classification of Nonallophanic Andosols in South Ecuadorian Alpine Grasslands (Páramo). *Geomorphology* **2006**, *73*, 207–221. [CrossRef]
86. Buytaert, W.; Sevink, J.; De Leeuw, B.; Deckers, J. Clay Mineralogy of the Soils in the South Ecuadorian Páramo Region. *Geoderma* **2005**, *127*, 114–129. [CrossRef]
87. Buytaert, W.; Wyseure, G.; De Bièvre, B.; Deckers, J. The Effect of Land-Use Changes on the Hydrological Behaviour of Histic Andosols in South Ecuador. *Hydrol. Process.* **2005**, *19*, 3985–3997. [CrossRef]
88. Páez-Bimos, S.; Villacís, M.; Morales, O.; Calispa, M.; Molina, A.; Salgado, S.; de Bievre, B.; Delmelle, P.; Muñoz, T.; Vanacker, V. Vegetation Effects on Soil Pore Structure and Hydraulic Properties in Volcanic Ash Soils of the High Andes. *Hydrol. Process.* **2022**, *36*, e14678. [CrossRef]
89. Podwojewski, P.; Poulenard, J.; Toulkeridis, T.; Gräfe, M. Polygenic Soils in the Southern Central Ecuadorian Highlands as the Result of Long-Lasting Pedogenesis, Geodynamic Processes and Climate Change. *J. S. Am. Earth Sci.* **2022**, *119*, 104119. [CrossRef]
90. Staß, S.; Dorau, K.; Aguirre, E.; Schwiebert, R.; Mansfeldt, T. Land Use Effects on Organic Carbon in Andean Volcanic Ash Soils. *J. Plant Nutr. Soil Sci.* **2025**, *188*, 196–208. [CrossRef]
91. Caurapan, B.; Zehetner, F.; Ottner, F.; Zúñiga, F.; Valle, S. Effects of Land-Use Change on Morphological, Physicochemical, and Mineralogical Properties of Volcanic Soils in Southern Chile. *Geoderma* **2025**, *462*, 117510. [CrossRef]
92. De Pue, J.; Rezaei, M.; Van Meirvenne, M.; Cornelis, W.M. The relevance of measuring saturated hydraulic conductivity: Sensitivity analysis and functional evaluation. *J. Hydrol.* **2019**, *576*, 628–638. [CrossRef]
93. Corral-Pazos-de-Provens, E.; Rapp-Arrarás, Í.; Domingo-Santos, J.M. Estimating Textural Fractions of the USDA Using Those of the International System: A Quantile Approach. *Geoderma* **2022**, *416*, 115783. [CrossRef]

94. Wainer, J.; Cawley, G. Nested Cross-Validation When Selecting Classifiers Is Overzealous for Most Practical Applications. *Expert Syst. Appl.* **2021**, *182*, 115222. [CrossRef]
95. Shi, W.; Zhang, M. Progress on Spatial Prediction Methods for Soil Particle-Size Fractions. *J. Geogr. Sci.* **2023**, *33*, 2303–2324. [CrossRef]
96. Rosso, P.; Kersebaum, K.C.; Groh, J.; Gerke, H.H.; Heil, K.; Gebbers, R. Pedotransfer Functions and Their Impact on Soil Water Dynamics Simulation and Yield Prediction: A HERMES Model Case Study. *Eur. J. Agron.* **2025**, *169*, 127730. [CrossRef]
97. Paschalis, A.; Bonetti, S.; Guo, Y.; Fatichi, S. On the Uncertainty Induced by Pedotransfer Functions in Terrestrial Biosphere Modeling. *Water Resour. Res.* **2022**, *58*, e2021WR031871. [CrossRef]
98. Kubát, J.; Strouhal, L.; Kavka, P. Estimation of Infiltration Parameters: The Role of Pedotransfer Functions and Initial Moisture Conditions. *J. Hydrol.* **2024**, *633*, 130954. [CrossRef]
99. Darmann, F.; Schwaighofer, I.; Kumpan, M.; Weninger, T.; Strauss, P. New Hydro-Pedotransfer Functions for Austrian Soil Mapping Applications. *Geoderma Reg.* **2024**, *39*, e00875. [CrossRef]
100. Valladares, C.; Boelens, R. Mining for Mother Earth. Governmentalities, Sacred Waters and Nature's Rights in Ecuador. *Geoforum* **2019**, *100*, 68–79. [CrossRef]
101. Hungerbühler, D.; Steinmann, M.; Winkler, W.; Seward, D.; Egüez, A.; Peterson, D.E.; Helg, U.; Hammer, C. Neogene Stratigraphy and Andean Geodynamics of Southern Ecuador. *Earth-Sci. Rev.* **2002**, *57*, 75–124. [CrossRef]

https://doi.org/10.3390/soilsystems10070068