



Soil information and soil property maps for the Kurdistan region, Dohuk governorate (Iraq)

Mathias Bellat^{1,2}, Mjahid Zebari^{3,4}, Benjamin Glissmann^{1,5}, Tobias Rentschler^{1,2,6}, Paola Sconzo⁷, Nafiseh Kakhani^{1,2}, Ruhollah Taghizadeh-Mehrjardi^{1,2,8}, Pegah Kohsrovani^{2,9}, Bekas Brifkany¹⁰, Peter Pfälzner^{1,5}, and Thomas Scholten^{1,2,11}

¹CRC 1070 “ResourceCultures”, University of Tübingen, Tübingen, 72070, Germany

²Department of Geosciences, Working group of Soil Science and Geomorphology,
University of Tübingen, Tübingen, 72070, Germany

³Ludwig-Maximilians-Universität München, München, 80634, Germany

⁴Nawroz University, Duhok, Iraq

⁵Institute for Ancient Near Eastern Studies (IANES), University of Tübingen, Tübingen, 72070, Germany

⁶Digital Humanities Center, University of Tübingen, 72074, Germany

⁷Department of Culture and Society, University of Palermo, Palermo, 90133, Italy

⁸Faculty of Agriculture and Natural resources, Ardakan University, Ardakan, Iran

⁹College of Agriculture, Shiraz University, Shiraz, Iran

¹⁰Dohuk Directorate of Antiquities and Heritages, Dohuk, Iraq

¹¹DFG Cluster of Excellence “Machine Learning: New Perspectives for Science”,
University of Tübingen, Tübingen, 72076, Germany

Correspondence: Mathias Bellat (mathias.bellat@uni-tuebingen.de)

Received: 18 July 2025 – Discussion started: 15 September 2025

Revised: 9 March 2026 – Accepted: 21 March 2026 – Published: 9 April 2026

Abstract. We present the first detailed soil property maps at multiple depths for the northwestern autonomous Kurdistan region of Iraq (Dohuk). A total of 532 soil samples from 122 sites were collected at five depth increments (0–10, 10–30, 30–50, 50–70, and 70–100 cm), and their mid-infrared (MIR) spectra were measured. A subset of 108 samples, selected via Kennard–Stone sampling, was analysed in a laboratory on ten soil properties. A Cubist model was trained and used from these measured values to predict all samples’ soil properties from their MIR spectra. Digital soil mapping was conducted using a machine learning regression techniques based on a quantile random forest model, trained on the predicted soil properties and using a total of 85 covariates at 30 m pixel resolution, resulting in 50 prediction maps in total. Results were compared with the SoilGrids 2.0 product and a regional texture model. Soil depth was also mapped using a similar model with 26 covariates. Our model outperformed global SoilGrids 2.0 predictions in resolution and accuracy, with texture RMSEs (sand: $\bar{x}$ RMSE 11.03; silt: $\bar{x}$ RMSE 8.82; clay: $\bar{x}$ RMSE 7.39) comparable to local models. Key predictors included Landsat 8 SWIR, EVI, SAVI, Sentinel 2 SWIR, PET and solar radiation. Spatial patterns reflected the contrast between the flat areas of the Selevani and Zakho plains, as opposed to the shallower and steeper Little Khabur Valley and anticline formations. Furthermore, the soil depth prediction model (R^2 0.39; RMSE 30.76 cm) showed strong correlation with slope and a similar pattern distribution with deeper soils in the flat areas of the Selevani and Zakho plains, while shallow soils were predicted in the anticline and strongly erodible areas. Our comprehensive dataset (<https://doi.org/10.1594/PANGAEA.973700>, Bellat et al., 2024a; <https://doi.org/10.1594/PANGAEA.973701>, Bellat et al., 2024b; <https://doi.org/10.1594/PANGAEA.973714>, Bellat et al., 2024c; <https://doi.org/10.6084/m9.figshare.31320958.v2>, Bellat et al., 2026a; <https://doi.org/10.57754/FDAT.d5h1h-4x027>, Bellat et al., 2026b) offers substantial insights for soil knowledge in the region, as well as for aridic and semi-aridic areas.

1 Introduction

Soils record chemical, physical and biological processes over extended temporal scales (Hillel and Hatfield, 2005; Schaetzl and Anderson, 2005; Duchaufour et al., 2020). They are part of global exchanges (Bossio et al., 2020; Lal et al., 2021; Telo da Gama, 2023) and exert significant influence on local ecosystems (Adhikari and Hartemink, 2016; Scholten et al., 2017; Zeraatpisheh et al., 2022; Webber et al., 2023; Guan et al., 2024). Soil texture provides insights into soil stability, water retention, carbon storage, and biomass production (Rabot et al., 2018), while pH regulates soil acidity and nutrient availability for plants (Thomas, 1996; Neina, 2019). Organic carbon (OC) reflects local organic production and functions as a major carbon sink for CO₂ at a global level (Trivedi et al., 2018; Bossio et al., 2020; Beillouin et al., 2023). Inorganic carbon – calculated as total carbon (C_t) minus organic carbon – also plays a critical role in carbon sequestration in semi-arid zones (Zamanian et al., 2016; Sharififar et al., 2023). Calcium carbonate (CaCO₃), abundant in calcareous soils of semi-arid climates, further influences both acidity (Yu et al., 2023) and carbon dynamics (Umer et al., 2020; Dou et al., 2023). Additional key soil properties include total nitrogen (N_t), which influences plant growth (Crawford and Forde, 2002; Anas et al., 2020), and electrical conductivity (EC), essential for assessing soil water content or capacity (Brevik et al., 2006), and soil salinity (Friedman, 2005), particularly problematic in arid and semi-arid regions such as Iraq (Smith and Robertson, 1962; Christen and Saliem, 2013; Azeez and Rahimi, 2017). Evaluating all of these properties and establishing a taxonomic classification of a soil gives information on its ability to fit agricultural purposes and helps to better understand the development of soils over time and under changing climatic conditions.

In the Dohuk Governorate (*Parêzgeha Dihok*) of north-western Kurdistan region of Iraq (*Herêmi Kurdistan*, Fig. 1), exploratory mapping efforts (Buringh, 1957, 1960; Altaie, 1968; Altaie et al., 1969; Barzanji, 1973; Muhaimeed et al., 2014; Muhaimeed, 2020) identified the presence of semi-arid and mountainous soils shaped by complex interactions between geomorphology, parent material and climate. The fluvial dynamics of the Tigris River (*Dicle*) have been recognised as a major factor in landscape formation, influencing salinity, clay deposition, and vertic properties through sedimentation and erosion (Buringh, 1960, pp. 51–54). However, critics (Wilkinson, 1990) have suggested that vertic features and horizons might have been overestimated (Buringh, 1957, 1960; Altaie, 1968; Abdulrahman et al., 2020). Gypsum is another critical factor in local soil development, either inherited from primary deposits such as alabaster formations (Buringh, 1960, p. 106), or formed secondarily through irrigation-induced precipitation and soil chemical processes (Buringh, 1960, p. 107). High gypsum concentrations are

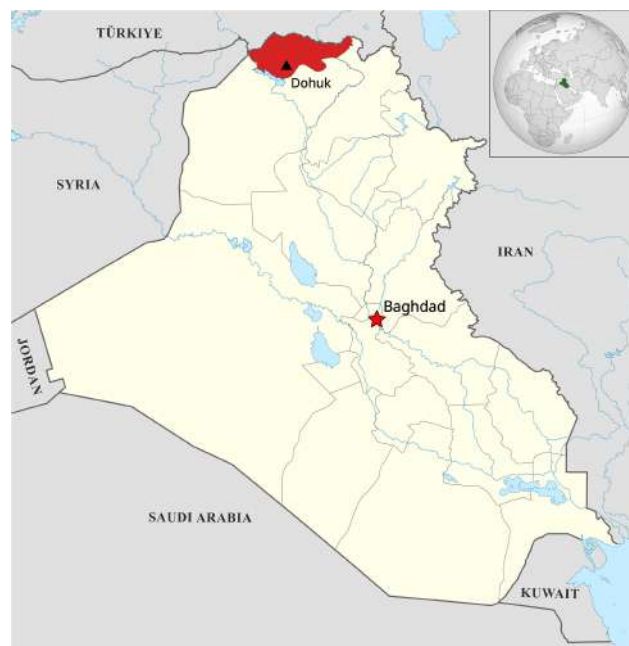


Figure 1. Location of Dohuk governorat in the Republic of Iraq (Wikimedia commons). Realised with QGIS 3.34.5.

commonly found in areas south and south-west of the Zagros and Taurus mountain chains (Smith and Robertson, 1962; Barzanji, 1973; Azeez and Rahimi, 2017), reflecting the influence of regional hydrogeology, and aquitard structures (Buringh, 1960, p. 108; Azeez and Rahimi, 2017). Favourable factors for soil development have been poorly explored outside of the alluvial plain area (Altaie et al., 1969; Barzanji, 1973). While some valley bottom soils may exhibit higher organic carbon content (Buringh, 1960, p. 78; Altaie, 1968) soils in upland areas are generally poorly developed due to severe erosion, leading to shallow, fragmented profiles (Muhaimeed et al., 2013) referred to as “broken soils” (Buringh, 1957; Altaie, 1968).

Quantitative soil property data for the region remain scarce. The global SoilGrids 2.0 product (Poggio et al., 2021) offers coarse-resolution (250 m) predictions of key soil attributes. While adequate at a national scale in some regions (Varón-Ramírez et al., 2022; Shi et al., 2025), its performance at finer scales is limited, particularly due to sparse calibration points in the Middle East and Iraq (Batjes et al., 2020; FAO and IIASA, 2023). At the regional scale, only one recent study has attempted digital texture mapping (Yousif et al., 2023), but it covers a different area and does not account for the full range of soil-forming factors described in the *scorpan* model (McBratney et al., 2003).

Previous classifications and soil descriptions in the region were mostly carried out at the national scale and do not reflect recent landscape changes (Forti et al., 2022), nor do they align with contemporary standards (IUSS Working Group

WRB, 2022). Moreover, no spatially explicit dataset currently exists for the most important chemical and physical soil attributes. While the previous mappings only used limited observations windows, with modern digital soil mapping (DSM), the spatial distribution of soils and their characteristics can now be described and modelled with increasing accuracy (Behrens and Scholten, 2006; Taghizadeh-Mehrjardi et al., 2014). Therefore, we have developed a meso-scale (1 : 200 000; 30 m pixel) DSM of key properties in the Dohuk region, alongside an updated classification based on the WRB taxonomy (IUSS Working Group WRB, 2022). Soil sampling campaigns conducted between 2022 and 2023 enabled the creation of 10 soil property maps across five depth intervals and a soil depth model for the western part of Dohuk governorate (Fig. 1). All data products follow the FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) and were adapted to physical geography specificities (Bailo et al., 2020). These outputs are relevant for application in agriculture, geography, and ecology, especially as climate change exacerbates desertification in Iraq (Eltaif and Gharaibeh, 2022; Eltaif et al., 2024). The production of a spatial dataset from unexplored region, and the digital soil maps from recent field observations became an asset for depicting actual soil situations and exploring potential solutions.

2 Material and Methods

The workflow (Fig. 2) followed a semi-standardised fully reproducible protocol (Malone et al., 2022). Using a conditioned Latin hypercube sampling design (cLHS), 532 soil samples were collected from 122 sites (Table 1) and analysed via mid-infrared (MIR) spectroscopy. A representative subset of 108 samples, including legacy material from older surveys, underwent laboratory analysis for detailed physical, biological, and chemical characterisation. These samples were used to calibrate a Cubist regression model, with one raw and three transformed MIR spectra as predictor variables. The resulting predictions of soil properties were then integrated into a digital soil mapping framework (Lagacherie et al., 2006; Behrens and Scholten, 2006; Brevik et al., 2016; Malone et al., 2017; Hengl and MacMillan, 2019), following the *scorpan* equation model (Eq. 1; McBratney et al., 2003), and computed with a quantile random forest regression for each of five soil depth increments. Additionally, a soil depth map was developed using a similar model, incorporating remote sensing covariates and field observations. The digital soil maps integrate field observations and spatial predictions derived from a suite of remote sensing and spatial datasets. All datasets used are listed in Table A1, with further methodological detail provided in the Supplement.

$$S_a = f(s, c, o, r, p, a, n) \quad (1)$$

2.1 Study area

The data were collected from the Dohuk governorate in the Kurdistan region of Iraq, specifically from the Simele (*Qezaya Sêmêl*) and Zakho districts (*Qezaya Zaxo*, Fig. 3) covering a total area of 2280 km². The region is often referred to as Eastern Khabur/^Habur (*Xabûr*, Pfälzner and Sconzo, 2016), though it is sometimes divided into two entities: the eastern Syrian al-Jazira (*Ġazîra*) for the western and southern part and the mountain chain of Khabur for the northern part (Abdulsalam and Schlaich, 1988).

2.1.1 Tectonic development and parent material

Our study area within the Dohuk governorate is located within the northwestern segment of the Zagros-fold thrust belt (ZFTB), a mountain belt that extends from southern Iran NW-ward to the Kurdistan Region of Iraq and SE Turkey. The ZFTB resulted from the ongoing convergence between the Arabian and Eurasian plates (Berberian, 1995; Agard et al., 2011; Mouthereau et al., 2012; Sembroni et al., 2024). The convergence started in the late Cretaceous with the subduction of the Neotethys oceanic crust beneath the Eurasia Plate and the obduction of the ophiolite sequences on Arabia's margin, followed by the subsequent continent-continent collision between the Arabian and Eurasian plates during the Oligocene-Early Miocene (Agard et al., 2011; Khadivi et al., 2012; Mouthereau et al., 2012; Koshnaw et al., 2017). Since the onset of continental deformation on the northeastern margin of the Arabian Plate (including the study area), it has propagated for 250–350 km (Blanc et al., 2003; Molinaro et al., 2005; Alavi, 2007; Agard et al., 2011; Mouthereau et al., 2012; Koshnaw et al., 2020; Zebari et al., 2020; Sembroni et al., 2024). Within the external part of the ZFTB, these zones include the Imbricated Zone, the High Folded Zone, and the Foothill (Low Folded) Zone (Berberian, 1995; Jassim and Goff, 2006; Fouad, 2012, 2014; Zebari et al., 2020).

The study area covers parts of the High Folded and Foothill zones (Fig. 4), where structures are mainly trending in a nearly E–W direction (Forti et al., 2021; Doski and McClay, 2022). The Bekhair/Ġebel Bihair Anticline (*Şax-e Bixêr*) is the main structural and morphological feature in the area and plunges at the western end of our study area. It separates the Selevani Plain (*Daşt-e Selevani*), which stretches from the northern bank of the Mosul Dam Lake to the anticline, in the south, from the Zakho-Cizre-Silopi Plain – later shorten into Zakho Plain – and Little Khabur Valley to the north (Forti et al., 2021; Doski and McClay, 2022).

The exposed rocks in the area include sedimentary units ranging in age from the Upper Cretaceous to the Pliocene (Sissakian and Al-Jiburi, 2012, 2014; Doski and McClay, 2022). The Upper Cretaceous units consist of platform carbonates and siliciclastic rocks (Jassim and Goff, 2006; Aqrabi et al., 2010). The Paleocene-Eocene units consist

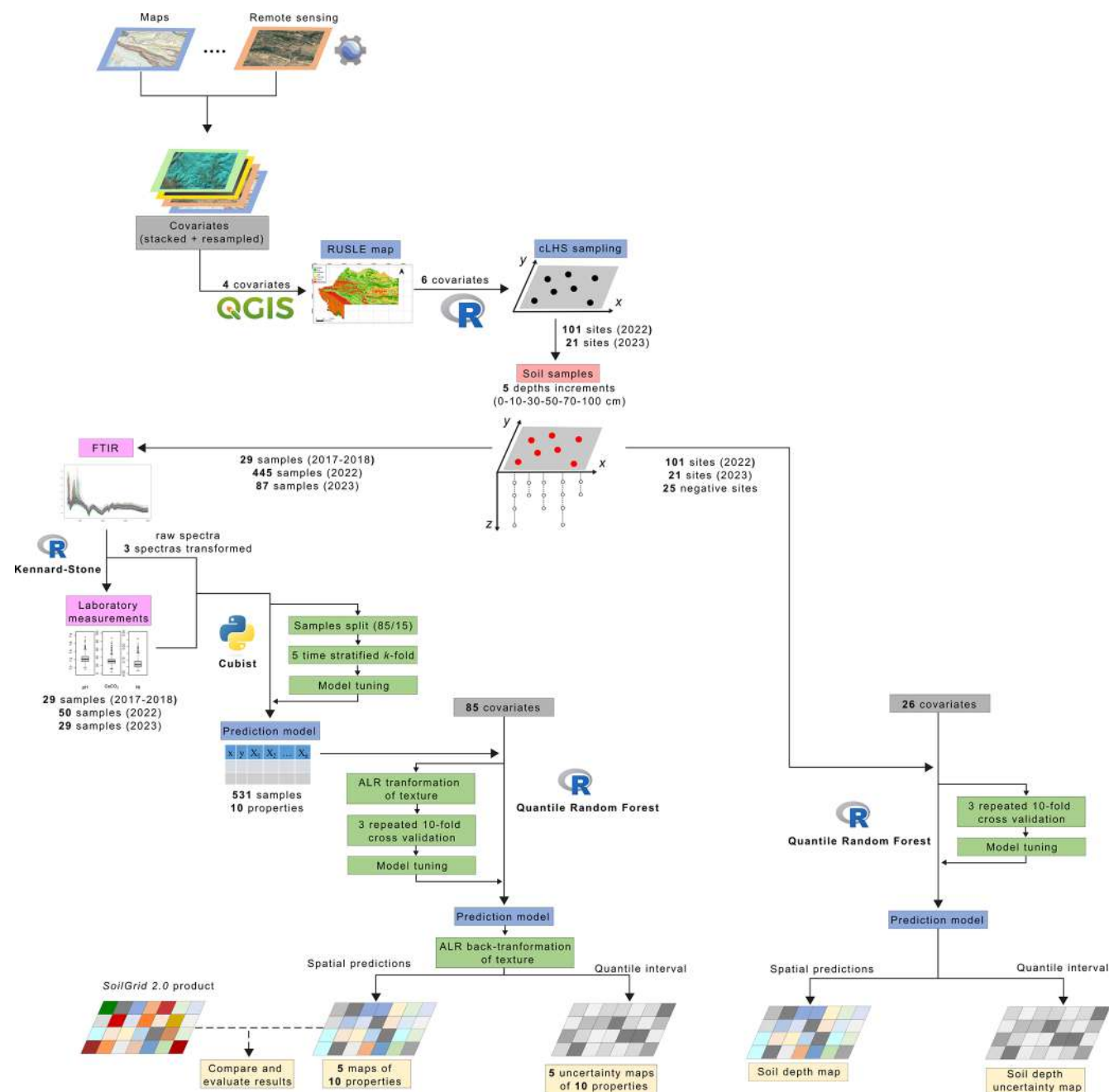


Figure 2. Workflow of the soil properties maps production based on Malone et al. (2022) protocol. Realised with Inkscape 1.4, and inspired by Shi et al. (2025) design.

mainly of marginal marine marls and shales that interfinger with rigid carbonate units, followed by red Eocene clays and carbonates. These Upper Cretaceous-Eocene units are exposed within the anticlinal structures in the area. The Oligocene units are missing in the area; thus, the Eocene carbonates underlie the Middle Miocene clays, evaporites, and limestones. The Upper Miocene–Pliocene units consist of fluvial sandy succession, clay, and conglomerate deposited in the Zagros foreland basin (Jassim and Goff, 2006; Aqrabi

et al., 2010). The Miocene–Pliocene units are exposed within the synclines and low-elevation area to the north and south of Bekhair Anticline.

The Quaternary deposits cover three different environments of the study area (Fig. 4). First, the flat area of the Selevani Plain and north of the Zakho Plain (Türkiye) is covered by residual clayey soil material, coming from the erosion of the Bekhair and Zagros anticlines. Second, along the riverbanks of the Tigris and Little Khabur rivers, sand

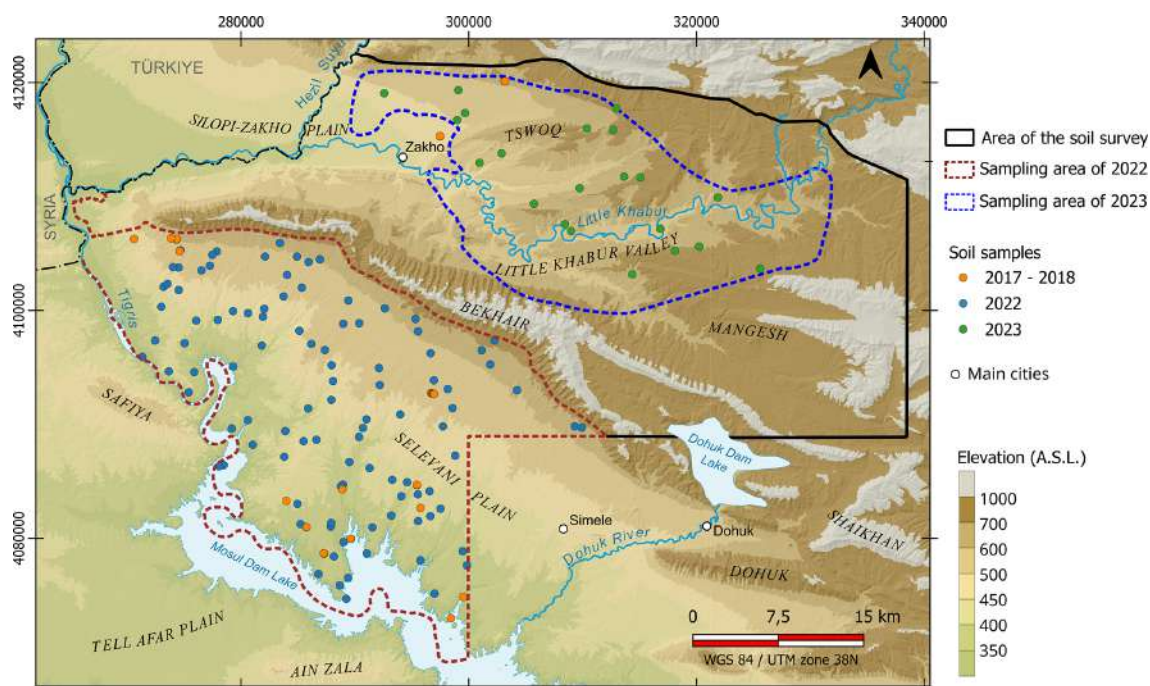


Figure 3. Map of the sampling areas and different sample locations (Background: Copernicus data ESA and Airbus, 2022). Realised with QGIS 3.34.5 and Inkscape 1.4.

and gravel-sized terrace deposits, as well as floodplain sediments of fine sand and clay, can be observed. Finally, Quaternary formations from alluvial fan sediments of clayey soil, combined with rock fragments coming from colluvial deposits, are visible in the foothills of the Bekhair and the shallow Little Khabur Valley. Sometimes, calcrete is also developed within these Quaternary deposits (Sissakian and Al-Jiburi, 2012, 2014; Forti et al., 2021).

2.1.2 Climate and vegetation

The central part of the study area falls within a Csa (Hot-summer Mediterranean) agro-climatic zone, according to the Köppen Geiger classification (Köppen, 1936; Beck et al., 2018; Alwan et al., 2019). Annual precipitation ranges from 200 to 500 mm, with an average yearly temperature exceeding 16°C (Fick and Hijmans, 2017; Salman et al., 2019; Najmaldin, 2023). Only the Little Khabur Valley, located north of the Bekhair anticline, experiences slightly cooler winters and receives higher rainfall, typically between 500 and 800 mm per year (Fick and Hijmans, 2017; Alwan et al., 2019). South of the Bekhair anticline, the Selevani Plain belongs to the Mesopotamian steppe floral complex, which supports a limited number of xerophytic shrubs and herbs, primarily *Artemisia herba-alba mesopotamica* often associated with *Aristida plumosa* (Guest and Al-Rawi, 1966, pp. 78–80; Zohary, 1973, p. 183). In contrast, the northern region, encompassing the Zakho Plain and Little Khabur Valley, falls within the Kurdo-Zagrosian climate zone, characterised

by a denser xerophilous deciduous steppe forest, driven by its higher elevation and more favourable climatic conditions (Fig. B1; Guest and Al-Rawi, 1966, p. 68). Dominant shrubs include *Anagyris foetida* or *Pistacia khinjuk* are associated with trees as *Quercus brantii*, or *Quercus boissieri*, which grows between 800 and 1700 m of altitude. Historical records mention the presence of pine forests (Zohary, 1973, pp. 183–190), though they are likely no longer extant. In both the foothills of the Mesopotamian Plain and the Kurdo-Zagrosian space, cultivated *Oliva europaeis* can be sporadically observed.

2.1.3 Geomorphology and soils

In the southern part of our study area, the Tigris floodplain and its Quaternary alluvial deposits have largely disappeared due to the construction of the Mosul Dam Lake (Forti et al., 2022). What remains are sporadic surface exposures of conglomerates and marls (Fig. B2; Forti et al., 2021) and three to four terraces levels (Al-Dabbagh and Al-Naqib, 1991; Forti et al., 2021, 2022). North of the Tigris river, in the Selevani Plain, combined action of wind and irregular water action of *wādīs*, have led to the formation of gullies on this depositional glacia (Yacoub et al., 2012; Forti et al., 2021), shaping a badland landscape (Fig. 5B). The Bekhair anticline and its imbricated zone form a structurally homogenous ridge dominated by exposed limestone and sandstone formations (Fig. 4, Forti et al., 2021), which are subject to lift-up process and tectonic action. The foothills on both sides of the ridge,

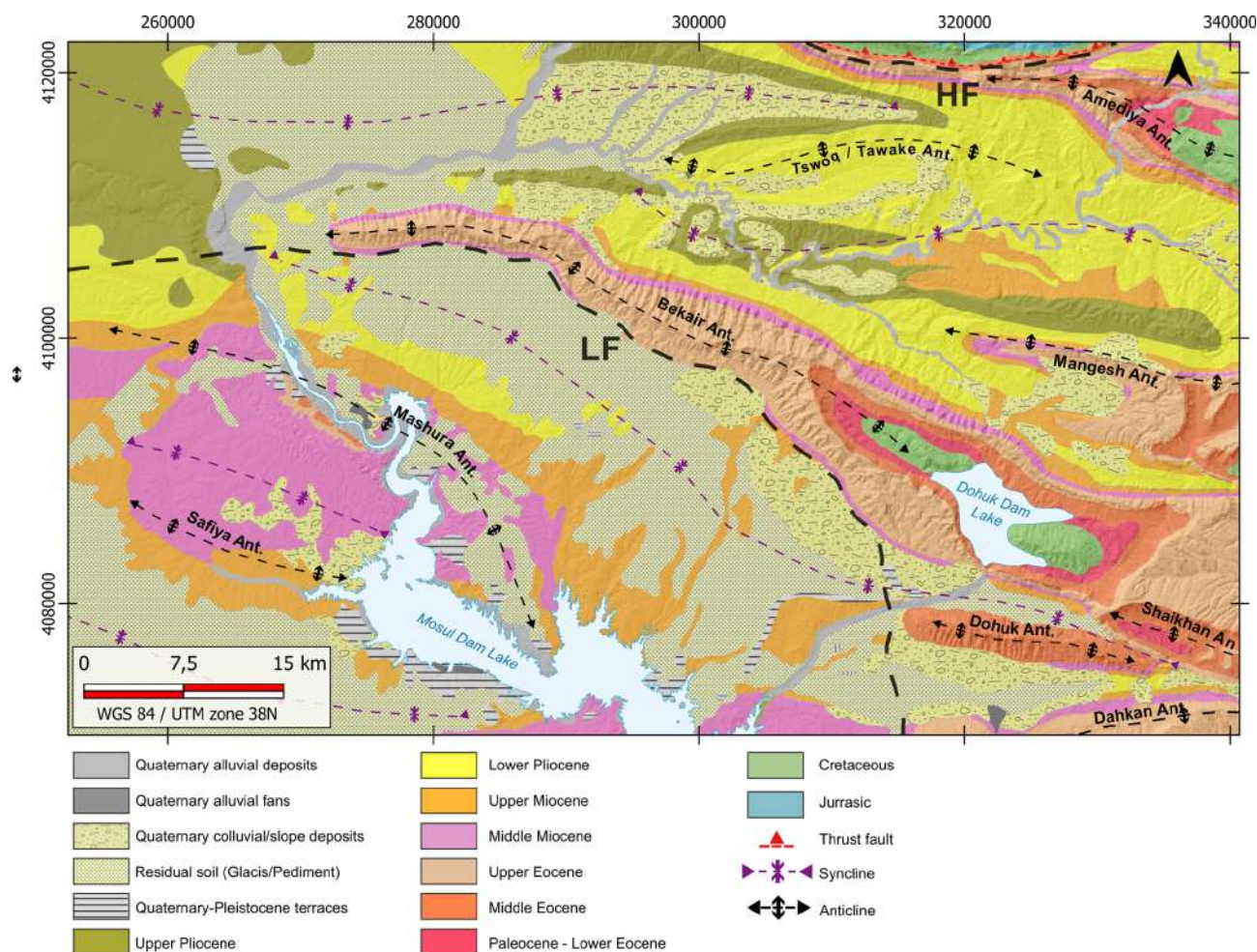


Figure 4. Geological map of the studied area compiled from: Ponikarov and Mikhailov (1986), Sissakian et al. (1995), Isiker et al. (2002), Al-Mousawi et al. (2007), Sissakian and Al-Jiburi (2012), Sissakian and Al-Jiburi (2014) and Doski and McClay (2022) (LF = Low-folded zone; HF = High-folded zone). Realised with QGIS 3.34.5 and Inkscape 1.4.

however, are subject to wind, water erosion, and gravitational processes, resulting in extensive colluvial deposits (Fig. 5C; Sissakian and Abdul Jab'bar, 2014; Sissakian et al., 2015). In the area of the Tswaq anticline and the Little Khabur Valley, the landscape is dominated by sandstone and conglomerate. Soil surface erosional processes are less pronounced in the Little Khabur Valley region due to the protective effect of denser vegetation cover. The Zakho Plain, located within a synclinal structure (Fig. 4), is a flat alluvial area, also less affected by erosion.

Soil mapping in the region was initially carried out in the 1950s and 1960s as part of the Iraq soil mapping project (Buringh, 1957; Altaie, 1968; Altaie et al., 1969). We adapted Buringh's classification to the WRB system (IUSS Working Group WRB, 2022), improved spatial detail using modern satellite imagery (Sentinel 2 ESA, 2022; DEM ESA and Airbus, 2022; and Bing Maps <https://www.bing.com/maps>, last access: 6 September 2024), and completed

the map with unrecorded Regosols and Fluvisols. This resulted in a new regional soil map built on an expert-based knowledge method (Fig. 6). The semi-arid climate, marked by sharp temperature variations and high subsurface CaCO_3 concentrations, has favoured the development of vertic and calcic features in many soil profiles (Abdulrahman et al., 2020). However, significant local variability exists. Soils adjacent to the Tigris River are typically calcic Vertisols, likely due to subsurface marl and conglomerate permeability. The Selevani Plain's glaci deposits are dominated by cambic and gypsic Calcisols (Figs. 5 and 6), with mediumly developed soil horizons, and a soil depth of 100 to 200 cm. North of the Selevani Plain, on the structural ridge and the steep slopes of the Bekhair anticline, soil development is minimal, due to active erosion, resulting in nudilithic Leptosols. On the northern side of the ridge, the Little Khabur Valley and its surroundings are dominated by poorly developed soil such as calcic Cambisols, Regosols and Leptosols, shaped by

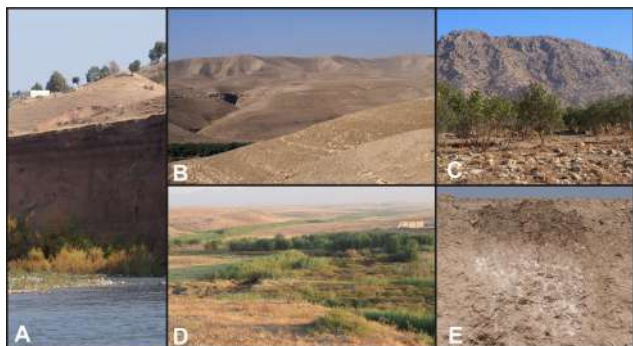


Figure 5. Examples of landscapes (A) Terraces of the Little Khabur (10–11 m) featuring a succession of colluvial and flood deposits. 37°05′14.46″ N 42°56′28.32″ E. (B) Hill and badland landscape on marl formation. *Wādīs* shape this landscape mainly used for grazing. 36°57′16.86″ N 42°28′38.53″ E. (C) Foothills landscape at the base of the Bekhair. Stones are visible on the surface, and olive trees (*Olivae europaeis*) are cultivated in these foothills. Lithosols, Cambisols or Calcisols dominate this landscape. 37°03′50.73″ N 42°34′50.71″ E. (D) *Wādī* landscape with heavily developed vegetation, shrubs and small trees. Fluvisols or Vertisols are usually associated with this environment. 36°54′35.33″ N 42°44′29.64″ E. (E) Calcisol developed on a conglomerate formation. The top 10–15 cm shows an A humic horizon, followed by 5–10 cm of a Bt horizon and a C calcitic horizon at the bottom. 37°01′24.61″ N 42°30′37.11″ E. Realised with Inkscape 1.4.

steep slopes and more erodible parent materials (conglomerate and sandstone), compared to the Simiele Plain. In contrast, the flat, irrigated Zakho alluvial plain, with higher precipitation, supports more developed soils, such as calcic isomeric Kastanozems (Fig. 6). Finally, Fluvisols occur sporadically along the Tigris and Little Khabur floodplains and major *wādīs* channels riverbanks (Figs. 5 and 6), identifiable by their ochric and/or umbric horizons.

2.2 Sampling campaign

The 2022 campaign primarily focused on the Selevani Plain and the Tigris riverbanks, resulting in collecting at 101 sites a total of 445 samples. In 2023, the mission was conducted in the Zakho district, where 21 sites were visited and 87 samples were collected. Due to ongoing violence and conflict between the Kurdistan Workers' Party (PKK) and the Turkish government in the mountainous areas of Zakho (Ertan, 2022), the 2023 survey coverage was reduced for safety reasons. To increase the number of training samples for model calibration (cf. Sect. 2.4.1), 16 additional sites and 29 samples were included from earlier 2017–2018 surveys, and were based on purposive, non-randomised sampling design. The total number of samples collected and analysed with the FTIR methods was 561 from 138 sites (Table 1).

2.2.1 Conditioned Latin hypercube sampling

Sampling design plays a critical role in ensuring that selected locations reflect the spatial and environmental variability of the study area (Brus, 2022). We adopted a conditioned Latin hypercube sampling (cLHS) approach, a method particularly suited to digital soil mapping applications (Minasny and McBratney, 2006; Stumpf et al., 2016). The cLHS method ensures that sampling points are distributed across the full range of values in selected environmental covariates by stratifying feature layers. For each covariate, a set of marginal strata is defined (intervals). The number of marginal strata for each covariate equals the sample size, resulting in p^n a total of marginal strata, where p is the number of covariates and n is the sample (Brus, 2022). Sampling was performed with R 4.4.0 (R Core Team, 2024) using the `clhs` package (Roudier, 2021).

We selected six covariates (Table A1) which represent a broad range of parameters influencing soil variability. These included physical characteristics, underlying geomorphological formations (Forti et al., 2021), potential soil properties and erosion process.

The potential soil properties layer was constructed using spectral indexes (clay minerals, ferrous minerals, rock outcrop, carbonate) derived from climatic and satellite datasets (Copernicus, 2019; EROS, 2020). Erosion risk was modelled using the Revised Universal Soil Loss Equation (RUSLE; Renard et al., 1991), incorporating five key factor: soil erodibility (K), soil coverage (C), topographic effect (LS), rainfall-runoff (R) and erosion control practices (P ; Cossart et al., 2020; Thapa, 2020; Abdi et al., 2023; Mehri et al., 2024). In our RUSLE model (Table C1), we set K as the soil strength factor based on texture and organic carbon values (Kouli et al., 2009), and C was set with values from Morgan (2005) and Swarnkar et al. (2018). Slope length and steepness factor LS were based on Desmet and Govers' method (Desmet and Govers, 1996), while the R factor from Morgan et al. (1984) was used. Finally, the conservation factor P , which could not be observed, has been set to 1 artificially, as suggested by Mehri et al. (2024).

2.2.2 Field measurements

At each site, samples were collected for the top 50 cm using a 3.5 cm $\varnothing$ auger and from depths up to 100 cm with a 2 cm $\varnothing$ auger. The different soil horizons' depths were measured, and colour was determined according to the Munsell soil colour chart. Samples were collected at five depth increments: 0–10, 10–30, 30–50, 50–70 and 70–100 cm. Bulk density was calculated for the topsoil using a 5.3 cm $\varnothing$ ring (Blake and Hartge, 1986). After removing the surface litter and loose sand, the sampling ring was used on the 0–10 cm soil layer. All samples were air-dried at 40 °C for 48 h before sieving at 2 mm for subsequent analysis.

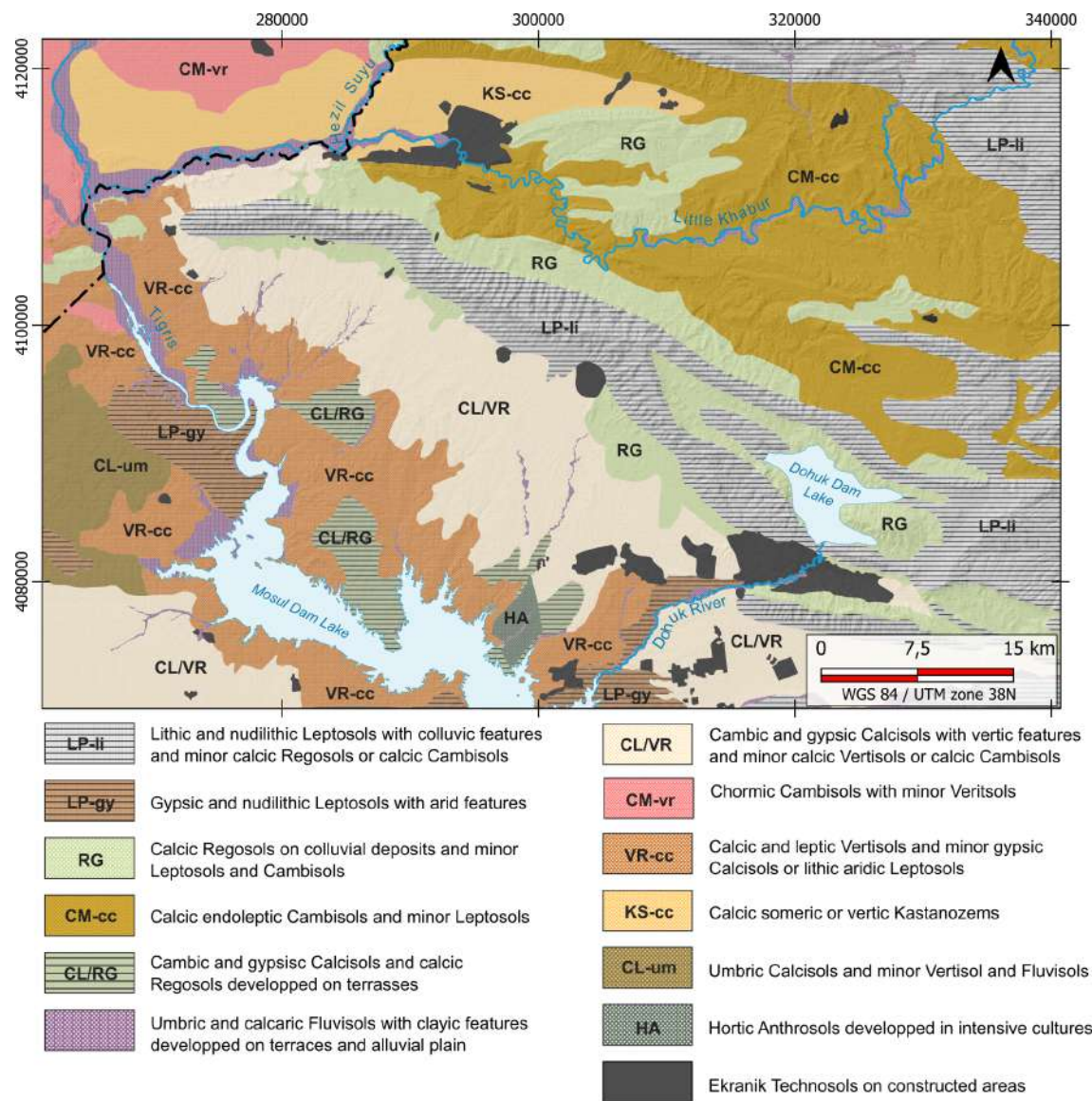


Figure 6. Soil type map based on the IUSS Working Group WRB (2022) classification. Observations come from survey informations and previous work of Buringh (1957), Altaie (1968) and European Soil Bureau Network and European Union (2005). Realised with QGIS 3.34.5 and Inkscape 1.4.

Table 1. Sampling campaigns.

Year	Number of sites	Number of samples	Samples selected for measuring soil properties	Depth explored	Area surveyed
2017–2018	16	29	29	0–10 cm	–
2022	101	445	50	0–100 cm	830 km ²
2023 ¹	21	87	29	0–100 cm	490 km ²
Total for soil properties prediction	138	560 ²	108	0–100 cm	–
Total for DSM	122	531 ³	–	0–100 cm	2280 km ²

¹ Original size of the non-reduced area is 1450 km². ² One sample did have FTIR spectra out of range, and therefore was not used. ³ Samples of the 2017–2018 campaigns were not used for the DSM due to the absence of several depth increments, and the highly clustered locations near archeological sites with modified soil characteristics.

2.3 Laboratory analysis

2.3.1 Mid-infrared spectroscopy

Mid-infrared spectroscopy to measure physical and chemical soil properties has significantly evolved over the past decades (Ng et al., 2022a) and offers reliable results while saving time and resources (Stenberg et al., 2010; Viscarra Rossel et al., 2022). All the 561 soils samples were ground under 1 µm with a *Pulverisette 5/4, classic line* (Fritsh, Idar-Oberstein, Germany) before being pressed into a tablet, mixing 1–1.3 mg of soil and 250 mg of potassium bromide (KBr). The spectra were analysed with a *Vertex 80v* (Bruker OPTIK GmbH, Germany), with a 4 cm⁻¹ resolution, on the 375–4500 cm⁻¹ interval.

The spectra were imported into R 4.4.0 and analysed using the `prospectr` (Stevens and Ramirez-Lopez, 2014) and `simplerspec` (Bauman, 2023) packages. To reduce noise interference, we decided to remove the measurements between 375–499 and 2451–2500 cm⁻¹ intervals and spectra value higher than 2 and lower than -2 (Curran et al., 1996; Ng et al., 2018). One sample was therefore removed due to its low values (< -2). The remaining 560 soil spectra were enhanced by applying three spectral transformations (Ng et al., 2018; Wadoux et al., 2021b; Ludwig et al., 2023), Savitzky-Golay with a polynomial order of 2 and a window size of 11 (SG 2.11), a moving average of 11 and standard normal variate transformation on the SG transformed spectra (SNV-SG). A total of 108 samples were selected for laboratory measurements (Table 2) using Kennard-Stone sampling (Kennard and Stone, 1969), ensuring a high diversity and variability of individuals, based on their spectral data (Ramirez-Lopez et al., 2014).

2.3.2 Soil properties

On these 108 selected samples, seven properties were measured: pH, CaCO₃, N_t, C_t, OC, EC and texture. The pH was measured using a potassium chloride (KCl) solution, with a *ProfiLine pH 3310* and a *WTW SenTix 81 pH* electrode (Fisher Scientific, Strasbourg, France). Carbonate calcium (CaCO₃) content was determined as a percentage using a calcimeter 08.33 (Royal Eijkelkamp, Giesbeek, Netherlands). Total nitrogen (N_t), total carbon (C_t) and total organic carbon (OC) were quantified as percentages with a CNS analyser, *Vario EL III* (Elementar, Hanau, Germany). The electroconductivity (EC) was measured in micro-siemens per centimeter (µS cm⁻¹) using a *Cond 330i/340i* (WTW, Weilheim in Oberbayern, Germany). Texture property was determined as a percentage and measured through wet sieving for sand fraction and a *SediGraph III* for finer fractions (Micromeritics, Norcross, USA). Additionally, we estimated the mean weight diameter in mm (MWD, Eq. 2) based on the texture results.

$$\text{MWD} = X_i \times W_i / 100 \quad (2)$$

where X_i is the mean diameter of i size fraction, and W_i is the weight percentage of i size fraction.

2.4 Models and pre-process

2.4.1 Spectra prediction

The Cubist model is a regression-based machine learning algorithm that extends the ideas of decision trees by combining rule-based predictive models with linear models at the leaves, enhancing both interpretability and predictive accuracy (Quinlan, 1992). This model excels at handling both continuous and categorical data, providing robust predictions even in the presence of complex interactions and non-linear relationships (Kuhn and Quinlan, 2024). Cubist's strength lies in its ability to partition the data space and fit separate linear models to each segment, making it particularly effective for problems with distinct patterns or heteroscedasticity (Wang and Witten, 1996). This model has been applied in a variety of studies for soil property prediction from spectral predictors, such as Viscarra Rossel et al. (2016), Padarian et al. (2020), and Behrens et al. (2022). We tested a Cubist regression model on the four spectral datasets from each of our 560 samples in a Python (Python Software Foundation, 2022) environment using the `Cubist` library (Aselin, 2024). The model used a stratified 5-folds cross-validation and a tuned grid (`n_rules`: 20, 30, 40 and `n_committees`: 5, 10, 15).

2.4.2 Digital soil properties mapping

We based our soil property model on the soil formation factors of the *scorpan* equation (Eq. 1) developed by McBratney et al. (2003). We included 85 covariates (Tables 2 and A1). The remote sensing variables were accessed through an API of Google Earth Engine (<https://earthengine.google.com>, last access: 15 January 2026) on Python, via the `ee` library (Google, 2025), and the different indices computed in R with the `terra` package (Hijmans et al., 2025). The terrain variables were computed on SAGA GIS 9.3.1 (Conrad et al., 2015) based on a filled and filtered DEM from GLO-30 (ESA and Airbus, 2022). All the computation was realised under R 4.4.0 environment (R Core Team, 2024). We included only the 2022 and 2023 samples to produce the DSM for two reasons. First, these campaigns followed a cLHS sampling strategy (cf. Sect. 2.2.1), whereas the 2017–2018 campaigns did not; including them would have reduced the consistency of the sampling design. Second, the 2017–2018 samples lacked depth information and were primarily limited to topsoil. To ensure data comparability, these earlier samples were therefore excluded, resulting in a dataset of 531 samples from 122 sites (Table 1). These samples were included in the DSM as input with: 122 samples for the 0–10 cm depth, 111 for the 10–30 cm increment, 108 for the 30–50 cm depth, 98 for the 50–70 cm increment and 92 for the 70–100 cm depth. We divided the mapping of each variable for each

sults were more conservative with the number of covariates selected (> 60 for each variables), longer in time computing capacities (800 %), and provided lower accuracy scores compared to Boruta selection, for the tested 0–10 cm depth increment.

For each soil depth, we used a 10 k-fold cross-validation repeated 3 times to tune the model and choose the final settings. This resampling strategy allowed us to avoid potential overfitting due to the small size of our training data set (< 100). We trained the models on a specific random forest algorithm called “quantile regression forest model” (QRF), which has shown good performance for digital soil mapping (Varón-Ramírez et al., 2022; Shi et al., 2025). The model was implemented with the *caret* (Kuhn, 2019), and *quantregForest* (Meinshausen and Michel, 2020) packages. We tuned the *mtry* hyperparameters ($1:n$, $by=1$; n being the number of covariates), which corresponds to the number of covariates randomly sampled as candidates at each split, and the minimum node size hyperparameter *nodesize* ($5:31$, $by=5$, Shi et al., 2025), which defines the minimum number of samples required to be at a leaf node. The number of trees was set as default at 500 (Liu et al., 2022). Best hyperparameters were selected based on the lowest RMSE (Table E1), before being applied to produce a final model with the whole training dataset.

Regression trees use a tree-based structure, splitting the data into different nodes. In the end, the model evaluates the leaves and selects those with the best performance. Their specificity in regression is to predict continuous values at the terminal nodes rather than classes, unlike classification trees. Random forests build upon this principle by combining many regression trees grown on bootstrapped samples of the data, which improves prediction performance and stability. Based on this random forest framework, the quantile regression forest (QRF) (Breiman, 2001), tracks each sample’s value at each node, providing a conditional response distribution. This allows the model to produce prediction intervals and to assess accuracy through quantiles (Vaysse and Lagacherie, 2017).

2.4.3 Soil depth mapping

To predict soil depth, we developed a prediction model using remote sensing data and ground-truth control points, collected during surveys to estimate soil depth. By soil depth, we defined the height of the active horizons in the pedogenesis process (IUSS Working Group WRB, 2022, p. 225), the so-called *solum*. The measurement is taken from the top of the *solum* downward to the mineral layer (C horizon) or consolidated rock (R horizon). The soil depth was measured from 0 to 100 cm on the 122 sampling sites; we added 25 zero values (negative sites) from remote sensing imagery observation on bare rock points. Soil depth is mainly determined by climate, terrain, parent material, vegetation, and land uses (Zhang et al., 2021; Liu et al., 2022). Con-

sequently, we used 26 environmental covariates to predict the soil depth (Table A1). Original soil depth data were first square root-transformed, similar to the DSM, we performed the training on a 10 k-fold cross-validation repeated three times. A quantile regression forest model (Meinshausen, 2006) was chosen and implemented in the R 4.4.0 environment (R Core Team, 2024) using the *caret* (Kuhn, 2019) and *quantregForest* (Meinshausen and Michel, 2020) packages. As described above (cf. Sect. 2.4.2), the QRF model is fitted for digital soil mapping. The tuning of the hyperparameters was similar to the DSM with a tuned grid for *mtry*, and *nodesize*, while the number of trees was leave by default at 500. The optimisation of the hyperparameters (Table E1), and training of the final model followed the same procedure as the DSM.

2.4.4 Evaluation criteria

To validate the performance of our models, we evaluated prediction accuracy on for both spectroscopy predictions (Bellon-Maurel et al., 2010; Williams et al., 2017), and DSMs (Lilburne et al., 2024). Prediction accuracy was assessed using complementary metrics capturing systematic error, error magnitude, and the strength of the relationship between predicted and observed values. Bias, was expressed as the mean error (ME, Eq. 3), used to quantify systematic over, or underestimation by the model. Values closer to zero indicate a lower systematic bias in the predictions. The root mean square error (RMSE, Eq. 4) was selected as the primary error metric, as it quantifies the average magnitude of prediction errors and indicates how far model predictions deviate from the observed values. To characterize the extent to which the model explains the variability of the observed data, we reported the coefficient of determination (R^2 , Eq. 5), which reflects the proportion of variance in the observations explained by the predictions. An R^2 value of 1 indicates that the model explains 100 % of the observed variability. For spectroscopy-based predictions, we additionally report the ratio of performance to interquartile range (RPIQ, Eq. 6), which normalizes the prediction error by the interquartile range of the observed values. This metric is commonly used in infrared spectroscopy to facilitate comparisons across skewed response variables, this enhanced the reliability of spectroscopic prediction models (Bellon-Maurel et al., 2010). Spatial models, such as DSMs, are prone to various sources of error, which should be quantified and represented spatially as prediction uncertainty (Schmidinger and Heuvelink, 2023; Lilburne et al., 2024). We evaluated the prediction uncertainty, for the quantile regression forest models of the DSM and soil depth, using the prediction interval coverage probability (PICP, Eq. 7). The PICP correspond to the probability that all observed values fall within their prediction intervals (Shrestha and Solomatine, 2006; Malone et al., 2011), here set at the 90 % level. Values of PICP under 0.9 suggest an overestimation of the uncertainty whereas a PICP above 0.9

indicates it was underestimated (Poggio et al., 2021). However, recent research in DSM has shown that PICP alone does not account for the distribution of the predictions inside the interval and can also be biased due to the PI boundaries (Schmidinger and Heuvelink, 2023). Other metrics can overcome these issues by capturing the full coverage of the reliability of a regression model such as quantile coverage probability (QCP) and probability integral transform (PIT). Their implementation, while simple, does rely on an independent test set and were limited due to our workflow based on the *caret* package. These limitations of the PICP have to be taken into account while reading at the quantification of the uncertainty.

$$ME = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (3)$$

where n is the number of observations, y_i the observed value for i , and $\hat{y}_i$ the predicted value for i .

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where n is the number of observations, y_i the observed value for i , and $\hat{y}_i$ the predicted value for i .

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where y_i the observed value for i , $\hat{y}_i$ the predicted value for i , and $\bar{y}$ is the mean of observed values.

$$RPIQ = (Q3 - Q1)/RMSE \quad (6)$$

$$PICP = \frac{1}{v} \text{count } j \quad i = PL_i^L \leq obs_i \leq PL_i^U \quad (7)$$

where v is the number of observations, obs_i is the observed value for i , PL_i^L is the lower prediction limit for i , and PL_i^U is the upper prediction limit for i .

3 Results

3.1 Laboratory measurements

Observations from the different laboratory samples show a large variability in soil property distributions (Table 3, Fig. 7). The pH values ranged from 6.93 to 8.2, with a mean of 7.3 ± 0.2 , indicating a slightly alkaline soil environment. The CaCO_3 content varied widely, from 3.61 % to 84.27 %, with an average of 30 ± 12 %, suggesting significant differences in carbonate content across the samples. Overall, only two samples had CaCO_3 content below 10 %, which indicates the strong relationship between soils and their parent material, mainly limestone, in this aridic environment. Total nitrogen (N_t) ranged from undetectable levels to 0.67 %, with a mean of 0.12 ± 0.1 %, with the higher

values concentrated in the upper 0–10 cm soil depth (20 on 20 of the highest measurements). While organic carbon (OC) content was generally low, with a mean of 1.1 ± 0.9 %, the total carbon (C_t) content was higher, with a mean of 4.8 ± 1.7 %, showing values approximately 360 % higher. This pronounced difference between organic and inorganic carbon is well known for aridic and semi-aridic environments (Zamanian et al., 2016). Electrical conductivity (EC) values included some outliers above $500 \mu\text{S cm}^{-1}$ ($n = 5$) – likely due to laboratory manipulation – explaining the high variability characterized by a standard deviation of $158 \mu\text{S cm}^{-1}$. The mean weight diameter (MWD) of soil aggregates was higher mostly in the upper part of the soil profile (17 of the 20 highest values). Soil texture was predominantly silty-clay and silty-clay-loam, with average sand, silt, and clay contents of 19.4 ± 15.3 %, 44 ± 10.5 %, and 36.5 ± 14.4 %, respectively. Clay presents a particularly homogenous distribution pattern (Fig. 7), indicating a high variability among the samples collected. The upper depth increments (0–10 and 10–30 cm) showed slightly sandier textures than the lower layers (14 of the 20 highest measurements).

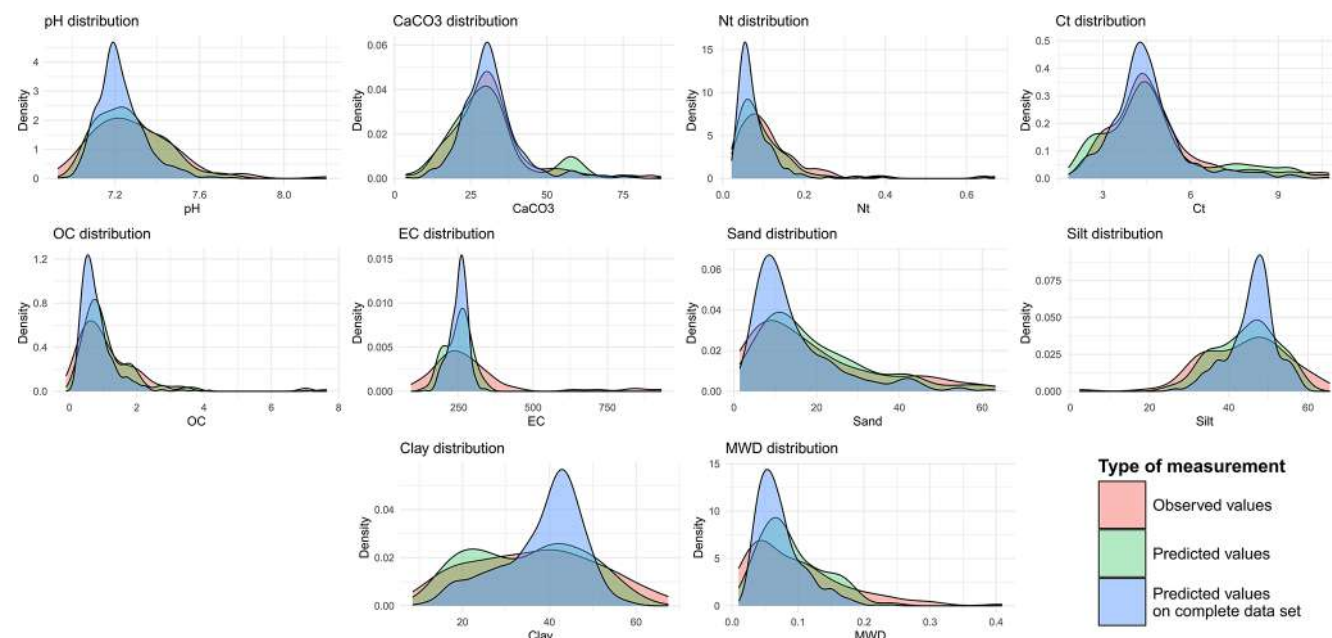
3.2 Soil properties spectra prediction

The SNV-SG transformed spectra provided the best performance for six out of the ten soil properties, while SG-SNV and raw spectra were optimal for two properties each (Table 4). Model performance varied substantially depending on the soil property. For pH, predictions showed a slight underestimation bias ($ME = -0.11$) and a moderate prediction error ($RMSE = 0.14$). The moderate R^2 (0.44) indicates a limited ability to reproduce the observed variability. Predicted values displayed a narrower range than the observations and tended to concentrate around the mean (Fig. 7). In contrast, CaCO_3 , N_t , and C_t were predicted with higher accuracy. These properties showed minimal bias and relatively small prediction errors ($RMSE = 3.88$ %, 0.04 %, and 0.5 %). Their high R^2 values (0.90, 0.83, and 0.91, respectively) indicate that a large proportion of the observed variability was captured by the models. For CaCO_3 in particular, the low bias is notable given the wide range of observed values (Table 3), and the predicted distributions closely matched the observed ones (Fig. 7). The interpretation of N_t remains more cautious due to the strong skewness of its observed distribution. Organic carbon predictions showed a slight overestimation bias ($ME = 0.17$) and moderate error ($RMSE = 0.6$ %). Given the high skewness of the observed values, this performance can be considered acceptable, although the R^2 (0.61) indicates that part of the variability remains unexplained. Regarding the electrical conductivity, the strong underestimation bias ($ME = -5.93 \mu\text{S cm}^{-1}$), large $RMSE$ ($59.22 \mu\text{S cm}^{-1}$), and low R^2 (0.29) indicate weak predictive performance. This suggests that EC is not well captured by the spectral information. The high variability and skewness of the observed EC values likely contributed to this difficulty. Pre-

Table 3. Descriptive statistics of soil properties observed and predicted.

	pH	CaCO ₃ [%]	N _t [%]	C _t [%]	OC [%]	EC [μS cm ⁻¹]	MWD [mm]	Sand [%]	Silt [%]	Clay [%]	Total texture [%]
Observed values											
Minimum	6.93	3.61	—*	1.89	0	90	0.01	2.24	2.3	8.5	98.9
Maximum	8.2	84.27	0.67	10.75	7.65	932	0.41	63.2	65.5	67.4	101.3
Mean	7.3 ± 0.2	30 ± 12	0.12 ± 0.1	4.8 ± 1.7	1.1 ± 0.9	287 ± 158	0.1 ± 0.07	19.4 ± 15.3	44 ± 10.5	36.5 ± 14.4	99 ± 0.3
1st quartile	7.14	23.9	0.07	3.76	0.49	200.7	0.04	7.65	36.1	25.6	99.7
3rd quartile	7.4	34.3	0.14	5.19	1.55	311	0.13	27.94	51.1	47.2	100
Skewness	1.23	1.47	3.16	1.49	3.23	2.43	1.42	1.13	−0.55	0.06	0.47
Predicted values											
Minimum	6.96	3.82	0.022	1.82	0.12	136.6	0.018	1.42	25.53	10.96	77.74
Maximum	7.75	87.53	0.64	10.69	7.02	345.36	0.241	62.4	58.92	59.06	116.29
Mean	7.23 ± 0.1	31.5 ± 10	0.078 ± 0.05	4.5 ± 1.3	0.83 ± 0.57	250 ± 30.2	0.07 ± 0.04	14.7 ± 10.8	45.9 ± 5.8	38.4 ± 9.4	99.1 ± 4.9
1st quartile	7.16	25.8	0.05	3.82	0.49	231.8	0.05	7.43	42.9	33.2	96.5
3rd quartile	7.28	35	0.09	4.9	0.9	269.9	0.08	18.09	49.3	44.8	102
Skewness	1.03	1.59	4.58	1.38	3.89	−0.43	1.28	1.71	−0.61	0.75	−0.25

* Device could not measure concentration below 0.03.

**Figure 7.** Density plot of the laboratory measurements, the predicted values, from the Cubist model, for the laboratory samples and the predictions of all samples. Realised with R 4.4.0.

dicted EC values were strongly concentrated in the lower range compared to the observations, as reflected by their smaller standard deviation and distribution pattern (Fig. 7). For MWD, predictions showed a slight underestimation bias and moderate error ($ME = -0.018$ mm, $RMSE = 0.05$ mm). However, the model failed to reproduce the upper range of observed values (Table 3). The predicted distribution showed reduced variability, with high MWD values systematically underrepresented. This indicates limited ability of the model to capture extreme aggregate stability conditions. Regarding soil texture, clay content was generally underestimated ($ME = -3.57$ %), whereas sand and silt showed slight over-

estimation biases ($ME = 2.04$ %, 0.19 %). Prediction errors were similar for all three fractions, but a good explanatory power was obtained for sand and clay ($R^2 = 0.81$ and 0.72), whereas silt showed weaker performance ($R^2 = 0.38$). The silt model particularly failed to represent the lower and higher ends of the distribution, resulting in a concentration of predicted values near the mean (Fig. 7). Overall, predicted texture fractions exhibited lower variability than the laboratory measurements, especially for silt.

Although the general patterns observed in laboratory measurements were also reflected in the predictions (Table 3), some differences remained. Predicted soil textures were

Table 4. Cubist model evaluation metrics.

	pH	CaCO ₃	N _t	C _t	OC	EC	MWD	Sand	Silt	Clay
Spectra	Raw	SG 2.11	SNV-SG	SNV-SG	SNV-SG	SNV-SG	Raw	SNV-SG	SNV-SG	SG 2.11
ME	−0.11	−0.05	−0.005	0.04	0.17	−5.93	−0.018	2.04	0.19	−3.57
RMSE	0.14	3.88	0.04	0.5	0.6	59.22	0.05	6.64	8.27	7.64
R ²	0.44	0.90	0.83	0.91	0.61	0.29	0.47	0.81	0.38	0.72
RPIQ	1.24	2.77	1.71	2.72	1.15	0.93	1.25	2.66	0.98	2.43

predominantly classified as silty-clay and silty-clay-loam (Fig. 8), with slightly lower clay and higher sand contents compared to observations. The silt and clay classes were proportionally less represented in the predictions than in the observed values.

3.3 Digital soil properties mapping

Feature selection using *Boruta* substantially reduced the number of covariates by 75 %–93 % (Table 5). Models for deeper soil layers selected fewer covariates than those for the upper layers. The most influential predictors were Landsat 8 SWIR bands and SAVI indices. Sentinel-2 SWIR bands, the EVI index, DEM-derived variables, potential evapotranspiration, wind speed, and solar radiation were also important. Overall, Landsat 8 products performed slightly better than Sentinel-2. These covariates mainly correspond to the *s* (soil), *o* (organisms), and *r* (relief) components of the *scorpan* framework.

The models generally showed low bias (Table 6), including for EC, where the largest ME (4.83) occurred only in the 50–70 cm layer. Sand and clay were the only variables showing systematic bias, with sand tending to be underestimated and clay overestimated. In contrast to the Cubist predictions, silt performed better than the other texture fractions in terms of bias, although clay had slightly lower RMSE. Overall explanatory power was relatively low, with *R*² values mostly below 0.5. Some depth-dependent variability in accuracy was observed. For instance, CaCO₃ and clay prediction errors increased with depth, whereas OC showed the opposite pattern, with lower RMSE at greater depths (Table 6). Prediction interval coverage was generally close to the nominal 90 % level, with PICP values mostly ranging between 85 %–95 %, indicating slight overestimation of uncertainty. A few exceptions were observed: pH at 50–70 cm, CaCO₃ at 10–30 cm, and C_t at 50–70 cm. Overall, this suggests that the models were reasonably well calibrated. PICP was not reported for texture fractions. Because log-ratio transformations are nonlinear, nominal coverage may not be strictly preserved after back-transformation.

The spatial distribution of soil properties reveals a clear contrast between the southern and western areas, Selevani and Zakho plains, Tigris riverbanks, and the northern and

eastern zones, Bekhair anticline and Little Khabur Valley (Figs. 9, 10, 11, 12 and 13).

Soil pH remained relatively stable with depth and tended to be higher in the anticline and valley areas, except at 50–70 cm where spatial variability increased (Fig. 12). CaCO₃ showed a generally uniform spatial distribution, with localized high values along the southern anticline foothills. N_t concentrations were highest in the topsoil (0–30 cm), particularly in the anticline and north-western areas (Figs. 9, and 10). C_t remained relatively consistent across the upper layers, with maxima near the anticline, while the 70–100 cm layer displayed additional hotspots in the Selevani Plain and appeared strongly influenced by water dynamics (Fig. 13). OC showed depth-dependent spatial shifts, with maxima located in the anticline in the surface layer and in the Little Khabur Valley at intermediate depths. EC was highest in the Selevani and Zakho plains, with secondary peaks in the Little Khabur Valley. Texture patterns indicated higher sand content near the surface and increased silt at 50–70 cm. Sand dominated in anticline and valley areas, whereas silt and clay were more prevalent in the plains. MWD broadly followed texture patterns, with higher values where sand content was greater. Uncertainty maps showed higher values in the Bekhair foothills, Little Khabur Valley, and along the Tigris River (Figs. 9, 10, 11, 12 and 13). For ALR-transformed texture variables, uncertainty values are not directly interpretable due to the nonlinear transformation and should be considered only for relative spatial comparison.

3.4 Soil depth mapping

The soil depth model explained about 39 % of the observed variability, with an RMSE of 30.76 cm. Uncertainty was slightly under-calibrated, as indicated by a PICP of 98.87 % (Table 5). DEM-derived predictors were dominant, with four of the five most influential variables accounting for roughly 25 % of explained variance (Table 6). Spatial predictions revealed two main patterns: shallow soils in mountainous areas and the Little Khabur Valley, and deeper soils in the Selevani and Zakho-Cizre-Silopi plains (Fig. 14).

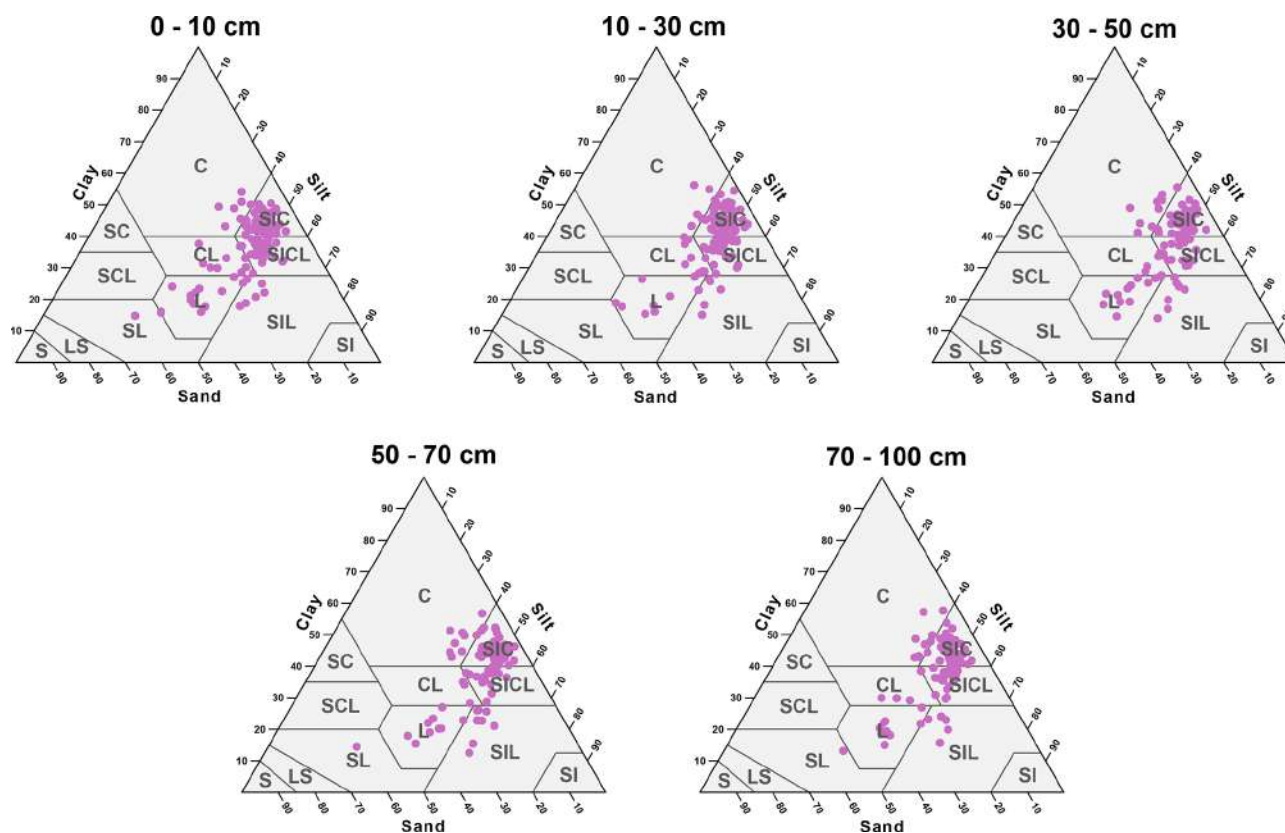


Figure 8. Particle size soil predictions, representation in a triangle diagram, according to USDA classification system (IUSS Working Group WRB, 2022), for each depth increments. C: clay; SC: sandy clay; SCL: sandy clay loam; CL: clay loam; SIC: silty clay; SICL: silty clay loam; L: loam; SIL: silty loam; SI: silt; SL: sandy loam; LS: loamy sand; S: sand. Realised with R 4.4.0.

4 Discussion

4.1 Spatial interpretations of soil properties distribution

The contrasting distribution of soil properties across the study area (Figs. 9, 10, 11, 12 and 13) can be attributed mostly to landscape differences. The Selevani and Zakho plains, together with the Tigris alluvial valley, largely comprise flat areas along rivers and depressions, which are mainly characterized by sedimentation processes, for example the deposition of flood sediments on river banks and terraces or the filling of depressions with erosion material. In contrast, the little Kabur Valley experiences stronger erosional processes with the formation of rills and gullies, and the Bekhair and Zagros anticlines are subject to uplift at the geological timescale.

The spatial correlation between pH and EC is particularly evident. Neutral pH values and elevated EC are mainly associated with Kastanozems in the Zakho Plain and calcareous Calcisols or Vertisols in the Selevani Plain. Organic carbon and N_t concentrations are higher in the Little Khabur Valley and mountainous areas, likely linked to denser vegetation cover (*Quercus brantii*, *Quercus boissieri*; Zohary, 1973, pp. 183–190) and a cooler climate regime. $CaCO_3$ content

appears to reflect the lithological composition of the Selevani Plain, particularly the carbonate-rich sandstone Injana Formation. Consequently, total carbon (C_t), closely follows $CaCO_3$ distribution, with the majority of carbon storage deriving from inorganic forms in deeper horizons (Moharana et al., 2021). Textural patterns reveal finer soil fractions (clay and silt) prevailing in the flat plains and the southern foothills of the Bekhair anticline. In contrast, coarse-textured soils with higher sand content are found in more eroded and badlands areas, such as the top of the anticline and the Little Khabur Valley.

Concerning the uncertainty of the predictions, higher values were observed in the Bekhair anticline and Little Khabur Valley, which are areas with greater geomorphological variability and microrelief. This is consistent with the fact that the terrain predictors derived from DEM, may not fully capture the complex topography of these areas. Additionally, the presence of active *wādīs* channels and badlands in these regions may contribute to increased spatial heterogeneity, further challenging the model's predictive performance.

Table 5. Number of covariates selected with *Boruta* and top five factors for every soil property.

Variable	Total number of covariates selected (For each soil depth increment)	Top covariates selected (Top five covariates present in several depth models)
pH	14; 17; 14; 8; 15	Landsat 8 SWIR2; TPI; Landsat 8 SAVI; Landsat 8 SWIR1; MODIS NDVI
CaCO ₃	12; 12; 7; 9; 9	Landsat 8 SWIR1; Landsat 8 SWIR2; Precipitations; Sentinel 2 SWIR2; Sentinel 2 SWIR1
N _t	21; 18; 7; 7; 7	PET; Solar radiations; Channel network base level; DEM; Flow accumulation
C _t	19; 11; 7; 10; 6	Landsat 8 Panchromatic; Landsat 8 SWIR1; Landsat 8 SWIR2; Solar radiations; Sentinel 2 SWIR1
OC	21; 19; 10; 11; 14	Positive openness; Channel network base level; DEM; PET; Wind
EC	11; 18; 13; 10; 9	MODIS red; Landsat 8 SWIR2; Geomorphology; Sentinel 2 EVI; Sentinel 2 SWIR1
MWD	20; 21; 19; 17; 14	Landsat 8 EVI; Landsat 8 SAVI; MODIS SAVI; PET; Sentinel 2 NDVI
alr_sand	18; 22; 14; 17; 18	Landsat 8 EVI; Landsat 8 SAVI; Landsat 8 LSWI; MODIS TVI; Sentinel 2 EVI
alr_silt	20; 20; 13; 12; 9	Solar radiations; Landsat 8 NDMI; Landsat 8 LSWI; Landsat 8 NIR; Landsat 8 SWIR1
Soil depth	26*	Slope; DEM; MrRTF; MrVBF; Wind speed

* No selection of the covariates was performed (see Supplement)

4.2 Soil depth distribution

Two distinct patterns emerged in the spatial distribution of soil depth (Fig. 14). Shallow soils are prevalent in foothill and mountainous regions, as expected (Patton et al., 2018), but are also common in badlands and along active *wādīs* channels, and riverbanks of the Tigris and Little Khabur.

In contrast, deeper soils were mapped in the Zakho Plain and on the plateaus of the Little Khabur Valley. These patterns may be explained by flat topography, active depositional process, and in the case of the plateaus, by denser vegetation zones typical of the Kurdo-Zagrosian climate formation.

The soil depth uncertainty map confirms the model's robustness in predicting both shallow and deep profiles. The highest uncertainty was found near major *wādīs*, badlands, and along the foothills – areas with greater geomorphological variability and microrelief. The top-ranked covariates, mainly derived from DEM, confirm the well-established relationship between topography and soil depth (Patton et al., 2018; Yan et al., 2018; Liu et al., 2022).

4.3 SoilGrids 2.0 product comparison

To assess model performance, we compared our results with the global SoilGrids 2.0 product (Poggio et al., 2021), focusing on pH, OC, N_t, and texture attributes (Table 7), for three generalised depth intervals (0–30, 30–60/70, and 60/70–100 cm). We scaled our predictions to match the 250 × 250 m resolution of SoilGrids 2.0, with a *bilinear* method from the *terra* package (Hijmans et al., 2025).

Our models predicted higher values of OC and N_t, with respective increases of ≈ 1000 % and 300 % over those from SoilGrids 2.0. Predicted sand values were also higher (by ≈ 25 %), while clay values were slightly lower (by ≈ 15 %). The differences in silt and pH values were negligible, with a 3 % higher value for silt and 5 % lower for pH in our predictions compared to SoilGrids 2.0.

The standard deviation of the SoilGrids 2.0 product is smaller, except for the pH, than for our prediction models (Table 7). This shows a narrower distribution of values, likely due to the wide range of input data used for the SoilGrids 2.0. The diversity of soil types and input data at the global scale makes the SoilGrids 2.0 model respond relatively homoge-

Table 6. Soil properties mapping models evaluation metrics.

Variable	Metric	0–10 cm	10–30 cm	30–50 cm	50–70 cm	70–100 cm
pH	ME	0.00	−0.01	0.00	−0.01	−0.01
	RMSE	0.09	0.09	0.09	0.11	0.09
	R^2	0.29	0.30	0.37	0.33	0.29
	PCIP	89.87	86.17	86.31	84.17	86.56
CaCO ₃	ME	−0.10	−0.41	−0.90	−1.26	−1.80
	RMSE	6.86	7.37	9.86	9.69	9.05
	R^2	0.30	0.32	0.27	0.33	0.20
	PCIP	86.45	84.07	87.38	89.33	85.58
N _t	ME	−0.01	0.00	0.00	0.00	0.00
	RMSE	0.05	0.03	0.02	0.02	0.02
	R^2	0.45	0.36	0.29	0.22	0.24
	PCIP	87.74	87.75	87.26	86.87	89.53
C _t	ME	−0.14	−0.04	−0.20	−0.13	−0.11
	RMSE	1.10	1.01	1.18	1.03	1.06
	R^2	0.32	0.39	0.33	0.34	0.23
	PCIP	90.15	89.24	89.81	83.61	84.24
OC	ME	−0.11	−0.07	−0.03	−0.04	−0.04
	RMSE	0.54	0.35	0.24	0.24	0.21
	R^2	0.49	0.37	0.37	0.44	0.43
	PCIP	87.79	88.18	88.46	85.45	90.76
EC	ME	1.82	0.58	2.92	4.83	1.11
	RMSE	26.02	22.74	24.03	28.01	23.64
	R^2	0.23	0.30	0.31	0.22	0.30
	PCIP	84.68	91.77	89.32	90.83	87.04
MWD	ME	0.00	0.00	−0.01	−0.01	−0.01
	RMSE	0.03	0.03	0.03	0.03	0.03
	R^2	0.39	0.44	0.37	0.39	0.29
	PCIP	88.76	86.65	86.72	88.70	85.04
alr_sand	ME	−0.10	−0.12	−0.08	−0.10	−0.13
	RMSE	0.67	0.70	0.78	0.81	0.72
	R^2	0.44	0.47	0.32	0.41	0.45
	PCIP	87.03	88.74	88.06	89.36	85.41
alr_silt	ME	−0.02	−0.04	−0.04	−0.02	−0.02
	RMSE	0.25	0.26	0.30	0.32	0.28
	R^2	0.35	0.34	0.31	0.32	0.40
	PCIP	90.41	88.89	89.03	89.11	87.95
Sand	ME	−2.91	−3.35	−2.93	−3.26	−3.08
	RMSE	11.17	9.90	11.15	11.98	10.97
	R^2	0.43	0.49	0.28	0.35	0.44
	PCIP	–	–	–	–	–
Silt	ME	1.09	1.01	0.77	1.51	1.30
	RMSE	8.65	7.76	9.13	10.02	8.54
	R^2	0.35	0.43	0.29	0.25	0.38
	PCIP	–	–	–	–	–
Clay	ME	1.82	2.34	2.17	1.75	1.78
	RMSE	6.85	6.88	7.68	8.04	7.54
	R^2	0.38	0.34	0.29	0.30	0.40
	PCIP	–	–	–	–	–
Depth	ME			3.27		
	RMSE			30.76		
	R^2			0.39		
	PCIP			98.87		

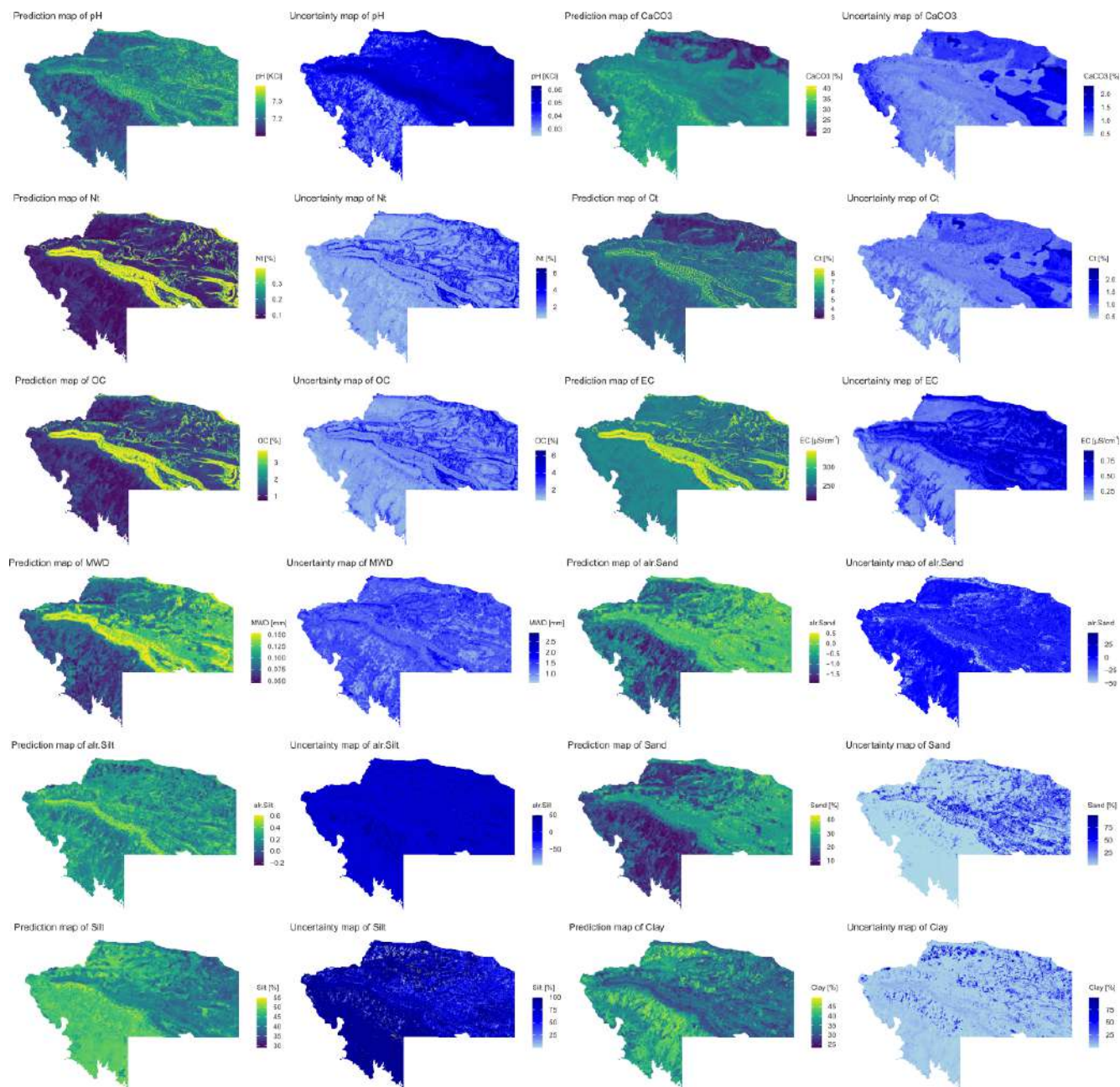


Figure 9. Prediction and uncertainty maps for the 0–10 cm depth increment. Realised with $R^2 = 0.4$.

neously at the regional scale. Furthermore, the SoilGrids 2.0 product shows a more skewed distribution for OC and N_t , with a higher concentration of values near the lower end of the distribution, which is consistent with the known underestimation of these properties in global models (Shi et al., 2025).

We also compared the evaluation metrics of our predicted values with those obtained from SoilGrids 2.0 on an independent data set (Table F1). Before training the models on a full data set, we splitted the data retaining 20 % for test and

80 % for training. The models were trained in similar conditions as our main prediction models (cf. Sect. 2.4.2), before evaluation were computed on the independent test set. Overall, our models outperformed SoilGrids 2.0 across all evaluation metrics for pH, N_t , OC, sand, silt, and clay at all depth intervals. The only exception was the sand QRF model RMSE score at the 10–30 cm depth interval, which was slightly higher for the SoilGrids 2.0 model at the 15–30 cm interval.

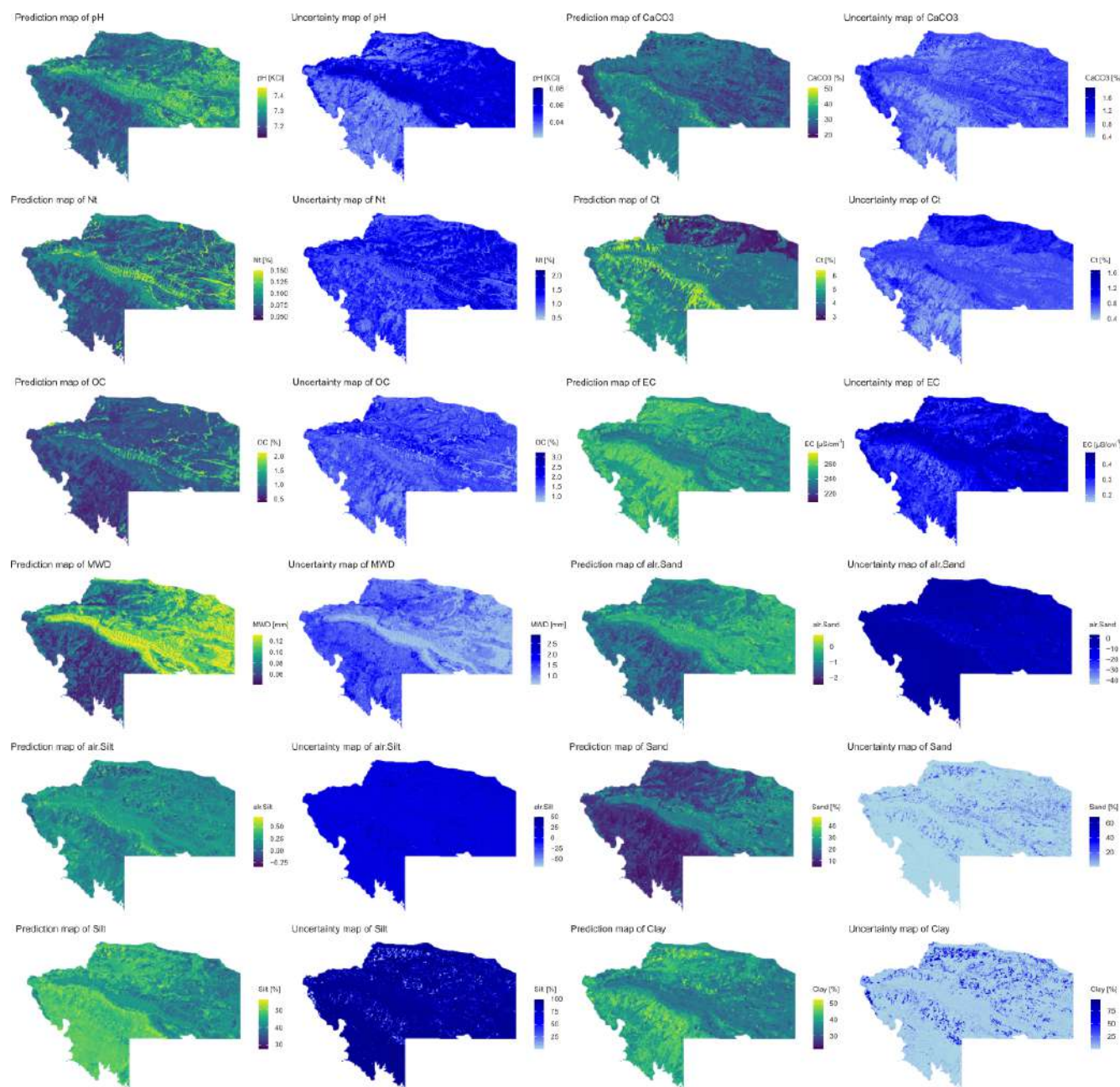


Figure 10. Prediction and uncertainty maps for the 10–30 cm depth increment. Realised with $R = 4.4 \times 10^{-1}$.

Bivariate comparison maps (Fig. 15) indicate that our model yields higher values of pH, N_t, OC, and sand in the Selevani Plain, while SoilGrids 2.0 shows higher values for the same properties in the Little Khabor Valley and the Bekhair anticline. The spatial patterns of silt and clay are inversely distributed. Areas of similarity between the models include pH in the eastern part of the Selevani Plain and Zakho area, OC in the Selevani Plain sporadically and eastern part of the Zakho Plain. Sand values are similar in the southern foothills of the Bekhair anticline while clay present similarity in the

eastern part of the Little Khabor Valley. Due to the use of different algorithms for each depth increment, a direct ensemble model comparison with SoilGrids 2.0 (Varón-Ramírez et al., 2022) was not feasible.

4.4 Data quality, limitations, and future applications

The laboratory-measured soil properties and their corresponding FTIR spectra constitute a valuable and reusable dataset that can contribute to improving predictive model performance over time (Viscarra Rossel et al., 2016;

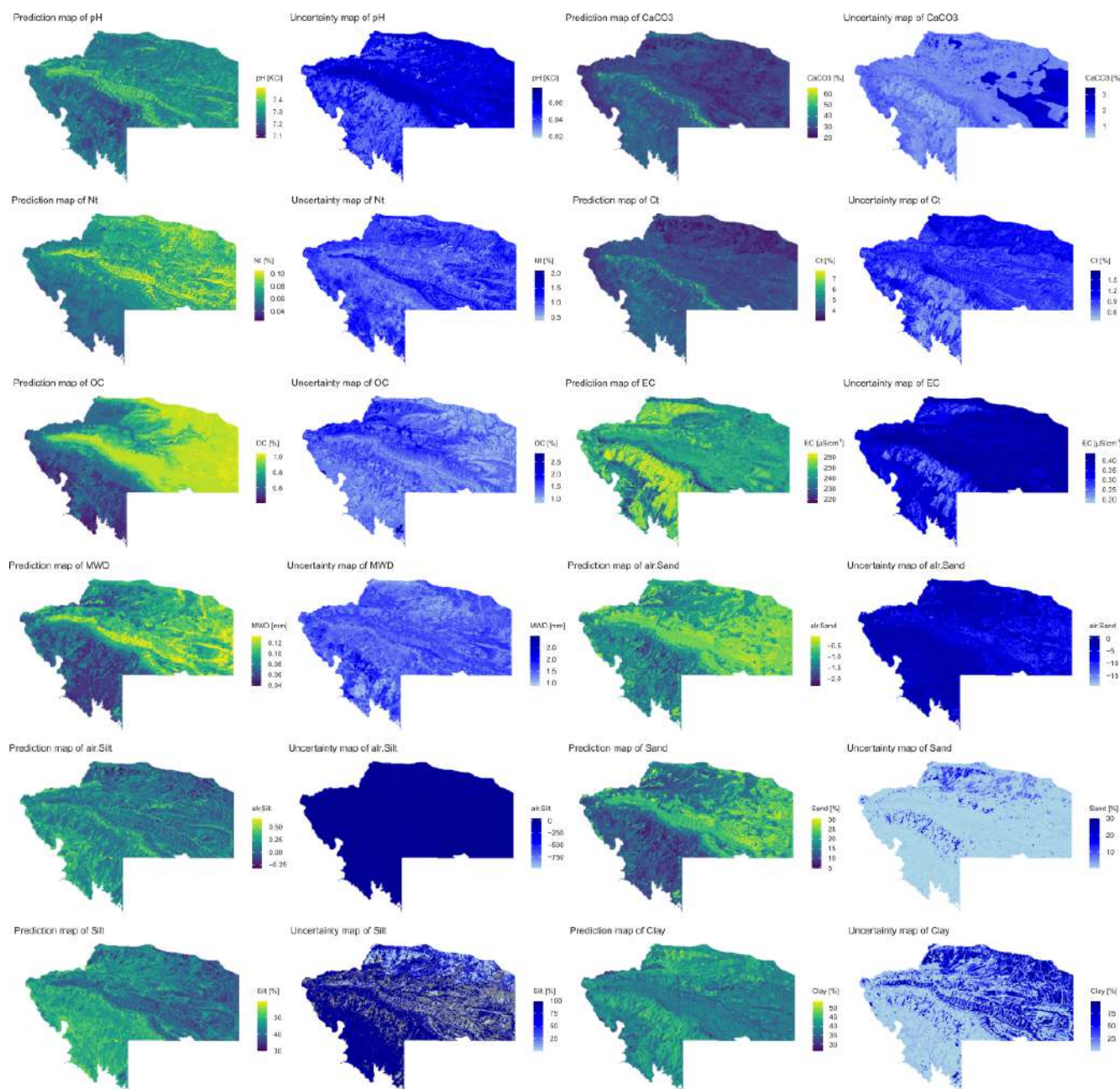


Figure 11. Prediction and uncertainty maps for the 30–50 cm depth increment. Realised with $R = 0.4 \pm 0.0$.

Safanelli et al., 2025). Assigning reliability categories to FTIR-based predictions is a common practice to help users interpret results (Wadoux et al., 2021a), and frameworks such as those proposed by Ng et al. (2022b) have supported this effort. However, such categories are not absolute and can vary depending on the study context, input data quality, and intended application. Therefore, although these categories (A–D) provide a useful reference framework, they were not formally applied in this study. Considering the results for the ten predicted properties (Table 4), CaCO_3 , C_t , sand, and

clay showed comparatively strong predictive performance, with low errors and good explanatory power, suggesting that these models may be transferable to similar contexts. In contrast, pH, N_t , OC, and MWD showed more moderate performances. For OC, this likely reflects the high variability and skewness of the observations, whereas for pH, N_t , and MWD it may indicate weaker relationships with spectral information. EC and silt predictions presented distinct challenges, consistent with known limitations of FTIR-based prediction (Hobley and Prater, 2019; Ng et al., 2022b). In addition, for

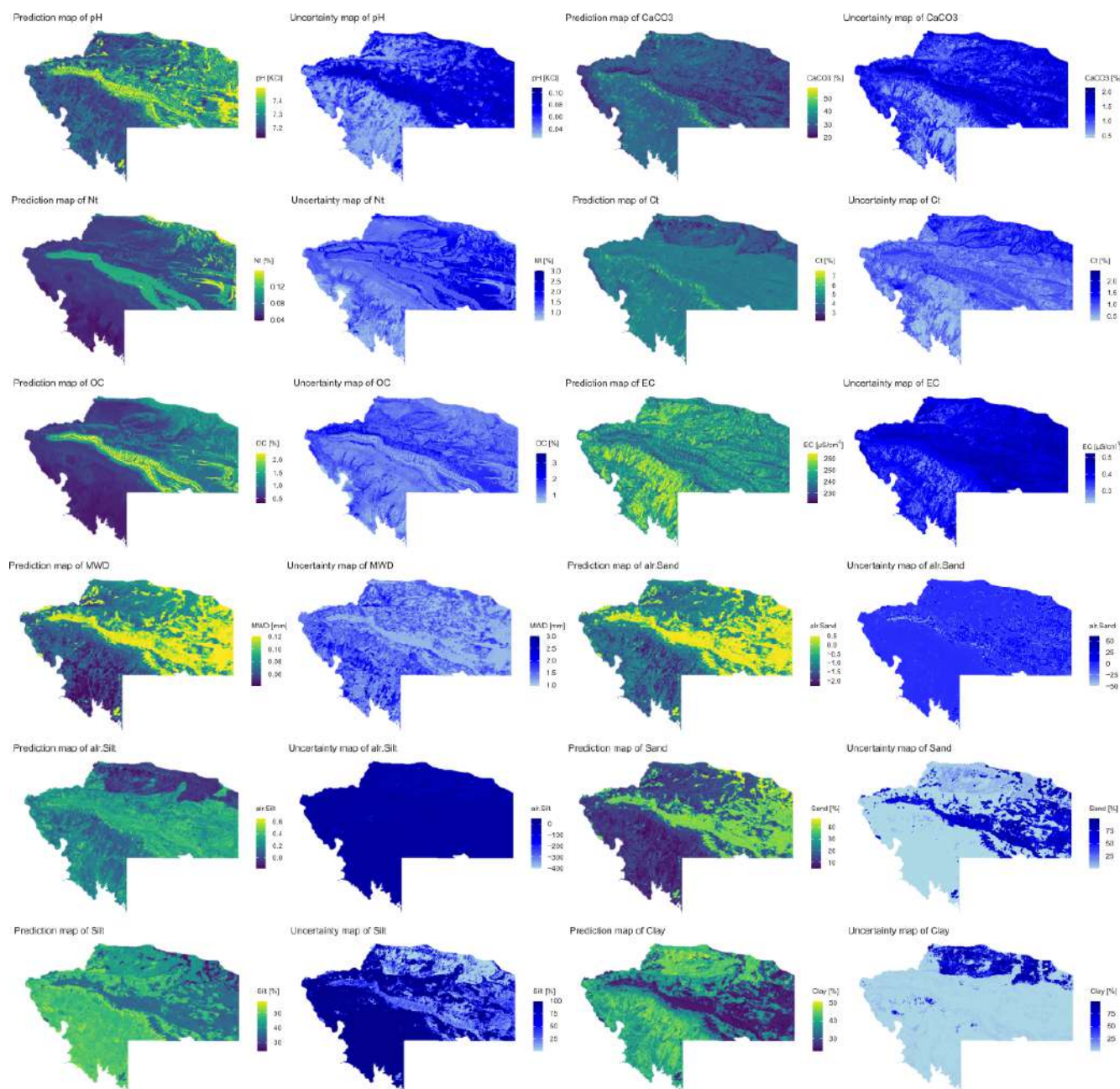


Figure 12. Prediction and uncertainty maps for the 50–70 cm depth increment. Realised with R 4.4.0.

the EC, the high variability and skewness of the measured values likely limited the model's ability to capture consistent patterns. Concerning the silt, lower performance may relate to the indirect nature of its spectral response, which is typically more difficult to detect than clay or sand (Janik et al., 2016; Lacerda et al., 2016). Additionally, for the silt, we used MIR spectra and Cubist model, while previous studies have reported improved performance when using vis-NIR data or alternative algorithms such as RF (Hobley and Prater, 2019).

The updated soil classification map (Fig. 6) must be interpreted with care, especially for local agricultural or construction planning. First, around 90 % of field observations were based on auger sampling rather than full-profile descriptions. Second, the 100 cm depth limit may omit deep horizons, although other local profile observations (Abdulrahman et al., 2020) tend to show that profiles below 100 cm are uncommon. Third, the Tigris right bank was mapped using remote sensing only, which may introduce higher uncertainty. But despite these limitations, the current product offers improved

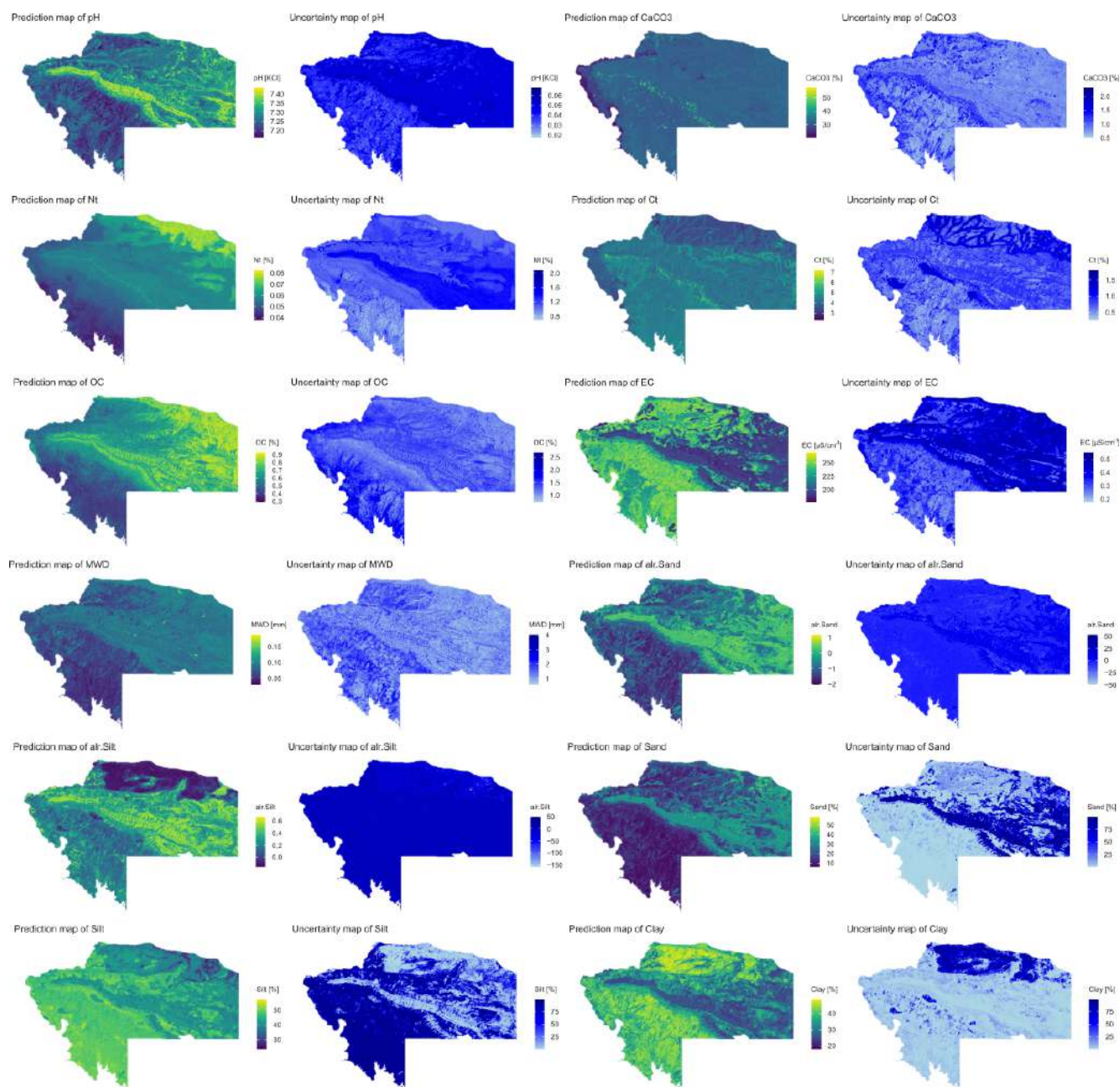


Figure 13. Prediction and uncertainty maps for the 70–100 cm depth increment. Realised with R 4.4.0.

detail over earlier maps (Buringh, 1957; Altaie, 1968), and adheres to modern WRB standards (IUSS Working Group WRB, 2022).

Direct comparison of our model's accuracy with global datasets such as SoilGrids 2.0 is inherently difficult due to differences in spatial resolution and input data. While SoilGrids 2.0 aims to provide consistent global coverage, our maps, with a resolution of 30 m, offer substantial improvements for regional applications. Notably, the density of training samples used in our study (53.5 per 1000 km²) greatly

exceeds that of the WoSIS dataset used for SoilGrids 2.0, which reports a density of only 0.032 per 1000 km² and includes no samples from the Kurdistan Region (Batjes et al., 2020). Furthermore, the use of a conditioned Latin hypercube sampling strategy further enhances spatial representativeness compared to legacy sampling methods (Brus, 2019; Ma et al., 2020).

One limitation lies in the harmonisation of our soil depth intervals with other models (Arrouays et al., 2014; Poggio et al., 2021; Varón-Ramírez et al., 2022; Shi et al., 2025).

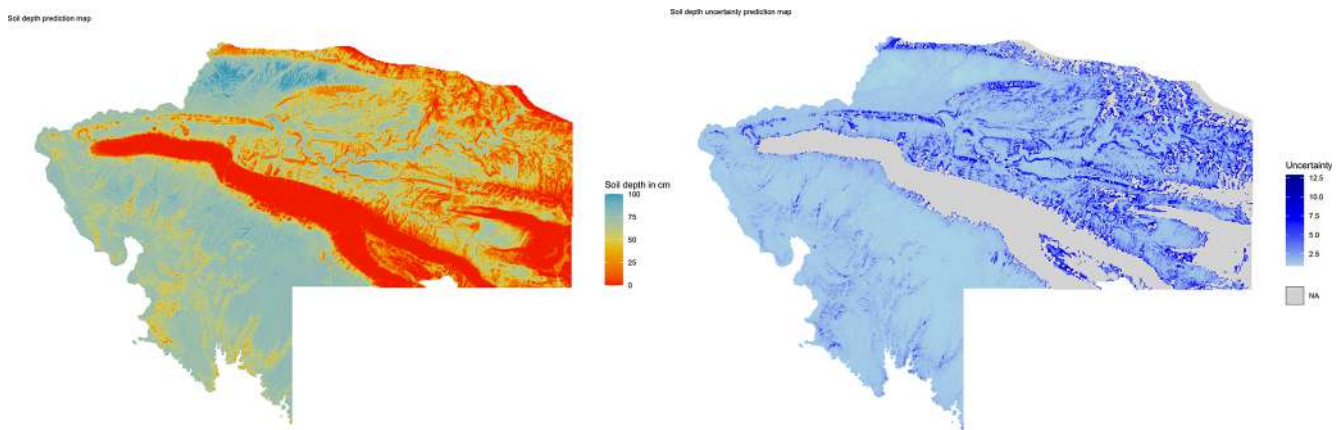


Figure 14. Left: Soil depth prediction map. Right: Soil depth prediction uncertainty map. Realised with R 4 . 4 . 0.

Table 7. Comparative statistics of prediction maps with SoilGrids 2.0 model.

Variable	Statistic	Top soil		Sub-soil		Lower soil	
		Prediction	SoilGrids 2.0	Prediction	SoilGrids 2.0	Prediction	SoilGrids 2.0
		0–30 cm	0–30 cm	30–70 cm	30–60 cm	70–100 cm	60–100 cm
pH	Minimum	8.12	−0.61	8.11	−0.61	8.16	−0.61
	Maximum	8.43	9.18	8.51	9.19	8.47	9.12
	Mean	8.26 ± 0.06	8.55 ± 1.28	8.29 ± 0.08	8.56 ± 1.28	8.29 ± 0.06	8.62 ± 1.29
	1st quartile	8.21	8.68	8.23	8.73	8.24	8.73
	3rd quartile	8.30	8.84	8.35	8.86	8.33	8.87
N _t [%]	Minimum	0.06	0.00	0.04	0.00	0.04	0.00
	Maximum	0.23	0.03	0.13	0.03	0.08	0.03
	Mean	0.11 ± 0.04	0.02 ± 0	0.07 ± 0.02	0.02 ± 0	0.06 ± 0.01	0.02 ± 0.01
	1st quartile	0.08	0.02	0.06	0.02	0.05	0.01
	3rd quartile	0.12	0.02	0.08	0.02	0.06	0.02
OC [%]	Minimum	0.57	0.00	0.40	0.00	0.32	0.00
	Maximum	2.37	0.40	1.63	0.19	0.90	0.18
	Mean	1.14 ± 0.36	0.14 ± 0.02	0.87 ± 0.29	0.07 ± 0.02	0.63 ± 0.14	0.06 ± 0.02
	1st quartile	0.89	0.13	0.61	0.06	0.52	0.05
	3rd quartile	1.29	0.15	1.08	0.09	0.75	0.07
Sand [%]	Minimum	9.29	0.00	8.63	0.00	9.85	0.00
	Maximum	40.71	25.69	39.43	26.20	54.06	26.89
	Mean	20.96 ± 6.56	15.06 ± 2.13	21.83 ± 8.20	14.70 ± 2.00	23.22 ± 9.03	15.06 ± 2.01
	1st quartile	15.23	14.01	15.14	13.77	15.41	14.13
	3rd quartile	26.86	15.84	30.69	15.49	31.75	15.83
Silt [%]	Minimum	29.94	0.01	26.67	0.01	23.68	0.01
	Maximum	49.10	46.98	52.90	45.73	52.21	46.00
	Mean	41.77 ± 3.42	40.17 ± 3.86	40.67 ± 4.62	39.48 ± 3.85	41.34 ± 4.34	39.44 ± 3.93
	1st quartile	38.92	39.50	36.48	38.74	38.10	38.58
	3rd quartile	44.90	41.71	44.61	41.06	44.92	41.12
Clay [%]	Minimum	27.09	0.01	27.85	0.01	20.54	0.01
	Maximum	50.19	49.78	50.51	52.90	48.71	51.73
	Mean	37.27 ± 3.98	43.59 ± 4.32	37.50 ± 4.59	44.62 ± 4.38	35.43 ± 6.51	44.31 ± 4.34
	1st quartile	34.05	42.89	33.22	43.92	29.68	43.60
	3rd quartile	40.29	45.35	41.24	46.36	41.05	46.02

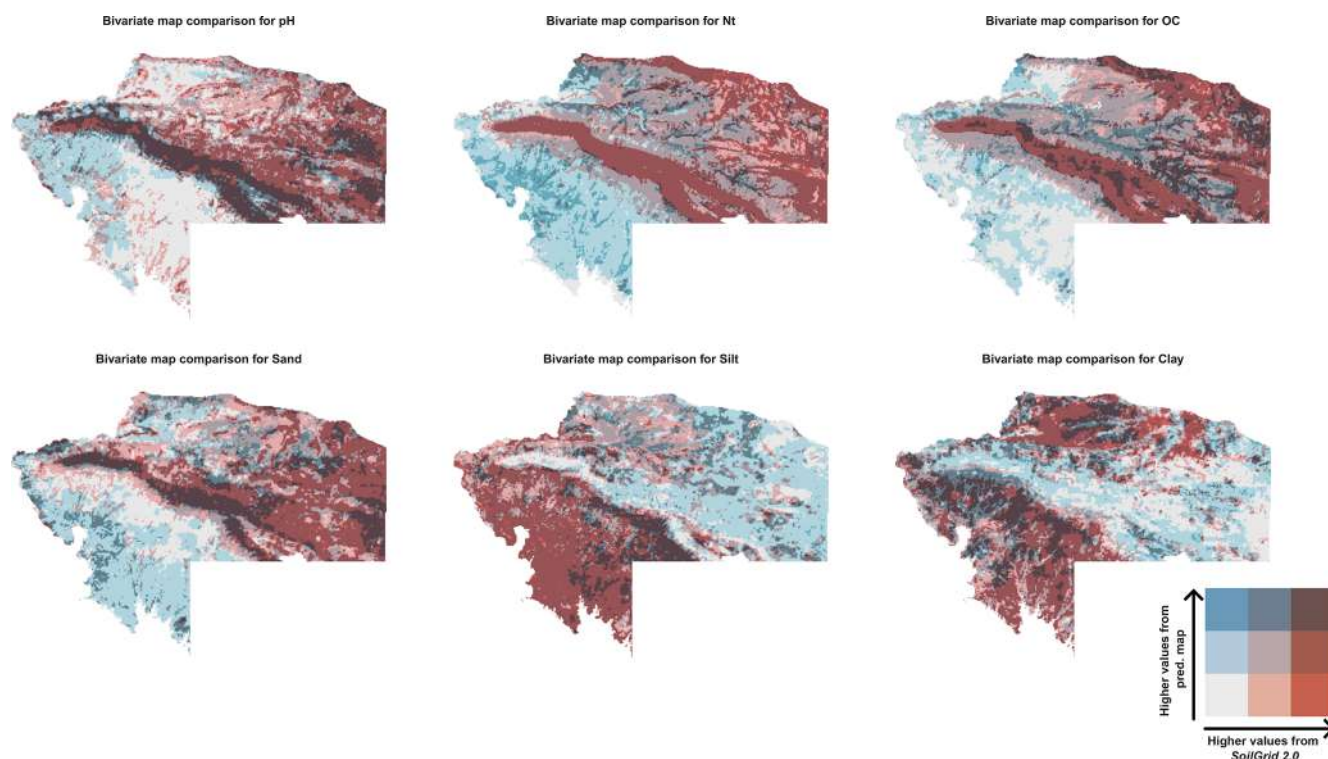


Figure 15. Bivariate map of pH, N_t, OC, and texture from prediction maps vs. SoilGrids 2.0. Realised with R 4.4.0.

A second limitation concerns our limited depth observation window of 100 cm, whereas global products such as SoilGrids 2.0 and the *Chinese Soil Atlas* extend to 200 cm (Shi et al., 2025).

Finally, compared to the prediction model by Yousif et al. (2023), our RMSE values for sand and silt (at 0–10 cm depth) are comparable, while slightly higher than theirs (sand = 9.14; silt = 7.18). However, their clay predictions are more accurate (RMSE = 3.70), and overall model R^2 is higher (sand = 0.91; silt = 0.85; clay = 0.90). Yet, their model is limited to topsoil (0–30 cm) and focuses only on “soil” areas (202 km²), based on the land use/cover classification (LU/C), while our model covers a broader range of landscapes.

5 Data and repository organisation

The Supplement of this paper contain all additional information and original product divided into eight folders:

1. RUSLE part contains all factor of the RUSLE model map and the map itself.
2. The cLHS part includes the R code and the soil profile points produced for 2022 and 2023 campaigns.
3. The field part contains all the photographs of sampling sites and the raw observations made during the cam-

paigns, including the soil classification map in .gpko format.

4. FITR element includes all the raw spectra in the .dpt format and the R code used to compile and filter these spectra.
5. Laboratory folder has only one item, the .csv of all the laboratory measurements detailed.
6. The spectra prediction folder includes the Python codes used to predict the soil properties based on the FTIR spectra with a Cubist model, and the predictions results and metrics.
7. DSM folder contains all the codes and exportations from the digital soil mapping made under R including the raw prediction and uncertainty maps.
8. The Soil depth folder is similar to the DSM folder, only with soil depth values.

6 Code and data availability

The final maps products are available at <https://doi.org/10.6084/m9.figshare.31320958.v2> (Bellat et al., 2026a) in both Network Common Data Form 4 (NetCDF4) and GeoTIFF (GTiff) formats. Profile depth measurements (<https://doi.org/10.1594/PANGAEA.973714>,

Bellat et al., 2024c), laboratory measurement (<https://doi.org/10.1594/PANGAEA.973701>, Bellat et al., 2024b) and MIR spectra and its predictions (<https://doi.org/10.1594/PANGAEA.973700>, Bellat et al., 2024a) are also accessible online. All the supplementary files and raw material is available at <https://doi.org/10.57754/FDAT.d5h1h-4x027>, Bellat et al., 2026b, and the interactive material visible at <https://mathias-bellat.github.io/DSM-Kurdistan/> (last access: 26 March 2026). Code is available in the Supplement but also at the GitHub deposit <https://github.com/mathias-bellat/DSM-Kurdistan.git> (last access: 26 March 2026) (<https://doi.org/10.5281/zenodo.19236909>, Bellat and Kakhani, 2026). Finally, we developed an online version of the prediction maps of the soil properties, adapted for colourblind persons accessible at <https://mathias-bellat.shinyapps.io/Northern-Kurdistan-map/> (last access: 26 March 2026). The sampling design and Cubist predictive model or spectra transformation were computed with: OS = *Windows 11*, CPU = Intel I7-9750H 2.60 GHz, RAM = 32 GB 2667 MHz DDR4 (1–2 h process). For the prediction maps models were computed on the *Humanum* cloud infrastructure: OS = *CentOS 8 (arch x86_64)*, CPU = 2 · AMD EPYC 7542 32-Core Processor 2.9 MHz (64 cores), RAM = 16 · 64 GB 3200 MHz DDR4 (1024 GB), GPU = 2 · Nvidia A100 40G, (6 h process per map, total 24 h).

7 Conclusions

We developed a complete workflow for digital soil mapping at a regional scale in the Dohuk governorate of the Kurdistan Region of Iraq, combining cLHS-driven sampling, MIR-based soil property prediction, and several machine-learning models to produce 50 soil property maps at 30 m resolution, as well as regional soil depth and soil class maps. Compared with SoilGrids 2.0 and earlier local products, our models offer more locally relevant predictions and improved spatial detail, while also covering a broader set of soil properties and depth increments than previous regional studies. The soil class map further aligns with current WRB standards and benefits from greater observational density than earlier exploratory works.

Beyond these technical achievements, the study highlights the importance of integrating local measurements with models tailored to regional environmental gradients. Global products provide consistent baselines, but they cannot fully capture the geomorphological and topographic contrasts that drive soil variability at fine scales. The superior performance of our regional models demonstrates the complementarity between global and local approaches: global datasets remain essential for broad-scale comparisons, whereas locally calibrated workflows are crucial for operational land management, agricultural planning, and resource assessments.

The proposed workflow is fully transferable to other regions of similar size (2000 km²). Areas with comparable environmental conditions – such as western Iran, northern Syria, or parts of the Mediterranean basin – represent suitable candidates for direct methodological transposition. In addition, the time investment required for the entire process, from sampling design to final DSM production, is relatively modest: in our case, approximately one year (235 person-days). The datasets produced in this study also offer broader reusability, for example, as calibration material for MIR/FTIR spectral libraries (Safanelli et al., 2025; Viscarra Rossel et al., 2016) or as part of a regional or global soil profile archive database (Lachmuth et al., 2025).

A further contribution of this work is the provision of the first regional FAIR-compliant soil dataset (Crystal-Ornelas et al., 2022). Soil science research remains highly geographically imbalanced, with five countries producing more than 80 % of global output (Cherubin et al., 2025). Southwestern Asia is sparsely represented, except for Iran, and no soil profiles from the Kurdistan Region of Iraq appear in the WoSIS database (Batjes et al., 2020) used by SoilGrids 2.0 (Poggio et al., 2021). Such data gaps reinforce global inequalities in environmental knowledge (Allik et al., 2020; Sonnenwald, 2007) and limit the capacity of data-poor regions to benefit from international modelling initiatives. By openly sharing our dataset and workflow, we help reduce this imbalance and contribute to greater transparency and reproducibility in regional soil information systems.

Overall, this project demonstrates that, locally informed digital soil mapping is feasible and highly effective in data-poor regions. The workflow presented here substantially advances soil knowledge in Dohuk Governorate and provides a generalisable and reproducible model for improving soil information in other regions facing similar environmental and data constraints.

Appendix A: Covariates description and features selection

Table A1. Covariates used for the different process and modelling (Cont.=Continuous data; Dis.=Discrete data; CLHs= Conditioned Latin Hypercube sampling; DSM=Digital soil mapping; NDWI=Normalised difference water index; NIR=Near-infrared; NDVI=Normalised difference vegetation index; SWIR=Short wavelength infrared; EVI=Enhanced vegetation index; SAVI=Soil adjusted vegetation index; NDMI=Normalised difference moisture index; CORSI=Combined spectral response index; LST=Land surface temperature; TVI=Transformed vegetation index; LSWI=Land surface water index; DEM=Digital elevation model; MrRTF= Multiresolution index of the ridge top flatness; MrVBF=Multiresolution index of the valley bottom flatness; TPI=Topographic position index; TWI=Topographic wetness index).

Name	Code	Type	Measure	Used for	Size (m)	Source
Landsat 8 Blue	LA.1	Cont.	0.45–0.51 μm	DSM	30	EROS (2020)
Landsat 8 Green	LA.2	Cont.	0.53–0.59 μm	CLHs; DSM	30	EROS (2020)
Landsat 8 NDVI	LA.3	Cont.	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	DSM	30	Rouse et al. (1974)
Landsat 8 NDWI	LA.4	Cont.	$\frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}}$	DSM	30	McFeeters (1996)
Landsat 8 NIR	LA.5	Cont.	0.85–0.88 μm	CLHs; DSM	30	EROS (2020)
Landsat 8 panchromatic	LA.6	Cont.	0.52–0.90 μm	DSM	15	EROS (2020)
Landsat 8 Red	LA.7	Cont.	0.64–0.67 μm	CLHs; DSM	30	EROS (2020)
Landsat 8 SWIR1	LA.8	Cont.	1.57–1.65 μm	CLHs; DSM	30	EROS (2020)
Landsat 8 SWIR2	LA.9	Cont.	2.11–2.29 μm	CLHs; DSM	30	EROS (2020)
Landsat 8 EVI	LA.10	Cont.	$2.5 \frac{\text{NIR} - \text{Red}}{(\text{NIR} + 6\text{Red} - 7.5\text{Blue}) + 1}$	DSM	30	Huete et al. (1994)
Landsat 8 SAVI	LA.11	Cont.	$1.5 \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.5}$	DSM	30	Huete (1988)
Landsat 8 TVI	LA.12	Cont.	$\sqrt{\text{NDVI} + 0.5}$	DSM	30	Deering (1975)
Landsat 8 NDMI	LA.13	Cont.	$\frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$	DSM	30	Gao (1996)
Landsat 8 CORSI	LA.14	Cont.	$\frac{\text{Blue} + \text{Green}}{\text{Red} + \text{NIR}}$	DSM	30	Fernández-Buces et al. (2006)
Landsat 8 LSWI	LA.15	Cont.	$\frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$	DSM	30	Chandrasekar et al. (2010)
Landsat 8 Brightness index	LA.16	Cont.	$\sqrt{\text{Red}^2 + \text{NIR}^2}$	DSM	30	Khan et al. (2001)
Landsat 8 Clay index	LA.17	Cont.	$\frac{\text{SWIR1}}{\text{SWIR2}}$	DSM	30	Bousbih et al. (2019)
Landsat 8 Salinity index	LA.18	Cont.	$\frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}}$	DSM	30	Abuelgasim and Ammad (2019)
Landsat 8 Carbonate index	LA.19	Cont.	$\frac{\text{Red}}{\text{Green}}$	DSM	30	Boettinger et al. (2008)
Landsat 8 Gypsum index	LA.20	Cont.	$\frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}}$	DSM	30	Nield et al. (2007)
Landsat 5 Green	LA5.1	Cont.	0.52–0.60 μm	Soil Depth	30	EROS (2013)
Landsat 5 Blue	LA5.2	Cont.	0.63–0.69 μm	Soil Depth	30	EROS (2013)
Landsat 5 Red	LA5.3	Cont.	0.76–0.90 μm	Soil Depth	30	EROS (2013)
Landsat 5 NIR	LA5.4	Cont.	2.08–2.35 μm	Soil Depth	30	EROS (2013)
Landsat 5 NDVI	LA5.5	Cont.	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	Soil Depth	30	Rouse et al. (1974)
Landsat 5 NDWI	LA5.6	Cont.	$\frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}}$	Soil Depth	30	McFeeters (1996)
LST Apr–May	LST.1	Cont.	Kelvin	Soil Depth	1000	Hulley and Hook (2018)
LST Feb–Mar	LST.2	Cont.	Kelvin	Soil Depth	1000	Hulley and Hook (2018)
LST Jun–Jul	LST.3	Cont.	Kelvin	Soil Depth	1000	Hulley and Hook (2018)
LST Oct–Nov	LST.4	Cont.	Kelvin	Soil Depth	1000	Hulley and Hook (2018)
MODIS EVI	MO.1	Cont.	$2.5 \frac{\text{NIR} - \text{Red}}{(\text{NIR} + 6\text{Red} - 7.5\text{Blue}) + 1}$	DSM	250	Huete et al. (1994)
MODIS LST day	MO.2	Cont.	Kelvin ¹	DSM	1000	Wan et al. (2021)
MODIS LST night	MO.3	Cont.	Kelvin ⁶	DSM	1000	Wan et al. (2021)
MODIS NDVI	MO.4	Cont.	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	DSM	250	Didan (2021)
MODIS NIR	MO.5	Cont.	0.841–0.876 μm	DSM	250	Vermote (2021)
MODIS Red	MO.6	Cont.	0.62–0.67 μm	DSM	250	Vermote (2021)
MODIS SAVI	MO.7	Cont.	$1.5 \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.5}$	DSM	250	Huete (1988)
MODIS TVI	MO.8	Cont.	$\sqrt{\text{NDVI} + 0.5}$	DSM	250	Deering (1975)
MODIS Brightness index	MO.9	Cont.	$\sqrt{\text{Red}^2 + \text{NIR}^2}$	DSM	250	Khan et al. (2001)
Distance rivers	OT.1	Cont.	m	DSM	25	ESA and Airbus (2022)
Geology	OT.2	Dis.	35 class	CLHs; DSM; Soil Depth	1 : 250 000	Sissakian et al. (1995); Al-Mousawi et al. (2007)
Geomorphology	OT.3	Dis.	17 class	CLHs; DSM; Soil Depth	1 : 250 000	Forti et al. (2021)
Landuses	OT.4	Dis.	11 class	RUSLE; DSM; Soil Depth	10	Zanaga et al. (2021)
Potential evapotranspiration	OT.5	Cont.	mm	DSM; Soil Depth	750	Zomer and Trabucco, 2022; Zomer et al., 2022
Precipitations	OT.6	Cont.	mm	RUSLE; DSM; Soil Depth	1000	Fick and Hijmans (2017)
Solar radiations	OT.7	Cont.	Kj m ⁻²	DSM; Soil Depth	1000	Fick and Hijmans (2017)
Diff. max. and min. temperature	OT.8	Cont.	°C	DSM	1000	Fick and Hijmans (2017)
Wind speed	OT.9	Cont.	m s ⁻¹	DSM; Soil Depth	1000	Fick and Hijmans (2017)
Temperature average	OT.10	Cont.	°C	Soil Depth	1000	Fick and Hijmans (2017)
RUSLE	OT.11	Cont.	Mg ha ¹ yr ⁻¹	CLHs	25	Mathias Bellat
Soil estimation	OT.12	Cont.	–	CLHs	30	Nafiseh Kakhani
HWSD V2	OT.13	Cont.	–	RUSLE	1000	FAO and IIASA (2023)
Sentinel 2 Blue	SE.1	Cont.	0.492–0.496 μm	DSM	10	ESA (2022)
Sentinel 2 Green	SE.2	Cont.	0.559–0.560 μm	DSM	10	ESA (2022)
Sentinel 2 NDVI	SE.3	Cont.	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	DSM	20	Rouse et al. (1974)
Sentinel 2 NDWI	SE.4	Cont.	$\frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}}$	DSM	20	McFeeters (1996)
Sentinel 2 NIR	SE.5	Cont.	0.833–0.835 μm	DSM	10	ESA (2022)
Sentinel 2 Red	SE.6	Cont.	0.664–0.665 μm	DSM	10	ESA (2022)

Table A1. Continued.

Name	Code	Type	Measure	Used for	Size (m)	Source
Sentinel 2 RedEdge1	SE.7	Cont.	0.738–0.739 μm	DSM	20	ESA (2022)
Sentinel 2 RedEdge2	SE.8	Cont.	0.739–0.740 μm	DSM	20	ESA (2022)
Sentinel 2 RedEdge3	SE.9	Cont.	0.779–0.782 μm	DSM	20	ESA (2022)
Sentinel 2 SWIR1	SE.10	Cont.	1.610–1.613 μm	DSM	20	ESA (2022)
Sentinel 2 SWIR2	SE.11	Cont.	2.185–2.202 μm	DSM	20	ESA (2022)
Sentinel 2 Water vapor	SE.12	Cont.	0.943–0.945 μm	DSM	90	ESA (2022)
Sentinel 2 EVI	SE.13	Cont.	$2.5 \frac{\text{NIR}-\text{Red}}{(\text{NIR}+6\text{Red}-7.5\text{Blue})+1}$	DSM	20	Huete et al. (1994)
Sentinel 2 SAVI	SE.14	Cont.	$1.5 \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}+0.5}$	DSM	20	Huete (1988)
Sentinel 2 TVI	SE.15	Cont.	$\sqrt{\text{NDVI} + 0.5}$	DSM	20	Deering (1975)
Sentinel 2 Clay index	SE.16	Cont.	$\frac{\text{SWIR1}-\text{SWIR2}}{\text{NIR}+\text{SWIR1}}$	DSM	20	Bousbih et al. (2019)
Sentinel 2 NDMI	SE.17	Cont.	$\frac{\text{NIR}-\text{SWIR1}}{\text{Blue}+\text{Green}+\text{NIR}}$	DSM	20	Gao (1996)
Sentinel 2 COSRI	SE.18	Cont.	$\frac{\text{Red}+\text{NIR}}{\text{NIR}-\text{SWIR1}}$	DSM	20	Fernández-Buces et al. (2006)
Sentinel 2 LSWI	SE.19	Cont.	$\frac{\text{NIR}-\text{SWIR1}}{\text{NIR}+\text{SWIR1}}$	DSM	20	Chandrasekar et al. (2010)
Sentinel 2 Brightness index	SE.20	Cont.	$\sqrt{\text{Red}^2 + \text{NIR}^2}$	DSM	20	Khan et al. (2001)
Sentinel 2 Salinity index	SE.21	Cont.	$\frac{\text{SWIR1}-\text{SWIR2}}{\text{SWIR1}-\text{NIR}}$	DSM	20	Abuelgasim and Ammad (2019)
Sentinel 2 Carbonate index	SE.22	Cont.	$\frac{\text{Red}}{\text{Red}+\text{NIR}}$	DSM	20	Boettinger et al. (2008)
Sentinel 2 Gypsum index	SE.23	Cont.	$\frac{\text{Green}}{\text{SWIR1}-\text{SWIR2}}$	DSM	20	Nield et al. (2007)
Aspect	TE.1	Cont.	Radian	DSM; Soil Depth	30	ESA and Airbus (2022)
Channel network base level	TE.2	Cont.	–	DSM	30	ESA and Airbus (2022)
Channel network distance	TE.3	Cont.	–	DSM	30	ESA and Airbus (2022)
Convexity	TE.4	Cont.	–	DSM	30	ESA and Airbus (2022)
DEM fill	TE.5	Cont.	m	RUSLE; CLHs; DSM; Soil Depth	3	ESA and Airbus (2022)
Flow accumulation	TE.6	Cont.	–	DSM	30	ESA and Airbus (2022)
General curvature	TE.7	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
MrRTF	TE.8	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
MrVBF	TE.9	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
Negative openness	TE.10	Cont.	Radian	DSM	30	ESA and Airbus (2022)
Normalized height	TE.11	Cont.	–	DSM	30	ESA and Airbus (2022)
Plan curvature	TE.12	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
Positive openness	TE.13	Cont.	Radian	DSM	30	ESA and Airbus (2022)
Profile curvature	TE.14	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
Slope height	TE.15	Cont.	–	DSM	30	ESA and Airbus (2022)
Slope	TE.16	Cont.	Radian	DSM	30	ESA and Airbus (2022)
Standardized height	TE.17	Cont.	–	DSM	30	ESA and Airbus (2022)
Surface landform	TE.18	Cont.	–	DSM	30	ESA and Airbus (2022)
Terrain ruggedness Index	TE.19	Cont.	–	DSM	30	ESA and Airbus (2022)
Terrain texture	TE.20	Cont.	–	DSM	30	ESA and Airbus (2022)
TPI	TE.21	Cont.	–	DSM; Soil Depth	30	ESA and Airbus (2022)
TWI	TE.22	Cont.	–	CLHs	30	ESA and Airbus (2022)
Total catchment area	TE.23	Cont.	–	DSM	30	ESA and Airbus (2022)
Valley depth	TE.24	Cont.	–	DSM	30	ESA and Airbus (2022)

Appendix B: Regional environment

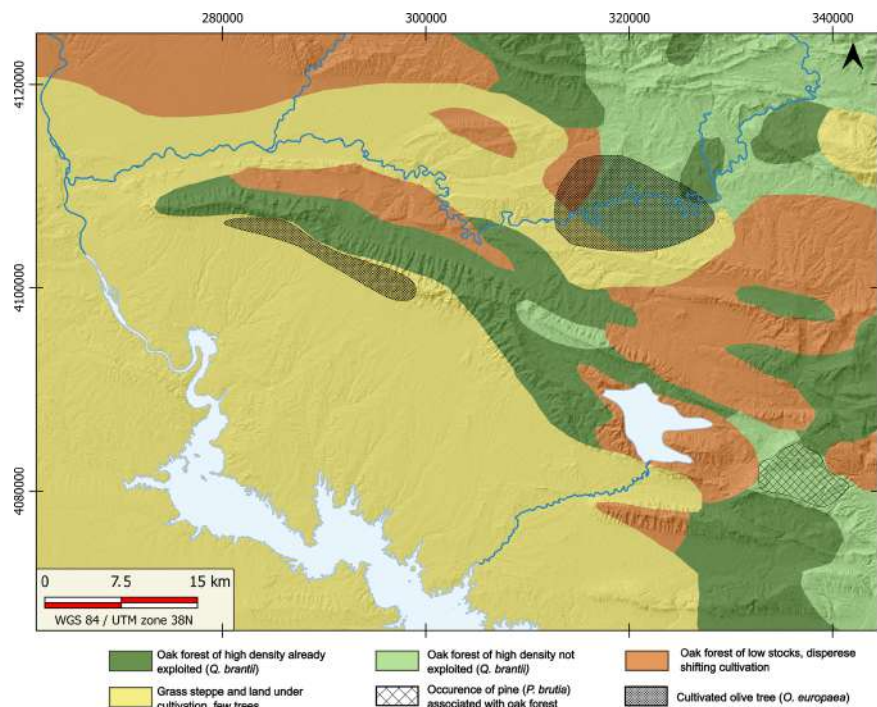


Figure B1. Vegetation map modified from Guest and Al-Rawi (1966). Realised with QGIS 3.34.5 and Inkscape 1.4.

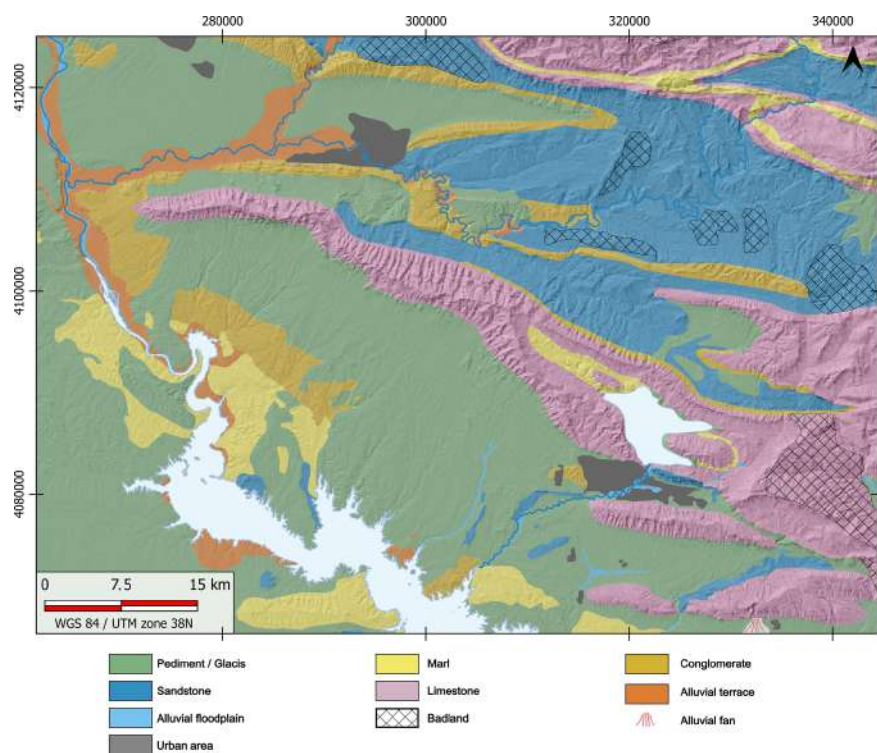


Figure B2. Geomorphological modified map from Forti et al. (2021). Realised with QGIS 3.34.5 and Inkscape 1.4.

Appendix C: RUSLE equations

Table C1. Equations for the performed RUSLE model.

Land use	Value
Water	0
Forest	0.001
Shrubland	0.01
Urban Area	0.1
Cropland	0.2
Wasteland	1

Land use (C) are defined based on (Copernicus, 2019).

$$K = F_{c_{sand}} \times F_{si-cl} \times F_{orgc} \times F_{hi_{sand}} \times 0.1317$$

where $F_{c_{sand}}$, F_{si-cl} , F_{orgc} and, $F_{hi_{sand}}$ are defined below.

$$F_{c_{sand}} = 0.2 + 0.3 \exp\left(-0.0256 \text{SAN} \left(1 - \frac{\text{SIL}}{100}\right)\right)$$

where SIL is the amount of silt in %.

$$F_{si-cl} = \frac{\text{SIL}^{0.3}}{\text{CLA} + \text{SIL}}$$

where SIL and CLA are the amount of silt and clay in %.

$$F_{orgc} = 1 - \frac{0.25C}{C + \exp(3.72 - 2.95C)}$$

where C is the amount of organic carbon in %.

$$F_{hi_{sand}} = 1 - \frac{0.7\left(\frac{\text{SAND}-1}{100}\right)}{\left(\frac{\text{SAND}-1}{100}\right) + \exp\left(-5.51 + 22.9\left(\frac{\text{SAN}-1}{100}\right)\right)}$$

where SAN is the amount of silt in %.

$$\text{LS}_j = \frac{S_j \left(\lambda_j^{m+1} - \lambda_{j-1}^{m+1} \right)}{\left(\lambda_j - \lambda_{j-1} \right) 22.13^m}$$

where S_j = Slope factor for the j th segment, λ_j = Distance from the lower boundary of the j th segment to the upslope (m) and m = length exponent of the RUSLE Ls factor.

$$R = 38.5 + 0.35P$$

where P in the sum of precipitations in one year in mm.

Appendix D: Boruta selections

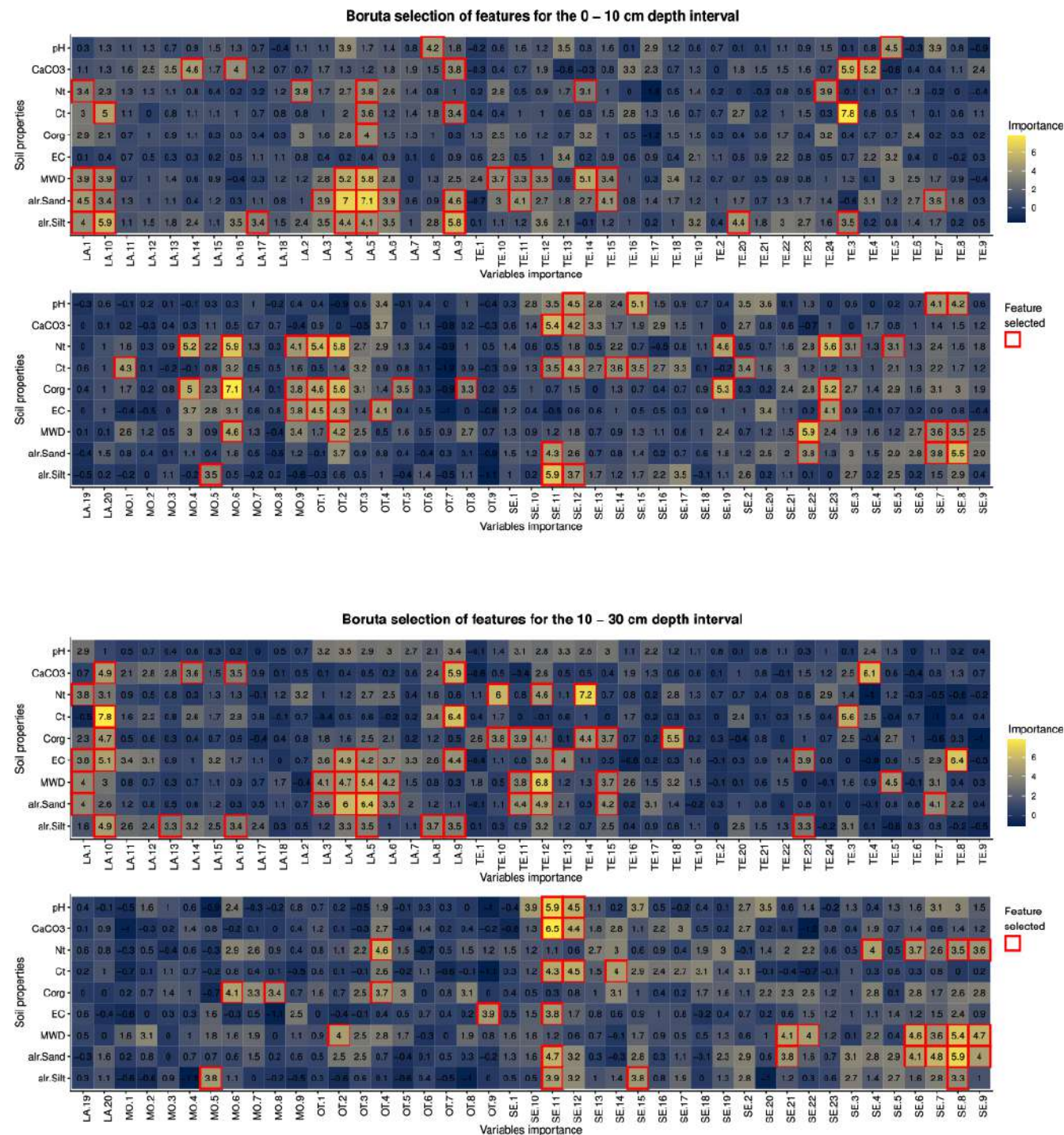
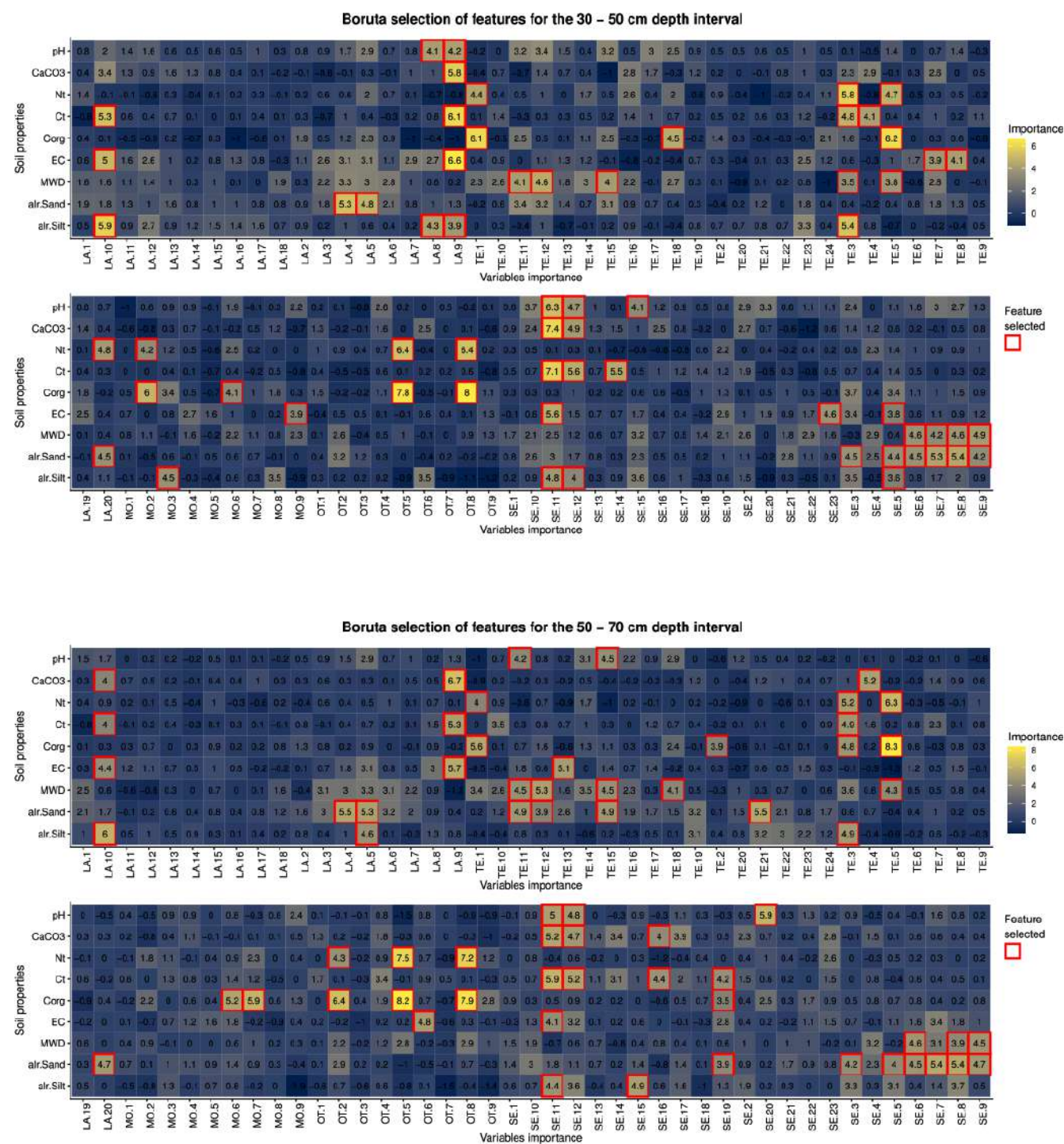


Figure D1.



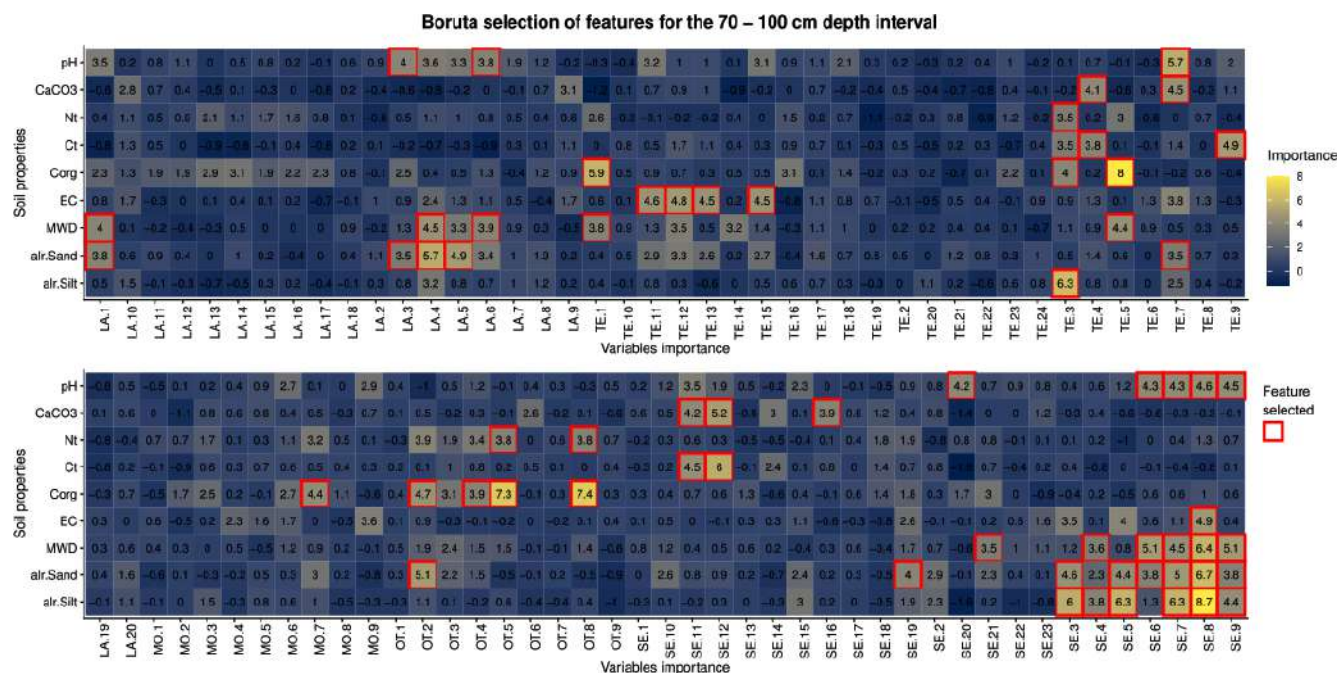


Figure D1. Boruta feature selection for each soil property. Realised with R 4.4.0.

Appendix E: Tuning hyperparameters for prediction maps

Table E1. Tuning hyperparameters used for the digital soil mapping of soil properties and depth (ntree = Number of trees; mtry = Number of predictors; nodesize = Minimum node size).

Variable	Increment	ntree	mtry	nodesize
pH	0–10 cm	500	1	6
	10–30 cm	500	1	1
	30–50 cm	500	13	1
	50–70 cm	500	7	6
	70–100 cm	500	4	6
CaCO ₃	0–10 cm	500	13	11
	10–30 cm	500	7	6
	30–50 cm	500	1	1
	50–70 cm	500	1	6
	70–100 cm	500	1	6
N _t	0–10 cm	500	10	11
	10–30 cm	500	4	1
	30–50 cm	500	7	6
	50–70 cm	500	1	11
	70–100 cm	500	1	21
C _t	0–10 cm	500	13	21
	10–30 cm	500	10	21
	30–50 cm	500	1	11
	50–70 cm	500	10	1
	70–100 cm	500	1	1
OC	0–10 cm	500	13	11
	10–30 cm	500	4	1
	30–50 cm	500	1	11
	50–70 cm	500	4	1
EC	70–100 cm	500	1	6
	0–10 cm	500	1	1
	10–30 cm	500	1	6
	30–50 cm	500	1	21
	50–70 cm	500	1	21
MWD	70–100 cm	500	1	1
	0–10 cm	500	1	1
	10–30 cm	500	4	6
	30–50 cm	500	1	6
	50–70 cm	500	10	21
alr_sand	70–100 cm	500	7	1
	0–10 cm	500	4	1
	10–30 cm	500	1	1
	30–50 cm	500	1	6
	50–70 cm	500	7	11
alr_silt	70–100 cm	500	13	1
	0–10 cm	500	1	6
	10–30 cm	500	1	1
	30–50 cm	500	10	11
	50–70 cm	500	10	16
Depth	70–100 cm	500	10	11
	–	500	5	16

Appendix F: SoilGrids.2

Table F1. Predictions from QRF model vs. SoilGrids.2 model metrics.

Variable	Metric	QRF model 0–10 cm	SoilGrids.2 0–5 cm	QRF model 10–30 cm	SoilGrids.2 5–15 cm	QRF model 30–50 cm	SoilGrids.2 15–30 cm	QRF model 50–70 cm	SoilGrids.2 30–60 cm	QRF model 70–100 cm	SoilGrids.2 60–100 cm
pH	ME	0.02	0.46	−0.02	0.28	0.03	0.51	0.05	0.35	0.01	−0.05
	RMSE	0.09	0.49	0.09	0.94	0.11	0.55	0.18	0.57	0.11	2.27
	R ²	0.22	0.01	0.55	0.01	0.16	0.02	0.09	0.05	0.30	0.00
N _t	ME	−0.02	−0.10	0	−0.05	0	−0.04	−0.01	−0.05	0	−0.03
	RMSE	0.05	0.12	0.02	0.05	0.02	0.04	0.04	0.06	0.02	0.04
	R ²	0.64	0.13	0.23	0.01	0.14	0.03	0.22	0.01	0.08	0.03
OC	ME	−0.24	−1.18	−0.11	−0.69	−0.08	−0.60	−0.07	−0.60	0.02	−0.51
	RMSE	0.61	1.39	0.27	0.77	0.30	0.68	0.34	0.76	0.17	0.55
	R ²	0.46	0.10	0.68	0.03	0.15	0.00	0.76	0.08	0.30	0.00
Sand	ME	−2.60	−2.23	−3.90	−2.83	−1.78	−1.82	0.23	−1.11	−2.96	−6.33
	RMSE	8.43	9.84	10.12	13.91	9.36	8.85	8.69	12.80	8.97	14.50
	R ²	0.29	0.00	0.55	0.11	0.01	0.00	0.51	0.02	0.51	0.01
Silt	ME	0.90	−1.56	0.65	−3.28	1.07	−3.05	1.06	−3.30	1.89	−4.79
	RMSE	5.10	5.52	5.10	8.37	6.27	6.33	4.45	8.34	5.31	13.88
	R ²	0.17	0.07	0.30	0.03	0.04	0.12	0.55	0.21	0.39	0.10
Clay	ME	1.70	3.63	3.25	4.02	0.71	4.86	−1.29	3.18	1.07	4.87
	RMSE	6.48	8.36	8.08	12.26	7.14	9.72	8.14	9.12	7.08	14.25
	R ²	0.28	0.00	0.45	0.03	0.24	0.00	0.23	0.05	0.48	0.03

Sample availability. All soil samples are conserved at the University of Tübingen in the Laboratory for Soil Science and Geoecology at Rümelinstraße 19–23, 72070, Tübingen, 72070, Germany. They can be consulted on demand and for reasonable requests.

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/essd-18-2507-2026-supplement>.

Author contributions. PF and TS secured the funding; MB, BG, RTM, PF and TS conceived the study protocol; BB and PF assured the security protocol during the sampling campaigns; MB, BG, TR and PS conducted the sampling campaigns; MB, TR, NK, RTM and PK performed the analysis; MB interpreted the results; MB, TR, NK, RTM curated the data; MB, MZ, NK and TS wrote the original draft of the manuscript; all authors contributed in the reviewing of the original draft manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors would like to thank all data contributors, especially the Directorate of Antiquities of Dohuk and its personnel, who assisted us during the sampling campaigns and without whom this project would not have been possible (Mohammed, Azad, Ali, and Walad) but also the students who helped on the field (Mathis, Marie-Amandine and Tom). We would also like to thank the *Humanum* consortium and Tara Beuzen-Waller who have left us access to their computing capacities. Finally, we would like to thank the referees for their important and constructive comments that helped us to improve the manuscript. We calculated the estimated carbon footprint for all the different processes and steps needed for this publication and the analysis related to it. All detail of the calculation are available in the Supplement. The total estimated carbon footprint is 3068.064 kg CO₂ equivalent. Furthermore, generative AI chatbot *GPT-4-turbo* and *Claude Sonnet4* were used to generate part of the R code and improve English writing.

Financial support. This work has received funding from the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center (CRC) 1070 “ResourceCultures” (grant no. 215859406).

This open-access publication was funded by the Open Access Publication Fund of the University of Tübingen.

Review statement. This paper was edited by Giulio G. R. Iovine and reviewed by David G. Rossiter, Bas Kempen, and three anonymous referees.

References

- Abdi, B., Kolo, K., and Shahabi, H.: Soil erosion and degradation assessment integrating multi-parametric methods of RUSLE model, RS, and GIS in the Shaqlawa agricultural area, Kurdistan Region, Iraq, *Environ. Monit. Assess.*, 195, 1149, <https://doi.org/10.1007/s10661-023-11796-4>, 2023.
- Abdulrahman, H. D., Mohammed, A., and Doski, J. A. H.: Formation And Development of Vertisols in Selivany Plain at Duhok Governorate, Kurdistan Region, Iraq, *The Journal Of Duhok University*, 23, 246–258, <https://doi.org/10.26682/ajuod.2020.23.2.28>, 2020.
- Abdulsalam, A. and Schlaich, F.: A VII 1 Vorderer Orient. Naturräumliche Gliederung, Dr. Ludwig Reichert Verlag, ISBN 9783882269246, https://reichert-verlag.de/en/tavo-maps/part-a/a_vii/9783882269246_a_vii_1_vorderer_orient_naturraeumliche_gliederung-detail (last access: 27 March 2026), 1988.
- Abuelgasim, A. and Ammad, R.: Mapping soil salinity in arid and semi-arid regions using Landsat 8 OLI satellite data, *Remote Sensing Applications: Society and Environment*, 13, 415–425, <https://doi.org/10.1016/j.rsase.2018.12.010>, 2019.
- Adhikari, K. and Hartemink, A. E.: Linking soils to ecosystem services – A global review, *Geoderma*, 262, 101–111, <https://doi.org/10.1016/j.geoderma.2015.08.009>, 2016.
- Agard, P., Omrani, J., Jolivet, L., Whitechurch, H., Vrielynck, B., Spakman, W., Monié, P., Meyer, B., and Wortel, R.: Zagros orogeny: a subduction-dominated process, *Geol. Mag.*, 148, 692–725, <https://doi.org/10.1017/S001675681100046X>, 2011.
- Aitchison, J.: The statistical analysis of compositional data, *Monographs on statistics and applied probability*, Chapman and Hall, London, 1st edn., ISBN 978-0-412-28060-3, 1986.
- Al-Dabbagh, T. H. and Al-Naqib, S. Q.: Tigris river terrace mapping in northern Iraq and the geotechnical properties of the youngest stage near Dao Al-Qamar Village, *Geological Society, London, Engineering Geology Special Publications*, 7, 603–609, <https://doi.org/10.1144/GSL.ENG.1991.007.01.59>, 1991.
- Al-Mousawi, H., Fouad, S., and Sissakian, V.: Geological map of Zakho quadrangle, State Establishment of Geological Survey and Mining, <https://geosurviraqi.industry.gov.iq/?page=47> (last access: 27 March 2026), 2007.
- Alavi, M.: Structures of the Zagros fold-thrust belt in Iran, *Am. J. Sci.*, 307, 1064–1095, <https://doi.org/10.2475/09.2007.02>, 2007.
- Allik, J., Lauk, K., and Realo, A.: Factors Predicting the Scientific Wealth of Nations, *Cross-Cult. Res.*, 54, 364–397, <https://doi.org/10.1177/1069397120910982>, 2020.
- Altaie, F. H.: The soils of Iraq: Bijdrage tot de kennis van de bodems van Irak/PhD thesis, Rijksuniversiteit Gent. Faculteit van de Wetenschappen, Ghent, <https://lib.ugent.be/catalog/rug01:000236543> (last access: 27 March 2026), 1968.
- Altaie, F. H., Sys, C., and Stoops, G.: Soils groups of Iraq: Their classification and characterization, *Pedologie*, 19, 65–148, https://libcatalog.ugent.be/permalink/32RUG_INST/1hqfkgf/alma990000104910409161 (last access: 27 March 2026), 1969.

- Alwan, I. A., Karim, H. H., and Aziz, N. A.: Agro-Climatic Zones (ACZ) Using Climate Satellite Data in Iraq Republic, IOP Conf. Ser.-Mat. Sci., 518, 022034, <https://doi.org/10.1088/1757-899X/518/2/022034>, 2019.
- Anas, M., Liao, F., Verma, K. K., Sarwar, M. A., Mahmood, A., Chen, Z.-L., Li, Q., Zeng, X.-P., Liu, Y., and Li, Y.-R.: Fate of nitrogen in agriculture and environment: agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency, *Biol. Res.*, 53, 47, <https://doi.org/10.1186/s40659-020-00312-4>, 2020.
- Aqrabi, A., Goff, J., and Horbury, A.: *The Petroleum Geology of Iraq*, Scientific Press Ltd., Singapore, ISBN 978-0-901360-36-6, 2010.
- Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., Mendonca-Santos, M. d., Minasny, B., Montanarella, L., Odeh, I. O., Sanchez, P. A., Thompson, J. A., and Zhang, G.-L.: GlobalSoilMap, in: *Advances in Agronomy*, vol. 125, 93–134, Elsevier, ISBN 978-0-12-800137-0, <https://doi.org/10.1016/B978-0-12-800137-0.00003-0>, 2014.
- Aselin, P.: A Python package for fitting Quinlan's Cubist regression model, *pypi* [code], <https://pypi.org/project/cubist/> (last access: 27 March 2026), 2024.
- Azeez, S. N. and Rahimi, I.: Distribution of Gypsiferous Soil Using Geoinformatics Techniques for Some Aridisols in Garmian, Kurdistan Region-Iraq, *Kurdistan Journal of Applied Research*, 2, 57–64, <https://doi.org/10.24017/science.2017.1.9>, 2017.
- Bailo, D., Paciello, R., Sbarra, M., Rabissoni, R., Vinciarelli, V., and Cocco, M.: Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures, *Front. Earth Sci.*, 8, <https://doi.org/10.3389/feart.2020.00003>, 2020.
- Barzanji, A. F.: Gypsiferous soils of Iraq, PhD thesis, Rijksuniversiteit Gent, Faculteit van de Wetenschappen, Ghent, https://libcatalog.ugent.be/permalink/32RUG_INST/1hqfqqf/alma990002358730409161 (last access: 27 March 2026), 1973.
- Batjes, N. H., Ribeiro, E., and van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), *Earth Syst. Sci. Data*, 12, 299–320, <https://doi.org/10.5194/essd-12-299-2020>, 2020.
- Baumann, P.: *philipp-baumann/simplerspec: simplerspec_v0.2.1 (simplerspec_v0.2.1)*, Zenodo [code], <https://doi.org/10.5281/zenodo.10075988>, 2023.
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, *Scientific Data*, 5, 180214, <https://doi.org/10.1038/sdata.2018.214>, 2018.
- Behrens, T. and Scholten, T.: Digital soil mapping in Germany – a review, *J. Plant Nutr. Soil Sc.*, 169, 434–443, <https://doi.org/10.1002/jpln.200521962>, 2006.
- Behrens, T., Viscarra Rossel, R. A., Ramirez-Lopez, L., and Baumann, P.: Soil spectroscopy with the Gaussian pyramid scale space, *Geoderma*, 426, 116095, <https://doi.org/10.1016/j.geoderma.2022.116095>, 2022.
- Beillouin, D., Corbeels, M., Demenois, J., Berre, D., Boyer, A., Fallot, A., Feder, F., and Cardinael, R.: A global meta-analysis of soil organic carbon in the Anthropocene, *Nat. Commun.*, 14, 3700, <https://doi.org/10.1038/s41467-023-39338-z>, 2023.
- Bellat, M. and Kakhani, N.: DSM-Kurdistan, Zenodo [code], <https://doi.org/10.5281/zenodo.19236909>, 2026.
- Bellat, M., Glissmann, B., Rentschler, T., Sconzo, P., Pfälzner, P., Brifkany, B., and Scholten, T.: Soil properties predicted on mid-infrared (MIR) spectroscopy measurements in North-Western Kurdistan region, Iraq, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.973700>, 2024a.
- Bellat, M., Glissmann, B., Rentschler, T., Sconzo, P., Pfälzner, P., Brifkany, B., and Scholten, T.: Soil properties in the North-Western Kurdistan region, Iraq, derived from laboratory measurements, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.973701>, 2024b.
- Bellat, M., Glissmann, B., Rentschler, T., Sconzo, P., Pfälzner, P., Brifkany, B., and Scholten, T.: Soil bulk density and soil depth from on-site observations in the North-Western Kurdistan region, Iraq, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.973714>, 2024c.
- Bellat, M., Glissmann, B., Rentschler, T., Sconzo, P., Pfälzner, P., Brifkany, B., and Scholten, T.: Digital soil mapping predicted on mid-infrared (MIR) spectroscopy measurements in North-Western Kurdistan region, Iraq (netCDF and GeoTIFF files) V3.0, figshare [data set], <https://doi.org/10.6084/m9.figshare.31320958.v2>, 2026a.
- Bellat, M., Zebari, M., Glissmann, B., Rentschler, T., Sconzo, P., Kakhani, N., Kohsraani, P., Taghizadeh-Mehrjardi, R., Brifkany, B., Pfälzner, P., and Scholten, T.: Material for: “Soil information in Kurdistan region, Dohuk governorate (Iraq)” (v.2.1.0), University of Tübingen [data set], <https://doi.org/10.57754/FDAT.d5h1h-4x027>, 2026.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., and McBratney, A.: Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy, *TRAC-Trend. Anal. Chem.*, 29, 1073–1081, <https://doi.org/10.1016/j.trac.2010.05.006>, 2010.
- Berberian, M.: Master “blind” thrust faults hidden under the Zagros folds: active basement tectonics and surface morphotectonics, *Tectonophysics*, 241, 193–224, [https://doi.org/10.1016/0040-1951\(94\)00185-C](https://doi.org/10.1016/0040-1951(94)00185-C), 1995.
- Blake, G. R. and Hartge, K. H.: Bulk Density, in: *Methods of Soil Analysis*, John Wiley & Sons, Ltd, 363–375, ISBN 978-0-89118-864-3, <https://doi.org/10.2136/sssabookser5.1.2ed.c13>, 1986.
- Blanc, E.-P., Allen, M., Inger, S., and Hassani, H.: Structural styles in the Zagros Simple Folded Zone, Iran, *J. Geol. Soc. London*, 160, 401–412, <https://doi.org/10.1144/0016-764902-110>, 2003.
- Boettinger, J., Ramsey, R., Bodily, J., Cole, N., Kienast-Brown, S., Nield, S., Saunders, A., and Stum, A.: Land-sat Spectral Data for Digital Soil Mapping, in: *Digital Soil Mapping with Limited Data*, edited by: Hartemink, A. E., McBratney, A., and Mendonça-Santos, M. d. L., Springer Netherlands, Dordrecht, 193–202, ISBN 978-1-4020-8592-5, https://doi.org/10.1007/978-1-4020-8592-5_16, 2008.
- Bossio, D. A., Cook-Patton, S. C., Ellis, P. W., Fargione, J., Sanderman, J., Smith, P., Wood, S., Zomer, R. J., von Unger, M., Emmer, I. M., and Griscom, B. W.: The role of soil carbon in natural climate solutions, *Nature Sustainability*, 3, 391–398, <https://doi.org/10.1038/s41893-020-0491-z>, 2020.
- Bousbih, S., Zribi, M., Pelletier, C., Gorraeb, A., Lili-Chabaane, Z., Baghdadi, N., Ben Aissa, N., and Mougenot, B.: Soil Texture Estimation Using Radar and Optical Data

- from Sentinel-1 and Sentinel-2, *Remote Sens.*, 11, 1520, <https://doi.org/10.3390/rs11131520>, 2019.
- Bouslihim, Y., John, K., Miftah, A., Azmi, R., Aboutayeb, R., Bouasria, A., Razouk, R., and Hssaini, L.: The effect of covariates on Soil Organic Matter and pH variability: a digital soil mapping approach using random forest model, *Annals of GIS*, 30, 215–232, <https://doi.org/10.1080/19475683.2024.2309868>, 2024.
- Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Brevik, E. C., Fenton, T. E., and Lazari, A.: Soil electrical conductivity as a function of soil water content and implications for soil mapping, *Precis. Agric.*, 7, 393–404, <https://doi.org/10.1007/s11119-006-9021-x>, 2006.
- Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., and Jordán, A.: Soil mapping, classification, and pedologic modeling: History and future directions, *Geoderma*, 264, 256–274, <https://doi.org/10.1016/j.geoderma.2015.05.017>, 2016.
- Brus, D. J.: Sampling for digital soil mapping: A tutorial supported by R scripts, *Geoderma*, 338, 464–480, <https://doi.org/10.1016/j.geoderma.2018.07.036>, 2019.
- Brus, D. J.: Spatial Sampling with R, Chapman and Hall/CRC, Boca Raton, 1st edn., ISBN 978-1-003-25894-0, <https://doi.org/10.1201/9781003258940>, 2022.
- Buringh, P.: Exploratory Soil Map of Iraq, Ministry of Agriculture, Baghdad, <https://esdac.jrc.ec.europa.eu/content/exploratory-soil-map-iraq-map-1> (last access: 27 March 2026), 1957.
- Buringh, P.: Soils And Soil Conditions In Iraq, Minstry of Agriculture (Iraq), the Netherlands, <http://archive.org/details/buringh-p-soils-and-soil-conditions-in-iraq-republic-of-iraq> (last access: 27 March 2026), 1960.
- Chandrasekar, K., Sesha Sai, M. V. R., Roy, P. S., and Dwevedi, R. S.: Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product, *Int. J. Remote Sens.*, 31, 3987–4005, <https://doi.org/10.1080/01431160802575653>, 2010.
- Cherubin, M. R., Pinheiro Junior, C. R., Souza, L. F. N., Canisares, L. P., Ferreira, T. O., Cerri, C. E. P., Minasny, B., and Smith, P.: Global blind spots in soil health research overlap with environmental vulnerability hotspots, *Communications Earth & Environment*, 6, 651, <https://doi.org/10.1038/s43247-025-02663-w>, 2025.
- Christen, E. and Saliem, K. A.: Managing Salinity in Iraq's Agriculture: Current State, Causes, and Impacts, Internal Report 1, International Center for Agricultural Research in the Dry Areas (ICARDA), Lebanon, <https://hdl.handle.net/20.500.11766/9487> (last access: 27 March 2026), 2013.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, *Geosci. Model Dev.*, 8, 1991–2007, <https://doi.org/10.5194/gmd-8-1991-2015>, 2015.
- Copernicus: ERA5-Land monthly averaged data from 1950 to present, Climate Data Store [data set], <https://doi.org/10.24381/CDS.68D2BB30>, 2019.
- Cossart, E., Fressard, M., and Chaize, B.: Spatial patterns of vineyard landscape evolution and their impacts on erosion susceptibility: RUSLE simulation applied in Mercurey (Burgundy, France) since the mid-20th century, *Erdkunde*, 74, 281–300, <https://doi.org/10.3112/erdkunde.2020.04.04>, 2020.
- Crawford, N. M. and Forde, B. G.: Molecular and Developmental Biology of Inorganic Nitrogen Nutrition, *The Arabidopsis Book*, 2002, <https://doi.org/10.1199/tab.0011>, 2002.
- Curran, P. J., West, S. G., and Finch, J. F.: The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis, *Psychol. Methods*, 1, 16–29, <https://doi.org/10.1037/1082-989X.1.1.16>, 1996.
- Deering, D. W.: Measuring forage production of grazing units from Landsat MSS data, in: Proceedings of 10th International Symposium on Remote Sensing of Environment, 1975, 1169–1178, ERIM, <https://ntrs.nasa.gov/citations/19760052035> (last access: 27 March 2026), 1975.
- Desmet, P. and Govers, G.: A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units, *J. Soil Water Conserv.*, 51, 427–433, <https://doi.org/10.1080/00224561.1996.12457102>, 1996.
- Didan, K.: MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061, EarthData [data set], <https://doi.org/10.5067/MODIS/MOD13Q1.061>, 2021.
- Doski, J. A. H. and McClay, K.: Tectono-stratigraphic evolution, regional structure and fracture patterns of the Zagros fold-thrust belt in the Duhok region, Kurdistan, northern Iraq, *Tectonophysics*, 838, 229506, <https://doi.org/10.1016/j.tecto.2022.229506>, 2022.
- Dou, X., Zhang, J., Zhang, C., Ma, D., Chen, L., Zhou, G., Li, J., and Duan, Y.: Calcium carbonate regulates soil organic carbon accumulation by mediating microbial communities in northern China, *CATENA*, 231, 107327, <https://doi.org/10.1016/j.catena.2023.107327>, 2023.
- Duchaufour, P., Faivre, P., Poulenard, J., and Gury, M.: Introduction à la science du sol: sol, végétation, environnement, Dunod, Malakoff (Hauts-de-Seine), 7th edn., ISBN 978-2-10-081992-8, 2020.
- Eltaif, N. I. and Gharaibeh, M. A.: Soil Erosion Catastrophe in Iraq-Preview, Causes and Study Cases, in: Environmental Degradation in Asia: Land Degradation, Environmental Contamination, and Human Activities, edited by: Al-Quraishi, A. M. F., Mustafa, Y. T., and Negm, A. M., Springer International Publishing, Cham, 179–207, ISBN 978-3-031-12112-8, https://doi.org/10.1007/978-3-031-12112-8_9, 2022.
- Eltaif, N. I., Gharaibeh, M. A., and Fadhil Al-Quraishi, A. M.: Impact of Climate Change on Iraq: Severe Water Scarcity and Desertification, in: The Handbook of Environmental Chemistry, edited by: Barcelo, D. and Kostianoy, A. G., Springer, Berlin, Heidelberg, 25 pp., https://doi.org/10.1007/978-3-031-12112-8_9, 2024.
- EROS: Landsat 4-5 Thematic Mapper Level-2, Collection 2, USGS [data set], <https://doi.org/10.5066/P9IAXOVV>, 2013.
- EROS: Landsat 8-9 Operational Land Imager/Thermal Infrared Sensor Level-2, Collection 2, USGS [data set], <https://doi.org/10.5066/P9OGBGM6>, 2020.
- Ertan, N.: Turkey launches offensive against PKK targets in northern Iraq, Al-Monitor, <https://www.al-monitor.com/originals/2022/04/turkey-launches-offensive-against-pkk-targets-northern-iraq> (last access: 27 March 2026), 2022.

- ESA: Sentinel-2 MSI Level-2A BOA Reflectance, European Space Agency, https://doi.org/10.5270/S2_-znk9xsj, 2022.
- ESA and Airbus: Copernicus DEM, European Space Agency, <https://doi.org/10.5270/ESA-c5d3d65>, 2022.
- European Soil Bureau Network and European Union: Soil atlas of Europe, Office for Official Publications of the European Communities, <https://esdac.jrc.ec.europa.eu/content/soil-atlas-europe> (last access: 27 March 2026), 2005.
- FAO and IIASA: Harmonized World Soil Database version 2.0, FAO; International Institute for Applied Systems Analysis (IIASA), ISBN 978-92-5-137499-3, <https://doi.org/10.4060/cc3823en>, 2023.
- Fernández-Buces, N., Siebe, C., Cram, S., and Palacio, J. L.: Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico, *J. Arid Environ.*, 65, 644–667, <https://doi.org/10.1016/j.jaridenv.2005.08.005>, 2006.
- Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, *Int. J. Climatol.*, 37, 4302–4315, <https://doi.org/10.1002/joc.5086>, 2017.
- Forti, L., Perego, A., Brandolini, F., Mariani, G. S., Zebari, M., Nicoll, K., Regattieri, E., Barbaro, C. C., Bonaccosi, D. M., Qasim, H. A., Cremaschi, M., and Zerboni, A.: Geomorphology of the northwestern Kurdistan Region of Iraq: landscapes of the Zagros Mountains drained by the Tigris and Great Zab Rivers, *J. Maps*, 17, 225–236, <https://doi.org/10.1080/17445647.2021.1906339>, 2021.
- Forti, L., Mariani, G. S., Brandolini, F., Pezzotta, A., and Zerboni, A.: Declassified intelligence satellite imagery as a tool to reconstruct past landforms and surface processes: The submerged riverscape of the Tigris River below the Mosul Dam Lake, Iraq, *Earth Surf. Proc. Land.*, 47, <https://doi.org/10.1002/esp.5389>, 2022.
- Fouad, S. F.: Western Zagros Fold – Thrust belt, part I: the Low Folded Zone, *Iraqi Bulltin of Geology and Mining*, 6, 39–62, <https://ibgm-iq.org/ibgm/index.php/ibgm/article/view/221> (last access: 27 March 2026), 2012.
- Fouad, S. F.: Western Zagros Fold – Thrust belt, part II: the High Folded Zone, *Iraqi Bulltin of Geology and Mining*, 6, 53–71, <https://ibgm-iq.org/ibgm/index.php/ibgm/article/view/274> (last access: 27 March 2026), 2014.
- Friedman, S. P.: Soil properties influencing apparent electrical conductivity: a review, *Comput. Electron. Agr.*, 46, 45–70, <https://doi.org/10.1016/j.compag.2004.11.001>, 2005.
- Gao, B.-C.: NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sens. Environ.*, 58, 257–266, [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3), 1996.
- Google: earthengine-api: Earth Engine Python API, pypi, <http://code.google.com/p/earthengine-api> (last access: 27 March 2026), 2025.
- Guan, Z.-H., Cao, Z., Li, X. G., Scholten, T., Kühn, P., Wang, L., Yu, R.-P., and He, J.-S.: Soil phosphorus availability mediates the effects of nitrogen addition on community- and species-level phosphorus-acquisition strategies in alpine grasslands, *Sci. Total Environ.*, 906, 167630, <https://doi.org/10.1016/j.scitotenv.2023.167630>, 2024.
- Guest, E. and Al-Rawi, A.: Flora of Iraq, vol. 1, Ministry of Agriculture of the Republic of Iraq, Baghdad, <https://floraofkurdistan.com/library> (last access: 27 March 2026), 1966.
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V.: Gene Selection for Cancer Classification using Support Vector Machines, *Mach. Learn.*, 46, 389–422, <https://doi.org/10.1023/A:1012487302797>, 2002.
- Hengl, T. and MacMillan, A. R.: Predictive Soil Mapping with R, OpenGeoHub foundation, Wageningen, the Netherlands, ISBN 978-0-359-30635-0, <http://soilmapper.org> (last access: 27 March 2026), 2019.
- Hijmans, R. J., Bivand, R., Chirico, M., Cordano, E., Dyba, K., Pebesma, E., Rowlingson, B., and Sumner, M. D.: terra: Spatial Data Analysis, CRAN, <https://doi.org/10.32614/CRAN.package.terra>, 2025.
- Hillel, D. and Hatfield, J. L. (Eds.): Encyclopedia of soils in the environment, Elsevier/Academic Press, Oxford, UK, Boston, 1st edn., ISBN 978-0-12-348530-4, 2005.
- Hobley, E. U. and Prater, I.: Estimating soil texture from vis-NIR spectra, *Eur. J. Soil Sci.*, 70, 83–95, <https://doi.org/10.1111/ejss.12733>, 2019.
- Huete, A.: A soil-adjusted vegetation index (SAVI), *Remote Sens. Environ.*, 25, 295–309, [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X), 1988.
- Huete, A., Justice, C., and Liu, H.: Development of vegetation and soil indices for MODIS-EOS, *Remote Sens. Environ.*, 49, 224–234, [https://doi.org/10.1016/0034-4257\(94\)90018-3](https://doi.org/10.1016/0034-4257(94)90018-3), 1994.
- Hulley, G. and Hook, S.: VIIRS/NPP Land Surface Temperature Daily L3 Global 1km SIN Grid Day V001, EarthData [data set], <https://doi.org/10.5067/VIIRS/VNP21A1D.001>, 2018.
- Isiker, A. K., Günay, Y., and Senel, M.: Geological map of Turkey Cizre, General Directorate of Mineral Research and Exploration, Ankara, <https://mta.gov.tr/en/mta/mission> (last access: 27 November 2023, no longer available online), 2002.
- IUSS Working Group WRB: World reference base for soil resources 2022: International soil classification system for naming soils and creating legends for soil maps, International Union of Soil Sciences, Vienna, Austria, 4th edn., ISBN 979-8-9862451-1-9, 2022.
- Janik, L. J., Soriano-Disla, J. M., Forrester, S. T., and McLaughlin, M. J.: Effects of soil composition and preparation on the prediction of particle size distribution using mid-infrared spectroscopy and partial least-squares regression, *Soil Res.*, 54, 889–904, <https://doi.org/10.1071/SR16011>, 2016.
- Jassim, S. Z. and Goff, J. C. (Eds.): Geology of Iraq, Dolin; Moravian Museum, Prague: Brno, 1st edn., ISBN 978-80-7028-287-8, 2006.
- Kennard, R. W. and Stone, L. A.: Computer Aided Design of Experiments, *Technometrics*, 11, 137–148, <https://doi.org/10.2307/1266770>, 1969.
- Khadivi, S., Mouthereau, F., Barbarand, J., Adatte, T., and Lacombe, O.: Constraints on palaeodrainage evolution induced by uplift and exhumation on the southern flank of the Zagros–Iranian Plateau, *J. Geol. Soc. London*, 169, 83–97, <https://doi.org/10.1144/0016-76492011-031>, 2012.
- Khan, N. M., Rastoskuev, V. V., Shalina, E. V., and Sato, Y.: Mapping salt-affected soils using remote sensing indicators—a simple approach with the use of GIS IDRISI, in: 22nd Asian Conference on Remote Sensing, Asian Association

- on remote Sensing, Singapore, <https://acrs-aars.org/proceeding/ACRS2001/Papers/AGS-05.pdf> (last access: 27 March 2026), 2001.
- Koshnaw, R. I., Horton, B. K., Stockli, D. F., Barber, D. E., Tamar-Agha, M. Y., and Kendall, J. J.: Neogene shortening and exhumation of the Zagros fold-thrust belt and foreland basin in the Kurdistan region of northern Iraq, *Tectonophysics*, 694, 332–355, <https://doi.org/10.1016/j.tecto.2016.11.016>, 2017.
- Koshnaw, R. I., Horton, B. K., Stockli, D. F., Barber, D. E., and Tamar-Agha, M. Y.: Sediment routing in the Zagros foreland basin: Drainage reorganization and a shift from axial to transverse sediment dispersal in the Kurdistan region of Iraq, *Basin Res.*, 32, 688–715, <https://doi.org/10.1111/bre.12391>, 2020.
- Kouli, M., Soupios, P., and Vallianatos, F.: Soil erosion prediction using the Revised Universal Soil Loss Equation (RUSLE) in a GIS framework, Chania, Northwestern Crete, Greece, *Environ. Geol.*, 57, 483–497, <https://doi.org/10.1007/s00254-008-1318-9>, 2009.
- Kuhn, M.: The caret Package, CRAN, <https://doi.org/10.32614/CRAN.package.caret>, 2019.
- Kuhn, M. and Quinlan, J. R.: Cubist: Rule- And Instance-Based Regression Modeling, r package version 0.4.4.9000, CRAN [code], <https://doi.org/10.32614/CRAN.package.Cubist>, 2024.
- Kursa, M. B. and Rudnicki, W. R.: Feature Selection with the Boruta Package, *J. Stat. Softw.*, 36, <https://doi.org/10.18637/jss.v036.i11>, 2010.
- Köppen, W. P.: Das geographische System der Klimate, *Handbuch der Klimatologie*, Gebrüder Borntraeger, Berlin, https://koeppen-geiger.vu-wien.ac.at/pdf/Koppen_1936.pdf (last access: 27 March 2026), 1936.
- Lacerda, M. P. C., Demattê, J. A. M., Sato, M. V., Fongaro, C. T., Gallo, B. C., and Souza, A. B.: Tropical Texture Determination by Proximal Sensing Using a Regional Spectral Library and Its Relationship with Soil Classification, *Remote Sens.*, 8, 701, <https://doi.org/10.3390/rs8090701>, 2016.
- Lachmuth, S., Dönmez, C., Hoffmann, C., Specka, X., Svoboda, N., and Helming, K.: Facilitating Effective Reuse of Soil Research Data: The BonaRes Repository, *Eur. J. Soil Sci.*, 76, e70103, <https://doi.org/10.1111/ejss.70103>, 2025.
- Lagacherie, P., McBratney, A. B., and Voltz, M.: Digital Soil Mapping, An Introductory Perspective, no. 31 in *Developments in Soil Science*, Elsevier, ISBN 978-0-444-52958-9, [https://doi.org/10.1016/S0166-2481\(06\)31059-8](https://doi.org/10.1016/S0166-2481(06)31059-8), 2006.
- Lal, R., Monger, C., Nave, L., and Smith, P.: The role of soil in regulation of climate, *Philos. T. R. Soc. B*, 376, 20210084, <https://doi.org/10.1098/rstb.2021.0084>, 2021.
- Lilburne, L., Helfenstein, A., Heuvelink, G. B. M., and Eger, A.: Interpreting and evaluating digital soil mapping prediction uncertainty: A case study using texture from SoilGrids, *Geoderma*, 450, 117052, <https://doi.org/10.1016/j.geoderma.2024.117052>, 2024.
- Liu, F., Yang, F., Zhao, Y.-G., Zhang, G.-L., and Li, D.-C.: Predicting soil depth in a large and complex area using machine learning and environmental correlations, *J. Integr. Agr.*, 21, 2422–2434, [https://doi.org/10.1016/S2095-3119\(21\)63692-4](https://doi.org/10.1016/S2095-3119(21)63692-4), 2022.
- Ludwig, B., Greenberg, I., Vohland, M., and Michel, K.: Optimised use of data fusion and memory-based learning with an Austrian soil library for predictions with infrared data, *Eur. J. Soil Sci.*, 74, e13394, <https://doi.org/10.1111/ejss.13394>, 2023.
- Ma, T., Brus, D. J., Zhu, A.-X., Zhang, L., and Scholten, T.: Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps, *Geoderma*, 370, 114366, <https://doi.org/10.1016/j.geoderma.2020.114366>, 2020.
- Malone, B., Stockmann, U., Glover, M., McLachlan, G., Engelhardt, S., and Tuomi, S.: Digital soil survey and mapping underpinning inherent and dynamic soil attribute condition assessments, *Soil Security*, 6, 100048, <https://doi.org/10.1016/j.soisec.2022.100048>, 2022.
- Malone, B. P., McBratney, A. B., and Minasny, B.: Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes, *Geoderma*, 160, 614–626, <https://doi.org/10.1016/j.geoderma.2010.11.013>, 2011.
- Malone, B. P., McBratney, A. B., and Minasny, B.: Using R for Digital Soil Mapping, *Progress in Soil Science*, Springer International Publishing: Imprint: Springer, Cham, 1st edn., ISBN 978-3-319-44327-0, <https://doi.org/10.1007/978-3-319-44327-0>, 2017.
- McBratney, A., Mendonça Santos, M., and Minasny, B.: On digital soil mapping, *Geoderma*, 117, 3–52, [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4), 2003.
- McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *Int. J. Remote Sens.*, 17, 1425–1432, <https://doi.org/10.1080/01431169608948714>, 1996.
- Mehri, A., Karimi, H., Mustafa, Y. T., Al-Quraishi, A. M. F., and Galalizadeh, S.: Predicting Soil Erosion Using RUSLE Model in Duhok Governorate, Kurdistan Region of Iraq, in: *Earth and Environmental Sciences Library*, Springer International Publishing, Cham, 171–187, ISBN 978-3-031-58314-8 978-3-031-58315-5, https://doi.org/10.1007/978-3-031-58315-5_9, 2024.
- Meinshausen, N. and Michel, L.: quantregForest: Quantile Regression Forests, CRAN [code], <https://doi.org/10.32614/CRAN.package.quantregForest>, 2020.
- Meinshausen, N.: Quantile Regression Forests, *J. Mach. Learn. Res.*, 7, 983–999, <https://jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf> (last access: 27 March 2026), 2006.
- Minasny, B. and McBratney, A. B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information, *Comput. Geosci.*, 32, 1378–1388, <https://doi.org/10.1016/j.cageo.2005.12.009>, 2006.
- Moeys, J., Shanguan, W., Petzold, R., Minasny, B., Rosca, B., Jelinski, N., Zelazny, W., Souza, R. M. S., Safanelli, J. L., and Caten, A. T.: soiltexture: Functions for Soil Texture Plot, Classification and Transformation, CRAN [code], <https://doi.org/10.32614/CRAN.package.soiltexture>, 2024.
- Moharana, P. C., Jena, R. K., Kumar, N., Singh, R. S., and Rao, S. S.: Assessment of soil organic and inorganic carbon stock at different soil depths after conversion of desert into arable land in the hot arid regions of India, *Carbon Manag.*, 12, 153–165, <https://doi.org/10.1080/17583004.2021.1893128>, 2021.
- Molinaro, M., Leturmy, P., Guezou, J.-C., Frizon de Lamotte, D., and Eshraghi, S. A.: The structure and kinematics of the southeastern Zagros fold-thrust belt, Iran: From thin-skinned to thick-skinned tectonics, *Tectonics*, 24, <https://doi.org/10.1029/2004TC001633>, 2005.

- Morgan, R. P. C.: Soil erosion and conservation, Blackwell Pub, Malden, MA, 3rd edn., ISBN 978-1-4051-1781-4, 2005.
- Morgan, R. P. C., Morgan, D. D. V., and Finney, H. J.: A predictive model for the assessment of soil erosion risk, *J. Agr. Eng. Res.*, 30, 245–253, [https://doi.org/10.1016/S0021-8634\(84\)80025-6](https://doi.org/10.1016/S0021-8634(84)80025-6), 1984.
- Mouthereau, F., Lacombe, O., and Vergés, J.: Building the Zagros collisional orogen: Timing, strain distribution and the dynamics of Arabia/Eurasia plate convergence, *Tectonophysics*, 532–535, 27–60, <https://doi.org/10.1016/j.tecto.2012.01.022>, 2012.
- Muhaimeed, A., Aziz, S., and Amin, I.: Genesis of argilic and calcic horizons in some northern Iraqi soils, *Journal of Agricultural and Biological Science*, 8, 29–34, https://www.arpnjournals.com/jabs/research_papers/rp_2013/jabs_0113_511.pdf (last access: 27 March 2026) 2013.
- Muhaimeed, A., Saloom, A., Saliem, K. A., Alani, K., and Muklef, W. M.: Classification and distribution of Iraqi soils, *International Journal of Agriculture Innovations and Research*, 2, 997–1002, https://www.ijair.org/administrator/components/com_jresearch/files/publications/IJAIR_670_Final.pdf (last access: 27 March 2026), 2014.
- Muhaimeed, A. S.: Using Remote Sensing to Predict Soil Properties in Iraq, in: *Environmental Remote Sensing and GIS in Iraq*, edited by: Al-Quraishi, A. M. F. and Negm, A. M., Springer International Publishing, Cham, 41–59, ISBN 978-3-030-21343-5 978-3-030-21344-2, https://doi.org/10.1007/978-3-030-21344-2_3, 2020.
- Najmaldin, H. E.: Statistical Analysis of Rainfall Variations in Duhok City and Semel District, Kurdistan Region of Iraq, *International Journal of Research in Environmental Science*, 9, 31–38, <https://doi.org/10.20431/2454-9444.0903004>, 2023.
- Neina, D.: The Role of Soil pH in Plant Nutrition and Soil Remediation, *Applied and Environmental Soil Science*, 2019, 5794869, <https://doi.org/10.1155/2019/5794869>, 2019.
- Ng, W., Minasny, B., Malone, B., and Filippi, P.: In search of an optimum sampling algorithm for prediction of soil properties from infrared spectra, *PeerJ*, 6, e5722, <https://doi.org/10.7717/peerj.5722>, 2018.
- Ng, W., Malone, B., Minasny, B., and Jeon, S.: Near and mid infrared soil spectroscopy, in: *Reference Module in Earth Systems and Environmental Sciences*, Elsevier, ISBN 978-0-12-409548-9, <https://doi.org/10.1016/B978-0-12-822974-3.00022-7>, 2022a.
- Ng, W., Minasny, B., Jeon, S. H., and McBratney, A.: Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions, *Soil Security*, 6, 100043, <https://doi.org/10.1016/j.soisec.2022.100043>, 2022b.
- Nield, S. J., Boettinger, J. L., and Ramsey, R. D.: Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data, *Soil Sci. Soc. Am. J.*, 71, 245–252, <https://doi.org/10.2136/sssaj2006-0049>, 2007.
- Padarian, J., Minasny, B., and McBratney, A. B.: Machine learning and soil sciences: a review aided by machine learning tools, *SOIL*, 6, 35–52, <https://doi.org/10.5194/soil-6-35-2020>, 2020.
- Patton, N. R., Lohse, K. A., Godsey, S. E., Crosby, B. T., and Seyfried, M. S.: Predicting soil thickness on soil mantled hillslopes, *Nat. Commun.*, 9, 3329, <https://doi.org/10.1038/s41467-018-05743-y>, 2018.
- Pfälzner, P. and Sconzo, P.: The Eastern Khabur Archaeological Survey in Iraqi Kurdistan. A Preliminary Report on the 2014 Season, in: *Zeitschrift für Orient-Archäologie*, no. 9 in *Zeitschrift für Orient-Archäologie*, Wasmuth, Tübingen, Berlin, 10–69, ISBN 9783803002235, <https://wasmuth-verlag.de/shop/archaeologie-alte-kulturen/zeitschrift-fuer-orient-archaeologie-9-2016/> (last access: 27 March 2026), 2016.
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, *SOIL*, 7, 217–240, <https://doi.org/10.5194/soil-7-217-2021>, 2021.
- Ponikarov, V. and Mikhailov, I.: Geological map of Syria, General Establishment of Geology and Mineral Ressource, Damascus, <https://searchworks.stanford.edu/view/2953240> (last access: 27 March 2026), 1986.
- Python Software Foundation: Python Language Reference, <https://www.python.org/> (last access: 27 March 2026), 2022.
- Quinlan, J. R.: Learning with continuous classes, in: *Proceedings AI'92, 5th Australian Conference on Artificial Intelligence*, edited by: Adams, A. and Sterling, L., World Scientific, Hobart, Tasmania, 343–348, ISBN 978-981-02-1250-6 978-981-4536-27-1, <https://doi.org/10.1142/9789814536271>, 1992.
- Rabot, E., Wiesmeier, M., Schlüter, S., and Vogel, H. J.: Soil structure as an indicator of soil functions: A review, *Geoderma*, 314, 122–137, <https://doi.org/10.1016/j.geoderma.2017.11.009>, 2018.
- Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Dematté, J. A., and Scholten, T.: Sampling optimal calibration sets in soil infrared spectroscopy, *Geoderma*, 226–227, 140–150, <https://doi.org/10.1016/j.geoderma.2014.02.002>, 2014.
- R Core Team, Venables, W. N., and Smith, D. M.: An introduction to R: a programming environment for data analysis and graphics, version 4.4.1, R Core Team, Auckland, New Zealand, 4th edn., <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (last access: 27 March 2026), 2024.
- Renard, K. G., Foster, G. R., Weesies, G. A., and Porter, J. P.: RUSLE: Revised universal soil loss equation, *J. Soil Water Conserv.*, 46, 30–33, <https://www.jswnonline.org/content/46/1/30> (last access: 27 March 2026), 1991.
- Roudier, P.: clhs: Conditioned Latin Hypercube Sampling, CRAN [code], <https://doi.org/10.32614/CRAN.package.clhs>, 2021.
- Rouse, J. W., Haas, R. H., Deering, D. W., Schell, J. A., and Harlan, J. C.: Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation, Tech. Rep. E75-10354, NASA, <https://ntrs.nasa.gov/citations/19750020419> (last access: 27 March 2026), 1974.
- Safanelli, J. L., Hengl, T., Parente, L. L., Minarik, R., Bloom, D. E., Todd-Brown, K., Gholizadeh, A., de Sousa Mendes, W., and Sanderman, J.: Open Soil Spectral Library (OSSL): Building reproducible soil calibration models through open development and community engagement, *PLoS One*, 20, e0296545, <https://doi.org/10.1371/journal.pone.0296545>, 2025.
- Salman, S. A., Shahid, S., Ismail, T., Ahmed, K., Chung, E.-S., and Wang, X.-J.: Characteristics of Annual and Seasonal Trends of Rainfall and Temperature in Iraq, *Asia-Pac. J. Atmos. Sci.*, 55, 429–438, <https://doi.org/10.1007/s13143-018-0073-4>, 2019.
- Schaetzl, R. J. and Anderson, S.: Soils: genesis and geomorphology, Cambridge University Press, Cambridge, New York, ISBN 978-0-521-81201-6, 2005.

- Schmidinger, J. and Heuvelink, G. B. M.: Validation of uncertainty predictions in digital soil mapping, *Geoderma*, 437, 116585, <https://doi.org/10.1016/j.geoderma.2023.116585>, 2023.
- Scholten, T., Goebes, P., Kühn, P., Seitz, S., Assmann, T., Bauhus, J., Bruelheide, H., Buscot, F., Erfmeier, A., Fischer, M., Härdtle, W., He, J.-S., Ma, K., Niklaus, P. A., Scherer-Lorenzen, M., Schmid, B., Shi, X., Song, Z., von Oheimb, G., Wirth, C., Wubet, T., and Schmidt, K.: On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems – a study from SE China, *J. Plant Ecol.*, 10, 111–127, <https://doi.org/10.1093/jpe/rtw065>, 2017.
- Sembroni, A., Reitano, R., Faccenna, C., and Callieri, P.: The geologic configuration of the Zagros Fold and Thrust Belt: an overview, *Mediterranean Geoscience Reviews*, 6, 61–86, <https://doi.org/10.1007/s42990-024-00118-6>, 2024.
- Sharififar, A., Minasny, B., Arrouays, D., Boulonne, L., Chevalier, T., van Deventer, P., Field, D. J., Gomez, C., Jang, H.-J., Jeon, S.-H., Koch, J., McBratney, A. B., Malone, B. P., Marchant, B. P., Martin, M. P., Monger, C., Munera-Echeverri, J.-L., Padarian, J., Pfeiffer, M., Richer-de Forges, A. C., Saby, N. P. A., Singh, K., Song, X.-D., Zamanian, K., Zhang, G.-L., and van Zijl, G.: Chapter Four – Soil inorganic carbon, the other and equally important soil carbon pool: Distribution, controlling factors, and the impact of climate change, in: *Advances in Agronomy*, edited by: Sparks, D. L., vol. 178, Academic Press, 165–231, <https://doi.org/10.1016/bs.agron.2022.11.005>, 2023.
- Shi, G., Sun, W., Shangguan, W., Wei, Z., Yuan, H., Li, L., Sun, X., Zhang, Y., Liang, H., Li, D., Huang, F., Li, Q., and Dai, Y.: A China dataset of soil properties for land surface modelling (version 2, CSDLv2), *Earth Syst. Sci. Data*, 17, 517–543, <https://doi.org/10.5194/essd-17-517-2025>, 2025.
- Shrestha, D. L. and Solomatine, D. P.: Machine learning approaches for estimation of prediction interval for the model output, *Neural Networks*, 19, 225–235, <https://doi.org/10.1016/j.neunet.2006.01.012>, 2006.
- Sissakian, V., Hagopian, D., and Hasan, E.: Geological map of Mosul quadrangle Sheet NJ-38-13, State Establishment of Geological Survey and Mining, <https://geosurviraqi.industry.gov.iq/?page=47> (last access: 27 March 2026), 1995.
- Sissakian, V. K. and Abdul Jab'bar, M. F.: Classification of the alluvial fan in Iraq, *Iraqi Bulletin of Geology and Mining*, 10, 43–67, <https://iasj.rdd.edu.iq/journals/uploads/2025/08/24/9fbf65b6df78dc8f152f9f30511a6440.pdf> (last access: 1 February 2024), 2014.
- Sissakian, V. K. and Al-Jiburi, B. S.: Stratigraphy of the Low Folded Zone, *Iraqi Bulletin of Geology and Mining*, 6, 63–132, <https://ibgm-iq.org/ibgm/index.php/ibgm/article/view/274> (last access: 27 March 2026), 2012.
- Sissakian, V. K. and Al-Jiburi, B. S.: Stratigraphy of the High Folded Zone, *Iraqi Bulletin of Geology and Mining*, 6, 73–161, <https://ibgm-iq.org/ibgm/index.php/ibgm/article/view/274> (last access: 27 March 2026), 2014.
- Sissakian, V. K., Abdul Jab'bar, M. F., Al-Ansari, N., and Knutson, S.: Old Alluvial Fan Relics in North and Northeast Iraq, *Journal of Earth Sciences and Geotechnical Engineering*, 5, 45–62, <https://www.diva-portal.org/smash/get/diva2:985298/FULLTEXT01.pdf> (last access: 27 March 2026), 2015.
- Smith, R. and Robertson, V. C.: Soil and irrigation classification of shallow soils overlying gypsum bed, northern Iraq, *J. Soil Sci.*, 13, 106–115, <https://doi.org/10.1111/j.1365-2389.1962.tb00687.x>, 1962.
- Sonnenwald, D. H.: Scientific collaboration, *Annu. Rev. Inform. Sci.*, 41, 643–681, 2007.
- Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., and Wetterlind, J.: Visible and Near Infrared Spectroscopy in Soil Science, in: *Advances in Agronomy*, vol. 107, Elsevier, 163–215, ISBN 978-0-12-381033-5, [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7), 2010.
- Stevens, A. and Ramirez-Lopez, L.: An Introduction to the prospectr package, CRAN, <https://doi.org/10.32614/CRAN.package.prospectr>, 2014.
- Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-Stitt, S., Buzzo, G., Dümperth, C., Wadoux, A., Xiang, W., and Scholten, T.: Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping, *J. Plant Nutr. Soil Sc.*, 179, 499–509, <https://doi.org/10.1002/jpln.201500313>, 2016.
- Suleymanov, A., Richer-de Forges, A. C., Saby, N. P. A., Arrouays, D., Martin, M. P., and Bispo, A.: National-scale digital soil mapping performances are related to covariates and sampling density: Lessons from France, *Geoderma Regional*, 37, e00801, <https://doi.org/10.1016/j.geodrs.2024.e00801>, 2024.
- Swarnkar, S., Malini, A., Tripathi, S., and Sinha, R.: Assessment of uncertainties in soil erosion and sediment yield estimates at ungauged basins: an application to the Garra River basin, India, *Hydrol. Earth Syst. Sci.*, 22, 2471–2485, <https://doi.org/10.5194/hess-22-2471-2018>, 2018.
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafyllis, J., and Omid, M.: Digital Mapping of Soil Classes Using Decision Tree and Auxiliary Data in the Ardakan Region, Iran, *Arid Land Res. Manag.*, 28, 147–168, <https://doi.org/10.1080/15324982.2013.828801>, 2014.
- Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., and Scholten, T.: Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space, *Remote Sens.*, 12, 1095, <https://doi.org/10.3390/rs12071095>, 2020.
- Telo da Gama, J.: The Role of Soils in Sustainability, Climate Change, and Ecosystem Services: Challenges and Opportunities, *Ecologies*, 4, 552–567, <https://doi.org/10.3390/ecologies4030036>, 2023.
- Thapa, P.: Spatial estimation of soil erosion using RUSLE modeling: a case study of Dolakha district, Nepal, *Environmental Systems Research*, 9, 15, <https://doi.org/10.1186/s40068-020-00177-2>, 2020.
- Thomas, G. W.: Soil pH and Soil Acidity, in: *Methods of Soil Analysis*, John Wiley & Sons, Ltd, 475–490, ISBN 978-0-89118-866-7, <https://doi.org/10.2136/sssabookser5.3.c16>, 1996.
- Trivedi, P., Singh, B. P., and Singh, B. K.: Chapter 1 – Soil Carbon: Introduction, Importance, Status, Threat, and Mitigation, in: *Soil Carbon Storage*, edited by: Singh, B. K., Academic Press, 28 pp., ISBN 978-0-12-812766-7, <https://doi.org/10.1016/B978-0-12-812766-7.00001-9>, 2018.
- Tsagris, M., Athineou, G., Alenazi, A., and Adam, C.: Compositional: Compositional Data Analysis, CRAN, <https://doi.org/10.32614/CRAN.package.Compositional>, 2025.

- Umer, M. I., Rajab, S. M., and Ismail, H. K.: Effect of CaCO_3 Form on Soil Inherent Quality Properties of Calcareous Soils, *Materials Sci. Forum*, 1002, 459–467, <https://doi.org/10.4028/www.scientific.net/MSF.1002.459>, 2020.
- Varón-Ramírez, V. M., Araujo-Carrillo, G. A., and Guevara Santamaría, M. A.: Colombian soil texture: building a spatial ensemble model, *Earth Syst. Sci. Data*, 14, 4719–4741, <https://doi.org/10.5194/essd-14-4719-2022>, 2022.
- Vaysse, K. and Lagacherie, P.: Using quantile regression forest to estimate uncertainty of digital soil mapping products, *Geoderma*, 291, 55–64, <https://doi.org/10.1016/j.geoderma.2016.12.017>, 2017.
- Vermote, E.: MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V061, EarthData [data set], <https://doi.org/10.5067/MODIS/MOD09Q1.061>, 2021.
- Viscarra Rossel, R., Behrens, T., Ben-Dor, E., Brown, D., Demattê, J., Shepherd, K., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B., Bartholomeus, H., Bayer, A., Bernoux, M., Böttcher, K., Brodský, L., Du, C., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C., Knadel, M., Morrás, H., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. R., Sanborn, P., Sellitto, V., Sudduth, K., Rawlins, B., Walter, C., Winowiecki, L., Hong, S., and Ji, W.: A global spectral library to characterize the world's soil, *Earth-Sci. Rev.*, 155, 198–230, <https://doi.org/10.1016/j.earscirev.2016.01.012>, 2016.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Chabrilat, S., Demattê, J. A. M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L., Shi, Z., Stenberg, B., Webster, R., Winowiecki, L., and Shen, Z.: Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century, *Eur. J. Soil Sci.*, 73, <https://doi.org/10.1111/ejss.13271>, 2022.
- Wadoux, A. M.-C., Heuvelink, G. B., Lark, R. M., Lagacherie, P., Bouma, J., Mulder, V. L., Libohova, Z., Yang, L., and McBratney, A. B.: Ten challenges for the future of pedometrics, *Geoderma*, 401, 115155, <https://doi.org/10.1016/j.geoderma.2021.115155>, 2021a.
- Wadoux, A. M. J.-C., Malone, B., Minasny, B., Fajardo, M., and McBratney, A. B.: Soil spectral inference with R analysing digital soil spectra, Springer Cham, S.I., ISBN 978-3-030-64895-4, <https://doi.org/10.1007/978-3-030-64896-1>, 2021b.
- Wan, Z., Hook, S., and Hulley, G.: MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V061, EarthData [data set], <https://doi.org/10.5067/MODIS/MOD11A2.061>, 2021.
- Wang, Y. and Witten, I. H.: Induction of model trees for predicting continuous classes, *Working Papers*, <https://hdl.handle.net/10289/1183> (last access: 27 March 2026), 1996.
- Webber, C. L., Bremer, U. F., Taghizadeh-Mehrjardi, R., Weber, B., Rosa, A., Scholten, T., and Seitz, S.: Biological soil crusts as a major ecosystem component in sandization areas of the Brazilian Pampa, *Geoderma Regional*, 34, e00682, <https://doi.org/10.1016/j.geodrs.2023.e00682>, 2023.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>, 2016.
- Wilkinson, T. J.: Soil Development and Early Land Use in the Jazira Region, Upper Mesopotamia, *World Archaeol.*, 22, 87–103, <https://www.jstor.org/stable/124805> (last access: 27 March 2026), 1990.
- Williams, P., Dardenne, P., and Flinn, P.: Tutorial: Items to be included in a report on a near infrared spectroscopy project, *J. Near Infrared Spec.*, 25, 85–90, <https://doi.org/10.1177/0967033517702395>, 2017.
- Yacoub, S. Y., Othman, A. A., and Kadhim, T. H.: Geomorphology of the Low Folded Zone, Iraqi Bulltin of Geology and Mining, 5, 7–37, <https://ibgm-iq.org/ibgm/index.php/ibgm/article/view/221> (last access: 27 March 2026), 2012.
- Yan, F., Shangguan, W., Zhang, J., and Hu, B.: Depth-to-Bedrock Map of China at a Spatial Resolution of 100 Meters, *Earth Syst. Sci. Data Discuss.* [preprint], <https://doi.org/10.5194/essd-2018-103>, 2018.
- Yousif, B. S., Mustafa, Y. T., and Fayyadh, M. A.: Digital mapping of soil-texture classes in Batifa, Kurdistan Region of Iraq, using machine-learning models, *Earth Sci. Inform.*, 16, 1687–1700, <https://doi.org/10.1007/s12145-023-01005-8>, 2023.
- Yu, X., Keitel, C., and Dijkstra, F. A.: Ameliorating soil acidity with calcium carbonate and calcium hydroxide: effects on carbon, nitrogen, and phosphorus dynamics, *J. Soil Sci. Plant Nut.*, 23, 5270–5278, <https://doi.org/10.1007/s42729-023-01400-6>, 2023.
- Zamanian, K., Pustovoytov, K., and Kuz'yakov, Y.: Pedogenic carbonates: Forms and formation processes, *Earth-Sci. Rev.*, 157, 1–17, <https://doi.org/10.1016/j.earscirev.2016.03.003>, 2016.
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., Ramoino, F., and Arino, O.: ESA WorldCover 10m 2020 v100, Zenodo [data set], <https://doi.org/10.5281/ZENODO.5571936>, 2021.
- Zebari, M., Balling, P., Grützner, C., Navabpour, P., Witte, J., and Ustaszewski, K.: Structural style of the NW Zagros Mountains and the role of basement thrusting for its Mountain Front Flexure, Kurdistan Region of Iraq, *J. Struct. Geol.*, 141, 104206, <https://doi.org/10.1016/j.jsg.2020.104206>, 2020.
- Zeraatpisheh, M., Bottega, E. L., Bakhshandeh, E., Owliaie, H. R., Taghizadeh-Mehrjardi, R., Kerry, R., Scholten, T., and Xu, M.: Spatial variability of soil quality within management zones: Homogeneity and purity of delineated zones, *CATENA*, 209, 105835, <https://doi.org/10.1016/j.catena.2021.105835>, 2022.
- Zhang, S., Liu, G., Chen, S., Rasmussen, C., and Liu, B.: Assessing soil thickness in a black soil watershed in north-east China using random forest and field observations, *International Soil and Water Conservation Research*, 9, 49–57, <https://doi.org/10.1016/j.iswcr.2020.09.004>, 2021.

- Zohary, M.: Geobotanical foundations of the Middle East, in: *Geobotanica selecta*, vol. 1, Swets & Zeitlinger, Stuttgart, ISBN 978-90-265-0157-9, 1973.
- Zomer, R. and Trabucco, A.: Global Aridity Index and Potential Evapotranspiration (ET0) Database: Version 3, figshare [data set], <https://doi.org/10.6084/M9.FIGSHARE.7504448.V6>, 2022.

- Zomer, R. J., Xu, J., and Trabucco, A.: Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, *Scientific Data*, 9, 409, <https://doi.org/10.1038/s41597-022-01493-1>, 2022.