Geoderma 443 (2024) 116838

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

High-resolution digital soil mapping of amorphous iron- and aluminium-(hydr)oxides to guide sustainable phosphorus and carbon management

Maarten van Doorn [a,b,1,*], Anatol Helfenstein [c,d,2], Gerard H. Ros [a,b,3], Gerard B.M. Heuvelink [c,e,4], Debby A.M.D. van Rotterdam-Los [a,5], Sven E. Verweij [a,6], Wim de Vries [b,7]

[a] Nutriënten Management Instituut, Nieuwe Kanaal 7C, 6709PA Wageningen, the Netherlands
[b] Environmental Systems Analysis Group, Wageningen University & Research, PO Box 47, 6700AA Wageningen, the Netherlands
[c] Soil Geography and Landscape Group, Wageningen University & Research, PO Box 47, 6700AA Wageningen, the Netherlands
[d] Soil, Water and Land Use Team, Wageningen Environmental Research, PO Box 47, 6700AA Wageningen, the Netherlands
[e] ISRIC – World Soil Information, PO Box 353, 6700AJ Wageningen, the Netherlands

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:
Soil health
Soil functions
Digital soil mapping
Aluminium
Iron
Agriculture

ABSTRACT

Amorphous iron- and aluminium-(hydr)oxides are key soil properties in controlling the dynamics of phosphorus availability and carbon storage. These oxides affect the potential of soils to retain phosphorus and carbon, thus affecting ecosystem services such as crop production, water quality and carbon sequestration. In this study, we spatially predicted oxalate-extractable Fe and Al (Fe_{OX}, Al_{OX}) contents in the Netherlands at 25 m resolution across six soil depth layers between 0 and 200 cm and quantified the associated prediction uncertainty using quantile regression forest. For model training and validation, geo-referenced data of Fe_{OX} and Al_{OX} contents were used including 12,110 wet-chemical observations and 102,393 NIR spectroscopy observations. Over 150 spatial covariates were selected that provide information about soil typology, climate, soil organisms, land use, relief, parent material and space (sampling depth and oblique coordinates). Map quality was assessed by comparing predictions with observations using an independent data set of 4841 soil samples from agricultural fields. Soil sample locations were selected by stratified random sampling, allowing us to assess map quality using design-based statistical inference. Map quality was evaluated using the metrics Model Efficiency Coefficient (MEC), Root Mean Square Error (RMSE) and Mean Error (ME). Map quality differed, depending on the target variable and soil depth, with MEC ranging from 0.19 to 0.80, RMSE from 13.5 to 56.9 mmol kg^{-1} and ME from −6.8 to 6.8 mmol kg^{-1}. Overall, map quality was better for topsoil than for subsoil and better for Al_{OX} contents than for Fe_{OX} contents. Prediction uncertainty quality was evaluated by calculating the Prediction Interval Coverage Probability of the 90 per cent Prediction Interval, which were close to 0.90 in all cases and slightly below 0.90 for Al_{OX}. Thus, prediction uncertainties were generally reliable, though for Al_{OX} contents uncertainty was slightly underpredicted. The maps are a valuable tool for site-specific manure and fertiliser management strategies aiming to balance crop production, water quality and carbon sequestration in agriculture.

* Corresponding author at: Nieuwe Kanaal 7C, 6709PA Wageningen, the Netherlands.
E-mail address: maarten.vandoorn@nmi-agro.nl (M. van Doorn).
[1] https://orcid.org/0000-0001-9197-5687.
[2] https://orcid.org/0000-0003-2432-2672.
[3] https://orcid.org/0000-0002-6062-9770.
[4] https://orcid.org/0000-0003-0959-9358.
[5] https://orcid.org/0000-0002-0582-6261.
[6] https://orcid.org/0000-0002-5573-3952.
[7] https://orcid.org/0000-0001-9974-0612.

https://doi.org/10.1016/j.geoderma.2024.116838
Received 5 December 2023; Received in revised form 22 February 2024; Accepted 25 February 2024
Available online 2 March 2024

1. Introduction

Improving the health of agricultural soils is a pivotal step in the transition towards a sustainable society. Healthy soils can be defined as soils having "the continued capacity to function as a vital living ecosystem that sustains plants, animals, and humans" (Lehmann et al., 2020). Such soils provide a wide range of ecosystem services, such as the production of food, retention of water and nutrients, sequestration of carbon and provision of habitat. Allowing soils to provide those ecosystem services is critical for meeting the Sustainable Development Goals of the United Nations, for example "Zero Hunger", "Clean Water and Sanitation" and "Life on Land" (Lal et al., 2021; Lehmann et al., 2020). In many regions, however, the ability of soils to provide the aforementioned ecosystem services is suppressed. In the European Union (EU), 60–70% of soils are estimated to be in an unhealthy state (European Commission, 2021a). Specific issues for agricultural soils include that 65–75% of agricultural soils have nutrient levels risking eutrophication of soils and water, and that croplands lose 0.5% carbon per year (European Commission, 2021a). In response to existing soil threats, a European Soil Deal has been developed with the mission to put "Europe on a trajectory towards sustainable soil management and restoration as part of a wider, green transition in rural and urban areas" (European Commission, 2021a). Objectives for agricultural soils include (i) reducing soil pollution and enhancing restoration by reducing nutrient losses by at least 50%; and (ii) conserving and increasing soil organic carbon stocks by 0.1–0.4% per year in 2030 (European Commission, 2021a). The EU Soil Deal interlinks with other EU strategies aiming to incorporate environmental objectives in agricultural management, such as the Farm to Fork strategy (European Commission, 2020) and the Zero Pollution Action Plan for air, water and soil (European Commission, 2021b).

Improving soil health requires tremendous effort in member states with intensively cultivated areas such as the Netherlands. In the Netherlands, more than half of the land area is used for agriculture and the livestock density is the highest of all EU member states (European Commission, 2022). This intensive land use affects environmental quality, illustrated by the fact that circa 70% of Dutch surface water bodies are not in a good ecological state (EEA, 2018, 2019), partly due to diffuse nutrient pollution from agriculture (Haase et al., 2023; van Grinsven et al., 2016; Vigiak et al., 2023).

To improve soil health, management decisions must be multifunctional and enhance several soil ecosystem services (Lehmann et al., 2020; Ros et al., 2022). For this reason, trade-offs between ecosystem services (e.g. crop production vs water quality) must be considered (Lehmann et al., 2020). To facilitate multifunction management decision-making, the identification and quantification of so-called soil health indicators which influence the trade-off between ecosystem services is needed (Lehmann et al., 2020). Such soil health indicators are likely to be adopted in agricultural policies. For example, the EU Soil Deal aims to include soil health indicators in a soil monitoring program to determine if soils can be considered to be healthy in terms of e.g. soil nutrients and carbon loss (European Commission, 2021a; European Environment Agency., 2022). In this context, soil contents of amorphous iron (Fe)- and aluminium (Al)-(hydr)oxides are important to consider.

Amorphous Fe- and Al-(hydr)oxides provide ecosystem functions relevant for balancing amongst others crop production, water quality and carbon sequestration. First, they determine the capacity of a soil to bind P and the associated speciation of P between soil sorption sites and soil solution. Soil P speciation, in turn, determines both the P supply to crops and the risk of P leaching to the water system (Kleinman, 2017; van Rotterdam et al., 2012). Second, amorphous Fe- and Al-(hydr)oxides stabilise organic matter in terrestrial ecosystems (Kirsten et al., 2021; Li et al., 2023; Masiello et al., 2004), which is relevant for assessing the potential of soils to sequester carbon (Fukumasu et al., 2021). The main reason for their impact on C and P is due to their small (nano-scale) particle size and high specific surface area. Specific surface areas of 200–1200 m^2 and 400–1750 m^2 per gram reactive metal-(hydr)oxide have been observed for Dutch agricultural topsoils (Hiemstra et al., 2010) and highly weathered tropical soils (Mendez et al., 2022), respectively. Quantitative insights into amorphous Fe- and Al-(hydr)oxide contents are crucial to balance crop production and the risk of P losses to the water system (van Doorn et al., 2023) and to assess the carbon sequestration potential of soils.

Despite the relevance of amorphous Fe- and Al-(hydr)oxides for crop production, water quality and carbon sequestration, spatially explicit information about their soil contents is scarce. This is largely a consequence of the fact that the oxides are not measured during routine agronomic soil testing protocols. Furthermore, soil monitoring networks like LUCAS only recently added the analysis of the oxides to their sampling protocols (Fernandez-Ugalde et al., 2022). Agronomic soil testing protocols currently follow the monofunctional, crop-production-focused rationale of fertilizer recommendations. Since insights into amorphous Fe- and Al-(hydr)oxide contents are not a necessity to derive crop-response curves and fertilizer recommendations which economically optimise crop yield versus nutrient inputs, the oxides are not measured. However, a shift from monofunctional to multifunctional management decision-making suggests that the rationale behind fertilizer recommendations and agronomic soil testing protocols must be re-assessed (van Doorn et al., 2023). This implies a need for spatially explicit predictions of these oxides at various spatial scales at high spatial resolution. In addition, insights into the spatial variation of the soil oxide contents should include insights into variation across the soil depth profile since the risk of P leaching from agricultural fields to the water system depends on the P loading of the oxides across the soil depth profile (Schoumans and Chardon, 2015; van der Zee et al., 1990a, 1990b). The soil carbon sequestration potential also depends on the amount and behaviour of the oxides across soil depth (Masiello et al., 2004). Amorphous Fe- and Al-(hydr)oxide contents have been shown to vary across the soil depth profile in previous studies (Lookman et al., 1995; Møller et al., 2023).

A feasible option to gain insight into the spatial variation in amorphous Fe- and Al-(hydr)oxide contents over the soil profile at a high spatial resolution is by predicting their contents using Digital Soil Mapping (DSM, McBratney et al., 2003). In DSM, statistical models are used to infer the relationship between a soil characteristic and spatially exhaustive environmental covariates (McBratney et al., 2003; Scull et al., 2003), which are typically proxies of the five soil forming factors, namely parent material, climate, organisms (including land use, land cover and soil organisms), topography and time (Jenny, 1941). In a generic DSM framework, covariates relating to space (spatial or geographic position) and other or previously measured attributes of the soil are also used (McBratney et al., 2003). In recent years, the predictive performance of DSM has improved due to the increasing availability of covariates, an extensive toolkit of geostatistical and machine learning algorithms, and a powerful computational infrastructure. Examples of commonly used covariates are climate data, spectral indices of land cover from remote sensing products, digital elevation models and their derivatives and geological maps. Statistical modelling approaches in DSM include regression kriging (Hengl et al., 2004; Odeha et al., 1994), random forests (Breiman, 2001; Helfenstein et al., 2022), gradient boosting (Friedman, 2001; Ma et al., 2021a), artificial neural networks (Behrens et al., 2005; Schmidhuber, 2015), or an ensemble of these methods (Hengl et al., 2021). On a global scale, a wide range of soil properties have been mapped at 250 m spatial resolution (Hengl et al., 2017; Poggio et al., 2021). In the Netherlands, a recent DSM study mapped soil pH at a high (25 m) spatial resolution (Helfenstein et al., 2022). More recently, soil contents of amorphous Fe- and Al-(hydr)oxides have been mapped at 30.4 m resolution for Denmark (Møller et al., 2023), a country similar to the Netherlands in terms of soil geography. Both DSM studies included predictions for multiple soil depth layers and quantified the associated prediction uncertainty. This shows that Digital Soil Mapping can play a key role in meeting the high demand for the

quantification of soil health indicators at various spatial scales.

In DSM studies, estimating prediction uncertainty and evaluating map quality is of interest due to the importance of uncertainty in decision-making. For example, Breure et al. (2022) presented a framework where the optimal nutrient application on agricultural soils – from a cost-benefit perspective – was computed using both the mapped soil nutrient contents as well as the prediction uncertainty. Examples of within-model prediction uncertainty measures are kriging standard deviations (Breure et al., 2022), prediction interval widths – for example the difference between the predicted 5th and 95th quantile (Helfenstein et al., 2022; Poggio et al., 2021) – and the ratio of the prediction interval width over the median prediction (Poggio et al., 2021). When evaluating map quality, both the quality of predictions (e.g. mean, median) as well as within-model uncertainty predictions are assessed. Generally, map quality is evaluated through cross-validation, or by comparing the predictions with observations in a dataset not used for model training (Brus et al., 2011). The most valid and unbiased method to assess map quality is by using data gathered with a design-based sample strategy involving probability sampling (Brus et al., 2011). However, such statistically sound evaluations of map quality are rare since such data are generally not available (Wadoux et al., 2020).

We identify that there is a high need to identify and quantify soil health indicators which are relevant for balancing several ecosystem functions in decision-making. Given the relevance of amorphous Fe- and Al-(hydr)oxides for balancing crop production, water quality and carbon sequestration in agriculture and the lack of spatially explicit data on their contents, we (i) predict soil amorphous Fe- and Al-(hydr)oxide contents for agricultural fields in the Netherlands across the soil depth profile (0–200 cm) at a 25 m spatial resolution using DSM; (ii) create uncertainty maps of the predictions; and (iii) assess map quality using design-based statistical inference.

2. Materials and methods

The methodological workflow consisted of five steps: (i) collect and pre-process soil amorphous Fe- and Al-(hydr)oxide data; (ii) collect and

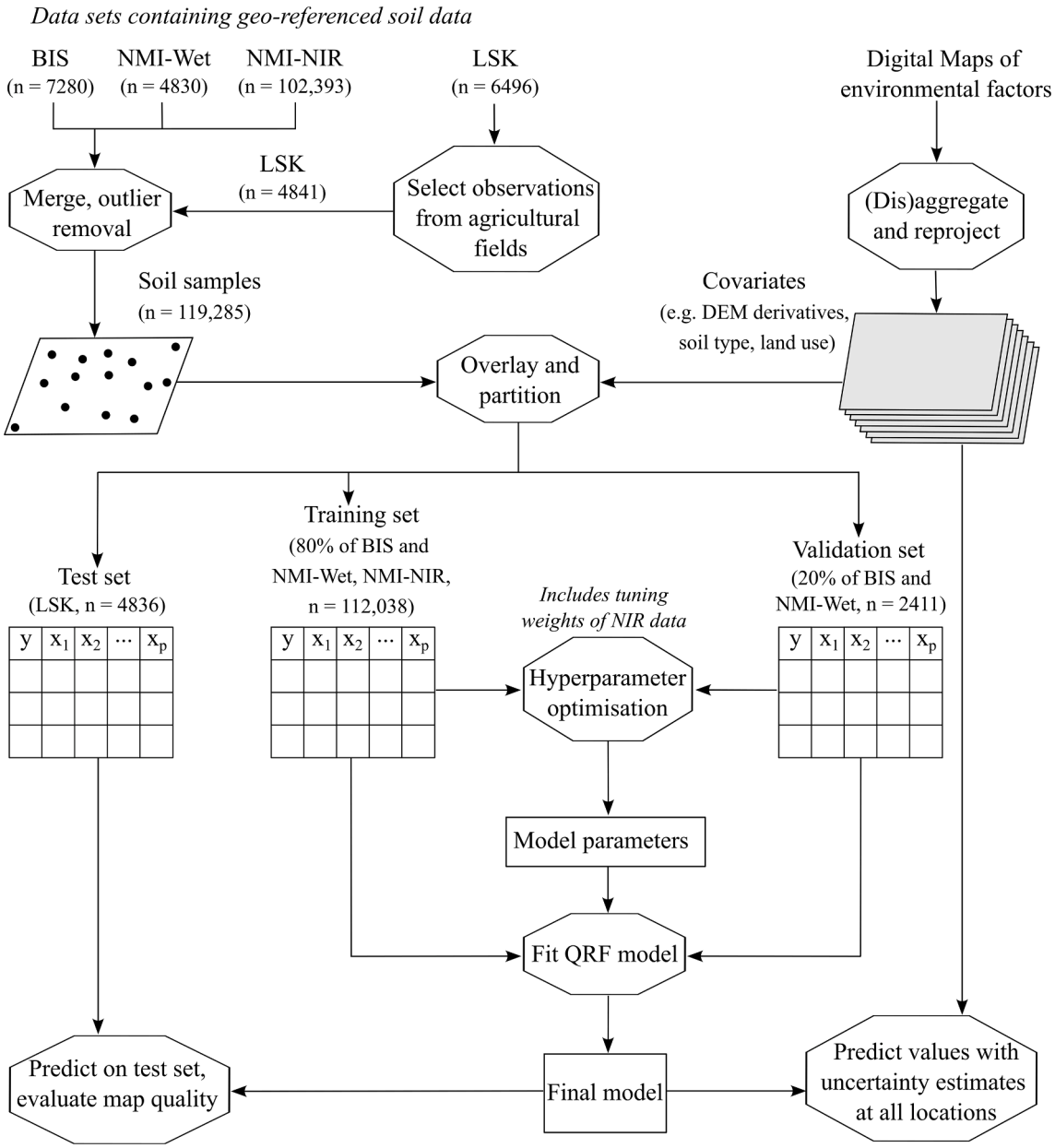

Fig. 1. Methodological framework used to predict soil amorphous Fe- and Al-(hydr)oxide contents for agricultural fields in the Netherlands.

pre-process covariates and prepare regression matrixes using a spatial overlay; (iii) optimise hyperparameters; (iv) fit and apply final spatial prediction models; and (v) assess map quality using design-based statistical inference (Fig. 1).

2.1. Collection and pre-processing of soil amorphous Fe- and Al-(hydr) oxide data

Soil amorphous Fe- and Al-(hydr)oxides contents were operationalised as Fe and Al measured using the 0.2 M acid ammonium oxalate extraction method of Schwertmann (1964), hereafter referred to as Fe_{OX} and Al_{OX} contents. The oxalate extraction, dithionite-citrate-bicarbonate extraction (Mehra and Jackson, 1958) and Na-pyrophosphate extraction (Bascomb, 1968) methods are used to operationally distinguish crystalline versus amorphous and inorganic versus organically complexed Fe and Al forms (McKeague et al., 1971). However, the extraction methods are not fully specific to the soil property of interest. For example, the oxalate extraction method approximates soil amorphous Fe- and Al-(hydr)oxide contents but has also been observed to dissolve organically complexed Fe and Al (McKeague, 1967). Oxalate-extractable Fe and Al contents in soils were thus considered operationalisations of soil amorphous Fe- and Al-(hydr)oxide contents.

For model training, geo-referenced data of Fe_{OX} and Al_{OX} contents were retrieved from three datasets. Firstly, soil data were retrieved from the Dutch soil information database (BIS, 2023), consisting of 7280 soil samples from 1953 locations (Fig. 2A, where data are referred to as 'BIS'). These soil sample locations were purposively selected between 1955 and 2011 in the context of specific research projects. Data were geo-referenced as points. However, it is likely that part of the soil samples were composite samples and thus represent a larger spatial area (e.g. agricultural field). Secondly, we retrieved soil data from the Dutch National Agronomical Soil Archive (NMI, 2022), consisting of 4830 soil samples from 2545 locations (Fig. 2A, where data are referred to as 'NMI-Wet'). These soil sample locations were also purposively selected and soil samples were collected between 2006 and 2023. Soil samples were taken as composites (n = 1875) or single samples (n = 2955). The common soil sampling practice is to sample the topsoil (0–25 cm) as a composite sample whereas samples from deeper soil layers (>25 cm) are taken as a single sample. Thirdly, indirect measurements of Fe_{OX} and Al_{OX} contents based on near-infrared (NIR) spectroscopy models were retrieved from the Dutch National Agronomical Soil Archive (NMI, 2022). Data consisted of 102,393 soil samples from 102,215 locations and were highly clustered in space (Fig. 2B, where data are referred to as 'NMI-NIR'). These soil sample locations were purposively selected between 2009 and 2018. Soil samples were taken as composite samples. The NIR predictions cohere to the wet-chemical measurements in the sense that the spectroscopy models used by the laboratory have been constructed using soil data where Fe_{OX} and Al_{OX} contents were measured using the same methodology, i.e. the 0.2 M ammonium oxalate extraction method (Schwertmann, 1964). Uncertainty estimates of the NIR predictions, however, were not provided by the laboratory and thus unknown. In the three datasets, there was generally a high proportion of topsoil samples (Fig. S1). This was especially observed for the NMI-NIR dataset (100% of soil samples taken from the 0–25 cm depth layer), and to a lesser degree for the NMI-Wet dataset (69% of samples taken from the 0–25 cm depth layer) and BIS dataset (32% of samples taken from the 0–25 cm depth layer). A higher proportion of topsoil observations follows from the standard soil sampling depth of 0–10 cm for grassland and 0–25 cm for arable land in routine agronomic soil testing protocols (Reijneveld et al., 2009).

For assessing map quality, we used data from an independent probability sample, which allowed us to perform design-based statistical inference. Design-based statistical inference using sampling theory is the preferred method of assessing map quality because it allows for unbiased estimation of map accuracy and quantification of the uncertainty of the accuracy metrics with confidence intervals (Brus et al., 2011). Soil sample locations in the three previously mentioned datasets, however, were purposively chosen in the context of research projects and therefore not a probability sample. Therefore, geo-referenced data of Fe_{OX} and Al_{OX} were retrieved from the Dutch national stratified soil survey 'Landelijke Steekproef Kaarteenheden', consisting of 6496 soil samples from 1514 locations (Finke et al., 2001). Soil sample locations were selected using a stratified simple random sampling design. Strata were determined based on a combination of soil type and groundwater regimes. Soil samples were collected between 1989 and 2000. At each location, soil horizons were individually sampled until a minimal depth of 1.2 m (lower limits of the soil depth layer range from 1.2 to 2.5 m), where horizons thicker than 0.25 m were subsampled in layers less than

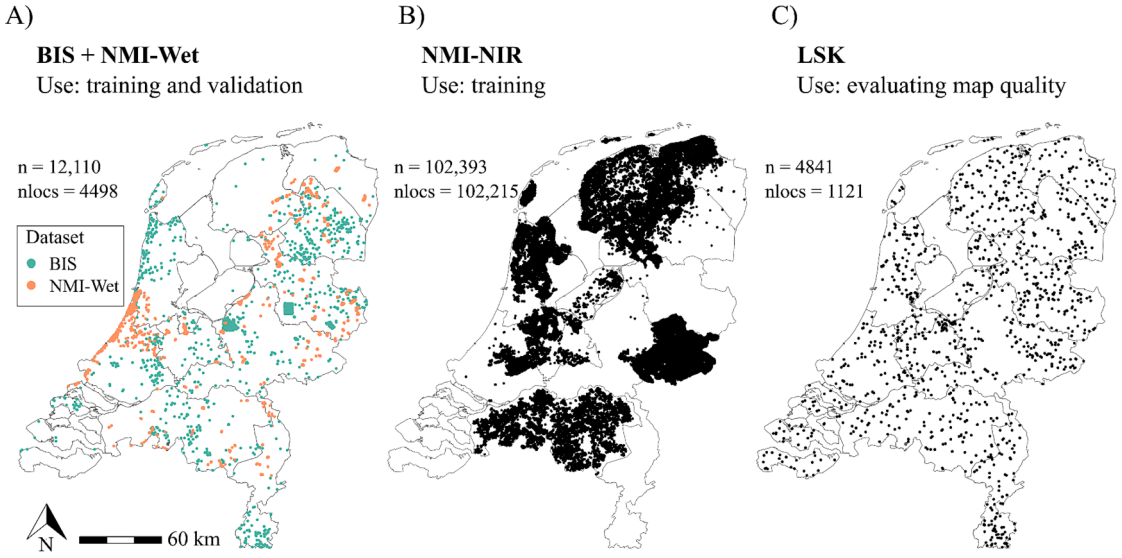

Fig. 2. Spatial distribution of oxalate-extractable Fe and Al content data. Observations are distinguished where: (A) Fe_{OX} and Al_{OX} were directly measured using wet-chemical analysis and where locations were purposively selected (BIS + NMI-Wet); (B) Fe_{OX} and Al_{OX} were indirectly measured using NIR spectroscopic models and where locations were purposively selected (NMI-NIR); and (C) Fe_{OX} and Al_{OX} were directly measured using wet-chemical analysis and where locations were selected using a stratified simple random sampling design (LSK). n refers to the total number of soil samples and nlocs to the total number of soil sample locations. Black solid lines refer to province borders.

20 cm each (Schoumans and Chardon, 2015). For this reason, observations in the LSK dataset are more equally distributed over the soil depth profile compared to observations in the other datasets (Fig. S1). From this dataset, observations from agricultural fields were selected using an agricultural soil mask, which was constructed using land-use data from 2005 to 2022 (Text S1). Selected observations consisted of 4841 soil samples from 1121 locations (Fig. 2C, where data are referred to as 'LSK').

In the aforementioned four datasets, observations were removed where the sampling depth was above the soil surface (litter layer, n = 3) and where data consisted of clear outliers (n = 56). The argumentation for why data were considered outliers is included in the Supporting Information (Text S2). An overview of the characteristics of all datasets is included in Table 1.

After outlier removal, both Fe_{OX} and Al_{OX} contents were positively skewed in all datasets (Table 1, Fig. S2 for histograms). For Fe_{OX}, skewness was highest in the BIS dataset (7.7), followed by NMI-Wet (3.1), LSK (2.3) and NMI-NIR (1.8). For Al_{OX}, skewness was highest in the NMI-Wet dataset (3.0), followed by BIS (2.2), LSK (2.1) and NMI-NIR (1.8). Notably, Fe_{OX} content data from the BIS dataset showed both the highest skewness and kurtosis (96.8) due to the occurrence of high reported Fe_{OX} contents. Specifically, 1% of observations consisted of reported Fe_{OX} contents between 730 and 2457 mmol kg^{-1}.

2.2. Collection and pre-processing of covariates

Over 150 spatial covariates were selected which provide information about parent material (including lithology), climate, organisms (including land use, land cover and soil organisms), topography, soil and space (McBratney et al., 2003). The full list of covariates is included in the SI (Table S1) and includes a digital elevation model (DEM) and its derivatives, soil typology, land-use, geomorphology, palaeogeography and spectral bands and indices from a least vegetated soil Landsat8 composite (the methodology on how the composite was created is described in Text S3). The factor 'space' was included by creating rasters of oblique coordinates across 11 different angles as this method has been shown to prevent the creation of orthogonal map artefacts and provide equal to better accuracies in comparison to alternative methods of including the factor 'space' in DSM studies (Møller et al., 2020).

Pre-processing of spatial covariates consisted of three steps. Firstly, covariates were resampled to an equal projection (Amersfoort, EPSG: 28992), spatial resolution (25 m), origin and extent using bilinear interpolation for numeric data and nearest neighbour interpolation for categorical data. Secondly, categorical variables were encoded to integer arrays, where the encoding method depended on the number of categories and whether the data were nominal or ordinal (overview of discussion on encoding methods discussed in Pargent et al., 2022). Nominal variables were either one-hot encoded (<15 categories) or binary encoded (>=15 categories) whereas ordinal variables were (incrementally) integer encoded.

In addition to spatial covariates, two covariates intrinsic to the data were included. Firstly, we included depth as the midpoint, upper limit and lower limit of the sampling depth layer, as done in various other DMS studies (e.g. Helfenstein et al., 2022) and reviewed by Ma et al. (2021b). Depth was included to allow for the prediction of Fe_{OX} and Al_{OX} contents across the three-dimensional space. Secondly, a covariate was added describing if Fe_{OX} and Al_{OX} contents were measured using wet-chemical analysis (Fig. 2A) or using NIR (Fig. 2B). This covariate was added since data gathered from different analytical techniques are not directly comparable. Correcting for this aspect by using a covariate describing the used analytical technique has shown to be valuable (Nussbaum et al., 2018).

Regression matrices were prepared by obtaining covariate values for the soil sample locations (Fig. 2) using a spatial overlay. If the locations consisted of polygons, extracted numeric data were summarised as the median and categorical data as the modal.

Table 1
Characteristics of the datasets from which geo-referenced data on Fe_{OX} and Al_{OX} were extracted. Summary statistics (mean, range, skewness, kurtosis) were calculated after outlier removal.

Characteristics		BIS	NMI-Wet	NMI-NIR	LSK
Samples (nr)		7280	4830	102,393	4841
Locations (nr)		1953	2545	102,215	1121
Years		1955–2011	2006–2023	2009–2018	1989–2000
Probability sample		No (purposive sample)	No (purposive sample)	No (purposive sample)	Yes (stratified random sample)
Method of analysing Fe_{OX} and Al_{OX} contents		Direct, wet-chemical	Direct, wet-chemical	Indirect, near-infrared	Direct, wet-chemical
Data used for		Training and validation	Training and validation	Training	Evaluating map accuracy
Mean (mmol kg^{-1})	Fe_{OX}	51	114	76	61
	Al_{OX}	43	43	43	31
Range (mmol kg^{-1})[a]	Fe_{OX}	0–2457	0–1262	0–682	0–615
	Al_{OX}	0–388	0–487	0–186	0–224
Skewness	Fe_{OX}	7.7	3.1	1.8	2.3
	Al_{OX}	2.2	3.0	1.8	2.1
Kurtosis	Fe_{OX}	96.8	14.6	6.5	10.2
	Al_{OX}	11.2	17.5	7.4	8.7

[a] Observations of Fe_{OX} and Al_{OX} contents of 0 mmol kg^{-1} (n = 32 for Fe_{OX}, n = 2 for Al_{OX}) were retained and expectedly correspond to measurements below the detection limit.

2.3. Hyperparameter optimisation

Models were tuned separately for Fe_{OX} and Al_{OX} contents using the ranger implementation of randomForest (Wright and Ziegler, 2017). Parameters used to control the learning process of the models (hyperparameters) were optimised using a training and validation set, i.e. various models differing in hyperparameter values were trained on a training set, and model performance was estimated by comparing predictions with measured values on a validation set. The training set consisted of (i) 80% of the BIS and NMI-Wet data (Fig. 2A); and (ii) all NMI-NIR data (Fig. 2B). The validation set consisted of 20% of the BIS and NMI-Wet data. Data were split into training and validation by random sampling of spatial clusters (see Text S4 for details on data-splitting procedures). The LSK data were not used during hyperparameter optimisation since these data were used as an independent test dataset for assessing map quality.

During hyperparameter optimisation, the Root Mean Square Error (RMSE) was used as a prediction metric to be minimized. To correct for sampling density bias (de Bruin et al., 2022), the inverse of the sampling density was used both (i) as an initial weight for sampling of training observations (case.weight); and (ii) as a weight in the derivation of RMSE to provide higher weights to locations in sparsely sampled areas. In line with de Bruin et al. (2022), sampling density was assessed using a two-dimensional kernel approach using the Cronie and van Lieshout criterion (Cronie and Van Lieshout, 2018) to select a bandwidth for the kernel estimation of point process intensity.

Hyperparameters were optimised using a Bayesian optimisation approach (Malkomes et al., 2016; Schratz et al., 2019). Bayesian optimisation efficiently optimises hyperparameters by sampling from a promising subset of a user-defined search space given the performance of previously trained models (Wu et al., 2019). This allows for more computationally efficient hyperparameter optimisation in comparison to methods where priors are not kept into account (such as full and random grid search). Details on how Bayesian optimisation was used, and which hyperparameters were optimised, are included in the SI (Text S4). In addition, we tuned the case.weights of NIR data to account for the

use of multi-source training data, i.e. data consisted of Fe_{OX} and Al_{OX} contents that were measured using wet-chemical analysis and measured using NIR (Fig. 2). One approach to deal with data differing in observational quality in DSM studies is the use of error-filtered machine learning where weights are set proportional to the inverse of the sum of the model residual variance and the measurement error variance (van der Westhuizen et al., 2022). Another approach is to tune those weights as hyperparameters using arbitrary values (Helfenstein et al., 2023). In our case, the measurement error variance of the NIR data was unknown and could not be computed since paired data on Fe_{OX} and Al_{OX} contents that were directly measured using wet-chemical analysis or indirectly measured using NIR were not available. For this reason, we optimised case.weights of NIR data during hyperparameter optimisation (case.weight multiplication factor between 0 and 1). Wet-chemical observations were considered to be "true" and their case.weight multiplication factors were set to 1 (in line with Wadoux et al., 2019). To investigate differences in the measurement error of NIR data, the effect of varying the case.weights of NIR data on the RMSE on the validation set was visualized by constructing Accumulated Local Effect (ALE) plots (Apley and Zhu, 2020).

2.4. Model fitting, prediction and uncertainty quantification

Final models were fitted on the combined training and validation sets (NMI-Wet, BIS and NMI-NIR) using the optimised hyperparameter values. We used quantile regression forest (QRF, Meinshausen, 2006), which produces quantiles of the conditional probability distribution of the target variable. During model fitting variable importance was computed using permutation (Breiman, 2001).

We predicted Fe_{OX} and Al_{OX} contents for all agricultural soils in the Netherlands by applying the final models on the covariate rasterstack. We created maps for the depth layers 0–5, 5–10, 10–25, 25–60, 60–100 and 100–200 cm. Note that these depth layers slightly deviate from GlobalSoilMap standard depths (Arrouays et al., 2014). This was done so that contents of Fe_{OX} and Al_{OX} can more easily be recalculated for the soil sampling depths used in routine agronomic soil testing for grassland (0–10 cm) and arable land (0–25 cm). Notably, the fitted models can be used to predict at any depth.

For each combination of the target variable, location and depth layer, the mean (extracted from the QRF model), median ($q_{0.5}$), 5th quantile ($q_{0.05}$) and 95th quantile ($q_{0.95}$) were predicted. Uncertainty was mapped as the 90 per cent prediction interval (PI90, Eq.1, see e.g. Helfenstein et al., 2022; Poggio et al., 2021) and the prediction interval ratio (PIR, Eq.2, see e.g. Poggio et al., 2021). PI90 and PIR were both included to provide measures of absolute and relative uncertainty.

$$PI90 = q_{0.95} - q_{0.05} \quad (1)$$

$$PIR = \frac{(q_{0.95} - q_{0.05})}{q_{0.5}} \quad (2)$$

To gain insight into differences in Fe_{OX} and Al_{OX} contents between soil types, summary statistics of predictions were computed for each soil physical unit (BOFEK) stratum in the Netherlands (Heinen et al., 2022). Strata consisted of peat, peaty sand, clay/loam and loess soils (Fig. S3). For each stratum, the mean and standard deviation of the predicted median, mean, PI90 and PIR of Fe_{OX} and Al_{OX} contents were computed. Summary statistics were computed on a random sample of grid cells (n = 100,000).

2.5. Assessing map quality using design-based statistical inference

Map quality was assessed with the LSK data and design-based statistical inference, following the approach in Helfenstein et al. (2022), Section 2.6.2. The LSK sampling locations are a probability sample in 2D space, but because Fe_{OX} and Al_{OX} contents were observed at multiple soil depths, design-based statistical inference needs to be conducted for separate soil depths. Firstly, observations were assigned to soil depth layers using the midpoint of the sampling depth. Secondly, Fe_{OX} and Al_{OX} contents were predicted at the locations and depths of the LSK soil samples as both mean and median predictions. Thirdly, data were filtered or averaged so that there was at most one observation per combination of soil sampling location and soil depth layer. Fourthly, accuracy metrics were computed for each depth layer. Accuracy metrics consisted of RMSE, Nash-Sutcliffe Model Efficiency Coefficient (MEC) and Mean Error or Bias (ME) adjusted for the probability sampling strategy (see Helfenstein et al., 2022 for equations).

Besides assessing the accuracy of the predictions of the model, the accuracy of the prediction uncertainty was assessed. The accuracy of the prediction uncertainty was assessed by comparing the PI90 of Fe_{OX} and Al_{OX} contents with the proportion of observations that fell into the 90 per cent prediction interval. Specifically, the prediction interval coverage probability (PICP) of the PI90 ideally equals 0.90, meaning that 90 per cent of the independent observations are between the predicted $q_{0.05}$ and $q_{0.95}$ (Goovaerts, 2001).

Map quality of the 0–5 cm soil depth layer could not be assessed due to an insufficient number of observations in the LSK dataset for this depth layer (n = 18).

2.6. Software

Most of the data processing, statistical procedures and data visualization were carried out in R version 4.2.1 (R Core Team, 2022) on a desktop with a 6-core processor and 32 GB random-access memory. In R, we used the packages "terra" (Hijmans, 2022) for the processing of raster data, "sf" (Pebesma, 2018) for the processing of vector data, "data.table" (Dowle and Srinivasan, 2021) for handling tabular data, "caret" (Kuhn, 2022) and "ranger" (Wright and Ziegler, 2017) for the construction of machine learning models, "ggplot2" (Wickham, 2016) for data visualization and "spatstat" (Baddeley et al., 2016) for creating maps of sampling density. R packages "foreach" (Microsoft and Weston, 2022a) and "doParallel" (Microsoft and Weston, 2022b) were used for parallelization. The DEM derivatives were computed in SAGA GIS (version 8.4.0), and Landsat data were pre-processed in the Google Earth Engine environment (Gorelick et al., 2017).

3. Results

3.1. Predictions of amorphous Fe- and Al-(hydr)oxide contents

Predictions of Fe_{OX} and Al_{OX} contents vary in space and with soil depth as shown for the predicted median values (Fig. 3 and Fig. 4). Predictions of both Fe_{OX} and Al_{OX} contents are highest in the topsoil of peat soils. Average predictions of (median) contents in the 0–5 cm soil depth layer of peat soils correspond to 169 mmol kg^{-1} for Fe_{OX} and 68 mmol kg^{-1} for Al_{OX} (Tables S2 and S3). For sand, loam/clay and loess soils and for the same soil layer, average predictions are lower and range from 30 to 87 mmol kg^{-1} for Fe_{OX} contents and from 26 to 37 mmol kg^{-1} for Al_{OX} contents, depending on soil type.

At all soil depths, predictions of Fe_{OX} contents generally decrease in the order of peat soils (highest), loam/clay soils, peaty soils, loess soils to sandy soils (lowest) whereas predictions of Al_{OX} contents generally decrease in the order of peat soils (highest), peaty soils, loess soils, sandy soils to loam/clay soils (lowest) (Tables S2 and S3). Predictions of Fe_{OX} contents are generally higher in soils with higher soil organic matter contents (Fig. 3) whereas predictions of Al_{OX} contents are generally lower in calcareous and younger soils in comparison to non-calcareous and older soils (Fig. 4).

Predictions of both Fe_{OX} and Al_{OX} contents generally decrease with soil depth. For Fe_{OX} contents specifically, exceptionally low predictions are observed in the deeper soil layers (>60 cm) of the sandy soils. To

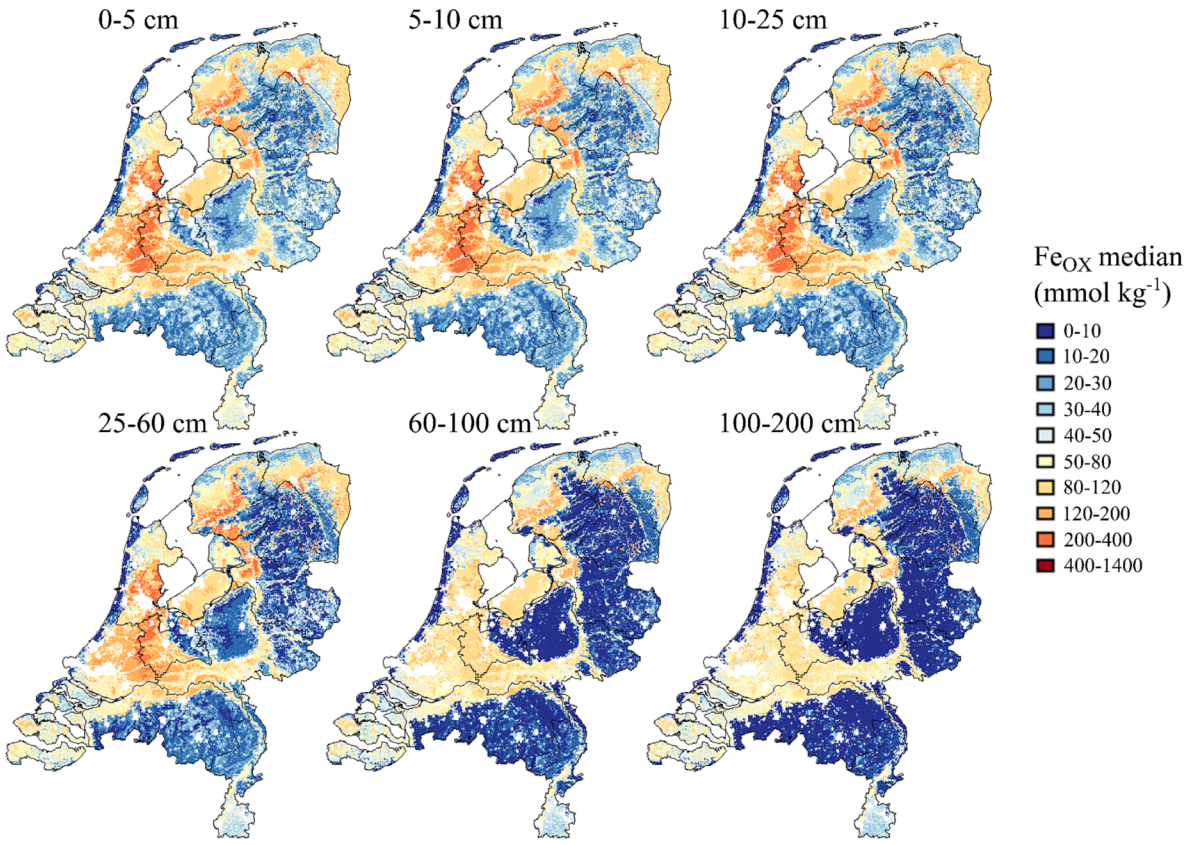

Fig. 3. Spatial predictions of the median Fe_{OX} contents across six depth layers.

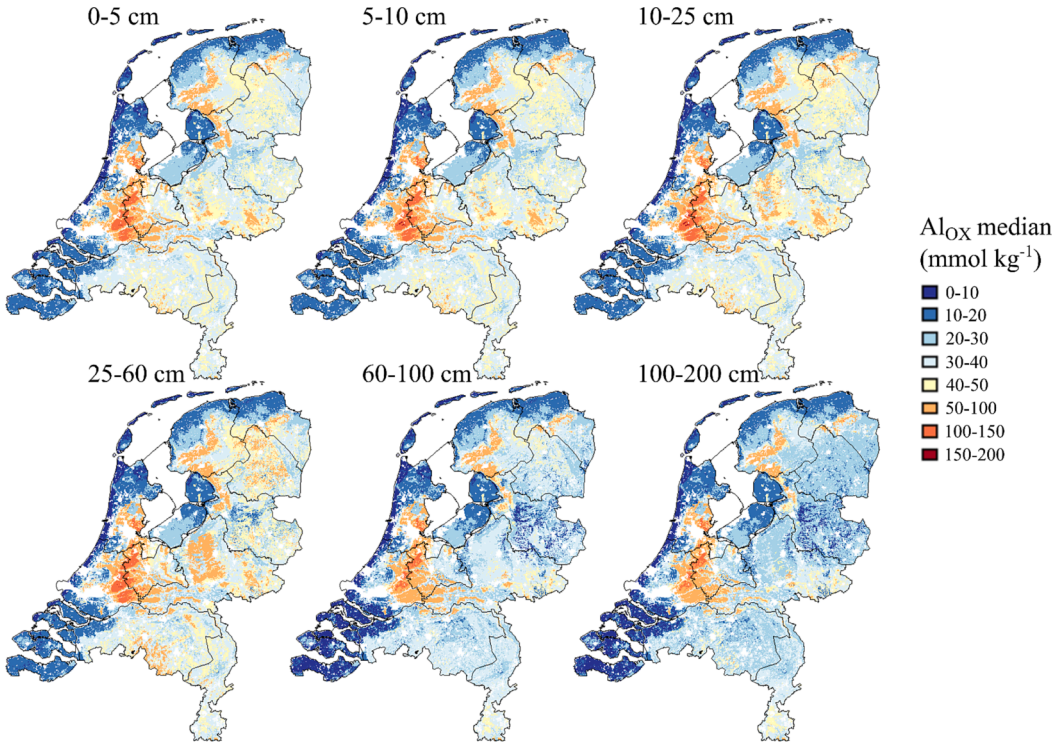

Fig. 4. Spatial predictions of the median Al_{OX} contents across six depth layers.

illustrate this, average (median) predictions of Fe_{OX} contents in the sandy soils decrease in the order of 0–10 cm (30 mmol kg^{-1}), 10–25 cm (29 mmol kg^{-1}), 25–60 cm (22 mmol kg^{-1}), 60–100 cm (10 mmol kg^{-1}) to 100–200 cm (8 mmol kg^{-1}) (Table S2). For Al_{OX} contents, decreasing predictions with soil depth are more pronounced for peat and peaty soils compared to other soil types (Table S3).

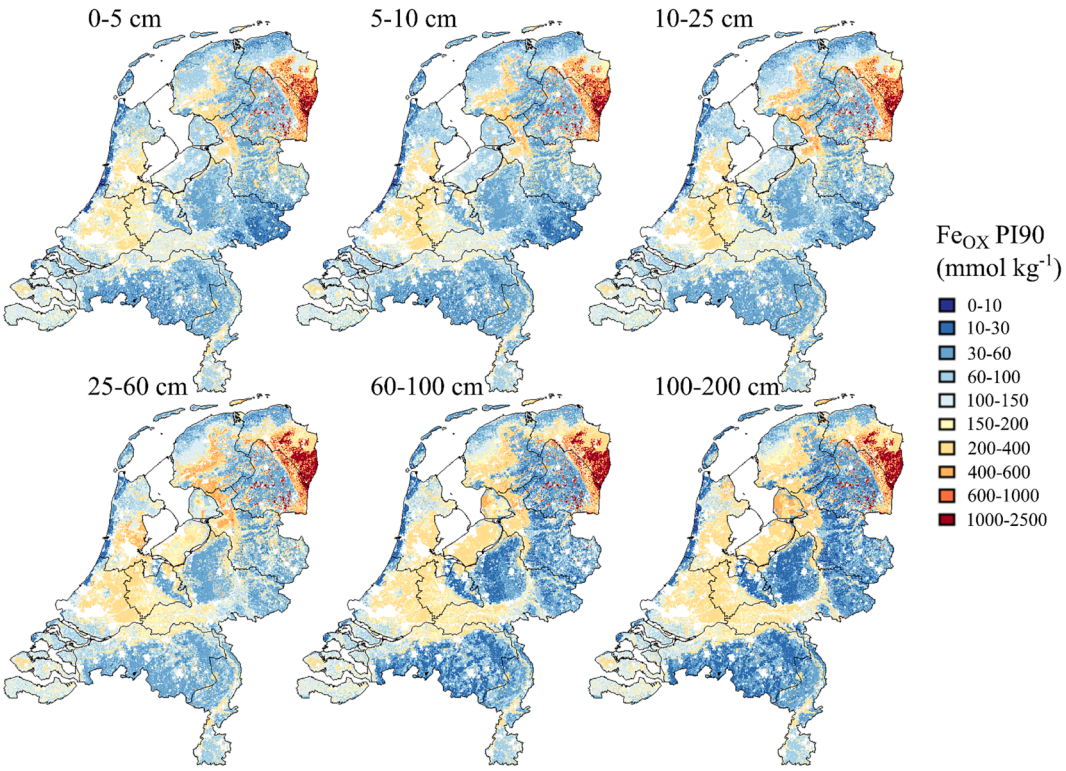

Fig. 5. Spatial predictions of the absolute prediction uncertainty (PI90) of Fe_{OX} contents across six depth layers.

3.2. Prediction uncertainty of amorphous Fe- and Al-(hydr)oxide contents

Absolute prediction uncertainty – operationalised as the PI90 – generally increased with soil depth for Fe_{OX} (Fig. 5) and Al_{OX} (Fig. 6). In addition, the PI90 was observed to depend on the predicted metal oxide content. For example, low prediction uncertainties are observed for Fe_{OX} in the sandy soils where predictions of Fe_{OX} contents are low (comparison of Fig. 3 and Fig. 5) and low prediction uncertainties are observed for Al_{OX} in the marine calcareous soils where predictions of Al_{OX}

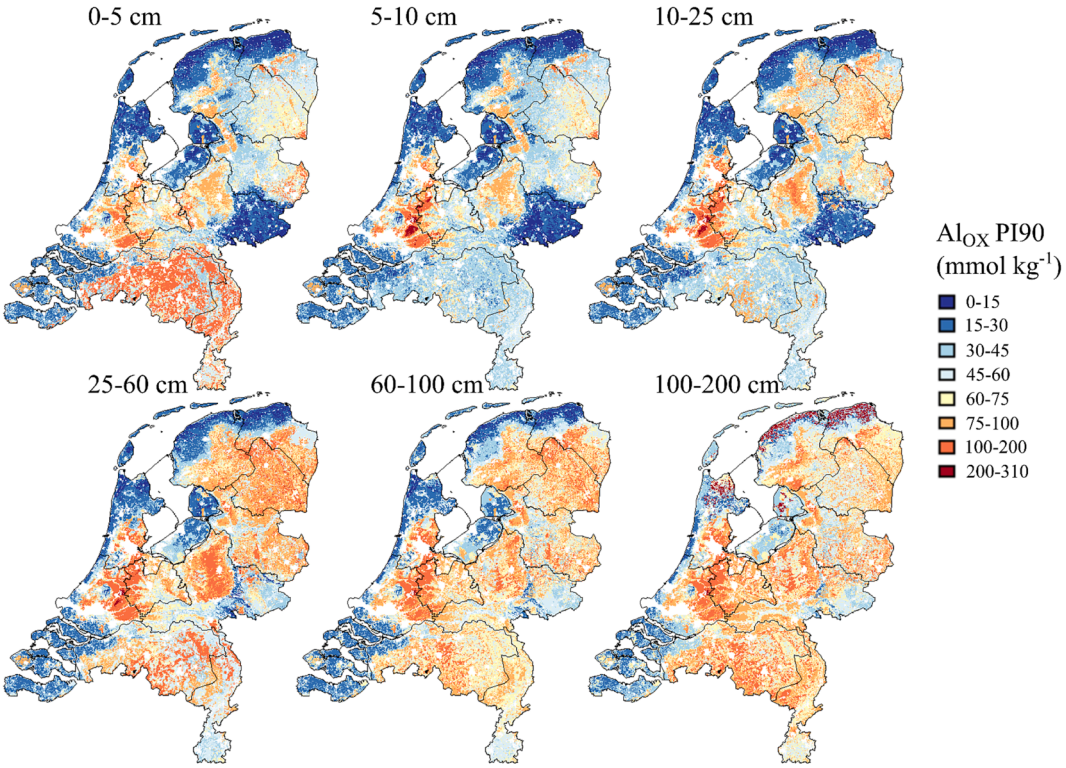

Fig. 6. Spatial predictions of the absolute prediction uncertainty (PI90) of Al_{OX} contents across six depth layers.

contents are low (comparison of Fig. 4 and Fig. 6). Quantifying prediction uncertainty as relative uncertainty (PIR) provides another perspective. For example, PIR (maps included in the SI, Fig. S4 for Fe_{OX} contents and Fig. S5 for Al_{OX} contents) is relatively low in the peat region whereas predicted oxide contents and PI90 are relatively high. In addition, there is a stronger relationship between PIR and soil sample density (Fig. 2) compared to PI90.

Especially in regions where both absolute prediction uncertainties (PI90) and relative prediction uncertainties (PIR) were highest, mean predictions were substantially higher than the median predictions. For Fe_{OX} contents, both PI90 and PIR were highest in the former reclaimed peat areas in the northeast of the Netherlands. In those areas, mean predictions were generally 100 to 630 mmol kg^{-1} higher than the median predictions (Fig. S6). For Al_{OX} contents, the same phenomenon was observed for sandy soils in the south (predictions for 0–5 cm) and clay soils in the north (predictions for 100–200 cm). In those areas, mean predictions were generally 20 to 40 mmol kg^{-1} higher compared to median predictions (Fig. S7).

3.3. Map quality of amorphous Fe- and Al-(hydr)oxide contents

Assessing map quality using design-based statistical inference on the LSK dataset showed that map quality differed for Fe_{OX} and Al_{OX} contents and across soil depth (Fig. 7, predicted vs measured plots included in Fig. S8). Map quality evaluation metrics were better for median predictions compared to mean predictions (Fig. 7 and Table S4). The remaining part of this section therefore focuses on the evaluation metrics of median predictions.

The model slightly underpredicted Fe_{OX} contents for the depth layers 5–10, 10–25, 25–60 and 60–100 cm, where bias decreased with soil depth (ME values of 6.8, 1.6, 0.8 and 0.7 mmol kg^{-1} for aforementioned depth layers, respectively). The model slightly overpredicted Fe_{OX} contents for the 100–200 cm depth layer (ME = −4.4 mmol kg^{-1}). Based on the combination of MEC and RMSE, map quality generally decreases with soil depth. To illustrate this, MEC is highest for the depth layer 5–10 cm (MEC = 0.66), followed by depth layers 10–25 cm (MEC = 0.63), 25–60 (MEC = 0.48), 60–100 (MEC = 0.40) and 100–200 cm (MEC = 0.19). RMSE varied between 46.1 and 56.9 mmol kg^{-1} across the soil depth layers. The model predicted uncertainty with high accuracy, shown by PICP of the PI90 values fluctuating around 0.90 across the soil depth layers (varying between 0.87 and 0.91).

The model slightly underpredicted Al_{OX} contents for the topsoil depth layers 5–10, 10–25 and 25–60 cm (ME values of 1.1, 2.5 and 1.8 mmol kg^{-1}, respectively) and slightly overpredicted Al_{OX} contents for the deeper soil layers 60–100 and 100–200 cm (ME values of −2.6 and −6.8 mmol kg^{-1}, respectively). Similar to Fe_{OX} contents, map quality is highest for the upper soil depth layers. To illustrate this, MEC was highest for the depth layer 5–10 cm (MEC = 0.80) followed by depth layers 10–25 (MEC = 0.61), 25–60 (MEC = 0.49), 60–100 (MEC = 0.38) and 100–200 cm (MEC = 0.33). Uncertainty was generally slightly underpredicted. This is shown by PICP of the PI90 values varying between 0.85 and 0.92 across the soil depth layers.

4. Discussion

4.1. Spatial variation in amorphous Fe- and Al-(hydr)oxide contents

The spatial variation of both Fe_{OX} and Al_{OX} predictions in the Netherlands is large. High predictions of Fe_{OX} and Al_{OX} contents in peat soils (Fig. 3 and Fig. 4) are likely the result of the concentration of Fe and Al as organic matter mineralizes from the topsoil and due to the oxalate-extraction extracting organically bound Fe and Al in addition to amorphous Fe- and Al-(hydr)oxides. To make agricultural production possible, peat soils have been drained. Due to the introduction of oxygen, peat mineralisation is enhanced. When the top peat layer is mineralized, amorphous Fe -and Al-(hydr)oxides are likely to remain in the topsoil, leading to an increase in their soil contents. Regarding the dissolution of organically bound Fe and Al during oxalate extraction, Schoumans (1999) investigated the Fe- and Al-speciation in Dutch peat soils by extracting Fe and Al from the soil samples using a sequential extraction scheme. In addition to oxalate extraction, a Na-pyrophosphate solution (Bascomb, 1968; McKeague, 1967) was used to estimate organically bound Fe and Al. In this study, Na-pyrophosphate extracted 59% and 89% of Fe_{OX} and Al_{OX} respectively in the soil samples with the highest organic matter contents. Indeed, the difference between oxalate-extractable and pyrophosphate-extractable Fe has been proposed to provide a more specific measure of amorphous Fe-(hydr)oxides in soils compared to extracting Fe with an oxalate solution only (McKeague et al., 1971). For Al, however, the pyrophosphate extraction method has been observed to be less specific for Al-organic complexes (McKeague et al., 1971). This implies that a considerable proportion of oxalate-extractable Fe and Al are organically bound, suggesting that an oxalate extraction overestimates the amount of amorphous Fe- and Al-(hydr)oxides in these soils.

In pure peat soils with organic matter contents above 90%, amorphous Fe- and Al-(hydr)oxide contents are expectedly generally low

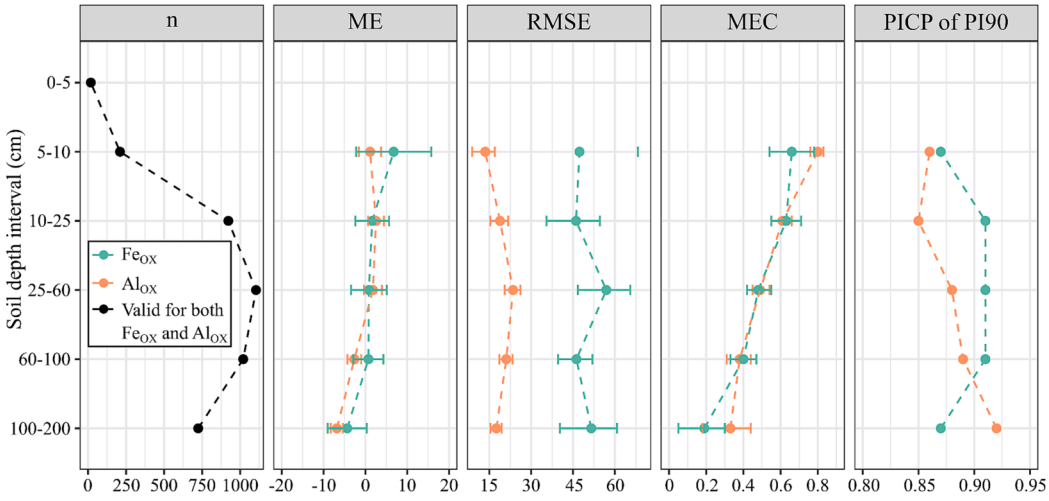

Fig. 7. Number of observations (n), Mean Error (ME), Root Mean Square Error (RMSE), Model Efficiency Coefficient (MEC) and the Prediction Interval Coverage Probability (PICP) of the 90 per cent Prediction Interval (PI90) using hold-out data from a probability sample (LSK dataset). Metrics are calculated with design-based statistical inference for both Fe_{OX} and Al_{OX} contents across soil depth. Error bars show the 95% confidence interval.

since there would be no primary minerals present which could weather to secondary minerals. However, in the Netherlands, soils are classified as peat from soil organic matter contents of 25–35%, depending on the type of peat (Heinen et al., 2022). In addition, most peat soils in the Netherlands are formed in coastal environments. Consequently, most peat soils contain a mixture of peat and clay. Over time, Al can be liberated from those clay minerals and form amorphous Al-(hydr)oxides or form complexes with soil organic matter. In addition, it is expected that there is a significant input of iron through seepage (Møller et al., 2023).

The observation that the spatial variation of Fe_{OX} predictions largely follows the spatial variation in soil organic matter (Fig. 3) is also observed in Denmark (Møller et al., 2023). Consequently, predicted contents are generally highest in peat soils and lowest in sandy soils (Fig. 3). For sandy soils specifically and in line with Lookman et al. (1995), we observed that predictions of Fe_{OX} contents were generally lower in poor to weak loamy sandy soils compared to strong loamy sandy soils.

Low predictions of Al_{OX} contents in clay/loam soils (Fig. 4) are also observed in Denmark where predictions of Al_{OX} contents were low in loamy morainic deposits and relatively young marine and aeolian deposits (Møller et al., 2023). However, this observation is in contrast with the observation in Belgium that predictions of Al_{OX} contents increased with an increase in clay and loam content (Lookman et al., 1995). Differences in observations may be explained due to soil calcareousness. Specifically, we observe that Al_{OX} contents are especially low in calcareous soils (e.g. calcareous clay soils) which could be explained by calcium carbonates inhibiting the weathering rate of aluminosilicates (commonly feldspars) to secondary minerals. To illustrate this, feldspars can chemically weather to the secondary mineral kaolinite through hydrolysis, especially under acid conditions and if base cations such as Ca^{2+} and Mg^{2+} are low (Schaetzl and Anderson, 2005). In turn, the (unstable) kaolinite can weather to (more stable) gibbsite (Schaetzl and Anderson, 2005). Indeed, predictions of Al_{OX} contents were generally low in the eastern part of Denmark where pH (measured in a suspension of soil with distilled water) is generally high (7.1 to 8.0) (Adhikari et al., 2014; Gomes et al., 2023).

General decreasing Fe_{OX} and Al_{OX} content predictions with soil depth are in line with observations in Belgium (Lookman et al., 1995). In Denmark, no clear decreases in Fe_{OX} contents with an increase in soil depth were observed, though Fe_{OX} content predictions were generally highest in topsoil (0–25 cm) and lowest in the deepest (75–100 cm) soil layer (Møller et al., 2023). Regarding Al_{OX}, Møller et al. (2023) observed that predictions of Al_{OX} contents were highest in the 25–50 cm layer and that Al_{OX} contents were lowest in the deepest soil layer. Lookman et al. (1995) describe that decreasing predictions of Fe_{OX} and Al_{OX} contents with an increase in soil depth are likely explained due to weathering differences. Specifically, Fe and Al might more easily weather from secondary minerals to amorphous Fe- and Al-(hydr)oxides in topsoil compared to deeper soil layers due to higher (i) soil temperature, (ii) levels of organic acids and (iii) organic matter contents in the topsoil (Lookman et al., 1995).

Ranges of the median predictions indicate smoothing of the distribution tails. Specifically, median predictions of Fe_{OX} and Al_{OX} contents ranged from 0–1400 and 0–200 mmol kg^{-1}, respectively (Fig. 3 and Fig. 4) whereas ranges of the observations in the data used for model training were wider, i.e. 0–2457 mmol kg^{-1} for Fe_{OX} and 0–388 mmol kg^{-1} for Al_{OX} (Table 1). Smoothing of distribution tails is a well-known observation in DSM studies and it remains a challenge to adequately deal with this in the analytical procedures (Nussbaum et al., 2023). However, in this study insights into possible non-smoothed 'truths' can be retrieved by investigating the predicted probability distribution and corresponding uncertainty measures, i.e. prediction of quantiles, PI90 and PIR. The ranges in soil Fe_{OX} and Al_{OX} contents in the datasets used for model training matched the range in the predicted probability distribution. Specifically, the lowest $q_{0.5}$ prediction corresponded to the lowest value of the range and the highest $q_{0.95}$ prediction corresponded to the highest value of the range.

4.2. Prediction uncertainty of amorphous Fe- and Al-(hydr)oxide contents

Absolute prediction uncertainty (PI90) generally increasing with soil depth (Fig. 5 and Fig. 6) is likely the result of a decreasing soil sampling density with depth (Fig. S1). However, an increase in PI90 with soil depth was not always observed due to decreasing predictions of soil oxide contents over depth and PI90 being partially dependent on the median prediction. Indeed, absolute prediction uncertainties (PI90) were generally higher at higher predictions of Fe_{OX} and Al_{OX} contents (Fig. 5 and Fig. 6) whereas relative prediction uncertainties (PIR) were related stronger to sample intensity (Fig. S4 and S5). For this reason, in the sandy soils decreasing predictions of Fe_{OX} contents with soil depth were paired with a decreasing PI90 (Fig. 5) and an increasing PIR (Fig. S4). Since PI90 and PIR provide different views on uncertainty (absolute vs relative), it is recommended that DSM users consider which uncertainty measure(s) best fit(s) with their objective and that DSM producers share information on both PI90 and PIR.

Mean predictions were substantially higher than median predictions especially in regions where both PI90 and PIR were high (Fig. S6 and S7). This may be the result of regional high short-distance spatial variations in soil oxide contents in combination with a low sample density (high nugget). To illustrate this, mean predictions of Fe_{OX} contents were substantially higher than the median predictions in the former reclaimed peat areas in the northeast of the Netherlands where both PI90 (Fig. 5) and PIR (Fig. S4) were highest (Fig. S6). In this area, sample intensity was low (Fig. 2). In addition, the history of anthropogenic activity in this area has led to a large short-distance variation in the presence of peat in the two and three-dimensional space (see e.g. Makken and Steur (1977) and Van Beek et al. (2015) for details). This has likely led to a strongly positively skewed distribution of both soil organic matter and soil Fe_{OX} contents. Indeed, we observed that Fe_{OX} predictions of individual trees in the model were positively skewed. Aggregating predictions using the mean would thus in most cases overestimate Fe_{OX} contents. Median predictions would generally be closer to the 'true' Fe_{OX} content but will at times (e.g. when a peaty soil layer is present) strongly underestimate Fe_{OX} contents.

4.3. Map quality of amorphous Fe- and Al-(hydr)oxide contents

The observation that map quality evaluation metrics were better for median compared to mean predictions was unexpected. Since data of Fe_{OX} and Al_{OX} contents are positively skewed (Table 1), median predictions were expected to systemically underpredict Fe_{OX} and Al_{OX} contents whereas mean predictions were expected to be less biased. However, mean predictions generally overpredicted Fe_{OX} and Al_{OX} contents, with ME values ranging from −26.3 to −6.2 mmol kg^{-1} for Fe_{OX} contents and −11.6 to 0.66 mmol kg^{-1} for Al_{OX} contents, depending on the soil depth layer (Table S4), whereas ME values for median predictions were closer to 0 mmol kg^{-1} (Fig. 7). Mean predictions overestimating Fe_{OX} and Al_{OX} contents could be explained by the presence of outliers in the data used for model training. Indeed, 51 outliers were removed from the datasets used for model training and validation due to extremely high reported Fe_{OX} and Al_{OX} contents which suggests that data were reported in the wrong unit (Text S2). Possibly, some outliers were still present, for example soil type-specific outliers which did not seem outliers when investigating the distribution of Fe_{OX} and Al_{OX} contents of all data where soil types were lumped (Fig. S2). Another possibility is that NIR estimates of Fe_{OX} and Al_{OX} contents were biased leading to biased model predictions. However, bias propagation is not expected since a covariate was included stating if Fe_{OX} and Al_{OX} using wet-chemical analysis or using NIR (Nussbaum et al., 2018).

Decreasing map quality with depth (Fig. 7) is likely the result of a

decreasing soil sampling density with depth (Fig. S1). Indeed, decreases in map quality were less pronounced in Møller et al. (2023), where observations of Fe_{OX} and Al_{OX} contents were distributed more equally over soil depth compared to our data. This suggests that decreases in map quality could be prevented by assigning larger weights to observations present in sparsely sampled depth intervals. In the two-dimensional space, cross-validation methods have already been proposed which correct for sample intensity, i.e. provide higher weights to observations present in sparsely sampled areas (de Bruin et al., 2022). Similar methodologies could be extended to the three-dimensional space. However, this might lead to an increased subsoil map quality at the cost of a decreased topsoil map quality, which may be undesirable.

For Al_{OX}, prediction uncertainty was generally slightly underestimated (Fig. 7). A slight underprediction of uncertainty can be explained by the use of NIR data for model training. A recent study observed that increasing the proportion of spectral data decreased the PI90 and to PICP of the PI90 dropping below 0.90 for some models (Chen et al., 2023). They argue that this is the result of spectral data having a smaller variability compared to wet-chemical data due to a smoothing effect of models predicting soil properties from spectra. Consequently, when NIR data are merged with wet-chemical data, the overall variability of the data decreases.

Whereas predictions of amorphous Fe- and Al-(hydr)oxide contents are paired with uncertainty predictions and evaluated using design-based statistical inference, accuracy metrics should be evaluated by map users to decide whether the maps are suitable for decision-making. We recommend users to consult overall map quality evaluation metrics and spatially explicit uncertainty estimates (PI90 and PIR) for their area of interest to decide whether the maps are suitable for their intention.

4.4. Hyperparameter optimisation and covariate importance

The combination of randomly sampling hyperparameter values using a random grid search followed by Bayesian optimisation showed to be a time-efficient approach for hyperparameter optimisation (Fig. S9). Final hyperparameter values were different for Fe_{OX} and Al_{OX} (final hyperparameter values reported in Table S5). For example, the number of covariates randomly selected as candidates for each split (mtry) was higher for Fe_{OX} than for Al_{OX}. In addition, the splitting rule (splitrule) was different (splitrules 'extratrees' and 'variance' for Fe_{OX} and Al_{OX}, respectively). Thus, the model for Fe_{OX} consisted of extremely randomized trees (Geurts et al., 2006). In addition, when the tree in a model for Fe_{OX} built a split for a covariate, 8 random split-points were chosen (num.random.splits) after which the best split was selected. For Al_{OX}, the best split was selected directly. For both Fe_{OX} and Al_{OX}, the final maximal tree depth was high (33 and 29, respectively) implying complex relationships between covariates and the soil oxide contents. Allowing trees to create many splits increases the risk of overfitting, but this was prevented by minimizing the weighted RMSE computed by comparing predictions with observations using a validation set where observations were located in different spatial clusters compared to the training set (predicted vs measured plots on training and validation set included in Fig. S10).

The most important covariates predicting Fe_{OX} and Al_{OX} contents, according to importance measures computed by permutation during model fitting, were covariates relating to the soil physical units (BOFEK), DEM, variables relating to soil (e.g. soil type, peat status, peat thickness, soil physical properties) and climate (e.g. long-term averages of temperature and precipitation) (Fig. S11 for Fe_{OX} contents and Fig. S12 for Al_{OX} contents). Relative importances of covariates relating to (historical) land use, palaeogeography, and geology were low. For Fe_{OX} contents, covariates relating to soil organisms (earthworm abundance and earthworm taxa) were included in the list of 30 covariates with the highest relative importance whereas this was not observed for Al_{OX} contents. Computed relative importance measures should however be interpreted with caution due to correlations of covariates influencing relative importance measures. Specifically, relative importance measures have been observed to decrease when the number of correlated covariates increases (Gregorutti et al., 2017). Interpretability of relative importance measures can be improved using de-correlation analysis and recursive feature elimination (Gregorutti et al., 2017; Poggio et al., 2021). De-correlation analysis and recursive feature elimination also lead to more parsimonious models. To illustrate this point, in a DSM study where several soil properties were mapped at the global scale (Poggio et al., 2021) de-correlation analysis decreased the number of covariates from over 400 to circa 150. Recursive feature elimination further decreased the number of covariates to 25 to 40 depending on the soil property (Poggio et al., 2021). However, in this study, those procedures were not followed due to the strong focus on model prediction performance, i.e. an increase in model prediction performance was accepted at the cost of a decrease in model parsimony and interpretability.

4.5. Dealing with data differing in observational quality in DSM studies

Spectroscopy is a cost-efficient alternative to wet-chemical analysis for analysing specific soil parameters (Nocita et al., 2015). Commercial laboratories are increasingly making use of NIR to provide a cheaper alternative for soil analysis compared to wet-chemical analysis (Reijneveld et al., 2022). The analytical error of the wet-chemical measure can be vastly different than the error in the spectral-derived measure. The latter has per definition a larger error, but the extent to which the error increases depends on the soil parameter. For DSM studies, the use of multi-source data with different measurement errors – such as the combination of wet-chemical data and NIR data – is likely to occur more frequently. DSM studies which use machine learning models rarely keep differences in measurement error into account, though this issue has recently gained more attention (Hengl et al., 2018; Takoutsing and Heuvelink, 2022; van der Westhuizen et al., 2022; Wadoux et al., 2019). Ideally, the measurement error variance of the data is known, so that it can be used in the statistical procedures, e.g. by using the inverse of the sum of the residual variance and the measurement error variance as weights (Hengl et al., 2018; van der Westhuizen et al., 2022; Wadoux et al., 2019). Commercial laboratories, however, rarely provide uncertainty estimates though efforts are being made by organizations such as the Global Soil Laboratory Network and The International Union of Soil Sciences to provide such information (Arrouays et al., 2020; van Leeuwen et al., 2022).

In this study, we did not know the measurement error variance of the NIR data and chose the pragmatic approach of tuning the case.weights during hyperparameter optimisation and including a covariate stating if Fe_{OX} and Al_{OX} were measured using wet-chemical analysis or using NIR. Notably, initial case.weights of NIR data were low since initial case.weights were set to the inverse of the sampling density while NIR data were strongly clustered in space (Fig. 2B). The inclusion of NIR data, however, led to only a slight increase in model performance. To illustrate this, tuning case.weights of NIR data did not lead to a significant decrease in RMSE on the validation set (Fig. S13). Final case.weight multiplication factors of NIR data were 0.28 and 0.17 for Fe_{OX} and Al_{OX}, respectively (Table S5). The largest effect on RMSE was found when setting case.weights to 0 (excluding NIR data from the training set), leading to small RMSE increases of roughly 1.5 and 0.3 mmol kg^{-1} for Fe_{OX} and Al_{OX}, respectively (Fig. S13). This negligible performance increase when incorporating NIR data might be explained by moderate accuracy of the NIR data. Chen et al. (2023) observed a negligible performance gain ($R^2 < 3\%$) when incorporating spectral data with an R^2 below 0.72, whereas higher performance gains (R^2 5–15%) were observed when incorporating spectral data with R^2s above 0.86. In addition and in contrast with observations in Nussbaum et al. (2018), inclusion of the covariate describing the analytical technique did not affect model performance. Relative importances of this covariate were 0.00% and 0.69% for Fe_{OX} and Al_{OX} contents, respectively. These

findings suggest that future research is needed on how to deal with datasets differing in observational quality but where uncertainty measures are not reported. In this study, incorporating NIR data with wet-chemical data did not lead to a substantial increase in model performance even though more than 100,000 NIR observations were included.

Observational quality of data may also be influenced by the time of the measurement. For example, more recent data may better reflect present conditions and have lower measurement uncertainties compared to older data. In the dataset used for model training, the year in which Fe_{OX} and Al_{OX} contents were measured ranged from 1955 to 2023 (Table 1). When the time of measurement is expected to significantly influence observational quality, some studies only use data collected in a given period for model training or include the time of the measurement as a covariate to construct a space–time model (Chen et al., 2022). The latter methodology allows for the construction of spatio-temporal digital soil maps. In this study, we did not expect a significant influence of the time of measurement on the observational quality since amorphous Fe- and Al-(hydr)oxide contents are rather constant in soils and spatial differences were far bigger than any improvement in the accuracy of wet-chemical methods.

4.6. Using the soil maps to guide sustainable phosphorus and carbon management in practice

The soil health concept pleas that healthy soils contribute to a wide range of sustainability goals and acknowledges the role of soil in water quality, climate change and human health (Lehmann et al., 2020). To assist policy-making and management decisions aiming to improve soil health (e.g. EU Soil Deal, European Commission, 2021a), there is a high need for quantification of soil-health indicators which are (i) relevant for ecosystem functions and services and (ii) informative for management (Lehmann et al., 2020). Digital Soil Mapping can contribute to quantifying those indicators as it's becoming the frontier of prediction and visualisation of soil function traits. In this study, we chose to map Fe_{OX} and Al_{OX} contents since those soil properties influence the retention of phosphorus and carbon in soils. Insights into the spatial variability of phosphorus and carbon retention in soils can be used to balance crop production, water quality and carbon sequestration in agriculture, thus aiding holistic agricultural management decisions where several ecosystem services are balanced.

To balance crop production and water quality, maps of Al_{OX} and Fe_{OX} contents are needed to assess the maximum P sorption capacity (PSC) of soils (van der Zee et al., 1987). Amongst others, information on the PSC can be used to translate agronomic soil P data to risks of P losses to the environment through subsurface leaching and erosion (van Doorn et al., 2023). They showed that the retention of the agronomic P surplus (input minus output) in the soil and the distribution of P between P pools differing in availability depends on the PSC. The relation between P quantity (measured as P_{AL}) could be translated to the Phosphorus Saturation Degree (PSD) and total reversibly bound P (P_{OX}) with a high accuracy when the PSC was known (MEC of 0.77). The Fe_{OX} and Al_{OX} maps thus give context to the agronomic soil P data, which are available at a high spatial and temporal resolution. In addition, van Doorn et al. (2023) proposed to decrease agronomic soil P target levels in soils with a low PSC and in soils with a high PSC. At low PSC, the lower soil P target levels are needed to prevent P from leaching to the water system. At high PSC, lower soil P levels are needed to prevent the build-up of substantial amounts of P in the soil related to the risk of erosion and the judicious use of finite P reserves. This has major implications for agricultural management. For example, in line with the soil health concept, it is reasonable to assume that the soil P status should be substantially reduced on soils where Fe_{OX} and Al_{OX} contents are low to safeguard water quality objectives. In other words, the potential P losses to the water system might outweigh the benefits of fertilizing those soils to an agronomic optimum.

Regarding carbon sequestration, maps of Fe_{OX} and Al_{OX} are needed to gain more insights into the carbon saturation potential of soils and soil carbon stability. Soil carbon can be protected by chemical (binding to soil particles), physical (encapsulation by aggregates) and biochemical (formation of recalcitrant compounds) processes (Six et al., 2002). The concept of carbon saturation is that physicochemical properties constrain the amount of carbon which can be stored in soils and that the relationship between carbon inputs and carbon storage is asymptotic (Moinet et al., 2023; Six et al., 2002; Stewart et al., 2007). In addition, some researchers propose that thresholds of soil organic carbon contents are dependent on management practices which influence stabilisation and decomposition rates such as tillage intensity and irrigation (Stewart et al., 2007). For this reason, soil organic carbon contents may plateau at values below the maximum storage limit. In the context of carbon saturation, amorphous Fe- and Al-(hydr)oxides are important physicochemical soil properties to consider as they influence carbon stability through chemical protection. Specifically, amorphous Fe- and Al-(hydr) oxides significantly contribute to the specific surface area of soils, which has been observed to positively correlate with soil organic carbon contents (Kaiser and Guggenberger, 2003; Kirschbaum et al., 2020). When paired with information on the spatial variability of other chemical, physical and biochemical protection mechanisms, this information can be used to approximate spatially explicit soil organic carbon saturation maxima and to better predict changes in soil organic carbon pools over time at certain management practices. Those insights can be used for strategic policy-making aiming to increase soil organic carbon contents on cultivated land, contributing to the objective of the European soil deal to reverse current carbon concentration losses on cultivated land to an increase of 0.1–0.4% per year (European Commission, 2021a).

5. Conclusions

In this study, we developed spatially explicit maps of the soil contents of amorphous Fe- and Al-(hydr)oxides at a 25 m resolution for Dutch agricultural fields across six soil depth intervals (0 to 5 cm, 5 to 10 cm, 10 to 25 cm, 25 to 60 cm, 60 to 100 cm and 100 to 200 cm). The probability distributions as well as the associated prediction uncertainty of soil Fe_{OX} and Al_{OX} contents were quantified. Map quality was assessed with design-based statistical inference. Map quality differed between Fe_{OX} and Al_{OX} contents and the soil depth interval. For Fe_{OX} contents, MEC ranged from 0.19 to 0.66, RMSE from 46.1 to 56.9 mmol kg^{-1}, ME from −4.4 to 6.8 mmol kg^{-1} and PICP of the PI90 from 0.87 to 0.91, depending on the soil depth interval. For Al_{OX} contents, MEC ranged from 0.33 to 0.80, RMSE from 13.5 to 23.5 mmol kg^{-1}, ME from −6.8 to 2.5 mmol kg^{-1} and PICP of the PI90 from 0.85 to 0.92. Overall, map quality was better for topsoil compared to subsoil and better for Al_{OX} contents compared to Fe_{OX} contents. We recommend future studies to use the maps to estimate the maximum phosphorus sorption capacity of soils and to use this information in combination with current agronomic soil phosphorus data to balance crop production and the risk of phosphorus losses from agricultural fields to the water system. Furthermore, we recommend using the maps to help determine the carbon saturation limits of soils, which is crucial to consider in soil carbon sequestration potential assessments.

Data availability statement

Digital Soil Maps of Fe_{OX} and Al_{OX} contents can be accessed through https://doi.org/10.4121/96c54816-4e36-4285-89fd-a63e478f9acd (van Doorn et al., 2024) for the mapped soil depth layers (0–5, 5–10, 10–25, 25–60, 60–100 and 100–200 cm). The maps include predictions of the mean, quantiles ($q_{0.05}$, $q_{0.5}$, $q_{0.95}$), absolute uncertainty (PI90) and relative uncertainty (PIR).

CRediT authorship contribution statement

Maarten van Doorn: Writing – review & editing, Writing – original

draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anatol Helfenstein:** Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis. **Gerard H. Ros:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Gerard B.M. Heuvelink:** Writing – review & editing, Methodology. **Debby A.M.D. van Rotterdam-Los:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Sven E. Verweij:** Writing – review & editing, Resources, Methodology. **Wim de Vries:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Maarten van Doorn reports financial support was provided by Nutriënten Management Instituut. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was financed by the Nutrient Management Institute, Wageningen, the Netherlands.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2024.116838.

References

Adhikari, K., Bou Kheir, R., Greve, M.B., Greve, M.H., Malone, B.P., Minasny, B., McBratney, A.B., 2014) Mapping soil pH and bulk density at multiple soil depths in Denmark. In: GlobalSoilMap: Basis of the Global Spatial Soil Information System – Proceedings of the 1st GlobalSoilMap Conference, pp. 155–160. https://doi.org/10.1201/b16500-31.

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. J. R. Stat. Soc. Ser. B Stat Methodol. 82 (4), 1059–1086. https://doi.org/10.1111/rssb.12377.

Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project specifications. In: GlobalSoilMap: Basis of the Global Spatial Soil Information System – Proceedings of the 1st GlobalSoilMap Conference, pp. 9–12. https://doi.org/10.1201/b16500-4.

Arrouays, D., Poggio, L., Salazar Guerrero, O.A., Mulder, V.L., 2020. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. Geoderma Regional 21, e00265.

Baddeley, A., Rubak, E., Turner, R., 2016. Spatial Point Patterns: Methodology and Applications with R. CRC Press.

Bascomb, C.L., 1968. Distribution of pyrophosphate-extractable iron and organic carbon in soils of various groups. Eur. J. Soil Sci. 19 (2), 251–268. https://doi.org/10.1111/j.1365-2389.1968.tb01538.x.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. J. Plant Nutr. Soil Sci. 168 (1), 21–33. https://doi.org/10.1002/jpln.200421414.

BIS, 2023. Bodem Informatie Systeem Nederland (BIS). https://bodemdata.nl/.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breure, T.S., Haefele, S.M., Hannam, J.A., Corstanje, R., Webster, R., Moreno-Rojas, S., Milne, A.E., 2022. A loss function to evaluate agricultural decision-making under uncertainty: a case study of soil spectroscopy. Precis. Agric. 23 (4), 1333–1353. https://doi.org/10.1007/s11119-022-09887-2.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62 (3), 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. Geoderma 409, 115567. https://doi.org/10.1016/j.geoderma.2021.115567.

Chen, S., Saby, N.P.A., Martin, M.P., Barthès, B.G., Gomez, C., Shi, Z., Arrouays, D., 2023. Integrating additional spectroscopically inferred soil data improves the accuracy of digital soil mapping. Geoderma 433 (116467). https://doi.org/10.1016/j.geoderma.2023.116467.

Cronie, O., Van Lieshout, M.N.M., 2018. A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. Biometrika 105 (2), 455–462. https://doi.org/10.1093/biomet/asy001.

de Bruin, S., Brus, D.J., Heuvelink, G.B.M., van Ebbenhorst Tengbergen, T., Wadoux, A. M.J.C., 2022. Dealing with clustered samples for assessing map accuracy by cross-validation. Eco. Inform. 69 (101665) https://doi.org/10.1016/j.ecoinf.2022.101665.

Dowle, M., Srinivasan, A., 2021. data.table: Extension of "data.frame". R package version 1.14.2. https://cran.r-project.org/package=data.table.

EEA, 2018. WISE Water Framework Directive (data viewer). Surface Water Bodies: Water Body Category and Ecological Status or Potential. https://www.eea.europa.eu/data-and-maps/dashboards/wise-wfd.

EEA, 2019. The European environment – state and outlook 2020. In: The European environment – state and outlook 2020. Publications Office of the European Union. pp. 92–111. https://doi.org/10.2800/96749.

European Commission, 2020. Farm to Fork Strategy. https://food.ec.europa.eu/system/files/2020-05/f2f_action-plan_2020_strategy-info_en.pdf.

European Commission, 2021a. A Soil Deal for Europe: 100 living labs and lighthouses to lead the transition towards healthy soils by 2030. https://research-and-innovation.ec.europa.eu/system/files/2021-09/soil_mission_implementation_plan_final_for_publication.pdf.

European Commission, 2021b. Pathway to a healthy planet for all. EU action plan: "Towards zero pollution for air, water and soil." https://eur-lex.europa.eu/resource.html?uri=cellar:a1c34a56-b314-11eb-8aca-01aa75ed71a1.0001.02/DOC_1&format=PDF.

European Commission, 2022. Key figures on the European food chain – 2022 edition. https://ec.europa.eu/eurostat/en/web/products-key-figures/w/ks-fk-22-001.

European Environment Agency, 2022. Soil monitoring in Europe : indicators and thresholds for soil quality assessments.

Fernandez-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., van Liedekerke, M., Marechal, A., Jonas, A., 2022. LUCAS 2018 Soil module – Presentation of dataset and results. https://doi.org/10.2760/215013.

Finke, P., De Gruijter, J., Visschers, R., 2001. Status 2001 Landelijke Steekproef Kaarteenheden en toepassingen; Gestructureerde bemonstering en karakterisering Nederlandse bodems. https://biblio.ugent.be/publication/306393.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Fukumasu, J., Poeplau, C., Coucheney, E., Jarvis, N., Klöffel, T., Koestel, J., Kätterer, T., Nimblad Svensson, D., Wetterlind, J., Larsbo, M., 2021. Oxalate-extractable aluminum alongside carbon inputs may be a major determinant for organic carbon content in agricultural topsoils in humid continental climate. Geoderma 402, 115345. https://doi.org/10.1016/j.geoderma.2021.115345.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Gomes, L.C., Beucher, A.M., Møller, A.B., Iversen, B.V., Børgesen, C.D., Adetsu, D.V., Sechu, G.L., Heckrath, G.J., Koch, J., Adhikari, K., Knadel, M., Lamandé, M., Greve, M.B., Jensen, N.H., Gutierrez, S., Balstrøm, T., Koganti, T., Roell, Y., Peng, Y., Greve, M.H., 2023. Soil assessment in Denmark: towards soil functional mapping and beyond. Front. Soil Sci. 3, 1090145. https://doi.org/10.3389/fsoil.2023.1090145.

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103 (1–2), 3–26. https://doi.org/10.1016/S0016-7061(01)00067-2.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. Stat. Comput. 27 (3), 659–678. https://doi.org/10.1007/s11222-016-9646-1.

Haase, P., Bowler, D.E., Baker, N.J., Bonada, N., Domisch, S., Garcia Marquez, J.R., Heino, J., Hering, D., Jähnig, S.C., Schmidt-Kloiber, A., Stubbington, R., Altermatt, F., Álvarez-Cabria, M., Amatulli, G., Angeler, D.G., Archambaud-Suard, G., Jorrín, I.A., Aspin, T., Azpiroz, I., Welti, E.A.R., 2023. The recovery of European freshwater biodiversity has come to a halt. Nature 620 (7974), 582–588. https://doi.org/10.1038/s41586-023-06400-1.

Heinen, M., Mulder, H.M., Bakker, G., Wösten, J.H.M., Brouwer, F., Teuling, K., Walvoort, D.J.J., 2022. The Dutch soil physical units map: BOFEK. Geoderma 427 (116123). https://doi.org/10.1016/j.geoderma.2022.116123.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D., 2023. 3D Space and Time Mapping Reveals Soil Organic Matter Decrease in Anthropogenic Landscapes. Under Review for Communications Earth and Environment.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J.P., 2022. Tier 4 maps of soil pH at 25 m resolution for the Netherlands. Geoderma 410 (115659). https://doi.org/10.1016/j.geoderma.2021.115659.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120 (1–2), 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018.

Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12 (2), 1–40. https://doi.org/10.1371/journal.pone.0169748.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6 (e5518). https://doi.org/10.7717/peerj.5518.

Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G. E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.M., Chamberlin, J., Silatsa, F.B.T., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11 (1) https://doi.org/10.1038/s41598-021-85639-y.

Hiemstra, T., Antelo, J., Rahnemaie, R., Riemsdijk, W.H.v., 2010. Nanoparticles in natural systems I: The effective reactive surface area of the natural oxide fraction in field samples. Geochim. Cosmochim. Acta 74 (1), 41–58. https://doi.org/10.1016/j.gca.2009.10.018.

Hijmans, R.J., 2022. terra: Spatial Data Analysis. https://cran.r-project.org/package=terra.

Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology. In: Geographical Review (vol. 35, Issue 2). McGraw-Hill. https://doi.org/10.2307/211491.

Kaiser, K., Guggenberger, G., 2003. Mineral surfaces and soil organic matter. Eur. J. Soil Sci. 54, 219–236. https://doi.org/10.1046/j.1365-2389.2003.00544.x.

Kirschbaum, M.U.F., Moinet, G.Y.K., Hedley, C.B., Beare, M.H., McNally, S.R., 2020. A conceptual model of carbon stabilisation based on patterns observed in different soils. Soil Biol. Biochem. 141 (107683) https://doi.org/10.1016/j.soilbio.2019.107683.

Kirsten, M., Mikutta, R., Vogel, C., Thompson, A., Mueller, C.W., Kimaro, D.N., Bergsma, H.L.T., Feger, K.H., Kalbitz, K., 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. Sci. Rep. 11 (5076) https://doi.org/10.1038/s41598-021-84777-7.

Kleinman, P.J.A., 2017. The persistent environmental relevance of soil phosphorus sorption saturation. Curr. Pollut. Rep. 3, 141–150. https://doi.org/10.1007/s40726-017-0058-4.

Kuhn, M., 2022. caret: Classification and Regression Training (6.0-93). https://cran.r-project.org/package=caret.

Lal, R., Bouma, J., Brevik, E., Dawson, L., Field, D.J., Glaser, B., Hatano, R., Hartemink, A.E., Kosaki, T., Lascelles, B., Monger, C., Muggler, C., Ndzana, G.M., Norra, S., Pan, X., Paradelo, R., Reyes-Sánchez, L.B., Sandén, T., Singh, B.R., Zhang, J., 2021. Soils and sustainable development goals of the United Nations: an International Union of Soil Sciences perspective. Geoderma Regional 25 (e00398). https://doi.org/10.1016/j.geodrs.2021.e00398.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. Nat. Rev. Earth Environ. 1, 544–553. https://doi.org/10.1038/s43017-020-0080-8.

Li, Q., Hu, W., Li, L., Li, Y., 2023. Interactions between organic matter and Fe oxides at soil micro-interfaces: quantification, associations and influencing factors. Sci. Total Environ. 855, 158710 https://doi.org/10.1016/j.scitotenv.2022.158710.

Lookman, R., Vandeweert, N., Merckx, R., Vlassak, K., 1995. Geostatistical assessment of the regional distribution of phosphate sorption capacity parameters (FeOX and AlOX) in northern Belgium. Geoderma 66 (3–4), 285–296. https://doi.org/10.1016/0016-7061(94)00084-N.

Ma, G., Ding, J., Han, L., Zhang, Z., Ran, S., 2021a. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. Regional Sustain. 2 (2), 177–188. https://doi.org/10.1016/j.regsus.2021.06.001.

Ma, Y., Minasny, B., McBratney, A., Poggio, L., Fajardo, M., 2021b. Predicting soil properties in 3D: should depth be a covariate? Geoderma 383 (114794). https://doi.org/10.1016/j.geoderma.2020.114794.

Makken, H., Steur, G.G.L., 1977. Toeliching bij de legenda van de concept bodemkaart en grondwatertrappenkaart van het reconstructiegebied oost-groningen en groningsdrentse veenkoloniën. https://edepot.wur.nl/500405.

Malkomes, G., Schaff, C., Garnett, R., 2016. Bayesian optimization for automated model selection. In: Advances in Neural Information Processing Systems, p. 29.

Masiello, C.A., Chadwick, O.A., Southon, J., Torn, M.S., Harden, J.W., 2004. Weathering controls on mechanisms of carbon storage in grassland soils. Global Biogeochem. Cycles 18 (4), 1–9. https://doi.org/10.1029/2004GB002219.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

McKeague, J.A., 1967. An evaluation of 0.1M pyrophosphate and pyrophosphate-dithionite in comparison with oxalate as extractants of the accumulation products in podzols and some other soils. Can. J. Soil Sci. 47 (2), 95–99. https://doi.org/10.4141/cjss67-017.

McKeague, J.A., Brydon, J.E., Miles, N.M., 1971. Differentiation of forms of extractable iron and aluminum in soils. Soil Sci. Soci. Am. 35 (1), 33–38. https://doi.org/10.2136/sssaj1971.03615995003500010016x.

Mehra, O.P., Jackson, M.L., 1958. Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. Clay Clay Miner. 7 (1), 317–327. https://doi.org/10.1346/ccmn.1958.0070122.

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Mendez, J.C., Van Eynde, E., Hiemstra, T., Comans, R.N.J., 2022. Surface reactivity of the natural metal (hydr)oxides in weathered tropical soils. Geoderma 406 (115517). https://doi.org/10.1016/j.geoderma.2021.115517.

Microsoft, Weston, S., 2022b. Foreach Parallel Adaptor for the "parallel" Package (1.0.17; p. 1.0.17). https://cran.r-project.org/package=doParallel.

Microsoft, Weston, S., 2022a. foreach: Provides Foreach Looping Construct (1.5.2). https://cran.r-project.org/package=foreach.

Moinet, G.Y.K., Hijbeek, R., van Vuuren, D.P., Giller, K.E., 2023. Carbon for soils, not soils for carbon. Glob. Chang. Biol. 29 (9), 2384–2398. https://doi.org/10.1111/gcb.16570.

Møller, A.B., Beucher, A.M., Pouladi, N., Humlekrog Greve, M., 2020. Oblique geographic coordinates as covariates for digital soil mapping. Soil 6 (2), 269–289. https://doi.org/10.5194/soil-6-269-2020.

Møller, A.B., Heckrath, G., Hermansen, C., Nørgaard, T., de Jonge, L.W., Greve, M.H., 2023. Mapping the phosphorus sorption capacity of Danish soils in four depths with quantile regression forests and uncertainty propagation. Geoderma 430 (116316). https://doi.org/10.1016/j.geoderma.2022.116316.

NMI, 2022. Nationaal Agrarisch Bodem Archief. https://www.nmi-agro.nl/works/nationaal-agrarisch-bodem-archief/.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E.B., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Wetterlind, J., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. Adv. Agron. 132, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002.

Nussbaum, M., Vogel, S., Oechslin, S., Tanner, S., Burgos, S., 2023. Smoothed predicted distributions in digital soil mapping – a comprehensive comparative study to predict soil texture for irrigation. EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-5543. https://doi.org/10.5194/egusphere-egu23-5543.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4, 1–22. https://doi.org/10.5194/soil-4-1-2018.

Odeha, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63 (3–4), 197–214. https://doi.org/10.1016/0016-7061(94)90063-9.

Pargent, F., Pfisterer, F., Thomas, J., Bischl, B., 2022. Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. Comput. Stat. 37 (5), 2671–2692. https://doi.org/10.1007/s00180-022-01207-6.

Pebesma, E., 2018. Simple features for R: standardized support for spatial vector data. R J. 10 (1), 439–446. https://doi.org/10.32614/RJ-2018-009.

Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. 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.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/.

Reijneveld, J.A., van Oostrum, M.J., Brolsma, K.M., Fletcher, D., Oenema, O., 2022. Empower innovations in routine soil testing. Agronomy 12 (1), 1–20. https://doi.org/10.3390/agronomy12010191.

Reijneveld, A., van Wensem, J., Oenema, O., 2009. Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. Geoderma 152 (3–4), 231–238. https://doi.org/10.1016/j.geoderma.2009.06.007.

Ros, G.H., Verweij, S.E., Janssen, S.J.C., De Haan, J., Fujita, Y., 2022. An open soil health assessment framework facilitating sustainable soil management. Environ. Sci. Tech. 56 (23), 17375–17384. https://doi.org/10.1021/acs.est.2c04516.

Schaetzl, R.J., Anderson, S., 2005. Soils – Genesis and Geomorphology. Cambridge University Press.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Schoumans, O.F., Chardon, W.J., 2015. Phosphate saturation degree and accumulation of phosphate in various soil types in the Netherlands. Geoderma 237–238, 325–335. https://doi.org/10.1016/j.geoderma.2014.08.015.

Schoumans, O.F., 1999. Beschrijving van het gedrag van anorganisch fosfaat in veengronden. https://edepot.wur.nl/299815.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Model. 406, 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002.

Schwertmann, U., 1964. Differenzierung der eisenoxide des bodens durch extraktion mit ammoniumoxalat-lösung. Zeitschrift Für Pflanzenernährung, Düngung, Bodenkunde 105 (3), 194–202. https://doi.org/10.1002/jpln.3591050303.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27 (2), 171–197. https://doi.org/10.1191/0309133303pp366ra.

Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of SOM implications for C saturation of soils.pdf. Plant Soil 241, 155–176. https://doi.org/10.1023/A:1016125726789.

Stewart, C.E., Paustian, K., Conant, R.T., Plante, A.F., Six, J., 2007. Soil carbon saturation: concept, evidence and evaluation. Biogeochemistry 86 (1), 19–31. https://doi.org/10.1007/s10533-007-9140-0.

Takoutsing, B., Heuvelink, G.B.M., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428 (116192). https://doi.org/10.1016/j.geoderma.2022.116192.

van Beek, R., Maas, G.J., Van Den Berg, E., 2015. Home turf: an interdisciplinary exploration of the long-term development, use and reclamation of raised bogs in the Netherlands. Landscape History 36 (2), 5–34. https://doi.org/10.1080/01433768.2015.1108024.

van der Westhuizen, S., Heuvelink, G.B.M., Hofmeyr, D.P., Poggio, L., 2022. Measurement error-filtered machine learning in digital soil mapping. Spatial Statistics 47 (100572). https://doi.org/10.1016/j.spasta.2021.100572.

van der Zee, S.E.A.T.M., van Riemsdijk, W.H., de Haan, F.A.M., 1990a. Het protokol fosfaatverzadigde gronden - Deel I: Toelichting. https://edepot.wur.nl/394261.

van der Zee, S.E.A.T.M., van Riemsdijk, W.H., de Haan, F.A.M., 1990b. Het protokol fosfaatverzadigde gronden - Deel II: Technische uitwerking. https://edepot.wur.nl/394250.

van der Zee, S.E.A.T.M., Fokkink, L.G.J., van Riemsdijk, W.H., 1987. A new technique for assessment of reversibly adsorbed phosphate. Soil Sci. Soc. Am. J. 51 (3), 599–604. https://doi.org/10.2136/sssaj1987.03615995005100030057x.

van Doorn, M., Helfenstein, A., Ros, G.H., Heuvelink, G.B.M., van Rotterdam-Los, D., Verweij, S.E., de Vries, W., 2024. Digital Soil Maps underlying the publication High-Resolution Digital Soil Mapping of Amorphous Iron- and Aluminium-(hydr)oxides to Guide Sustainable Phosphorus and Carbon Management. https://doi.org/0.4121/96c54816-4e36-4285-89fd-a63e478f9acd.

van Doorn, M., van Rotterdam, D., Ros, G., Koopmans, G., Smolders, E., de Vries, W., 2023. The phosphorus saturation degree as a universal agronomic and environmental soil P test. Crit. Rev. Environ. Sci. Technol. 54 (5), 385–404. https://doi.org/10.1080/10643389.2023.2240211.

van Grinsven, H.J.M., Tiktak, A., Rougoor, C.W., 2016. Evaluation of the Dutch implementation of the nitrates directive, the water framework directive and the national emission ceilings directive. NJAS – Wageningen J. Life Sci. 78, 69–84. https://doi.org/10.1016/j.njas.2016.03.010.

van Leeuwen, C.C.E., Mulder, V.L., Batjes, N.H., Heuvelink, G.B.M., 2022. Statistical modelling of measurement error in wet chemistry soil data. Eur. J. Soil Sci. 73 (1), 1–17. https://doi.org/10.1111/ejss.13137.

van Rotterdam, A.M.D., Bussink, D.W., Temminghoff, E.J.M., van Riemsdijk, W.H., 2012. Predicting the potential of soils to supply phosphorus by integrating soil chemical processes and standard soil tests. Geoderma 189–190, 617–626. https://doi.org/10.1016/j.geoderma.2012.07.003.

Vigiak, O., Udías, A., Grizzetti, B., Zanni, M., Aloe, A., Weiss, F., Hristov, J., Bisselink, B., de Roo, A., Pistocchi, A., 2023. Recent regional changes in nutrient fluxes of European surface waters. Sci. Total Environ. 858 https://doi.org/10.1016/j.scitotenv.2022.160063. Part(160063).

Wadoux, A.M.J.C., Padarian, J., Minasny, B., 2019. Multi-source data integration for soil mapping using deep learning. Soil 5 (1), 107–119. https://doi.org/10.5194/soil-5-107-2019.

Wadoux, A.M.J.C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020. A note on knowledge discovery and machine learning in digital soil mapping. Eur. J. Soil Sci. 71 (2), 133–136. https://doi.org/10.1111/ejss.12909.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org.

Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of random forests for high dimensional data in C++ and R. R J. Stat. Software 77 (1), 1–17. https://doi.org/10.18637/jss.v077.i01.

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. J. Electron. Sci. Technol. 17 (1), 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120.