






Accounting for analytical and proximal soil sensing errors in digital soil mapping

Bertin Takoutsing^{1,2}  | Gerard B. M. Heuvelink^{1,3}  | Jetse J. Stoorvogel¹  |
Keith D. Shepherd⁴  | Ermias Aynekulu⁵ 

¹Soil Geography and Landscape Group,
Department of Environmental Sciences,
Wageningen University, Wageningen, The
Netherlands

²Land Health Decisions, World
Agroforestry (ICRAF), Yaoundé,
Cameroon

³ISRIC-World Soil Information,
Wageningen, The Netherlands

⁴Innovative Solutions for Decision
Agriculture, Nairobi, Kenya

⁵Land Health Decisions, World
Agroforestry (ICRAF), Nairobi, Kenya

Correspondence

Bertin Takoutsing, Land Health
Decisions, World Agroforestry Centre, BP
16317, Yaoundé, Cameroon.
Email: b.takoutsing@cgiar.org

Funding information

CGIAR Research Program on Water, Land
and Ecosystems (WLE)

Abstract

Digital soil mapping (DSM) approaches provide soil information by utilising the relationship between soil properties and environmental variables. Calibration of DSM models requires measurements that may often have substantial measurement errors which propagate to the DSM outputs and need to be accounted for. This study applied a geostatistical-based DSM approach that incorporates measurement error variances in the covariance structure of the spatial model, weights measurements in accordance with their measurement accuracies and assesses the effects of measurement errors on the accuracies of DSM outputs. The method was applied in the Western Cameroon, where soil samples from 480 locations were collected and analysed for pH, clay and soil organic carbon (SOC) using conventional and mid-infrared spectroscopy methods. Variogram parameters and regression coefficients were estimated using residual maximum likelihood under two scenarios: with and without taking measurement errors into account. Performance of the spatial models in the two scenarios was compared using validation metrics obtained with three types of cross-validation. Acknowledging measurement errors impacted the regression coefficients and influenced the variogram parameters by reducing the nugget and sill variance for the three soil properties. Validation metrics including mean error, root mean square error and model efficiency coefficient were quite similar in both scenarios, but the prediction uncertainties were more realistically quantified by the models that account for measurement errors, as indicated by accuracy plots. There were relatively small absolute differences in predicted values of soil properties of up to 0.1 for pH, 1.6% for clay and 2 g/kg for SOC between the two scenarios. We emphasised the need of incorporating measurement errors in DSM approaches to improve uncertainty quantification, particularly when applying spectroscopy for estimating soil properties. Further development of the approach is the extension to non-linear machine learning regression methods.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *European Journal of Soil Science* published by John Wiley & Sons Ltd on behalf of British Society of Soil Science.

Highlights

- Errors in soil measurements are usually not accounted for and may affect DSM results.
- Measurement error variances were incorporated in the geostatistical models of three soil properties.
- Quantifying measurement errors in DSM allows to weigh measurements in accordance with their accuracy.
- Accounting for measurement errors in DSM better assesses prediction accuracy.

KEYWORDS

measurement error, partial least squares regression, prediction uncertainty, proximal soil sensing, regression kriging, residual maximal likelihood, spatial prediction

1 | INTRODUCTION

Soil spatial information is crucial to address global issues such as food security, climate change and land degradation (McBratney et al., 2003; Shepherd et al., 2015). Soil information is also important at various scales in helping policy makers, extension agents, and land users whose decisions impact land management interventions, particularly those designed to support agricultural production (Stoorvogel et al., 2015). The successes of digital soil mapping (DSM) in providing such information are ascribed to recent technological and computational advances, availability of high-resolution remote sensing data, advancement of proximal soil sensing (PSS), and the development of machine-learning algorithms (MLA) (Minasny & McBratney, 2016). The quest for large soil datasets for the development and application of DSM models, as well as the increasing demand for soil spatial information to efficiently manage agronomic inputs such as fertilisers (Stoorvogel et al., 2015), has led to the increased use of PSS (Viscarra Rossel, Behrens, et al., 2016). Diffuse reflectance spectroscopy is a rapid and low-cost method to generate soil measurements for use in DSM (Shepherd et al., 2015). Despite the potentials of PSS in generating larger amounts of spatially explicit soil data, soil spectral estimation of soil properties tends to have larger measurement errors than wet chemistry measurements that eventually propagate to DSM outputs (Heuvelink, 2018; Somarathna et al., 2018).

One aspect of DSM that has received little attention so far is the errors in soil measurements used for calibration and prediction. Although modellers may be aware that measurements are not error-free, most DSM studies ignore this fact and consider only the limited predictive power of environmental covariates and spatial interpolation error as sources of uncertainties. Analysis and

quantification of uncertainties in soil measurements is a subject of interest and should be incorporated in DSM, to weigh measurements in accordance with their accuracy and provide end-users with reliable information about the accuracy of the prediction maps (Arrouays et al., 2017; Heuvelink, 2018). The lack of consideration of uncertainties may lead to suboptimal models and systematic underestimation or overestimation of the uncertainties of DSM outputs (Heuvelink, 2018; Poggio et al., 2016). Decisions based on suboptimal models and poor quality maps whose accuracy is overestimated may have extensive and profound impacts on the design of land management interventions (Takoutsing et al., 2017), as well as on soil amendment practices, such as fertiliser application. End-users may increase their investments in obtaining accurate soil maps, for instance by increasing soil sampling density or getting better covariates, if they are reliably informed about the accuracy of the available maps. Recent studies have demonstrated that measurement errors may have significant impacts on subsequent spatial analyses (Somarathna et al., 2018). However, to the best of our knowledge, there are no published studies that explicitly considered how uncertainty in PSS data affects DSM outputs.

The recent expansion of DSM approaches has resulted in the shift from geostatistics to machine learning (ML). Although ML has overtaken kriging to become the most popular DSM method due to its flexibility and tendency to improve predictions (Hengl et al., 2015; Veronesi & Schillaci, 2019), kriging has important advantages over ML. First, kriging can better account for spatial autocorrelation than ML, which is not a spatial model (Hengl, Walsh, et al., 2018). Second, it yields an interpretable parametric model of the soil spatial variation. Third, kriging does not need as large a dataset as ML for calibration and can be used in a case of just 100 measurements

or more. Fourth, kriging does not only characterise the prediction uncertainty using a prediction error variance, but it also quantifies the spatial correlation in the kriging prediction errors. At best, ML characterises the prediction error at prediction locations, for example using Quantile Regression Forests (Vaysse & Lagacherie, 2017), but not the spatial correlation of that error, which is needed to quantify uncertainties of spatial averages. Fifth, from a statistical perspective, it is feasible to incorporate measurement errors in model calibration and prediction with kriging (Chilès & Delfiner, 2012; Kotters et al., 1995; Viscarra Rossel, Brus, et al., 2016; Viscarra Rossel & Brus, 2018).

There is need for DSM approaches that account for uncertainties in soil measurements generated using analytical and PSS methods. While geostatistical methods are available to handle this in a realistic manner, they have not so far been used. Therefore, the objectives of this study are to quantify measurement errors in analytical and PSS soil data, incorporate them into a state-of-the-art geostatistical method for spatial interpolation and compare the results with a case in which measurement errors are ignored. We illustrate the methods with a case study in which we map pH, clay and soil organic carbon (SOC) for a study area in the Western Highlands of Cameroon. More specifically, we use regression kriging (RK) supported by restricted maximum likelihood (REML) parameter estimation.

2 | MATERIALS AND METHODS

2.1 | Study area

The study area covers parts of the West region of Cameroon that spans 1053 km² and features the major characteristics of the highlands of Cameroon (Figure 1), dominated by subsistence agricultural systems. The climate is of tropical humid mountain type with average rainfall that varies from 1000 to 2000 mm per year. The mean daily minimum and maximum temperatures are 18 and 30°C, respectively. The topography is undulating with altitudes ranging between 600 and 1800 m above sea level, and the vegetation is of savannah type with patches of gallery and montane forests. The soils are predominantly Ferralsols, of volcanic origin and suitable for the production of a range of annual and perennial crops, though soil tends to be generally acidic (Takoutsing et al., 2016).

2.2 | Sampling design

The study area was sampled using a spatially stratified and hierarchical sampling approach based on the

concept of 10 km × 10 km sentinel sites (Vågen & Winowiecki, 2020). The soil sampling was limited to the sentinel sites, which is suboptimal for kriging since parts of the study area are poorly covered (Figure 1), but this was done to save travel time and accessibility costs. The site locations were established using convenience sampling, while accounting for land cover/land use and topography of the study area to capture the variation of landscape conditions (i.e., feature space coverage). The sampling design for establishing the sentinel sites was initially conceived for a larger area that covers the entire southern parts of the Republic of Cameroon, but for this study, we used only three sites namely Bamendjou, Bana and Kekem located in the study area. As explained in detail in Vågen and Winowiecki (2013), each site was subdivided into 16 square 2.5 km × 2.5 km tiles within which random centroid locations for clusters within each tile were generated, but buffered to avoid overlapping with neighbouring tiles. Each cluster consists of 10 circular sampling plots (1000 m² each) randomly located within a 1 km² circular area randomly placed within the tile, giving 160 sampling plots per sentinel site.

Within each of the sampling plots, four subplots were established, one at the centre of the plot and the three others surrounding the centre plot at 12.2 m and disposed at 120°. Topsoil (0–20 cm) samples (~500 g) were collected at the four subplots, pooled together and thoroughly mixed to obtain a composite sample for each plot, yielding 160 soil samples per site and 480 for the entire study area.

2.3 | Soil data

Soil samples of 480 numbers were collected (10 samples from each cluster, 160 samples from each site) between 2015 and 2017 within the framework of the Cameroon land health project (Takoutsing et al., 2017). One out of ten soil samples in each cluster were randomly selected and subjected to conventional laboratory analyses for pH, clay content and SOC and referred to as 'reference samples' ($n = 48$). Soil pH was determined using a pH meter (1:2.5 soil to water ratio), clay by the hydrometer method, and SOC concentration using the potassium dichromate oxidation method.

Next, all samples ($n = 480$) were processed and analysed by mid-infrared spectroscopy (MIRS), following standard procedures described in Terhoeven-Urselmans et al. (2010). The measured mid-infrared (MIR) reflectances were first converted to apparent absorbance units [$\log(1/\text{Reflectance})$] and then preprocessed with the Savitzky–Golay smoothing method (Sila et al., 2016). The reference samples were used for both calibration and

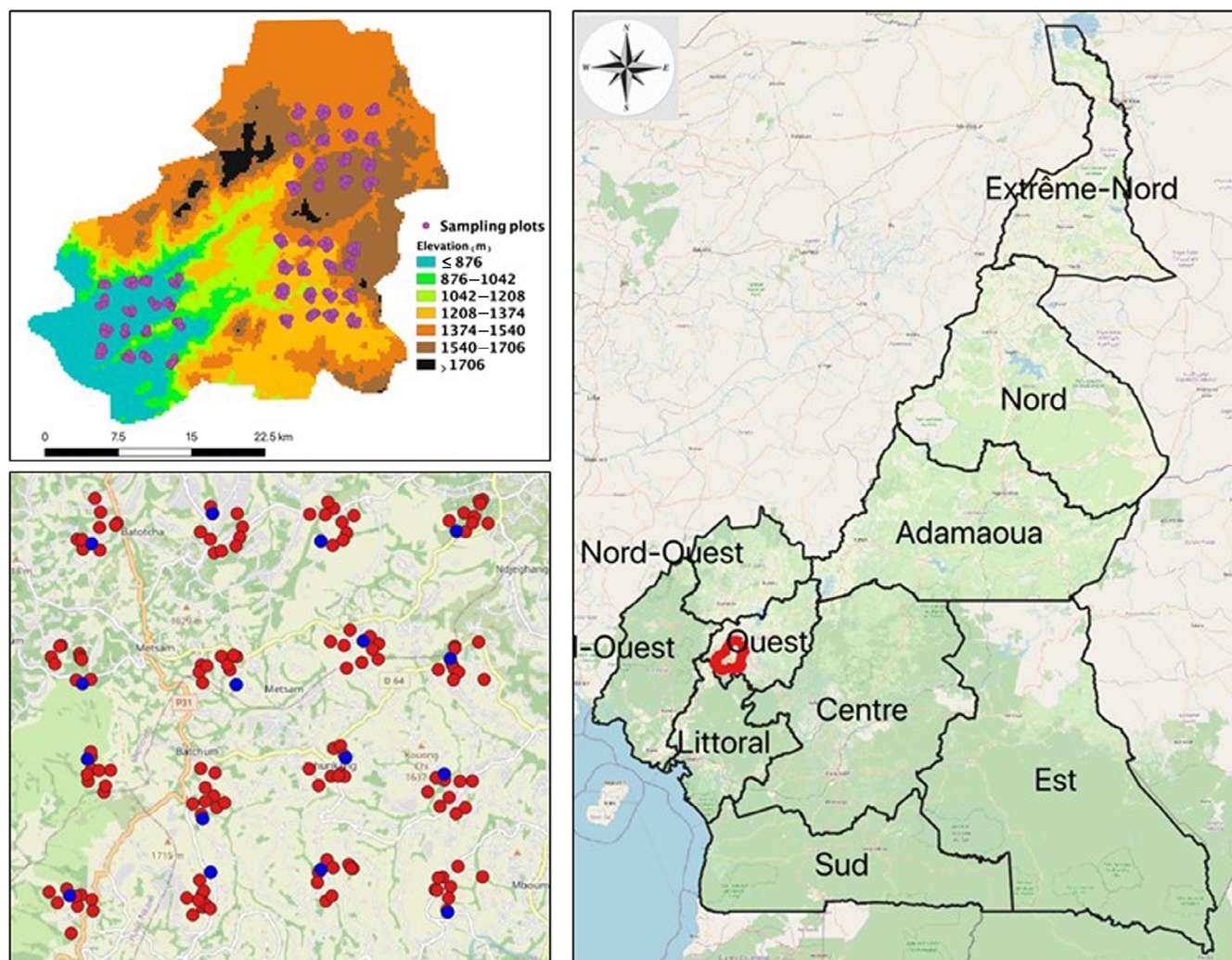


FIGURE 1 Map of Cameroon showing the study area. Soil sampling was done in three 10 km × 10 km sentinel sites. Each sentinel site has 160 sampling locations (violet dots). Bottom-left panel zooms in on the most northern sentinel site (red dots represent spectral data, blue dots analytical data)

validation of the prediction models. All spectral replicates for each sample were averaged and regression models were built to relate the processed spectra to the reference samples using partial least squares regression (PLSR). The PLSR used spectra data as independent variables and the analytical data as dependent variables. The fitted regression models were used to predict the targeted properties of all the samples.

All the 48 paired observations of analytical and spectral data were used for the calibration of the PLSR model and validation metrics were computed using leave-one-out cross-validation. The accuracies of the models were assessed using the mean error (ME), the root mean squared error (RMSE) and the model efficiency coefficient (MEC). See Section 2.5.6 for definitions of these accuracy metrics. The fitted PLSR models were applied to obtain soil property predictions based on MIR data at all

480 locations. Soil measurements obtained through conventional laboratory methods are referred to as analytical data while predicted soil values using MIR spectroscopy/PLSR are referred to as spectral data.

2.4 | Environmental variables

Soil spatial variation is influenced by environmental factors including climate (e.g., precipitation and temperature), organisms (e.g., land cover), relief (e.g., terrain attributes), and parent materials (McBratney et al., 2003). The study derived these factors from several spatial datasets to effectively represent each key soil-forming factor.

We initially considered an extensive stack of over 170 environmental layers downloaded from the

International Soil Reference and Information Centre (ISRIC) repository. The relief represented by a digital elevation model was obtained from Shuttle Radar Topography Mission (SRTM), from which various topographic parameters were derived (e.g., elevation, slope and topographic wetness index). Land use/cover classes were obtained from the global land cover map (GlobeLand30) for the year 2015. The MODIS near and mid-infrared reflectance (NIR, MIR), and Enhanced Vegetation Index (EVI) products were derived using a stack of MOD13Q1 products. Climatic data made up of annual temperature and precipitation averages were obtained from the CHELSA Bioclimatic images (<https://chelsa-climate.org/bioclim/>). Landform classes (breaks/foothills, flat plains, high mountains/deep canyons, hills, low hills, low mountains, smooth plains) were based on the USGS's Map of Global Ecological Land Units. Since the environmental layers were from different sources, they were all resampled to 250 m spatial resolution before they were used as independent variables in the spatial models.

2.5 | Statistical modelling

The DSM model selection, calibration and prediction were fully implemented in the R environment for statistical computing (R Core Team, 2016). The process consisted of the following main steps (Figure S1): (1) model definition, (2) quantification of measurement errors in analytical and spectral data, (3) model selection, (4) model calibration (parameter estimation), (5) spatial prediction and (6) cross-validation. The sub-sections below explain these six steps for two scenarios. In scenario 1, measurement errors are ignored, while scenario 2 accounts for measurement error variances.

2.5.1 | Model definition

In RK, the dependent variable is modelled as the sum of a deterministic trend and a spatially correlated stochastic residual as described in chapter 9 of Webster and Oliver (2007)

$$Z(\mathbf{s}) = m(\mathbf{s}) + \varepsilon(\mathbf{s}) = \sum_{j=0}^p \beta_j \cdot x_j(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad \mathbf{s} \in A \quad (1)$$

Here, $Z(\mathbf{s})$ represents the soil property of interest at any location \mathbf{s} in the geographic domain A , m is the trend, taken as a linear combination of covariates x_j ($j = 1, \dots, p$), the β_j are regression coefficients (β_0 is the intercept, by setting $x_0(\mathbf{s}) = 1$ for all $\mathbf{s} \in A$), and ε is a

zero-mean, normally distributed, stationary stochastic residual, whose spatial covariance structure is defined by a variogram γ . The normal distribution and residual stationarity are stringent assumptions that need to be justified and possibly adjusted in real-world applications. For instance, in the case study we will apply log-transformation to SOC before invoking the normal distribution assumption.

We also have observations y_i of the dependent variable at a finite number of locations \mathbf{s}_i ($i = 1, \dots, n$) in A . These are interpreted as realisations of an observation process:

$$Y_i = Z(\mathbf{s}_i) + \delta_i, \quad i = 1, \dots, n, \quad (2)$$

where δ_i are measurement errors, assumed jointly normally distributed with zero mean and having an $n \times n$ variance-covariance matrix \mathbf{V} . Note that the zero-mean assumption signifies that we ignore systematic errors in soil measurements.

2.5.2 | Quantification of measurement errors in analytical and spectral data

In this study, we considered errors in the measurement of analytical and spectral data of the soil samples. By assuming that measurement errors of different soil samples are uncorrelated, the variance-covariance matrix \mathbf{V} reduces to a diagonal matrix so that errors in soil measurements are completely summarised by their variances. These variances were assumed constant for a given measurement method but were different for analytical and spectral measurements. They were assessed as follows. Let Z_T be the 'true' value of the soil property, Z_A the value of the soil property obtained through laboratory analysis and Z_S the value of the soil property obtained using PSS (i.e., the spectroscopy model predictions of Z_A). We now have:

$$\begin{aligned} \sigma_S^2 &= \text{var}(Z_S - Z_T) = \text{var}(Z_S - Z_A + Z_A - Z_T) \\ &= \text{var}(Z_S - Z_A) + \text{var}(Z_A - Z_T) \\ &\quad + 2 \text{cov}(Z_S - Z_A, Z_A - Z_T) \\ &= \text{var}(Z_S - Z_A) + \text{var}(Z_A - Z_T), \end{aligned} \quad (3)$$

where the latter equality holds because the PLSR fitting error is not correlated with the laboratory measurement error.

For laboratory data, the measurement error variances $\sigma_A^2 = \text{var}(Z_A - Z_T)$ were derived using laboratory repeatability procedures (Libohova et al., 2019). In this context, repeatability describes the variation of a mean result obtained in successive measurements of the same sample

analysed in the same laboratory under the same conditions (Libohova et al., 2019). For this study, soil samples were analysed following standard processing procedures for pH (10 duplicates), clay (33 duplicates) and SOC (10 duplicates). Next, the SD of the analytical measurement error σ_A was estimated from the differences between the measurements on the same sample. We found unrealistic values (outliers) for clay which might be caused by other sources of errors (i.e., blunders, gross errors). These outliers were removed before computing the estimate of σ_A so that our focus remained on real measurement errors.

For spectral data, the measurement error variances σ_S^2 were obtained using Equation (3) by adding up σ_A^2 and $\text{var}(Z_S - Z_A)$. The latter was estimated from the residual variance of the PLSR models. For the case of SOC that showed positive skewness, the analytical data were first log-transformed (logSOC) before running the PLSR. The PLSR residual variance was estimated using:

$$\text{var}(Z_S - Z_A) \cong \frac{1}{n} \sum_{i=1}^n (z_A(\mathbf{s}_i) - \hat{z}_A(\mathbf{s}_i))^2, \quad (4)$$

where $\hat{z}_A(\mathbf{s}_i) = z_S(\mathbf{s}_i)$ is the PLRS predicted soil property at location \mathbf{s}_i and n is the number of paired observations of analytical and spectral data (i.e., $n=48$ in the case study). Both the analytical and spectral measurement error variances estimated above were incorporated in the kriging model as shown in the sections below.

2.5.3 | Model selection

Based on literature, pedological information, and their relevance to specific soil properties, 50 layers were selected out of 170 initial layers, to represent key soil-forming factors. These covariates were processed and overlaid with sample locations to construct a matrix of covariate values for each sample point. Initially, a correlation analysis was performed to reduce redundancy between the selected layers. Some pairs of environmental variables were highly correlated with each other. For statistical models it is preferred that environmental variables retained are weakly correlated with each other, because it increases the potential for fitting a combination of environmental variables to explain the variation in the soil properties. Only layers with a correlation coefficient ≤ 0.75 with all other layers were retained for subsequent analysis (Hanchuan et al., 2005). For each pair of covariates correlated above the set threshold, we arbitrarily retained the first one in alphabetical order for inclusion in the model. This reduced the number of

covariates to 23. Next, the best combination of covariates for each soil property was selected by combined forward and backward stepwise regression using the Bayesian Information Criterion (BIC) (Gao & Song, 2010). Regression models, their coefficients and p-values were examined to derive quantitative data on the relative roles and behaviour of each covariate in the model. During the model selection procedure, measurement errors in soil data and spatial correlation were ignored.

2.5.4 | Model calibration by REML method

Using matrix notation for compactness, Equations (1) and (2) combined can be written as $\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\epsilon} + \boldsymbol{\delta}$, where \mathbf{Y} is an n -vector, \mathbf{X} an $n \times (p+1)$ matrix, $\boldsymbol{\beta}$ a vector of $p+1$ regression coefficients, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ vectors containing the stochastic residual and measurement errors at the n observation locations, respectively. The parameters of this model are $\boldsymbol{\beta}$, the parameters of the variogram γ (i.e., the nugget, sill and range, assuming a known shape), and the error variance–covariance matrix \mathbf{V} (i.e., σ_A^2 and σ_S^2). Estimation of σ_A^2 and σ_S^2 was explained in Section 2.5.2. The variograms were all fitted with exponential models. Note that the measurement error parameters were estimated and fixed prior to estimation of other parameters through REML.

It is a common practice in geostatistics to estimate γ using the method of moments. However, this is sub-optimal and has additional bias problems in case of RK, where $\boldsymbol{\beta}$ also needs to be estimated (Lark et al., 2006). For this study, variogram parameters and regression coefficients were estimated using REML. We give a brief description and refer to section 9.2.1 in Webster and Oliver (2007) for details. REML is computationally demanding in a case of large datasets but was quick in our case study, where we have 480 observations.

Since all stochastic components of the model are normally distributed, \mathbf{Y} has a multivariate normal distribution, with probability density:

$$f_Y(\mathbf{y}) = (2\pi)^{-n/2} \cdot |\mathbf{C} + \mathbf{V}|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{C} + \mathbf{V})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right), \quad (5)$$

where \mathbf{y} is the vector of observations y_i ($i=1, \dots, n$) and \mathbf{C} is the variance–covariance matrix of $\boldsymbol{\epsilon}$, derived from the variogram γ and the distances between the observation locations. The idea of maximum likelihood is to choose the model parameters (i.e., $\boldsymbol{\beta}$ and the nugget, sill, and range of γ) such that the probability density $f_Y(\mathbf{y})$ is maximised. REML is a particular form of maximum

likelihood estimation that estimates the model parameters in two steps. First the variogram parameters are estimated by maximising a conditional likelihood, in which the dependence of the variogram parameters on the regression coefficients is removed (Lark & Cullis, 2004). For this a numerical optimisation technique is used. In a second step the regression coefficients are estimated, conditional on the already estimated variogram parameters. This can be done analytically, because given the variogram parameters, the maximum likelihood estimate of the vector of regression coefficients equals the conventional generalised least squares solution:

$$\hat{\beta} = (\mathbf{X}^T(\mathbf{C} + \mathbf{V})^{-1}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{C} + \mathbf{V})^{-1}\mathbf{y} \quad (6)$$

Note that the model calibration returns only estimates of the ‘true’ regression coefficients and variogram parameters. To simplify the subsequent analysis, we will ignore these estimation errors and assume that all model parameters are perfectly known. While it is not difficult to include estimation errors of regression coefficients in kriging (Brus & Heuvelink, 2007), it is much more difficult to account for variogram estimation errors in prediction.

2.5.5 | Spatial prediction

Our aim is to predict $Z(\mathbf{s}_0)$ given the measurements y_i and the covariates $x_j(\mathbf{s}_0)$. Note that the prediction location \mathbf{s}_0 could be any location in the study area for which covariates are available. In practice, the prediction locations are the nodes of a fine grid that are visited one by one. Under the assumptions made, the best prediction (i.e., the one that has the smallest expected squared prediction error) is the conditional mean:

$$\hat{Z}(\mathbf{s}_0) = E[Z(\mathbf{s}_0)|\mathbf{Y} = \mathbf{y}] = \mathbf{x}_0^T \hat{\beta} + \mathbf{c}_0^T (\mathbf{C} + \mathbf{V})^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}), \quad (7)$$

where \mathbf{x}_0 is a $(p+1)$ -vector of covariates at \mathbf{s}_0 and \mathbf{c}_0 is the vector of covariances between $Z(\mathbf{s}_0)$ and the $Z(\mathbf{s}_i)$. The prediction is unbiased and has prediction error variance:

$$\sigma_K^2(\mathbf{s}_0) = \text{Var}(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)) = c_{00} - \mathbf{c}_0^T (\mathbf{C} + \mathbf{V})^{-1} \mathbf{c}_0, \quad (8)$$

where c_{00} is the variance of $Z(\mathbf{s}_0)$ (i.e., the sill of γ). Note that regression coefficient estimation errors are ignored in Equations (7) and (8).

Spatial interpolation techniques such as kriging are sensitive to skew distributions due to the high impact of extreme values on variogram parameter estimation that may render outputs unstable. From the summary statistics presented in Section 3.1, soil pH and clay satisfactorily met the assumption of a normal distribution. SOC exhibited a positively skewed distribution, and the normal distribution assumption was made after a log-transformation to logSOC. The back-transformed estimate of SOC and local variance for each interpolated location was obtained as described in section 8.10 of Webster and Oliver (2007) and in Laurent (1963). RK uses the regression model and the variogram parameters to estimate the values of soil properties at all locations and generate maps of kriging predictions, and those of the kriging SDs for both scenarios. The kriging SD was obtained by taking the square root of the kriging variance. In addition, we subtracted the final prediction results of scenario 1 from scenario 2 using raster calculation and generated the prediction difference maps between the two scenarios. Recall that scenario 1 ignores measurement errors. In other words, it uses the model calibration and prediction approach described above but enforces that both σ_A^2 and σ_S^2 are zero.

2.5.6 | Cross-validation

Because the sampling locations were clustered in three sentinel sites within the study area, a conventional cross-validation might produce overoptimistic results (Roberts et al., 2017). The accuracy of the model predictions was therefore assessed using leave-one-out, leave-cluster-out and leave-sentinel-site-out cross-validation. Density scatter plots were used to compare the predicted values in the two scenarios at validation points. For each soil property, we derived three validation metrics: the ME, the RMSE and the MEC (Janssen & Heuberger, 1995). The MEC is equal to one minus the ratio between the residual sum of squares and the total sum of squares, as defined in Equation (11). In hydrology it is known as the Nash-Sutcliffe Model Efficiency (Nash & Sutcliffe, 1970). The MEC equals 1 in case of a perfect model, while it is 0 for a model that is as good as taking the mean of all observations as a prediction. MEC can be negative for models that are severely biased. To evaluate the kriging SD, the prediction interval coverage probability (PICP) was computed and used to derive accuracy plots for the leave-one-out cross-validation case. The section below describes how these cross-validation metrics were computed in case of uncertain validation data.

As before, let $z_T(\mathbf{s})$ be the true value of the soil property at location \mathbf{s} , and let $z_M(\mathbf{s})$ be the value of the soil

property at \mathbf{s} obtained through a measurement (using either laboratory or spectral analysis). Note that we use lower-case notation here because we now treat these as the actual values, not as random variables. Let $\hat{z}(\mathbf{s})$ be the prediction of the soil properties obtained using RK in cross-validation mode and $\sigma_K^2(\mathbf{s})$ be the associated kriging variance. In other words, $\hat{z}(\mathbf{s})$ and $\sigma_K^2(\mathbf{s})$ are derived as explained in Section 2.5.5, using that part of the measurements that were not put aside for validation. This is done for all measurement locations \mathbf{s}_i , $i = 1, \dots, n$.

Given n validation locations we derive the ME as

$$\begin{aligned} ME &= \frac{1}{n} \sum_{i=1}^n \hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i) \\ &= \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_M(\mathbf{s}_i)) + \frac{1}{n} \sum_{i=1}^n (z_M(\mathbf{s}_i) - z_T(\mathbf{s}_i)) \\ &\approx \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_M(\mathbf{s}_i)), \end{aligned} \quad (9)$$

where the latter approximation holds because we assume that the measurement method has no systematic error.

For the mean squared error (MSE) we get

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i) + z_T(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i))^2 + \frac{1}{n} \sum_{i=1}^n (z_T(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^n ((\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i)) \cdot (z_T(\mathbf{s}_i) - z_M(\mathbf{s}_i))) \\ &\approx \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i))^2 + \frac{1}{n} \sum_{i=1}^n (z_T(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2, \end{aligned} \quad (10)$$

where the latter approximation holds because the kriging prediction error and the measurement error are uncorrelated. Equation (10) shows that an estimate of the $MSE = \frac{1}{n} \sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i))^2$ is obtained by subtracting the measurement error variance (i.e., a weighted average of σ_A^2 and σ_S^2 , with weights equal to the fraction of analytical and spectral validation measurements, respectively) from the MSE computed on error-contaminated validation data. In practice, we are more interested in the RMSE than the MSE. This is derived by taking the square root after the measurement error variance has been

subtracted from the MSE computed on error-contaminated validation data.

Similarly, the MEC under uncertain validation data can be derived as

$$\begin{aligned} MEC &= 1 - \frac{\sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_T(\mathbf{s}_i))^2}{\sum_{i=1}^n (z_T(\mathbf{s}_i) - \bar{z}_T)^2} \approx 1 \\ &\quad - \frac{\sum_{i=1}^n (\hat{z}(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2 - \sum_{i=1}^n (z_M(\mathbf{s}_i) - z_T(\mathbf{s}_i))^2}{\sum_{i=1}^n (z_M(\mathbf{s}_i) - \bar{z}_M)^2 - \sum_{i=1}^n (z_T(\mathbf{s}_i) - z_M(\mathbf{s}_i))^2}, \end{aligned} \quad (11)$$

where $\bar{z}_T = \frac{1}{n} \sum_{i=1}^n z_T(\mathbf{s}_i)$ and $\bar{z}_M = \frac{1}{n} \sum_{i=1}^n z_M(\mathbf{s}_i)$ and where as before $\sum_{i=1}^n (z_T(\mathbf{s}_i) - \bar{z}_T)^2$ is derived from a weighted average of σ_A^2 and σ_S^2 .

The PICP evaluates how often the validation data are within a $(1 - \alpha)$ prediction interval for various values of α (Shrestha & Solomatine, 2006). Assuming a normal distribution for the kriging prediction error and analytical error, these prediction intervals can be derived from the variances of both errors. Since:

$$Z_M(\mathbf{s}) = \hat{Z}(\mathbf{s}) + (Z_M(\mathbf{s}) - Z_T(\mathbf{s})) + (Z_T(\mathbf{s}) - \hat{Z}(\mathbf{s})) \quad (12)$$

and measurement and kriging errors are uncorrelated we have that $Z_M(\mathbf{s})$ should lie between $\hat{Z}(\mathbf{s}) - z_{1-\alpha/2} \cdot \sqrt{\sigma_M^2 + \sigma_K^2(\mathbf{s})}$ and $\hat{Z}(\mathbf{s}) + z_{1-\alpha/2} \cdot \sqrt{\sigma_M^2 + \sigma_K^2(\mathbf{s})}$ in $(1 - \alpha) \times 100\%$ of all cases. Here, $z_{1-\alpha/2}$ refers to the $1 - \alpha/2$ quantile of the standard normal distribution. Note that σ_M^2 equals σ_A^2 in the case of analytical measurements and equals σ_S^2 in the case of spectral measurements. A plot of PICP against α boils down to an accuracy plot which visualises the assessment of quality of the estimated prediction uncertainty (Goovaerts, 2001; Wadoux et al., 2018).

3 | RESULTS

3.1 | Mid infrared spectroscopy models

Figure 2 shows PLSR predictions against wet chemistry observations for the 48 soil samples where both types of analysis were carried out. Note that SOC data were log-transformed to logSOC prior to running the PLSR (see also Section 3.2). Accurate predictive models were obtained for soil pH (ME = 0.004, RMSE = 0.219,

MEC = 0.87, clay (ME = 0.216, RMSE = 5.47, MEC = 0.83) and logSOC (ME = 0.006, RMSE = 0.192, MEC = 0.82). Note that these metrics were computed on only 48 observations and are only approximations of the population validation metrics.

The soil analytical and spectral data put together constitute the dataset for the study. For samples analysed with both conventional and MIRS, only analytical data were retained, resulting in 48 analytical and 432 spectral observations and making a total of 480 observations. Although the accuracies of the predictive models are acceptable, spectrally soil estimated data are not as accurate as the analytical data because the PLSR prediction error adds to the uncertainty.

3.2 | Descriptive statistics

The basic statistical parameters for both datasets and the merged dataset are summarised in Table 1. Considering the total dataset, soil pH was low and varied from strong acidity to near neutral (3.54 to 6.91) with a mean of 5.23, which falls within the optimum range for the production of priority crops in the tropics, such as maize (i.e., pH between 5.5 and 6.5). Textural analysis revealed that the study area is dominated by clay-rich soils with a mean of 65.5% and values ranging from 37.9% to 100%. SOC concentrations ranged from 6.7 to 84.5 g/kg with a mean of 26.4 g/kg. The analysis of the distribution of the soil properties indicates that pH and clay approximated normality, while SOC values were positively skewed (skewness coefficient of 1.12), as shown by the histogram (Figure S2). SOC was therefore log-transformed to logSOC, which had a much more symmetric distribution. The statistical modelling hereafter was applied to pH, clay and logSOC.

3.3 | Quantification of measurement errors in analytical and spectral data

The measurement error SDs for the analytical and spectral data were obtained using the methodology described in Section 2.5.2 and are given in Table 2. The estimated analytical measurement error SD for pH was 0.083, for clay 3.33% and for logSOC 0.038. The PLSR prediction error variance was added to the analytical error variance to get a total measurement error SD for spectral data of 0.234, 6.40%, and 0.196 for pH, clay and logSOC, respectively. As expected, the PLSR prediction errors for the three soil properties had substantially larger SDs than the analytical data (Table 2).

3.4 | Model selection

The stepwise model selection procedure using BIC resulted in the selection of 9 variables for pH, 4 variables for clay and 5 variables for SOC. A brief description of the 12 covariates retained for the three models is summarised in Table 3. pH and logSOC were primarily influenced by precipitation, terrain morphology, landform classes, MODIS net productivity and land cover. The selected covariates for clay were climate variables and landform classes.

3.5 | Model calibration

The linear regression models fitted using REML showed significant correlations between soil properties and the retained covariates. The relationships were of moderate statistical strength for pH ($R^2 = 0.60$) and logSOC ($R^2 = 0.49$) and of weak statistical strength for clay content ($R^2 = 0.21$). The model residuals had a symmetric

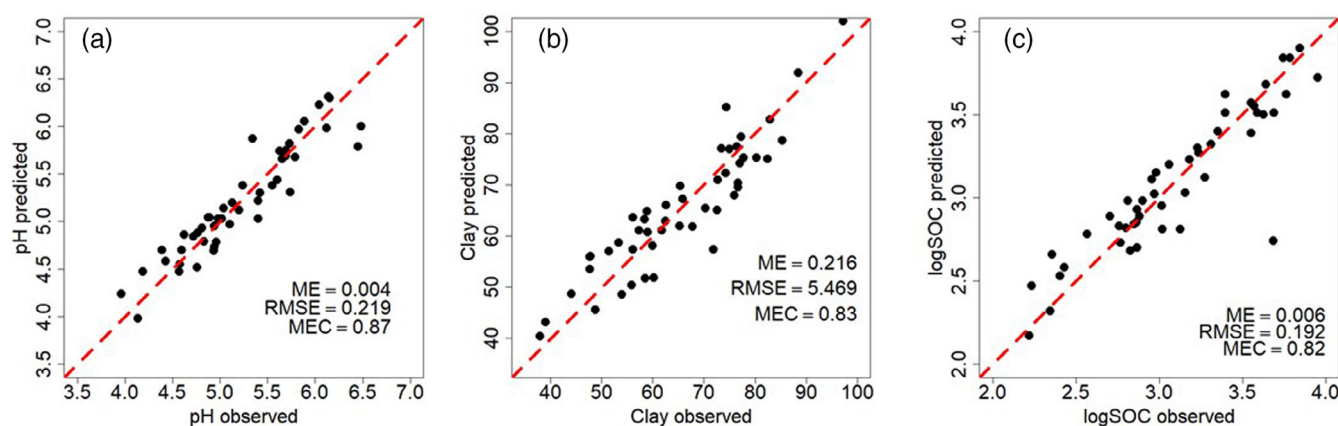


FIGURE 2 Scatter plots of PLSR predictions against observations: (a) pH, (b) clay, (c) logSOC. Red dashed lines represent the 1:1-line

Variable	Min.	Mean	Max.	SD	CV (%)	Skewness
pH						
Analytical	3.96	5.21	6.48	0.61	11.7	0.13
Spectral	3.54	5.23	6.91	0.56	10.8	0.43
Analytical + spectral	3.54	5.23	6.91	0.57	10.9	0.37
Clay (%)						
Analytical	37.9	65.4	97.3	13.3	20.3	−0.01
Spectral	41.7	65.4	100.0	10.9	16.7	0.36
Analytical + spectral	37.9	65.5	100.0	11.2	17.1	0.33
SOC (g/kg)						
Analytical	9.0	24.7	52.0	11.1	45.1	0.64
Spectral	6.7	26.8	84.4	12.9	48.1	1.15
Analytical + spectral	6.7	26.4	84.4	12.6	47.9	1.12

Abbreviation: CV, coefficient of variation.

TABLE 1 Summary statistics of soil properties for the analytical ($n = 48$), spectral ($n = 432$) and merged data sets ($n = 480$)

Soil properties	σ_A	PLSR prediction error standard deviation	σ_S
pH	0.083	0.219	0.234
Clay	3.33	5.47	6.40
LogSOC	0.038	0.192	0.196

TABLE 2 Standard deviation of analytical and spectral soil measurement errors

TABLE 3 Description of environmental variables (covariates) used in the stepwise linear regression models

	Covariate codes	Descriptions	Sources	pH	Clay	logSOC
1	CLM_CHE_PYRSUM	Total annual precipitation	CHELSEA (Karger et al., 2016)	+	+	+
2	CLM_MOD_CCYRAVG	Mean annual cloud cover	EarthEnv (Wilson & Jetz, 2016)	+	−	+
3	CLM_MOD_LSTDYRAVG	Mean annual surface temperature	MODIS (Wan, 2006)	−	+	−
4	MOR_MRG_CRU	DEM-parameters: local upslope curvature	SRTM (Rabus et al., 2003)	+	−	−
5	MOR_MRG_TPI	DEM-parameters: Topographic Position Index	SRTM (Rabus et al., 2003)	+	−	−
6	MOR_MRG_VDP	DEM-parameters: valley depth	SRTM (Rabus et al., 2003)	+	−	+
7	MOR_USG_F02	Landform class: flat plains	USGS (Sayre et al., 2014)	+	+	−
8	MOR_USG_F04	Landform class: hills	USGS (Sayre et al., 2014)	−	−	+
9	MOR_USG_F06	Landform class: low mountains	USGS (Sayre et al., 2014)	+	+	−
10	SAT_L07_B4NIR14	Band 4 (NIR) for year 2014	Landsat (Zanter, 2019)	+	−	−
11	LUC_GFC_BARLY10	30 m global land cover: bare soil	ESA (Hansen et al., 2013)	−	−	+
12	VEG_MOD_NPPY00	Net primary productivity in 2000	MODIS (Savtchenko et al., 2004).	+	−	−

Note: The plus (+) and minus (−) signs indicate whether a covariate was selected for a soil property.

distribution and were fairly normally distributed (Figure S3).

The regression coefficient estimates without (scenario 1) and with (scenario 2) accounting for measurement

errors, as well as the accompanying p -values computed using Wald tests and the relative change in coefficient estimates between scenarios are presented in Table 4. Note that the coefficients represent the mean change in

TABLE 4 Estimated regression coefficients for the environmental variables under scenarios 1 and 2

	Scenario 1 (without measurement errors)		Scenario 2 (with measurement errors)		Changes per covariate (%)
Covariates	Estimate	p-value	Estimate	p-value	
pH					
Intercept	7.27	5.99E-09	7.6	6.74E-09	−4.5
CLM_CHE_PYRSUM	−8.172E-04	5.45E-06	−8.244E-04	5.82E-06	−0.9
CLM_MOD_CCYRAVG	−1.446E-04	1.10E-01	−1.804E-04	1.13E-01	−24.8
MOR_MRG_CRU	8.370E-05	4.04E-01	5.917E-05	4.06E-01	29.3
MOR_MRG_TPI	−3.759E-04	4.78E-03	−3.517E-04	4.81E-03	6.4
MOR_MRG_VDP	−1.132E-04	2.67E−10	−1.102E-04	2.93E−10	2.7
MOR_USG_F02	−1.567E-03	1.41E-02	−1.936E-03	1.43E-02	−23.5
MOR_USG_F06	1.289E-03	3.11E-02	1.413E-03	3.13E-02	−9.6
SAT_L07_B4NIR00	1.845E-02	7.21E-04	1.867E-02	7.13E-04	−1.2
VEG_MOD_NPPY00	−4.133E-05	1.20E-01	−4.664E-05	1.20E-01	−12.9
Clay					
Intercept	653.5	3.64E-12	694.9	3.70E-12	−6.3
CLM_CHE_PYRSUM	−8.691E-03	3.22E-01	−8.585E-03	3.21E-01	1.2
CLM_MOD_LSTDYRAVG	−1.918E-01	6.04E-12	−2.058E-01	6.15E-12	−7.3
MOR_USG_F02	−5.544E-02	1.30E-02	−5.592E-02	1.31E-02	−0.9
MOR_USG_F06	−2.668E-02	1.30E-01	−2.758E-02	1.32E-01	−3.4
logSOC					
Intercept	3.547	6.57E-03	3.568	6.71E-03	−0.6
CLM_CHE_PYRSUM	−5.592E-04	8.12E-04	−5.637E-04	8.38E-04	−0.8
CLM_MOD_CCYRAVG	1.361E-04	2.09E-01	1.344E-04	2.12E-01	1.2
LUC_GFC_BARLY10	−4.213E-02	5.32E-02	−4.316E-02	5.33E-02	−2.5
MOR_MRG_VDP	−8.827E-05	5.84E-09	−8.804E-05	6.71E-09	0.3
MOR_USG_F04	1.340E-04	6.37E-01	1.384E-04	6.39E-01	−3.3

the dependent variable for one unit of change in the covariate while holding other covariates in the model constant.

Precipitation, cloud cover, valley depths and net primary productivity had negative regression coefficients for pH, indicating that areas with high rainfall and rich in biomass tend to have lower pH. This is typical for soils of humid climates which are commonly acidic. Increase in precipitation contributes in leaching many of the alkaline basic cations from the topsoil, leading to soil acidification (Chytrý et al., 2007). Clay was also negatively influenced by precipitation and tends to be lower in hills and low mountains. Cloud cover positively influenced logSOC, attesting high values of SOC in areas with high vegetation cover rate. Since SOC is related to organic matter content, areas rich in biomass, humus and associated organisms responsible for biological activities tend to have higher SOC values (Lei et al., 2019; Takoutsing et al., 2018).

The inclusion of measurement errors modified the regression coefficient estimates. Particularly for pH, coefficients changes were of up to 29% for some of the covariates. This may be because pH has the largest number of covariates and is more sensitive to incorporation of measurement error SDs, because of collinearity effects (Figure S4).

The variogram parameters estimated using REML in both scenarios are presented in Table 5. The fitted variograms are shown in Figure 3. While the variograms in both scenarios exhibited similar structures and patterns, the nuggets and sills are much smaller in scenario 2. This is because the nugget represents only spatial variation at short distances and does not include measurement error variance (Chilès & Delfiner, 2012). Part of the observed variation in soil properties is therefore explained by measurement errors, meaning that the spatial variation of the true

TABLE 5 Parameters of exponential variogram models of pH, clay and logSOC fitted using REML for scenarios 1 and 2

Model parameters	pH		Clay		logSOC	
	Scenario 1	Scenario 2	Scenario 1	Scenario 2	Scenario 1	Scenario 2
Nugget (C_0)	0.077	0.037	51.98	18.34	0.067	0.039
Partial sill (C)	0.08	0.08	61.78	59.93	0.07	0.07
Total sill ($C_0 + C$)	0.157	0.117	113.76	78.27	0.137	0.109
$C_0/(C_0 + C)$ (%)	49.04	31.62	45.69	23.43	48.91	35.78
Range parameter (m)	3000	3000	1577	1644	3000	3000

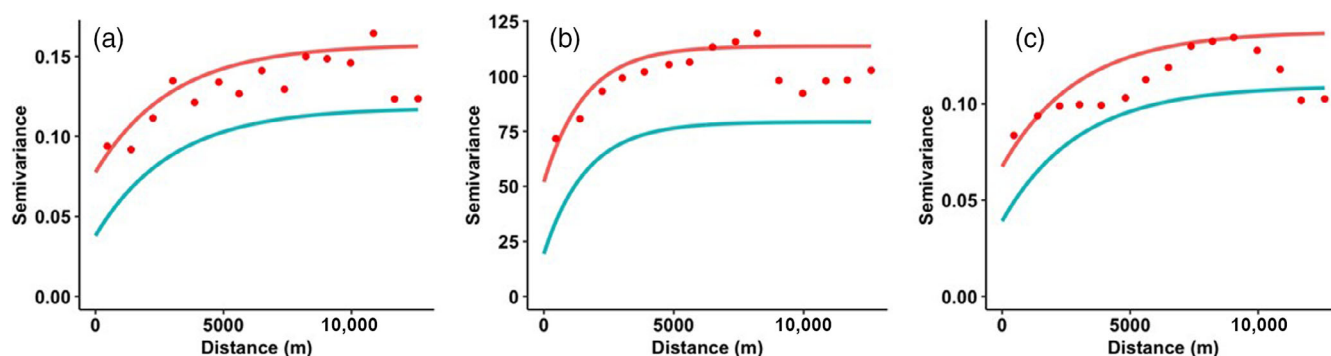


FIGURE 3 Residual variograms for (a) pH, (b) clay and (c) logSOC. Red lines represent scenario 1, where measurement errors are not explicitly modelled. Blue lines represent scenario 2, where measurement uncertainty is accounted for. Red dots are the sample variogram values

(error-free) soil properties is lower than that of the observations.

All residual variograms indicated presence of spatial structure, which means that there is added value in residual kriging. The range parameters for clay slightly increased from 1.57 to 1.64 km from scenario 1 to scenario 2, while those of pH (3 km) and logSOC (3 km) remained constant and were not affected by the incorporation of measurement errors. Note that the effective variogram range, which is about three times the range parameter for the exponential model (Webster & Oliver, 2007, section 5.2.2), is fairly small compared to the extent of the study area for all three soil properties. In other words, the residual spatial structure is somewhat limited and residual kriging will only improve prediction in the neighbourhood of measurement locations.

3.6 | Model validation

Three validation methods were used in this study: leave-one-out (LOO), leave-cluster-out (LCO) and leave-site-out (LSO) cross-validation. Validation metrics computed for both scenarios show that the models in general provided good (acceptable) predictive ability except for clay (Table 6). ME values were close to zero for the three soil

properties. Among the soil properties, clay content was poorly modelled with the lowest MEC value, as also revealed by the lowest coefficient of determination of the regression models (see Section 3.5).

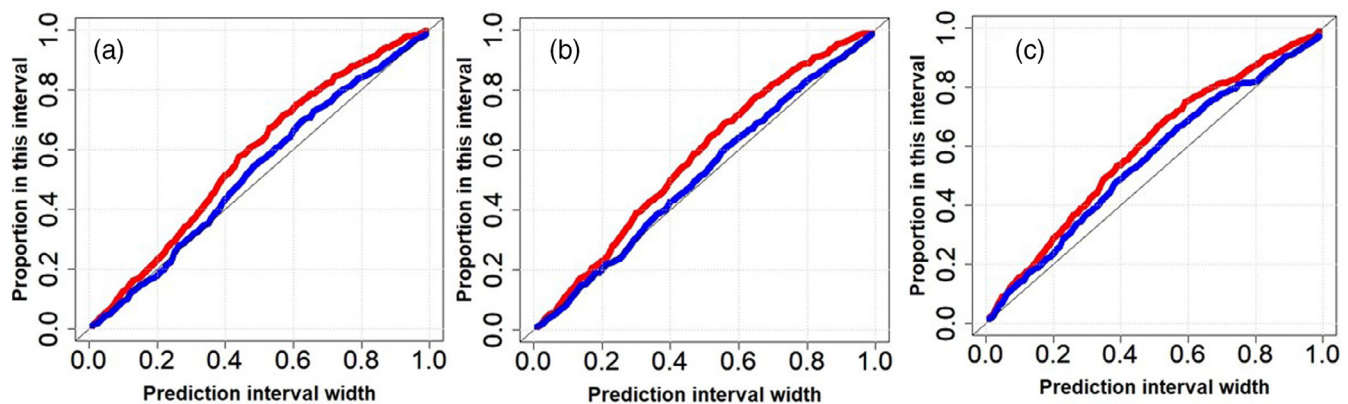
RK was able to explain the spatial variation between 68 and 83% for pH, between 18% and 62% for clay, and between 53% and 72% for logSOC. RMSE values increased while MEC values decreased from LOO to LCO to LSO cross validation, especially for clay. The decrease in model performances from LOO to LCO to LSO cross validation is as expected and due to the decrease in neighbouring values used when making predictions in each case. For LSOCV, this effectively led to spatial extrapolation rather than spatial interpolation, which is more challenging and susceptible to larger prediction errors. In all cases, validation metrics were practically the same between scenarios, attesting no significant change with the incorporation of measurement uncertainty.

As revealed by the PICP plots shown in Figure 4, which were based on LOO cross-validation, the curves deviate from the 1:1 line and show that both scenarios tend to overestimate the prediction interval widths. However, the deviation from the 1:1 line is much larger for scenario 1 than for scenario 2. For example, for pH, we find that for scenario 1, 64% of the observations is included in the 50% prediction interval, while it is only

TABLE 6 Statistical validation metrics obtained by leave-one-out, leave-cluster-out and leave-site-out cross-validation

Cross-validation methods	Scenario 1 (without measurement errors)			Scenario 2 (with measurement errors)		
	ME	RMSE	MEC	ME	RMSE	MEC
pH						
LOO	−0.001	0.213	0.834	−0.002	0.214	0.832
LCO	0.012	0.428	0.695	0.010	0.428	0.696
LSO	−0.029	0.435	0.675	−0.031	0.434	0.678
Clay						
LOO	0.020	5.73	0.621	0.021	5.73	0.621
LCO	0.177	11.69	0.297	0.160	11.68	0.300
LSO	−0.495	12.11	0.182	−0.458	12.11	0.184
logSOC						
LOO	0.000	0.217	0.729	0.001	0.218	0.729
LCO	−0.004	0.375	0.594	−0.003	0.374	0.594
LSO	0.015	0.388	0.536	0.018	0.387	0.537

Abbreviations: LCO, leave cluster out cross validation; LOO, leave one out cross validation; LSO, leave site out cross validation.

**FIGURE 4** Accuracy plots for (a) pH, (b) clay and (c) logSOC obtained using leave-one out cross validation. Red line represents scenario 1, blue line scenario 2

57% for scenario 2. This indicates that scenario 2 has a more realistic assessment of prediction uncertainty than scenario 1. For clay, scenario 2 has a negligible deviation from the 1:1 line as compared to other soil properties.

Overall, both scenarios are similar in their predictive performance (Table 6), but the prediction uncertainty is more realistically quantified in scenario 2 than in scenario 1. The best modelling approach would therefore be the one that accounts for measurement errors in soil observations.

3.7 | Spatial prediction

The fitted parameters of the regression models (covariates and regression coefficients) and the variograms (nugget, sill and range) were used by RK to predict the values of the soil properties at all locations. The logSOC

predictions and prediction error SDs were back-transformed to SOC values following Laurent (1963). The differences between the predicted values in both scenarios assessed by the scatter density plots (Figure 5) showed no systematic differences between predictions. The absolute differences were never bigger than 0.1, 1.6% and 2 g/kg for pH, clay and SOC respectively.

The maps of the predicted values for the three soil properties at 250 m spatial resolution, as well as maps of the kriging SDs for scenarios 1 and 2 are presented in Figures 6 and 7, respectively. Generally, there are similarities in the spatial distribution of predicted values as the maps showed comparable ranges of predicted values, and similar spatial patterns and features such as areas of low and high concentrations. Therefore, only the maps of scenario 2 and the maps differences between scenarios are presented.

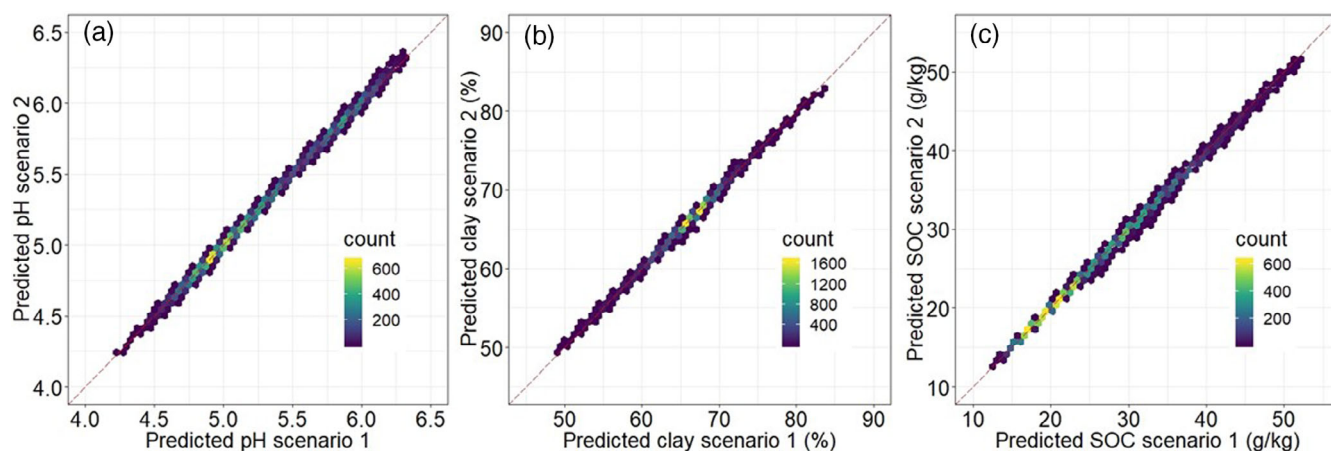


FIGURE 5 Scatter density plots of predicted values for scenario 1 and scenario 2 for (a) pH, (b) clay, and (c) SOC. The dashed red line is the 1:1 line

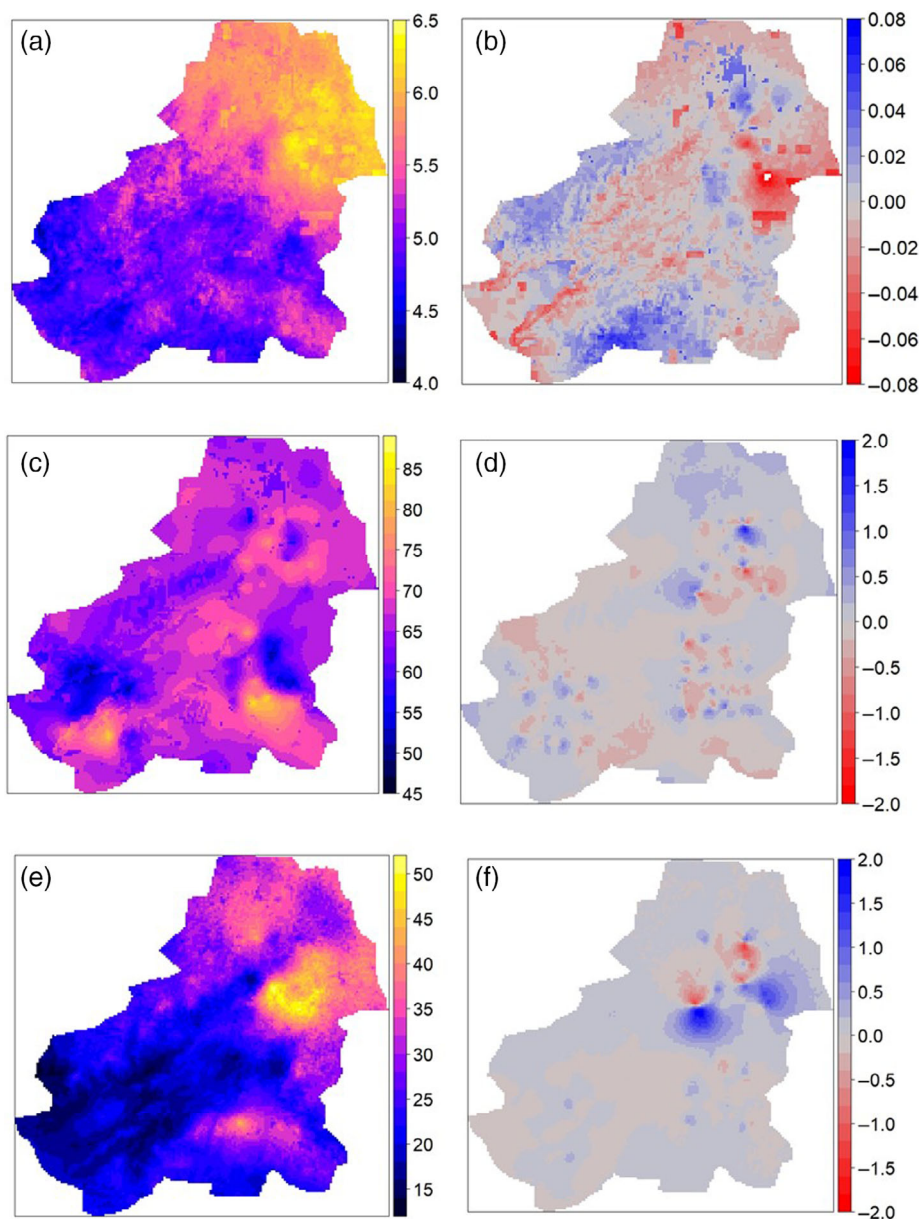


FIGURE 6 Maps of soil property predictions and prediction differences: (a) pH scenario 2, (b) difference between pH scenario 1 and pH scenario 2, (c) clay scenario 2 (%), (d) difference between clay scenario 1 and clay scenario 2 (%), (e) SOC scenario 2 (g/kg), (f) difference between SOC scenario 1 and SOC scenario 2 (g/kg). Prediction maps of scenario 1 not shown because these are very similar to those of scenario 2

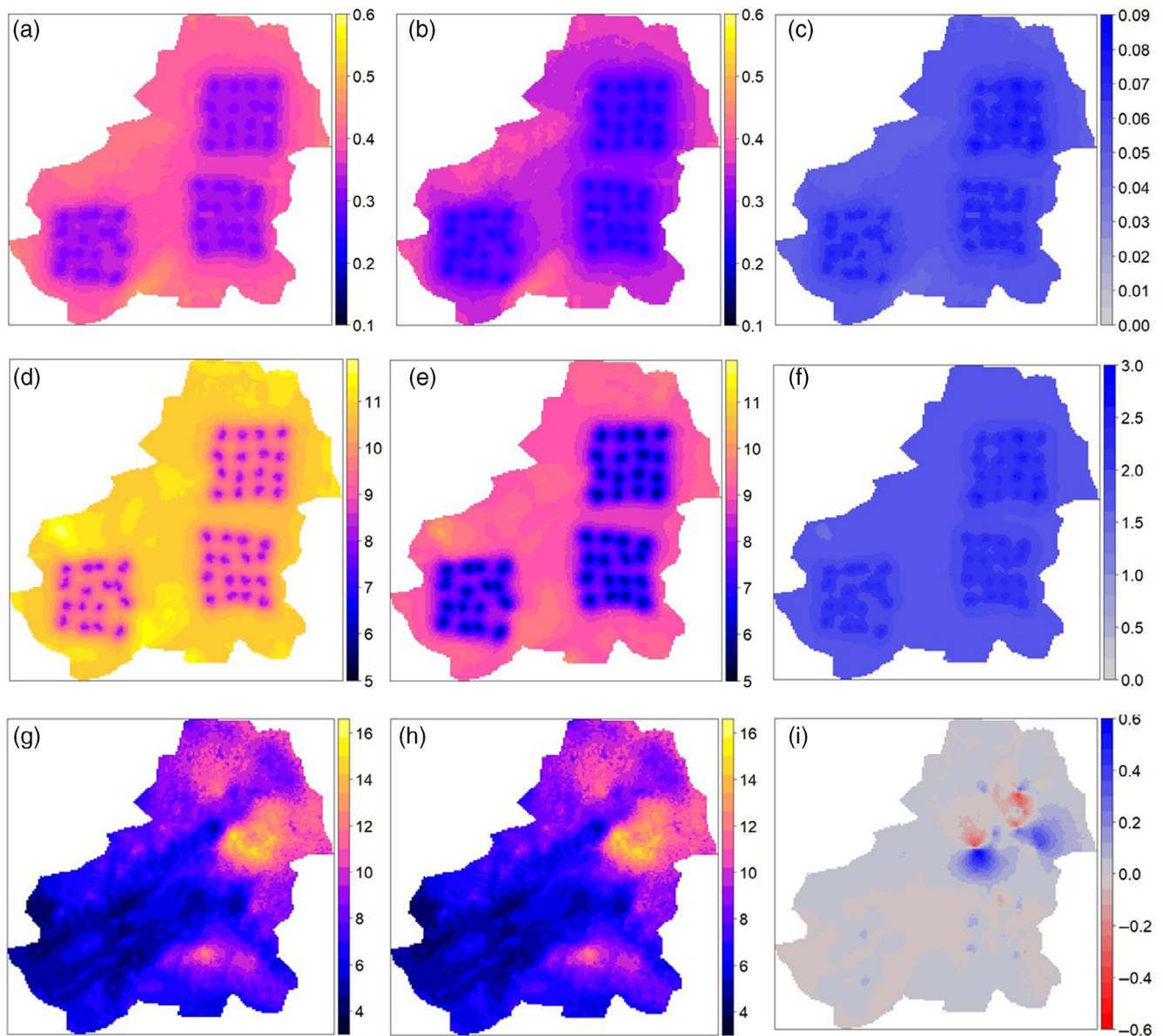


FIGURE 7 Maps of kriging SDs and differences: (a) pH scenario 1, (b) pH scenario 2, (c) difference between pH scenario 1 and pH scenario 2, (d) clay scenario 1 (%), (e) clay scenario 2 (%), (f) difference between clay scenario 1 and clay scenario 2 (%), (g) SOC scenario 1 (g/kg), (h) SOC scenario 2 (g/kg), (i) difference between SOC scenario 1 and SOC scenario 2 (g/kg)

The kriging SD maps for pH and clay clearly showed the spatial sampling design that was used, which was expected from the fairly small effective variogram ranges, which implies that the kriging SD is small only close to measurement locations (Figure 7). For SOC, the spatial sampling design does not appear in the SD map because after back-transformation this map is more influenced by the logSOC prediction than the logSOC SD.

We observed relative differences in SDs in some areas of the study area between the two scenarios of up to 0.08 for pH, 2.7% for clay and 0.5 g/kg for SOC between the two scenarios (Figure 4). The pH and clay SD maps for scenario 2 had lower values than those for scenario 1 for

the entire study area, and this corroborates well with the results obtained using the accuracy plots (Figure 4). For SOC the SD differences between scenarios 1 and 2 are both positive and negative.

4 | DISCUSSION

4.1 | Measurement errors and implications for spatial modelling

The primary aim of this study was to quantify the uncertainty in soil measurements, incorporate measurement

error variances in the covariance structure of spatial models and evaluate the influence of this on the prediction and prediction uncertainties. We illustrated the methodology and applied it to a case study, where we mapped three soil properties using soil data obtained using conventional laboratory analytical methods (analytical data) and MIRS (spectral data).

Among the three soil properties, the clay content model had poor performance compared with pH and SOC, with the lowest MEC value. This could be attributed to one or a combination of the following: (a) the spatial resolution of some of the environmental variables was not detailed enough to capture the variation (Maleki et al., 2020); (b) the set of covariates retained was not suitable and other environmental variables need to be included; or (c) the sampling protocol was too clustered to capture variability across the study area, since observations were taken from three main clusters and various processes operate at different spatial scales (Hendriks et al., 2021).

Results showed that taking measurement uncertainty into account had a small to moderate effect on the estimated regression coefficients of the RK model (Table 4) and a large effect on the residual variograms (Figure 3), which had much smaller nugget variances and sills when measurement errors were explicitly accounted for. This is as expected because measurement error variance was separately modelled in scenario 2 and hence not included in the residual variogram.

Depending on the soil property and the type of cross-validation used (leave-one-out, leave-cluster-out and leave-site-out), the amount of variance explained varied between 18 and 83% and showed that resulting maps are useful in assessing the spatial variation in pH and SOC and provided a first approximation for clay. Large differences were observed between the three types of cross-validation applied, particularly for clay, with $LOO > LCO > LSO$ cross-validation. This is largely due to the decrease in the number of nearby available observations as we move from one type of cross-validation to another (Lagacherie et al., 2020; Loiseau et al., 2021). The reduction in the number of observations may have also weakened the relationships between environmental variables and soil properties in LSO cross-validation. LOO cross-validation likely gives a too optimistic view of model performance, especially when the data are spatially clustered as in our case study, while LSO cross-validation is likely too pessimistic, because it applies spatial extrapolation instead of interpolation. LCO cross-validation may be the best compromise for evaluating prediction performance in this study. But there is a need to investigate how best to carry out cross-validation in the case of clustered data, to get a realistic estimate of

model performance that does not lead to biased estimates of the validation metrics (Chartin et al., 2017; Poggio et al., 2021; Roberts et al., 2017; Styc & Lagacherie, 2021).

Spatial models with and without measurement errors were comparable in predictive performance (Table 6). The ME, RMSE and MEC values between scenarios were quite similar, and so were the kriging prediction maps (Figure 6). This was contrary to our expectations, also taking into consideration the influence of the measurement error variances on the model parameters and regression coefficients. The insignificant differences between the validation metrics and prediction maps of the two scenarios could perhaps be explained as follows. If all observations had the same measurement error variance, then the same performance would have been achieved because in such case all observations would carry equal weight and scenarios 1 and 2 are effectively the same. In this study we used data from two different sources (analytical and spectral) with very different measurement error variances. The analytical data had much smaller measurement error variances than the spectral data, and therefore get much larger weights in the estimation of regression coefficients and in kriging. However, analytical data represented only 10% of the data set, and the spatial distribution of the analytical data was similar to that of the spectral data. As shown in Figure 1, the analytical data were in the same clusters (one observation per cluster) as the spectral data. If the analytical and spectral data had been located in different parts of the study area, we likely would have obtained larger differences between the two scenarios (Meyer et al., 2018). The results that we obtained refer to just one case study, and it is worthwhile to investigate the sensitivity of the DSM models and maps to incorporation measurement errors in other studies and in other contexts.

While differences in prediction maps and cross-validation metrics of predictions did not differ much between the two scenarios, we did get substantial differences in prediction error SD maps and in the evaluation of the prediction uncertainty. As shown by the accuracy plots (Figure 4), ignoring measurement error variances led to a large deviation from the 1:1 line. The line for scenario 1 was much above it, which indicates that the kriging prediction error SDs were unrealistically high. Though we obtained deviations in both scenarios, the problem is much more pronounced for the variance models in scenario 1, when measurement errors are ignored. Prediction intervals were larger in scenario 1 than in scenario 2, attesting that the quantification of uncertainties had significantly improved and was more realistic when measurement errors were accounted for.

The kriging SDs maps for pH and clay in scenario 2 had lower values than those in scenario 1, and this

corroborates well with the findings derived from comparison of the accuracy plots. Accounting for measurement errors decreased the kriging variance because the 'true' soil properties had less spatial variation than the measured soil properties, which means that they are easier to predict, even in case of presence of measurement errors (Chilès & Delfiner, 2012). For SOC this did not occur, even though the logSOC kriging SDs were all smaller for scenario 2 than for scenario 1 (results not shown). This was because the back-transformation of logSOC not only depends on the kriging SDs of logSOC but also on the logSOC predictions (Laurent, 1963; Webster & Oliver, 2007).

In practice, the usefulness of DSM lies in its ability to quantify and map prediction uncertainties (Malone et al., 2015), and ignoring measurement errors leads to poor assessment of the accuracy of digital soil maps (Takoutsing et al., 2017). Soil scientists have made considerable efforts in quantifying prediction uncertainties in their work, much more than in other disciplines, but this has not always been as systematic as it should be (Piikki et al., 2021). Whenever uncertainties are quantified with SD maps, one of course has to make sure that these are realistic assessments of the map error. In this study, this does not happen in scenario 1, while scenario 2 does it much better. In other words, this stressed the importance of taking measurement error into account to accurately quantify the prediction uncertainties. It is unfortunate that most end-users of DSM products are only interested in the prediction maps from which soil information are extracted, often ignoring the uncertainty maps, meanwhile prediction maps with large errors could have important economic and environmental consequences for the design and implementation of land restoration initiatives (Styc & Lagacherie, 2021; Takoutsing et al., 2017).

The results of this study overall indicate that the additional investment in quantification of measurement error variances and the incorporation in the spatial models is worth the effort, as shown by the improvement in the quantification of the prediction uncertainties. There is a need to create awareness among end-users on the importance of realistic and reliable uncertainties of the maps they intend to use, so that in case of large uncertainty, investments can be made to obtain more accurate soil information (Heuvelink, 2014).

4.2 | Limitations and recommendations for future research

Despite the successful incorporation of measurement error variances in RK and the improvement in the

quantification of prediction uncertainties, there are several aspects worthy of attention and further development.

The soil sampling design used was not initially intended for geostatistical mapping, but rather to provide a biophysical baseline, and a monitoring and evaluation framework for assessing processes of land degradation and the effectiveness of rehabilitation measures (Vågen & Winowiecki, 2013). Consequently, the design was not optimised to properly account for spatial dependence over large distances (Brus et al., 2011). In Africa, cluster sampling is often favoured due to accessibility problems and limited resources; however, the method is prone to biases and large prediction errors in unsampled areas. For future sampling, it is highly recommended to combine cluster sampling with other sampling methods to account for variation at both short and larger distances. Moreover, sampling should also cover the feature space well (Brus & Heuvelink, 2007; Wadoux et al., 2019).

One important motive that measurement errors variances have not been systematically incorporated in DSM is the challenge in their quantification in laboratories. These are rarely reported systematically with the results of analyses, probably due to the lack of interests from clients (Li et al., 2019). In this study, we had one sample analysed in duplicate under the same conditions to quantify the measurement error variances of analytical data. We assumed constant measurement error variances for each of the soil properties, but in many practical cases measurement errors are proportional to the measured values (Libohova et al., 2019). Since measurement error variance for SOC was estimated on logarithmic scale, we assumed a proportional error model for SOC, but not for pH and clay. Quantification of measurement errors in analytical data can be improved if laboratories pay more attention and systematically quantify the uncertainties of their measurements and benchmark against standards to minimise systematic bias.

There are many sources of errors that propagate during the DSM process, and each contributes to uncertainty in the final prediction (Robinson et al., 2015). The fact that this study focused on the uncertainty in soil observations does not mean that the influence of other sources of errors can be ignored. From the modellers and end-users perspectives, possible improvements would be to quantify these other sources of errors and assess their implications so that measures are taken to improve the uncertainties of the DSM. Efforts have already been made in this line and error quantification methods have already been broadly discussed for some of the sources in Nelson et al. (2011); Bishop et al. (2015); Robinson et al. (2015).

This study used resampled environmental variables at 250 m resolutions as covariates, which might be too coarse and not able to capture some of the variation

across the study area (Taylor et al., 2013). The use of high spatial resolution covariates that can provide more detailed information and reflect the distribution characteristics of the targeted soil properties, particularly for small study area is recommended (Samuel-Rosa et al., 2015).

This study has shown that it is relatively easy to incorporate measurement error variances in RK once these are quantified. Further development of the approach is the extension to machine learning DSM models (Hengl, Nussbaum, et al., 2018; Wadoux, 2019). This is critical because of the rapid uptake for the ML algorithms in DSM that is transforming the process of spatial modelling and generating more accurate predictions (Wadoux et al., 2020).

5 | CONCLUSION

We applied a geostatistical DSM approach to derive prediction and prediction uncertainty maps after quantifying and incorporating measurement error variances in the covariance structure of the spatial model. Accounting for measurement errors resulted in changes in regression coefficients of up to 29% and influenced the variogram parameters by reducing the nuggets and sill variances. Validation metrics were quite similar in the two scenarios, but prediction uncertainties were more realistically quantified by the models that account for measurement errors, as indicated by accuracy plots. Prediction maps were similar between scenarios, but we observed slight differences in predicted values in some parts of the study area of up to 0.1 for pH, 1.6% for clay and 2 g/kg for SOC. Differences in regression kriging SDs were up to 0.08 for pH, 2.7% for clay and 0.5 g/kg for SOC. For pH and clay the kriging SDs were systematically smaller when measurement errors were explicitly accounted for.

The study stressed the importance of quantifying prediction uncertainties, particularly when the issue of uncertainty propagation in the modelling processes becomes essential. This will help end-users to be aware of the real prediction uncertainties and their implications for the design and implementation of land restoration interventions. It is advised that the methodology used in this work is also tested in other case studies and further developments of the approach should include its extension to non-linear ML regression methods, such as Random Forest.

ACKNOWLEDGEMENTS

The authors highly appreciate the support of the staff of the ICRAF Soil–Plant Spectral Diagnostic Laboratory, particularly Andrew Sila for providing support for the predictions of spectral data. Funding support for this

study was provided by the CGIAR Research Program on Water, Land and Ecosystems (WLE). We are grateful to the *International Soil Reference and Information Centre* (ISRIC) for providing the environmental variables used as covariates in the study. We sincerely thank the two anonymous reviewers for their critical and constructive evaluation of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

AUTHOR CONTRIBUTIONS

Research concept and design: Bertin Takoutsing and Gerard B. M. Heuvelink. *Drafting of manuscript:* Bertin Takoutsing and Gerard B. M. Heuvelink. *Critical revision of the manuscript:* Jetse J. Stoorvogel, Keith D. Shepherd and Ermias Aynekulu. *Revision, approval and agreement to the published version of the manuscript for submission:* All authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Bertin Takoutsing  <https://orcid.org/0000-0002-2822-6453>

Gerard B. M. Heuvelink  <https://orcid.org/0000-0003-0959-9358>

Jetse J. Stoorvogel  <https://orcid.org/0000-0003-4297-122X>

Keith D. Shepherd  <https://orcid.org/0000-0001-7144-3915>

Ermias Aynekulu  <https://orcid.org/0000-0002-1955-6995>

REFERENCES

- Arrouays, D., Lagacherie, P., & Hartemink, A. E. (2017). Digital soil mapping across the globe. *Geoderma Regional*, 9, 1–4.
- Bishop, T. F. A., Horta, A., & Karunaratne, S. B. (2015). Validation of digital soil maps at different spatial supports. *Geoderma*, 241–242, 238–249.
- Brus, D. J., & Heuvelink, G. B. M. (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138, 86–95.
- Brus, D. J., Kempen, B., & Heuvelink, G. B. M. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62, 394–407.
- Chartin, C., Stevens, A., Goidts, E., Krüger, I., Carnol, M., & van Wesemael, B. (2017). Mapping soil organic carbon stocks and estimating uncertainties at the regional scale following a legacy

- sampling strategy (southern Belgium, Wallonia). *Geoderma Regional*, 9, 73–86.
- Chilès, J. P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. John Wiley & Sons, Inc.
- Chytrý, M., Danihelka, J., Ermakov, N., Hájek, M., Hájková, P., Kočí, M., Kubešová, S., Lustyk, P., Otýpková, Z., Popov, D., Roleček, J., Řezníčková, M., Šmarda, P., & Valachovič, M. (2007). Plant species richness in continental southern Siberia: Effects of pH and climate in the context of the species pool hypothesis. *Global Ecology and Biogeography*, 16, 668–678.
- Gao, X., & Song, P. X. K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association*, 105, 1531–1540.
- Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, 103, 3–26.
- Hanchuan, P., Fuhui, L., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 1226–1238.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342, 850–853.
- Hendriks, C. M. J., Stoorvogel, J. J., Álvarez-Martínez, J. M., Claessens, L., Pérez-Silos, I., & Barquín, J. (2021). Introducing a mechanistic model in digital soil mapping to predict soil organic matter stocks in the Cantabrian region (Spain). *European Journal of Soil Science*, 72, 704–719.
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Mendes de Jesus, J., Tamene, L., & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS One*, 10, e0125814.
- 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.
- Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., & Prentice, I. C. (2018). Global mapping of potential natural vegetation: An assessment of machine learning algorithms for estimating land potential. *PeerJ*, 6, e5457.
- Heuvelink, G.B.M. 2014. Uncertainty quantification of GlobalSoilMap products. In: *GlobalSoilMap. Basis of the global spatial soil information system*, D. Arrouays, N. McKenzie, J. Hempel, A. Richer de Forges & McBratney, A. Taylor & Francis, CRC Press, London. pp. 335–340.
- Heuvelink, G. B. M. (2018). Uncertainty and uncertainty propagation in soil mapping and modelling. In A. B. McBratney, B. Minasny, & U. Stockmann (Eds.), *Pedometrics* (pp. 439–461). Springer International Publishing.
- Janssen, P. H. M., & Heuberger, P. S. C. (1995). Calibration of process-oriented models. *Ecological Modelling*, 83, 55–66.
- Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., Zimmermann, N., Linder, H., & Kessler, M. (2016). CHELSA climatologies at high resolution for the earth's land surface areas (Version 1.1). *World Data Center for Climate*, 2016.
- Knotters, M., Brus, D. J., & Oude Voshaar, J. H. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67, 227–246.
- Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., & Nkuba-Kasanda, L. (2020). Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. *Geoderma*, 375, 114503.
- Lark, R. M., & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, 55, 799–813.
- Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, 57, 787–799.
- Laurent, A. G. (1963). The lognormal distribution and the translation method: Description and estimation problems. *Journal of the American Statistical Association*, 58, 231–235.
- Lei, Z., Yu, D., Zhou, F., Zhang, Y., Yu, D., Zhou, Y., & Han, Y. (2019). Changes in soil organic carbon and its influencing factors in the growth of *Pinus sylvestris* var. *mongolica* plantation in Horqin Sandy Land, Northeast China. *Scientific Reports*, 9, 16453.
- Li, C., Ashlock, J. C., & Wang, X. (2019). Quantifying repeatability reproducibility sources of error and capacity of a measurement: Demonstrated using laboratory soil plasticity tests. *Advances in Civil Engineering*, 2019, 4539549.
- Libohova, Z., Seybold, C., Adhikari, K., Wills, S., Beaudette, D., Peaslee, S., Lindbo, D., & Owens, P. R. (2019). The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers and practitioners. *European Journal of Soil Science*, 70, 185–199.
- Loiseau, T., Arrouays, D., Richer-de-Forges, A. C., Lagacherie, P., Ducommun, C., & Minasny, B. (2021). Density of soil observations in digital soil mapping: A study in the Mayenne region, France. *Geoderma Regional*, 24, e00358.
- Maleki, S., Khormali, F., Mohammadi, J., Bogaert, P., & Bagheri Bodaghabadi, M. (2020). Effect of the accuracy of topographic data on improving digital soil mapping predictions with limited soil data: An application to the Iranian loess plateau. *Catena*, 195, 104810.
- Malone, B. P., Kidd, D. B., Minasny, B., & McBratney, A. B. (2015). Taking account of uncertainties in digital land suitability assessment. *PeerJ*, 3, e1366.
- McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117, 3–52.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, 101, 1–9.
- Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, 301–311.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10, 282–290.

- Nelson, M. A., Bishop, T. F. A., Triantafyllis, J., & Odeh, I. O. A. (2011). An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science*, 62, 417–430.
- Piikki, K., Wetterlind, J., Söderström, M., & Stenberg, B. (2021). Perspectives on validation in digital soil mapping of continuous attributes—A review. *Soil Use and Management*, 37, 7–21.
- 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. *The Soil*, 7, 217–240.
- Poggio, L., Gimona, A., Spezia, L., & Brewer, M. J. (2016). Example of Bayesian uncertainty for digital soil mapping. In G.-L. Zhang, D. Brus, F. Liu, X.-D. Song, & P. Lagacherie (Eds.), *Digital soil mapping across paradigms, scales and boundaries* (pp. 181–193). Springer Singapore.
- R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission—A new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57, 241–262.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Robinson, N. J., Benke, K. K., & Norng, S. (2015). Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Research*, 53, 592–604.
- Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M., & Anjos, L. H. C. (2015). Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma*, 243–244, 214–227.
- Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., & Nickless, D. (2004). Terra and AquaMODIS products available from NASA GES DAAC. *Advances in Space Research*, 34, 710–714.
- Sayre, R., Dangermond, J., Frye, C., Vaughan, R., Aniello, P., Breyer, S. P., Cribbs, D., Hopkins, D., Nauman, R., Derrenbacher, W., Wright, D. J., Brown, C., Convis, C., Smith, J. H., Benson, L., VanSistine, D. P., Warner, H., Cress, J. J., Danielson, J. J., ... Comer, P. (2014). *A new map of global ecological land units—An ecophysiological stratification approach*. Association of American Geographers.
- Shepherd, K. D., Shepherd, G., & Walsh, M. G. (2015). Land health surveillance and response: A framework for evidence-informed land management. *Agricultural Systems*, 132, 93–106.
- Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19, 225–235.
- Sila, A. M., Shepherd, K. D., & Pokhariyal, G. P. (2016). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems*, 153, 92–105.
- Somarathna, P. D. S. N., Minasny, B., Malone, B. P., Stockmann, U., & McBratney, A. B. (2018). Accounting for the measurement error of spectroscopically inferred soil carbon data for improved precision of spatial predictions. *Science of the Total Environment*, 631–632, 377–389.
- Stoorvogel, J. J., Kooistra, L., & Bouma, J. (2015). Managing soil variability at different spatial scales as a basis for precision agriculture. In R. Lal & B. A. Stewart (Eds.), *Soil-specific farming: Precision agriculture* (pp. 37–72). London: CRC Press. <https://doi.org/10.1201/b18759-6>
- Styc, Q., & Lagacherie, P. (2021). Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. *Geoderma*, 391, 114968.
- Takoutsing, B., Martín, J. A. R., Weber, J. C., Shepherd, K., Sila, A., & Tondoh, J. (2017). Landscape approach to assess key soil functional properties in the highlands of Cameroon: Repercussions of spatial relationships for land management interventions. *Journal of Geochemical Exploration*, 178, 35–44.
- Takoutsing, B., Weber, J., Aynekulu, E., Rodríguez Martín, J. A., Shepherd, K., Sila, A., Tchoundjeu, Z., & Diby, L. (2016). Assessment of soil health indicators for sustainable production of maize in smallholder farming systems in the highlands of Cameroon. *Geoderma*, 276, 64–73.
- Takoutsing, B., Weber, J. C., Rodríguez Martín, J. A., Shepherd, K., Aynekulu, E., & Sila, A. (2018). An assessment of the variation of soil properties with landscape attributes in the highlands of Cameroon. *Land Degradation & Development*, 29, 2496–2505.
- Taylor, J. A., Jacob, F., Galleguillos, M., Prévot, L., Guix, N., & Lagacherie, P. (2013). The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). *Geoderma*, 193–194, 83–93.
- Terhoeven-Urselmans, T., Vagen, T. G., Spaargaren, O., & Shepherd, K. D. (2010). Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal*, 74, 1792–1799.
- Vågen, T.-G., & Winowiecki, A. (2020). *The land degradation surveillance framework field guide*. ICRAF.
- Vågen, T.-G., & Winowiecki, L. A. (2013). Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential. *Environmental Research Letters*, 8, 015011.
- Vaysse, K., & Lagacherie, P. (2017). Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma*, 291, 55–64.
- Veronesi, F., & Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of top-soil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators*, 101, 1032–1044.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., ... Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230.
- Viscarra Rossel, R. A., & Brus, D. J. (2018). The cost-efficiency and reliability of two methods for soil organic C accounting. *Land Degradation & Development*, 29, 506–520.
- Viscarra Rossel, R. A., Brus, D. J., Lobsey, C., Shi, Z., & McLachlan, G. (2016). Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. *Geoderma*, 265, 152–163.

- Wadoux, A., Brus, D. J., & Heuvelink, G. B. M. (2018). Accounting for non-stationary variance in geostatistical mapping of soil properties. *Geoderma*, 324, 138–147.
- Wadoux, A. M. J. C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, 351, 59–70.
- Wadoux, A. M. J. C., Brus, D. J., & Heuvelink, G. B. M. (2019). Sampling design optimization for soil mapping with random forest. *Geoderma*, 355, 113913.
- Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359.
- Wan, Z. (2006). *MODIS land surface temperature products users' guide*. Santa Barbara, CA: Institute for Computational Earth System Science, University of California.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). John Wiley & Sons, Inc.
- Wilson, A. M., & Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biology*, 14, e1002415.

Zanter, K. (2019). *Landsat 4-7 Surface Reflectance Code (LEDAPS) Product Guide, landsat-normalized-difference-vegetation-index*. www.usgs.gov/land-resources/nli/landsat/landsat-normalized-difference-vegetation-index

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Takoutsing, B., Heuvelink, G. B. M., Stoorvogel, J. J., Shepherd, K. D., & Aynekulu, E. (2022). Accounting for analytical and proximal soil sensing errors in digital soil mapping. *European Journal of Soil Science*, 73(2), e13226. <https://doi.org/10.1111/ejss.13226>