Modelling and mapping maize yields and making fertilizer recommendations with uncertain soil information

Bertin Takoutsing[1,2*], Gerard B. M. Heuvelink[1,3], Ermias Aynekulu[4], Keith D. Shepherd[5],

[1]Soil Geography and Landscape Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands
[2]Land Health Decisions, World Agroforestry (CIFOR-ICRAF), Yaoundé, Cameroon
[3]ISRIC-World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands
[4]Land Health Decisions, World Agroforestry (CIFOR-ICRAF), Nairobi, Kenya
[5]Innovative Solutions for Decision Agriculture, Harpenden, UK

[*]Corresponding author: Bertin Takoutsing, Land Health Decisions, World Agroforestry (CIFOR-ICRAF), Yaoundé, Cameroon. Email: b.takoutsing@cifor-icraf.org

Abstract

Crop models can improve our understanding of crop responses to environmental conditions and farming practices. However, uncertainties in model inputs can notably impact the quality of the outputs. This study aimed at quantifying the uncertainty in soil information and analyse how it propagates through the Quantitative Evaluation of Fertility of Tropical Soils model using to affect yield and fertilizer recommendation rates using Monte Carlo simulation. Additional objectives were to analyse the uncertainty contributions of the individual soil inputs to model output uncertainty and discuss strategies to communicate uncertainty to end-users. The results showed that the impact of soil input uncertainty on model output uncertainty was significant and varied spatially. Comparison of the results of a deterministic model run with the mean of the Monte Carlo simulation runs showed systematic differences up to 1.0 tons ha^{-1} for maize yield and up to 59, 42, and 20 kg ha^{-1} for nitrogen (N), phosphorous (P) and potassium (K) fertilizers, respectively. Stochastic sensitivity analysis showed that pH was the main source of uncertainty for K fertilizer (81.6%) and that soil organic carbon contributed most to the uncertainty of N fertilizer (97%). Uncertainty in P fertilizer mostly came from uncertainty in extractable phosphorus (55%) and exchangeable potassium (20%). A threshold probability map designed using statistical predictions served as a visual aid that could enable farmers to swiftly make informed decisions about fertilizer application locations. The study highlights the importance of refining the accuracy of soil maps, which improves QUEFTS model predictions and offers valuable insights into the relationship between soil information accuracy and reliable crop modeling for sustainable agricultural decisions.

Keywords

Uncertainty propagation; Monte Carlo simulation; stochastic sensitivity; QUEFTS model; digital soil mapping; model input uncertainty

1. Introduction

Food security is one of the major challenges faced by populations in most sub-Saharan African countries (SSA) (Mesfin *et al.*, 2021) and agriculture remains the foremost food supplier in most of these countries. Substantial increase in agricultural production has been achieved over the past decades (FAOSTAT, 2020), but largely due to expansion into new land areas rather than increases in land productivity. The current yield gaps point to opportunities to increase food production through the efficient use of fertilizers (van Ittersum *et al.*, 2016). Obviously, the imbalance between current fertilizer application rates and the nutrient requirements for crops, leads to inefficient use of fertilizers and reduction in crop yields. A major challenge has been to recommend fertilizers that account for soil spatial variability (Rurinda *et al.*, 2020) and differences in landscape attributes (Takoutsing *et al.*, 2018).

Current fertilizer formulation in SSA has been conventionally promoted through blanket recommendations often developed based on a limited number of field-experimental data (Rurinda *et al.*, 2020). These recommendations are not able to provide the required optimal application rates because they ignore the often-high spatial variability in soil nutrient supply. This is bound to create imbalanced crop nutrition in heterogenous fields (Kihara *et al.*, 2016), leading to either over- or under-fertilization in different parts of the fertilized area. Site-specific fertilizer recommendations have been proposed to account for this variation by optimizing fertilizer use based on specific soil conditions and actual soil nutrient supply (Mesfin *et al.*, 2021). Current decision-making on fertilizer application rates is based primarily on yield responses, meanwhile farmers are eager to have detailed information on the nutrient status and fertilizer requirement specific to their fields (Breure *et al.*, 2022a). Though site-specific fertilization is much preferred over blanket fertilization, it can only achieve its objective if soil conditions are adequately known and uncertainty in soil properties does not prohibit deriving fertilizer recommendation.

One way to provide fertilizer recommendations that is site-specific is the use of decision support tools such as crop models that combine soil nutrient supply and crop nutrient demand to recommend fertilizer application rates to achieve a target yield. While crop models have become common within agricultural research domains, they have traditionally been hampered by their complexity and high demand for input data that are seldomly available in SSA. In addition, some of these models are unable to predict nutrient-limited yield. The QUEFTS model can quantify the nutrient requirements of crops based on the target yield and nutrient uptake (Janssen *et al.*, 1990). The most prominent feature of QUEFTS is that it estimates soil N, P and K supply on the basis of soil data, and predicts N, P and K fertilizer rates to achieve target yields at specific locations (Smaling and Janssen, 1993). The model has been calibrated and validated for different crops in varying soils, climate and management conditions in SSA (Ezui *et al.*, 2017) and other regions (Sattari *et al.*, 2014). Therefore, it can also be used to estimate the nutrient requirements of maize in Cameroon.

Most recent developments in crop modelling have acknowledged the need to quantify model uncertainties (Wallach and Thorburn, 2017). The sources of uncertainties are associated with errors in model structure, inputs and parameters, and can overshade the spatiotemporal variability of simulated model outputs, thus limiting predictability (Ramirez-Villegas *et al.*, 2017; Chapagain *et al.*, 2022). Input uncertainty arises from uncertainty in climate (*e.g.* temperature), soil (*e.g.* soil properties), initial conditions and crop management practices, which are typical inputs required for most crop models (Ojeda *et al.*, 2021; Chapagain *et al.*, 2022). In particular, soil information embodies a substantial degree of error, because laboratory analyses are imperfect (van Leeuwen *et al.*, 2022) and most soil information is obtained from maps that suffer from prediction and interpolation errors. The effects of errors introduced due to soil sampling and chemical analysis procedures on fertilizer recommendations has been evaluated, with conclusions that large uncertainty exists in estimates of soil nutrient supply based on soil property measurements (Schut and Giller, 2020). Uncertainty in soil information can be quantified by probability distributions (Heuvelink, 2014) and several methods for uncertainty propagation analysis have been developed and applied at various scales. Monte Carlo analysis has often been used to compute output probability distributions by repeated model simulations with input variables randomly sampled from their probability distribution. Importantly, with the increased generation of spatially explicit gridded crop model simulations, failure to account for model output uncertainty may lead to poor decision-making by policy makers and stakeholders. Uncertainty in model outputs should therefore be communicated effectively so as to enable the end-users to draw valid conclusions and make sound decisions (Breure *et al.*, 2022b; Lark *et al.*, 2022).

For a case study on maize (*Zea mays L.)* in the Western Highlands of Cameroon, this study aimed to: 1) quantify the uncertainty (probability distributions) in soil information used by the QUEFTS model; 2) analyse how uncertainty in soil information propagates through QUEFTS and affects yields and fertilizer recommendation rates; 3) compare the results of the uncertainty propagation analysis with a case where uncertainty in soil inputs is ignored; 4) analyse the uncertainty contributions of the individual soil inputs to model output uncertainty; 5) discuss strategies to communicate uncertainty in QUEFTS outputs to end-users and advise them on how uncertainty can be incorporated in their decision-making process.

2. Case study

2.1 Description of the study area

The study area is the west region of Cameroon that forms part of the Western Highlands and spans over 14,000 km^2 (Fig. 1). The climate is tropical humid with two seasons: a long, wet season of eight months from March to October, and a short, dry season of four months from November to February. The average annual temperature ranges between 20 °C and 28 °C, while annual average rainfall ranges from 1,200 mm to 2,300 mm (Neba, 1999). The area is characterized by accidented relief of massifs and mountains that consist of plains, undulating hills, and gentle sloping areas. Altitudes reach as high as 2,400 and as low as below 450 masl in valleys. The dominant soil types are Ferralsols and Nitisols of the World Reference Base system (IUSS, 2015). The area is largely an agrarian area subjugated by subsistence agricultural systems where high-skill farmers exploit virtually every strip of land available to grow a range of annual and perennial crops. Maize is the major staple crop, and farmers grow it either in association or rotation with other crops.

2.2. Field sampling using the Land Degradation Surveillance Framework

The Land Degradation Surveillance Framework (LDSF) was used to collect soil samples across the study area. The LDSF is a systematic methodology to conduct landscape-level assessments of soil and land health based on a consistent set of indicators and field protocols and uses the concept of sentinel sites (Vågen and Winowiecki, 2020). A sentinel site is a 100 km^2 area stratified into 16 clusters of 1 km^2 size, each containing 10 plots of size 1000 m^2, while each plot is further subdivided into 4 subplots of size 100 m^2. Four sentinel sites were randomly placed within the study area. The distribution of the 640 soil sampling plots is shown in Fig. 1. Topsoil samples (0 – 20 cm) were collected at the four subplots of each plot and thoroughly mixed to form a composite sample for each plot, resulting in 160 soil samples per site and 640 soil samples in total.

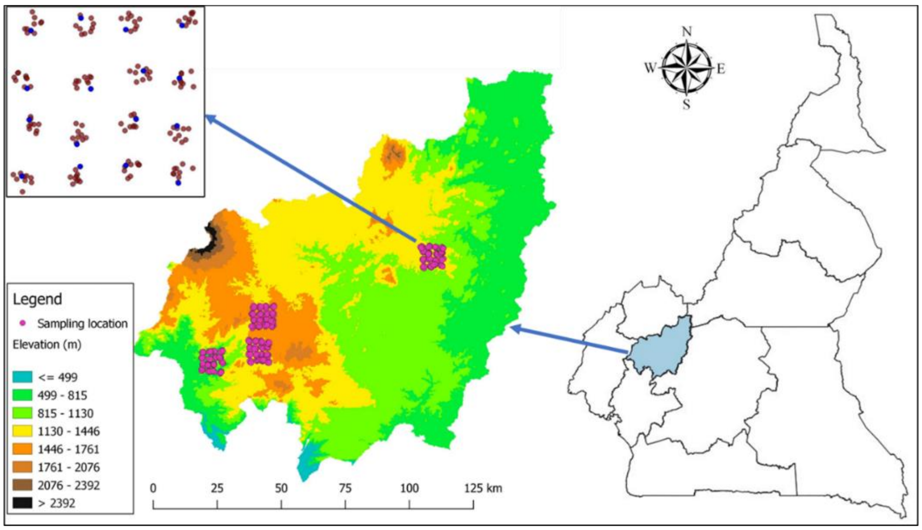

Fig. 1 Digital elevation map of the study area in Cameroon showing the soil sampling locations. Upper-left panel zooms in on one sentinel site. Red dots represent locations with spectral data, blue dots locations with analytical and spectral data

2.3 Soil data

The soil samples were air dried, crushed by rolling pins and sieved through a 2-mm sieve. About 10 g of each soil sample was milled to pass a 75-µm sieve using a Retsch RM 200 mill (Retsch, Düsseldorf, Germany) and analysed using a high-throughput Bruker Tensor 27 Fourier Transform MIR spectrometer. MIR diffuse reflectance spectra were recorded at a waveband range of 601 to 4000 cm^{-1} with a resolution of 4 cm^{-1}. Ten per cent of the soil samples (n = 64) were randomly selected as references and also analysed using conventional wet chemistry methods for the determination of SOC (dry combustion), pH (1:1 solution in water), macro- and micro-elements including base cations (Melich-3 extraction), and texture (Laser diffraction particle size analysis). Calibration models were developed using the paired observations of MIR spectra and analytical laboratory measurements to predict the soil properties at locations where only spectra were available. SOC, pH, P and K were used as inputs for crop modelling, while Al, Fe, and Ca were used for pedotransfer functions as explained below. Prior to applying the MIRS, Olsen Phosphorus (POlsen) and exchangeable potassium (KExch) in mmol kg^{-1} were estimated from available Mehlich 3 data using transfer functions proposed by Breure *et al.* (2022a) and shown in Equations 1 and 2:

$$ln\_POlsen = 0.77 * ln(P_{M3}) + 0.62 * ln(Al_{M3}) + 0.13 * ln(Fe_{M3}) + 0.10 *$$

$$ln(Ca_{M3}) - 0.19 * pH - 4.31 \quad (1)$$

$$KExch = 0.028 * K_{M3} + 0.015 \quad (2)$$

Note that the nutrients in Mehlich 3 (P, Al, Ca, and Fe) in mg kg^{-1} were log-transformed for application of the P transfer function (Equation 1). The predicted POlsen values were obtained by back-transformation following Lark and Lapworth (2012):

$$POlsen = exp(ln\_POlsen + 0.5 * \sigma^2_{pred}) \quad (3)$$

where ln_POlsen is the predicted value with the transfer function (Equation 1) and σ^2_{pred} denotes the associated prediction error variance, which can be obtained from the residual variance of the multiple linear regression model and variance-covariance matrix of the estimation errors of the regression coefficients.

3. Methodology

3.1 Spatial modelling of soil properties using random forest

The spatial modelling of soil properties has been described in detail in Takoutsing and Heuvelink (2022), and we only repeat the essentials here. Since soil properties are correlated with environmental variables (Minasny and McBratney, 2016), a set of 191 spatially distributed environmental variables was collected from different sources to represent the major soil forming processes and surface characteristics of the study area. Firstly, we carried out a correlation analysis to address multicollinearity of the 191 environmental layers. Only covariate layers with a pairwise correlation coefficient <= 0.75 with all the other covariates were retained for further analyses. In case two covariates were correlated above this threshold; we only retained the first one in alphabetical order for use in the modelling process. This first step reduced the number of covariates to 99 layers. Next, we performed a selection procedure using the Recursive Feature Elimination algorithm as implemented in the caret package (Kuhn, 2008) to remove the least important covariates. The Recursive Feature Elimination procedure is an iterative process that starts by fitting a model using all covariates, assesses its performance and ranks the covariates according to their importance (Gomes *et al.*, 2019). The least important covariate is removed, and the process is repeated, until only one covariate is left. From all evaluated models the one with the most favourable cross-validation statistic (RMSE or R^2) is selected. For each soil property, an optimal set of covariates was selected, which can differ between soil properties.

The fitted models were applied to the stack of retained environmental variables to predict the four soil properties required for further modelling, i.e. SOC, pH, POlsen and KExch across the study area at 250 m resolution. The performance of each predictive model was evaluated by calculating the Mean Error (ME), the Root Mean Squared Error (RMSE) and the model efficiency coefficient (MEC) (Janssen and Heuberger, 1995) using leave-cluster-out cross-validation (LCOCV). The LCOCV has previously been demonstrated to be a suitable method to evaluate the performance of prediction models in the case of clustered data (Takoutsing *et al.*, 2022). We used quantile regression forest (QRF) (Meinshausen (2006) to obtain model predictions and the associated prediction uncertainty. QRF derives quantiles of the conditional probability distribution at each prediction location. For each soil property, we computed the mean, 0.05 quantile and 0.95 quantile, i.e. the lower and upper limits of a 90 % prediction interval. The 90% prediction interval coverage probability (PICP) was used to validate the quantified prediction uncertainty (Shrestha and Solomatine, 2006). Ideally, the PICP should be closed to 0.90. A PICP value substantially greater than 0.90suggests that the uncertainty is underestimated, while a value substantially smaller than 0.90 indicates that it was overestimated.

3.2 The QUEFTS model

The QUEFTS model was originally developed by Janssen *et al.* (1990) to estimate yield responses based on nutrients present in the soil and those added through application of NPK fertilizers. Further improvements were made by Smaling and Janssen (1993) to estimate the nutrient requirements for crop yield. The model estimates expected yields given limitations in nitrogen, phosphorus and potassium uptake and takes into account the relationship between N, P, and K rather than the demand for individual nutrient elements alone. The model can also be used to generate fertilizer recommendations given soil properties for target yields in such a way that N, P and K are not limiting for yield (Rurinda *et al.*, 2020) . The QUEFTS model entails a four-step process (Fig. S1) that simulates the potential supply of nutrients, plant nutrient uptake, yield ranges and the final yield based on nutrient accumulation and dilution. The four steps are described in detail in Ravensbergen *et al.* (2021), and we only repeat the essentials here:

Step 1: First the potential soil supplies of N, P and K are calculated, by applying relationships between four chemical soil properties, namely pH, SOC, POlsen and Kexch, and the maximum quantity of these nutrients that can be taken up by maize if no other nutrients and no other growth factors are yield-limiting. In addition to the nutrient supply from the soil, nutrient supply from fertilizer application is obtained by accounting for the fertilizer recovery of applied fertilizers.

Step 2: In the second step the actual uptake of each of the three nutrients N, P and K is calculated as a function of the potential supply of that nutrient, taking into account the potential supplies of the other two nutrients. In QUEFTS the actual uptake of a nutrient is calculated twice for each nutrient, where each time only one of the other two nutrients is considered. For instance, the actual uptake of nitrogen is calculated once as a function of its own supply and the supply of phosphorus and once as a function of its own supply and the supply of potassium.

Step 3: QUEFTS converts the estimated uptake of N, P and K into maize yield. For each nutrient, the upper and lower bounds yields are calculated based on the actual uptake of each nutrient. The upper bound yield refers to the yield attainable when for instance N is maximally diluted in the plant. The lower bound yield refers to the yield that could be obtained when N is maximally accumulated in the plant. The actual yield lies somewhere in-between these yields.

This leads to six combinations describing the uptake of one nutrient given maximum dilution or accumulation of another nutrient.

Step 4: In the fourth step, the yield estimates are calculated in pairs on the basis of the actual uptake of each nutrient and the yield ranges calculated in step 3. This will result in six paired estimations which are averaged to obtain a final yield estimate.

In case the calculated yield is below a required level, either because no fertilizer was applied or because too little was applied, then QUEFTS can calculate the required NPK application rates to achieve the required yield level. To compute the required application rates the relation between fertilizer application and yield as outlined in the four steps above is mathematically inversed. The QUEFTS model was run to compute the amount of NPK fertilizer for target yield of 5.0 tons ha^{-1} under two scenarios: the NPK application rates with all soil inputs assumed certain (Scenario 1), and the NPK application rate and associated yield with uncertain soil information (Scenario 2).

The R version of QUEFTS based on Sattari *et al.* (2014) was used and run with the various input parameters: 1) soil properties as described in Section 3.1 (i.e., pH, SOC, KExch and POlsen); 2) QUEFTS model default values for nutrient recovery fractions (0.5 for N and K, 0.1 for P fertilizer); 3) QUEFTS model default values for maximum physiological efficiency for maize (IE borderline) (i.e., 70, 600, 120 kg biomass per kg N, P and K, respectively, and for minimum physiological efficiency (IE borderline) (i.e., 30, 200, and 30 kg biomass per kg N, P and K, respectively; 4) maximum crop season potential yield for the Western Highlands of Cameroon of 5.0 tons ha^{-1} ; and 5) average temperature for the study area of 25 ^{o}C.

3.3 Uncertainty propagation using the Monte Carlo method

Model output uncertainty is generally determined from three main sources, namely input uncertainty, model structure uncertainty, and model parameter uncertainty. In this study, we solely focussed on model input uncertainty, and only considered uncertainty in soil inputs. QUEFTS requires four soil properties (i.e., pH, SOC, POlsen and KExch) as inputs to derive fertilizer recommendation rates to achieve a target yield of 5.0 tons ha^{-1}. Soil input uncertainties propagate through the model, resulting in uncertainties in yield and fertilizer recommendation rates.

To assess how the uncertainty in soil input variables affects QUEFTS outputs, we performed a Monte Carlo (MC) simulation approach. Attractive characteristics of the MC method are easy implementation, general applicability and that it yields the entire probability distribution of the model output (Heuvelink, 1998). The aim of the uncertainty analysis is to quantify the uncertainty of the yield and fertilizer recommendation rates as the result of uncertainties in the soil input variables. It can also be used to analyse the uncertainty contribution of each individual uncertainty source. The MC method repeatedly samples realizations from the probability distributions of the uncertain input variables and runs the model for all realizations. The results of the model simulations are then analysed to estimate the probability distribution of the outputs and quantify the uncertainty. The method as applied here thus consisted of the following steps:

1. Define the mathematical model (QUEFTS) that is to be simulated, including the inputs, parameters, and the targeted output variables.

2. Quantify the uncertainty of the soil inputs by probability distributions (Section 3.1) and draw a large number of realisations (i.e., 500, see below) from them using a pseudo-random number generator.
3. Run the QUEFTS model repeatedly (we used 500 Monte Carlo runs), each time using one of the simulated sets of inputs and store the model outputs. Running the MC analysis several times while changing the seed confirm that 500 runs were sufficient to reach stable results.
4. Construct an empirical probability distribution of the 500 output values. From this distribution summary statistics can be derived, such as the mean, standard deviation, and percentiles.

The frequency distributions of the Monte Carlo simulations of fertilizer recommendation rates and yield represent the propagation of input uncertainty to the model output uncertainty. In particular, the width of these distributions (in this study characterized by the difference between the 0.95 and 0.05 quantiles) signifies the uncertainty of the QUEFTS predicted yield and fertilizer recommendation rates.

3.4 Contributions of soil input variables to the uncertainty of model outputs

The contribution of individual soil input uncertainty to the overall uncertainty of QUEFTS outputs was analysed using a stochastic sensitivity analysis (Saltelli *et al.*, 2008) The uncertainty contribution was expressed as the percentage of the output variance accounted for by each uncertain input. If *m* is the number of uncertain inputs, then we need *m* +1 MC analyses to compute the uncertainty contributions. Initially, the total output uncertainty MC_{tot} is computed by stochastically varying all input variables considered (as explained in Section 3.3). The uncertainty associated with the first input variable x_1 is next obtained through another MC simulation MC_1 in which x_1 is set equal to its deterministic value, while the other input variables vary stochastically. Similarly, the other MC analyses MC_2, MC_3, . . ., MC_m are used to quantify the uncertainty contribution for the other uncertain inputs. The contribution of individual input variables to the uncertainty of the model output, that is the stochastic sensitivity S_i for each uncertain input x_i is then computed as shown in Torres-Matallana *et al.* (2021):

$$S_i = 1 - \frac{var(MC_i)}{var(MC_{tot})} \tag{4}$$

The larger the index S_i, the high the contribution of the input uncertainty x_i.

3.5 Communicating uncertainty and its integration in decision-making

Scientists may be familiar with the concept of uncertainty and methods to quantify it (Brown and Heuvelink, 2005), but end-users are often less familiar and used to deterministic solutions. They often assume or require error-free model outputs to facilitate decision-making. One of the consequences of this desire for simplicity is that modellers do not pay much attention to the uncertainty of their outputs, and therefore do not communicate it to end-users (Verstegen *et al.*, 2012). In practice, achieving error-free model outputs is impossible, and end-users should be assisted to understand uncertainty and how to take decision based on uncertain data. Our interest in this study was to communicate the uncertainty of QUEFTS outputs to a range of end-users including policy makers, extension service agents, and staff of non-governmental organizations. These individuals may in turn be required to communicate uncertainty to the farmers or to the general public and integrate it in decision-making.

Quantification of uncertainty can be straightforward, but communicating uncertainty to a range of users of information and its integration in decision-making processes is less so (Verstegen *et al.*, 2012; Chagumaira *et al.*, 2021). The challenges may include the difficulty to grasp uncertainty for users without basic knowledge of statistics, the fact that uncertainty is input-dependent, the variation of uncertainty over space, and the lack of software that can integrate spatial modelling, uncertainty analysis and visualization (Verstegen *et al.*, 2012). In addition, the success of a method to communicate uncertainty may depend on the subject matter and on the background of the users of information (Milne *et al.*, 2015). Uncertainty can be communicated using an empirical probability distribution, a probability interval, or can be simply described using words, for example, on a verbal scale (Milne *et al.*, 2015; Spiegelhalter, 2017). It is important that the uncertainty statistics are communicated in an efficient way that is both informative and understandable to users with varied backgrounds. There should be ample explanation of the results and a visual representation of the uncertainty to facilitate communication. In this study, we used the 0.05 and 0.95 quantiles of the probability distribution of the spatially distributed yield and fertilizer recommendation rates as a 90 % prediction interval, which expresses the uncertainty about the true values. Furthermore, we opted to visualize uncertainty in the form of maps which show the mean as well as the upper and lower bounds of the prediction interval separately.

Agricultural business decisions, e.g., decisions on whether or not to apply fertilizer in a specific field, depend heavily on model-based recommendations, which are associated with uncertainty. Decisions based on misinterpreted or erroneous model outputs can be costly due to the irreversibility of such decisions (Verstegen *et al.*, 2012). Without understanding this risk, it is not possible for end-users to draw proper conclusions and making optimal decisions. Farmers are aware that fertilizer application improves yield, but fertilization comes at a cost, and they will want to be sufficiently certain that there will be a positive return on investment. Farmers could establish ‘rules’ which account for uncertainty. For example, they could state that they are only willing to apply fertilizer if the probability of obtaining a yield gain above a threshold (e.g., 2.0 tons ha^{-1}) is greater than 90 per cent. Using this example rule, we evaluated which part of the study area would be fertilized. We first derived the optimal fertilizer application for Scenario 1 (no uncertainty in soil inputs) and used these recommendation rates to derive the probability distribution of the yield gain at each location, by subtracting the simulated yield without fertilizer application from that obtained with fertilizer application. Note that this leads to a probability distribution of the yield gain because both the yield with and without fertilizer application are uncertain due to uncertainty about the soil properties. Next, we delineated the area where the expected yield gain (i.e., the mean of the probability distribution) is bigger than the threshold as well as the area where we are at least 90% certain that the yield gain is bigger than the threshold.

3.6 Data processing and statistical computing

All analysis and modelling in this study were conducted in the R statistical open-source software (R Core Team, 2021). The R packages used in this study included “ranger” (Wright and Ziegler, 2017) for fitting the random forest and quantile regression forest models; “raster” (Hijmans, 2021) for handling raster layers; and spplot (Pebesma and

Bivand, 2005) and ggplot (Wickham, 2016) for plotting. All soil maps were produced in QGIS version 3.22 using the Albers equal area projection (EPSG:22832).

4. Results and discussion

4.1 Descriptive statistics of soil property data

The dataset used in this study combined both analytical and spectral soil data. For soil samples with both analytical and spectral data, only analytical data were retained. This resulted in 64 analytical and 546 spectral observations, making a total of 640 observations. Summary statistics of the soil properties are provided in Table S1. The SOC content ranged from 8.7 to 45.6 g kg^{-1} and had a mean of 25.5 g kg^{-1}. This range indicates high potential of crop responses and recovery of applied fertilizers and therefore, a high productivity of crops such as maize could be expected in the area. Soils are generally acidic with pH ranging from 4.2 to 6.3 and a mean of 5.2 and show moderate to severe limitations to crop production. For a large proportion of the soil samples (> 60%), POlsen values were below the critical value of 10 mg kg^{-1} for optimum maize yield (Ussiri *et al.*, 1998; Bai *et al.*, 2013). The deficiency could be due to inherent low P concentration in the parent material and P-fixation (Eijk *et al.*, 2006). The values of KExch were far below the threshold value of 2 mmol kg^{-1} required for the growth of major crops (Chilimba *et al.*, 1999). These findings indicate that P and K fertilizers are needed to enhance maize production in the study area.

4.2 Spatial prediction of soil properties and uncertainty quantification

The correlation analysis and variable selection procedure indicated that not all 191 covariates were useful in explaining the spatial variation of the soil properties of interest. The optimal number of covariates included in the RF predictive model for pH, SOC, POlsen and KExch were 9, 3, 22 and 57, respectively. Plots of recursive feature elimination showing model performance as a function of number of covariates are shown in Fig. S1. The LCOCV statistics used to evaluate the performance of each predictive model are presented in Table 1. The ME values were close to zero for all properties. The degree to which the spatial variation of soil properties was predicted from the available covariates varied substantially. Spatial variation in pH, POlsen and SOC was best described, with MEC values above 0.60, while variation in KExch was less described, with a MEC value of 0.41. The low MEC value for KExch is likely caused by the weak relation between the soil property and the covariates, which leads to a higher prediction uncertainty. Table 1 also summarizes the PICPs for the models. For all four soil properties, the PICP ranged between 0.89 and 0.92. This indicates that the prediction intervals obtained with QRF were a realistic representation of the various model prediction uncertainties, as these were close to 0.90.

Table 1 Leave-cluster-out cross-validation metrics for the random forest predictions of the four soil properties

Soil property	ME	RMSE	MEC	PICP
pH	-0.001	0.157	0.832	0.91
SOC (g kg^{-1})	-0.020	5.620	0.618	0.89
POlsen (mg kg^{-1})	0.010	3.450	0.640	0.89
KExch (mmolc kg^{-1})	-0.001	0.130	0.410	0.92

For each soil property, we spatially predicted the mean value using RF, as well as the lower and the upper limits of the 90% PI with QRF. The results for SOC are presented in Fig. 2, while those of the three other soil properties are provided in Fig. S3. Predicted pH values ranged from 4.68 to 6.04, SOC values from 12.30 to 43.85 g kg^{-1}, POlsen from 1.36 to 11.81 mg kg^{-1} and KExch from 0.10 to 0.74 mmol kg^{-1}. It can be seen from the maps that lower values of pH and SOC are found in the south-western parts, while lower values of POlsen and KExch are found in the central and the northern parts of the study area. These two areas are dominated by Ferralsols which have a low nutrient retention capacity. High values of pH, SOC and KExch are found in a region in between these two areas. This region is a mountainous volcanic area dominated by fertile Nitisols which are well-drained with very favorable chemical and physical properties.

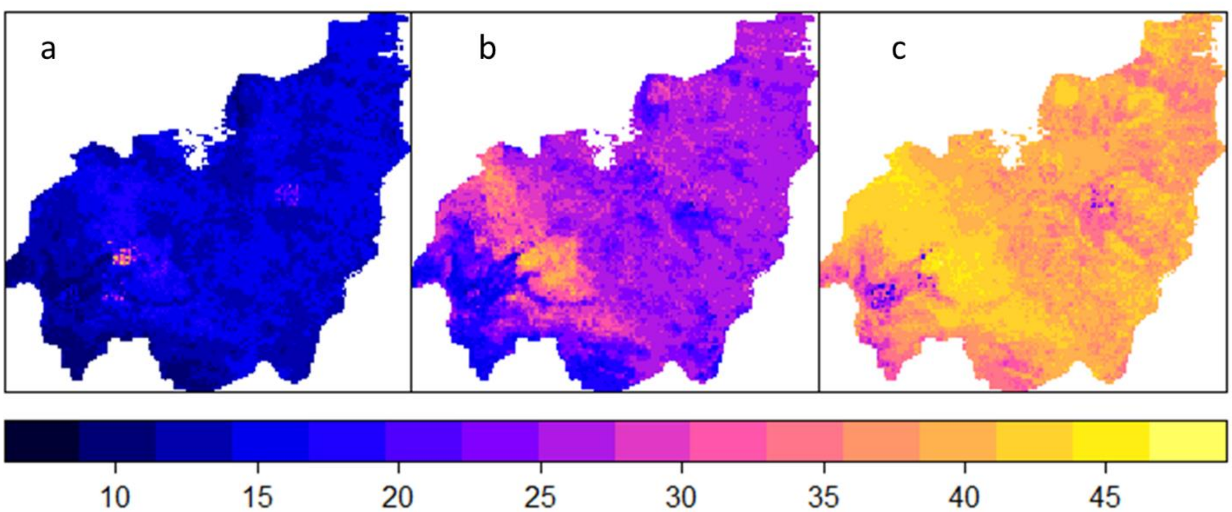

Fig. 2 Spatial distribution of SOC predictions and associated limits of the 90% prediction interval (g kg^{-1}): a) lower limit, b) mean values, c) upper limit

4.3 QUEFTS model outputs ignoring uncertainty in soil inputs

Maps of the potential soil supply of the three macro-nutrients nitrogen, phosphorus and potassium estimated by QUEFTS under Scenario 1 are presented in Fig. S4. It can be seen from the maps that the south-western and central parts of the study area have low soil N and P supplies, whereas in between these areas there is a region with high N and P supply. The soil K supply map has a very different spatial pattern and shows high supply of K in the southern part of the study area. With no fertilizer applied, the predicted yield ranged from 1.44 to 4.94 with a mean of 3.26 tons ha^{-1} (Table 2).

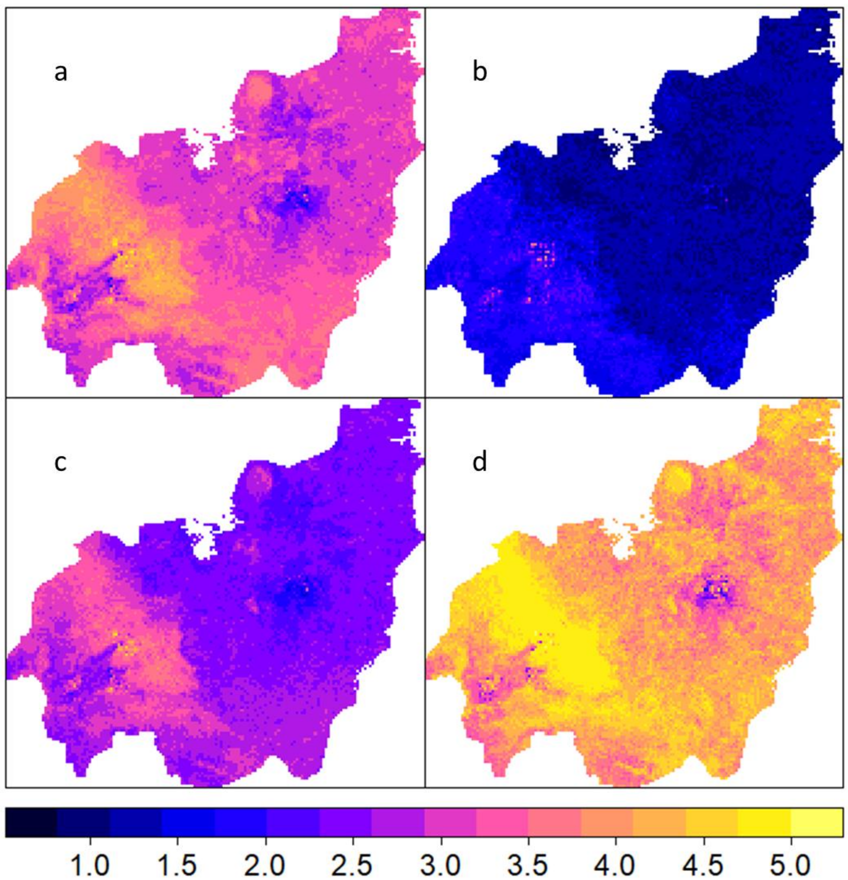

Fig. 3 Spatial distribution of yield as predicted by QUEFTS without fertilizer application: a) yield ignoring uncertainty in soil inputs, b) 0.05 quantile of 500 MC yield maps; c) mean of 500 MC yields; d) 0.95 quantile of 500 MC yields

The spatial distribution of the yield as predicted by the QUEFTS model without fertilizer application is presented in Fig. 3a. Predicted yield values in some areas varied from low to high within a very short distance, e.g., in the central and south-west parts of the study area. The spatial distribution of the predicted yield agreed with those of the N and P soil supply maps and much less with the K soil supply map (Fig. 4). With the estimated soil nutrient supplies, the QUEFTS model was run to estimate the optimum NPK fertilizer recommendation rates required to achieve the target yield of 5.0 tons ha^{-1}, assuming the predicted soil properties to be certain (Scenario 1). Summary statistics of the predicted NPK recommendation rates are presented in Table 2. Fertilizer application allowed to reach the target yield across the entire study area. There were considerable yield gains across the area when NPK fertilizer were applied ranging from 0.06 to 3.56 with a mean of 1.74 tons ha^{-1}.

The spatial distribution of recommended NPK fertilizers to achieve a target yield of 5.0 tons ha^{-1} is presented in Fig. 4 and shows high variability across the study area. Summary statistics of the predicted yield gain and NPK recommendation rates are presented in Table 2. As expected, the maps are a mirror image of the soil nutrient supply maps shown in Fig. S4. Some parts of the study area have sufficient N and P supply from the soil and do not require N and P application to reach the target yield, whereas other parts have low soil supply and need a high fertilizer rate to reach the target yield. The spatial variation of the optimized NPK application is very large indicating that blanket fertilizer recommendation would not be a suitable policy to achieve the target yield of 5.0 tons ha^{-1} across the study area. The maximum N application rate is higher than that of P and K (Fig. 4, Table 2), while the mean recommended application rate is highest for K. This is probably the result of spatial variation in soil nutrient supplies, which are influenced by other environmental factors, not accounted for by QUEFTS.

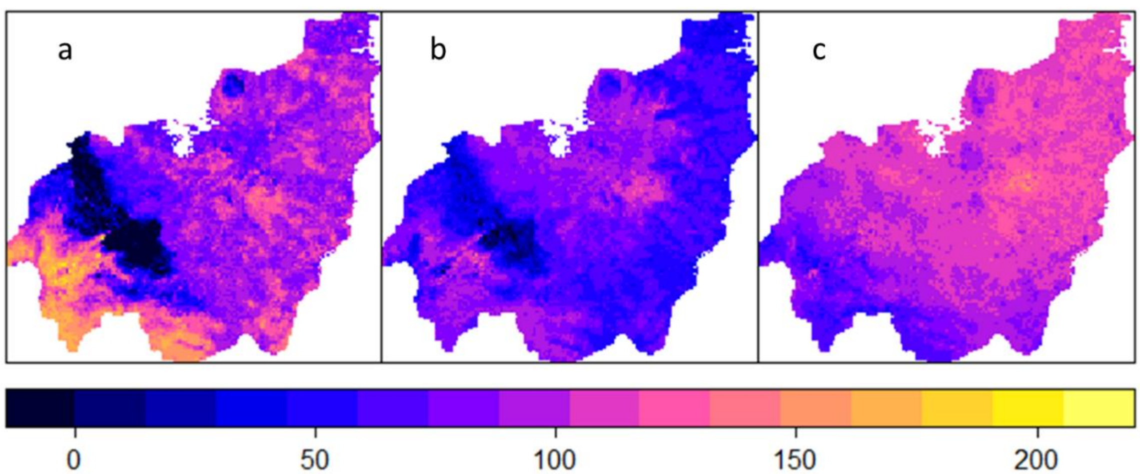

Fig. 4 Fertilizer recommendation rates required to obtain the targeted yield of 5.0 tons ha^{-1} as predicted by the QUEFTS model while ignoring uncertainty in soil inputs (in kg ha^{-1}): a) N fertilizer; b) P fertilizer; c) K fertilizer

Table 2 Summary statistics of the QUEFTS model predicted yield without fertilizer application and NPK recommendation rates required to achieve a target yield of 5.0 tons ha^{-1} under Scenario 1 (ignoring soil input uncertainty)

	Tons ha^{-1}	kg ha^{-1}			
Statistic	Yield without NPK application (tons ha^{-1})	N Fertilizer	P fertilizer	K fertilizer	Yield gain (tons ha^{-1})

		(kg ha^{-1})	(kg ha^{-1})	(kg ha^{-1})	
Maximum	4.94	205.8	143.2	157.0	3.56
75th percentile	3.47	107.2	84.4	115.2	2.00
Mean	3.26	89.8	69.1	102.6	1.74
25th percentile	3.02	77.4	56.9	93.8	1.52
Minimum	1.44	0.0	0.0	32.9	0.06

4.4 QUEFTS model outputs with uncertain soil inputs

We applied the same procedure as in Section 4.3 and used QUEFTS to calculate the yield without fertilizer application and predict the fertilizer recommendation rates required to achieve the target yield of 5.0 tons ha^{-1}, but now assuming the four soil inputs to be uncertain (Scenario 2). Each of the 500 MC runs produced a realisation of the yield and fertilizer recommendations, so that the frequency distribution of the 500 outputs approximates the probability distribution of the uncertain yield and fertilizer recommendation rates. We verified that 500 MC runs was sufficient to obtain stable results, by redoing the analysis using a different random seed. Differences were indeed small (results not shown). In a case without fertilizer application, we compared the yield map of the deterministic run with the mean of the results of the 500 MC simulations and found substantial differences (Fig. 3). For some locations, differences in predicted yield between the deterministic run and the mean of the 500 MC runs were close to 1.0 tons ha^{-1}.

We observed a spatial average of 3.26 tons ha^{-1} for the deterministic run (Fig. 3a) and 2.64 tons ha^{-1} for the mean of the MC runs (Fig. 3b). This shows that the yield is systematically overestimated when uncertainty in soil inputs is ignored. These systematic differences arise from the non-linear nature of the QUEFTS model, where yield increments are more pronounced for soil supply increments in the low range compared to the high range of soil nutrient supply (Dhakal and Lange, 2021).

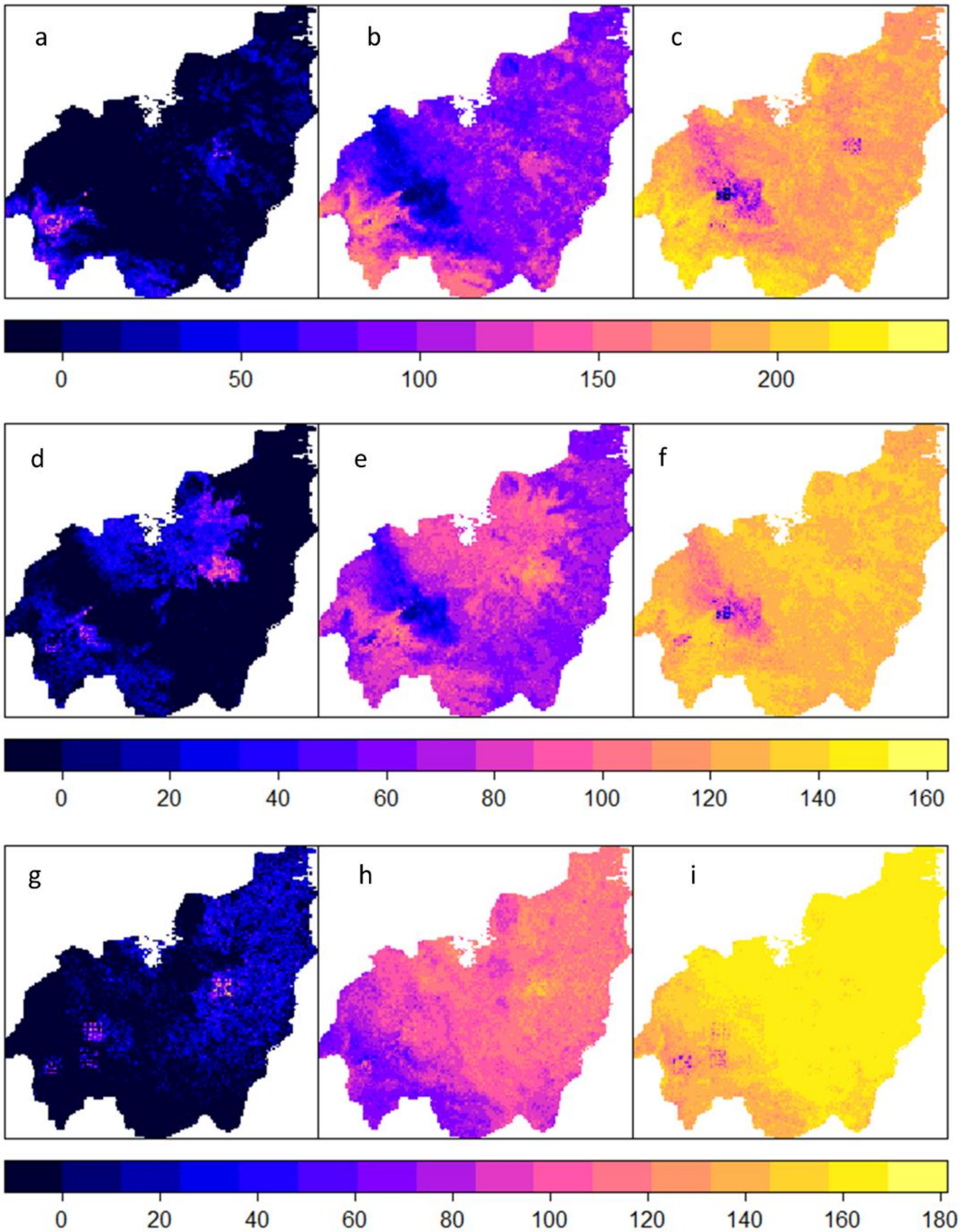

Fig. 5 Spatial distribution of fertilizer recommendation rates calculated using QUEFTS (kg ha^{-1}) under Scenario 2: lower limit, mean values, and upper limit for N (a, b, c); P (d, e, f); and K (g, h, i)

Maps of the mean and 0.05 and 0.95 quantiles obtained by propagating uncertainty in soil input variables through the QUEFTS model are presented in Fig. 5, while those of the 90% prediction interval widths are presented in Fig. S5. Uncertainty in soil inputs resulted in a large uncertainty of the NPK fertilizer recommendation rates required to reach the target yield. As shown in Fig. 5, there is a large difference between the lower and upper limits of the prediction intervals. As a result, the 90% prediction interval widths are very large. High uncertainty of soil properties leads to high uncertainty of soil nutrient supplies, and this in turn means that we are highly uncertain about how much fertilizer should be applied to reach the target yield 5.0 tons ha^{-1}. The uncertainty maps of Fig. S5 show that the spatial distribution of uncertainty is not homogeneous and varies substantially across the study area. Relatively low values are found close to the sentinel sites, where soil input uncertainty is lower than elsewhere in the study area. The 90% prediction interval was generally wider for N and K fertilizers and narrower for P fertilizer, particularly in the western and northern parts of the study area. For instance, lower uncertainty for N was observed in areas where N fertilizer recommendation is low (see Fig. 4a). This corresponds to areas with fairly high soil supply of N. In these areas, we are less uncertain about the recommended N fertilization, because in many MC runs the soil supply is sufficient to reach the target yield, so that the N application rate is zero.

The spatial variation of the QUEFTS model outputs can also be displayed as spatial cumulative frequency distributions. In Scenario 2, this yields a cumulative distribution for each single MC run. We represented the 500 MC curves by curves of the mean and the 0.05 and 0.95 quantiles (Fig. 6). Incorporation of uncertainty leads to a systematic shift of the cumulative distributions of the three fertilizer applications to the right (higher values) while that of the yield shifted to the left (lower values) when uncertainty in soil inputs is accounted for. This indicate an overestimation of the yield and slight underestimation of the fertilizer recommendation when uncertainty in soil inputs is ignored. Though the spatial variation is not large, as shown by the steepness of the curves, large uncertainty was observed, as indicated by the wide prediction intervals for the four model outputs, in particular for N fertilizer recommendation.

Thus, Fig. 6 nicely combines and allows to compare spatial variability and uncertainty in one figure (Heuvelink and Pebesma, 1999).

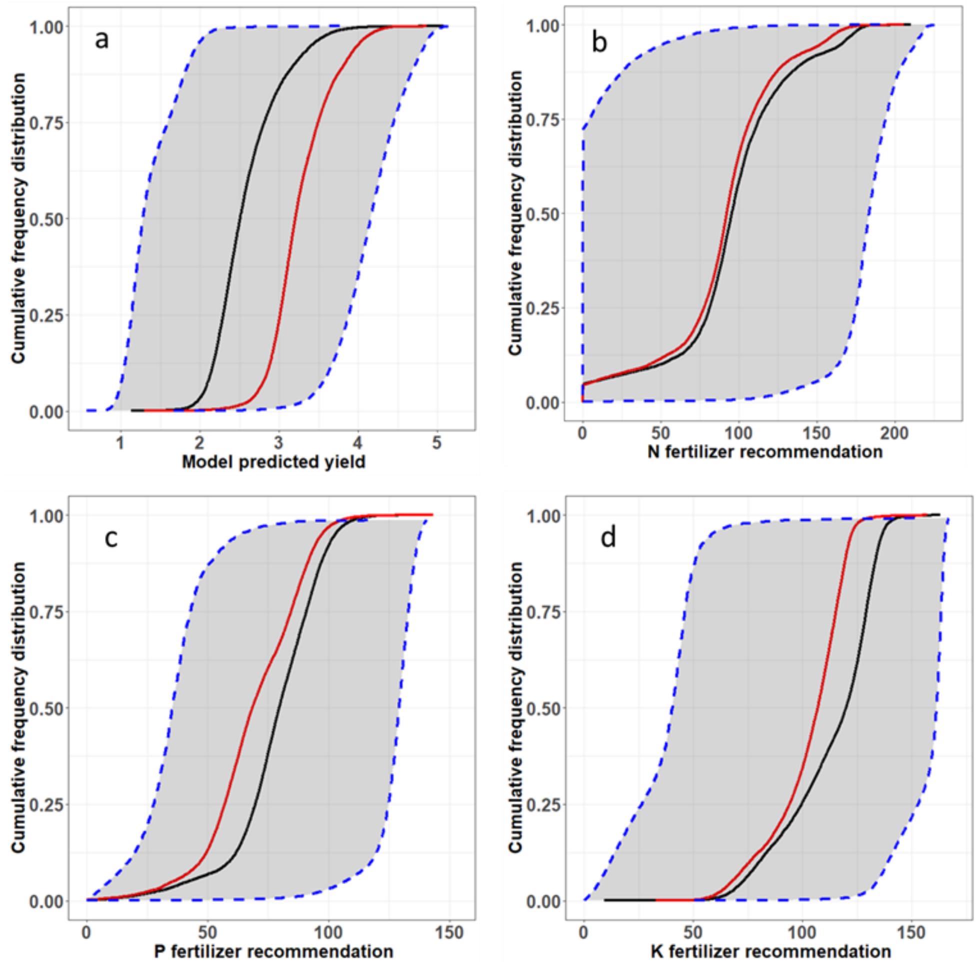

Fig. 6 Spatial cumulative frequency distribution of the yield (tons ha^{-1} and the N, P and K fertilizer recommendations (kg ka^{-1}) for cases with and without soil uncertainty: red lines represent deterministic run, black lines the mean of all MC runs, dashed blue lines represent the 0.05 and 0.95 quantiles. The grey area indicates uncertainty about the position of the 'true' cumulative distribution, while the steepness of the curves shows the degree of spatial variation

4.5 Contribution of soil input variables to the uncertainty of fertilizer recommendation rates

We decomposed the uncertainty of each of the three fertilizer recommendation rates into the contribution of the four soil input variables using the stochastic sensitivity analysis as described in Section 3.4. An additional four MC simulations with 500 runs each were performed to estimate the stochastic sensitivity S_i of the input variables pH, SOC, POlsen and KExch (Table 3). The relative contributions of input variables varied greatly among the four soil input variables.

Soil pH is the main source of uncertainty for K fertilizer (82%) while uncertainty in SOC is by far the most dominant for N fertilizer (97%), and a substantial source of uncertainty for P fertilizer (25%) and K fertilizer (18%). As expected, POlsen is the main source of uncertainty for P fertilizer (55%). The second-most important variable that contributes to the uncertainty of P fertilizer is KExch, with a stochastic sensitivity coefficient of about 21%. KExch is the least important uncertainty contributor for N and P fertilizers. From these results, we can infer that pH, SOC and POlsen were the dominant sources of uncertainty for the N, P, and K fertilizer recommendation rates. K fertilizer uncertainty is more influenced by pH uncertainty as compared to N and P. At low pH, the soil's ability to keep supplying potassium to plants is decreased, therefore potentially increasing the need for additional P fertilizer application. SOC uncertainty contributes strongly to N fertilizer uncertainty due to its direct relationship with organic matter and the strongly correlation between the two properties (Takoutsing *et al.*, 2017).

Table 3 Relative contributions of soil input variables to the uncertainty of QUEFTS model outputs in terms of percentage of total variance

	Stochastic sensitivity (S_i) of input variable (%)		
Soil inputs	N fertilizer	P fertilizer	K fertilizer
pH	0.1	0.0	81.6
SOC	96.7	24.8	18.2
POlsen	0.1	54.8	0.0
KExch	3.2	20.5	0.2

4.6 Communicating model output uncertainty to end-users and integration in decision-making processes

The uncertainty propagation analysis showed that the fertilizer recommendation rates derived by the QUEFTS model were highly uncertain due to uncertainty about soil inputs. The uncertainty was visualised by jointly plotting the lower and upper limits of prediction intervals (Figs. 2, 5, S3) and the prediction interval width maps (Fig. S5). The probability that the true value lies within a prediction interval might not be easily interpreted by a range of users, particularly farmers. Moreover, it is also challenging to include the reported uncertainty in decision-making on fertilizer recommendations. Section 3.5 explained how end-users can be supported by providing maps that show where it is sufficiently certain that fertilizer application will pay off.

We considered a yield gain of 2.0 tons ha^{-1} as the minimum threshold that will encourage farmers to apply fertilizers. Fig. 7 shows which parts of the study area are expected to have a yield gain above the threshold, as well as parts where the probability of a yield gain above the threshold is at least 90%. The map shown in Fig. 7 could be a simple and useful tool to communicate uncertainty to end-users for a rapid decision on whether to apply fertilizer in a specific plot or not. For instance, farmers are advised to apply fertilizer in the green areas, but not to the orange areas. However, the decision to apply in the yellow areas is risky for the farmers because it is far from certain that the threshold will be reached. Farmers who are risk-aversive would probably not apply fertilizer in these areas, meanwhile those who can afford the risk can apply. The yellow areas could be targeted by farmers to reduce uncertainty in soil properties through additional soil sampling, hence also reducing uncertainty in yield gain.

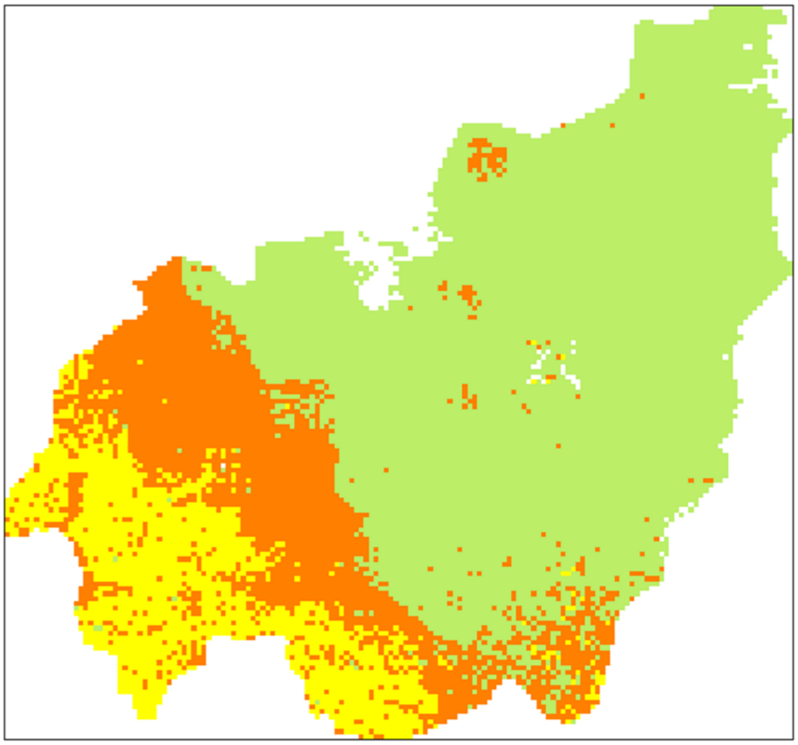

Fig. 7 Maps showing areas where yield gains are above the threshold of 2.0 tons ha^{-1}: green = probability that yield gains are above threshold ≥ 0.90, yellow = average of yield gain over 500 MC runs not certain to be > 2.0 tons ha^{-1}, orange = average of yield gain over 500 MC runs < 2.0 tons ha^{-1}

5. General discussion

5.1 Uncertainty propagation analysis and implication for QUEFTS modelling

One of the key objectives of this study was to quantify the uncertainty in soil information and analyse how this uncertainty propagates through the QUEFTS model to affect yield and fertilizer recommendations. We illustrated the methodology by combining QRF with Monte Carlo simulation to quantify uncertainty in soil inputs and explore their impacts on the overall uncertainty of the model outputs. We also quantified the contributions of individual soil input uncertainty to the QUEFTS outputs. Then, we summarised the model output probability distributions in a map that shows where fertilizer application has almost certainly a large effect, thus providing a tool to communicate uncertainty to end-users and supporting the integration of uncertain information in decision-making. The methodology applied in a case study in the Western Highlands of Cameroon yielded satisfactory results in recommending fertilizer application rates required to achieve a target maize yield across the study area, along with associated uncertainty. To the best of our knowledge, this study represents the first analysis of the uncertainty in QUEFTS model outputs in Cameroon.

However, there is general consensus among modellers that uncertainty originating from inputs could be a major source of uncertainty in the uncertainty of crop model outputs, given that inputs constitute the most substantial data source (Dokoohaki et al., 2021). However, soil input uncertainty is only one of many sources of uncertainty in crop models, and other sources of uncertainty might well be more important sources. But in this study, we focused only on the uncertainty in four soil inputs, which in itself is worthwhile but of course other uncertainty sources should also be considered in future research. Moreover, we found that just uncertainty in the four soil inputs already resulted in large uncertainty in yield and fertilizer recommendation, thus underlining the importance of this study and the importance of improving the accuracy of the soil maps that are used as input in the QUEFTS model. The results of this study showed a large contribution of soil input to the uncertainty of yield and fertilizer recommendation rates, warranting the need to further reduce crop modelling uncertainty by optimizing the accuracy of soil property maps. Improving soil map accuracy could be achieved by increasing data collection through extensive field sampling, using superior covariates, and employing advanced DSM models like ensemble approaches.

Not all soil inputs had equal contribution to the output uncertainty, and this varied greatly depending on the output variable considered (Table 3). Clearly, the overall output uncertainty and uncertainty contributions will vary from one case study to another. This is because the magnitude of the soil map uncertainty is heavily influenced by the soil spatial variability, the spatial sampling design and sampling density, and by the ability of the covariates to explain soil spatial variation (Heuvelink and Webster, 2022; Pusch *et al.*, 2023). In other case studies, it is conceivable that the model output may exhibit varying degrees of sensitivity to specific inputs, with the potential for greater sensitivity in one scenario compared to another. Therefore, the contribution of uncertainty is significant only when the input uncertainty is large, and the model demonstrates sensitivity to that specific input in a specific case study (Nol *et al.*, 2010). Understanding the apportioned uncertainty from different soil inputs helps recognize the key soil properties influencing model outcomes. For instance, certain soil properties account for over 90% of the uncertainty in fertilizer recommendations This knowledge guides efforts to map these inputs more accurately, focusing resources on components with the most significant impact on model uncertainty (Brown and Heuvelink, 2005).

Though we ignored other sources of uncertainty, we contend that comprehensive uncertainty quantification that account for all sources of uncertainty is vital for crop model validation and reproducibility. Other inputs, such as climate, crop parameters, and management practices are very uncertainty and could have major contributions to QUEFTS output uncertainty, The uncertainty in soil inputs was translated into large uncertainty in fertilizer recommendations (Fig. 5) to reach the target yield of 5 tons ha-1 across the entire study area. We think that aiming for a constant target yield for all locations may not be fair, given the high uncertainty in fertilizer recommendations. The results indicated that some parts of the study area do not require fertilizer to reach the targeted yield due to inherent favourable soil conditions (Fig. 3). A higher yield could be aimed for such areas, while investment to increase yield in areas with very poor soil conditions should be limited.

5.2 Communicating model uncertainty and integration in decision-making processes

One of the objectives of this study was to provide a method for communicating QUEFTS model outputs to end-users and how it can be integrated in decision-making. We argue that instead of reporting a single mean value, i.e. result from a deterministic run, the entire probability distribution in model outputs should be recognized and reported, since it represents the uncertainty about the output, and how confident modellers are with the results of their predictions. The uncertainty in soil inputs resulted in large uncertainty in fertilizer recommendation rates and yield gains (Fig 6). We summarized the output probability distributions in a map that shows where fertilizer application has almost certainly a large effect on yield gains (Fig. 7). This map provided a simple visual aid to show the spatial distribution of the uncertainty for a threshold of yield gains. By showing the areas where there is sufficient certainty that fertilization pays off, the map provides an intuitive way for end-users to integrate uncertainty in decision-making. This is under the assumption that the only source of uncertainty is the four soil properties considered in this study. In the real world, there are many other factors including climate, pest and diseases, and crop management practices that affect yield, so we are not certain that the yield gain will always be above the threshold at all locations across the study area. This mode of visualization for uncertain spatial data could be very suitable for end-users without profound knowledge of statistics. Communicating the impacts of uncertainty in this spatially explicit way is a contribution to the ongoing dialogue taking place between modellers, policy makers, and farmers on how uncertainty about model outputs could be communicated (Getson *et al.*, 2022).

A significant factor contributing to the gap between a modellers' comprehension of model output uncertainty and the end-users' utilization concerns the uncertainty communication method and tools. Ineffective communication strategies, especially concerning risk communication (Begho *et al.*, 2022), can result in heightened misinterpretation of uncertainty, potentially fostering disbelief in the impact of model outputs. This is especially critical in the context of fertilizer recommendations that entail substantial investments on the part of small-scale farmers with limited resources (Islam *et al.*, 2022). Drawing from experiences in other fields, particularly the health sector, where effective communication and visualization approaches for conveying uncertainty have been developed and used in critical situations such as the COVID-19 pandemic (McCabe *et al.*, 2021), it becomes evident that scientists often do not convey uncertainty through appropriate communication channels or presented in a format easily digestible by farmers for integration into decision-making (Spiegelhalter, 2017). This could also be the lack of skills on the part of scientists to design suitable uncertainty communication tools that cater to a diverse range of stakeholders. Therefore, efforts to enhance the capacities of both modellers and end-users could help overcome the challenge (Milne *et al.*, 2015). Well-formulated fertilizers can boost yields but pose risks for small-scale farmers. Our QUEFTS study showed that large uncertainty in fertilizer recommendations can lead to risk-averse decisions (Monjardino et al., 2015). Due to financial constraints, farmers tend to be risk-averse because fertilizer application is a high-risk, high-return agricultural business

(Haile et al., 2020). In a single input case like fertilizer, a risk-averse farmer uses fewer inputs than a risk-neutral counterpart if the input increases output variability, assuming all other factors constant (Begho et al., 2022).

5.3 Limitations to our study and possible improvements

While the QUEFTS model demonstrated satisfactory performance based on the objectives of this study, it is essential to acknowledge certain limitations and challenges. These areas provide opportunities for further development and improvement in the model, as well as enhancing the results of uncertainty quantification and propagation analysis for crop modelling.

Non-consideration of other sources of uncertainty: The DSM and crop modelling processes encompass multiple error sources, contributing to overall output uncertainty. These errors stem from input data inaccuracies, model limitations, environmental variations, and challenges in scaling. Human errors, imprecise parameterization, and the dynamic nature of cropping system further add complexity to error propagation. While this study specifically addressed soil input uncertainties, neglecting other inputs and QUEFTS model uncertainties, it is crucial to recognize that overlooking an important uncertainty source may lead to an underestimation of the overall output uncertainty. To enhance future studies, careful consideration of all sources of uncertainties is essential for a comprehensive estimation of the accuracy of the model outputs.

Not-accounting for cross-correlation between the uncertainty of soil inputs: In our uncertainty propagation analyses, we did not account for interactions between the uncertainty of soil inputs, a factor we recognize as a limitation. We posit that incorporating such correlations would have provided a more realistic model uncertainty, given the known strong correlations among some of the soil variables. Employing methods that simultaneously model multiple soil properties, such as multivariate random forest or regression kriging models could be a valuable approach to quantify the correlations between the uncertainties of multiple variables, and even simulate from the joint probability distribution (van der Westhuizen et al., 2023). If such correlation can be quantified, then it can also be incorporated in the MC analysis, and this can be explored in further research.

6. Conclusion

Our study proposed a methodological procedure for quantifying and propagating uncertainty in soil inputs to QUEFTS outputs, focusing on its impacts on yield and fertilizer recommendation rates in the Western Highlands of Cameroon. The results showed that uncertainty in soil inputs resulted in a large uncertainty in the NPK fertilizer recommendation rates required to reach the target yield of 5 tons ha^{-1}. There was an overestimation of the yield and underestimation of the fertilizer recommendations when uncertainty in soil inputs was ignored. Comparison of the results of a deterministic model run with the mean of the Monte Carlo simulation runs showed systematic differences up to 1.0 tons ha^{-1} for maize yield and up to 59, 42, and 20 kg ha^{-1} for N, P, and K fertilizers, respectively in some parts of the study area. Stochastic sensitivity analysis showed that pH was the main source of uncertainty for K fertilizer (81.6%) and that SOC contributed most to the uncertainty of N fertilizer (97%). Uncertainty in P fertilizer mostly came from uncertainty in POlsen (55%) and KExch (20%). A threshold probability map of yield gains designed using statistical predictions served as a visual tool that empowers farmers to swiftly make informed decisions about fertilizer application locations. The apportionment of the contribution of different soil inputs to model outputs facilitates prioritizing future efforts in crop modelling to reduce uncertainty around yield gains and enhance agricultural intensification. This approach developed in this study provides valuable spatial insights into crop nutrient requirement estimations with associated uncertainty, enabling farmers to adopt tailored fertilizer application based on specific soil conditions.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability statement

The data used in this study can be made available upon request to the corresponding author.

Funding statement

We receive no funding for the study.

References

Bai, Z., Li, H., Yang, X., Zhou, B., Shi, X., Wang, B., Li, D., Shen, J., Chen, Q., Qin, W., Oenema, O., Zhang, F., 2013. The critical soil P levels for crop yield, soil fertility and environmental safety in different soil types. Plant and Soil 372, 27-37.

Begho, T., Eory, V., Glenk, K., 2022. Demystifying risk attitudes and fertilizer use: A review focusing on the behavioral factors associated with agricultural nitrogen emissions in South Asia. Frontiers in Sustainable Food Systems 6.

Breure, M.S., Kempen, B., Hoffland, E., 2022a. Spatial predictions of maize yields using QUEFTS – A comparison of methods. Geoderma 425, 116018.

Breure, T.S., Haefele, S.M., Hannam, J.A., Corstanje, R., Webster, R., Moreno-Rojas, S., Milne, A.E., 2022b. A loss function to evaluate agricultural decision-making under uncertainty: a case study of soil spectroscopy. Precision Agriculture.

Brown, J.D., Heuvelink, G.B.M., 2005. Assessing Uncertainty Propagation through Physically Based Models of Soil Water Flow and Solute Transport. Encyclopedia of Hydrological Sciences.

Chagumaira, C., Chimungu, J.G., Gashu, D., Nalivata, P.C., Broadley, M.R., Milne, A.E., Lark, R.M., 2021. Communicating uncertainties in spatial predictions of grain micronutrient concentration. Geosci. Commun. 4, 245-265.

Chapagain, R., Remenyi, T.A., Harris, R.M.B., Mohammed, C.L., Huth, N., Wallach, D., Rezaei, E.E., Ojeda, J.J., 2022. Decomposing crop model uncertainty: A systematic review. Field Crops Research 279, 108448.

Chilimba, A.D.C., Mughogho, S.K., Wendt, J., 1999. Mehlich 3 or modified olsen for soil testing in Malawi. Communications in Soil Science and Plant Analysis 30, 1231-1250.

Dhakal, C., Lange, K., 2021. Crop yield response functions in nutrient application: A review. Agronomy Journal 113, 5222-5234.

Dokoohaki, H., Kivi, M.S., Martinez-Feria, R., Miguez, F.E., Hoogenboom, G., 2021. A comprehensive uncertainty quantification of large-scale process-based crop modeling frameworks. Environmental Research Letters 16, 084010.

Eijk, D., Janssen, B., Oenema, O., 2006. Initial and residual effects of fertilizer phosphorus on soil phosphorus and maize yields on phosphorus fixing soils. A case study in South-West Kenya. Agriculture, Ecosystems & Environment 116, 104-120.

Ezui, K.S., Franke, A.C., Ahiabor, B.D.K., Tetteh, F.M., Sogbedji, J., Janssen, B.H., Mando, A., Giller, K.E., 2017. Understanding cassava yield response to soil and fertilizer nutrient supply in West Africa. Plant and Soil 420, 331-347.

FAOSTAT, 2020. Statistical Database. Food and Agriculture Organization of the United Nations, Rome.

Getson, J.M., Church, S.P., Radulski, B.G., Sjöstrand, A.E., Lu, J., Prokopy, L.S., 2022. Understanding scientists' communication challenges at the intersection of climate and agriculture. PLOS ONE 17, e0269927.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337-350.

Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: D. Arrouays, N. McKenzie, J. Hempel, A. Richer de Forges, McBratney, A. (Eds.), GlobalSoilMap. Basis of the Global Spatial Soil Information System, pp. 335 - 340.

Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modelling. Geoderma 89, 47-65.

Heuvelink, G.B.M., Webster, R., 2022. Spatial statistics and soil mapping: A blossoming partnership under pressure. Spatial Statistics 50, 100639.

Hijmans, R.J., 2021. raster: Geographic Data Analysis and Modeling. R package version 3.5-2

Islam, M.S., Bell, R.W., Miah, M.A.M., Alam, M.J., 2022. Farmers' fertilizer use gaps relative to government recommendations in the saline coastal zone of the Ganges Delta. Agronomy for Sustainable Development 42, 59.

IUSS, W.G.W., 2015. World reference base for soil resources 2014. FAO, Rome.

Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J., van Reuler, H., 1990. A system for quantitative evaluation of the fertility of tropical soils (QUEFTS). Geoderma 46, 299-318.

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecological Modelling 83, 55-66.

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., Palm, C., Huising, J., 2016. Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa. Agriculture, Ecosystems & Environment 229, 1-12.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28, 1 - 26.

Lark, R.M., Chagumaira, C., Milne, A.E., 2022. Decisions, uncertainty and spatial information. Spatial Statistics 50, 100619.

Lark, R.M., Lapworth, D.J., 2012. Quality measures for soil surveys by lognormal kriging. Geoderma 173-174, 231-240.

McCabe, R., Kont, M.D., Schmit, N., Whittaker, C., Løchen, A., Walker, P.G.T., Ghani, A.C., Ferguson, N.M., White, P.J., Donnelly, C.A., Watson, O.J., 2021. Communicating uncertainty in epidemic models. Epidemics 37, 100520.

Meinshausen, N., 2006. Quantile Regression Forests. Journal of Machine Learning Research 7, 983-999.

Mesfin, S., Haile, M., Gebresamuel, G., Zenebe, A., Gebre, A., 2021. Establishment and validation of site specific fertilizer recommendation for increased barley (Hordeum spp.) yield, northern Ethiopia. Heliyon 7, e07758-e07758.

Milne, A.E., Glendining, M.J., Lark, R.M., Perryman, S.A., Gordon, T., Whitmore, A.P., 2015. Communicating the uncertainty in estimated greenhouse gas emissions from agriculture. J Environ Manage 160, 139-153.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301-311.

Neba, A., 1999. Modern Geography of the Republic of Cameroon. Neba Publishers, Bamenda.

Nol, L., Heuvelink, G.B.M., Veldkamp, A., de Vries, W., Kros, J., 2010. Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. Geoderma 159, 9-23.

Ojeda, J.J., Rezaei, E.E., Kamali, B., McPhee, J., Meinke, H., Siebert, S., Webb, M.A., Ara, I., Mulcahy, F., Ewert, F., 2021. Impact of crop management and environment on the spatio-temporal variance of potato yield at regional scale. Field Crops Research 270, 108213.

Pebesma, E., Bivand, R., 2005. S Classes and Methods for Spatial Data: the sp Package.

Pusch, M., Samuel-Rosa, A., Sergio Graziano Magalhães, P., Rios do Amaral, L., 2023. Covariates in sample planning optimization for digital soil fertility mapping in agricultural areas. Geoderma 429, 116252.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Ramirez-Villegas, J., Koehler, A.-K., Challinor, A.J., 2017. Assessing uncertainty and complexity in regional-scale crop model simulations. European Journal of Agronomy 88, 84-95.

Ravensbergen, A.P.P., Chamberlin, J., Craufurd, P., Shehu, B.M., Hijbeek, R., 2021. Adapting the QUEFTS model to predict attainable yields when training data are characterized by imperfect management. Field Crops Research 266, 108126.

Rurinda, J., Zingore, S., Jibrin, J.M., Balemi, T., Masuki, K., Andersson, J.A., Pampolino, M.F., Mohammed, I., Mutegi, J., Kamara, A.Y., Vanlauwe, B., Craufurd, P.Q., 2020. Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa. Agric Syst 180, 102790.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis: the primer. John Wiley & Sons.

Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L., Janssen, B.H., 2014. Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model. Field Crops Research 157, 35-46.

Schut, A.G.T., Giller, K.E., 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. Geoderma 380, 114680.

Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. Neural Networks 19, 225-235.

Smaling, E.M.A., Janssen, B.H., 1993. Calibration of quefts, a model predicting nutrient uptake and yields from chemical soil fertility indices. Geoderma 59, 21-44.

Spiegelhalter, D., 2017. Risk and Uncertainty Communication. Annual Review of Statistics and Its Application 4, 31-60.

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.

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.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.

Torres-Matallana, J.A., Leopold, U., Heuvelink, G.B.M., 2021. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. Hydrol. Earth Syst. Sci. 25, 193-216.

Ussiri, D.A., Mnkeni, P.N.S., MacKenzie, A.F., Semoka, J.M.R., 1998. Soil test calibration studies for formulation of phosphorus fertilizer recommendations for maize in Morogoro district, Tanzania. I. evaluation of soil test methods. Communications in Soil Science and Plant Analysis 29, 2801-2813.

Vågen, T.-G., Winowiecki, A., 2020. The Land Degradation Surveillance Framework Field Guide. ICRAF, Nairobi.

van Ittersum, M.K., van Bussel, L.G.J., Wolf, J., Grassini, P., van Wart, J., Guilpart, N., Claessens, L., de Groot, H., Wiebe, K., Mason-D'Croz, D., Yang, H., Boogaard, H., van Oort, P.A.J., van Loon, M.P., Saito, K., Adimo, O., Adjei-Nsiah, S., Agali, A., Bala, A., Chikowo, R., Kaizzi, K., Kouressy, M., Makoi, J.H.J.R., Ouattara, K., Tesfaye, K., Cassman, K.G., 2016. Can sub-Saharan Africa feed itself? Proceedings of the National Academy of Sciences 113, 14964-14969.

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. European Journal of Soil Science 73, e13137.

Verstegen, J.A., Karssenberg, D., van der Hilst, F., Faaij, A., 2012. Spatio-temporal uncertainty in Spatial Decision Support Systems: A case study of changing land availability for bioenergy crops in Mozambique. Computers, Environment and Urban Systems 36, 30-42.

Wallach, D., Thorburn, P.J., 2017. Estimating uncertainty in crop model predictions: Current situation and future prospects. European Journal of Agronomy 88, A1-A7.

Wickham, H., 2016. Programming with ggplot2. In: Wickham, H. (Ed.), ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing, Cham, pp. 241-253.

Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77, 1 - 17.