Geoderma 425 (2022) 116018

Contents lists available at ScienceDirect

Geoderma

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

Spatial predictions of maize yields using QUEFTS – A comparison of methods

Mirjam S. Breure [a,*], Bas Kempen [b], Ellis Hoffland [a]

[a] Soil Biology Group, Wageningen University, Wageningen, The Netherlands
[b] ISRIC — World Soil Information, Wageningen, the Netherlands

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:
QUEFTS
Transfer functions
Yield
Yield-limiting nutrient
Random forest modelling
Soil maps

ABSTRACT

Using fertilisers is indispensable for closing yield gaps in Sub Saharan Africa. Current fertiliser recommendations, however, are often blanket recommendations which do not take spatial variation in soil conditions within a region or country into account. Soil maps can potentially support fertiliser recommendations at a higher spatial resolution. The QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model is a decision support tool that predicts crop yields as an indicator of soil fertility and can be used to evaluate yield responses to fertilisers. It was designed for field level output and runs on field-specific soil information. The aim of this study was to compare two methods for developing maps of QUEFTS output, i.e. maize yield and the yield-limiting nutrient, with Rwanda as a case study. We used a database containing soil analysis results of 999 samples collected across Rwanda. Transfer functions were applied to predict the required P-Olsen and Exchangeable K input for QUEFTS based on the soil data. For the "Calculate-then-Interpolate" (CI) method, transfer functions and QUEFTS were applied to point data, and the final output was then interpolated using random forest modelling. For the "Interpolate-then-Calculate" (IC) method, maps of the soil parameters were developed first, before applying calculations. Implications of the chosen method (i.e. CI or IC) on QUEFTS predictions on a national scale were evaluated using set-aside locations. Results showed low precision and accuracy of QUEFTS maize yield predictions across Rwanda. The CI method performed better in predicting QUEFTS yield and yield-limiting nutrient than the IC method. Correlations between mapped yield predictions and predictions on set-aside evaluation locations were similar for the CI ($r = 0.444$) and IC ($r = 0.439$) methods. The poorer performance of the IC method was mostly due to overestimation of yields, which was most likely caused by the effect of smoothing on the soil maps used as input for QUEFTS. We conclude that the CI method is the preferred method for spatial application of QUEFTS.

1. Introduction

Agricultural productivity in Sub Saharan Africa should increase in order to sustain its growing population. In African soils, often multiple nutrients are depleted and the use of fertilisers is indispensable for closing yield gaps (Giller et al., 2011; Ichami et al., 2019; Shehu et al., 2019). Current fertiliser recommendations are often blanket recommendations that do not take into account spatial heterogeneity of soils and other site-specific factors, leading to either a waste of resources or low productivity of land (Giller et al., 2011; Vanlauwe et al., 2015). Site-specific fertiliser recommendations based on soil testing have been shown to lead to increased revenues over blanket recommendations (Njoroge et al., 2015). However, high costs, limited access to soil testing services and difficulty in interpreting results (Chianu et al., 2012), as well as uncertainties associated with sampling and analytical procedures (Schut and Giller, 2020) complicate the use of soil testing to increase productivity on a large scale.

Soil maps in combination with information on crop response to nutrient availability can potentially be used to refine blanket fertiliser recommendations. The QUantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model is a simple yet versatile tool that can be used to predict yield and yield response to fertilisers, taking interactions among nitrogen (N), phosphorus (P) and potassium (K) into account (Janssen et al., 1990; Sattari et al., 2014). QUEFTS requires few input parameters: soil pH, soil organic carbon (SOC), P measured in an Olsen extract (P-Olsen), exchangeable K measured in an ammonium acetate extract (Exch-K) and crop-specific physiological efficiency parameters. It has been developed and validated for a wide range of crops and regions (e.g. Das

* Corresponding author at: P.O box 47, 6700 AA Wageningen, the Netherlands.
E-mail address: mirjam.breure@wur.nl (M.S. Breure).

https://doi.org/10.1016/j.geoderma.2022.116018
Received 30 November 2021; Received in revised form 9 May 2022; Accepted 20 June 2022
Available online 22 July 2022

et al., 2009; Ezui et al., 2017; Shehu et al., 2019; Tabi et al., 2008) and has proven adequate for developing fertiliser recommendations (e.g. Maiti et al., 2006; Mesfin et al., 2021; Xu et al., 2013). QUEFTS was designed for field level output and runs on field-specific soil information, but can be applied to soil maps as well (Leenaars et al., 2018b). In this study, different methods will be explored to arrive at maps of maize yield and the yield-limiting nutrient at a country level using QUEFTS.

There are two pathways to develop maps of a target variable: model calculations can be applied to soil maps or to the point data underpinning these maps. When models are applied to data points, the model output is calculated for each data point, followed by spatial interpolation of these points to develop a map of the target variable. When models are applied to soil maps, the point data underpinning the maps are interpolated before applying model calculations. Both methods were previously described as the calculate-then-model and model-then-calculate methods (Kempen et al., 2019; Orton et al., 2014) or calculate-then-interpolate and interpolate-then-calculate methods (Heuvelink and Pebesma, 1999). We will refer to both methods as the calculate-then-interpolate (CI) and interpolate-then-calculate (IC) methods. The IC method is computationally more intensive, but application of the CI method may be limited in case of missing point data for one or more input parameters, or when only maps are available as input data and not the underpinning data points (Orton et al., 2014).

The aim of this work is to evaluate the implications of the chosen method (i.e. CI or IC) on QUEFTS predictions on a national scale, using Rwanda as a case study. In our case, two consecutive steps of model calculations are needed: First, transfer functions developed by Breure et al. (2022) are used to calculate QUEFTS input parameters P-Olsen and Exch–K from available data. Second, QUEFTS model calculations are made. Minor differences between outcomes of the CI and IC methods have been reported for simple linear models (Kempen et al., 2019; Orton et al., 2014; Styc and Lagacherie, 2019), but substantial differences for non-linear models (Addiscott and Tuck, 1996). Based on Heuvelink and Pebesma (1999), we hypothesise that the IC method will result in the most accurate spatial predictions. As QUEFTS is a non-linear model involving several interacting calculation steps, we expect that results of the CI and IC methods will differ more compared to the work of Kempen et al. (2019), Orton et al. (2014) and Styc and Lagacherie (2019). The outcomes of this work will help to evaluate whether QUEFTS can be used to develop fertiliser recommendations at scale.

2. Materials & methods

2.1. Data availability

A geo-referenced dataset containing 999 topsoil (0–20 cm) samples with a good spatial coverage across Rwanda (Fig. 1) was provided by the International Fertiliser Development Center (IFDC). Soils were sampled in 2014 as part of the CATALIST-2 programme and analysed by the Crop Nutrition Laboratory Services, Nairobi, Kenya. The dataset contained pH-H_2O measured in a 1:2 soil:water suspension, organic matter content (%) determined with the Walkley-Black method and several nutrients measured in a Mehlich 3 extraction (mg kg^{-1}). SOC values were calculated from organic matter, assuming 50% of organic matter is carbon (Pribyl, 2010). The dataset was randomly split into a calibration (n = 699) and evaluation (n = 300) dataset. Part of the data were excluded based on selection criteria (Fig. 1; sections 2.3 and 2.4).

In addition, 134 covariate layers at 250 m spatial resolution were available for random forest modelling that we used for spatial interpolation (Section 2.5). These layers were previously prepared for and described in Kempen et al. (2015). Briefly, the covariate layers were acquired from six sources: the Africa Soil Information Service (AfSIS), ISRIC WorldGrids, USGS Africa Ecosystems Mapping database, Soil and Terrain database of Central Africa (SOTERCAF), soil and terrain database for north-eastern Africa (SOTERNEA) and the Multipurpose Africover Databases on Environmental Resources (MADE). Covariate layers contained information on soil type, terrain, climate, land cover, vegetation indices, primary production, spectral reflectance, albedo and soil moisture.

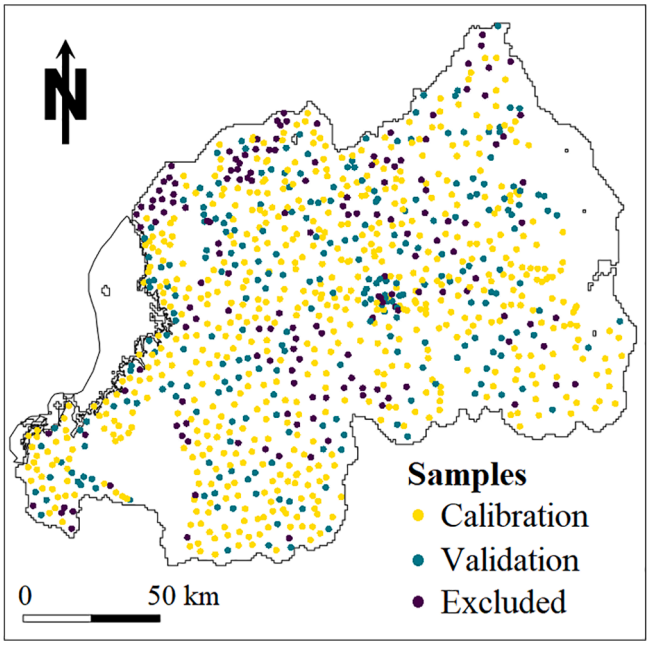

Fig. 1. Distribution of calibration, evaluation and excluded data points across Rwanda. Note that the white areas are lake Kivu (west) and national parks (south-west and east).

2.2. Workflow

Two methods were used to develop maps with QUEFTS for Rwanda, in a two-step approach (Fig. 2). In Step 1, QUEFTS input data were derived from available data. Besides pH-H_2O and SOC (in g kg^{-1}), QUEFTS requires P-Olsen (in mg kg^{-1}) and Exch-K (in mmol kg^{-1}) measured in ammonium acetate at pH 7.0, which were not available in the dataset and were therefore predicted using transfer functions (Section 2.3). In Step 2, QUEFTS was applied to compute yield and the most yield-limiting nutrient (considering N, P and K), based on pH, SOC, P–Olsen and Exch-K data (Section 2.4). QUEFTS was run for the situation that no fertilisers are applied.

For the CI method, Steps 1 and 2 were applied to the calibration dataset. The yield and most yield-limiting nutrient predictions at these locations were spatially interpolated using random forest modelling (Section 2.5). For the IC method, the input parameters of the transfer function and QUEFTS model were first interpolated (using the calibration dataset) to produce maps for each of these parameters. Subsequently, Steps 1 and 2 were applied to the soil maps to derive maps of maize yield and the yield-limiting nutrient.

2.3. Transfer functions for QUEFTS inputs

P-Olsen and Exch-K were estimated from available Mehlich 3 (M3) data using the transfer functions of Breure et al. (2022), as presented in Equations 1 and 2. Note that Eq. (1) was adjusted from Breure et al. (2022) to estimate Exch-K values in mmol kg^{-1} instead of mg kg^{-1}.

$$Exch-K\ \left(mmol\ kg^{-1}\right) = 0.028{*}K-M3\ \left(mg\ kg^{-1}\right) + 0.015 \quad (1)$$

$$\begin{aligned} ln(P-Olsen) = {} & 0.769{*}ln(P-M3) + 0.620{*}ln(Al-M3) \\ & +0.131{*}ln(Fe-M3) + 0.095{*}ln(Ca-M3) \\ & -0.191{*}pH - 4.307 \end{aligned} \quad (2)$$

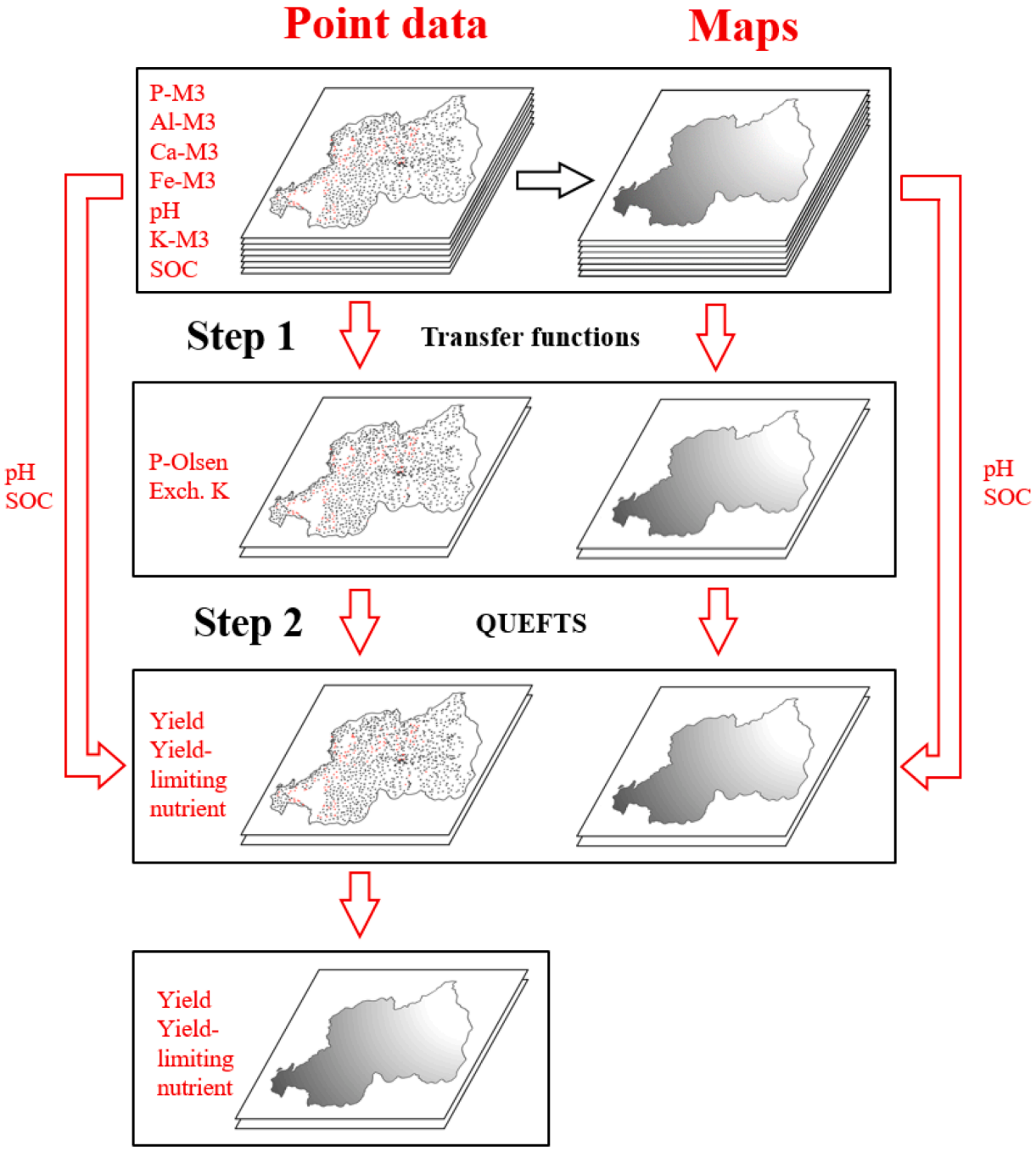

Fig. 2. Two methods of applying P and K transfer functions and QUEFTS to data points: Calculate-then-Interpolate (vertical pathway on the left) or Interpolate-then-Calculate (right). Black horizontal arrow indicates the process of spatial interpolation of data points to maps. Red arrows represent input that is used for either the P and K transfer functions (Step 1) or for QUEFTS (Step 2).

The nutrients in Mehlich 3 (P, Al, Ca and Fe; in mg kg^{-1}) were transformed to natural logarithms for application of the P transfer function (Equation 2). The predicted P-Olsen values were obtained by back-transformation of the ln(P-Olsen) predictions following Lark and Lapworth (2012):

$$\mathrm{P-Olsen} = \exp\left(\ln(\mathrm{P-Olsen}) + 0.5 * \sigma^2_{pred}\right) \quad (3)$$

where ln(P-Olsen) is the predicted value with the P transfer function and σ^2_{pred} denotes the prediction error variance. For the CI method, where the transfer function was applied to the data points, the prediction error variance was computed using Equation (4):

$$\sigma^2_{pred} = \sigma^2(\hat{y}) + \sigma^2_{\varepsilon} \quad (4)$$

where $\sigma^2(\hat{y})$ is the variance of the regression estimate $\hat{y}$ and σ^2_{ε} the residual variance (Hastie et al., 2009 Chapter 3, Eq. 3.22). The prediction error variance is specific for each calibration point, its magnitude depending on the values of the input parameters of the P transfer function. For the IC method the prediction error variance was approximated using a quantile regression forest (Equation (6); Section 2.5).

The transfer functions of Breure et al. (2022) were derived from data within certain ranges. It is currently unknown whether application of the transfer functions to data outside these ranges leads to reliable Exch-K and P–Olsen predictions. Limits were therefore applied to the data

used in this study to select only those data points that fell within the transfer function calibration range. All data points fell within the ranges of Breure et al. (2022) for Al-M3 and Al-Fe, though some data points exceeded the maximum calibration values for P-M3 (950 *vs* 94 mg kg^{-1}) and Ca-M3 (9800 *vs* 3283 mg kg^{-1}). Therefore, cut–off values of 150 mg kg^{-1} for P-M3 and 3500 mg kg^{-1} for Ca-M3 were applied, which were somewhat higher than the maximum values in the P transfer dataset, but minimised data exclusion. Although K-M3 values in the dataset extended beyond the range of the K transfer function data (2960 *vs* 710 mg kg^{-1}), no limit was applied, as relations between K-M3 and Exch-K are expected to be linear also for higher concentrations (Mamo et al., 1996). Finally, pH limits of 4.0 – 8.0 were also applied to the data.

For the CI method, applying these limits led to exclusion of 70 calibration and 31 evaluation data points. This left 629 samples as input for Step 2 in the workflow; a number of 269 samples remained for evaluation. For the IC method, limits were applied to the maps that were developed based on the calibration data. As a consequence of applying these limits, no P-Olsen predictions could be made for 2.2% of the total number of grid cells.

2.4. QUEFTS

The latest version of QUEFTS calibrated for maize was used here (Sattari et al., 2014). QUEFTS calculates nutrient-limited yield and does not account for other potentially yield-limiting factors, such as water availability or presence of pests and diseases. On a site level, these factors can be accounted for by reducing the maximum yield (Y_{max}) parameter. In this study however, Y_{max} was fixed at 10 Mg ha^{-1} across Rwanda. The yield-limiting nutrient was calculated using QUEFTS output and Eq. (5) (Heinen, pers. comm., 2020):

$$\mathrm{DFX} = \frac{U_X - U_{Xd}}{U_{Xa} - U_{Xd}} = \frac{U_X - \frac{Y}{d_X} - r_X}{\frac{Y}{a_X} - \frac{Y}{d_X}} \quad (5)$$

DF_X is the dilution factor of nutrient X (i.e. N, P or K), ranging between 0 and 1. U_X refers to uptake of nutrient X (kg ha^{-1}). U_{Xd} refers to the uptake of X in the case this nutrient is maximally diluted in the crop, which is calculated as yield (Y) divided by the maximum physiological efficiency parameter *d* for X, corrected with *r*. Parameter *r* refers to the minimum uptake of a nutrient needed to produce any yield. U_{Xa} refers to the uptake of X in the case this nutrient is maximally accumulated in the crop, which is calculated as yield (Y) divided by the minimum physiological efficiency parameter *a* for X, corrected with *r*. The crop-specific *a*, *d* and *r* values specified by Sattari et al. (2014) for maize were used. The principle behind Equation 5 is to estimate how much the shoot concentrations differ from the physiological minimum. The nutrient with the lowest dilution factor is considered to be yield-limiting.

A second set of limits was applied to the data, according to the limits set for application of QUEFTS (Janssen et al., 1990): SOC < 70 g kg^{-1}, P-Olsen < 30 mg kg^{-1} and Exch-K < 30 mmol kg^{-1}. For the CI method, applying these limits left 591 calibration and 251 evaluation data points. For spatial interpolation of the QUEFTS output, predictions were made on locations with values outside set limits. Locations of these grid cells were identified using the IC maps and values were set to NULL. As a consequence, for the IC method, together with previously applied limits to the P–Olsen and Exch-K maps, QUEFTS predictions could not be made for 2.3% of the total number of grid cells.

2.5. Random forest modelling

Gridded maps of the yield and the most yield-limiting nutrient for the CI method, and P-M3, Al-M3, Ca-M3, Fe-M3, K-M3, pH and SOC for the IC method were developed using random forest modelling (Breiman, 2001; Strobl et al., 2009). A model was fitted for each variable using the data points included in the calibration dataset and the 134 covariate layers that served as the explanatory variables.

It is considered good practice to remove redundant covariates (i.e. covariates with limited predictive power) prior to modelling, for instance using recursive feature elimination (RFE) (Hounkpatin et al., 2018; Poggio et al., 2021). In addition, correlated covariates should preferably be removed (Poggio et al., 2021). Redundant covariates were not removed in this study, as this would increase the computational load substantially without having direct benefits. First, removing redundant covariates can have computational advantages when data sets are large, which is not the case here. Second, when one aims to understand what drives the model it can be advantageous to remove redundant variables so that these do not confound predictive relationships. Eliciting and understanding predictive relationships, however, is outside the scope of this study. Last, Poggio et al. (2021) show that the effect of removing redundant covariates with RFE on model performance is only very marginal. We therefore do not expect relevant differences in this study in the performance of models fitted with the full covariate stack and models fitted with a reduced stack.

Subsequently, the fitted random forest models were applied to the stack of covariate layers to predict each target variable across Rwanda at 250 m spatial resolution. Each fitted random forest model was composed of 1000 trees. The number of randomly selected candidate covariates for splitting a node was set as the square root of the total number of covariates (default setting of the *mtry* argument of the ranger function in R). Model residuals were inspected for presence of spatial correlation (Oliver and Webster, 2015); variogram analysis, however, indicated residual kriging was not required.

Variables P-M3, Ca-M3 and K-M3 were transformed to natural logarithms before fitting random forest models, as these maps had better evaluation statistics than maps based on models fitted to variables on the original scale. Maps of the log-transformed variables were back-transformed using Eq. (3). Unlike kriging methods, the random forest model does not provide an estimate of the prediction error variance required for back-transformation of log-scale predictions to the original scale (Lark and Lapworth, 2012). The variance was therefore approximated using the values of the 0.05 ($Q_{0.05}$) and 0.95 ($Q_{0.95}$) quantile outputs of a quantile regression forest (Meinshausen, 2006), assuming the prediction error is normally distributed:

$$\mathrm{variance} = \left(\frac{Q0.95 - Q0.05}{1.645*2}\right)^2 \quad (6)$$

For the IC method, the P and Ca maps on log-scale were used directly as input for the P transfer function, which requires log-transformed input for the nutrients measured in M3.

2.6. Evaluation

To evaluate the maps, grid predictions at the set-aside evaluation locations were compared to the observed (predicted for P-Olsen, Exch-K, yield and yield-limiting nutrient) values. No grid predictions could be derived at 29 out of 251 evaluation locations, which were located outside the prediction area. The remaining 222 evaluation locations were used to compute a set of evaluation statistics, including the mean error (ME) which is a measure of prediction bias, the root mean squared error (RMSE) as a measure for prediction accuracy, and the Modelling Efficiency Coefficient (MEC; Janssen and Heuberger, 1995). The MEC is a unitless goodness-of-fit statistic that measures the deviation of the predicted values from the 1:1 line and allows comparison with similar

models from other studies. The MEC is a measure of how well the model performs compared to using the mean of the calibration dataset as a predictor (Equation (7)):

$$MEC = 1 - \frac{\sum (P_i - O_j)^2}{\sum (O_i - \overline{O})^2} \quad (7)$$

In which P refers to grid predictions extracted on evaluation locations i, Oi refers to values observed at evaluation locations i, and $\bar{O}$ to the mean value of the calibration data.

2.7. Software

All analyses for this study were done with the statistical software R (version 3.6.3; R core team, 2020). Plots were made using the spplot (sp package, version 1.4–2; Bivand et al., 2013; Pebesma and Bivand, 2005) and ggplot (ggplot2 package, version 3.3.2; Wickham, 2016) functions. For application of the P and K transfer functions and the random forest models, the generic predict function of the stats package (version 3.6.2), was used. Random forest models and quantile regression forests were fitted with the ranger package (version 0.12.1; Wright and Ziegler, 2017). Spatial data processing was done with the sp and raster (version 3.3–13; Hijmans, 2020) packages.

3. Results

3.1. Input point data

Data points that were excluded based on P-M3, Ca-M3, pH or QUEFTS limits, were located across Rwanda, but tended to be clustered in the north-west and centre (Fig. 1). Calibration and evaluation data represented the total dataset well in terms of spatial coverage. Data distributions of the soil parameters were similar between the calibration and evaluation datasets (Table 2); distributions were furthermore relatively similar to the total dataset (Table 1), except for Ca-M3 and P-M3, of which most descriptive values were lower than the original dataset because of the application of limits.

Table 1
Descriptive statistics of the soil parameters in the original IFDC dataset (n = 999).

	-	g kg^{-1}	in Mehlich 3 (mg kg^{-1})				
	pH	SOC	P	K	Al	Ca	Fe
Minimum	3.2	2.2	0.2	14	342	51	61
1st Q	4.9	16.8	5.4	83	913	398	125
Median	5.6	20.6	9.3	153	1140	792	177
Mean	5.6	21.1	29.6	226	1182	1192	201
3rd Q	6.2	25.2	22.1	267	1400	1480	260
Maximum	8.5	42.2	950	2960	2590	9800	764

3.2. Soil maps

For the IC method, the degree to which the individual soil parameters could be predicted from available covariates, varied substantially (Table 3). Variation in Ca-M3 and Fe-M3 was described best, with MEC values above 0.50, followed by pH (MEC = 0.45), Al-M3 (MEC = 0.39) and SOC (MEC = 0.37). Variation in K-M3 and P-M3 was described less well, with MEC values of 0.23 and 0.07 respectively. For most properties, ME values were small compared to the mean values in the calibration and evaluation sets (Table 2) and relatively small compared to the RMSEs. For some soil properties, such as Ca and pH, the ME was relatively large compared to the RMSE. RMSE values furthermore are relatively high compared to mean values in the calibration and evaluation sets (Table 2), especially for P, K and Ca, indicating predictions are associated with large uncertainty.

3.3. Application of transfer functions

For the CI method, the P and K transfer functions were applied to the calibration data. The median predictions were 7.8 mg kg^{-1} for P-Olsen and 4.0 mmol kg^{-1} for Exch-K (Table 5 and Fig. 4). For the IC method, P-Olsen and Exch. K maps were developed by applying the transfer functions to maps of the input soil properties. Evaluation of the P-Olsen map showed that 16% of the variation in the (modelled) P-Olsen observations was explained (MEC = 0.16; Table 4), compared to a MEC of 0.07 for P-M3 (Table 3), the main parameter in the P transfer function. P-Olsen predictions had a ME of −1.2 mg kg^{-1}, indicating that they were somewhat underestimated. P-Olsen predictions were furthermore associated with a high degree of uncertainty, as the RMSE of 11.1 mg kg^{-1} is higher than the mean value of P-Olsen predictions (9.8 mg kg^{-1}, Table 5). Exch-K predictions were a linear transformation of the K-M3 grid. Not surprisingly, the MEC value of Exch-K predictions (0.22; Table 4) was almost identical to that of K-M3 (0.23; Table 3). Exch-K predictions were slightly underpredicted with an ME of −0.07 mmol kg^{-1}. Similar to P-Olsen, the accuracy of Exch–K predictions was low, with RMSE being higher than the mean of predicted values (7.3 *vs* 5.9 mmol kg^{-1}).

Table 2
Descriptive statistics of Al, Ca, Fe, P and K in Mehlich 3, pH and SOC of the calibration (C; n = 591) and evaluation (E; n = 251) datasets.

	mg kg^{-1}										-		g kg^{-1}	
	Al		Ca		Fe		P		K		pH		SOC	
	C	E	C	E	C	E	C	E	C	E	C	E	C	E
Minimum	342	379	57	107	61	66	0.2	1.2	14	31	4.0	4.0	2.2	9.4
1st Q	939	875	389	360	123	128	5.0	5.0	80	79	4.9	4.9	16.5	16.5
Median	1160	1160	705	691	171	178	8.2	8.2	142	145	5.4	5.4	19.9	20.2
Mean	1190	1184	913	898	195	195	13.7	12.7	194	185	5.5	5.5	20.7	20.4
3rd Q	1410	1465	1275	1220	255	248	15.9	16.2	246	215	6.0	5.9	25.0	23.8
Maximum	2590	2390	3330	3490	533	449	105	66.2	1070	952	7.9	7.9	42.2	34.4

Table 3
Evaluation statistics for the maps developed with the interpolate-then-calculate (IC) method. Statistics for K, P and Ca are given on log-transformed and original scale, after back-transformation. K, P, Al, Ca and Fe in Mehlich 3 (in mg kg^{-1}).

	ln(K)	ln(P)	ln(Ca)	K	P	Al	Ca	Fe	pH	SOC
ME	0.03	0.07	0.07	1.4	0.26	−7.1	123	2.3	0.10	0.03
RMSE	0.75	1.04	0.73	257	71.6	324	927	72	0.70	4.7
MEC	0.27	0.24	0.47	0.23	0.07	0.39	0.51	0.53	0.45	0.37

Table 4
Evaluation statistics for the P-Olsen (mg kg^{-1}) and Exch-K (mmol kg^{-1}) maps developed with the interpolate-then-calculate (IC) method. Statistics for P-Olsen are given on log-transformed and original scale, after back-transformation.

	ln(P-Olsen)	P-Olsen	Exch.K
ME	0.15	−1.2	−0.07
RMSE	0.70	11.1	7.3
MEC	0.19	0.16	0.22

3.4. Application of QUEFTS

3.4.1. Data distribution

Spatial predictions show that pH values are lowest (between 4.2 and 5.2) in the west and highest (between 6.1 and 7.2) in the east of Rwanda (Fig. 3). Regions with the highest SOC values (between 25 and 38 mg kg^{-1}) are clustered in the south-west, north-west and south-east. For a large proportion of the surface area of Rwanda (51%), P-Olsen predictions are below the critical value of maize of 10 mg kg^{-1} (Bai et al., 2013; Ussiri et al., 1998), whereas for Exch-K, only a negligible part of the predictions are below the critical value of 2 mmol kg^{-1} (0.03%; Chilimba et al., 1999). Smoothing is visible for the gridded maps of each of the four QUEFTS input parameters: compared to the calibration data of the CI method, distributions have become narrower as the lowest and highest values are under-represented compared to mean values (Fig. 4). As a consequence of smoothing, median values of SOC (20.6 *vs* 19.9 g kg^{-1}), pH (5.53 *vs* 5.44), P–Olsen (8.8 *vs* 7.8 mg kg^{-1}) and Exch-K (5.0 *vs* 4.0 mmol kg^{-1}) were higher for the gridded maps than for the calibration data (Table 5). Similar to the QUEFTS input parameters, gridded maps of the QUEFTS yield showed smoothing and a higher median value than direct QUEFTS predictions on the calibration points (4.0 *vs* 3.4 Mg ha^{-1}).

Table 5
Data distribution of input for QUEFTS and predicted yield values at calibration locations (C; n = 591) and extracted from the grids at the calibration locations.

	SOC (g kg^{-1})		pH (−)		P-Olsen (mg kg^{-1})		Exch-K (mmol kg^{-1})		Yield (Mg ha^{-1})	
	C	Grid	C	Grid	C	Grid	C	Grid	C	Grid
Minimum	2.2	11.5	4.02	4.40	0.4	1.3	0.4	1.3	0.4	1.2
1st Q	16.5	18.1	4.91	5.10	5.2	6.7	2.3	3.7	2.3	3.1
Median	19.9	20.6	5.44	5.53	7.8	8.8	4.0	5.0	3.4	4.0
Mean	20.7	21.1	5.51	5.53	9.8	9.8	5.5	5.9	3.5	4.0
3rd Q	25.0	23.9	6.05	5.96	12.5	12.0	6.9	7.1	4.4	4.7
Maximum	42.2	37.7	7.86	7.01	29.8	23.8	29.9	20.8	7.8	7.4

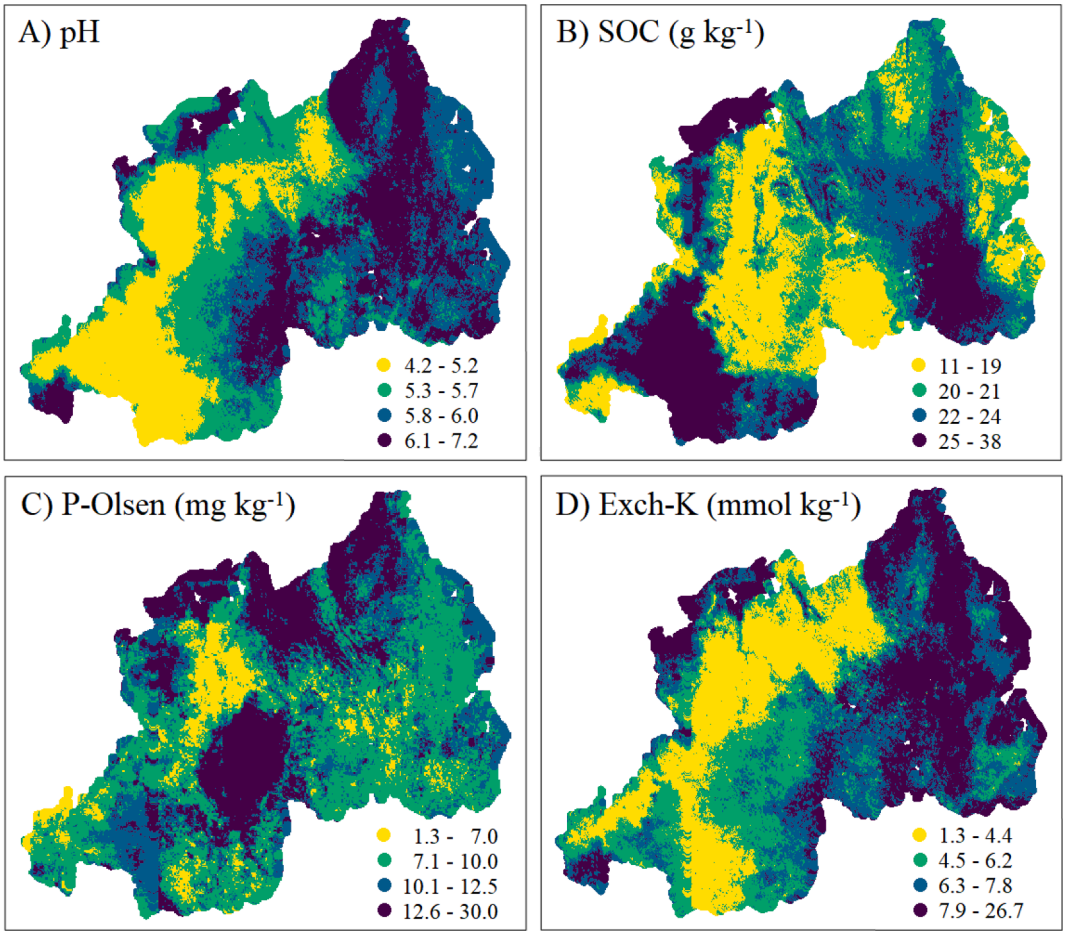

Fig. 3. Spatial plots of pH (A), SOC (B), P-Olsen (C) and Exch-K (D) inputs required by QUEFTS, for the interpolate-then-calculate (IC) method. Class boundaries are based on the quantile distribution of the grid values, except for the P-Olsen map, where boundaries were adjusted to clearly represent values below the critical limit for maize (10 mg kg^{-1}).

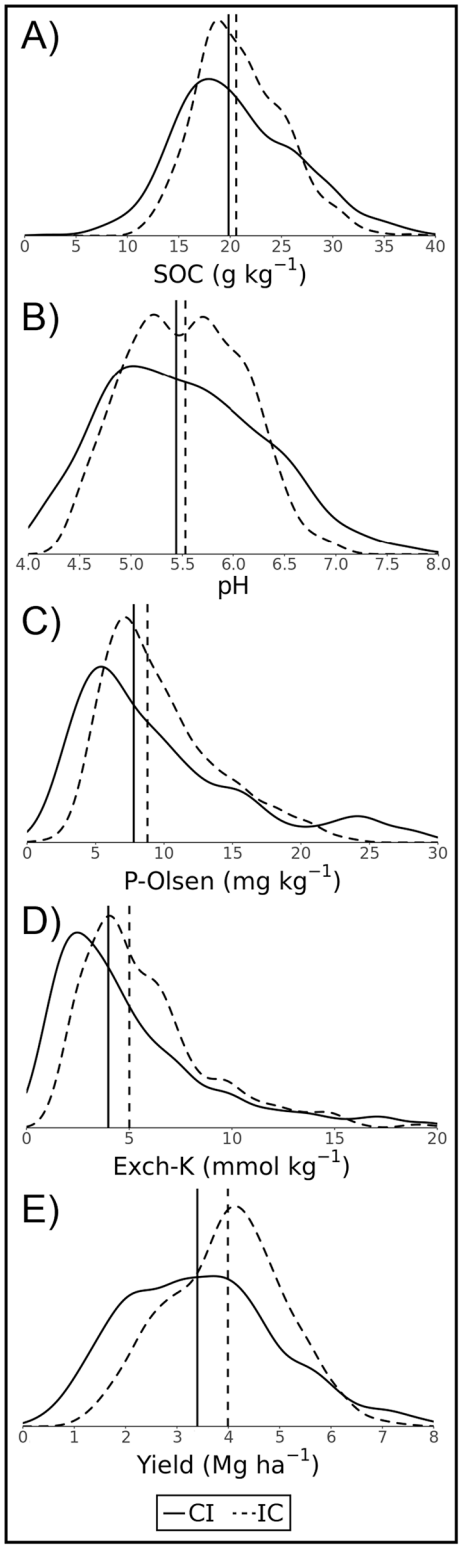

Fig. 4. Density plots of SOC (A), pH (B), P–Olsen (C) and Exch-K (D) inputs and yield (E) on calibration locations of the calculate-then-interpolate (CI) method, and extracted from the interpolate-then-calculate (IC) maps at calibration locations. Vertical lines represent median values.

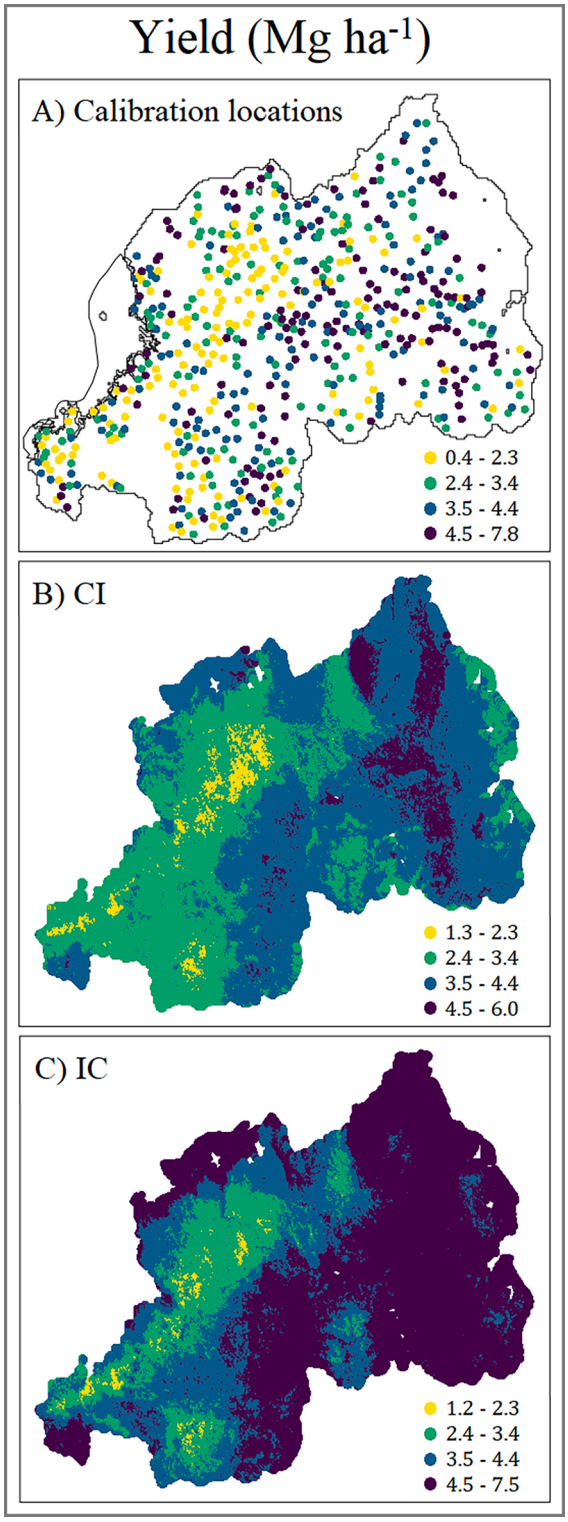

Fig. 5. Yield predictions for (A) the calibration data (n = 591) and for maps developed with (B) the calculate-then-interpolate (CI) and (C) interpolate-then-calculate (IC) methods. Categories are based on the quantile distribution of the calibration points. Note that distributions have different minimum and maximum values.

Table 6
Data distributions of the yield predictions on the point data (total dataset, calibration locations C, evaluation locations E), as well as the CI and IC predictions of the grids and on evaluation locations.

Yield (Mg ha^{-1})	Point data			CI		IC	
	Total	C	E	Grid	E	Grid	E
Minimum	0.4	0.4	0.8	1.3	1.8	1.2	2.1
1st Q	2.3	2.3	2.4	3.0	3.0	3.6	3.5
Median	3.4	3.4	3.3	3.6	3.5	4.5	4.3
Mean	3.5	3.5	3.5	3.5	3.5	4.3	4.2
3rd Q	4.4	4.4	4.3	4.0	4.0	5.0	4.9
Maximum	8.2	7.8	8.2	6.0	4.8	7.5	5.9
n	842	591	251	355,178	251	355,178	251

3.4.2. *Yield predictions*

Yield maps produced with the CI and IC methods show similar patterns, with lower yields in the west of Rwanda and higher yields in the east (Fig. 5). The low-yielding locations in the west are underrepresented in both maps however, compared to the yield predictions at the calibration locations, while for the IC method, the high-yielding locations in the east are overrepresented. The median yield prediction of the maps developed with the IC method was substantially higher than for maps developed with the CI method (4.5 *vs* 3.6 Mg ha^{-1}, respectively; Table 6). In addition, despite smoothing of each of the four QUEFTS input variables for the IC method (Fig. 4), the interquartile range in yield predictions was broader compared to the CI method (i.e. 3.0–4.0 *vs* 3.6–5.0 Mg ha^{-1}; Table 6).

The patterns of spatial yield predictions correspond strongly with the distribution of pH predictions across Rwanda, and to some extent with Exch-K predictions (Fig. 5 *vs* Fig. 3). The lowest yields (below 2.3 Mg ha^{-1}) are found in areas with low pH, and low P-Olsen and Exch-K values. The high yield predictions in the east of Rwanda furthermore correspond to high pH, SOC and Exch-K values.

The IC method overpredicted maize yields (Fig. 6), resulting in a ME of 0.8 Mg ha^{-1}, compared to a relatively small overestimation of 0.04 Mg ha^{-1} for the CI method. Evaluation statistics furthermore indicate that IC yield predictions are less accurate compared to results of the CI method (MEC = -0.03 *vs* 0.28 and RMSE = 1.54 *vs* 1.27 Mg ha^{-1} respectively). Despite the poor MEC and RMSE results for IC, which are mostly caused by overprediction of yields, correlation coefficients of both evaluation plots are similar, with 0.444 and 0.439 for the CI and IC methods, respectively (Fig. 6). This indicates that the IC and CI methods differentiate equally well between higher and lower yielding locations. In accordance with the data distribution of the grids, yield predictions by the IC method at the evaluation locations furthermore spanned a broader range than for those by the CI method (Table 6).

3.4.3. *Yield-limiting nutrient*

For both the CI and IC methods, P was predicted most often as the yield-limiting nutrient, for 76% and 65% of the grid cells, respectively (Fig. 7). These are higher percentages compared to the calibration data, for which P was predicted to be yield-limiting at 52% of the locations. N and K were yield-limiting at 22% and 26% of the calibration locations. The CI method underestimated both N and K as yield-limiting nutrient compared to the calibration data, as only 12% of the grid cells were predicted to be yield-limiting for each nutrient. The IC method overestimated N as yield-limiting (34%), but strongly underpredicted K as yield-limiting (2%).

Patterns of P-Olsen predictions strongly overlap with predictions of the yield-limiting nutrient (Fig. 7 *vs* Fig. 3). Generally speaking, P was identified as yield-limiting in regions with P-Olsen values below 12.5 mg kg^{-1}, irrespective of soil pH and SOC. In areas with P-Olsen values above 12.5 mg kg^{-1}, N tended to be yield-limiting. Although K was strongly underpredicted to be yield-limiting, areas with K as yield-limiting nutrient tended to have SOC values above 22 g kg^{-1}.

Evaluation of QUEFTS yield-limiting nutrient predictions showed that the CI method predicted the yield-limiting nutrient correctly at 133 out of 222 (60%) evaluation locations, whereas IC predicted the yield-limiting nutrient correctly at 120 out of 222 (54%) evaluation locations (Tables 7 and 8). The CI method predicted P as yield-limiting more accurately than the IC method (110 *vs* 87 out of 126 locations), whereas the opposite was true for N (17 *vs* 31 out of 56 locations). Both methods perform poorly when it comes to predicting K as the yield-limiting nutrient, with only 6 and 2 out of 40 cases predicted correctly for the CI and IC method, respectively.

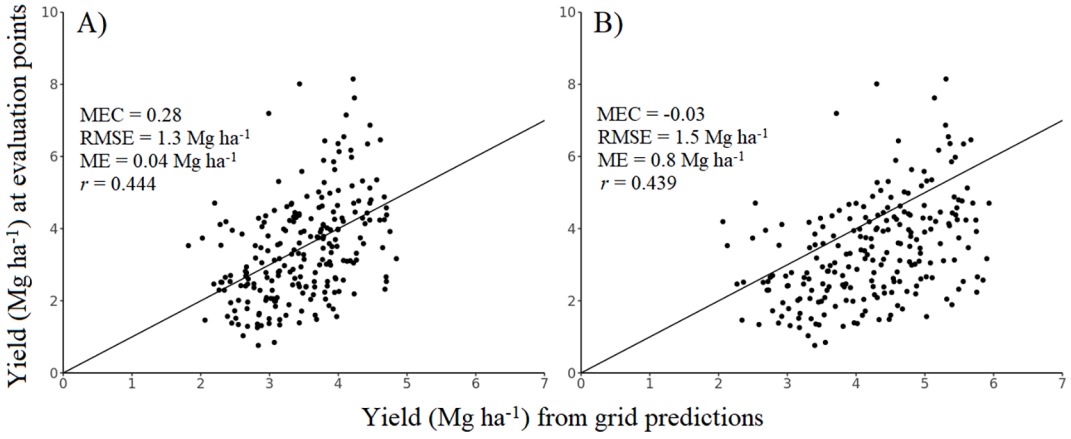

Fig. 6. Evaluation plots of yield predictions by the (A) calculate-then-interpolate (CI) and (B) interpolate-then-calculate (IC) methods. The x-axis represents grid predictions extracted on evaluation locations. Line represents the 1:1 line.

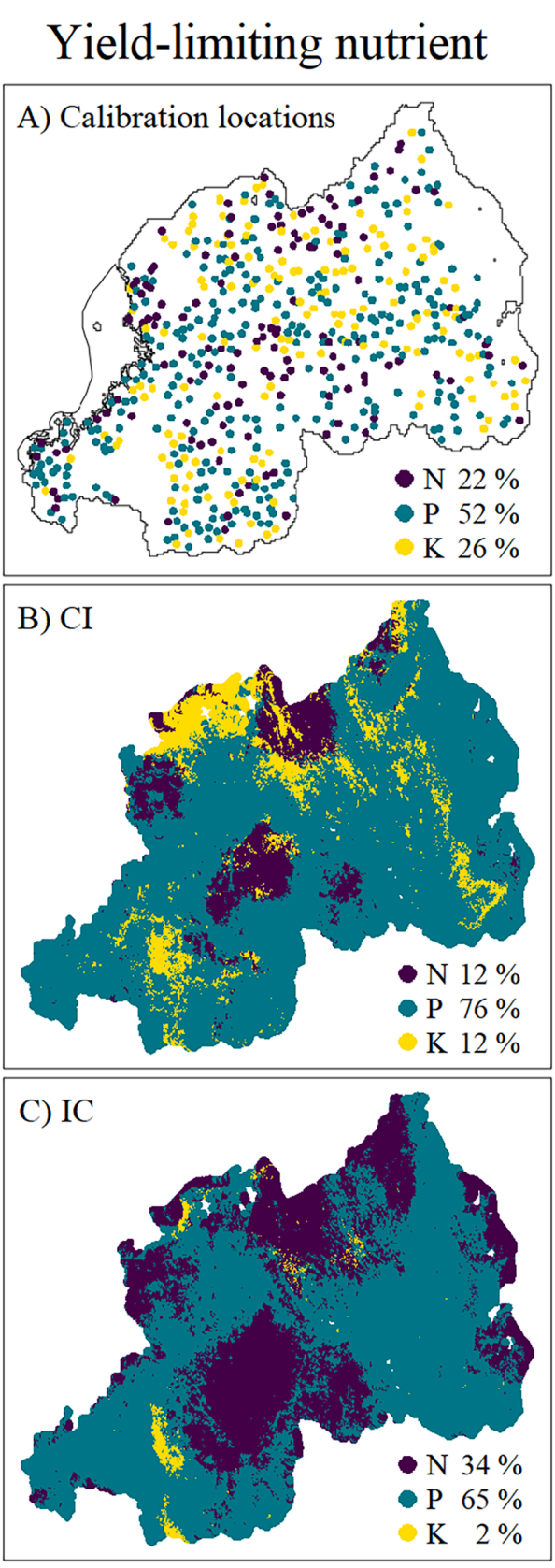

Fig. 7. Predicted yield-limiting nutrient for (A) the calibration locations (n = 591) and for maps developed with (B) the calculate-then-interpolate (CI) and (C) interpolate-then-calculate (IC) method. Percentages indicate the proportion of data being yield-limited by N, P or K.

Table 7
Confusion matrix of the observed versus predicted yield-limiting nutrient at the evaluation locations for the CI method.

Yield-limiting nutrient - CI	Predicted					
		N	**P**	**K**	**Total**	**Total**
Observed	**N**	**17**	32	7	56	25%
	P	6	**110**	10	126	57%
	K	3	32	**5**	40	18%
	Total	26	174	22	222	
	Total	12%	78%	10%		**59%**

Table 8
Confusion matrix of the observed versus predicted yield-limiting nutrient at the evaluation locations for the IC method.

Yield-limiting nutrient - IC	Predicted					
		N	**P**	**K**	**Total**	**Total**
Observed	**N**	**31**	25	0	56	25%
	P	38	**87**	1	126	57%
	K	11	27	**2**	40	18%
	Total	80	139	3	222	
	Total	36%	63%	1%		**54%**

4. Discussion

4.1. Comparing methods

This study showed three main findings: i) the choice of method has large effect on the outcomes: predicted yields and the yield-limiting nutrient were more accurate for the CI method than the IC method, as the latter method caused a large overprediction of yields across Rwanda, ii) the range in yield predictions was wider for the IC method than for the CI method, and iii) large errors were associated with spatial predictions of both methods.

Compared to Kempen et al. (2019) and Orton et al. (2014), differences between the CI and IC methods were more pronounced. This confirmed our hypothesis and is most likely related to the non-linear nature of QUEFTS (Heuvelink and Pebesma, 1999). In contrast to our findings, Heuvelink and Pebesma (1999) argued that the IC method will generally lead to more accurate results than the CI method. They attribute this to the optimal use of available information, as each input parameter of the IC method has a specific correlation structure with the covariates. The CI method does not take these individual correlation structures into account, as only the model output is interpolated. Although in this study the IC method performed worse than the CI method, due to an overprediction of yields, we hypothesise that the broader range in yield predictions can be attributed to the more optimal use of covariate data.

We hypothesise that the CI method performed better than the IC method, because of the effects of smoothing and applications of limits (Sections 4.1.1 and 4.1.2). However, although spatial yield predictions of the CI method were better compared to the IC method, RMSE still was 1.27 Mg ha^{-1}. This level of error corresponds to variability of maize yields within smallholder farms (Tittonell et al., 2007; Vanlauwe et al., 2006). This raises the question whether spatial application of QUEFTS, can be used to develop fertiliser recommendations on a national scale.

4.1.1. Smoothing

A striking difference between both methods was that IC yields were overestimated compared to CI yield predictions (Table 6). We believe this is caused by smoothing: data distributions of predictions (i.e. the soil maps) were more narrow than the datasets on which the predictions were based. For IC, smoothing as a result of spatial interpolation occurred early on in the workflow (Fig. 2) and for each of the seven soil parameters used as input for the transfer functions and QUEFTS. For the

CI method, on the other hand, smoothing occurred only once, at the end when the final yield and yield-limiting nutrient predictions at the calibration locations were interpolated.

Smoothing led to a smaller range in predictions for the QUEFTS input parameters, but also to higher median values for P–Olsen (7.8 *vs* 8.8 mg kg^{-1}) and Exch-K (4.0 *vs* 5.0 mmol kg^{-1}; Table 5). Although smoothing had a limited effect on median pH and SOC values, predictions were relatively more centred around the median after spatial interpolation. Low pH values (<4.7) are unfavourable for N availability, high pH values (greater than6.8) are unfavourable for K availability and P availability is suboptimal below pH 6 and above pH 6.7 (Sattari et al., 2014). After interpolation, the data distribution had become centred around pH 5.0–6.5, which could be considered the optimal range for availability of each of the three nutrients. In combination with the overestimation of the lowest P–Olsen and Exch-K values, we hypothesise that smoothing of pH predictions is the main reason that yields were overpredicted for the IC method compared to the CI method.

The way and extent to which smoothing affects spatial predictions, depends on the predictive power of the covariates. In regression, variation in the data is explained as the sum of the variation among the regression estimates and the residual variation (Snedecor and Cochran, 1989). In case covariate data explain the variation of a soil parameter completely, the residual variance equals zero and the variance of the predictions thus equals the variance of the data, hence no smoothing will occur. In other words, the lower the degree of variation in the target variable that can be explained by the covariates (i.e. MEC), the higher the degree of smoothing. The MEC value was 0.37 for SOC, 0.45 for pH, 0.16 for P-Olsen and 0.22 for Exch-K, while the MEC of the QUEFTS yield predictions with the CI method was 0.28. Variation in QUEFTS yield is thus explained less well by the covariates than variation in SOC and pH, but better than variation in P–Olsen and Exch-K. As SOC and pH play a prominent role in QUEFTS yield predictions, we hypothesise that the relatively good predictability of these soil characteristics is the reason for the wider range in yield predictions for the IC method.

From an agronomic perspective, predicting low P-Olsen and Exch-K values accurately is more relevant than predicting high values accurately. Although depending on other soil and agro-ecological properties that control crop yield, the critical P-Olsen concentration below which P is expected to be yield-limiting for maize, is 10 mg kg^{-1} (Bai et al., 2013; Ussiri et al., 1998). The Exch K concentration below which K is expected to be yield-limiting is 2 mmol kg^{-1} (Chilimba et al., 1999). Smoothing thus led to an overprediction of values in the agronomically relevant range. As a consequence, QUEFTS P and K fertiliser recommendations in those regions would be inadequate for sustainably increasing yields.

Smoothing can be avoided by using stochastic simulation (Heuvelink and Pebesma, 1999). Such approach, however, requires a geostatistical modelling framework instead of machine learning and was beyond the scope of the current study.

4.1.2. Data limits

Application of data limits for use of the transfer functions and QUEFTS increased differences between outcomes of both methods. For the CI method, application of limits led to the exclusion of 108 out of 699 data points for the calibration set (591 points remaining). For the IC method, maps were developed based on all 699 calibration data points and limits were applied to the maps. Application of limits mostly led to exclusion of data points with high values. Basing maps based on the complete calibration set thus resulted in higher yield predictions for the IC method. Developing soil maps based on the subset of 591 calibration points, resulted in a lower overprediction of yields for the IC method, although the average yield still exceeded that of the CI method (results not presented). The yield predictions by the IC method furthermore still spanned a broader range than the yields predicted by the CI method, indicating that it is smoothing rather than application of limits that caused the relatively narrow range in CI yields.

4.2. Limitations

4.2.1. Evaluation

To evaluate the extent to which QUEFTS spatial predictions correspond to reality, external evaluation of the yield maps developed in this study is indispensable. To this end, a dataset with unfertilised maize yields from a substantial number of replicated, geo-referenced field trials is needed. Trial locations should ideally be spread across Rwanda, covering a wide range in values for each of the relevant soil properties. Such dataset for Rwanda was not available to us and compiling an evaluation dataset from literature is complicated by e.g. different study designs or missing information. As more data are becoming (publicly) available, evaluation of QUEFTS spatial predictions may be possible in the future.

4.2.2. Sources of uncertainty

In the process of developing maps with QUEFTS, several sources of uncertainty can be identified. Although error propagation analysis was considered outside the scope of this study, the different sources of uncertainty will be discussed here.

Firstly, soil analysis data contain measurement errors, which can be attributed to e.g. varying measurement conditions, methods and measuring instruments (Van Leeuwen et al., 2021). As a consequence, lab results can differ among laboratories. Uncertainty associated with the Mehlich 3 extraction method is of considerable importance for this study, as five of the seven soil properties (P, Al, Ca, Fe and K) used as input for QUEFTS were determined with this analytical procedure. As part of a different study, a small subset ($n = 19$) of the IFDC samples was analysed at the CBLB laboratory of Wageningen University for M3 extractable nutrients using the same protocol as Breure et al. (2022). Mehlich 3 measurements by CropNuts were roughly 5% (P), 23% (Al), 10% (Ca) and 17% (Fe) higher compared to CBLB results. For K, no significant differences were found. Application of the P transfer function to the IFDC data led to an 19% overestimation of P-Olsen predictions compared to using CBLB measurements. This confirms that inter-laboratory variability for a given method, when used as input for models, can have a substantial effect on model output, as also shown by Schut and Giller (2020).

Secondly, the accuracy of random forest predictions were modest to poor for most soil properties, as RMSE values were relatively high compared to mean values (Tables 2 and 3). Available covariate data generally explained variation in soil properties only to a limited extent. In line with Hengl et al. (2017a), variation in P–M3 was described poorly compared to other nutrients. Reasons may include historic management which can affect P availability (Njoroge et al., 2019), P fixation in tropical soils (de Campos et al., 2018) and other soil properties which affect P extractability, such as pH (e.g. Penn et al., 2018). As P-M3 is the most relevant parameter in the P transfer function, P-Olsen predictions were not modelled well (MEC = 0.16). Exch-K predictions were only slightly better (MEC = 0.22). Although variation in pH (MEC = 0.45) and SOC (MEC = 0.37) was explained better based on covariate data, the combination of uncertainty associated with each of the QUEFTS input parameters culminated in a large error in yield predictions (RMSE of 1.54 Mg ha^{-1}; Fig. 5).

Thirdly, the use of transfer functions is necessary in case required input data are not available, but introduce additional uncertainty (Heuvelink and Pebesma, 1999). The uncertainty associated with the P and K transfer functions used in this study, can cause deviations in QUEFTS yield of more than 10%, compared to using measured P-Olsen and Exch-K values (Breure et al., 2022). Finally, QUEFTS yield predictions contain uncertainty; when observed maize yields were compared to QUEFTS predictions, MEC values of 0.84 and 0.67 were reported by Tabi et al. (2008) and by Shehu et al. (2019) after partial reparameterization of QUEFTS.

4.2.3. Spatial-temporal heterogeneity

Agronomic practices such as fertiliser application, liming and organic matter management impact soil nutrient availability, pH and SOC and can differ between farmers depending on socio-economic status, as well as within a farm depending on distance from the homestead (Chikowo et al., 2014; Zingore et al., 2007). The work of Njoroge et al. (2019) furthermore has shown that historic management can impact yields even seven growing seasons after changing practices. These differences in management practices, that typically vary at short spatial scales, are difficult to capture with (spatially exhaustive) environmental covariates. Hence it is very challenging to capture the effect of management on soil conditions in digital soil mapping models. Lack of soil management related information contributes to the poor performance of the prediction models for the soil nutrients. Soil maps developed at regional or national level may therefore lack the precision to describe short-distance spatial heterogeneity (Vanlauwe et al., 2015) and one should be careful to use these to derive information at field level. Maps furthermore are not able to capture temporal change and maps of soil parameters that are subject to short-term change should be used and interpreted with caution (Hengl et al., 2017b).

4.2.4. Data transformation

Adding half the prediction variance is required to ensure an unbiased estimation of the mean (Lark and Lapworth, 2012) when back-transforming a log-transformed variable. In some situations, existing soil maps are available, but not the underpinning data points. If these maps were created by back-transformation of log-scale predictions, variables would be log-transformed a second time in the process of applying the P transfer function. As a consequence, distributions of these log-scale predictions will have a higher mean than the initial log-scale predictions. In this study, log-scale predictions of P-M3 and Ca-M3 were used directly as input for the P transfer function. When using log-transformed P-M3 and Ca-M3 grid predictions on the original scale (after back-transformation) as input for the P transfer function, the median P-Olsen predictions increased from 9.9 to 16.8 mg kg^{-1}. The critical P-Olsen range, below 10 mg kg^{-1} for maize is thus strongly under-represented as a result of this 'transformation error' and may potentially have a large effect on fertiliser recommendations. The magnitude of this error depends on which and how many nutrients were log-transformed twice. As P–M3 is the most relevant parameter in the P transfer function (Breure et al., 2022), it had a substantial effect. When using maps developed with unknown methods as input for the P transfer function, or any model that requires log-transformed parameters as input, one should take into account that the model output may be substantially overestimated as a result of the transformation error.

4.3. Opportunities

Several developments may improve QUEFTS spatial predictions of yield and most yield-limiting nutrient. The large error associated with spatial yield predictions can mainly be attributed to the low predictive power of the covariates to explain variation in the soil and (for this study modelled) yield observations. As high (50–100 m) resolution covariate layers are becoming more and more available, spatial prediction models for soil and agronomic variables are likely to be improved (Hengl et al., 2017b; Poggio et al., 2021).

QUEFTS spatial predictions can also improve strongly in case P-Olsen measurement data become available. Variation in P–Olsen predictions was described least well of the four QUEFTS input parameters. This can partly be attributed to the use of a transfer function rather than measurements, and due to the uncertainty associated with each of the input maps for the P transfer function. Alternatively, Shehu et al. (2019) reparametrized QUEFTS using P-M3 instead of P–Olsen, which enables (spatial) application of QUEFTS without the use of the P transfer function.

QUEFTS spatial predictions can be adjusted to local conditions by using covariate layers directly as input. In this study, the QUEFTS maximum yield (Y_{max}) was assumed to be constant at 10 Mg ha^{-1}, but is likely to vary strongly in practice depending on local growing conditions. Covariate layers containing information on e.g. rooting depth, water availability or expected yield gap, currently not accounted for in this study, could be used to adjust the Y_{max} parameter to local conditions, thereby improving the accuracy of QUEFTS predictions (Leenaars et al., 2018a; Leenaars et al., 2018b; Steinbuch et al., 2016).

5. Conclusions

Spatial predictions of QUEFTS yield and yield-limiting nutrient were more accurate for the CI method than the IC method. The IC method overestimated yields, caused by the effect of smoothing on the distributions of the QUEFTS input (soil) parameters. Based on the results of this study, the CI method would thus be the preferred method for spatial application of QUEFTS. However, although the CI method performed better than the IC method, yield predictions were associated with large errors in our case study. This indicates that QUEFTS should be applied spatially with caution.

The large error of QUEFTS spatial predictions is caused by the low predictive power of the covariates to explain variation in the QUEFTS input parameters. When higher quality input soil maps become available, spatial application of QUEFTS could provide a low-cost, science-based alternative to national blanket recommendations. Evaluation with independent geo-referenced field data of measured yields remains necessary however, to gain insights in the current performance, as well as opportunities for improvement of QUEFTS spatial predictions.

Funding

This study was funded by NWO-AES Open Technology Programme, grant number 14688, "Micronutrients for better yields".

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We want to acknowledge the contribution of IFDC Rwanda, who provided the soil database on which this study was based. We also want to thank Job de Pater for his explorative work on soil mapping with QUEFTS. Many thanks to Marius Heinen for providing valuable input for determining the yield-limiting nutrient using QUEFTS. Finally, we want to thank Prof. Gerard Heuvelink and Johan Leenaars for providing valuable feedback on this work.

References

Addiscott, T.M., Tuck, G., 1996. Sensitivity analysis for regional-scale solute transport modeling, in: Corwin, D.L., Loague, K. (Eds.), Applications of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone. SSSA Special Publication Number 48. Soil Science Society Association, Madison, Madison, pp. 153–162.

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 Soil 372, 27–37. https://doi.org/10.1007/s11104-013-1696-y.

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied spatial data analysis with R, Second. ed. Springer, New York.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1201/9780429469275-8.

Breure, M.S., Van Eynde, E., Kempen, B., Comans, R.N.J., Hoffland, E., 2022. Transfer functions for phosphorus and potassium soil tests and implications for the QUEFTS model. Geoderma 406, 115458. https://doi.org/10.1016/j.geoderma.2021.115458.

Chianu, J.N., Chianu, J.N., Mairura, F., 2012. Mineral fertilizers in the farming systems of sub-Saharan Africa. A review. Agron. Sustain. Dev. 32, 545–566. https://doi.org/10.1007/s13593-011-0050-0.

Chikowo, R., Zingore, S., Snapp, S., Johnston, A., 2014. Farm typologies, soil fertility variability and nutrient management in smallholder farming in Sub-Saharan Africa. Nutr. Cycl. Agroecosyst. 100, 1–18. https://doi.org/10.1007/s10705-014-9632-y.
Chilimba, A.D.C., Mughogho, S.K., Wendt, J., 1999. Mehlich 3 or Modified Olsen for soil testing in Malawi. Commun. Soil Sci. Plant Anal. 30, 1231–1250. https://doi.org/10.1080/00103629909370280.
Das, D.K., Maiti, D., Pathak, H., 2009. Site-specific nutrient management in rice in Eastern India using a modeling approach. Nutr. Cycl. Agroecosyst. 83, 85–94. https://doi.org/10.1007/s10705-008-9202-2.
de Campos, M., Antonangelo, J.A., van der Zee, S.E.A.T.M., Ferracciú Alleoni, L.R., 2018. Degree of phosphate saturation in highly weathered tropical soils. Agric. Water Manag. 206, 135–146. https://doi.org/10.1016/j.agwat.2018.05.001.
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 Soil 420, 331–347. https://doi.org/10.1007/s11104-017-3387-6.
Giller, K.E., Tittonell, P., Rufino, M.C., van Wijk, M.T., Zingore, S., Mapfumo, P., Adjei-Nsiah, S., Herrero, M., Chikowo, R., Corbeels, M., Rowe, E.C., Baijukya, F., Mwijage, A., Smith, J., Yeboah, E., van der Burg, W.J., Sanogo, O.M., Misiko, M., de Ridder, N., Karanja, S., Kaizzi, C., K'ungu, J., Mwale, M., Nwaga, D., Pacini, C., Vanlauwe, B., 2011. Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. Agric. Syst. 104, 191–203. https://doi.org/10.1016/j.agsy.2010.07.002.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Second ed. Springer-Verlag New York. 10.1007/978-0-387-84858-7.
Heinen, M., 2020. Wageningen Environmental Research.
Hengl, T., Leenaars, J.G.B., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B.M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I., Kwabena, N.A., 2017a. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutr. Cycl. Agroecosyst. 109, 77–102. https://doi.org/10.1007/s10705-017-9870-x.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017b. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, e0169748.
Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modelling. Geoderma 89, 47–65. https://doi.org/10.1016/S0016-7061(98)00077-9.
Hijmans, R.J., 2020. Raster: Geographic Data Analysis and Modeling. R package version 3.3-13.
Hounkpatin, K.O.L., Schmidt, K., Stumpf, F., Forkuor, G., Behrens, T., Scholten, T., Amelung, W., Welp, G., 2018. Predicting reference soil groups using legacy data: A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). Sci. Rep. 8, 9959. https://doi.org/10.1038/s41598-018-28244-w.
Ichami, S.M., Shepherd, K.D., Sila, A.M., Stoorvogel, J.J., Hoffland, E., 2019. Fertilizer response and nitrogen use efficiency in African smallholder maize farms. Nutr. Cycl. Agroecosyst. 1–19. https://doi.org/10.1007/s10705-018-9958-y.
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. https://doi.org/10.1016/0016-7061(90)90021-Z.
Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecol. Modell. 83, 55–66. https://doi.org/10.1016/0304-3800(95)00084-9.
Kempen, B., Dalsgaard, S., Kaaya, A.K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., Walsh, M.G., 2019. Mapping topsoil organic carbon concentrations and stocks for Tanzania. Geoderma 337, 164–180. https://doi.org/10.1016/j.geoderma.2018.09.011.
Kempen, B., Vereijken, P., Keizer, P., González, M.R., Bindraban, P., Wendt, J., 2015. Preliminary evaluation of the feasibility of using geospatial information to refine soil fertility recommendations. VFRC Report 2015/6. Virtual Fertilizer Research Center, Washington, D.C.
Lark, R.M., Lapworth, D.J., 2012. Quality measures for soil surveys by lognormal kriging. Geoderma 173–174, 231–240. https://doi.org/10.1016/j.geoderma.2011.12.008.
Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez González, M., van Bussel, L.G.J., Guilpart, N., Yang, H., Cassman, K.G., 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324, 18–36. https://doi.org/10.1016/j.geoderma.2018.02.046.
Leenaars, J.G.B., Gonzalez, M.R., Kempen, B., 2018b. Extrapolation of fertilizer nutrient recommendations for major food crops in West Africa. Project report (draft) for IFDC, USAID - West Africa Fertilizer Program, Accra. ISRIC - World Soil Information, Wageningen, the Netherlands.
Maiti, D., Das, D., Pathak, H., 2006. Simulation of fertilizer requirement for irrigated wheat in eastern India using the QUEFTS model. Arch. Agron. Soil Sci. 52, 403–418. https://doi.org/10.1080/03650340600768706.
Mamo, T., Richter, C., Heiligtag, B., 1996. Comparison of extractants for the determination of available phosphorus, potassium, calcium, magnesium and sodium in some Ethiopian and German soils. Commun. Soil Sci. Plant Anal. 27, 2197–2212. https://doi.org/10.1080/00103629609369697.
Meinshausen, N., 2006. Quantile Regression Forests. J. Mach. Learn. Res. 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.
Njoroge, R., Birech, R., Arusey, C., Korir, M., Mutisya, C., Scholz, R.W., 2015. Transdisciplinary processes of developing, applying, and evaluating a method for improving smallholder farmers' access to (phosphorus) fertilizers: the SMAP method. Sustain. Sci. 10, 601–619. https://doi.org/10.1007/s11625-015-0333-5.
Njoroge, S., Schut, A.G.T., Giller, K.E., Zingore, S., 2019. Learning from the soil's memory: Tailoring of fertilizer application based on past manure applications increases fertilizer use efficiency and crop productivity on Kenyan smallholder farms. Eur. J. Agron. 105, 52–61. https://doi.org/10.1016/j.eja.2019.02.006.
Oliver, A.M., Webster, R., 2015. Basic steps in geostatistics: The variogram and kriging. SpringerBriefs in Agriculture. https://doi.org/10.1007/978-3-319-15865-5.
Orton, T.G., Pringle, M.J., Page, K.L., Dalal, R.C., Bishop, T.F.A., 2014. Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. Geoderma 230–231, 119–130. https://doi.org/10.1016/j.geoderma.2014.04.016.
Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. R News 5, 9–13.
Penn, C.J., Rutter, E.B., Arnall, D.B., Camberato, J., Williams, M., Watkins, P., 2018. A discussion on Mehlich-3 phosphorus extraction from the perspective of governing chemical reactions and phases: Impact of soil pH. Agriculture 8, 106. https://doi.org/10.3390/agriculture8070106.
Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240. https://doi.org/10.5194/soil-7-217-2021.
Pribyl, D.W., 2010. A critical review of the conventional SOC to SOM conversion factor. Geoderma 156, 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003.
R core team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna https://www.R-project.org.
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 Res. 157, 35–46. https://doi.org/10.1016/j.fcr.2013.12.005.
Schut, A.G.T., Giller, K.E., 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. Geoderma 380, 114680. https://doi.org/10.1016/j.geoderma.2020.114680.
Shehu, B.M., Lawan, B.A., Jibrin, J.M., Kamara, A.Y., Mohammed, I.B., Rurinda, J., Zingore, S., Craufurd, P., Vanlauwe, B., Adam, A.M., Merckx, R., 2019. Balanced nutrient requirements for maize in the Northern Nigerian Savanna: Parameterization and validation of QUEFTS model. Field Crops Res. 241, 107585. https://doi.org/10.1016/j.fcr.2019.107585.
Snedecor, G.W., Cochran, W.G., 1989. Statistical methods, 8th ed. Iowa State University Press.
Steinbuch, L., Brus, D.J., van Bussel, L.G., Heuvelink, G.B., 2016. Geostatistical interpolation and aggregation of crop growth model outputs. Eur. J. Agron. 77, 111–121. https://doi.org/10.1016/j.eja.2016.03.007.
Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychol. Methods 14, 323–348. https://doi.org/10.1037/a0016973.
Styc, Q., Lagacherie, P., 2019. What is the best inference trajectory for mapping soil functions: An example of mapping soil available water capacity over Languedoc Roussillon (France). Soil Syst. 3, 34. https://doi.org/10.3390/soilsystems3020034.
Tabi, F.O., Diels, J., Ogunkunle, A.O., Iwuafor, E.N.O., Vanlauwe, B., Sanginga, N., 2008. Potential nutrient supply, nutrient utilization efficiencies, fertilizer recovery rates and maize yield in northern Nigeria. Nutr. Cycl. Agroecosyst. 80, 161–172. https://doi.org/10.1007/s10705-007-9129-z.
Tittonell, P., Vanlauwe, B., de Ridder, N., Giller, K.E., 2007. Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil fertility gradients or management intensity gradients? Agric. Syst. 94, 376–390. https://doi.org/10.1016/j.agsy.2006.10.012.
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. Commun. Soil Sci. Plant Anal. 29, 2801–2813. https://doi.org/10.1080/00103629809370155.
Van Leeuwen, C.C.E., Mulder, V.L., Batjes, N.H., Heuvelink, G.B.M., 2021. Statistical modelling of measurement error in wet chemistry soil data. Eur. J. Soil Sci. 1–17 https://doi.org/10.1111/ejss.13137.
Vanlauwe, B., Descheemaeker, K., Giller, K.E., Huising, J., Merckx, R., Nziguheba, G., Wendt, J., Zingore, S., 2015. Integrated soil fertility management in sub-Saharan Africa: Unravelling local adaptation. Soil 1, 491–508. https://doi.org/10.5194/soil-1-491-2015.
Vanlauwe, B., Tittonell, P., Mukalama, J., 2006. Within-farm soil fertility gradients affect response of maize to fertiliser application in western Kenya. Nutr. Cycl. Agroecosyst. 76, 171–182. https://doi.org/10.1007/s10705-005-8314-1.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Wright, M.N., Ziegler, A., 2017. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77. 10.18637/jss.v077.i01.
Xu, X., He, P., Pampolino, M.F., Chuan, L., Johnston, A.M., Qiu, S., Zhao, S., Zhou, W., 2013. Nutrient requirements for maize in China based on QUEFTS analysis. Field Crops Res. 150, 115–125. https://doi.org/10.1016/j.fcr.2013.06.006.
Zingore, S., Murwira, H.K., Delve, R.J., Giller, K.E., 2007. Influence of nutrient management strategies on variability of soil fertility, crop yields and nutrient balances on smallholder farms in Zimbabwe. Agric. Ecosyst. Environ. 119, 112–126. https://doi.org/10.1016/j.agee.2006.06.019.