Geospatial modelling of soil geochemistry at national-scale for improved human nutrition

Christopher Chagumaira

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

School of Biosciences

October 2022

Abstract

Mineral micronutrient deficiencies (MND), prevalent in Ethiopia and Malawi among most countries in sub-Saharan Africa, are linked to soil type. Dietary mineral intake is influenced by mineral content of edible portions of crops, and there is strong evidence that cereals grown in these regions have low uptake of micronutrients. The low nutrient uptake is attributed to soil conditions. Spatial information on soil and crop properties is therefore required to improve local estimates of MND risk in order to implement targeted and efficient interventions. Obtaining spatial information on soil micronutrients status and other relevant properties that affect their uptake requires substantial effort, and there are uncertainties in the resulting information, which depend, in part on the methods used for prediction and the sampling design. Therefore, it is necessary to use robust and efficient methods for spatial prediction which characterise the uncertainty of the predictions reliably. Furthermore, it is necessary that these uncertainties can be communicated effectively to stakeholder groups so that they can account for them at all stages from commissioning the survey through to making decisions based on the information.

In this study, it was important first to understand how uncertain spatial information can be communicated to stakeholders (e.g., those in public health or nutrition and agronomy or soil science) through a systematic evaluation in the forms of maps. Evaluation of the test methods were done through a structured elicitation of the opinions of members of a stakeholder group about the usefulness of the methods. Stakeholders found that general measures of uncertainty, such as prediction error variances (e.g., kriging variance) were less clear than measures which integrated the uncertainty explicitly with the decision–e.g., the probability that the true value of a variable at a site if interest falls below a critical threshold. There was no evidence that they found verbal phrases these (e.g., “very uncertain”) clearer than numerical values (i.e., a probability in the interval [0,1]).

Following on this finding, it was necessary to examine how stakeholders interpret such probability information in more detail. Specifically, is it possible to estimate a probability threshold which a stakeholder group would choose to intervene, reflecting their assessment of the costs attached to errors of commission and omission? Further does this probability depend on framing of the problem (e.g., probability that a threshold is exceeded or that it is not exceeded) and does it depend on professional background of the stakeholder? In a designed experiment, stakeholders were presented with uncertain information on micronutrient supply from a crop, with the uncertainty expressed as a probability with positive framing (probability of adequate supply) or negative framing (probability of insufficient supply). The results showed that probabilities presented in a negative framing led to more conservative decisions, i.e., deciding to intervene at a much smaller probability of deficiency than if the equivalent probability of sufficiency were presented. The elicited probability threshold is prone to framing effects (i.e., how the question is posed), and that this effect interacts with professional group.

The two components of this thesis described above showed how uncertain information can be effectively communicated to stakeholders to support decisions. The next task was to develop a framework for the planning, execution and evaluation of surveys to address specific requirements of these stakeholders. This was based on a decision-theory approach to analyse the particular task, to identify the key uncertainties and their implications and so to enable stakeholders to ensure that an approach to survey would meet their needs. A particular task, based on research practices within the GeoNutrition project (Bill & Melinda Gates funded) was identified–the selection of the study sites to evaluate agronomic biofortification strategies for MND at selected sites based on soil soluble Selenium ($\mathrm{Se_{sol}}$). The information required were analysed, and then the outputs of spatial prediction at national-scale of $\mathrm{Se_{sol}}$ by ordinary kriging, indicator kriging, linear mixed models and random forest were evaluated. There were substantial uncertainties by all three methods, and challenges

with dealing with a complex statistical distribution. This work showed the importance of validation–internal and independent for understanding the suitability of spatial prediction to support decision making.

Uncertainty should be considered when planning sampling for a geostatistical survey. It is important to consider how stakeholders can assess the implications of uncertainty in spatial predictions to determine appropriate sampling grid space for a geostatistical survey. Four approaches (offset correlation, prediction intervals, conditional probabilities and implicit loss functions), that can be used to assess the implications of uncertainty in spatial predictions using prior information on variability of the target properties, were presented to a diverse group of stakeholders in order to determine an appropriate grid spacing. There were variations in the selection made by each method. Some were not well understood. The one which stakeholder favoured, offset correlation, is not directly linked to decision making. More work is needed to develop sound but accessible ways to engage stakeholders with uncertainty consistent across planning and interpretation. Findings from this research will help in better understanding of uncertainties in the data obtained in the GeoNutrition projects thereby facilitating improved use and uptake of that information by decision makers in Ethiopia and Malawi. Better decisions will be made on sampling for such surveys in other countries which decide to undertake those using better methodologies for national-scale surveys of soil properties or similar environmental variables.

Dedication

I dedicate this thesis to my sons Tawana, Tayana and Tafara!

Acknowledgements

This research was supported by the Nottingham-Rothamsted Future Food Beacon International Doctoral Training Program. This work was also supported in part, by the Bill & Melinda Gates Foundation funded GeoNutrition [INV-009129] and MAPS projects INV-002855].

I would like to express my sincere gratitude to my supervisors Professor Murray Lark, Dr Alice Milne, Dr Joseph Chimungu, Dr Patson Nalivata and Professor Martin Broadley for their time, guidance and support during this study. A special mention goes to Murray for the “graffiti” and Alice for the jokes during the meetings. It has been a memorable journey and I learnt a lot from you two! I want to thank School of Biosciences Research Ethics Committee, especially Professor Kate Millar for the comments and suggestions on research ethic applications for the elicitation exercises conducted in this research. I would also like to thank Joanna Smuga-Lumatz, for the managing all the logistics and finances during the entire project period.

I would not have done this without the love and support of my wife Kudzai. She did an amazing job taking caring of Tawana and Tayana when I was away. It was difficult to have been living in different continents and the COVID-19 pandemic made it worse. I am glad my family joined me during my studies. I am grateful to my siblings and my sisters-in-law for their love and support. Finally, I thank God for the gift of life!

Contents

List of Figures

Chapter 1

Introduction and Background

1.1 Problem statement

Micronutrient deficiencies (MND) are a widespread health problem in sub-Saharan Africa (Hurst et al., 2013; Joy et al., 2014), underlying many non-communicable diseases such as anemia, goitre and thyroid dysfunction (Fairweather-Tait et al., 2011; Winther et al., 2020). Most illnesses from MND are associated with deficiency of selenium (Se), iron (Fe), iodine (I) and zinc (Zn) in diets (World Health Organisation, 2004; Chilimba et al., 2011; Joy et al., 2015). Micronutrient deficiencies can reduce immune functions and in some cases impair growth and cognitive development in women and infants (Rayman, 2012; Winther et al., 2020). There is large prevalence of MND in Ethiopia and Malawi, and this is attributed to reliance on predominantly cereal-based diets (Joy et al., 2014; Gashu et al., 2020). This is thought to be, in part, because the availability of micronutrients in crops is limited by the soil in which they are grown (Chilimba et al., 2011; Hurst et al., 2013; Ligowe et al., 2020).

The capacity of the soils in farming regions to supply adequate quantities of micronutrients for optimal crop growth varies widely (Tittonell et al., 2011). Gashu et al. (2021) reported the location of residence had the largest influencing factor in determining the dietary intake of micronutrients from cereals. The cereals commonly cultivated, by people residing in these areas, include maize (*Zea mays* L.), sorghum (*Sorghum bicolor* (L.) Moench), wheat (*Triticum aestivum* L.) and teff (*Eragrostis tef* (Zucc.) Trotter). Joy et al. (2015) showed evidence of restricted uptake of micronutrients into edible portions of crops in countries such as Ethiopia and Malawi, in sub-Saharan Africa. The concentration of micronutrients in these crops vary spatially (Gashu et al., 2021). Spatial information on soil and crop properties is required to improve estimates of MND risk for implementing targeted and efficient interventions. Micronutrient deficiencies can be addressed through the various types of intervention such as agronomic biofortification and food fortification (Joy et al., 2019; Botoman et al., 2022). These interventions should be made at a location where there is a problem. Sound and interpretable spatial information on soil and crop micronutrient variation is therefore needed to support decisions to address MND. These interventions are costly, and their efficient deployment requires that they are targeted to address local needs (Brown et al., 2015).

Spatial information about soil and crop micronutrient concentration would be useful to many stakeholders such as farmers, and agricultural practitioners to inform policy decisions and address knowledge gaps (Lark et al., 2014). This information is needed to identify locations where there will be likely risks of problems in order to plan for mitigatory measures. For example, farmers may require this information to make a decision about the soil, such as application of fertilisers, at a local level (e.g. farm). Agricultural practitioners such as land managers and extension workers are mostly concerned about the management of nutrient supply to improve both crop and livestock quality (Lark et al., 2016).

Low and middle income countries, such as Ethiopia and Malawi, rely on information from past soil surveys, more than 30 years ago, to make decisions on soil conditions at sites of interest. This data is commonly referred to as legacy data. Although generalised soil spatial information is available for sub-Saharan Africa at small-scale (cartographic), provided by African Soil Information Services (AfSIS, 2015), this is largely based on legacy analytical data and key soil chemistry properties relevant to micronutrient status of crops (e.g. particular fraction) as opposed to fertiliser management are not available.

1.2 Soil surveys, their objectives and execution

Conventional soil survey is based on classification of the soil, and the delineation of map units which are each identified with one soil class, or an association of soil classes, as explained in the map legend (Dent and Young, 1981). Most maps are produced with 'general-purpose' classes, defined on genetic principles and intended for many uses (White, 2006). These aim to group together soils produced by similar factors in comparable landscape conditions. 'Special-purpose' classes (e.g. engineering properties, land irrigability classification) may also be mapped, although commonly only at the largest scales and based on one or a few soil properties (White, 2006). In most instances the user's requirement is the basis for separating the classes. Information on the soil is organized with respect to classes (e.g. estimation of class means for soil properties, or provision of a description and analyses of a 'representative profile' (not always statistically robust). Information from soil surveys has mainly been used for planning and managing different agricultural land parcels such as croplands, grazing areas and forestry (Dent and Young, 1981). It has been critical in making sound decisions about to soil management and land-use planning.

Conventional soil survey should begin with reconnaissance and semi-detailed surveys, for general fact findings and these heavily on remote-sensor data or aerial photograph interpretation (White, 2006). The survey start with studying aerial photographs or satellite images, and this would be followed by drawing boundaries. Limited field work activities, such as transects that cross the boundaries, would then be done to validate and interpret the maps. For detailed and semi-intensive surveys, free surveys are undertaken involving interpretation of the landscape on the ground, defining mapping units in terms of landform, vegetation and other clues, and using observations to corroborate a mental model of the landscape in which these map units correspond to one soil class (simple map unit) or an association of classes (complex map units) (Malone et al., 2018). Soil map units are never perfect, inclusions of unpredicted classes will occur because only 0.001% of the survey region is observed (Burrough et al., 1971).

Intensive soil surveys, to map properties affecting a high value crop, are usually done by sampling on a grid, and typically specific soil properties are measured by laboratory analysis. Some methods used to interpolate the soil class at intervening points or to draw contours for properties of interest. Originally, interpolation was done by hand or general contouring program but the discovery of geostatistical methods (Burgess and Webster, 1980) to support such interpolation, gave rise to the use of geostatistical methodology and ultimately to the development of digital soil mapping (DSM).

1.3 Digital soil mapping

The simplest geostatical method, ordinary kriging, uses only information of the target variable but, as described below, can be generalised to a wider class of methods which uses covariates to model the local mean of the property. Other methods, including machine learning, also use covariates. These approaches are generalised into a framework called the ‘scorpan model’ for empirical quantitative description of soil landscape relationships for spatial prediction (McBratney et al., 2003) a generalisation of Jenny’s (1994) factors of soil formation. The ‘scorpan’ factors are: s–soil class or property, c–climate, o–organisms other biotic environmental factors, r–relief or topography, p–parent materials including lithology, a–age, and n–space or spatial position. This is usually represented as $S_c = f\{s, c, o, r, p, a, n\}$, and this conceptual model is very useful in DSM. Many models can represent $f()$, and these can be classified into two broad groups (i) geostatistical methods, and (ii) machine learning (ML) algorithms. Digital soil mapping has three core elements, (i) data input from field and laboratory measurements (e.g. use of soil maps, collection of new samples), (ii) spatial and non-spatial inference through building statistical or algorithmic models relating soil properties and environmental variables (covariates) and (iii) the output in the form of spatial soil information systems in the form of raster prediction maps and uncertainty of the prediction (Minasny et al., 2008; Minasny and McBratney, 2016).

1.3.1 Geostatical models for spatial prediction

Kriging has become the generic term for a range of best unbiased linear predictor (BLUP) methods for spatial prediction of soil properties in geostatistics (Lark et al., 2006; Oliver, 2010). The BLUP of a variable computed from a linear mixed model (LMM) comprises an additive combination of one or more fixed-effects, one or more random effects and an independent random error variable. When the single fixed effect is just an unknown constant mean then the BLUP is equivalent to ordinary kriging and when the fixed effects of the LMM are a combination of spatial coordinates then the BLUP is equivalent to universal kriging (Lark and Cullis, 2004; Lark et al., 2006). When the fixed effects are one or more covariates such as remotely sensed data then BLUP is equivalent to kriging with an external drift or regression kriging (Webster and Oliver, 2007). Estimated variance parameters are required to implement the BLUP. Maximum likelihood (ML) and residual maximum likelihood (REML) methods can be used to estimate the variance parameters of a BLUP. REML allows the estimation of a variance structure for the LMM, and this will be used to obtain the estimates of the model coefficients to form the empirical best unbiased predictor (E-BLUP) (Stein, 1999). Both ML and REML are based on the assumption that random effects have a joint Gaussian distribution, therefore it is important to study the residuals from an exploratory fit of the fixed effects and to transform the data when necessary (Kerry and Oliver, 2007).

The theory of the REML in combination with empirical best linear unbiased predictor (E-BLUP) for spatial prediction is described in greater detail by Lark et al. (2006). The LMM takes the form

$$\mathbf{z} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{z}$ contains n observations of a variable (e.g. soil property) at sampled locations, $\mathbf{X}$ is a design matrix for the fixed effects (e.g., spatial trend, environmental covariates), with $\boldsymbol{\tau}$ the vector of regression coefficients or fixed effects parameters, and $\mathbf{Z}$ is the design matrix for random effects. $\mathbf{Z}$ is typically a n x q identity matrix, n–number of observations and q-number of locations. The vector $\mathbf{u}$ contains random effects, realisations of a variable u, is a Gaussian variable which has zero mean with a covariance, matrix $\mathbf{G}$ that expresses its spatial dependence. The term ε is an independently and identically distributed Gaussian residual with mean zero and a variance σ^2. The error represents both independent error measurements and variation that arises over shorter distances than separate samples. Such that it can be described as the 'nugget' in geostatistical terms because it represents both variation that arises over short distances and measurement errors (Matheron, 1963; Webster and Oliver, 2007). Vectors $\mathbf{u}$ and ε are independent of each other and jointly Gaussian, thus

$$\begin{bmatrix} \mathbf{u} \\ \varepsilon \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2\xi\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix} \right), \tag{1.2}$$

where ξ as the ratio of the variance of u to σ^2 and $\mathbf{I}$ is an identity matrix. Correlation matrix $\mathbf{G}$ will depend only on the relative locations of observations with some specified correlation function, $C(\cdot)$, with one or more parameters that characterise spatial dependence:

$$\mathbf{G}_{i,j} = \mathrm{Corr}\left[\mathbf{u}(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_j)\right] = C(\mathbf{x}_i - \mathbf{x}_j), \tag{1.3}$$

where $\mathbf{u}$ is assumed to be drawn from a second-order stationary random process. The correlation function may take various forms which the spherical or exponential are commonly used. The exponential function has a single distance parameter, a, and

$$C(\mathbf{x}_i - \mathbf{x}_j) = \exp\left\{ \frac{-|\mathbf{x}_i - \mathbf{x}_j|}{a} \right\}. \tag{1.4}$$

With the spherical model, the distance parameter becomes the range, a, at which the correlation goes exactly to zero and

$$\begin{aligned} C(\mathbf{x}_i - \mathbf{x}_j) &= 1 - \frac{3|\mathbf{x}_i - \mathbf{x}_j|}{2a} - \frac{1}{2}\left(\frac{|\mathbf{x}_i - \mathbf{x}_j|}{a}\right)^3 \text{ if } |\mathbf{x}_i - \mathbf{x}_j| < a \\ &= 0 \text{ otherwise.} \end{aligned} \tag{1.5}$$

Other correlation functions such as the Matérn are particularly useful when the variation is locally smooth (Stein, 1999; Minasny and McBratney, 2007). The Matérn covariance function can be written as:

$$C(\mathbf{x}_i - \mathbf{x}_j) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{2\kappa^{\frac{1}{2}}|\mathbf{x}_i - \mathbf{x}_j|}{a}\right)^{\kappa} \mathcal{K}_{\kappa}\left(\frac{2\kappa^{\frac{1}{2}}|\mathbf{x}_i - \mathbf{x}_j|}{a}\right), \tag{1.6}$$

where κ is the smoothness parameter and $\mathcal{K}_\kappa$ is the modified Bessel function of the second kind order κ (Matérn, 1960; Stein, 1999). These correlation functions describe isotropic variation and the variogram will only depend on the lag distance between $|\mathbf{x}_i - \mathbf{x}_j|$ (Lark and Cullis, 2004). The parameters of the covariance function are estimated from observations, $\mathbf{z}$, by maximum likelihood or REML (Marchant and Lark, 2007). Estimation by REML is preferred because it reduces bias in the estimates of the random effects parameters due to uncertainty in the fixed effects parameters (Lark and Cullis, 2004). The residual log-likelihood is conditional on the data and selected fixed effects, contains the unknown parameters of the correlation function in vector $\boldsymbol{\theta}$. The unknown parameters in $\boldsymbol{\theta}$ are σ^2 and ξ. The residual log-likelihood is

$$\ell_R(\sigma^2, \xi, \boldsymbol{\theta}|\mathbf{z}) = -\frac{1}{2}\left\{\log|\mathbf{H}| + \log|\mathbf{X}^{\mathrm{T}}\mathbf{H}\mathbf{X}| + (n-p)\sigma^2 + \frac{1}{\sigma^2}(\mathbf{I} - \mathbf{W}\mathbf{C}^{-1}\mathbf{W}^{\mathrm{T}})\mathbf{z}\right\}, \tag{1.7}$$

where $\mathbf{W} = [\mathbf{Z}, \mathbf{Z}]$, $\mathbf{H} = \xi\mathbf{ZGZ}^{\mathrm{T}} + \mathbf{I}$. The estimates of σ^2 ξ, and $\boldsymbol{\theta}$ that maximise $\ell_R(\sigma^2, \xi, \boldsymbol{\theta}|\mathbf{z})$ will be determined numerically (Lark et al., 2006).

In instances where $\mathbf{z}$ is contaminated by independent errors from a long-tailed distribution, this may affect estimated model parameters and a ‘robustified’ REML may be needed (Künsch et al., 2013; Papritz, 2021). A robust REML identifies the outlying observations and down-weights the outliers when estimating model parameters (Nussbaum et al., 2012). A slightly different residual log-likelihood will be used to estimate the correlation function parameters. A “robustified” REML means that an algorithm to find the maximised ℓ_R is substituted by:

$$\ell_R = \sum_i f\left(\psi\left(\frac{\mathbf{z}_i - \mathbf{X} - u}{\tau}|\sigma^2, \xi, \boldsymbol{\theta}|\mathbf{z}\right)\right), \tag{1.8}$$

where ψ down-weights the outliers (Künsch et al., 2013), to replace the general log-likelihood

$$\ell_R = \sum_i f\left(\frac{\mathbf{z}_i - \mathbf{X} - u}{\tau}|\sigma^2, \xi, \boldsymbol{\theta}|\mathbf{z}\right). \tag{1.9}$$

The estimates of variance parameters obtained by REML, will be used to compute the estimated covariance matrix for the random effects at the sampled points. When the matrix is computed, the next step would be to compute the estimates of fixed effects, $\hat{\boldsymbol{\tau}}$, and predictions of the random effect, $\tilde{\mathbf{u}}$ by solution of mixed model equation:

$$\mathbf{C}\begin{bmatrix}\hat{\boldsymbol{\tau}} \\ \tilde{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathrm{X}^{\mathrm{T}}\mathbf{z} \\ \mathrm{Z}^{\mathrm{T}}\mathbf{z}\end{bmatrix}. \tag{1.10}$$

The covariance of the error of the estimates is

$$\mathrm{Cov}\begin{bmatrix}\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \\ \tilde{\mathbf{u}} - \mathbf{u}\end{bmatrix} = \sigma^2\mathbf{C}^{-1} = \sigma^2\begin{bmatrix}\mathbf{C}^{1,1} & \mathbf{C}^{1,2} \\ \mathbf{C}^{2,1} & \mathbf{C}^{2,2}\end{bmatrix} \tag{1.11}$$

When the fixed effects parameters and covariance matrix have been estimated, they will be used in E-BLUP to make a prediction at unsampled locations,

$$\begin{aligned}\tilde{z}_p &= \mathbf{s}_p^{\mathrm{T}}\hat{\boldsymbol{\tau}} + \tilde{u}_p \\ &= \mathbf{s}_p^{\mathrm{T}}\hat{\boldsymbol{\beta}} + \mathbf{g}_{o,p}^{\mathrm{T}}\mathbf{G}^{-1}\tilde{\mathbf{u}},\end{aligned} \tag{1.12}$$

where $\mathrm{Cov}[\mathbf{u}, \mathrm{u_p}] = \xi\sigma^2\mathbf{g}_{\mathrm{o,p}}$ and $\mathbf{s}_p$ is a p x 1 vector containing the fixed effects. The prediction error variance for the E-BLUP is

$$\mathrm{Var}\left[\tilde{z}_p - z_p\right] = \sigma^2\left\{\left[\mathbf{s}_p, \mathbf{g}_{o,p}^{\mathrm{T}}\mathbf{G}^{-1}\right]^{\mathrm{T}}\left[\mathbf{s}_p, \mathbf{g}_{o,p}^{\mathrm{T}}\mathbf{G}^{-1}\right] - \xi\left(g_{p,p}, \mathbf{g}_{o,p}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{g}_{o,p}\right)\right\}. \tag{1.13}$$

The prediction error variances or kriging variances, $\mathrm{Var}\left[\tilde{z}_p - z_p\right]$ is a 'prior' measure of uncertainty for the E-BLUP. A prior measure of uncertainty is output directly from the prediction process and depends on the predictive model. The kriging variance can be mapped and the map will show how uncertainty varies spatially. The values of kriging variance will be smaller in the neighbourhood of sample points and larger further away. How well the kriging variances characterise uncertainty can be assessed by computing the standardised squared prediction error (SSPE), $\theta(\mathbf{x})$ after cross-validation. The SSPE is computed as

$$\theta(\mathbf{x}) = \frac{\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x}_i)\}^2}{\hat{\sigma}_K^2(\mathbf{x}_0)} \tag{1.14}$$

where $\hat{Z}(\mathbf{x_i})$ is the cross-validation kriging prediction of $z(\mathbf{x_i})$ and $\hat{\sigma}_K^2(\mathbf{x_0})$ is the kriging variance, $\theta(\mathbf{x})$ is expected to have a χ^2 distribution with one degree of freedom if the kriging errors are assumed to follow a Gaussian distribution (Lark, 2000). The median value of $\theta(\mathbf{x})$ over all data is expected to be 0.455 and Lark (2000) showed that this is a more reliable summary of SSPE than its mean (with an expected value of 1).

After cross-validation, the kriging standard errors can be examined. If the kriging standard errors plausibly be regarded as normally distributed, then other posterior measures of uncertainty can be computed such as prediction intervals. The E-BLUP prediction and prediction error variance (kriging variance) are parameters of the prediction distribution and conditional probabilities. A prediction intervals contains the unknown value of a prediction site with specific probability (Heuvelink, 2018). If there is a value of the variable of management significance (e.g. regulatory threshold for soil contamination or a concentration of a micronutrient in a staple grain which corresponds to adequate intake for health), then the conditional probability that the unknown values exceeds this, or does not, can be obtained from the prediction distribution. This is called conditional distribution because it depends on the prediction distribution and hence the model, the data and the location of interest.

With advances in technology, there is now an abundance of auxiliary data which can be used to improve spatial prediction of soil and crop properties as covariates. Sources of the auxiliary data includes satellites (remote sensing) and digital elevation models and maps of soils, vegetation and land use. Using all available covariates for spatial prediction may present some problems. Problems that can be encountered include (i) risk of propagating error in the regression coefficients when weak covariates are included in a model, and (ii) including two or more strongly correlated covariates in a linear model which would result in numerical problems for estimation of regression coefficients (Lark et al., 2007). Methods such as stepwise and backward regression are sometimes used to select covariate(s) from a pool of potential predictors. Stepwise methods consists of sequentially adding and removing covariates from a spatial random model. However, these procedure are not robust to variations in data because the model may depend on the first variable added or the variables deleted. Lark et al. (2007) suggested that the problem of variable selection is considered in terms of multiple hypothesis testing.

Lark (2017) suggested the use of α-investment proposed by Foster and Stine (2008) to select significant covariates for predictive models. Under the α-investment, hypotheses are tested in an ordered sequence with hypothesis j tested against a threshold P_j. The threshold P-value of the jth test depends on the α-wealth after the previous test, W(j - 1). If the previous null hypotheses have been rejected we can set a larger threshold P-value to test hypothesis j while controlling false discovery rate and the sequence of tests will end either at the kth test or the jth test when W(j) goes to zero. So it is important to test the most plausible hypotheses first as this increases the statistical power. A set of k hypotheses in a particular order, will be advanced by considering how available covariates may be considered as explanatory variables through literature review and reflection on the underlying process. The selection and ordering of hypotheses is done without examining relationships between the covariates and other data, but exploratory analysis to identify redundancy between correlated covariates is useful. Having proposed hypotheses in order, they are tested by fitting the corresponding models in order testing each additional covariate by the log-likelihood ratio.

1.3.2 Machine learning for spatial prediction

Machine learning refers to a large class of non-linear data-driven algorithms originally developed for pattern recognition, data mining, regression and classification problems with the ultimate goal for prediction and they don't entail explicit statistical assumptions about the distribution of a soil property (Witten et al., 2016; Wadoux, 2019; Arrouays et al., 2020). Random forest are the most commonly used ML approaches in DSM. Random forests are an ensemble of decision trees (Breiman, 2001). A decision tree algorithm recursively partitions data into several homogenous and non-overlapping regions using a set of splitting rules (Hastie et al., 2009; James et al., 2013). A set of p predictors, $X_1, X_2, \ldots, X_p$, for a dependent variable Y will be split into J regions, $R_1, \ldots, R_j$, that are distinct and non-overlapping. The mean value for observations in each region will be calculated and assigned as the prediction for all the observation that fall in that region (James et al., 2013).

The decision tree algorithm automatically decides on the best split point, s, and the splitting variable, j in the data set. The best pair of j and s are the ones that minimizes the residual sum of squares given by:

$$\sum_{j=1}^{J} \sum_{i=\epsilon R_j} \left(y - \bar{y}_{R_j}\right), \tag{1.15}$$

where $\bar{y}_{R_j}$ is the mean response for the observations in R_j. Fig 1a illustrates a simple decision tree, were the data are partitioned into two homogenous region using the split point, s_i. The initial split to give two regions, R_1 and R_2, also known as terminal nodes. The two regions produced from the first split can then themselves be split in turn, each by

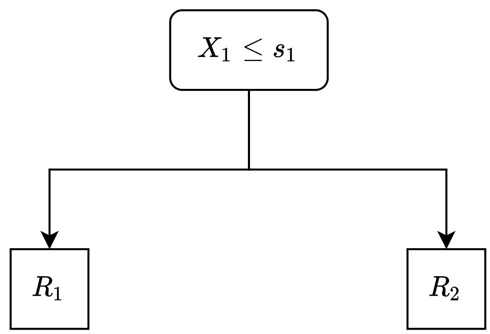


Figure 1a: Example of subtree illustrating the initial split to give two regions.

the selection of the pair (j, s) which minimizes Equation (1.15). This produces two new nodes, as shown in Fig. 1b. The tree now has four terminal nodes, namely: R_1, R_2, R_3 and R_4. These four regions can be further split until as stopping rule is invoked. For example, the splitting process may stop when a minimum node size is reached (Hastie et al., 2009). The splitting of the trees is stopped when the resulting trees achieves the lowest residual sum of squares. At each node, the algorithm, chooses split points that reduce the residual sum of squares for that particular subgroup, rather than optimising splits to reduce the overall residual sum of squares (James et al., 2013). Hence decision trees are referred to as a 'greedy algorithm' and this leads to poor predictions and over-fitting of the observation (Bramer, 2020). Decision trees are unstable and have large 'variance' (Hastie et al., 2009). Large variance in this setting, refers to large changes in the prediction and the model upon making a small change in the observations of Y. Due to problems with decision trees, other tree-structured models have been developed to counter these challenges and use trees as building blocks to construct powerful prediction models such as bagging and random forest.

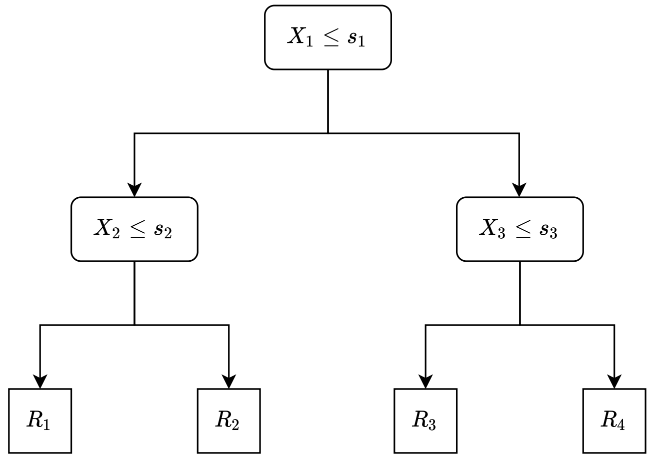


Figure 1b: Example of subtree illustrating further split of the nodes to give four regions

Bagging or bootstrap aggregation was introduced by Breiman (1996) to address some challenges of decision trees. Bagging reduces the variance of the model without increasing bias (Hastie et al., 2009). This is achieved by random sampling with replacement from the training observation of Y (bootstrap). This procedure is repeated several times to produce different bootstrap samples of the training dataset. Decision trees can now fit the bootstrapped samples and a prediction is made from the average of all the trees (James et al., 2013). Some observations can occur more than once in the bootstrap sample, and this may lead to building individual trees in the model which are highly correlated (Boehmke and Greenwell, 2019).

Random forest reduces the strong correlation amongst the trees by using a set of rules when the trees are split (Breiman, 2001). Each tree is built from the bootstrapped samples from the training dataset. At the terminal node, the tree will only consider a random subset of m_{try} predictors from the full list of predictors, p, when splitting. When using random forest for regression, usually $m_{\text{try}} \leq p$ and it is recommended to use $m_{\text{try}} = \frac{p}{3}$ for the split-point, and when $m_{\text{try}} = p$, the algorithm is equivalent to bagging (Boehmke and Greenwell, 2019).

A large number of single trees, k, are grown by the random forest algorithm with identically and independently distributed vector $\Theta_t, t = 1, \ldots, k$. The vector Θ determines how the trees are grown (Meinshausen, 2006). The prediction of a single tree is the weighted sum over all the observation of Y (Breiman, 2001)

$$\hat{\mu}(x) = \sum_{i=1}^{n} w_i(x, \Theta) Y_i. \tag{1.16}$$

The weighted average, $w_i(x)$, is obtained by

$$w_i = k^{-1} \sum_{t=1}^{k} w_i(x, \Theta_t) Y_i. \tag{1.17}$$

After the trees are constructed, they vote for the popular class (Breiman, 2001). Random forest prediction is given by

$$\hat{\mu}(x) = \sum_{i=1}^{n} w_i(x) Y_i. \tag{1.18}$$

In random forest, all other information about the observations in a node of a tree are disregarded except for the conditional mean. The conditional mean, $\mathrm{E}(\mathrm{Y}|\mathrm{X} = \mathrm{x})$, predicted by random forest algorithm is the average predictions of k single trees grown. Quantile random forest is an extension of random forest. The quantile regression forest keeps the value of all observations and assesses the conditional distribution based on this information (Meinshausen, 2006). The full conditional distribution function $\mathrm{F}(Y|X = x)$ will be given by

$$\begin{aligned} \mathrm{F}(y|X = x) &= P(Y \leq y|X = x) \\ &= \mathrm{E}(1_{\{Y \leq y\}}|X = x). \end{aligned} \tag{1.19}$$

The estimate of $\mathrm{E}(1_{\{Y \leq y\}}|X = x)$ is given by the weighted mean of all observations of $1_{\{Y \leq y\}}$,

$$\hat{\mathrm{F}}(y|X = x) = \sum_{i=1}^{n} w_i 1_{\{Y \leq y\}}, \tag{1.20}$$

using the weights in Equation (1.17). The variable $1_{\{Y \leq y\}}$ is an indicator variable,

$$1_{\{Y \leq y\}} = \begin{cases} 1 & \text{If Y} \leq \text{y}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.21}$$

and this is equivalent of indicator kriging. The quantile regression forest can be used to obtain prediction interval for any given α. The estimate $\hat{Q}_\alpha(x)$ of the conditional quantile will then be obtained by $\hat{\mathrm{F}}(y|X = x)$ in Equation (1.22). For example, a 95% prediction interval for the value of Y will be given by

$$I(x) = [Q_{0.025}(x), Q_{0.975}(x)]. \tag{1.22}$$

Random forest builds trees from bagged bootstrapped samples from the training dataset, not all the data will be used for this. The data that is left out when building the model are the out-of-bag observations. Spatial predictions of the target variable at the locations of each out-of-bag observations will be made and the out-of-bag mean square error (MSE) and the out-of-bag R^2 will be computed (Breiman, 2001). The out-of-bag MSE is the prior measure of uncertainty and is regarded as the unbiased estimator of the prediction error (Breiman, 2001). The out-of-bag MSE should be close to zero and be positive. Whilst the out-of-bag R^2 is used to measure the strength and correlation of the model.

The conditional distribution, obtained through quantile regression, of the predicted soil property are very important. The conditional quantiles from the distribution are another prior measure of uncertainty for random forest. The conditional quantiles can be used to compute other measures of uncertainty such as prediction intervals and conditional probability that a value of the target variable falls below or above a threshold.

Validation with an independent dataset is important because they provide posterior measures of uncertainty. The validation usually compare predictions of a soil property $\tilde{Y}$ at location $\mathbf{x}_0$ with independent observations of soil property y at location $\mathbf{x}_i$, from the validation dataset. Then the universal predication accuracy measures such mean error ; mean square error, and root mean square error can be computed from the validation dataset.

The aim of covariate selection is to remove redundant covariates when calibrating ML algorithms. This is important in-order not to over-fit the model and thereby reduce the complexity. Variable importance is used in random forests and decision trees in order to remove the least important predictors. This variable importance is the overall summary of the importance of each predictor by comparing their residual sum of squares and those with low importance can be omitted from the model (James et al., 2013). Two main approaches are used for variable selection when using random forest (i) filter methods and (ii) wrapper methods.

In first approach, commonly referred to as ‘filter’ methods, covariate selection is taken as a pre-processing step before calibrating the ML algorithm. The selection of covariates will be independent from any ML algorithm and this involves exploratory statistical analysis of the covariates. This can done by quantifying the correlation between the covariates by correlation coefficients such as Pearson’s in order to discard those highly correlated from being potential predictors. Then when the best subset of predictors are selected, the training of the algorithm commences. The second approach is the use of 'wrapper' methods and this involves training the ML algorithm several times and based on inferences from previous model covariates can be added or removed from the subset. Commonly used wrapper methods include forward, backward, and recursive feature elimination.

Forward selections begins with a null model that contains no predictors, and then one at a time a covariate is added to the model until all the predictors are added. The best model will be the one with the smallest residual sum of squares or the highest R^2, computed from the calibrated models. Then using an independent validation data, the universal measures are computed for the best model. Backward stepwise selection involves starting with full model with all predictors and then iteratively removes the least useful covariates one at a time (James et al., 2013).

1.3.3 Posterior measures of uncertainty

It is important to validate a model in order to assess the accuracy of predictions. Validation compares the prediction at a site and the true value at that site and this provides a 'posterior' measure of uncertainty. Several approaches can be used to validate predictions and these include jackknifing or data splitting, cross-validation and collection of additional independent data. Brus et al. (2011) recommended the collection of additional independent data by probability sampling for validating predictions. In the event of data being sparse or for it being too expensive to collect additional samples cross-validation is recommended. Cross-validation can be carried out by using leave-one-out-cross-validation and k-fold cross-validation, amongst other methods. With leave-one-out cross-validations, each observation is removed in turn and the remaining will be used for prediction of that observation, using the variance model fitted to all data. k-fold cross-validation involves partitioning the data set into k sets, then one set is removed and the remaining are used for prediction. The procedure is repeated for each of the k sets. With jackknifing, an observation is removed in turn and variance parameters estimated with the remaining data. At the end, there will a set of predictions and model parameters that can be used to compute posterior measures of uncertainty (Journel and Huijbregts, 1976).

After the validation, posterior measures of uncertainty can be computed. These include mean error (bias), mean squared error and the root mean square error. The mean error is computed by

$$\frac{1}{N}\sum_{i=1}^{N}\left(\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x_0})\}\right), \tag{1.23}$$

where $\hat{Z}(\mathbf{x_0})$ is the prediction and $z(\mathbf{x}_i)$ is the true value at that location or value of the observation at that location. The mean error assesses the accuracy of the prediction. The mean error value ideally should be close to zero indicating an absence of bias. The mean square error is computed by

$$\frac{1}{N}\sum_{i=1}^{N}\left(\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x_0})\}\right)^2. \tag{1.24}$$

The mean square error on the other hand should be small. Often the mean square error is expressed as its root, the root mean square error which is in the same units as the measurements. This is computed as

$$\left(\frac{1}{N}\sum_{i=1}^{N}(\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x_0})\})^{\mathbf{2}}\right)^{\frac{1}{2}}. \tag{1.25}$$

If the prediction is biased (i.e. mean error is not close to zero) then this will inflate the mean square error and root mean square error. Imprecise but unbiased predictions also result in large values. These measures of uncertainty can be assessed by comparing them with the variance or standard deviation of the data. The mean square error skill score (Wilks, 2011). It is computed as

$$1 - \frac{\sum_{i=1}^{n}\left(\{z(\mathbf{x}_i) - \hat{Z}(\mathbf{x_0})\}\right)^2}{\sum_{i=1}^{n}\left(z(\mathbf{x}_i) - \frac{1}{n}\sum_{i=1}^{n} z(\mathbf{x}_i)\right)^2}, \tag{1.26}$$

Here, a score of 1 indicates perfect predictions, whilst a score less than 0 indicate predictions with large variance (Wilks, 2011). A score of 0 indicates the predictions being the same with overall mean.

1.3.4 How to distribute sample points in space?

Sampling designs in the context of geostatistical surveys refers to the set of rules for selection of sampling points. Geostatistical methods for spatial prediction depend on covariance parameters estimated from a variogram, and at least 100-150 sampling points are required to estimate an accurate variogram (Webster and Oliver, 1992). Kriging minimises the prediction error variance given the data and the model of the variogram (Webster and Lark, 2013). Therefore, a variogram is needed to determine sampling effort to reduce kriging error and achieve required precision. It is possible to draw some conclusion about spatial variation to support decisions on subsequent sampling from an approximate variogram of a region (Lark et al., 2017) or from a comparable region (Alemu et al., 2022). An approximate variogram can be obtained through reconnaissance surveys. When we have kriging variances of an area we are able to obtain sampling intervals through an optimal sampling scheme for local estimation and mapping regional variables (OSSFIM) proposed by Burgess and Webster (1980) and McBratney et al. (1981). One will able to deduce the required sampling interval for a maximum kriging variances deemed acceptable, by drawing a horizontal line to cut the curve and a perpendicular line from this the intersection will give the required interval (Webster and Lark, 2013). When the sampling interval is determined by a fixed budget, the maximum kriging variance can also be deduced from the graph (Oliver and Webster, 2015). Once the sampling density is obtained the next issue will be how to distribute the samples in the study region.

Systematic sampling on a regular grid is ideal for kriging, because the distance between a target point and nearest sample point is minimised. Square grids are more convenient than triangular grids (Webster and Lark, 2013). Square grids have the largest distance between the target point and sampling points for the same density when compared with triangular grids. However, the kriging variance is only slightly larger because there are four near points on a square grid compared to three on a triangular grid. Regular grids are restrictive in finite irregular regions (i.e. real world) because the prediction error variances will be larger since there are no measurement outside the border. Therefore, the best configuration of the sampling points can be achieved through the use optimised sampling design.

Spatial coverage sampling aims at distributing the grid points in the study area as uniformly as possible (Royle and Nychka, 1998). This involves shifting of grid points to the undersampled areas thereby making the pattern irregular and selection of sampling locations is based on the spatial coordinates of the locations (Brus, 2019). A further step may involve partitioning of the study region into geographic compact blocks that will be used as strata for random sampling (Walvoort et al., 2010). The feature space coverage sampling measurements evenly spread out measurements by utilising k-means algorithms that minimise the feature space distance criterion between sampling points and prediction points (Brus, 2019; Wadoux et al., 2019). Although the sampling on a grid is good for kriging, it does not provide information on spatial dependence over short interval. This can be resolved by additional points near the grid nodes, close-pairs (Lark and Marchant, 2018). The close-pairs are need to give reliable estimates of variogram parameters, and to increase spatial coverage to minimise kriging variances especially at the border of a study region. In this way, sampling points will be set out in systematic way and the grid will be optimised on an estimate of an underlying model.

These methods are favoured for DSM because they allow for building on the relationship of the soil property and one or more several environmental covariates available. When planning sampling for mapping with covariates, when using geostatical methods or ML, it is useful to spread out sample points over the range of covariate values, as well as spreading them out in space, to estimate fixed effects coefficients more precisely.

Conditioned Latin Hypercube sampling (cHLS) is a stratified random sampling procedure that provide a full coverage of the range of each variable by maximally stratifying the marginal distribution (Minasny and McBratney, 2006). It is not straightforward process to optimise the sample design for ML because there is no statistical model in which a relation between sample size/distribution and precision of predictions is implied. This is also due to the fact that ML do not rely on rigid statistical assumptions about the distribution of soil property unlike model based methods which map soil using a known model of spatial variation. However, despite the challenges (Brus, 2019) recommended methods such as feature space coverage sampling and cHLS for ML because they optimised to spread out measurements uniformly in both geographic and feature spaces. Conditioned Latin Hypercube sampling manage this by defining a marginal strata for each covariate and the breaks with each interval are chosen such that the number of pixels in each strata are equal by using quantiles corresponding with evenly spaced cumulative probabilities (Minasny and McBratney, 2006). Also, the marginal distributions of the covariates in the sample are close to the distribution of the entire population, making it advantageous for regression trees and random forest that rely on non-linear relations (Brus, 2019).

1.4 Aims of this research

There is urgent need for spatial information to support decisions on interventions to address MND in sub-Saharan Africa and elsewhere. This has been shown by the GeoNutrution project, amongst others (Gashu et al., 2020, 2021). Obtaining spatial information on soil and crop micronutrient status and other properties that affect micronutrient uptake from the soil, such as pH, requires substantial effort, and there are uncertainties, which depend, in part on the methods used for prediction and the sampling design. Therefore, there is need robust methods for spatial prediction of key information to support decisions making and quantification of the uncertainties. It is essential that uncertainty can be effectively communicated with stakeholders at key stages in soil surveys. These uncertainties must be managed as they will have implications for the outcome of management or policy decisions based on the predictions.

The following issues have been identified. First, the sampling and prediction methods must be robust and efficient. Second, the uncertainties must be quantified appropriately. Finally, the spatial predictions and their uncertainty must be communicated in ways which enable the information-user to make sound decisions which account for uncertainty in a transparent and justifiable way. This includes decisions such as whether to implement an intervention to tackle MND at a particular location, but may also include decision making earlier in the process; for example, how much survey effort is justified given the diminishing returns in reduced uncertainty to increased survey effort and cost. In order to address these challenges, this PhD considers the following research questions:

1. How can spatial uncertainty be most effectively communicated to a range of key decision-makers to support their decision making both on interventions to address MND and on how much resources should be used for sampling?

2. What are the implications of uncertainty in spatial predictions of soil and crop micronutrient status for their use by key decision-makers, and how does this affect various information needs of diverse stakeholders?

3. How is soil and crop micronutrient status most efficiently mapped, in terms how good the predictions have to be and trade off between sampling and degree of precision?

The specific objectives are:

1. To establish how best to effectively communicate uncertainty in spatial prediction of grain micronutrients concentration that fall below threshold levels.

2. To understand how stakeholders can best be helped to make decisions from uncertain spatial information and how they interpret probabilistic information to decide whether to recommend an intervention.

3. To define a “decision process” for selecting the most suitable spatial prediction method for mapping soil micronutrients to address information needs of diverse stakeholders.

4. To establish the optimum sampling densities when planning for a geostatistical survey at national-scale.

1.5 Thesis structure

Chapter 1 gives an overview of the problem, literature review and the objectives of this study. The literature review describes the two groups of methods used in digital soil mapping. Chapter 2 is concerned with how the uncertainty in spatial information about environmental variables can be communicated to stakeholders who must use this information to make decisions. Five methods for communicating the uncertainty in spatial predictions were tested by eliciting the opinions of end-users about the usefulness of the methods, through formal elicitation exercises in Ethiopia and Malawi. Chapter 3 examines how different professional groups (agricultural scientists or health and nutrition experts) interpret uncertain information conveyed with probabilities, when making a decision about interventions to address human Se deficiency. The information provided was a map, either of the probability that Se concentration in local staple grain falls below a nutritionally-significant threshold (negative framing) or of the probability that grain Se concentration is above the threshold (positive framing). In chapter 4, a “decision process” which serves as a to guide to address some of the data needs for end-users of spatial information will be defined. The decision process is used comparing machine learning approaches and geostatistical methods using a case study to identify locations for trials for agronomic biofortification in Malawi. Chapter 5 is concerned with how stakeholders use uncertain information to make decision on sampling strategies. The optimum strategy to obtain sampling density was obtained through formal elicitation with end-users and survey sponsors from Malawi, Ethiopia and the wider GeoNutrition network such as United Kingdom, Zimbabwe and Zambia. The overall discussion and synthesis will be presented in chapter 6. The conclusions and recommendations of the study will also be presented in this chapter.

1.6 Ethical statement

This research used results from the laboratory analysis of soil and crop samples collected in Ethiopia and Malawi in Work Package 1 of the GeoNutrition project. Ethical approvals for the collection of these data allows for their reuse was provided by the University of Nottingham, School of Sociology and Social Policy Research Ethics Committees (REC); BIO-1718-004 and BIO-1819-001 for Ethiopia and Malawi, respectively. The approval ensured that the farmers who's soils and crops were sampled gave informed consent, including subsequent reuse of the data, as in this PhD. Ethical approval to conduct formal elicitation presented in Chapter 2 and Chapter 3 was granted by the University of Nottingham, School of Sociology and Social Policy REC; BIO-1920-007 and BIO-1920- 004 for Ethiopia and Malawi respectively. Ethical approval to conduct the online workshop for eliciting sampling densities, in Chapter 5, was granted by University of Nottingham, School of Biosciences REC; SBREC202122018FEO. These REC approvals, with the exception of the one for Chapter 5 conducted online by University of Nottingham staff only, were recognized formally by the Directors of Research at Addis Ababa University (Ethiopia) and Lilongwe University of Agriculture and Natural Resources (Malawi), who also reviewed the study protocols.

Chapter 2

Communicating uncertainties in spatial predictions of grain micronutrient concentration

This chapter has been published as:

Chagumaira, C., Chimungu, J. G., Gashu, D., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M. (2021). Communicating uncertainties in spatial predictions of grain micronutrient concentration. *Geoscience Communication*, 4 (2), 245–265, https://doi.org/10.5194/gc-4-245-2021

Geosci. Commun., 4, 245–265, 2021
https://doi.org/10.5194/gc-4-245-2021

Geoscience Communication
Open Access
EGU

Communicating uncertainties in spatial predictions of grain micronutrient concentration

Christopher Chagumaira[1,2,3], Joseph G. Chimungu[3], Dawd Gashu[4], Patson C. Nalivata[3], Martin R. Broadley[1], Alice E. Milne[2], and R. Murray Lark[1]

[1]Future Food Beacon of Excellence and School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK
[2]Sustainable Agriculture Sciences Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK
[3]Crop and Soil Sciences Department, Bunda College, Lilongwe University of Agriculture and Natural Resources, P.O. Box 219, Lilongwe, Malawi
[4]Center for Food Science and Nutrition, Addis Ababa University, Addis Ababa, Ethiopia

Correspondence: Christopher Chagumaira (christopher.chagumaira@nottingham.ac.uk, chris.chagumaira@gmail.com)

Received: 6 October 2020 – Discussion started: 27 November 2020
Revised: 19 March 2021 – Accepted: 22 March 2021 – Published: 29 April 2021

Abstract. The concentration of micronutrients in staple crops varies spatially. Quantitative information about this can help in designing efficient interventions to address micronutrient deficiency. Concentration of a micronutrient in a staple crop can be mapped from limited samples, but the resulting statistical predictions are uncertain. Decision makers must understand this uncertainty to make robust use of spatial information, but this is a challenge due to the difficulties in communicating quantitative concepts to a general audience. We proposed strategies to communicate uncertain information and present a systematic evaluation and comparison in the form of maps. We proposed testing five methods to communicate the uncertainty about the conditional mean grain concentration of an essential micronutrient, selenium (Se). Evaluation of the communication methods was done through a questionnaire by eliciting stakeholder opinions about the usefulness of the methods of communicating uncertainty. We found significant differences in how participants responded to the different methods. In particular, there was a preference for methods based on the probability that concentrations are below or above a nutritionally significant threshold compared with general measures of uncertainty such as the prediction interval. There was no evidence that methods which used pictographs or calibrated verbal phrases to support the interpretation of probabilities made a different impression than probability alone, as judged from the responses to interpretative questions, although these approaches were ranked most highly when participants were asked to put the methods in order of preference.

1 Introduction

Micronutrient deficiencies are an important issue in developing countries such as Ethiopia and Malawi. Deficiencies in micronutrients underlie many non-communicable diseases. For example, deficiencies in selenium (Se) can cause thyroid dysfunction, suppressed immune response and increase disease progression and mortality rates, especially in people with already compromised immunity (Fairweather-Tait et al., 2011; Rayman, 2012; Winther et al., 2020).

Micronutrients are largely derived from dietary sources, and there is evidence of a suboptimal intake of Se below recommended levels in Ethiopia and Malawi (Gashu et al., 2020; Ligowe et al., 2020a). Interventions to improve the dietary intake of Se are possible. These include agronomic biofortification, food diversification and fortification (Broadley et al., 2010; Chilimba et al., 2011; Joy et al., 2019; Ligowe et al., 2020b).

Micronutrient deficiencies and the factors that cause them vary spatially (Phiri et al., 2019, 2020; Belay et al., 2020; Gashu et al., 2020). For example, the intake of Se in Ethiopia and Malawi is linked to soil type and other factors (Chilimba et al., 2011; Hurst et al., 2013; Joy et al., 2015). Belay et al. (2020) showed that the risk of Se deficiency is widespread and spatially dependent across Ethiopia. So, spatial information (e.g. on grain micronutrients) can be used to design more efficient interventions to address this micronutrient deficiency.

Soil and crops cannot be sampled everywhere and measurements can only be made directly at a few locations. Using statistical models, interpolations at unsampled locations can be made, but the predictions are uncertain. Predictions are subject to uncertainty because of spatial variability resulting from multiple factors operating at different scales (Lark et al., 2014a). In addition to environmental factors (geology and climate), there is also uncertainty due to measurement error, in the analysis of material, and sampling error, in the field where a single crop or soil sample is collected. When using spatial information, it is important to report this uncertainty and make sure that decision makers understand it in order to make informed decisions.

In geostatistical prediction, the uncertainty of a predicted value is quantified directly by the kriging variance, the mean squared error of the prediction. The prediction is a linear combination of the data, sometimes after a nonlinear transformation, which is optimal in the sense of minimising the kriging variance, given a variogram function which models the spatial dependence of the variable of interest. The kriging variance depends on the spatial distribution of observations. The kriging prediction and variance can be regarded as parameters of a prediction distribution at an unsampled site of interest, which represents our uncertain knowledge about the value of the variable there (conditional on our data and the variogram model). If we assume that the prediction errors are normally distributed, then we can find the interval bounded by the 0.025 and 0.975 quantiles of the prediction distribution as a 95 % prediction interval, which expresses our uncertainty about the true value. It is, therefore, possible to represent the uncertainty in a map of micronutrient concentrations in grain by a corresponding map, which shows the kriging variance, or by the upper and lower bounds of the prediction interval, which can also be mapped.

Other approaches can be taken to communicate the uncertainty in a prediction when the prediction is to be interpreted relative to some threshold value of the mapped variable (e.g. a threshold concentration below which typical intake of grain does not provide adequate intake of a nutrient). While the predicted value may lie above the threshold because the prediction is uncertain, it is possible that the true value is actually below the threshold. This probability, conditional on the data and on the geostatistical model, can be obtained in various ways. A common geostatistical approach is to use indicator kriging (e.g. Webster and Oliver, 2007).

The quantification of uncertainty is generally straightforward, but the communication of this uncertainty to a range of users of information is less so. As Milne et al. (2015) found, the success of a method to present uncertainty may depend on the subject matter and on the background of the interpreter. The probability that the true value lies below a threshold might not be easily interpreted by the policy maker or manager who needs to make a decision based on a map. Probability is often not easily interpreted by a range of end-users of information (Spiegelhalter et al., 2011), and for this reason, in addition to the "raw" probability, verbal interpretations of probability based on "calibrated phrases" (e.g. "unlikely") have been proposed, e.g. the Intergovernmental Panel for Climate Change (IPCC) scale (Mastrandrea et al., 2010). Pictographs may also be used to communicate probabilities by enabling the interpreter to visualise them as proportions (e.g. Spiegelhalter et al., 2011).

Statistical predictions can be used to support decision-making to identify areas of sufficiency or insufficiency. A simple decision model could be based on a threshold value of a variable, with the aim that the user should act if the variable of interest falls below or exceeds the threshold. In our study, we chose a threshold of 38 $\mu g\,kg^{-1}$, based on the assumption that a mean daily intake of 330 g of grain flour should provide a third of the daily estimated average requirement (EAR) of Se for an adult woman. The EAR is a commonly used measure of intake when assessing nutritional status and planning intervention.

In this study, we propose methods for communicating uncertainty in mapped concentrations of micronutrients in grain, using Se as a case study. These methods are based on the kriging variance or on the probability that concentration falls below a nutritionally significant threshold. Maps using these methods, and based on real data collected in Ethiopia and Malawi, were presented to panels of stakeholders in those countries, and their experience of using the maps and their evaluation of the different methods were recorded using questionnaires.

2 Materials and methods

This study was conducted in Ethiopia and Malawi. Ethiopia is located in the horn of Africa (9.1450° N, 40.4897° E), while Malawi is in southern Africa (13.2543° S; 34.3015° E). Primarily, these are research sites for the GeoNutrition project (http://www.geonutrition.com/, last access: 3 July 2020) to inform strategies on addressing micronutrient deficiencies commonly referred to as "hidden hunger". We proposed testing five methods to communicate the uncertainty about predictions of Se concentration in grain (see Sect. 2.1).

In order to determine how best to communicate the uncertainty in our predictions, we recruited participants to evaluate our five candidate methods at two workshops held in Lilongwe, Malawi (November 2019), and Addis Ababa, Ethiopia (January 2020). Each method was presented on a poster, with the same format, consisting of (1) predicted nutrient concentration in map form and (2) a map communicating the uncertainty about the predictions. Examples of the posters are shown in Figs. S1–S5 in the Supplement. Formal evaluations were done through a structured questionnaire that participants completed during the workshops. Ethical approval to conduct this study was granted by the University of Nottingham School of Sociology and Social Policy Research Ethics Committees (BIO-1920-004 for Malawi and BIO-1920-007 for Ethiopia).

2.1 Test methods

2.1.1 Statistical modelling and spatial prediction of grain Se concentration

Field sampling in Amhara, Ethiopia, was previously conducted to support the spatial prediction of Se concentration in grain crops, including the staple crops teff (*Eragrostis tef* (Zucc.) *Trotter*) and wheat (*Triticum aestivum* L; Gashu et al., 2020). The sample frame was defined with reference to the Africa Soil Information Service map of croplands in Amhara region (AfSIS, 2015) so that all sample sites were expected to have a crop or to be near a cropped site. The sample points were selected to give good spatial spread across the sample frame and to be spatially balanced. This procedure was implemented in the `BalancedSampling` library for the `R` platform (`R` Core Team, 2020; Grafström and Lisic, 2016). A total of 25 additional sample sites, closely paired with one of those selected as described above, were added to the sample design to support the estimation of the parameters of the spatial linear mixed model (Gashu et al., 2020). In total, 455 sampling points were obtained, including 136 and 113 locations where teff and wheat were sampled, respectively. The sample support for these data consisted of a bulk grain sample formed from aliquots collected from grain samples within a single field, as described by Gashu et al. (2020). The predictions, and quantifications of uncertainty, therefore relate to grain nutrient concentrations at individual field scale. This is appropriate when considering possible health implications for smallholder and subsistence producers.

In Malawi, the objective of field sampling was to support the spatial prediction of Se concentration in maize (*Zea mays* L), the staple crop. The location of sample points were obtained with the `spcosa` package for the `R` platform (Walvoort et al., 2010). This finds sample points which give good spatial coverage of a sample domain and can incorporate the location of fixed prior points. We had 820 prior points from the 2015–2016 micronutrient survey of Malawi (Phiri et al., 2019) and added a further 890 spatial coverage points with `spcosa`. Of these 1710 sites, 190 were selected at random for a duplicate "close pair" sample to support spatial modelling, with 10 % of the total samples following Lark and Marchant (2018).

We first undertook an exploratory data analysis using simple summary statistics and plots, notably quantile–quantile (QQ) plots, to check whether we needed to transform the data to make the assumption of normality reasonable. In order to check for any spatial trends, we plotted classified post plots which show the spatial location of data and use symbols to indicate quantiles. We found no evidence of the spatial trend in the Malawi data. The data were very skewed, and we transformed them to logarithms to make the assumption of normality plausible. However, for the Amhara data set, we observed a spatial trend. Exploratory analysis indicated that a linear trend model in the spatial coordinates accounted for this, and the exploration of the residuals from the trend indicated that a transformation to logarithm was necessary.

After the exploratory data analysis, we used ordinary kriging to obtain the kriging prediction and kriging variance of grain Se concentration in the Malawi data set for every prediction location on the transformed (log) units. However, for the Amhara data, we used universal kriging, which also makes predictions at unsampled locations, $\boldsymbol{x}_0$, by a weighted linear combination of available sample data designed to minimise prediction error whilst filtering the trend (Webster and Oliver, 2007). The variance parameters for both Amhara and Malawi data sets were estimated by the residual maximum likelihood (Diggle and Ribeiro, 2010) with the `likfit` procedure for the `R` platform.

The kriging predictions were on the log scale and need to be back-transformed for ease of interpretation. For such strongly skewed variables, while an unbiased back-transformation is available, it has been proposed by Pawlowsky-Glahn and Olea (2004) that the median, rather than the mean, of the conditional distribution on the original scale of measurement is obtained by back-transformation (i.e. by simple exponentiation of the kriging prediction). We followed this proposal, and so we refer to our predicted values as the conditional median rather than the conditional mean. The back-transformation of the limits of the prediction interval is straightforward.

We used indicator kriging to obtain the conditional probability that grain Se concentration at the unsampled location exceeds the threshold value, $38\,\mu\mathrm{g\,kg}^{-1}$. Indicator kriging predictions are made by ordinary kriging of an indicator variable created by a transformation of the data on a variable of interest, z, to an indicator variable, w, given a threshold value of interest, z_T. The indicator variable at location $\boldsymbol{x}$ takes the value 0 if $z(\boldsymbol{x}) \leq z_\mathrm{T}$ and 1 otherwise. The estimate of the indicator variable at some location $\boldsymbol{x}_0$ can be interpreted as the conditional probability that $z(\boldsymbol{x}_0) \leq z_\mathrm{T}$ (Webster and Oliver, 2007). While exceedance probabilities could be computed on the assumption of normally distributed errors, we chose to use the widely applied nonparametric method, i.e. indicator kriging, which requires no such assumption.

2.1.2 Kriging variance

In statistical predictions, some unknown quantity (e.g. grain Se concentration at a location) has a prediction distribution conditional on the data and a statistical model. The kriging variance at an unsampled location, $\boldsymbol{x}_0$, is defined as follows:

$$\sigma_K^2 = E[\{Z(\boldsymbol{x}_0) - \tilde{Z}(\boldsymbol{x}_0)\}^2], \tag{1}$$

where the random variable $Z(\boldsymbol{x}_0)$ is predicted by $\tilde{Z}(\boldsymbol{x}_0)$, a kriging prediction. We noted above that the kriging prediction and variance can be regarded as parameters of a prediction distribution at an unsampled site of interest. The dispersion of this distribution reflects our uncertainty about the true value of the variable there, which is, therefore, quantified by the kriging variance. The kriging variance is evaluated at each unsampled site and so can be displayed as a map alongside the map of predictions.

The map of kriging variance is a summary of the uncertainty about our predictions in the study area and shows areas that need further sampling to resolve uncertainty for decision-making. In ordinary kriging, the kriging variance has smaller values near the sample locations and so reflects the distribution of sampling points. For universal kriging, the kriging variance is smallest near the sample location where the values of covariates are close to their respective mean. Because the kriging variance is a direct output of kriging algorithms, it is common to see it mapped alongside kriging predictions and referred to as a measure of local prediction uncertainty (e.g. Holmes et al., 2007; Goovaerts, 2014; Hatvani et al., 2021). However, the interpretation of the kriging variance may be challenging, particularly for a non-specialist user of spatial information. One could take its square root and present it as a kriging standard error with the same units as the target variable. However, the interpretation of the raw standard error can clearly be helped by rescaling it to a prediction interval, and we considered this option in the next section.

The interpretation of the kriging variance is particularly difficult in the case of a variable which must be transformed prior to analysis. The kriging variance cannot be back-transformed to the original units (except for simple kriging). In this setting then, the kriging variance can serve as little more than a general "uncertainty index", indicating in general where uncertainty is large and where it is small. However, such generalised indices have been developed for 3-D geological information to serve the needs of engineering stakeholders (e.g Lelliott et al., 2009; Lark et al., 2014b). For this reason, and because of the long-standing use of kriging variance as an uncertainty measure (see above), we included it as a measure of uncertainty in this experiment. One poster showed a map of the conditional median of Se concentration in grain (Sect. 2.1.1) with a map of kriging variance on the transformed units (see Table 1; Fig. S3).

2.1.3 Prediction intervals

We computed cross-validation predictions from our geostatistical model and exploratory analysis of the kriging errors, $\{z(\boldsymbol{x}_0) - \tilde{Z}(\boldsymbol{x}_0)\}$, and showed that these can be regarded as a normal random variable. Because the kriging predictor is unbiased, the mean of the errors is zero, and their standard deviation is equal to kriging standard deviation $\sigma_K(\boldsymbol{x}_0)$. On this basis, we computed a 95 % prediction interval at each prediction location as $\tilde{Z}(\boldsymbol{x}_0) \pm 1.96\sigma_K(\boldsymbol{x}_0)$. One poster showed a map of the conditional median of Se concentration in grain plus the lower and upper bounds of the 95 % prediction intervals mapped separately to communicate the uncertainty (see Table 1; Fig S1).

Table 1. The designated poster number for each method of communicating uncertain information.

Poster	Method of communication
Poster 1	Prediction interval
Poster 2	IPCC verbal scale
Poster 3	Kriging variance
Poster 4a	Raw probability
Poster 4b	Raw probability plus pictograph

2.1.4 Conditional probability

Using indicator kriging allowed us to quantify uncertainty of the prediction in terms of the probability that the true value exceeds or lies below the threshold. This is a conditional probability, which is conditional on the data and indicator variogram. The probability provides a basis for decisions on interventions, given the threshold value. For example, if the conditional probability that grain Se is below the threshold is very large, then a decision might be made to promote an intervention such as dietary supplementation or agronomic biofortification.

Probability can be presented in a number of different ways – at the first instance, on a raw probability scale, from 0 to 1 or 0 % to 100 %. However, raw probabilities are not very useful to non-specialists as they are often misinterpreted (Spiegelhalter et al., 2011). Given this shortfall, the IPCC (Mastrandrea et al., 2010) introduced a verbal scale for communicating probabilistic information from uncertain results using calibrated verbal phrases. For example, an event with probability < 1 % will be described as "exceptionally unlikely" and an event with probability in the interval 90 %–99 % is described as "very likely". However, the scale is not always interpreted consistently among different individuals. Budescu et al. (2009) observed a tendency for a "regressive" interpretation in which large or small probabilities are interpreted as being close to 50 %. Therefore, we followed Lark et al. (2014a) in supplementing the calibrated verbal phrases with the definition of the probability range.

Graphics, such as pictographs, can be used to report the probability of an event exceeding a threshold. Graphics can be tailored to the target audience and can help those with low numeracy. Zikmund et al. (2008) showed that pictographs significantly improved people's understanding of disease

Table 2. The composition of participants during the meetings in Ethiopia and Malawi.

Meeting/country	Number of participants: Agronomist	Soil scientist	Nutritionist/health practitioner	Total
Ethiopia meeting				
Ethiopia	6	13	17	36
Malawi meeting				
Ethiopia	–	1	1	2
Malawi	6	5	4	15
Pakistan	–	2	–	2
Zambia	–	2	–	2
Zimbabwe	–	2	2	4
Total	12	25	24	61

Figure 1. Use of pictographs reporting a probability of an event exceeding a threshold.

risks compared with other graphics. However, Spiegelhalter et al. (2011) suggested that graphics such as pictographs can be misinterpreted, particularly by people with low numeracy. Therefore, in this study, we proposed to combine raw probabilities and graphics to communicate uncertainty to address these setbacks. In the exercise, we did it by showing the probability map and the pictograph for locations of interest. We used pictographs to report the probability of grain Se concentration exceeding the threshold value, as shown in Fig. 1.

Therefore, we presented three posters, each showing a map of the conditional median of Se concentration in grain (Sect. 2.1.1.), plus probability, and presenting the (1) raw probability scale (see Fig. S4), (2) IPCC verbal scale (see Fig. S2) and (3) raw probability scale plus pictographs (see Fig. S5) to communicate the uncertainty (see Table 1).

2.2 Format of the exercise

We wanted to elicit stakeholder opinions about the usefulness of the communication methods presented as posters described in Sect. 2.1. We invited participants working in the following sectors: agriculture, nutrition and health, non-governmental organisations (NGOs), and universities and government departments from Ethiopia, Malawi and other areas in the wider GeoNutrition project sites. In Ethiopia, through a contact person in the GeoNutrition project, we recruited participants who fitted in the above criterion, and these were mainly local professionals. In Malawi, through contact persons at the Lilongwe University of Agriculture and Natural Resources, we invited participants who fitted the above criterion. Many of the participants were already engaged with the GeoNutrition project. In total, we had 61 participants, with 36 at the Ethiopia meeting and 25 at the Malawi meeting (see Table 2). We asked our participants to assign themselves into one of the three professional groups, i.e. (1) "agronomist", (2) "nutritionist/health practitioner" and (3) "soil scientist". We then asked them to record their level of mathematical education and level of use of statistics or mathematics in their job role.

Evaluation of communication methods was done through a questionnaire, as shown in Table 3, but without putting the participants in a situation where they felt they were being tested on their mathematical skills and understanding. The first part of the questionnaire was an interpretative task (questions 1–3, i.e. Q1–Q3). We presented the participants with true statements about the confidence in the information presented on the maps at different locations (x, y and z). We asked whether the communication of uncertainty was clear. Then, we had the decision-focused task, Q4, in which we asked whether each poster (prediction plus uncertainty) provided adequate information to support a given decision. We then had reflective tasks Q5 and Q6. In Q5, we asked whether in each case the uncertainty about grain Se concentration was straightforward to interpret. We asked if the method of communication helped them understand uncertainty in the predictions in Q6. At the end of the questionnaire, we wanted the participants to assess the methods (Q7) by ranking the posters in order of their effectiveness at communicating uncertainty in the predictions.

Table 3. The list of questions used to elicit stakeholder opinions about the usefulness of the communication methods presented as posters in the workshops in Ethiopia and Malawi.

Question		Response
Question 1 (Q1)	Is it clear from the poster that this statement is true? "Our confidence that grain Se concentration exceeds 38 µg kg^{-1} is greater at x than at z"	(1) Not clear (2) Took a while (3) Can be misinterpreted (4) More information needed (5) Message clear
Question 2 (Q2)	Is it clear from the poster that this statement is true? "Our confidence that grain Se concentration does not exceed 38 µg kg^{-1} is greater at z than at y"	(1) Not clear (2) Took a while (3) Can be misinterpreted (4) More information needed (5) Message clear
Question 3 (Q3)	Is it clear from the poster that this statement is true? "Our confidence that grain Se concentration does not exceed 38 µg kg^{-1} is greater at y than at x"	(1) Not clear (2) Took a while (3) Can be misinterpreted (4) More information needed (5) Message clear
Question 4 (Q4)	Does the poster provide adequate information for you to determine how likely it is that an intervention programme is needed at any given location?	(1) Inadequate information (2) Adequate information (3) More than what I wanted
Question 5 (Q5)	Is the way this poster communicates the uncertainty about grain Se concentration straightforward to interpret?	(1) Not clear (2) Took a while (3) Can be misinterpreted (4) More information needed (5) Message clear
Question 6 (Q6)	Do you think that the poster helped you understand the uncertainty in the predictions?	(1) Yes (2) No
Question 7 (Q7)	Comparing all methods, please rank the posters in order of their effectiveness, in your experience, at communicating uncertainty in the predictions.	Rank 1 being most effective and Rank 5 the least effective.

In each workshop, we started out with an introductory talk to explain the objectives of the exercise. During the talk, we also explained the structure of the questionnaire and how we expected the participants to complete it. After being handed the questionnaires, the participants were directed into a room with the five methods displayed on A0-sized posters. Participants visited each poster in a randomised order to avoid any bias resulting from the carry-over effects from one poster to another when the individual responses were pooled for analysis. For example, if participants found a particular method easier to interpret, this might help them understand the next poster that they examined. Participants were not allowed to speak to one another when they were completing their questionnaires to avoid bias. When completing the last two questions on the questionnaire, participants were allowed to revisit the posters without following the randomised order to revise their answers. A non-specialist facilitator was stationed at the poster, to check that participants were on the correct pages on the colour-coded questionnaire, to check that all questions were completed and to help with any problems (e.g. translating language).

2.3 Data analysis

We presented our results for Q1 to Q6 as contingency tables, where the selected responses in the rows (of which there are n_r) and the columns (of which there are n_c) are the posters (i.e. methods of communication), separated either between the location of the meeting (Ethiopia or Malawi) or between professional group (agronomist, soil scientist or nutritionist/health practitioner) of the respondent. Analysis of the contingency table allowed us to test the null hypothesis of the random association of the responses with the factor in columns (i.e. that the proportion of participants indicating a particular response to the question is independent of the poster which they are considering). The description of how we partitioned contingency tables to evaluate whether there were differences between the location of the meeting and professional groups is given in the Appendix.

The null hypothesis for a contingency table is equivalent to an additive log-linear model of the table under which the expected number of responses in cell $[i, j]$, $e_{i,j}$, is the product of the row and column totals (n_i and n_j) divided by the total number of responses, N. An alternative log-linear model, the so-called "saturated" model for the table, has an extra $(n_\mathrm{r}-1)\times(n_\mathrm{c}-1)$ term which allows an interaction between the rows and columns of the table, such that the proportions of different responses may differ among all the posters.

The evidence for the saturated model, as a better model for the data than the additive model, is provided by the likelihood ratio statistic or the deviance for the two models, L, where, in the following:

$$L = 2\sum_{i=1}\sum_{j=1} o_{i,j}\log\frac{o_{i,j}}{e_{i,j}}, \quad (2)$$

and $o_{i,j}$ are the number of observed responses in cell $[i, j]$. Under the null hypothesis of random association between the rows and columns of the table, L has an approximate χ^2 distribution, with $(n_\mathrm{r}-1)\times(n_\mathrm{c}-1)$ degrees of freedom (Christensen, 1997; Lawal, 2014). We fitted the log-linear models using the `loglm` function from the `MASS` package in the `R` platform (Venables and Ripley, 2002).

Our primary interest is whether there are differences in the responses recorded by our participants depending on the method of communicating uncertainty. However, it was first necessary to consider whether there was evidence for differences in the responses between the two sets of respondents at different locations. Such differences might arise because of differences in the composition of the groups (Table 2), differences between the examples presented (a map from the Amhara region in Ethiopia and a map of Malawi), differences between the contexts (in Ethiopia, many were local professionals recruited for the exercise; in Malawi, many of the participants were already engaged with the GeoNutrition project) and the possibility of unconscious changes in how the second meeting, in Ethiopia, was conducted (adapting from the experience of conducting the exercise in Malawi). Because our participants were drawn from different professional groups, we thought this would affect their responses, and if this was the case, then this would also be of interest because it would suggest that people from different professional backgrounds find some methods better than others.

For this reason, we first tested whether there were differences in the overall responses between the locations of the meetings, using a contingency table in which the responses to different posters by people from different professional groups are pooled within the two meeting locations. This gave us a five (responses) by two (locations) contingency table, with 4 degrees of freedom for each poster (Q1–Q3 and Q5), or a three (responses) by two (locations) contingency table, with 2 degrees of freedom (Q4), or a two (responses) by two (locations) contingency table, with 1 degree of freedom (Q6). We next tested whether there were differences in the overall responses between the different professional groups, using a contingency table in which the responses to different posters were pooled within each of those groups.

For some questions, there were differences in the responses between the location of the meeting. But for no questions was there any evidence to reject the null hypothesis of random association between responses and the professional group of the participants. We, therefore, proceeded to consider a set of prior hypotheses about differences in the responses between posters and the methods which they employed to communicate uncertain information, based either on a partition of the separate subtables for each location (where the locations differed) or of a table in which the responses from the different locations were pooled.

The first hypothesis which we considered is that participants would respond differently to a threshold-based approach to uncertainty (in which the poster presents the probability that the Se concentration in grain at an unsampled site falls below or above a threshold – posters 2, 4a and 4b) than they would to a general measure of uncertainty (the kriging variance, poster 3, or the prediction interval for the prediction, poster 1). We call this hypothesis H^1, and the evidence against the corresponding null hypothesis, H^1_0, was evaluated by the deviance in the subtable for which the responses to posters 2, 4a and 4b were pooled in one column (threshold based) and the responses to posters 1 and 3 were pooled in a second.

The second hypothesis that we considered, H^2, was that the respondents' views on the posters that used kriging variance would differ from their views on the posters that used prediction intervals. The evidence against the corresponding null hypothesis, H^2_0, was tested by the subtable comprising the responses to poster 1 in one column and the responses to poster 3 in a second.

The deviances for the tables testing null hypotheses H^1_0 and H^2_0 are two components of the deviance for the overall table (whether this is pooled over several locations or a subtable for one location). The remaining deviance component is for a subtable with all the separate responses to threshold-based methods. This can be partitioned into two further components, which address our two remaining hypotheses.

The first of these, hypothesis H^3, was that respondents would have different opinions about poster 4a (raw probability values) than the posters (4b and 2) in which guides to the interpretation of the probability are given (pictographs or a partition of the probability into intervals corresponding to the calibrated phrases of the IPCC scheme). The null hypothesis, H^3_0, is tested by the deviance of a table in which one column comprises responses to poster 4a and the second contains pooled responses to posters 4b and 2.

The final hypothesis, H^4, was that respondents would have different opinions on the poster which used the calibrated phrases of IPCC (poster 2) and the rather different approach

Table 4. Analysis of Q1 according to the location of the meeting, professional group and methods that the latter tested on separate location subtables.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	P*
Full contingency table analysis				
Full table		93.33	36	< 0.001
Pooled within location of meeting		22.83	4	< 0.001
Pooled within professional group		11.71	8	0.16
Subtable – Ethiopia meeting				
Poster effects		21.78	16	0.15
Threshold based vs. general	H_0^1	9.61	4	0.05
Within general	H_0^2	7.10	4	0.13
Within threshold based		5.07	8	0.75
Poster 4a vs. guided	H_0^3	2.64	4	0.62
Poster 4b vs. poster 2	H_0^4	2.43	4	0.66
Subtable – Malawi meeting				
Poster effects		48.72	16	< 0.001
Threshold based vs. general	H_0^1	31.95	4	< 0.001
Within general	H_0^2	6.53	4	0.16
Within threshold based		10.24	8	0.25
Poster 4a vs. guided	H_0^3	8.87	4	0.06
Poster 4b vs. poster2	H_0^4	1.37	4	0.85

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

of poster 4b, with pictographs imposed on a map of probabilities.

The approaches above were applied for Q1 to Q6.

We tabulated the responses for Q7, with ranks as the rows and posters as the columns. Participants were asked to rank the preferred poster first, but we reversed this for the analysis, giving a rank of 5 to the most preferred poster and of 1 to the least. We considered only those responses in which a complete ranking was provided by the respondent. The mean rank was calculated for each poster, and this was done over all respondents and then separately for locations and for professional groups.

For a set of rankings of k items, under a null hypothesis of random ranking, the expected mean rank for each item is $(k+1)/2$. The evidence against this null hypothesis can be measured by the following statistic:

$$\frac{12n}{k(k+1)}\sum_{i=1}^{k}\left\{\overline{r_i}-\frac{k+1}{2}\right\}^2, \qquad (3)$$

where $\overline{r_i}$ is the mean rank of the ith item, and a total of n rankings comprise the data. Under the null hypothesis, this statistic is distributed as $\chi^2(k-1)$ (Marden, 1995).

3 Results

At the Ethiopia meeting, we had fewer participants (64 %) who had studied mathematics and statistics up to degree level and above than at the Malawi meeting (88 %; see Fig. S9). We had more participants using statistics or mathematics regularly in their job at the Malawi meeting (52 %) than at the Ethiopian meeting (18 %). Most of the participants at the Ethiopian meeting (58 %) occasionally use mathematics or statistics in their jobs. There were more soil scientists (48 %) at the meeting in Malawi than agronomists and nutritionists/health practitioners. While, in Ethiopia, there were more nutritionists/health practitioners (47 %) compared to the other professional groups.

3.1 Interpretative tasks

The full tables for responses over both locations and all posters to Q1 are shown in Table A1 in the Appendix. The responses pooled for both meeting locations are shown in Table A2. There is strong evidence for differences among the columns of the full table ($P < 0.001$) and strong evidence ($P < 0.001$) against the null hypothesis of random association between posters and responses pooled within locations and responses (Table 4). However, there was no evidence

Table 5. Analysis of Q2 according to the location of the meeting, professional group and methods that the latter tested on separate location subtables.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	$P*$
Full contingency table analysis				
Full table		60.66	36	0.01
Pooled within location of meeting		24.42	4	< 0.001
Pooled within professional group		14.95	8	0.06
Subtable – Ethiopia meeting				
Poster effects		16.21	16	0.44
Threshold based vs. general	H_0^1	7.59	4	0.11
Within general	H_0^2	2.18	4	0.70
Within threshold based		6.44	8	0.60
Poster 4a vs. guided	H_0^3	3.91	4	0.42
Poster 4b vs. poster 2	H_0^4	2.52	4	0.64.
Subtable – Malawi meeting				
Poster effects		20.02	16	0.22
Threshold based vs. general	H_0^1	5.34	4	0.25
Within general	H_0^2	6.93	4	0.14
Within threshold based		7.76	8	0.46
Poster 4a vs. guided	H_0^3	4.04	4	0.40
Poster 4b vs. poster2	H_0^4	3.72	4	0.45

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

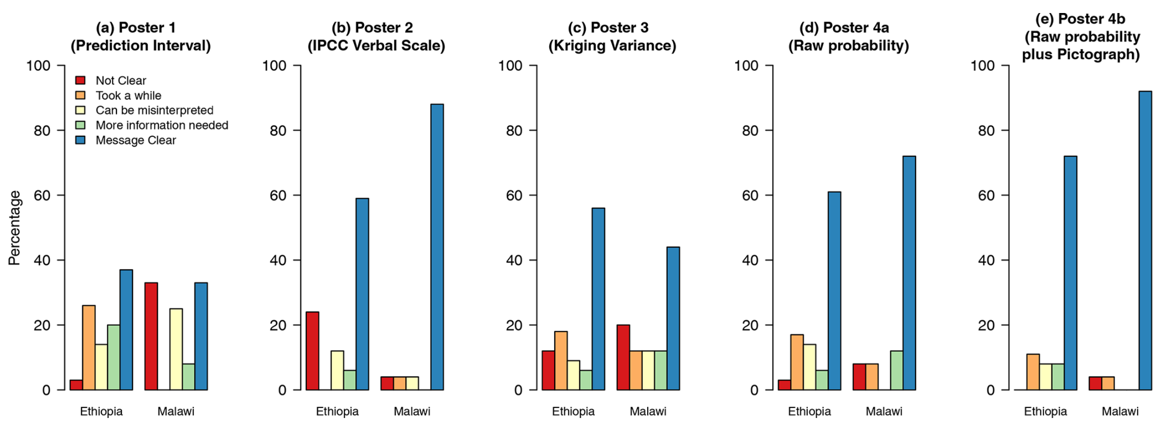


Figure 2. Bar charts showing how participants when pooled within the location of the meeting responded to the interpretive task (Q1).

to reject the null hypothesis of random association between posters and responses pooled within professional groups. On this basis, further analysis of responses to posters was based on the separate subtables for the Ethiopia and Malawi meeting locations. Similar results were obtained for Q2 and Q3, as shown in Tables 5 and 6, respectively.

For Q2, while there is evidence for a difference in responses between the two meeting locations, there is no evidence, either for the responses from Ethiopia or from Malawi, to reject the null hypothesis for any of the focussed questions about differences between posters (see Table 5). For Q3, however, there is evidence for a difference in the responses for the threshold-based methods and the general methods in the responses from Ethiopia ($P = 0.009$) and from Malawi ($P = 0.02$; see Table 6).

Figure 2 shows the responses to Q1 for the separate posters for each subtable. Threshold-based methods were found to be clearer by a larger proportion of the participants. In both

Table 6. Analysis of Q3 according to the location of the meeting, professional group and methods that the latter tested on separate location subtables.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	P^*
Full contingency table analysis				
Full table		60.36	36	0.006
Pooled within location of meeting		21.93	4	0.0002
Pooled within professional group		10.01	8	0.26
Subtable – Ethiopia meeting				
Poster effects		16.60	16	0.41
Threshold based vs. general	H_0^1	13.48	4	0.009
Within general	H_0^2	0.51	4	0.97
Within threshold based		2.61	8	0.96
Poster 4a vs. guided	H_0^3	2.03	4	0.73
Poster 4b vs. poster 2	H_0^4	0.58	4	0.97
Subtable – Malawi meeting				
Poster effects		21.83	16	0.15
Threshold based vs. general	H_0^1	11.67	4	0.02
Within general	H_0^2	4.07	4	0.40
Within threshold based		6.09	8	0.64
Poster 4a vs. guided	H_0^3	4.07	4	0.40
Poster 4b vs. poster2	H_0^4	2.03	4	0.73

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

countries, there was a marked difference between poster 1 (prediction intervals) and the rest, with a much smaller proportion of respondents selecting the response "message clear". In Malawi, a large proportion of respondents selected "not clear" as their response for this poster. The figures which summarise responses for Q2 and Q3 are shown in the Supplement (Figs. S10 and S11).

3.2 Decision-focused task

There was no evidence for differences among the columns of the full table ($P = 0.11$) and strong evidence ($P = 0.01$) against the null hypothesis of random association between posters and responses pooled within locations and responses for Q4 (Table 7). However, there was no evidence to reject the null hypothesis of random association between posters and responses pooled within professional groups. Therefore, further analysis of responses to posters was based on the separate subtables for the Ethiopia and Malawi meeting locations.

For Q4, we have no evidence to reject the null hypothesis of the random association between poster and response for any of our set of four focussed hypotheses in Ethiopia. In Malawi, however, there is evidence ($P = 0.03$) to reject the H_0^1 and not the other focussed hypotheses.

Figure 3 shows the responses to Q4 for the separate posters for each subtable graphically. The larger proportion of the participants found threshold-based methods to provide adequate information for decision-making. In Ethiopia, poster 3 (kriging variance) was different from all other posters, with a large proportion of respondents selecting "inadequate information".

3.3 Reflective task

There is no evidence for differences among the columns of the full table ($P = 0.26$) for Q5 (Table 8). Also, there is no evidence ($P = 0.63$) against the null hypothesis of the random association between posters and responses pooled within locations. Table 9 shows that there is strong evidence for differences among the columns of the full table ($P = 0.001$) for Q6. However, the evidence is marginal ($P = 0.05$) against the null hypothesis of random association between posters and responses pooled within locations and responses. However, there was no evidence to reject the null hypothesis of random association between posters and responses pooled within professional groups for both Q5 and Q6. On this basis, further analysis of responses to posters was based on pooled counts for the Ethiopia and Malawi meetings. The responses for Q5 are shown in Table A3.

Table 7. Analysis of Q4 according to the location of the meeting, professional group and methods that the latter tested on separate location subtables.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	P^*
Full contingency table analysis				
Full table		25.70	18	0.11
Pooled within location of meeting		9.14	2	0.01
Pooled within professional group		8.96	4	0.06
Subtable – Ethiopia meeting				
Poster effects		6.47	8	0.59
Threshold based vs. general	H_0^1	4.34	2	0.11
Within general	H_0^2	0.28	2	0.87
Within threshold based		1.85	4	0.76
Poster 4a vs. guided	H_0^3	1.22	2	0.54
Poster 4b vs. poster 2	H_0^4	0.63	2	0.73
Subtable – Malawi meeting				
Poster effects		10.09	8	0.26
Threshold based vs. general	H_0^1	6.94	2	0.03
Within general	H_0^2	1.61	2	0.45
Within threshold based		1.53	4	0.82
Poster 4a vs. guided	H_0^3	0.63	2	0.73
Poster 4b vs. poster2	H_0^4	0.90	2	0.64

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

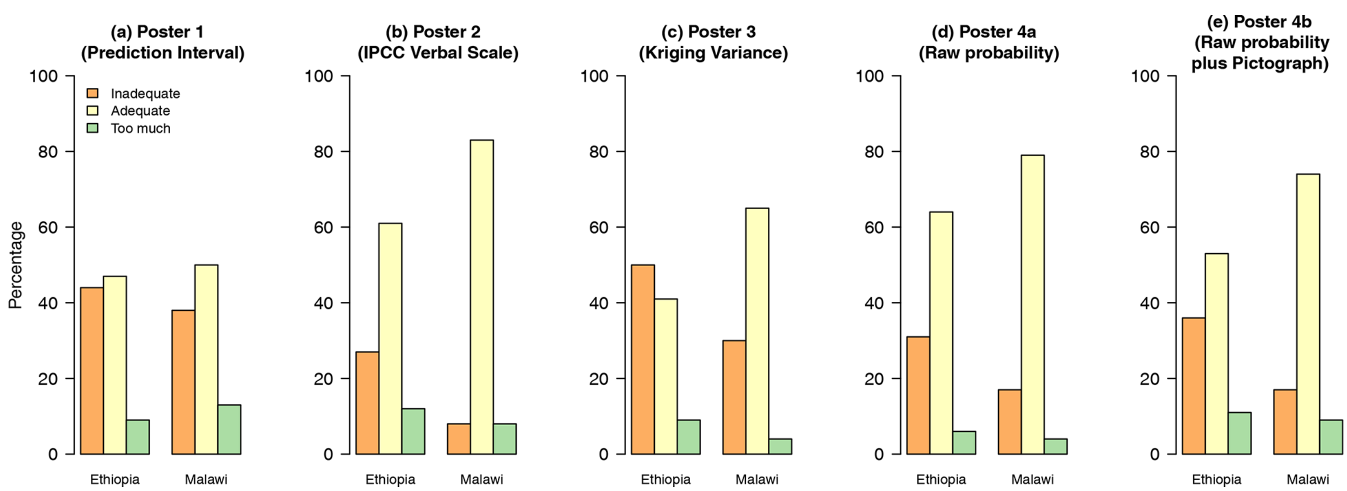


Figure 3. Bar charts showing how participants, when pooled according to the location of the meeting, responded to whether a method provided adequate information or not (Q4).

As shown in Table 8, we have evidence ($P = 0.02$) to reject the null hypothesis of contrasting the threshold-based methods with the general uncertainty measures for Q5. For Q6, there is evidence for a difference in the responses for the threshold-based methods and the general methods ($P < 0.001$). However, we have no evidence for the second, third and fourth focussed hypotheses in both Q5 and Q6.

Figure 4 shows the responses to Q5 for the separate posters for pooled counts graphically. We can see that there is a greater proportion of respondents selecting the response "message clear" for threshold-based methods, i.e. posters 2

Table 8. Analysis of Q5 according to the location of meeting, professional group and methods that the latter tested on pooled counts over Ethiopia and Malawi.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	P^*
Full contingency table analysis				
Full table		40.93	36	0.26
Pooled within location of meeting		2.55	4	0.63
Pooled within professional group		2.35	8	0.99
Pooled counts over Ethiopia and Malawi				
Poster effects		17.74	16	0.34
Threshold based vs. general	H_0^1	12.23	4	0.02
Within general	H_0^2	1.11	4	0.89
Within threshold based		4.40	8	0.82
Poster 4a vs. guided	H_0^3	2.34	4	0.67
Poster 4b vs. poster 2	H_0^4	2.06	4	0.72

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

Table 9. Analysis of Q6 according to the location of the meeting, professional group and methods that the latter tested on pooled counts over Ethiopia and Malawi.

	Specified null hypothesis	Deviance (L^2)	Degrees of freedom	P^*
Full contingency table analysis				
Full table		29.08	9	0.001
Pooled within location of meeting		23.69	1	0.05
Pooled within professional group		0.39	2	0.82
Pooled counts over Ethiopia and Malawi				
Poster effects		24.13	4	< 0.001
Threshold based vs. general	H_0^1	3.60	1	< 0.001
Within general	H_0^2	0.002	1	0.97
Within threshold based		0.53	2	0.77
Poster 4a vs. guided	H_0^3	0.34	1	0.56
Poster 4b vs. poster 2	H_0^4	0.18	1	0.67

Note: each row of this table presents a test of a null hypothesis of random association between the rows and columns of a contingency table, but the four highlighted here correspond to the prior hypotheses about differences among posters which are of primary interest. The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random association of the rows and columns of the table holds.

(IPCC verbal scale), 4a (raw probability) and 4b (raw probability plus pictograph), than on general based. We also see more people selecting the response "not clear" for posters 1 (prediction intervals) and 3 (kriging variance), the general-based methods. Figure 5 shows how participants responded to Q6. There was a marked difference between poster 3 (kriging variance) and the rest, with a much larger proportion of respondents selecting the response "no".

3.4 Assessment of the method

For Q7, first, we computed the mean ranks for all the participants and measured the evidence against the null hypothesis of random ranking using Eq. (3). Table 10 shows that there is strong evidence ($P = 0.002$) against the null hypothesis of random ranking.

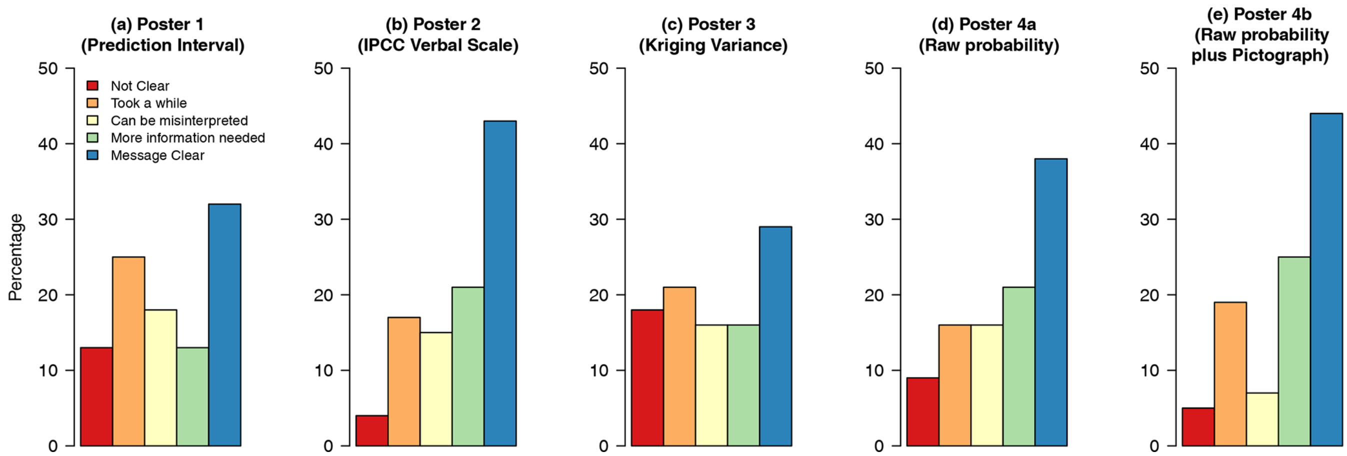


Figure 4. Bar charts showing how participants responded to whether a method is straightforward to interpret (Q5).

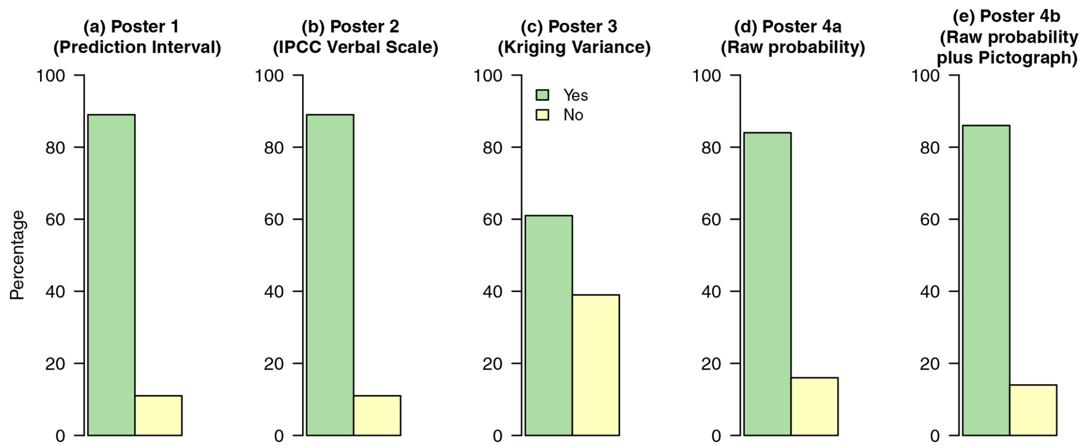


Figure 5. Bar charts showing how participants responded to how each poster helped them understand uncertainty in the spatial predictions (Q6).

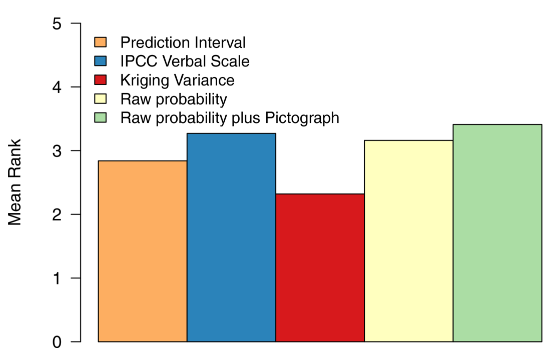


Figure 6. Ranking of posters in terms of the most effective at communicating uncertainty about spatial predictions.

Second, we computed mean ranks for each location of the meeting. After the test, we found no evidence ($P = 0.12$) against the null hypothesis in Ethiopia. However, at the Malawi meeting, there was strong evidence ($P = 0.001$). The difference may be because the set of stakeholders at the Malawi meeting was more homogenous in terms of professional group (a less even distribution among them) and level of mathematical education than the stakeholders at the Ethiopia meeting.

Last, we computed mean ranks for the different professional groups. We found strong evidence against the null hypothesis of random ranking for the nutritionists/health practitioners ($P = 0.017$) and not for soil scientists ($P = 0.16$) and agronomists ($P = 0.23$).

Figure 6 shows the mean rankings for the separate posters for all the respondents graphically. Posters 4b (raw probability plus pictograph) and 2 (IPCC verbal scale) had the largest mean ranks, and poster 3 (kriging variance) had the least. Threshold-based methods were found to be more effective at communicating uncertainty about spatial predictions of grain Se concentration.

Table 10. Analysis of Q7 according to all the respondents, the location of the meeting and the professional group.

	Test statistic (X^2)	Degrees of freedom	P^*
All respondents	16.90	4	0.002
Location of meeting			
Ethiopia	7.44	4	0.12
Malawi	18.21	4	0.001
Professional group			
Agronomist	5.60	4	0.23
Soil scientist	6.51	4	0.16
Nutritionist/health Practitioner	12.10	4	0.017

The asterisk (*) indicates the probability of obtaining a deviance statistic this large or larger if the null hypothesis of random ranking of the rows and columns of the table holds.

4 Discussion

In this study, we tested strategies to communicate uncertain information through a systematic evaluation and comparison with distinct groups of data end-users. We found significant differences between participants' responses to the posters which employed general measures of uncertainty (kriging variance or prediction interval) and those which presented the probability that the Se concentration in grain falls below or above a threshold. The interpretative task that participants undertook was based on interpretation of the information relative to a nutritional threshold. The presentation of uncertainties in terms of probabilities framed with respect to this threshold was found more accessible by data users than the general measures of uncertainty, despite the general view (see Spiegelhalter et al., 2011) that users of information commonly find probabilities hard to interpret. Our results suggest that users of information can find information presented in terms of probabilities accessible and clear.

There was no evidence that the participants responded more positively to communication of uncertainty in the form of probabilities when these were supported with pictographs, or the calibrated phrases of the IPCC scheme, in contrast to the simple map of probability, although the maps with pictographs were highest ranked. These methods to assist the interpretation of probability are widely used because of the assumption that many users of information find probabilities hard to interpret. However, there is evidence that calibrated phrases are themselves not without problems. Budescu et al. (2009) reported substantial inconsistencies in how people interpret scales of calibrated phrases, with a tendency to have a "regressive" interpretation (interpreting large or small probabilities as close to 0.5). Jenkins et al. (2019) found that presentations of probability in numerical formats were consistently perceived as more credible than verbal expressions. While the posters using pictographs were ranked highest (Fig. 6) in our study, we have not shown that they are markedly preferred. We note that our study focussed on stakeholders' preferences and opinions and did not include tests of how correctly the information was interpreted. We, therefore, suggest that further work is needed before a definitive assessment can be made of the value of calibrated phrases or pictographs in supplementing raw probability, while noting that we have not found them to be markedly more congenial to the user.

Kriging variances were the lowest-ranked poster in the participants' overall assessment (Fig. 6). The kriging variance is fundamental to the geostatistical approach for predicting spatial variables. It is the quantity which is minimised by the kriging predictor, and its virtues as a built-in measure of the uncertainty of point predictions have been widely acknowledged. Nonetheless, it is clear that the kriging variance in itself is not an accessible measure of uncertainty for most end-users. Along with prediction intervals, the kriging variance is a general measure of uncertainty which reflects the spatial variability of the target variable and the local density of the sample. Although the kriging variance is a valid statistic, in this context it has very little value as a means for communicating uncertainty for a general audience. That is particularly true in this case, where the kriging variance must remain on transformed units, and so it serves as little more than a general uncertainty index. This was clear a priori and is confirmed by the responses we received. Our findings here cannot, therefore, be regarded as definitive, and a similar experiment for variables which do not require transformation would be necessary in further research. In such cases, one could also include the kriging standard error as an uncertainty measure to assess (i) whether the fact that it is presented in the units of the target variable makes it preferable to kriging variance and (ii) whether it is regarded as less interpretable than its rescaled form as a prediction interval. That said, our results do show that the communication of prediction intervals requires more attention.

These considerations aside, kriging variances, standard error and prediction intervals must be interpreted by the user along with other information (for example, is the predicted value close to the threshold or substantially different from it?) in order to make a judgement at a particular location. Our results do show that the probability measure, tied directly to the interpretative task, is clearer to the user than general measures of uncertainty.

Prediction intervals were not ranked highly by our participants, and we had no evidence that they were found any clearer than the kriging variance. In part, this might be because of the limitations in presenting the predictions and upper and lower bounds of the prediction interval as three separate maps. The task of interpreting the information at one location or comparing two, when this entails examining three maps, may have influenced the participants' responses. In other settings, the prediction intervals might be more effective for interpretation, for example where the user of information can display the prediction intervals for a prediction at a site of interest as a single figure (e.g. a bar against a scale) with the threshold value of concern indicated. Further work is needed on different ways to present the prediction intervals for interpretative tasks.

We only found strong evidence of differences between the meeting location for questions on interpretative and decision-focused tasks. This can be attributed to the composition of each group. Participants at the Malawi meeting comprised researchers and stakeholders already somewhat engaged with the GeoNutrition project, whereas those in Ethiopia were mainly local stakeholders not previously involved with the project.

The participant groups from the two locations differed in their self-assessed level of mathematical education and use of mathematics and statistics in their work. We had more participants with mathematical components in their education up to degree level in Malawi than in Ethiopia. We had fewer people who had mathematical education only to secondary/high school level in Malawi than in Ethiopia. There were fewer participants who used mathematics and statistics regularly at the Ethiopia meeting. This, along with the differences in role noted in the previous paragraph, might contribute to differences between the locations. However, our data cannot support a more detailed assessment of the effects of mathematical background because they are strongly unbalanced. For example we only had 3 % of participants educated up to certificate/diploma level at the Ethiopia meeting. Further work to address this question and examine how stakeholders interpreted each poster will require an elicitation with sufficient numbers of participants with different mathematical backgrounds. This would be useful for a better understanding of how different learning styles influence the interpretation of uncertain information.

No map is perfect (Heuvelink, 2018), but maps must be used as a basis for decisions. It is, therefore, important to ensure that the user of spatial information is aware of the uncertainty in these predictions, and that these are communicated in a clear way. The user must be aware that the predictions have an attached uncertainty, and it is therefore possible that a decision they make might be judged as being incorrect in the light of perfect information. Given this, the user must have a clear enough understanding of the uncertainty attached to a prediction so as to be confident that the decision they make will be robust given the uncertainty. For example, the predicted concentration of a nutrient in a staple crop at a location may be such that the intake of the nutrient should be sufficient to meet the needs of those who eat that crop. The user should consider the uncertainty in that prediction. If the probability that the threshold concentration is exceeded is just 0.6 (about as likely as not on the IPCC scale), then they may conclude that a decision on whether or not to proceed with an intervention at that location requires further information. If, on the other hand, the probability is 0.95 (very likely) then they may be confident in deciding to prioritise interventions elsewhere. However, if the uncertainty is not communicated clearly, then the data user might be overconfident in predictions where the probability that the threshold is exceeded is only just over 0.5 and may waste resources in further investigation or unnecessary interventions at locations where the prediction was well supported and indicated adequate local concentrations of the nutrient.

The findings of this study complement work that has been done on cartography and visualisation for spatial information (Kunz et al., 2011; Beven et al., 2015). Our findings show the importance of finding cartographic solutions to represent probability information and to develop interactive methods for interpretation in a geographic information system (GIS) environment (e.g. to produce pictographs, like those we have used, for sites of interest or to find more effective ways of representing the 95 % prediction interval). It is good practice to use a consistent colour scale for the three legends showing the lower and upper 95 % prediction interval and the conditional median. However, in our study, we could not use one colour legend for the three maps for Fig. S1 (poster 1) because of the marked differences in the predicted values on back-transformation. This made it difficult to find a working colour scale from the minimum value in the lower bound to the maximum in the upper bound on which one would see the variation in all three maps. We opted to use a continuous legend on the map of the mean and discrete ones for the lower and upper limits. This might have hindered interpretation. However, we suspect that there is a need for fundamentally different ways of visualising prediction intervals, perhaps by using interactive methods to display them in a GIS environment.

We accept that a possible source of bias in any such study is that a participant feels that they are being tested on their interpretative skills and so might select a response which suggests, in a general sense, that they understand the input (e.g."message clear" for the case in Table 3). However, all participants were aware that their responses were strictly anonymous, and it was emphasised that the task involved their evaluation of several methods for the communication of an interpretation which was provided. In future studies, it might be useful to include some final questions which actually are "tests of interpretation", secondary to the main task, to see whether this affects the responses given for different methods.

5 Conclusions

Despite the general expectation that users of spatial information do not generally find probabilities a congenial way to express uncertainty, we found that when probability is used to quantify the uncertainty in a specific interpretation of spatial information, based on a nutritionally significant threshold, end-users largely found the approach clear and preferable to general measures of uncertainty which are not directly linked to the specific interpretation (prediction intervals and kriging variance). In the general assessment and ranking of how methods to present uncertainty succeeded, the methods based on a specific interpretation of the information, using probability, were again preferred. There was no significant evidence for a difference in assessment by users of presentations which used probability alone or those which used pictographs or verbal phrases to aid in the interpretation of the raw probability values, although these latter methods were ranked highest among all methods.

Because decisions on interventions to address nutrient deficiencies may have positive and negative effects on peoples' health and well-being, the interpretation of information such as that we have used is not value neutral, and uncertainty in information has ethical implications (given that all spatial information is uncertain, how much uncertainty is ethically acceptable in the decision-making process?). While these considerations are outside the scope of the study reported here, it would be interesting, in future research, to examine how individual attitudes to the ethics of fortification interventions affect their responses and whether individuals' perspectives on the ethical implications of basing decisions on uncertain information differs between different methods to communicate that uncertainty.

To conclude, we suggest that the challenge of communicating the significance of uncertain information to a range of stakeholders should be considered in the context of specific interpretations of the information (e.g. nutrient concentrations relative to thresholds of nutritional significance) and that, in this setting, probabilities can be accessible to a wide range of end-users. Calibrated phrases or pictographs seem to have some value (given the rankings by our participants), although there is no strong evidence that they should be preferred to a simple map of the probability. While general measures of uncertainty (kriging variance and prediction intervals) are valid ways of quantifying uncertainty, they are less effective for communication, although other ways of presenting prediction intervals for spatial data in interactive formats online or in a GIS may merit further investigation.

Appendix A

In this section, we describe how we partitioned contingency tables to evaluate whether there were differences between the location of the meeting and professional groups. A full table, such as the one shown in Fig. A1, may be hard to interpret. Table A1 shows a full table of how many individuals selected a given response to Q1, the interpretive task. It is possible to partition the table, and its deviance statistic and degrees of freedom, into components corresponding to pooled tables and subtables of the full table. This is illustrated in Fig. A1. Here the full table is partitioned into a subtable for responses from Malawi and another subtable for responses from Ethiopia, as shown also in Table A2. A pooled table, in which the responses pooled over all posters in Malawi were compared with the responses similarly pooled from Ethiopia, completes the partition. As shown in Fig. A1, the deviance statistics for these three tables, and their degrees of freedom, sum to the deviance and degrees of freedom for the full table. In this case, we could conclude whether there are differences in the responses between the two locations (if not, then we might pool the responses for any poster at the two locations), and whether there are differences in responses to the posters at each location in turn. As described below, we used this approach to evaluate whether there were differences between the two locations. We also used it to examine evidence for differences in the responses for professional groups. Having done this, we then analysed either pooled tables or separate subtables (e.g. for responses in Ethiopia and responses in Malawi) to examine a priori contrasts between particular posters and groups of posters.

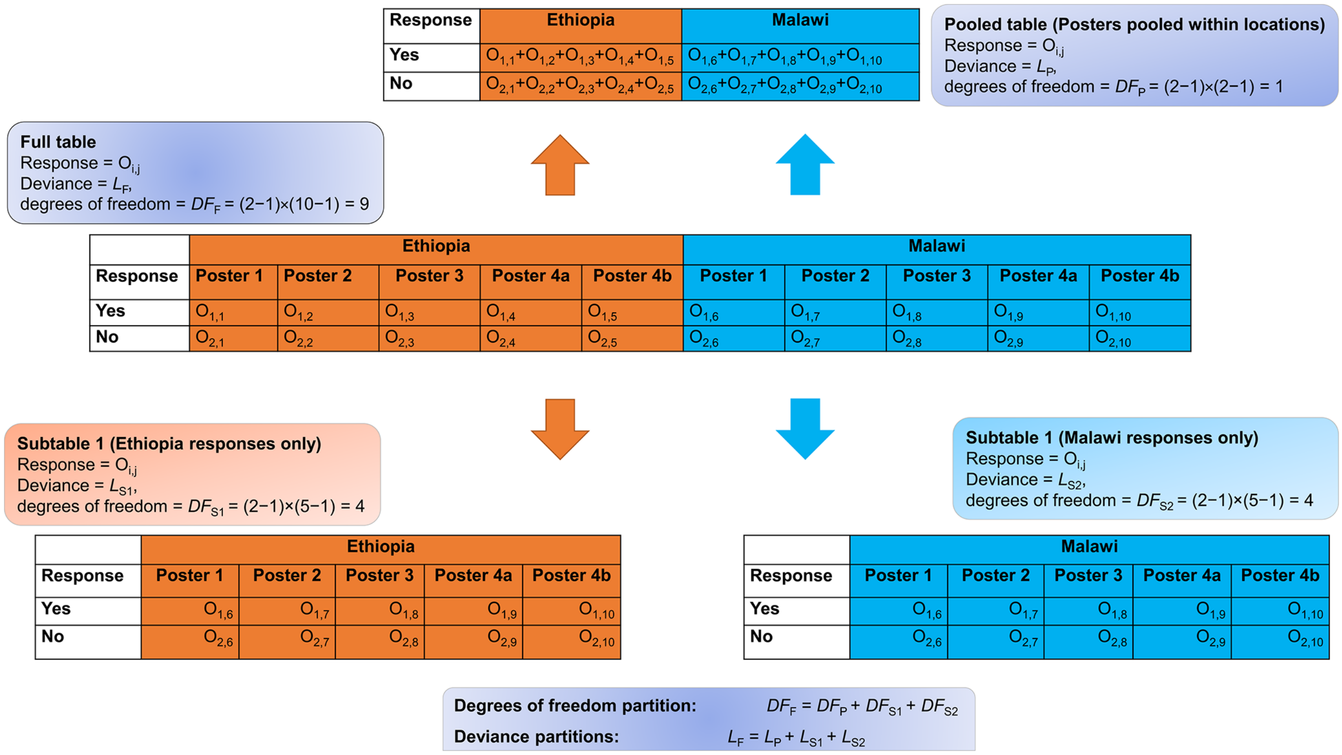


Figure A1. An illustration of how the log likelihood ratio can be partitioned into subtables and pooled tables.

Table A1. The full contingency table showing how many individuals selected a given response to Q1 (interpretive task). The table is presented according to the location of the meeting and the method of communication. The figures in parentheses are the expected numbers, and $e_{i,j}$ is the product of the row and column totals (n_i and n_j) divided by the total number of responses (N).

Response	Ethiopia					Malawi				
	Poster 1	Poster 2	Poster 3	Poster 4a	Poster 4b	Poster 1	Poster 2	Poster 3	Poster 4a	Poster 4b
Not clear	1(1)	0(1)	4(1)	1(1)	0(1)	8(3)	1(3)	5(3)	2(3)	1(3)
Took a while	9(7)	8(6)	6(6)	6(7)	4(7)	0(1)	1(1)	3(1)	2(1)	1(1)
Can be misinterpreted	5(4)	4(4)	3(4)	5(4)	3(4)	6(2)	1(2)	3(2)	0(2)	0(2)
More information needed	7(3)	2(3)	2(3)	2(3)	3(3)	2(2)	0(2)	3(2)	3(2)	0(2)
Message clear	13(20)	20(19)	19(19)	22(21)	26(21)	8(16)	22(16)	11(16)	18(16)	22(16)

Table A2. A subtable showing how many individuals selected a given response to Q1 when columns are pooled within the location of the meeting.

Response	Ethiopia	Malawi
Not clear	6	17
Took a while	33	7
Can be misinterpreted	20	8
More information needed	16	8
Message clear	100	81

Table A3. Responses to Q5 pooled responses from the Ethiopia and Malawi meetings.

Response	Pooled counts
Not clear	27
Took a while	55
Can be misinterpreted	40
More information needed	53
Message clear	103

Code and data availability. The data and code that support this research are available at https://doi.org/10.6084/m9.figshare.14465736.v2 (Chagumaira et al., 2021).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/gc-4-245-2021-supplement.

Author contributions. The project was conceived, designed and implemented by CC, AEM and RML. PCN, JGC and DG supervised the data collected. RML and MRB were responsible for project administration and funding acquisition. All authors contributed to the preparation of the article.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. The funders of this paper were not involved in the study design or the collection, management, analysis and interpretation of the data, the writing of the report or the decision to submit the report for publication.

Acknowledgements. The authors gratefully acknowledge the contributions made to this research by the participating farmers and field sampling teams. In Ethiopia, field sampling teams were from the Amhara National Regional Bureau of Agriculture. In Malawi, field sampling teams were from the Department of Agricultural Research Services and the Lilongwe University of Agriculture and Natural Resources.

We also would like to acknowledge the contributions made by E. Louise Ander, Edward J. M. Joy and Alexander A. Kalimbira during the design of the experiment. We also appreciate Diriba B. Kumssa, Adamu Belay, Mesfin Kebede, Demeke Teklu Senbetu, Kidist Gemechu and Hasset Tamirat for their roles as facilitators during the elicitation in Malawi and Ethiopia.

Financial support. This research has been supported by the Bill and Melinda Gates Foundation (grant no. INV-009129) and the Nottingham–Rothamsted Future Food Beacon (International Agriculture DTP).

Review statement. This paper was edited by Kirsten von Elverfeldt and reviewed by Hazel Napier and one anonymous referee.

References

AfSIS: New cropland and rural settlement maps of Africa, available at: http://africasoils.net/2015/06/07/new-cropland-and-rural-settlement-maps-of-africa (last access: 25 April 2020), 2015

Belay, A., Joy, E., Chagumaira, C., Zerfu, D., Ander, E. L., Young, S. D., Bailey, E. H., Lark, R. M., Broadley, M. R., and Gashu, D.: Selenium Deficiency Is Widespread and Spatially Dependent in Ethiopia, Nutrients, 12, 1565, https://doi.org/10.3390/nu12061565, 2020.

Beven, K., Lamb, R., Leedal, D., and Hunter, N.: Communicating uncertainty in flood inundation mapping: a case study, Int. J. River Basin Manage., 13, 285–295, https://doi.org/10.1080/15715124.2014.917318, 2015.

Broadley, M. R., Alcock, J., Alford, J., Cartwright, P., Foot, I., Fairweather-Tait, S. J., Hart, D. J., Hurst, R., Knott, P., McGrath, S. P., Meacham, M. C., Norman, K., Mowat, H., Scott, P., Stroud, J. L., Tovey, M., Tucker, M., White, P. J., Young, S. D., and Zhao, F. J.: Selenium biofortification of high-yielding winter wheat (*Triticum aestivum* L.) by liquid or granular Se fertilisation, Plant Soil, 332, 5–18, https://doi.org/10.1007/s11104-009-0234-4, 2010.

Budescu, D. V., Broomell, S. B., and Han, P.: Improving communication of uncertainty in the reports of Intergovernmental Panel on Climate Change, Psychol. Sci., 20, 299–308, 2009.

Chagumaira, C., Murray L. R., and Milne, A. E.: Data and Code for Chagumaira et al. 2021 [Dataset], figshare, https://doi.org/10.6084/m9.figshare.14465736.v2, 2021.

Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., and Broadley, M. R.: Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi, Sci. Rep., 1, 72, https://doi.org/10.1038/srep00072, 2011.

Christensen, R.: Log-Linear Models and Logistic Regression, Springer, Springer-Verlag, New York, 1997.

Diggle, P. and Ribeiro, P. J.: Model-based geostatistics, Springer-Verlag, New York, 2010.

Fairweather-Tait, S. J., Bao, Y. P., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., and Hurst, R.: Selenium in Human Health and Disease, Antioxid. Redox Sign., 14, 1337–1383, https://doi.org/10.1089/ars.2010.3275, 2011.

Gashu, D., Lark, R., Milne, A., Amede, T., Bailey, E., Chagumaira, C., Dunham, S., Gameda, S., Kumssa, D., Mossa, A., Walsh, M., Wilson, L., Young, S., Ander, E., Broadley, M., Joy, E., and McGrath, S.: Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia, Sci. Total Environ., 733, 139231, https://doi.org/10.1016/j.scitotenv.2020.139231, 2020.

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Oxford University Press, 1997.

Goovaerts, P.: Geostatistics: a common link between medical geography, mathematical geology, and medical geology, J. S. Afr. I. Min. Metall., 114, 605–612, 2014.

Grafström, A. and Lisic, J.: BalancedSampling: Balanced and Spatially Balanced Sampling. R package version 1.5.2, available at: https://CRAN.R-project.org/package=BalancedSampling (last access: 26 March 2020), 2016.

Hatvani, I. G., Szatmàri, G., Kern, Z., Erdélyi, D., Vreča, P., Kanduč, T., Czuppon, G., Lojen, S., and Kohán , B.: Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary, Environ. Int., 146, 106263, https://doi.org/10.1016/j.envint.2020.106263, 2021.

Heuvelink, G. B. M.: Uncertainty and uncertainty propagation in soil mapping and modelling, in: Pedometrics (Progress in

Soil Science), edited by: McBratney A. B., Minasny, B., and Stockmann, U., Springer International Publishing, 439–461, https://doi.org/10.1007/978-3-319-63439-5-14, 2018.

Holmes, K. W., Van Niel, K. P., Kendrick, G. A., and Radford, B.: Probabilistic large-area mapping of seagrass species distributions, Aquat. Conserv., 17, 385–407, 2007.

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., and Broadley, M. R.: Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi, Sci. Rep., 3, 1425, https://doi.org/10.1038/srep01425, 2013.

Jenkins, S. C., Harris, A. J. L., and Lark, R. M.: When unlikely outcomes occur: the role of communication format in maintaining communicator credibility, J. Risk Res., 22, 537–554, https://doi.org/10.1080/13669877.2018.1440415, 2019.

Joy, E. J. M., Kumssa, D. B., Broadley, M. R., Watts, M. J., Young, S. D., Chilimba, A. D. C., and Ander, E. L.: Dietary mineral supplies in Malawi: spatial and socioeconomic assessment, BMC Nutrition, 1, 1–25, 2015.

Joy, E. J. M., Kalimbira, A. A., Gashu, D., Ferguson, E. L., Sturgess, J., Dangour, A. D., Banda, L., Chiutsi-Phiri, G., Bailey, E. H., Langley-Evans, S. C., Lark, R. M., Millar, K., Young, S. D., Matandika, L., Mfutso-Bengo, J., Phuka, J. C., Phiri, F. P., Gondwe, J., Ander, E. L., Lowe, N. M., Nalivata, P. C., Broadley, M. R., and Allen, E.: Can selenium deficiency in Malawi be alleviated through consumption of agro-biofortified maize flour? Study protocol for a randomised, double-blind, controlled trial, Trials, 20, 795, https://doi.org/10.1186/s13063-019-3894-2, 2019.

Kunz, M., Grêt-Regamey, A., and Hurni, L.: Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system, Nat. Hazards, 59, 1735–1751, https://doi.org/10.1007/s11069-011-9864-y, 2011.

Lark, R. M. and Marchant, B. P.: How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters?, Geoderma, 319, 89–99, https://doi.org/10.1016/j.geoderma.2017.12.022, 2018.

Lark, R. M., Ander, E. L., Cave, M. R., Knights, K. V., Glennon, M. M., and Scanlon, R. P.: Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland, Geoderma, 226, 64–78, https://doi.org/10.1016/j.geoderma.2014.03.002, 2014a.

Lark, R. M., Mathers, S. J., Marchant, A., and Hulbert, A.: An index to represent lateral variation of the confidence of experts in a 3-D geological model, P. Geologists' Assoc.,125, 267–278, 2014b.

Lawal, B.: Applied Statistical Methods in Agriculture, Health and Life Sciences, Springer International Publishing, Switzerland, 2014.

Lelliott, M. R., Cave, M. R., and Wealthall, G. P.: A structured approach to the measurement of uncertainty in 3D geological models, Q. J. Eng. Geol. Hydroge., 42, 95–105, 2009.

Ligowe, I. S., Phiri, F. P., Ander, E. L., Bailey, E. H., Chilimba, A. D., Gashu, D., Joy, E. J., Lark, R. M., Kabambe, V., Kalimbira, A. A., Kumssa, D. B., Nalivata, P. C., Young, S. D., and Broadley, M. R.: Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages, P. Nutr. Soc., 79, 457–467, https://doi.org/10.1017/S0029665120006904, 2020a.

Ligowe, I. S., Young, S. D., Ander, E. L., Kabambe, V., Chilimba, A. D., Bailey, E. H., Lark, R. M., and Nalivata, P. C.: Selenium biofortification of crops on a Malawi Alfisol under conservation agriculture, Geoderma, 369, 114–315, 2020b.

Marden, J.: Analyzing and modeling rank data, CRC Press, Boca Raton, 1995.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Ottmar, E., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., and Zwiers, F. W.: Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties, Intergovernmental Panel on Climate Change (IPCC), IPCC Cross-Working Group Meeting on Consistent Treatment of UncertaintiesJasper Ridge, CA, USA, 6–7 July, 2010.

Milne, A. E., Glendining, M. J., Lark, R. M., Perryman, S. A. M., Gordon, T., and Whitmore, A. P.: Communicating the uncertainty in estimated greenhouse gas emissions from agriculture, J. Environ. Manage., 160, 139–153, https://doi.org/10.1016/j.jenvman.2015.05.034, 2015.

Pawlowsky-Glahn, V., and Olea, R. A.: Geostatistical Analysis of Compositional Data, Oxford University Press, 2004

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., and Broadley, M. R.: The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales, Sci. Rep., 9, 6566, https://doi.org/10.1038/s41598-019-43013-z, 2019.

Phiri, F. P., Ander, E. L., Lark, R. M., Bailey, E. H., Chilima, B., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Phuka, J. C., Suchdev, P. S., Middleton, D. R. S., Hamilton, E. M., Watts, M. J., Young, S. D., and Broadley, M. R.: Urine selenium concentration is a useful biomarker for assessing population level selenium status, Environ. Int., 134, 105218, https://doi.org/10.1016/j.envint.2019.105218, 2020.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/, last access: 3 March 2020.

Rayman, M. P.: The importance of selenium to human health, Lancet, 356, 233–241, https://doi.org/10.1016/s0140-6736(00)02490-9, 2000.

Spiegelhalter, D., Pearson, M., and Short, I.: Visualizing Uncertainty About the Future, Science, 333, 1393–1400, https://doi.org/10.1126/science.1191181, 2011.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S. Fourth Edition, Springer-Verlag, New York, 2002.

Walvoort, D. J. J., Brus, D. J., and de Gruijter, J. J.: An R package for spatial coverage sampling and random sampling from compact geographical strata by *k*-means, Comput. Geosci., 36, 1261–1267, https://doi.org/10.1016/j.cageo.2010.04.005, 2010.

Webster, R. and Oliver, M. A.: Geostatistics for Natural Environmental Scientists, 2nd edn., John Wiley & Sons Chichester, 2007.

Winther, K. H., Rayman, M. P., Bonnema, S. J., and Hegedus, L.: Selenium in thyroid disorders – essential knowledge for clinicians, Nat. Rev. Endocrinol., 16, 165–176, https://doi.org/10.1038/s41574-019-0311-6, 2020.

Zikmund-Fisher, B. J., Fagerlin, A., and Ubel, P. A.: Improving Understanding of Adjuvant Therapy Options by Using Simpler Risk Graphics, Cancer, 113, 3382–3390, https://doi.org/10.1002/cncr.23959, 2008.

Supplement for: Chagumaira et al 2021 Communicating uncertainties in spatial predictions of grain micronutrient concentration

Chagumaira, C., Chimungu, J. G., Gashu, D., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M. (2021). Communicating uncertainties in spatial predictions of grain micronutrient concentration.*Geoscience Communication*, 4 (2), 245–265, https://doi.org/10.5194/gc-4-245-2021

Supplement of Geosci. Commun., 4, 245–265, 2021
https://doi.org/10.5194/gc-4-245-2021-supplement

Supplement of

Communicating uncertainties in spatial predictions of grain micronutrient concentration

Christopher Chagumaira et al.

Correspondence to: Christopher Chagumaira (christopher.chagumaira@nottingham.ac.uk, chris.chagumaira@gmail.com)

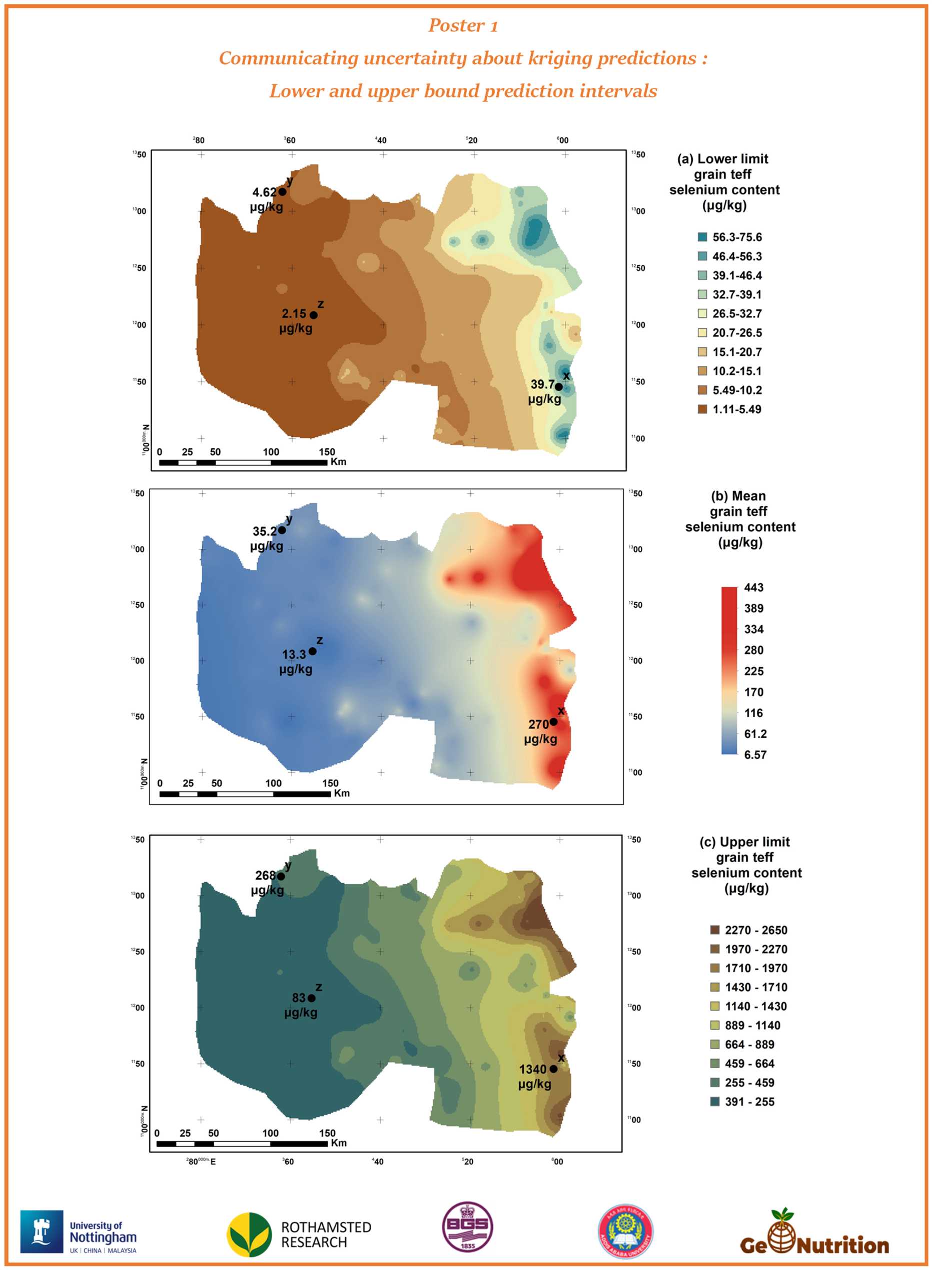


Figure S1. Poster showing the prediction intervals (lower and upper limit) and the predicted selenium concentration in teff grain in Amhara region, Ethiopia

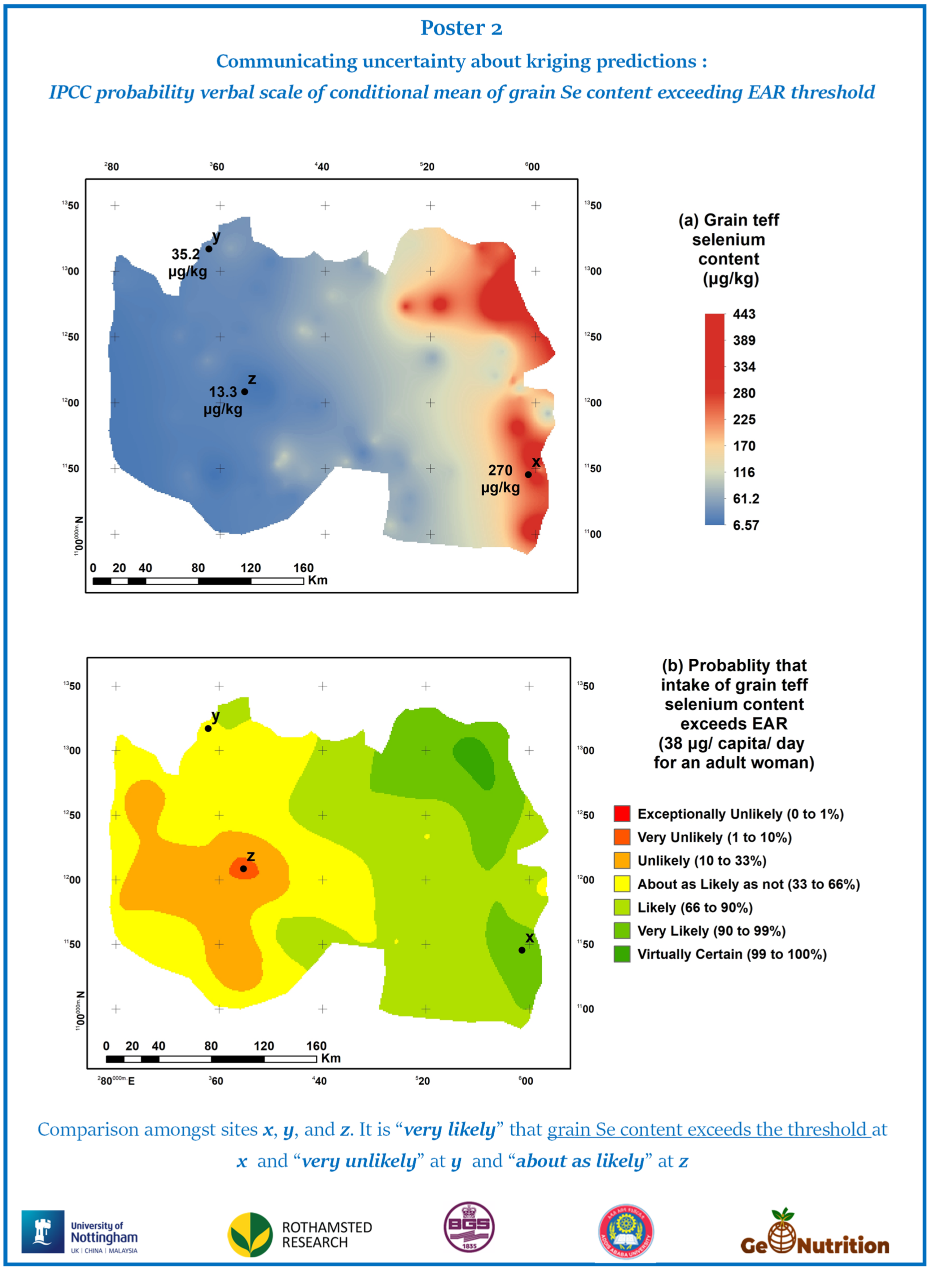


Figure S2. Poster on verbal probability scale (with probabilities indicated as percentages) of the probability that intake of grain teff selenium concentration is less than the threshold 38 μg kg^{-1} in Amhara region, Ethiopia

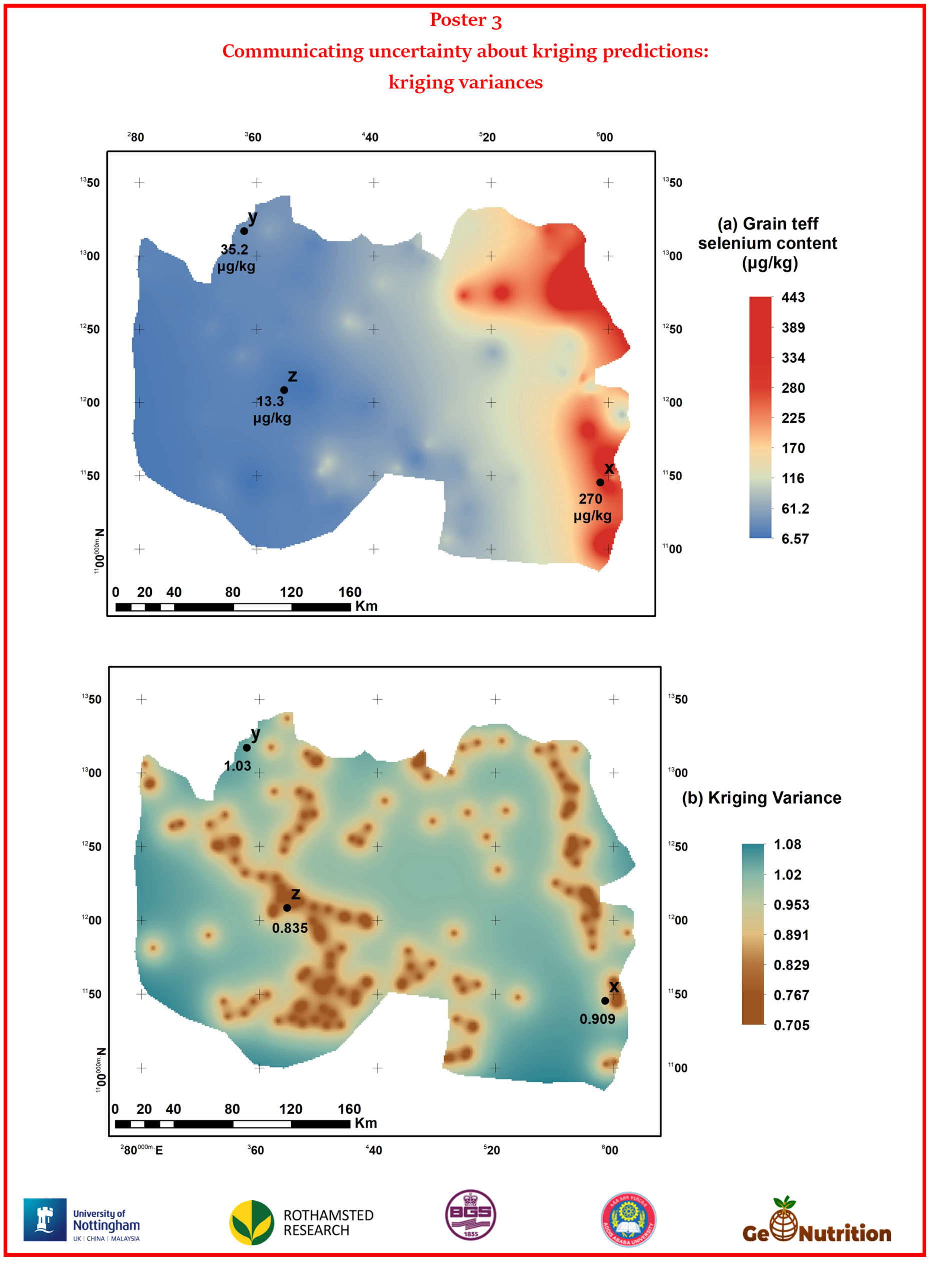


Figure S3. Poster showing (a) predicted selenium concentration in teff grain and (b) kriging variance (expected squared prediction error) in Amhara region, Ethiopia

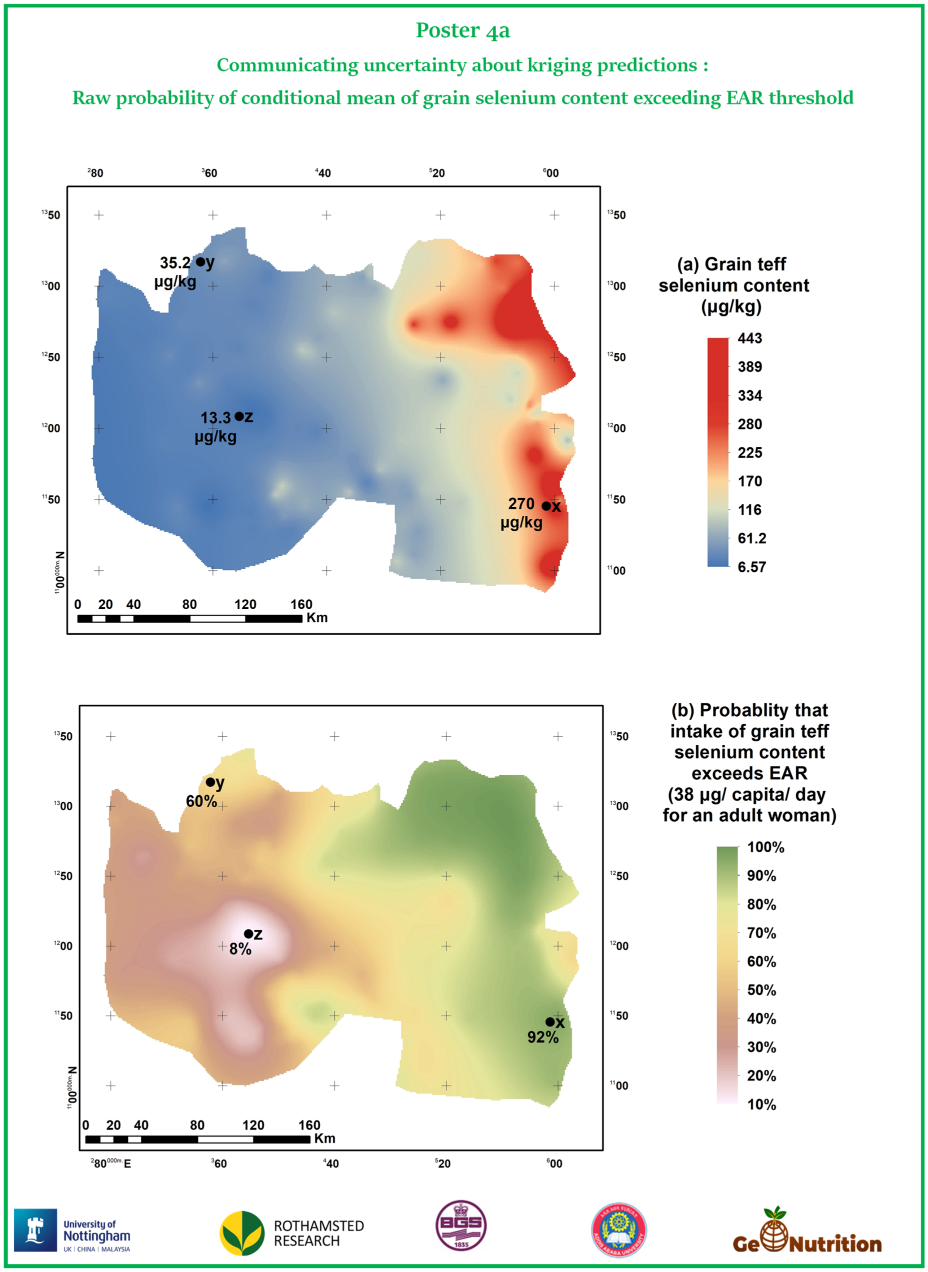


Figure S4. Poster on raw probability scale of the probability that intake of grain teff selenium concentration is less than the threshold 38 μg kg^{-1} in Amhara region, Ethiopia

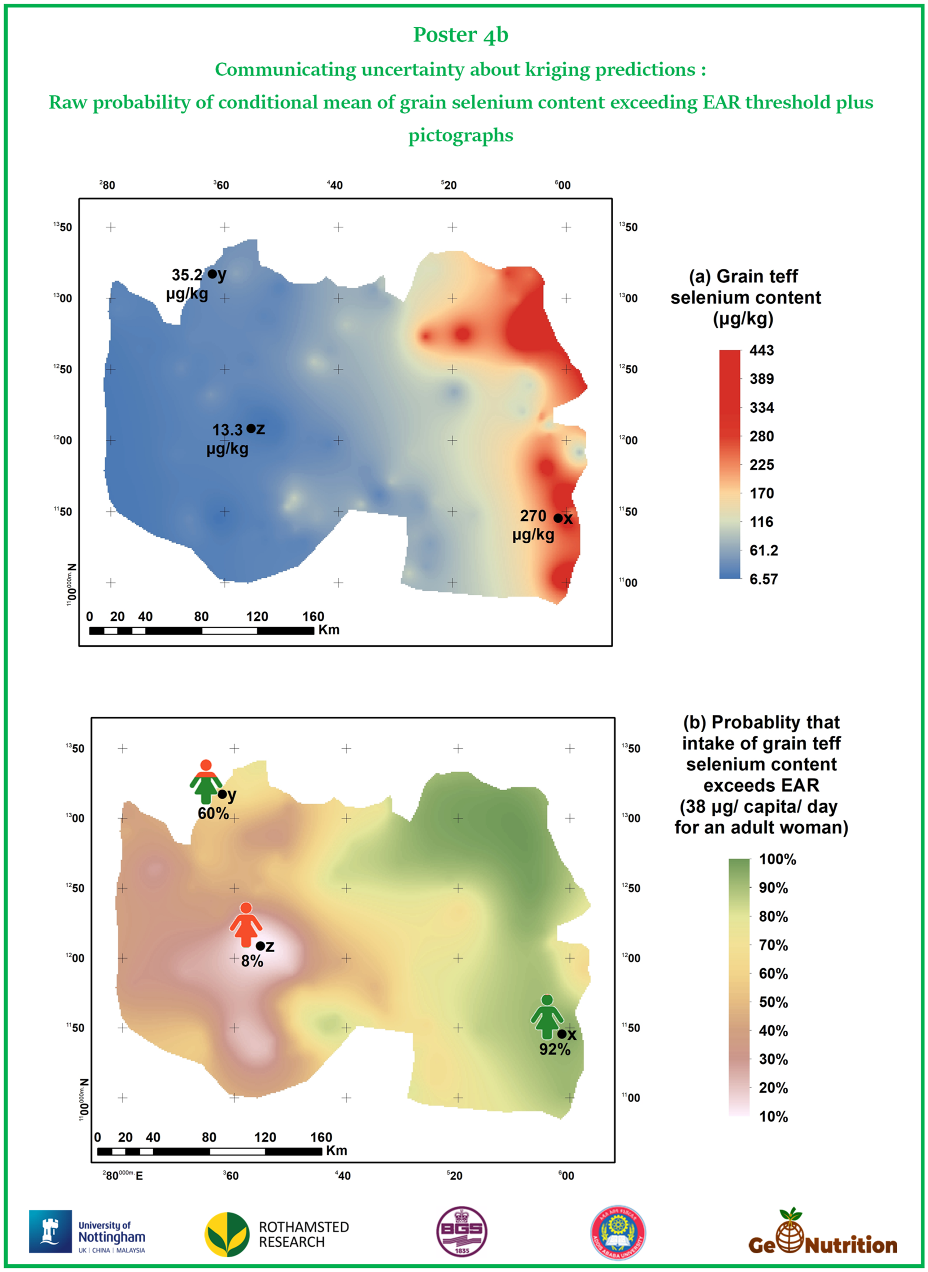


Figure S5. Poster on raw probability scale, with pictographs, of the probability that intake of grain teff selenium concentration is less than the threshold 38 µg kg^{-1} in Amhara region, Ethiopia

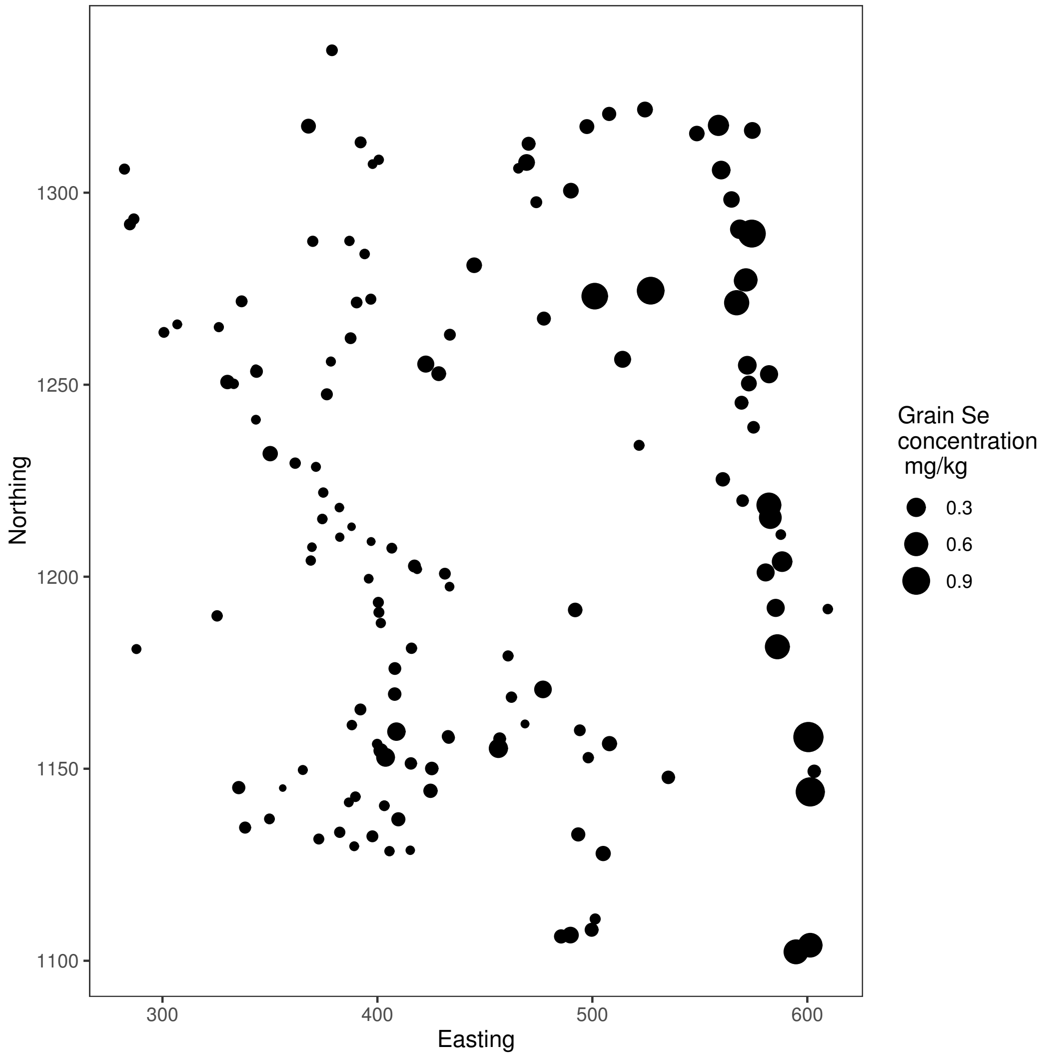


Figure S6. Post-plot of Amhara dataset

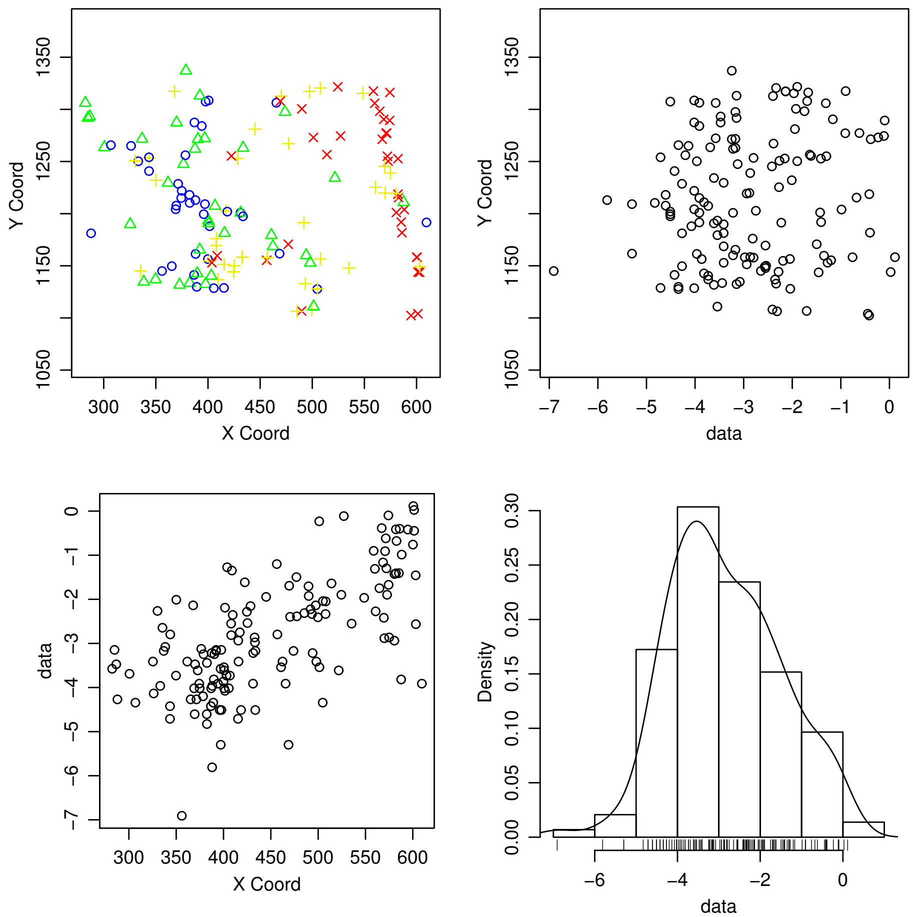


Figure S7. Post-plot of Amhara dataset

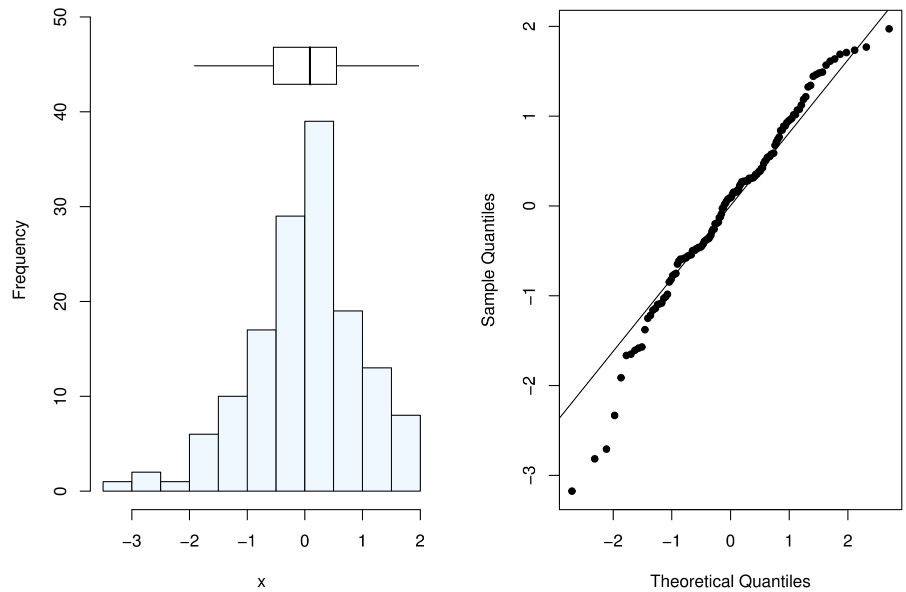


Figure S8. Post-plot of kriging errors in the cross-validated Amhara dataset

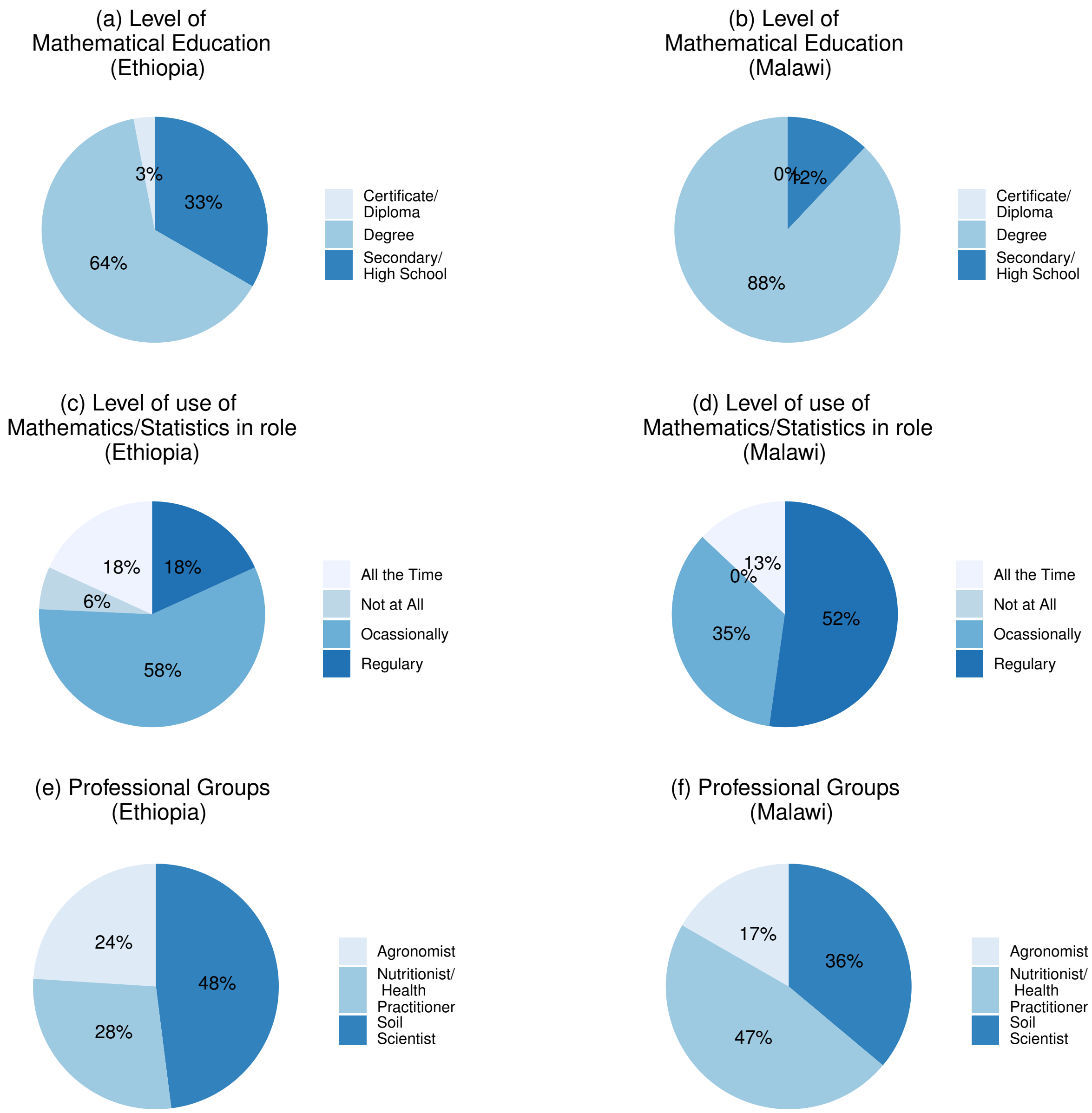


Figure S9. The percentage of participants by level of mathematical education and use of mathematics or statistics in their role.

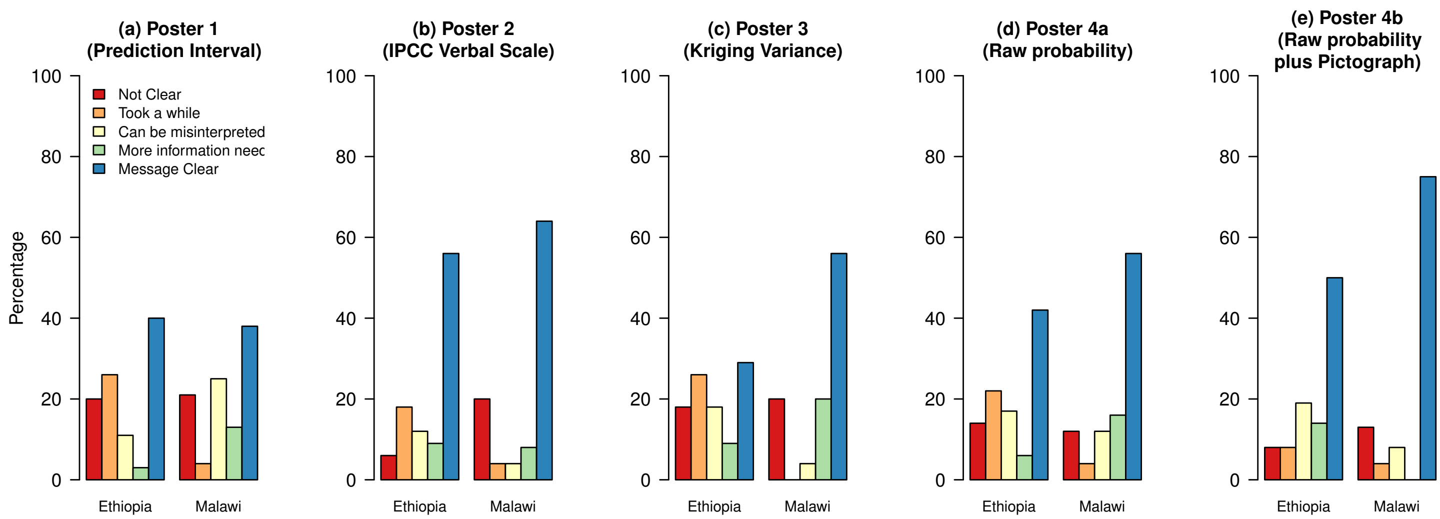


Figure S10. Bar charts showing how participants when pooled within location of meeting responded to the interpretive task on Question 2

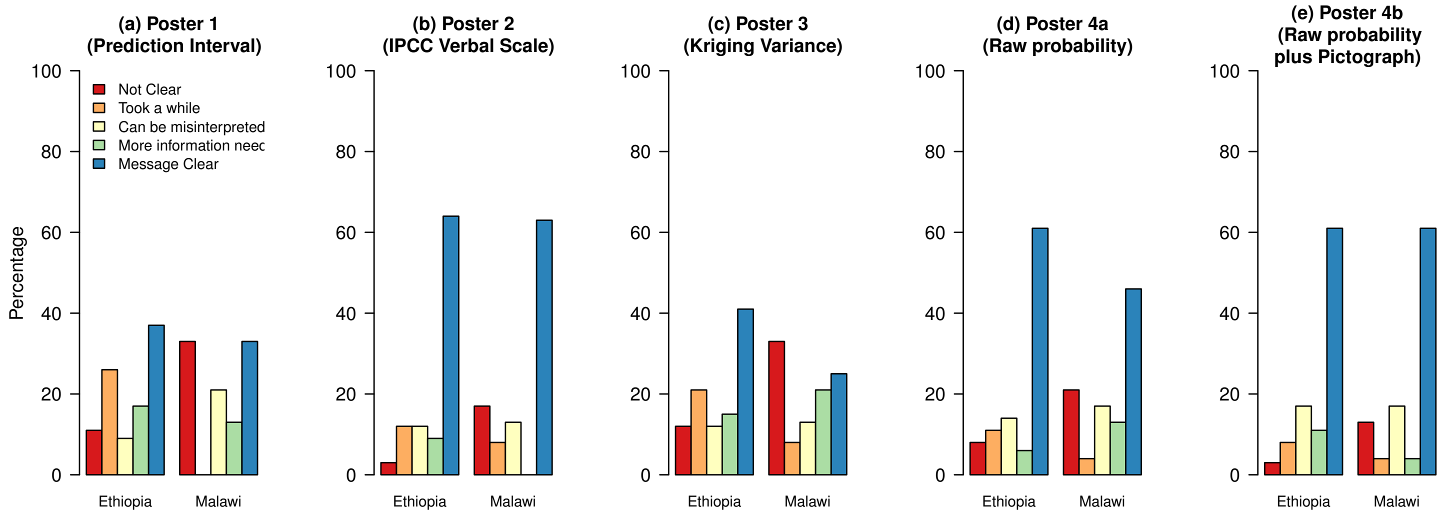


Figure S11. Bar charts showing how participants when pooled within location of meeting responded to the interpretive task on Question 3

Questionnaire

Purpose of Survey Spatial information is critical to many important decisions made by stakeholders in the area of food and nutrition, for example about whether and where interventions are required to address nutritional deficiencies. In this study we consider the example of information on micronutrient concentrations in staple crops. These concentrations vary spatially because of many factors. We can make direct measurements only at limited numbers of sites and use statistical models to make predictions elsewhere as a basis for mapping. Because of this the information presented in maps has attendant uncertainty. It is important that this uncertainty is communicated effectively to users of the information, and the objective of this exercise is to elicit information from stakeholder groups about the success or otherwise of different approaches to the problem.

This questionnaire aims to identify the best method(s) for communicating the uncertainty in spatial prediction of grain Se concentration. We hope to identify the most appropriate methods of communicating uncertainty for different groups, and so define the outputs we need from our uncertainty analysis.

We will show you five methods that could be used to communicate uncertainty. Please consider each in turn and answer the associated sets of questions. The two central questions ask:

1. Is the information that you need on uncertainty represented?
2. Is the method used to present uncertainty clear and not misleading?

Section A: Questions about you

1. Country where you work
2. Which group do you represent
   (a) Agronomist
   (b) Soil Scientist
   (c) Nutritionists/Health Practitioners
3. What level of mathematical education do you have?
   (a) Very Little
   (b) Secondary/ High school qualifications
   (c) Certificate/Diploma
   (d) Degree level and above
4. How much do you use mathematics or statistics in your role?
   (a) Not at all
   (b) Occasionally
   (c) Regularly
   (d) All the time

Section B: Questions about communicating uncertainty about spatial predictions of grain Se concentration

In all posters, the threshold Se concentration in grain to which we refer is 38 µg kg^{-1} (micrograms per kilogram), such that a serving of 330g of grain flour provides a third of the daily EAR of Selenium for an adult woman.

Poster 1

Please look at the poster and find locations x, y, and z

1. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration exceeds 38 μg kg^{-1} is greater at x than at z”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
2. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 μg kg^{-1} is greater at z than at y”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
3. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 μg kg^{-1} is greater at y than at x”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

   Suppose you would want to design programmes to address micronutrient deficiency in your region
4. Does the poster provide adequate information about the selenium content of grain for you to identify locations where programme is most needed?
   (a) Inadequate Information
   (b) Adequate information
   (c) More than what I wanted
5. Is the way this poster communicates the uncertainty about grain selenium content straightforward to interpret
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Poster 2

Please look at the poster and find locations x, y, and z

1. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration exceeds 38 µg kg^{-1} is greater at x than at z”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
2. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at z than at y”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
3. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at y than at x”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

   Suppose you would want to design programmes to address micronutrient deficiency in your region
4. Does the poster provide adequate information about the selenium content of grain for you to identify locations where programme is most needed?
   (a) Inadequate Information
   (b) Adequate information
   (c) More than what I wanted
5. Is the way this poster communicates the uncertainty about grain selenium content straightforward to interpret
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Poster 3

Please look at the poster and find locations x, y, and z

1. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration exceeds 38 μg kg^{-1} is greater at x than at z”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
2. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 μg kg^{-1} is greater at z than at y”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
3. Is it clear from the poster, that the statement below is true? “Our confidence that grain selenium concentration does not exceed 38 μg kg^{-1} is greater at y than at x”
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Suppose you would want to design programmes to address micronutrient deficiency in your region

4. Does the poster provide adequate information about the selenium content of grain for you to identify locations where programme is most needed?
   (a) Inadequate Information
   (b) Adequate information
   (c) More than what I wanted
5. Is the way this poster communicates the uncertainty about grain selenium content straightforward to interpret
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Poster 4a

Please look at the poster and find locations x, y, and z

1. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration exceeds 38 µg kg^{-1} is greater at x than at z"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
2. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at z than at y"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
3. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at y than at x"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

   Suppose you would want to design programmes to address micronutrient deficiency in your region
4. Does the poster provide adequate information about the selenium content of grain for you to identify locations where programme is most needed?
   (a) Inadequate Information
   (b) Adequate information
   (c) More than what I wanted
5. Is the way this poster communicates the uncertainty about grain selenium content straightforward to interpret
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Poster 4b

Please look at the poster and find locations x, y, and z

1. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration exceeds 38 µg kg^{-1} is greater at x than at z"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
2. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at z than at y"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear
3. Is it clear from the poster, that the statement below is true? "Our confidence that grain selenium concentration does not exceed 38 µg kg^{-1} is greater at y than at x"
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Suppose you would want to design programmes to address micronutrient deficiency in your region

4. Does the poster provide adequate information about the selenium content of grain for you to identify locations where programme is most needed?
   (a) Inadequate Information
   (b) Adequate information
   (c) More than what I wanted
5. Is the way this poster communicates the uncertainty about grain selenium content straightforward to interpret
   (a) No it is not clear at all
   (b) I understand it but took me a while to figure it out
   (c) I think it is good but can be misinterpreted
   (d) Good but needs more information
   (e) Message is clear

Comparing all methods

Once you have completed all the posters, which poster did you find easy to interpret and communicated uncertainty the best?

6. Do you think that the poster helped you understand the uncertainty in the predictions?
   (a) Poster 1
      i. Yes
      ii. No
   (b) Poster 2
      i. Yes
      ii. No
   (c) Poster 3
      i. Yes
      ii. No
   (d) Poster 4a
      i. Yes
      ii. No
   (e) Poster 4b
      i. Yes
      ii. No
7. Please rank the posters in order of their effectiveness, in your experience, at communicating uncertainty in the predictions, Rank1 being MOST effective and Rank 5 the LEAST
   (a) Rank 1 : Poster ....
   (b) Rank 2: Poster ....
   (c) Rank 3: Poster ....
   (d) Rank 4: Poster ....
   (e) Rank 4: Poster ....

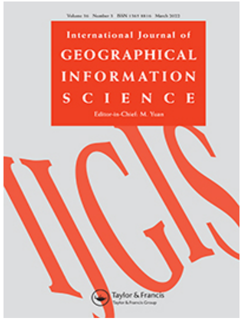
   Thank you for completing this questionnaire. If you have any further comments about the best ways to communicate uncertainty, please write below.

Chapter 3

Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops

This chapter has been published as:

Chagumaira, C., Nalivata, P. C., Chimungu, J. G., Gashu, D., Broadley, M. R., Milne, A. E., and Lark, R. M. (2022). Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops. *International Journal of Geographical Information Science*, 1–27, https://doi.org/10.1080/13658816.2021.2020278.



Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops

Christopher Chagumaira, Patson C. Nalivata, Joseph G. Chimungu, Dawd Gashu, Martin R. Broadley, Alice E. Milne & R. Murray Lark

To cite this article: Christopher Chagumaira, Patson C. Nalivata, Joseph G. Chimungu, Dawd Gashu, Martin R. Broadley, Alice E. Milne & R. Murray Lark (2022): Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2021.2020278

To link to this article: https://doi.org/10.1080/13658816.2021.2020278

Published online: 01 Mar 2022.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2021.2020278

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops

Christopher Chagumaira[a,b,c], Patson C. Nalivata[c], Joseph G. Chimungu[c], Dawd Gashu[d], Martin R. Broadley[a,b], Alice E. Milne[b] and R. Murray Lark[a]

[a]Future Food Beacon of Excellence and School of Biosciences, University of Nottingham, Loughborough, UK; [b]Sustainable Agriculture Sciences Department, Rothamsted Research, Hertfordshire, UK; [c]Crop and Soil Sciences Department, Bunda College, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi; [d]Center for Food Science and Nutrition, Addis Ababa University, Addis Ababa, Ethiopia

ABSTRACT

Spatial information, inferred from samples, is needed for decision-making, but is uncertain. One way to convey uncertain information is with probabilities (e.g. that a value falls below a critical threshold). We examined how different professional groups (agricultural scientists or health and nutrition experts) interpret information, presented this way, when making a decision about interventions to address human selenium (Se) deficiency. The information provided was a map, either of the probability that Se concentration in local staple grain falls below a nutritionally-significant threshold (negative framing) or of the probability that grain Se concentration is above the threshold (positive framing). There was evidence for an effect of professional group and of framing on the decision process. Negative framing led to more conservative decisions; intervention was recommended at a smaller probability that the grain Se is inadequate than if the question were framed positively, and the decisions were more comparable between professional groups under negative framing. Our results show the importance of framing in probabilistic presentations of uncertainty, and of the background of the interpreter. Our experimental approach could be used to elicit threshold probabilities which represent the preferences of stakeholder communities to support them in the interpretation of uncertain information.

ARTICLE HISTORY
Received 7 April 2021
Accepted 15 December 2021

KEYWORDS
Decision-making; framing effect; Interpret probability; values mean probability value; Uncertainty

CONTACT Christopher Chagumaira christopher.chagumaira@nottingham.ac.uk, chris.chagumaira@gmail.com

Supplemental data for this article can be accessed here.

1. Introduction

1.1. The problem

There is increasing awareness that, while much progress has been made to address malnutrition with respect to energy and protein supply, micronutrients (such as zinc, iron, iodine and selenium) may remain deficient among populations of many countries (Ligowe *et al*. 2020). This micronutrient deficiency (MND) or 'hidden hunger' has implications for human health, growth and cognitive function. In the GeoNutrition project, funded by the Bill and Melinda Gates Foundation, micronutrient studies in soil, crops and the human population are being conducted in Malawi and Ethiopia (Gashu *et al*. 2020, 2021). There is interest in how MND problems may vary spatially due to variation in soil and other environmental conditions. If this occurs, then interventions might be more effectively targeted where particular MND are prevalent.

Through the GeoNutrition project, a large dataset has been collected on soil and crop micronutrient status in Malawi and Ethiopia. This allows the micronutrient concentration in soil and staple crops to be mapped. The spatial predictions are uncertain, but the statistical models on which they are based allow us to compute the probability that a particular micronutrient concentration falls below or above a nutritionally relevant threshold at some unsampled location. It is often suggested that mapping this probability will help interpret the information while allowing for its uncertainty in the spatial data. However, it remains unclear how various stakeholders, for whom such information is required to support decisions on interventions to address MND, would use the probabilities in order to account for uncertainty.

In this paper we describe a study to examine how stakeholders interpret probability that local grain micronutrient concentration falls below a threshold. Groups of stakeholders were provided with different scenarios, in which this probability took different values, and were asked to indicate in which they would recommend an intervention (such as campaign to promote fertiliser to increase crop micronutrient concentration, or the deployment of nutrient supplements or fortified food). We used these responses to estimate and compare the mean probability value at which different stakeholder groups chose to recommend an intervention. We also examined how the framing of the question affected the responses. That is to say, whether the responses of stakeholders presented with a positive framing (probability that the grain Se content is sufficient) would be different to those who were presented information with negative framing (probability that the grain Se content is inadequate). On this basis we aimed to assess the feasibility of using formal elicitation to estimate the threshold probability at which groups of stakeholders would recommend an intervention, as a basis both for examining critically how they interpret probabilistic information and developing rules for interpretation which reflect stakeholder opinion and assumptions.

1.2. The general context

Spatial information has uncertainty, which arises from error (location error, measurement error), environmental heterogeneity, and our uncertainty about the interpretation of information (e.g. the vagueness of concepts such as a 'deep soil', which play a part in data interpretation) (Li *et al*. 2018). For this reason it is widely recognized in geographical information science (GIScience) that the uncertainty about spatial information must be communicated to its end-users if they are to apply it effectively (Li *et al*. 2012, Greiner *et al*. 2018). Heuvelink and Burrough (2002) suggested that it is necessary to address how stakeholders deal with problems of uncertainty in spatial information as part of a decision making process. The study reported here fits into that research agenda, and is concerned with how stakeholders make decisions based on comparison of spatial variables to threshold values when the uncertainty about the true value of the variable relative to the threshold is expressed in terms of probability.

A simple and common decision model is where some action is taken at a location if the value of a variable there exceeds (or falls below) a threshold. For example, action must be taken to remediate soil where the concentration of a contaminant exceeds a soil guideline value (Cole and Jeffries 2009) or legislative thresholds (Marchant *et al*. 2017). Fertilisers might be recommended where the measured concentration of a nutrient in soil is smaller than an index value and liming might be recommended where soil pH is less than a threshold. For example, in Malawi, it is recommended that liming should be done when soil pH is below 5.0 (Chilimba *et al*. 2013); whereas, in the UK if soil pH falls below 6.0 in pasture land, then liming is recommended to maintain yield and forage quality (DEFRA 2010). Interventions to address micronutrient deficiencies in human populations can be recommended where measurements of a biomarker (such as concentration of the nutrient in blood serum or urine) falls below a threshold (e.g. Likoswe *et al*. 2020, Phiri *et al*. 2020) or where inferred intake is less than a quantity such as the recommended daily allowance (RDA) or estimated average requirements (EAR) (e.g. Joy *et al*. 2014, 2015).

Such management decisions are usually made in the face of uncertainty because the variable concerned is estimated or predicted from partial data or a model (Goovaerts 1997). Spatial uncertainty can be quantified in a number of ways. In geostatistical mapping, the spatial uncertainty of the predictions is quantified directly by the prediction error variance or the kriging variance. The kriging variance varies spatially, and its values are small in the neighbourhood of sample points and larger further away. The kriging variance is the variance of the prediction distribution at an unsampled site of interest, or the conditional distribution given the data and the geostatistical model. The width of this prediction distribution (indicated by its variance) represents the uncertainty of the predicted value there (Heuvelink 2018). The kriging variance might be mapped directly as an indicator of uncertainty (e.g. Hatvani *et al*. 2021). Alternatively, it might be more accessible to compute prediction intervals from the prediction distribution, that is to say an interval of values which contains the true value at the location with some specified probability (e.g. Karl 2010). These methods are useful to experts familiar with the underlying concepts, but may be inaccessible for decision makers who do not necessarily understand kriging variance. Prediction intervals and kriging variance were the methods of communicating and quantifying uncertainty least-preferred by end-users (Chagumaira *et al*. 2021).

When there are decisions to be made relative to thresholds, spatial uncertainty can be quantified by using probabilities. This uncertainty can be quantified by the probability that the threshold is exceeded or not. Ideally this probability can be obtained from the prediction distribution of the variable from data and an appropriate statistical model (conditional probability). Marchant *et al*. (2017) took this approach to compute probabilities that arsenic and mercury concentration exceeds soil guidance values and to map this across France. Lark *et al*. (2014) similarly computed the probability that local soil conditions indicate a risk of cobalt deficiency in grazing sheep across part of the north of Ireland. Approaches such as disjunctive kriging (DK) and indicator kriging (IK) are commonly used to compute conditional probabilities (Webster and Oliver 2007). Ordinary kriging may be used, along with an assumption of normal errors. However, indicator kriging is more robust to any failures of this assumption, and is also more resistant to local outliers. Lark *et al*. (2016) used DK to map the probability that soil pH under pasture in the north of Ireland is below 6.0, to indicate where liming would be advised. Goovaerts *et al*. (1997) used IK to map the probability that cadmium concentration exceeds a regulatory threshold at sites across the Swiss Jura, to indicate where remediation might be necessary. Other approaches have been used to compute local probabilites that variables exceed thresholds of environmental significance. These include copulas, conditional simulation and Bayesian methods to compute or sample from a local posterior distribution (Goovaerts 2001, Marchant *et al*. 2011, Greiner *et al*. 2018).

Much work has focused on computing the conditional probability that a variable exceeds a threshold, and there is an implicit assumption that if the stakeholder has been given the probability they will be able to use it to make decisions with the uncertain information (Lark *et al*. 2016). Little attention has been given to how stakeholders might use such information and how they might be helped to do so more consistently and effectively. The use of probability to communicate uncertainty is not straightforward (Milne *et al*. 2015) and probabilities are not always easily interpreted by stakeholders who have to make the decision (Spiegelhalter *et al*. 2011). Because of this, verbal interpretations of probability based on 'calibrated phrases' (e.g. 'unlikely') have been proposed — e.g. the Intergovernmental Panel for Climate Change (IPCC) scale due to Mastrandrea *et al*. (2010). Although calibrated phrases have been widely used, Budescu *et al*. (2009) showed that they may be interpreted regressively (i.e. any phrase indicating uncertainty about an outcome is thought to indicate that its probability is around 0.5). Furthermore, calibrated phrases may be subject to severity bias, depending on how the outcome of interest is expressed (e.g. if it is stated that 'severe flooding is very unlikely' the adjective 'severe' influences the assessment of risk more than does the phrase indicating the uncertainty). However, Jenkins *et al*. (2019) showed that stakeholders regard probabilities expressed in numerical form as more credible than calibrated phrases. Chagumaira *et al*. (2021) found that, despite these challenges in interpretation of probabilities, varied stakeholders preferred statements of uncertainty expressed as probabilities to more general measures such as prediction intervals or a prediction error variance.

Spatial uncertainty is an important subject in GIScience (Heuvelink and Burrough 2002, Li *et al*. 2012, 2018) and presenting spatial datasets together with their uncertainties is necessary because it adds to the quality of spatial information used in decision making. As we have noted, a common approach to presenting uncertain information about the value of a variable relative to a threshold is to compute the probability that the variable exceeds (or falls below) that threshold. However, we contend that insufficient attention has been given to how stakeholders incorporate such uncertain information into decision-making processes.

A stakeholder, using uncertain information to support a decision, must in effect decide on the probability threshold at or above which they would choose to act as if the threshold was exceeded/not exceeded. Taking the concentration of Se in staple grain as an example, would a stakeholder approve an intervention at a certain location where there were a 50% probability that the concentration of Se falls below the threshold? Would they make the same decision if the probability were 25%, or 75%?

A stakeholder deals with an unknown state, the true value of the environmental variable either indicates that the action should be taken or it does not. They also have a choice of two actions to intervene or not. We might expect that the threshold probability at which a stakeholder would choose to intervene will reflect their assessment of the loss attached to each possible outcomes—the intervention was necessary or not, as determined by the unknown state, under each decision (intervene or not). These losses may reflect factors such as the social, economic, individual and political consequences of failing to address a problem, and the opportunity costs of resources expended on unnecessary intervention. In some cases these losses may be quantified, and used in a formal analysis e.g. Ramsey *et al.* (2002) who considered the losses associated with different decisions and outcomes in the management of contaminated land. However, for many applications the different losses under decisions and outcomes may be complex and hard to quantify. The question that we address in this paper is how and whether one might identify a threshold probability that consistently reflects the perception of the losses by a stakeholder group, and how they weight these, tacitly if not explicitly. Before refining this question, we consider a theoretical framework.

1.3. Theory

Let L_1 be the loss incurred if we intervene unnecessarily, where with perfect knowledge we would intervene only if the variable (nutrient concentration) $z<z_t$, where z is the unknown true value and z_t is the threshold of interest. In this treatment we regard the loss as zero if we intervene appropriately. Let L_2 be the loss incurred if we choose not to intervene, but should have done so. Again, we regard loss as zero if we correctly choose not to intervene. If P is the probability that the concentration is below the threshold, $z<z_t$, then expected loss if we choose to intervene is

$$(1-P)L_1. \tag{1}$$

If we choose not to intervene then the expected loss is

$$PL_2. \tag{2}$$

If we wish to make the decision with the smaller expected loss, a rational assumption, then it follows that we should intervene if P takes a value such that

$$(1-P)L_1 \leq PL_2, \tag{3}$$

and not intervene otherwise. By simple algebraic rearrangement of Equation (3) we can show that we should intervene if

$$P \geq \frac{L1}{(L_1 + L_2)}, \tag{4}$$

and not otherwise, that is to say if P exceeds or equals a threshold value, P_t where,

$$P_t = \frac{L_1}{L_1 + L_2}. \tag{5}$$

The larger the loss from an unnecessary intervention relative to a failure to intervene where necessary, the larger P_t must be.

In a situation where L_1 and L_2 can be quantified directly, P_t could be computed from Equation (5). However, complex real-world problems components of the loss associated with outcomes maybe difficult to quantify (e.g. the political cost of a failure to address a public health problem) and controversial (e.g. do disability adjustment life years, DALYs, lost really capture all the social loss from a failure to act where a nutritional deficiency pertains?) and may not be commensurable. The value of P_t at which an agent chooses to act therefore reflects a complex judgement.

This study is based on two principles. First, while the provision of conditional probabilities is a natural way to communicate the uncertainty associated with the information of a variable which users of that information will interpret relative to threshold values, the problem of decision-making is not solved by those probabilities. As we have seen, a judgement must still be made. Second, we suggest that one approach to this problem is to elicit a threshold probability from members of relevant stakeholder communities. We assume that an individual stakeholder has a least a tacit sense of the values of L_1 and L_2 that they would assume in making a judgement from conditional probabilities. In principle, then, a suitable process might be used to elicit a value of P_t from individuals or groups of stakeholders that represent an individual opinion or a group consensus. Such an elicitation would be analogous to the process by which probabilities of unknown states or distribution for uncertainty quantification are formally elicited from expert panels (O'Hagan *et al.* 2006).

The aim of the study reported here was to address the following:

- *Can a consistent (i.e. reasonably precise) estimate of P_t be elicited from a stakeholder group?*
- *Does the estimated P_t depend on the specific interests of the group (e.g. does it differ between nutritionists and agronomists)?*
- *Is the estimated P_t prone to framing effects (i.e. does the estimate depend on how the question is posed)?*

These are practical and useful questions to address. If decisions are to be based on uncertain information then a value of P_t is required for a decision making and should be obtained by some transparent process in which the underlying questions are examined. The findings of this study should provide a basis for designing a formal procedure to elicit a value of P_t for this and similar problems. In this study we address these questions, considering a core study concerned with decision on interventions to improve micronutrient supply based on estimates of the amount provided locally by staple crops. We asked two stakeholder groups individually to identify a threshold probability at which an intervention would be recommended, and used these to estimate an underlying mean value for each group. Furthermore, we investigated whether the framing of the question influenced the responses.

2. Method

2.1. Basic approach

The approach was to offer respondents a set of scenarios for which the probability that concentration of Se in staple crop is less than the threshold Se concentration ($\mathrm{Se_{grain}} < t_{\mathrm{Se}}$) took a series of values over the range 0–1. For each one they were invited to respond as to whether the intervention would be recommended or not.

The respondents were asked to self-identify as either (i) A public health and nutrition specialist, or (ii) an agronomists and soil scientist. Each respondent was also allocated at random to one of two groups. The first group was presented with a positive framing of the question (i.e. to select a probability that $\mathrm{Se_{grain}} > t_{\mathrm{Se}}$ below which an intervention would be recommended). The second group was presented with a negative framing of the question (i.e. to select a probability that $\mathrm{Se_{grain}} < t_{\mathrm{Se}}$ above which an intervention would be recommended).

More detail on the practical organization of the experiment is given in section 2.2. The threshold Se concentration, t_{Se}, in grain to which we referred is 38 μg kg^{-1}, such that a serving of 330 g of grain flour provides a third of the daily EAR of Se for an adult woman. We used EAR because it is one of the commonly-used measure of intake when assessing nutritional status and planning intervention.

The respondents were presented with probabilities that Se concentration in grain falls below or above a threshold from specific locations on maps of Amhara, Ethiopia or Malawi dependent on the location of the particular session. These maps were derived by indicator kriging (see Webster and Oliver 2007) from data collected in the GeoNutrition project (Gashu *et al*. 2021). Indicator kriging was used because it requires no specific assumption that the kriging errors are normally distributed (Rivoirard 1994). More detail on this is provided by Chagumaira *et al*. (2021). Note that the grain samples in this project, in both Ethiopia and Malawi, were collected on a consistent sample support: a 0.1-ha circular plot in the centre of the sampled field. The probabilities therefore relate to mean values of grain concentration across such a support within a field at a specified location.

2.2. Organization of the experiment

The experiment was done in two sessions at Lilongwe, Malawi (November 2019) and Addis Ababa, Ethiopia (January 2020). Ethical approval to conduct this study was granted by the University of Nottingham School of Sociology and Social Policy Research Ethics Committees (BIO-1920-004 for Malawi, and BIO-1920-007 for Ethiopia), as approved by Lilongwe University of Agriculture and Natural Resources (LUANAR), and Addis Ababa University (AAU).

Table 1. Composition of different professional groups during the experiment in Ethiopia and Malawi.

Professional group	Location: Ethiopia	Location: Malawi	Total
Agronomist	4	5	9
Soil scientist	12	13	25
Public health and nutrition specialist	12	5	17
Total	28	23	51

We invited participants from among professionals working in agriculture, nutrition and health, at NGOs, universities and government departments from Ethiopia, Malawi and in the wider GeoNutrition project. Recruitment was undertaken by the local GeoNutrition Project team. In total we had 51 participants, 34 were agronomists and soil scientists and 17 were public health and nutrition specialists, see Table 1.

In each workshop, we started by randomly allocating participants to one of two groups one for positive framing and the other for negative. This was done by asking each participant to draw a shuffled card from a pot of cards bearing group labels. Cards were not replaced. We did not explain why we were grouping them until after the exercise had been completed.

We presented the first group with a map of probability that $\mathrm{Se}_{\mathrm{grain}} > t_{\mathrm{Se}}$. The locations were identified on the map, and at each probability that $\mathrm{Se}_{\mathrm{grain}} > t_{\mathrm{Se}}$ was also illustrated by a pictograph (see Figure 1(a)). The questions were targeted to their areas of expertise. Specifically, agronomists and soil scientists were asked to decide whether or not they would recommend an intervention to provide and promote Se-fortified fertiliser. The public health and nutrition specialists to decide whether or not they would recommend a programme to provide Se-fortified food at that site. In both cases we asked the participants to assume that checks would be undertaken before the intervention took effect to ensure that no one was exposed to toxic levels of Se. The map showed nine locations, labelled **a**, **b**, **c**, **d**, **e**, **f**, **g**, **h** and **i**, at which probability that $\mathrm{Se}_{\mathrm{grain}} > t_{\mathrm{Se}}$ was 7%, 25%, 33%, 41%, 58%, 76%, 82%, 92%, 99%, respectively.

For each location in turn and by referring to the probability (as shown on map with pictograph, and explicitly stated in words), each participant recorded in a questionnaire whether or not they would recommend an intervention at the site given the probability. Using location **a** as an example, we phrased our question as follows: *'At site* **a** *there is 7% probability that the concentration of grain Se concentration exceeds the threshold, would you approve this intervention?'* We chose a range of probabilities giving coverage of the interval [0,1] so as not to limit the responses participants could give.

When the first group had completed filling in the questionnaires we invited participants from the second group into the room. To this group we presented a map of probability that $\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}$. At each location, probability that $\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}$ was also illustrated by a pictograph (see Figure 1(b)). The map showed the same nine locations but with 93%, 75%, 67%, 59%, 42%, 24%, 18%, 8%, 1% probability that $\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}$. The participants answered the same questions as the first group, for the same location, but with a negative framing. For example, we asked them, *'At site* **a** *there is 93%*

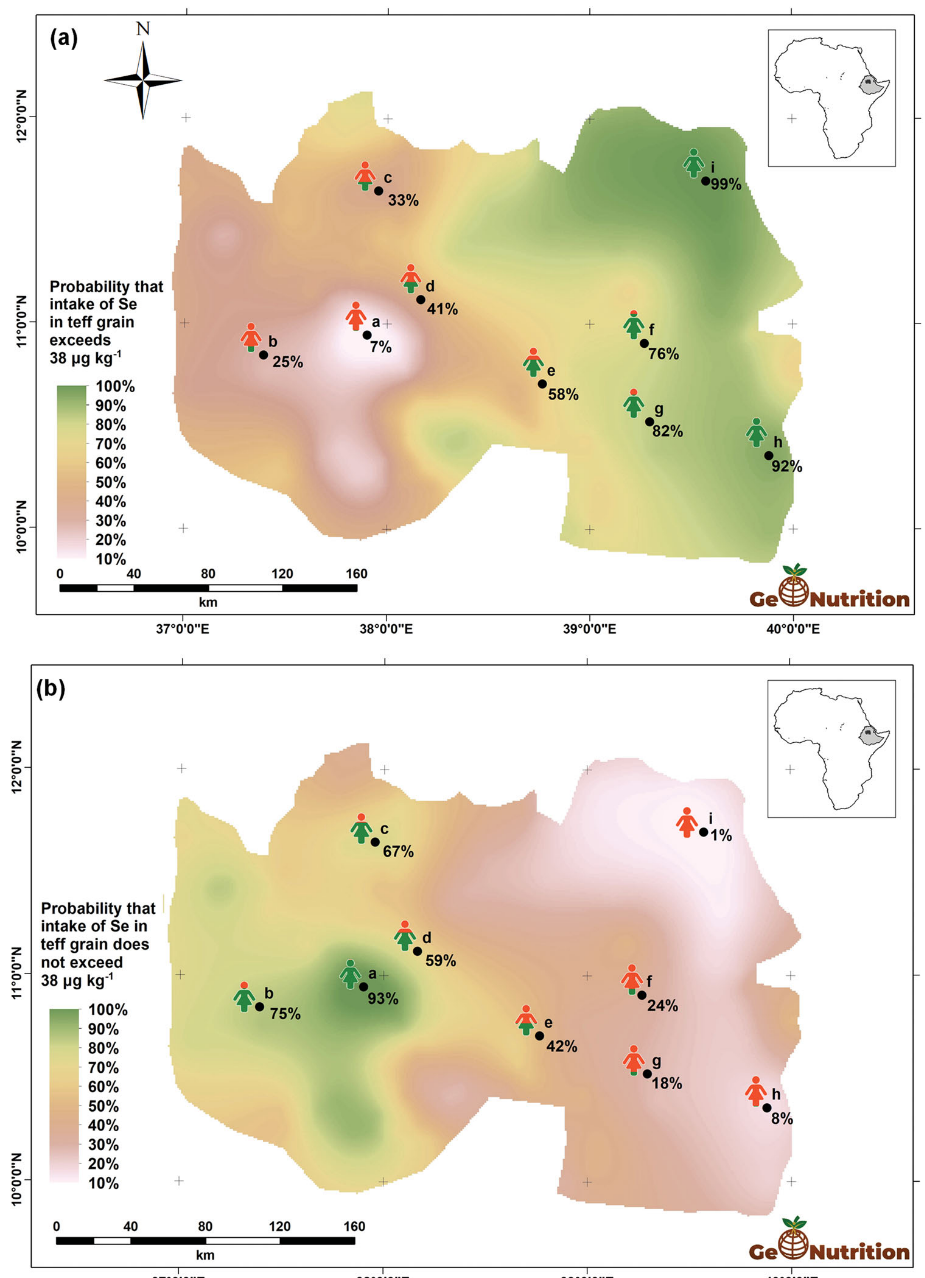


Figure 1. (a) Probability that concentration of Se in teff grain is greater than 38 µg kg^{-1} ($Se_{grain} > t_{Se}$) in Amhara region, Ethiopia. This was presented to the first group, with a positive framing of the question. The locations labelled **a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**, and **i** at which probability that $Se_{grain} > t_{Se}$ is also illustrated with a pictograph. (b) Probability that the concentration of Se in teff grain is less than 38 µg kg^{-1} ($Se_{grain} < t_{Se}$) in Amhara region, Ethiopia. This was presented to the second group of participants, with a negative framing of the question. The locations labelled **a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**, and **i** at which probability that $Se_{grain} < t_{Se}$ is also illustrated with a pictograph.

probability that the concentration of grain Se concentration does not exceed the threshold, would you approve this intervention?'

Participants did this exercise independently, and were asked not discuss the questions with each other until they had completed the exercise. In the introduction to this exercise, it was pointed out to the participants that errors could go on both directions, resulting in an intervention where it was not needed (error of commission), or failing to intervene where the nutritional supply from staple foods was inadequate (error of omission). We encouraged participants to consider the sources of losses under errors of commission or omission. For example, the agronomists and soil scientists group should consider the costs of buying Se-enriched fertilisers especially given that Se does not improve crop yield. For public health and nutrition specialists, there would be costs associated with failing to intervene when there is need because of increased risk of health complications and mortality especially with people with compromised immunity due Se deficiency (e.g. thyroid disfunction and suppressed immune response), but that unnecessary interventions are likely to represent a loss as resources are used which could address other public health initiatives. However, we did not ask the participants to attempt to calculate any of these costs. Rather, the aim was that having considered the possible outcomes, they should make a judgement in the light of their experience. This would be expected to reduce any framing effect (Almashat *et al*. 2008). When both groups had completed the exercise, we brought them together and we then explained the objectives of the exercise and background of the loss functions.

2.3. Model and analysis

The following sections describe the statistical methodology used in this paper to analyse the data from the experiment. We summarize the methods briefly here for the benefit of readers for whom the mathematical content is of limited interest. We propose a statistical model for a set of responses to the questionnaires. Under the model any individual respondent is assumed to advocate intervention once the probability that grain Se concentration is less than 38 μg kg^{-1} exceeds some value p'. We assume that the values of p' for a set of respondents can be treated as a random variable with a Beta distribution, a distribution particularly suited to modelling values which are constrained on an interval, and able to accommodate a wide range of behaviours. The two parameters of the Beta distribution can be estimated for a set of observations by a maximum likelihood method. Of interest is an estimate of the mean of the distribution, which we refer to as P_t, the expected value of, p' for an individual from the population of which the set of respondents is a sample. The maximum likelihood estimation allows us to evaluate evidence that, for example, it is necessary to model the responses from positive or negative framing with different parameter sets. This is done by means of the log-likelihood ratio test to compare a null model (in which responses with the two framings are pooled) with an alternative (in which distinct parameters are estimated for each framing). We used this approach to test the effect of framing, location (Ethiopia or Malawi), and professional group (agronomists and soil scientists or public health and nutrition specialists).

Having explored the data by modelling we decided that we wished to estimate the mean value P_t for all professional groups and locations pooled, for the responses to the negatively framed question. We did this by Bayesian estimation, using very uninformative prior distributions for the Beta parameters (that is, priors that have very little influence on the posterior distribution, which is dominated by the data.

2.3.1. Form of the data and their interpretation

Our data are a set of responses to questions, asking whether an intervention would be recommended in a situation given the probability that Se concentration in grain exceeds a nutritionally-significant threshold (positive framing) or is below the threshold (negative framing). The probabilities were expressed as percentages. Let the ordered set of percent probabilities (negatively framed) be $\{P_1, P_2, \ldots, P_m\}$. The positively-framed question set was directly equivalent, referring to the same scenarios, and so the percent probabilities presented with the positively framed questions were $\{100 - P_m, \ldots, 100 - P_2, 100 - P_1\}$.

For purposes of analysis the probabilities were scaled to $[0, 1]$, and the positively-framed probabilities were converted to the equivalent probability that $\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}$. We denote these probabilities by $\{p_1, p_2, \ldots, p_m\}$.

A response to the question is deemed to be consistent only if the respondent indicated that, for some $i \in \{1, 2, \ldots, m\}$, an intervention should be considered for all scenarios where the probability that $\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}$ was greater than or equal to p_i, and that the intervention should not be considered otherwise. If a response was not consistent in this sense, then it was discarded. Our data therefore comprise a set of n index values, ϱ, where $\varrho[j] = i$ if the j^{th} respondent stated that interventions would be recommended in all cases where $P(\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}) \geq p_i$. Of the 51 responses five were inconsistent (for example, the respondent recommended an intervention in a case where the probability of deficiency took some value, but did not recommend it in cases with both larger and smaller probabilities of deficiency). Three of the responses were anomalous, the respondent advocated an intervention for cases with a small probability of deficiency, and did not recommend intervention in cases with a large probability of deficiency. These 8 returns were discarded, leaving 43 for analysis, but they do illustrate the difficulties that stakeholders can have with the interpretation of probabilities.

We assume that each respondent has a latent 'personal' probability, p' such that, given all available information, they would advocate an intervention at a site where $P(\mathrm{Se}_{\mathrm{grain}} < t_{\mathrm{Se}}) \geq p'$. Furthermore, we assume that, if the respondent indicates that an intervention should be recommended for all scenarios in the set for which the probability equals or exceeds p_i, then the lower and upper bounds on p' are given by

$$\begin{aligned} l_i &= \frac{p_i + p_{i-1}}{2} \quad & i \neq 1, \\ &= 0 \quad & i = 1, \end{aligned} \tag{6}$$

and

$$\begin{aligned} u_i &= \frac{p_i + p_{i+1}}{2} \quad & i \neq m, \\ &= 1 \quad & i = m. \end{aligned} \tag{7}$$

2.3.2. The statistical model and its estimation

We assume that the distribution of p' within any group of respondents has a Beta distribution, such that the probability density function for some value $x \in [0, 1]$ is given by

$$f_\beta(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \qquad 0<x<1, \\ = 0 \qquad\qquad \text{otherwise,} \tag{8}$$

where

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

and $\Gamma(\cdot)$ denotes the gamma function. The Beta distribution is particularly appropriate for modelling probabilities as random variables, because a Beta random variable is continuous but constrained to a fixed interval (here [0,1]), and it is very flexible, accommodating a wide range of behaviours: bell-shaped, symmetrical with large or small kurtosis, uniform, strongly positively or negatively skew, straight-line or U-shaped (Tjims 2018).

The parameters of the gamma distribution are α and β but a convenient reparameterization (because of the correlation of these parameters) is to the mean U and a precision parameter V which is smaller the more dispersed the distribution of x;

$$U = \frac{\alpha}{\alpha+\beta}, \tag{9}$$

and

$$V = \alpha + \beta. \tag{10}$$

We denote the probability density function for some set of parameters $\boldsymbol{\theta} = \{U, V\}$ by $f_\beta(x|\boldsymbol{\theta})$ (McDonald and Xu 1995).

If the value of p' for the jth respondent can be regarded as a Beta random variable with probability density function (PDF) $f_\beta(x|\boldsymbol{\theta}_k)$ then the probability of observing $\varrho[j] = i$ can be obtained as the integral of the Beta PDF over the limits l_i and u_i:

$$\text{Prob}\{\varrho[j] = i\} = \int_{l_i}^{u_i} f_\beta(x|\boldsymbol{\theta}_k)\mathrm{d}x. \tag{11}$$

If we treat all our respondents as members of a single population of interest, then the log-likelihood for a proposed set of parameters $\boldsymbol{\theta}$ for that population can be obtained by computing, for each entry in ϱ the probability for the observed value of i by evaluating Equation (11). The sum of the logarithms of these probabilities gives the log likelihood. A maximum likelihood estimate of $\boldsymbol{\theta}$ can be found numerically, as described below.

For our purposes we want to estimate models for our observations which assumes that there are different sub-populations from which the they are drawn, and that different values of the Beta parameters may be estimated for such a sub-population. For example, we might choose to fit a model in which we assume that all responses from individuals who were presented with information with positive framing are drawn from a sub-population with a set of Beta parameters, and that those responses where the framing was negative constitute a second sub-population. The likelihood, as described above, must be extended to this more complex model.

Consider a set of responses from a group of n subjects. The subjects can each be assigned to one of Q sub-populations, and our hypothesis is that a particular set of values of the parameters $\boldsymbol{\theta}_k = \{U_k, V_k\}$ can be proposed for the k^{th} sub-population. We denote the full set of Q parameters by $\boldsymbol{\Theta} = \left[\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_Q^{\mathrm{T}}\right]$.

Given the assumptions set out in Equations (6) and (7) above, the log-likelihood for proposed values of the parameters $\boldsymbol{\Theta}$, given a set of n responses can be obtained as

$$\updownarrow(\varrho; \boldsymbol{\Theta}) = \sum_{k=1}^{Q} \sum_{j=1}^{n} \sum_{i=1}^{m} I_{k,j,i} \log \int_{l_i}^{u_i} f_\beta(x|\boldsymbol{\theta}_k) \mathrm{d}x, \tag{12}$$

where $I_{k,j,i}$ is an indicator variable which takes the value 1 if $\varrho[j] = i$ and the j^{th} respondent belongs to the k^{th} sub-population of respondents. In all other cases $I_{k,j,i} = 0$. This indicator variable allows us to simplify the notation. The three nested summations implies that we compute the log of the probability for every sub-population parameter set over every set of bounds for each observation, but the indicator takes the value zero for any combination where the jth respondent is not in the kth sub-population, and $\varrho[j] \neq i$. Equation (12) therefore allows us to compute the log likelihood for a proposed set of Beta parameters, $\boldsymbol{\Theta}$ for a corresponding model of a set of responses.

In this study we found maximum likelihood estimates of the parameters $\hat{\boldsymbol{\theta}}_k, k \in \{1, 2, \ldots, P\}$ which minimized $-\ell(\varrho; \boldsymbol{\Theta})$ given the data in ϱ. This was done using the *optim* function in base R (R Core Team 2020), using the default optimizer which is the simplex algorithm of Nelder and Mead (1965).

A series of nested models were fitted to the data. In the first, model M_0, all respondents were considered as a single population. In the second, model M_1, respondents who were presented with a negative framing were treated as a distinct sub-population from respondents presented with a positive framing. These two models were compared by computing the log-likelihood ratio statistic:

$$L = 2(\ell_{M_1} - \ell_{M_0}), \tag{13}$$

where ℓ_{M_1} and ℓ_{M_0} denote the maximized log-likelihood for models M_1 and M_0 respectively. Under a null-hypothesis where the parameters for the two sub-populations can be regarded as equal (as in M_0, termed the 'null model') L is asymptotically distributed as $\chi^2(2)$, the degrees of freedom being equal to the number of additional parameters in M_1 relative to M_0.

Further models were considered in which sub-populations were defined by (i) the location of the experiment and (ii) the broad professional group, both tested with the groups with positive and negative framing. The first of these was considered in case there were some differences in the way the meetings in two locations were conducted. Differences could also be due to composition of the participants group (see Table 1), we had fewer public health and nutrition specialists in the Malawi meeting. For familiarity and engagement, we used a probability map from Ethiopia's Amhara region in the experiment in Ethiopia and a map of Malawi in the experiment in Malawi. The comparison between the groups (agronomists and soil scientists or public health and nutrition specialists), was considered to test the hypothesis that cultural differences between the two professional groups contribute to differences in sensitivity to the framing effect, and in the relative weighting of the cost of errors of commission and omission.

2.3.3 Bayesian estimation

After examining the alternative models described in the previous section, it was decided to make a final estimate of the mean value of U for all respondents (both locations and professional groups) within the sub-sets presented with negative framing. A Bayesian approach was taken for this final step so as to quantify uncertainty in the parameter estimates without the assumptions of linearity required in methods based on the information matrix or the assumption that estimation errors are normal (Spiegelhalter and Rice 2009).

The Bayesian approach requires prior distributions for the parameters U and V. A uniform prior distribution over (0, 1) was assumed for U. This is entirely uninformative about the parameter. The prior for V was a gamma distribution with parameters (1, 20). This is a weakly informative prior, so the posterior distribution is dominated by the data.

The prior predictive density for the data was obtained by integrating out the parameters, this was done with the *adaptIntegrate* function from the cubature library for the R platform (Narasimhan *et al*. 2020). The posterior joint density of U and V is then straightforward to evaluate. The posterior density of U was then evaluated at a fine set of locations by integrating out V with the *integrate* function of base R. The highest posterior density credible interval for U (95%) was then evaluated by applying the *hdi* function from the *HDInterval* library for R (Meredith and Kruschke 2018) to the set of density values. Finally the mean of U was obtained by integration over its posterior density.

3. Results

We had similar numbers of attendees whose professional background was agronomy and soil science in both workshops (see Table 1). However, we had more professionals who where public health and nutrition specialists in the Ethiopian experiment.

Table 2. Fitted models for respondent data and maximized log-likelihood.

	Model	Number of parameters	ℓ
M_0	All respondents pooled	2	−81.58
M_1	Respondents separated by framing	4	−73.22
M_2	Respondents separated by framing within location	8	−69.15
M_3	Respondents separated by framing within professional group	8	−67.35

Table 3. Log-likelihood ratio tests to compare models.

Null model	Model	L	Degrees of freedom	p
M_0	M_1	16.71	2	0.0002
M_1	M_2	8.13	4	0.087
M_1	M_3	11.74	4	0.019

3.1. Nested model analysis

Table 2 shows the fitted models for the combined respondent data and their maximised log-likelihood.

Table 3 shows log-likelihood tests to compare the models. There is strong evidence to reject the model with all respondents pooled (M_0) and to accept an overall difference between the groups with different framing (M_1) ($p = 0.0002$). However, there is no strong evidence to reject M_1 by comparison to the more complex model with locations (M_2) ($p = 0.087$).

When comparing a more complex model with professional group (M_3) with model with respondents separated only by framing, there is some evidence ($p = 0.019$) to reject M_1. Therefore, further analysis of the respondent data was based on M_1 and M_3.

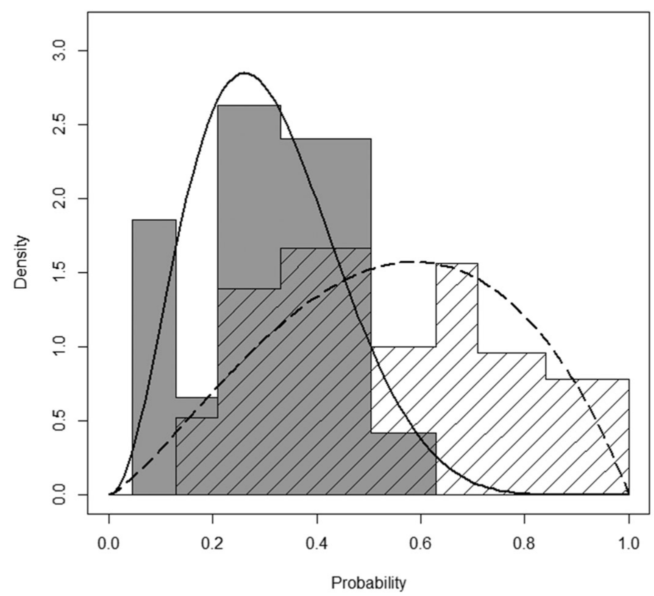


Figure 2. Fitted beta distributions for model M_1 (negative or positive framing, professional groups and locations pooled) superimposed on histograms for the results. The solid line and dark grey histogram corresponds to the respondents with negative framing. The broken line and hachured histogram are for respondents with positive framing.

Table 4. Maximum likelihood estimates of parameters U and V for models M_1 and M_3.

		Parameters	
Model	Sub-group	U	V
M_1	All respondents with negative framing	0.307	10.55
M_1	All respondents with positive framing	0.547	4.26
M_3	All public health and nutrition specialists with negative framing	0.310	30.19
M_3	All public health and nutrition specialists with positive framing	0.712	3.80
M_3	All agronomists and soil scientists with negative framing	0.303	7.69
M_3	All agronomists and soil scientists with positive framing	0.462	6.81

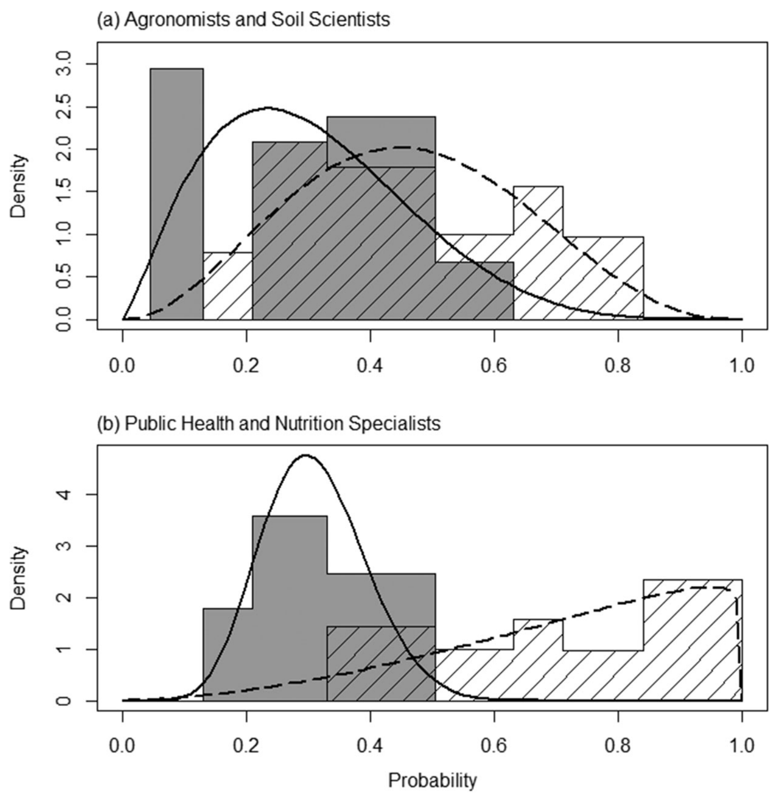


Figure 3. Fitted beta distributions for model M_3 (negative or positive framing, locations pooled) superimposed on histograms for the results for professionals from (a) agronomists and soil scientists and (b) public health and nutrition specialists. The solid line and dark grey histogram corresponds to the respondents with negative framing. The broken line and hachured histogram are for respondents with positive framing.

3.2. Model fitting

Figure 2, shows the beta probability densities for positive and negative framing under model M_1. The histograms show empirical densities from the responses over the probability ranges in each group. The solid line and dark grey histogram corresponds to the respondents with negative framing. The broken line and hachured histogram are respondents with positive framing. The figure also shows that negative framing results in a decision to intervene at a smaller probability that the threshold is not exceeded than does the positive framing.

Table 4, shows the estimated parameters for M_1. Figure 3 shows fitted beta distributions for model M_3. Here again, decisions to intervene are at a smaller probability for the respondents with negative framing in both professional groups, although the difference is most marked for the public health and nutrition specialists.

Table 4, shows the estimated parameters for M_3. The mean values for U are very similar in both professional groups with negative framing. The estimates of U under positive framing in the public health and nutrition specialists group is close to the

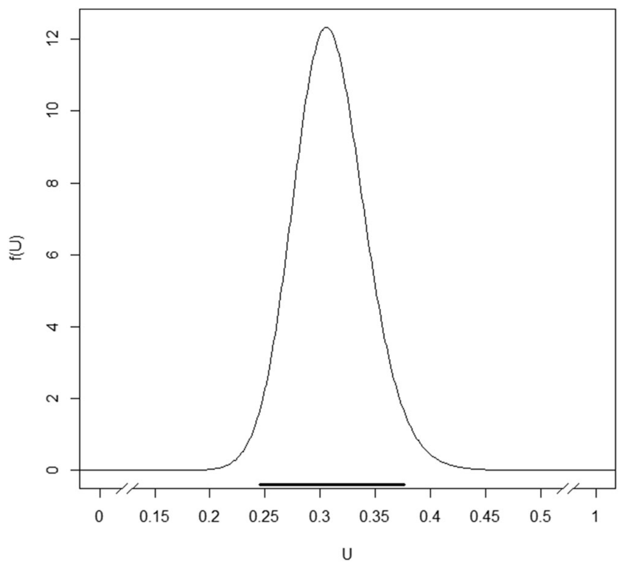


Figure 4. Posterior density for U and (solid bar) the highest posterior density credible interval (95%) estimated from pooled data for all respondents with negative framing.

complement of this value under negative framing, and the dispersion is large. It is possible that this reflects some misunderstandings of the probabilities with this group. On this basis we pooled the negatively framed responses for further analysis.

The mean of U from the posterior distribution for the pooled (over professional group) responses to the negatively-framed question was 0.31 (similar to the ML estimate). The posterior density is shown in Figure 4.

Close to symmetrical, the highest-posterior density credible interval for U, is $[0.25-0.38]$, so comfortably below 0.5. For positive framing, further analysis was based on the separate professional groups. The mean of U from the posterior distribution for the public health and nutrition specialists group to the positively framed question was 0.70 (very close to the ML estimate 0.71) with a highest-posterior density credible interval for U, is $[0.55-0.85]$. Whilst for the agronomists and soil scientists group it was 0.46 (similar to the ML estimate) with a highest-posterior density credible interval for U, is $[0.37-0.55]$.

Figure 5(a) shows a map of the probability that the concentration of Se in teff grain less than the threshold, 38 μg kg^{-1} in Amhara region, Ethiopia. The dashed line is the probability isoline or contour at which the probability is equal to the estimated mean value of P_t for the pooled (over professional group) responses to the negatively- framed question. If this value is used as a guide to decisions, then interventions would be recommended where probabilities mapped on this figure exceed the specified isoline. In these circumstances intervention would be recommended over 50% of the mapped area (34,672 km^2).

Figure 5(b) shows the same probabilities as 5a, but this time with two probability isolines, one (black) is the estimated mean value of P_t for the response of the public health and nutrition specialists group to the positively framed to the positively-framed

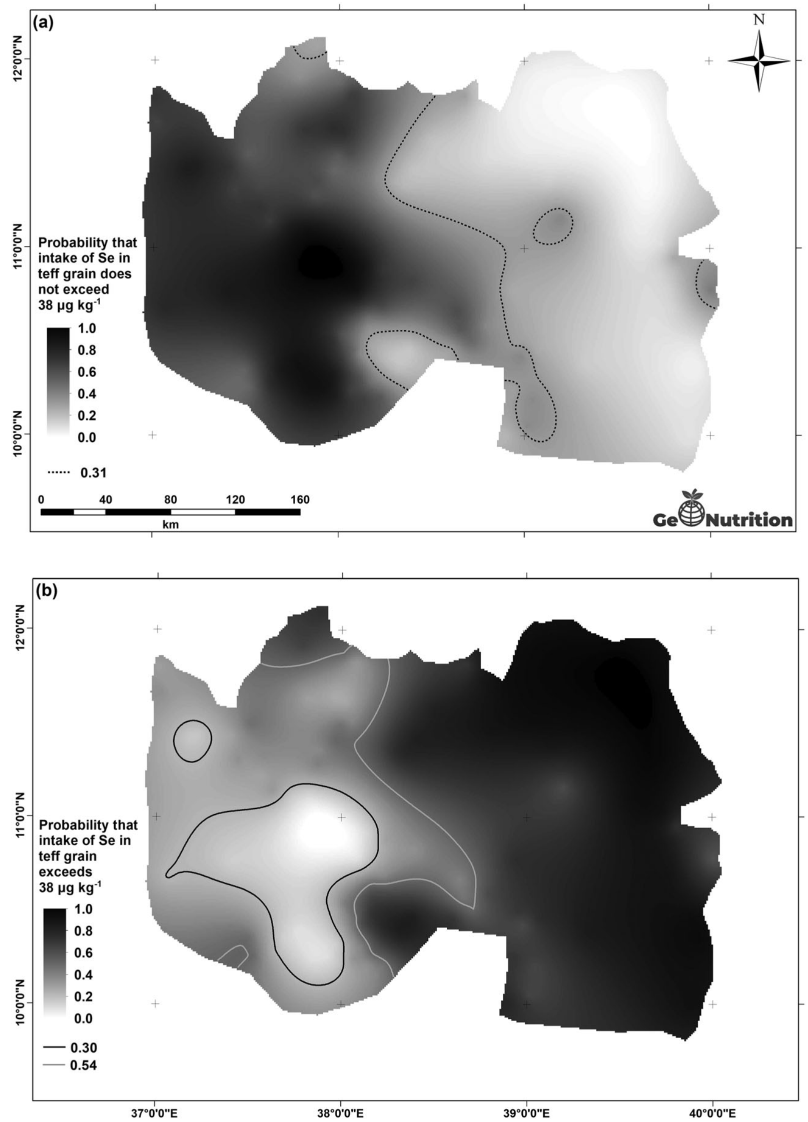


Figure 5. (a) Probability that the concentration of Se in teff grain less than 38 µg kg^{-1} in Amhara region, Ethiopia. The dashed probability isoline is the mean probability value, P_t, at which a stakeholder would judge that an intervention should be made. This is the probability at which either professional group would recommend an intervention in Amhara region, Ethiopia which the question was framed negatively. (b) Probability that the concentration of Se in teff grain exceeds 38 µg kg^{-1} in Amhara region, Ethiopia. The grey probability isoline is the mean probability value, P_t, at which agronomists and soil scientists would judge that an intervention should be made which the question was framed positively. The black probability isoline is the mean probability value, P_t, at which public health and nutrition specialists would judge that an intervention should be made which the question was framed positively.

question, this encloses an area where an intervention would be recommended corresponding to proportion, 12% of the mapped area (7,792 km^2). The second isoline (grey) is the estimated mean value of P_t for the response of the agronomist and soil scientist group to the same question. Decisions based on this value of P_t would see interventions over proportion, 40% of the mapped area (26,596 km^2).

4. Discussion

4.1. Our findings

Our results have shown (Figure 4) that a reasonably precise estimate of the mean probability value, P_t, at which a stakeholder would judge that an intervention should be made, can be elicited from a stakeholder group. The estimated mean value of P_t from a group of stakeholders in Malawi and Ethiopia, 0.31, is shown visually as a contour on the map of probabilities for Amhara region in Ethiopia (Figure 5(a)). This is the estimated mean probability at which either professional group would recommend an intervention in Amhara, Ethiopia and Malawi, if the question were framed negatively (i.e. in terms of deficiency). This P_t should not be interpreted as an objective optimal threshold value for the decision. Rather, it reflects the judgement of some group of stakeholders and their tacit assessment of losses and costs associated with making a choice with uncertain information. The methodology provided here to elicit this quantity from a stakeholder group allows us to identify a threshold P_t to use so as to present uncertain information with an interpretation which reflects the assumptions and decision-making of a particular stakeholder group. The elicitation method may also help to make that tacit process of judgement more explicit.

We also examined whether the elicited P_t depended on the specific interests of the group, and whether it is prone to framing effects (i.e. how the question is posed). With or without the effects of professional group (both M_1 and M_3), our results show that the negative framing resulted in a decision to intervene at a much smaller probability than positive framing. We also observed similar estimates of U for both professional groups within the negative framing. With the public health and nutrition specialists group positive framing resulted in a much larger threshold probability of deficiency for intervention than was the case with the agronomists and soil scientists group.

Framing effects are well known in the psychology of decision-making. Decisions are influenced by irrelevant aspects of the way information is presented, even though the same information is presented with different framings (Tversky and Kahneman 1981). In this example, a negative framing of the question draws the participant's attention to deficiency, rather than to sufficiency, and hence to a more conservative decision. We see such an effect despite preparatory activities in the experiment to draw the attention of participants to the possibility, and the implications, of interpretative errors in both directions, as suggested by Almashat *et al.* (2008). The greater consistency of responses across professional groups with negative framing may indicate that stakeholders find this easier to interpret. This maybe because stakeholders are accustomed to think about the specific problem in terms of nutrient deficiency. This shows the importance of framing spatial information, and statements of its uncertainty, in terms with which the user of the information is familiar.

We noted above that our samples and predictions, with associated probabilities were on a consistent, fixed support. A change of support (e.g. to predict a mean value across a ward or other small region, or a cell in raster GIS) will reduce the local uncertainty of the prediction. It would be interesting to see whether awareness that a probability refers to a mean across a local administrative unit, rather than a small bulk sample from within a field (which is particularly relevant to the nutrient supply to subsistence farmers) changes stakeholder's interpretation, and whether any such effect interacts with framing.

4.2. Generalizability, and topics for further work

The probability threshold which we estimated here is for a very specific problem, micronutrient concentration in staple crops, and is unlikely to serve as a general one for interpretation of spatial information. We would expect the threshold probability to differ between settings depending on the particular stakeholder perspective on the costs entailed if an intervention is not recommended where it should be, or is implemented unnecessarily. The approach which we have used could be applied to different groups and different problems and settings where decisions are based on uncertain information.

The framing effect which we have seen has been identified in other studies on decision-making under uncertainty (e.g. Chen *et al.* 2014), and so is likely to apply in other cases where probabilities are used to indicate whether the state of affairs at a location requires an intervention. In our case negative framing led to a more conservative outcome because the stakeholders are directed to think in terms of nutrient deficiency. This cannot be generalised for different problems and settings. For example, in the case of assessing concentrations of a potentially harmful element in soil against soil guideline values, a positive framing (probability that the threshold is exceeded) might be expected to result in more conservative decisions.

It would be interesting to see whether the interaction of professional group and framing holds more generally for other problems (e.g. the interpretation of information on environmental contaminants). In particular our finding in this instance, that the interpretation of probabilities was more consistent between professional groups under the framing which led to more conservative decisions, would be of practical significance if it is found to hold consistently.

Probabilities are not straightforward to interpret. As noted above, our experimental procedure included presentations to participants about uncertainty and its implications for decision making prior to their completing the exercise. However, it would have been possible to spend more time in 'priming' participants before the exercise. This could be achieved by discussion of probability problems from everyday life, like weather forecasts, when decisions are made. This might reduce the framing effect, as well as the rate of rejection due to inconsistent or anomalous interpretations. However, the responses based on minimal priming are perhaps of more practical interest, because they may better represent how a stakeholder approaches probabilistic information in the course of their ordinary working life. The fact that eight returns received from our experiment had to be discarded because they were inconsistent or anomalous underlines the difficulties that stakeholders with professional expertise in their own fields may have with the interpretation of probability. This has already been recognized (e.g. Spiegelhalter *et al*. 2011), although paradoxically, Jenkins *et al*. (2019) found that stakeholders seem to attach greater authority to numerical statements of probability than to calibrated phrases.

Some professional groups may have been able to handle and interpret probabilities better than others because of the content of education and training programmes which they typically complete. Further work to assess this, with a more varied range of professional groups, would be interesting, and might help to show how professional skills in the interpretation of uncertain spatial information could be best be developed, either in higher education curricula or in particular professional training.

When decisions are made, stakeholders weigh up the pros and cons for the decision they make. We suggest that this process might be better-emulated in an experiment such as ours if more time could be spent in engagement with stakeholder groups to co-create scenarios for decision-making, and outcomes which are possible given the uncertainty in the spatial information which is used and the stakeholders' professional experience.

4.3. Implications for practice in GIScience

The mean value of P_t obtained in this experiment will be used for practical purposes to aid interpretation of maps of nutrient supply from staple crops produced in the GeoNutrition project. We shall add a contour line to probability maps (for negative framing), as in Figure 5(a), annotating the legend to indicate that the mean threshold value applied by our stakeholder group means that interventions would be recommended where the probability takes larger values. The value can also be used as a starting point for discussion with other stakeholder groups, at national and local level, about the implications of the spatial information provided by the project.

In GIScience, it is common to validate prediction distributions by assessing the coverage of prediction intervals for validation data at different probabilities. Lark *et al*. (2019) provide an example from the study of soil nutrients. The coverage of the prediction intervals may be consistent with their probability over some ranges of values but not others. One value of this study for practical purposes in the GeoNutrition project is that we shall be able to focus our assessments of methods for spatial mapping on the validity of prediction intervals for probabilities close to P_t.

If decisions are based on uncertain information, presented in terms of the probability that a variable exceeds or falls below a threshold, then, other factors being equal, the decision process is equivalent to selecting a value of P_t. We suggest that this be done through a transparent process in which the underlying questions are examined by relevant stakeholders. Our experimental procedure, supplemented by standardized processes to co-create scenarios and to set the scene on uncertainties, could provide the basis for a formal elicitation methodology to achieve this. There is increasing interest in the use of elicitation methods to formalize the decision processes and conceptual models which individuals and communities of stakeholders may hold and use, at least tacitly, when forming expert judgements. Methods for expert elicitation have been applied to problems in medical diagnosis, the interpretation of data on natural hazards and engineering design (e.g. O'Hagan *et al*. 2006).

The development of an elicitation procedure should take account of our findings with respect to framing effects, differences between professional groups and the interaction of professional group with framing. In our particular study there was greater consistency between the two professional groups with negative framing, and a more conservative outcome. These would be reasons for using negative framing when eliciting P_t for this particular problem, but as we note above further work is needed to see how far this finding can be generalized. At the very least it is important to ensure that framing is done consistently (i.e. we do not use mix positive and negative framing for the same problem) and that framing is coherent with standard terminology in the relevant stakeholder community, e.g. whether nutrient supply is generally described in terms of deficiency (deficient or not) or sufficiency (sufficient or not).

In the theoretical framework for this study we noted that a threshold probability, P_t, can be expressed in terms of the relative losses of contrasting decisions relative to those made with perfect information. We also noted that these losses, in general, are not accessible as they may be complex and have multiple components including actual costs (e.g. money required for interventions, the economic value of disability-adjusted life years saved or not saved) but also losses which are less tangible, and which may not be directly commensurable, (the value of public health, political and reputational losses). It is possible that the elicitation of a value of P_t could help to make public or community discussions of these losses more explicit. For example, if a stakeholder group decides that interventions to address micronutrient deficiency be recommended if probability of deficiency is ≥ 0.1 then it could be pointed out that this implies that the losses arising from a failure to intervene where intervention is required are nine times larger than the losses arising from an unnecessary intervention. Stakeholders might then reflect on whether this undervalues the opportunities to apply resources to other better-focussed interventions. This discussion could be built into a group elicitation process on the lines of the behavioural elicitation methods proposed by Reagan-Cirincione (1994) under which, after initial modelling of values returned by individuals, a group works together to arrive at a consensus.

We note one further development of our approach, which could be of practical relevance. In our conceptual framework we assume discrete states: an intervention happens or does not in response to whether or not a spatial variable exceeds a threshold. In practice spatial information might be used to set a continuous value at which some intervention is applied (e.g. a rate of fortification of a foodstuff, or a rate for a fertilizer or other agronomic input). In such a case, rather than discrete losses, there may be a continuous loss function of the error of the prediction, which is zero at zero error and increases with both under- and over-estimation of the target variable. If we assume that the loss function is piece-wise linear with error in the target variable, and that α_1 is the loss per unit of error of overestimation and α_2 is the loss per unit of error of underestimation, then the expected loss is minimized at a location with some particular prediction distribution for the target variable if we use as our estimate of the target variable the value $\breve{X}$

$$\breve{X} = F^{-1}(P_o), \tag{14}$$

where $F^{-1}(p)$ denotes the quantile of the prediction distribution corresponding to probability P_o and

$$P_o = \frac{\alpha_2}{\alpha_1 + \alpha_2}, \tag{15}$$

(Journel 1984). The formal similarity with our conceptual model for P_t in the case of discrete decisions (intervene or not) is apparent. Lark and Knights (2015) showed how the continuous loss-function model could be used to compute an implicit loss function, the loss function implied by a particular level of effort to obtain spatial information, and suggested that this could be used to support decision making about sampling effort. However, it requires a value for the ratio of α_1 and α_2. One approach to obtaining this would be to provide stakeholders with scenarios in which the predicted value of the target variable is at the threshold for intervention, and to elicit a value of P_t which, under negative framing, could be regarded as an approximation to P_o in Equation (15) above.

Visualization of spatial uncertainty is important in GIScience. It is important to use appropriate colour scales to visualize spatial information, including uncertainty (Kunz *et al*. 2011, Kinkeldey *et al*. 2014). Uneven colour scales, such as rainbows, can distract from the information content of the image, and even generate artefacts (Crameri *et al*. 2020). Probabilities are ordered, continuous quantities, and we have no particular interest in values relative to a centric value (as we might for a variable on a scale from −1 to +1). For this reason, following Crameri *et al*. (2020), we decided that a sequential colour scale was appropriate. Because we wish to have good discrimination across the range of probabilities, a two-hue sequential scale is preferred. We therefore selected the 'terrain' HCL (hue-chroma-luminance) colour scale (Zeileis *et al*. 2020) to present probabilities to participants.

5. Conclusions

Much effort in GIScience and spatial statistics has focused on how to obtain prediction distributions, and probabilities from these (disjunctive kriging, indicator kriging, Bayesian methods), but it is clear (e.g. Chagumaira *et al*. 2021) that the task of communicating the uncertainty in spatial information is not complete when that is achieved, at least if the objective is that a general range of stakeholders should be able to use the information. This paper is a step towards that development. In our study we have shown we can go beyond just computing probabilities, and consider how uncertainty can be communicated to a diverse group of end-users for decision making for interventions. We also have shown that a reasonably precise estimate of the mean probability value at which a stakeholder would judge that an intervention should be made, can be elicited from a stakeholder group with particular expertise and interests.

There were more consistent estimates of the mean probability value under negative framing. This might not apply generally, whether it is should be a matter for further research. Note that 'negative' framing relative to a threshold in this setting gives rise to a conservative response, but that in other contexts (e.g. if the threshold is a pollutant), the positive framing might be expected to do so. Hence the framing effect can be pronounced in the interpretation of probabilistic representation of uncertainty presented as maps, and that this effect interacts with professional group.

Acknowledgements

The authors gratefully acknowledge the contributions made to this research by the participating farmers and field sampling teams. In Ethiopia, field sampling teams were from the Amhara National Regional Bureau of Agriculture. In Malawi, field sampling teams were from the Department of Agricultural Research Services, and Lilongwe University of Agriculture and Natural Resources.

Data and codes availability statement

The data and code that support the research are available at https://doi.org/10.6084/m9.figshare.14339987.v4

Disclosure statement

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.

Funding

This work was supported by GeoNutrition projects funded by the Bill & Melinda Gates Foundation (BMGF) [INV-009129], and the Nottingham-Rothamsted Future Food Beacon Studentships in International Agricultural Development. The funders were not involved in the study design, the collection, management, analysis, and interpretation of data, the writing of the report or the decision to submit the report for publication.

Notes on contributors

Christopher Chagumaira is a PhD candidate at the Future Food Beacon of Excellence International Agriculture Doctoral Training Programme at the University of Nottingham, Rothamsted Research and Lilongwe University of Agriculture and Natural Resources. His research focuses on spatial statistics, understanding and communicating uncertainty in spatial information of micronutrients to design site-specific interventions to address 'hidden hunger' in sub-Saharan Africa regions.

Patson Nalivata is an Associate Professor of Soil Science and head of Crop and Soil Science Department at Lilongwe University of Agriculture and Natural Resources, Malawi. His research focuses on understanding biogeochemical dynamics of selenium and iodine in tropical soils; doubled-up legume technologies; optimization of fertiliser recommendations; biochar influences on soil fertility; effects of water stress on root architecture among Malawi maize germplasm; composting systems and processes in Malawi and the quality of human compost (ecosan toilets).

Joseph Chimungu is a Senior Lecturer in Soil Science at Lilongwe University of Agriculture and Natural Resources, Malawi. His research interest is on understanding the genetic, physiological, and ecological basis of plant adaptation to drought and low soil fertility.

Dawd Gashu is an Associate Professor of Food Science and Nutrition at Addis Ababa University. His research focuses on micronutrient nutrition and its link to human health and finding feasible ways of addressing micronutrient deficiency.

Martin Broadley is Professor of Plant Nutrition at the University of Nottingham. His research seeks to increase our understanding of the movement of micronutrients and trace elements in food systems. Potential outcomes of the research include improving the nutritional quality of soils and crops for human and livestock diets.

Alice Milne is a mathematical modeller in the Sustainable Agricultural Sciences department at Rothamsted Research. Her research focuses on the mathematical analysis and modelling of agro-ecological systems using various statistical and geostatistical techniques. She also has keen interest in quantifying and communicating uncertainty in model-based predictions.

Murray Lark is Professor of Environmetrics at the University of Nottingham. His research is concerned with spatial statistics, sampling and experimental design, and capacity strengthening in agricultural and environmental research. Most of his current work addresses issues of sustainability and human health in relation to food and agricultural systems in Africa.

ORCID

Christopher Chagumaira http://orcid.org/0000-0003-0590-0350
Patson C. Nalivata http://orcid.org/0000-0001-8429-2829
Joseph G. Chimungu http://orcid.org/0000-0002-0330-0872
Dawd Gashu http://orcid.org/0000-0002-6067-4432
Martin R. Broadley http://orcid.org/0000-0003-3964-7226
Alice E. Milne http://orcid.org/0000-0002-4509-0578
R. Murray Lark http://orcid.org/0000-0003-2571-8521

References

Almashat, S., *et al.*, 2008. Framing effect debiasing in medical decision making. *Patient Education and Counseling*, 71 (1), 102–107.

Budescu, D.V., Broomell, S., and Por, H.H., 2009. Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, 20 (3), 299–308.

Chagumaira, C., *et al.*, 2021. Communicating uncertainties in spatial predictions of grain micronutrient concentration. *Geoscience Communication*, 4 (2), 245–265.

Chen, S.Y., Ross, B.H., and Murphy, G.L., 2014. Decision making under uncertain categorization. *Frontiers in Psychology*, 5 (991), 991–913.

Chilimba, A.D.C., *et al.*, 2013. *Agricultural lime application for improved soil and crop production in Malawi*. Malawi: Soil Health Consortium of Malawi (SOHCOM).

Cole, S., and Jeffries, J., 2009. *Using soil guideline values*. Bristol, United Kingdom: Environmental Agency.

Crameri, F., Shephard, G.E., and Heron, P.J., 2020. The misuse of colour in science communication. *Nature Communications*, 11 (1), 7.

DEFRA 2010., *Fertiliser Manual (RB209)*. Norwich, United Kingdom. Available from: https://ahdb.org.uk/rb209.

Gashu, D., *et al.*, 2020. Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia. *The Science of the Total Environment*, 733, 139231.

Gashu, D., *et al.*, 2021. The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, 594 (7861), 71–76.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. New York, Oxford: Oxford University Press.

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma*, 103 (1–2), 3–26.

Goovaerts, P., Webster, R., and Dubois, J.P., 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environmental and Ecological Statistics*, 4 (1), 49–48.

Greiner, L., *et al.*, 2018. Uncertainty indication in soil function maps – transparent and easy-to-use information to support sustainable use of soil resources. *SOIL*, 4 (2), 123–139.

Hatvani, I., *et al.*, 2021. Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary. *Environment International*, 146, 106263.

Heuvelink, G.B.M., 2018. Uncertainty and uncertainty propagation in soil mapping and modelling. In: A.B. McBratney, B. Minasny and U. Stockmann, eds. *Pedometrics*. Cham: Springer International Publishing, 439–461.

Heuvelink, G.B.M., and Burrough, P.A., 2002. Developments in statistical approaches to spatial uncertainty and its propagation. *International Journal of Geographical Information Science*, 16 (2), 111–113.

Jenkins, S.C., Harris, A.J.L., and Lark, R.M., 2019. When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. *Journal of Risk Research*, 22 (5), 537–554.

Journel, A.G., 1984. mad and conditional quantile estimators. In: G. Verly, M. David, A.G. Journel and A. Marechal, eds. *Geostatistics for natural resource characterisation part 1. Dordrecht*. Netherlands: Springer Netherlands, 261–270.

Joy, E.J.M., *et al.*, 2014. Dietary mineral supplies in Africa. *Physiologia Plantarum*, 151 (3), 208–229.

Joy, E.J.M., *et al.*, 2015. Dietary mineral supplies in Malawi: spatial and socioeconomic assessment. *BMC Nutrition*, 1 (1), 25.

Karl, J.W., 2010. Spatial predictions of cover attributes of rangeland ecosystems using regression kriging and remote sensing. *Rangeland Ecology & Management*, 63 (3), 335–349.

Kinkeldey, C., MacEachren, A.M., and Schiewe, J., 2014. How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51 (4), 372–386.

Kunz, M., Grêt-Regamey, A., and Hurni, L., 2011. Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Natural Hazards*, 59 (3), 1735–1751.

Lark, R.M., *et al.*, 2014. Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. *Geoderma*, 226–227, 64–78.

Lark, R.M., Ander, E.L., and Broadley, M.R., 2019. Combining two national-scale datasets to map soil properties, the case of available magnesium in England and Wales. *European Journal of Soil Science*, 70 (2), 361–377.

Lark, R.M., Ander, L., and Knights, K., 2016. Information for agriculture from regional geochemical surveys: the example of soil pH in the Tellus and Tellus Border data. In: M. Young, ed. *Unearthed: impacts of the Tellus surveys of the North of Ireland*. Dublin, Ireland: Royal Irish Academy, 193–204.

Lark, R.M., and Knights, K.V., 2015. The implicit loss function for errors in soil information. *Geoderma*, 251–252, 24–32.

Li, D.R., Zhang, J.X., and Wu, H.Y., 2012. Spatial data quality and beyond. *International Journal of Geographical Information Science*, 26 (12), 2277–2290.

Li, L., *et al.*, 2018., 1.22 - spatial data uncertainty. In: B. Huang, ed. *Comprehensive geographic information systems*. Amsterdam: Elsevier, 313–340.

Ligowe, I.S., *et al.*, 2020. Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, 79 (4), 457–467.

Likoswe, B.H., *et al.*, 2020. Inflammation adjustment by two methods decreases the estimated prevalence of zinc deficiency in Malawi. *Nutrients*, 12 (6), 1563.

Marchant, B.P., *et al.*, 2011. Spatial prediction of soil properties with copulas. *Geoderma*, 162 (3–4), 327–334.

Marchant, B.P., Saby, N.P.A., and Arrouays, D., 2017. A survey of topsoil arsenic and mercury concentrations across France. *Chemosphere*, 181, 635–644.

Mastrandrea, M.D., *et al.*, 2010. *Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties*. Intergovernmental Panel on Climate Change (IPCC), Report October. Available from: https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf.

McDonald, J.B., and Xu, Y.X., 1995. A generalization of the beta distribution with applications. *Journal of Econometrics*, 66 (1–2), 133–152.

Meredith, M., and Kruschke, D., 2018. *HDInterval: Highest (Posterior) Density Intervals. R package version 0.2.0*. Available from: https://CRAN.R-project.org/package=HDInterval.

Milne, A.E., *et al.*, 2015. Communicating the uncertainty in estimated greenhouse gas emissions from agriculture. *Journal of Environmental Management*, 160, 139–153.

Narasimhan, B., *et al.*, 2020., *cubature: Adaptive Multivariate Integration over Hypercubes. R package version 2.0.4*. Available from: https://CRAN.R-project.org/package=cubature.

Nelder, J.A., and Mead, R., 1965. A simplex method for function minimization. *The Computer Journal*, 7 (4), 308–313.

O'Hagan, A., *et al.*, 2006., *Uncertain judgements:eliciting experts' probabilities*. Chichester, UK: John Wiley & Sons.

Phiri, F.P., *et al.*, 2020. Urine selenium concentration is a useful biomarker for assessing population level selenium status. *Environment International*, 134, 105218.

R Core Team. 2020., *R: A language and environment for statistical computing.* Available from: https://www.r-project.org/.

Ramsey, M., Taylor, P., and Lee, J., 2002. Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose . *Journal of Environmental Monitoring*, 4 (5), 809–814.

Reagan-Cirincione, P., 1994. Improving the accuracy of group judgement: a process intervention combining group facilitation, social judgement analysis and information technology. *Organizational Behavior and Human Decision Processes*, 58 (2), 246–270.

Rivoirard, J., 1994. *Introduction to disjunctive kriging and non-linear geostatistics*. Oxford, UK: Oxford University Press.

Spiegelhalter, D., and Rice, K., 2009. Bayesian statistics. *Scholarpedia*, 4 (8), 5230.

Spiegelhalter, D., Pearson, M., and Short, I., 2011. Visualizing uncertainty about the future. *Science*, 333 (6048), 1393–1400.

Tjims, H., 2018. *Probability: a lively introduction*. Cambridge: Cambridge University Press.

Tversky, A., and Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science*, 211 (4481), 453–458.

Webster, R., and Oliver, M.A., 2007. *Geostatistics for natural environmental scientists*. 2nd ed. West Sussex: John Wiley & Sons Chichester.

Zeileis, A., *et al.*, 2020. Colorspace: A toolbox for manipulating and assessing colors and palettes. *Journal of Statistical Software*, 96 (1), 1–49.

Supplement for: Chagumaira et al 2022 Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops

Chagumaira, C., Nalivata, P. C., Chimungu, J. G., Gashu, D., Broadley, M. R., Milne, A. E., and Lark, R. M. (2022). Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops. *International Journal of Geographical Information Science*, 1–27, https://doi.org/10.1080/13658816.2021.2020278.

Supplementary Material

Equation (3) states that we should intervene when

$$(1 - P)L1 \leq PL2,$$

i.e., if

$$\frac{(1 - P)}{P} \leq \frac{L_2}{L_1},$$

i.e., if

$$\frac{1}{P} \leq \frac{(L_1 + L_2)}{L_1},$$

i.e., if

$$P \geq \frac{L1}{(L_1 + L_2)}.$$

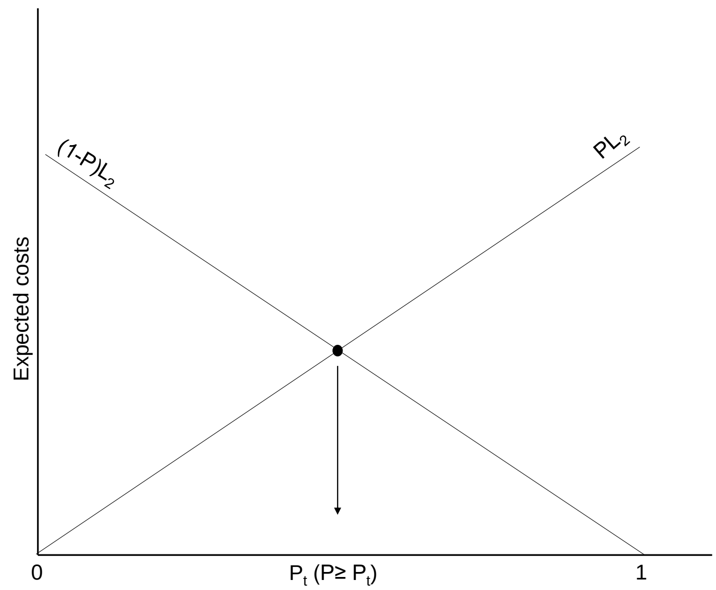


Figure S1. The expected costs as a function of P.

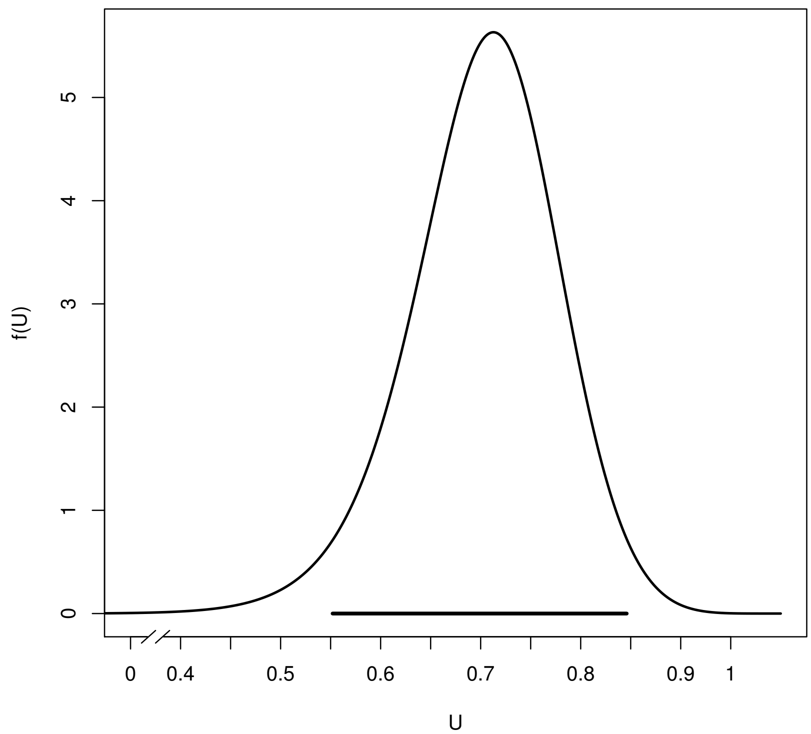


Figure S2. Posterior density for U and (solid bar) the highest posterior density credible interval (95%) estimated from the public health and nutrition group with positive framing.

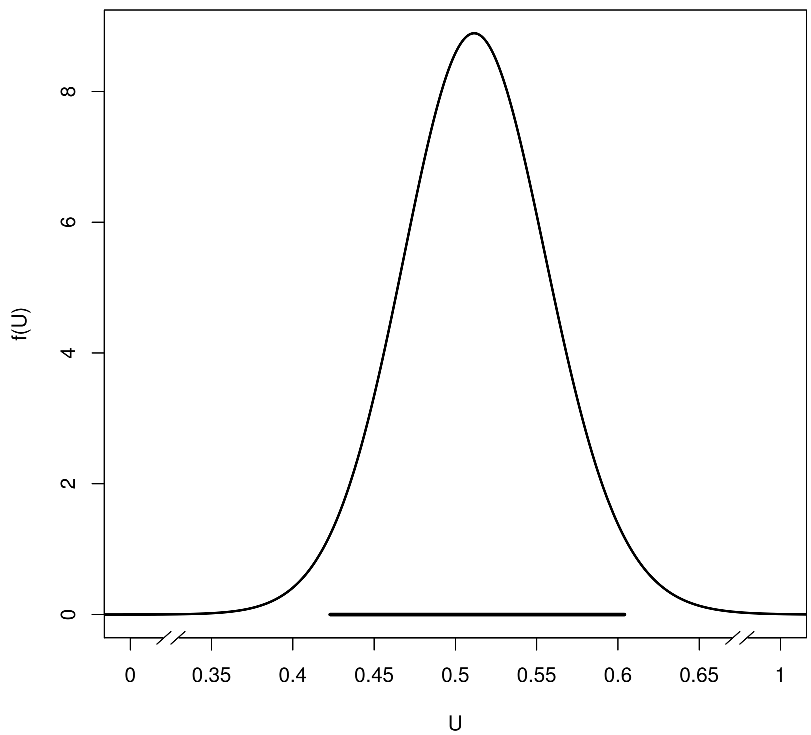


Figure S3. Posterior density for U and (solid bar) the highest posterior density credible interval (95%) estimated from the agronomy and soil science group with positive framing.

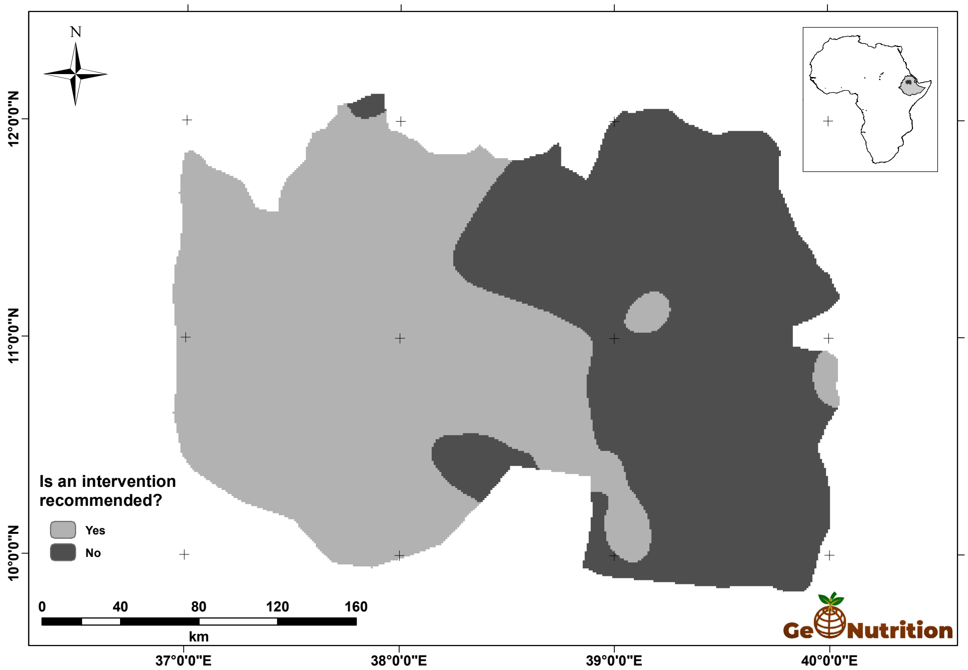


Figure S4. The area where either professional group recommended an intervention in Amhara Region, Ethiopia in the case that the question was framed negatively. The area recommended for intervention is 34,672 km^2.

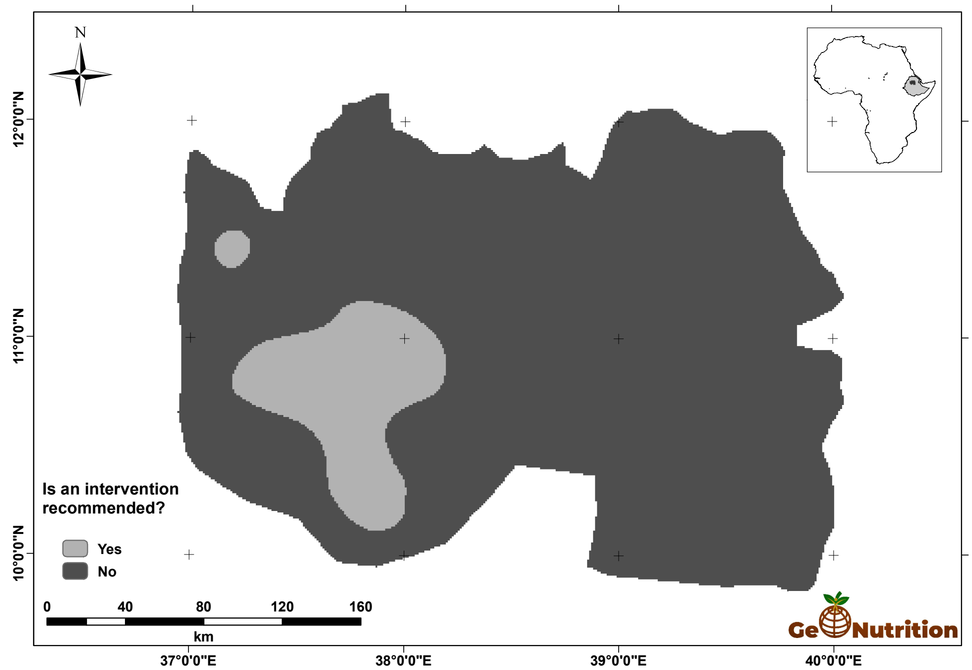


Figure S5. The area where public health and nutrition specialists recommended an intervention in Amhara Region, Ethiopia in the case that the question was framed positively. The area recommended for intervention is 7,792 km^2.

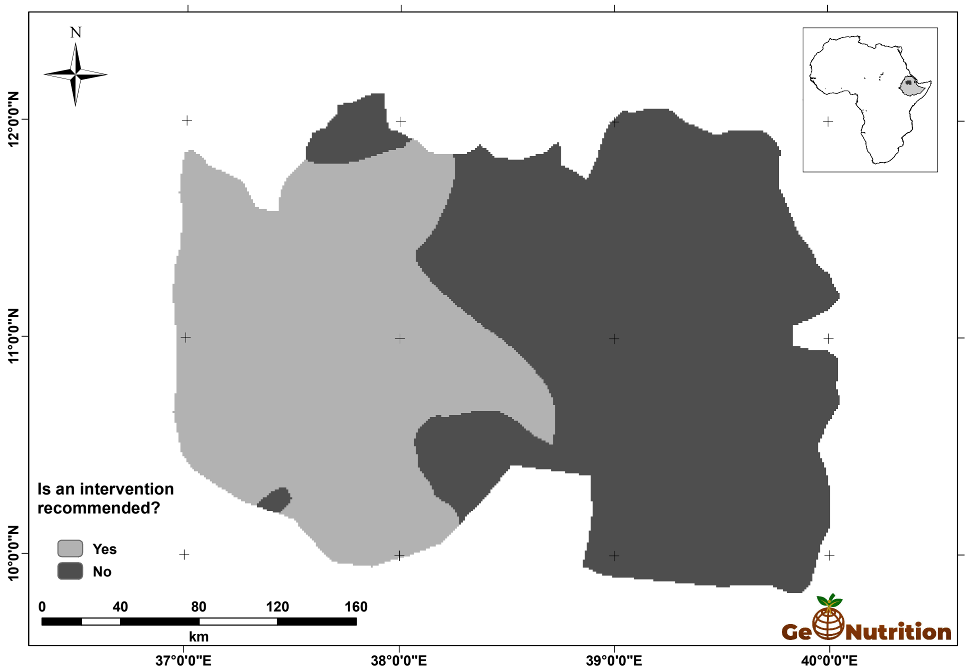


Figure S6. The area where agronomists and soil scientists recommended an intervention in Amhara Region, Ethiopia in the case that the question was framed positively. The area recommended for intervention is 26,596 km^2.

Chapter 4

Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework

This manuscript was posted as preprint on EGUsphere for public peer review for consideration for publication with the SOIL journal :

Chagumaira, C.., Chimungu, J. G., Nalivata, P. C., Broadley, M. R., Nussbaum, M., Milne, A. E., and Lark, R. M. (2022). Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework, *EGUsphere [preprint]*, https://doi.org/10.5194/egusphere-2022-583.

Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework

Christopher Chagumaira[1,2,3,4], Joseph G. Chimungu[4], Patson C. Nalivata[4], Martin R. Broadley[2,3], Madlene Nussbaum[5], Alice E. Milne[3], and R. Murray Lark[1,2]

[1]Future Food Beacon of Excellence, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom
[2]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom
[3]Rothamsted Research, Harpenden, AL5 2JQ, United Kingdom
[4]Lilongwe University of Agriculture and Natural Resources, Bunda College, P.O. Box 219, Lilongwe, Malawi
[5]Bern University of Applied Sciences (BFH), School of Agricultural, Forest and Food Sciences (HAFL), Switzerland

Correspondence: Christopher Chagumaira (christopher.chagumaira@nottingham.ac.uk; chris.chagumaira@gmail.com)

Abstract. Mineral micronutrient deficiencies (MND), prevalent in many countries, are linked to soil type. Stakeholders in Malawi, with different information needs, require spatial information about soil micronutrients in order to design efficient interventions. These stakeholders require reliable evidence for them to act, in most cases the outcome of their decisions involves financial costs and implications for farmers' livelihoods, food security and public health. They would not want to intervene where it is unnecessary to do so or not fail to intervene where it is needed. Information about the concentration of micronutrient in soil is needed by stakeholders for decision-making. In practice this information is uncertain. Geostatistical methods and those based on algorithmically driven machine learning (ML) generate predictions of soil properties with measures of uncertainty, these measures are rarely linked to the decision-making process for which spatial information is required and it may not be clear to the stakeholders how to make use of the uncertainty information in decision-making. In this study we start from an analysis of how stakeholders, in Malawi, may use uncertain spatial information to support decisions, providing the decisions about the acceptable quality of the information and how it should be collected. We then use this analysis as a framework to compare options for spatial prediction of micronutrients in soil by ML (e.g. random forest) and geostatistical methods (e.g. linear mixed models).

1 Introduction

Mineral micronutrient deficiencies (MND) prevalent in many countries, including Malawi, are linked to soil type (Hurst et al., 2013; Joy et al., 2015; Gashu et al., 2021), therefore the concentration of micronutrients in staple cereals is spatially dependent (Gashu et al., 2020). This is of particular concern in countries where diets are sourced locally as there is a greater risk of deficiency. For example, Gashu et al. (2020), showed that the concentration of Se in grain teff (*Eragrostis tef* (Zucc.) Trotter) was dependent on soil properties such as pH and organic carbon content. Joy et al. (2015) showed that there is strong evidence that cereals grown in Malawi have restricted uptake of micronutrients and dietary mineral intake is influenced by mineral content of edible portions of crops. Studies done in Ethiopia and Malawi have shown positive relationships between the concentration of Se in grain and Se-biomarker values in women of reproductive age (Phiri et al., 2019; Belay et al., 2020).

Spatial information about soil properties is needed to design site-specific interventions to address MND such as the promotion of practices like agronomic biofortification with additions of micronutrients to fertilisers (Botoman et al., 2022; Joy et al., 2022). Soil properties cannot be measured everywhere, and to map the variation in a soil property one must interpolate from measurements made on samples taken at a number of locations across the area of interest (Webster, 1977). However, the resulting predictions are uncertain due to the inherent variation of soil at multiple scales and resulting sampling error, measurement error and uncertainty arising from predictive factors in our spatial models.

Spatial information about the soil can be derived from soil survey. Conventionally, a soil survey was almost always based on classification of the soil, and the delineation of map units which are each identified with one soil class, or an association of soil classes, as explained in the map legend (Dent and Young, 1981). Information on the soil was organised with respect to the classes, e.g. estimation of class means for soil properties, or provision of a description and analyses of a 'representative profile'.

The soil surveyor would usually draw sharp boundaries between the map units they recognise (Webster, 2015). The variation of soil properties within each mapping unit is treated as an independent and identically distributed random variable for purpose of quantifying uncertainty (Webster and Beckett, 1968; Webster and Lark, 2013). However, the implicit model of spatial variation–sharp boundaries between map units– is not adequate to fully capture soil variation. One might expect more efficient predictions or similar map accuracy with fewer samples (Nussbaum et al., 2018) from a model in which soil variation occurs continuously and at multiple spatial scale in space. Such a model is provided by the regionalised variable theory of Matheron (1965) which underpins geostatistical methods. The pioneering work of Burgess and Webster (1980) introduced this methodology to soil science. The approach has been taken up and developed substantially over the intervening period (Malone et al., 2018) leading to the development of digital soil mapping (DSM, McBratney et al., 2003).

Geostatistical methods aim to capture the spatial dependence, by treating soil variation as an outcome of a random process (Webster, 2000), through predicting soil classes or properties onto grid points or cells (raster in Geographic Information System terms) from a set of point observations which might be on a systematic grid, or assembled from past surveys with different designs. Additional points near the grid nodes (close pairs) are needed to give reliable estimates of the fine-scale covariance in the soil property (captured formally in the variogram model) for geostatistical methods (Webster and Lark, 2013; Lark and Marchant, 2018). The simplest geostatistical method, ordinary kriging (OK), uses only data on the target soil property and entails the assumption that its unknown mean value is locally constant. However, the assumption can be relaxed by modelling the mean as a function of covariates. These might just be coordinates, to capture a simple trend (universal kriging) or could include other variables such as remote sensor data (kriging with an external drift). All the kriging approaches can be regarded as forms of the Empirical Best Linear Unbiased Predictor (E-BLUP, Webster and Oliver, 2007). The E-BLUP is based on the linear mixed model (LMM) with covariates as the fixed effects, spatially correlated random effects and uncorrelated residuals.

More recently, DSM practitioners have turned their attention to machine learning (ML) methods for spatial prediction of soil properties. Machine learning algorithms refers to a large class of data-driven algorithms originally developed for data mining and pattern recognition. Most ML methods do not assume a pre-defined functional form of the response-covariate relationship and therefore do not require or have relaxed assumptions of the model errors to follow a pre-specified distribution (Wadoux, 2019; Arrouays et al., 2020). DSM by ML methods can use the same covariates as LMM for spatial prediction with the E-BLUP. For spatial prediction with OK, sample points are best distributed on a grid or some other design which achieves spatial coverage (de Gruijter et al., 2006). If covariates are incorporated through an E-BLUP then the estimation of fixed effects coefficients must also be considered in the design of sampling. Brus et al. (2006) showed how requirements for estimating both the fixed and random effects components of the E-BLUP influence an optimal sample design. If spatial prediction does not directly exploit spatial dependence, the sample selection may be based on the variation of the covariates. For example, conditioned Latin hypercube sampling (Minasny and McBratney, 2006) aims to spread the sample points over the covariate space. Ma et al. (2020) suggested that the feature space coverage sampling design is optimal for ML because it covers the multivariate covariate space equally. Once soil data are collected the prediction model must be built. The incorporation of covariates which are poor predictors may inflate the uncertainty of the final prediction. In geostatistics, a variable selection procedure may be used while ML methods aim to weight all covariates appropriately.

Variability, sampling effort and modelling of the relationship between soil and covariate all contribute to the uncertainty in spatial predictions. Most studies in pedometrics provide some measure of uncertainty alongside spatial predictions, but all too often these are done in a “vacuum” without considering the particular requirements of a specific end-user in mind. The question therefore usually remains open whether an advance has been achieved from the perspective of the user of information (Lark et al., 2022). Although attempts have been made to quantify and communicate uncertainty (Chagumaira et al., 2021), pedometricians have realised that these measures are rarely linked explicitly to the decision-making process for which spatial information is required (Wadoux et al., 2021) and it may not be clear to the stakeholders how to make use of the uncertainty information in decision-making. With increasing technological advances, there is a demand to provide soil information from new soil surveys and using legacy data. We aim to analyse how a ‘decision process’ can be used to address various information needs by different stakeholders. This has implications for how good predictions have to be, i.e. the trade-off between sampling effort and degree of precision. This study fits well in the research agenda focusing on the decision process of key users of soil information to quantify impacts of uncertainty in spatial information raised by Wadoux et al. (2021) among the ten challenges for future of pedometrics.

2 Decision process

Spatial information about environmental variables is required to serve stakeholders with different needs. These stakeholders require reliable evidence for them to act, in most cases the outcome of their decisions may involve financial costs and implications for farmers’ livelihoods, food security and public health. They would not want to intervene where it is unnecessary to do so or not to intervene where it is needed. The needs of stakeholder are unsystematically reported in most pedometrics research, perhaps a lot of consideration of their needs go unaddressed. Without a full analysis of the implications of decisions end-users make, we cannot fully address their requirements. In this section we will discuss different clusters of decisions and concepts, the ‘decision process’, that should be considered when planning for sampling and spatial prediction. The general decision theory concepts which we use can be found in standard texts such as Peterson (2017).

2.1 Decisions and losses

As stakeholders have different information needs, it is important to consider the use of soil information. This can be characterised in terms of a set, I, of five questions about the information.

$$\mathrm{I}\begin{cases}\text{(I1) What decision is to be made with the information?}\\ \text{(I2) At what management unit is the decision made (e.g. field, farm, district)?}\\ \text{(I3) How is soil information used in the decision?}\\ \text{(I4) What are the possible outcomes from the decisions given uncertainties?}\\ \text{(I5) What is the potential legacy value of the survey?}\end{cases}$$

There is a need to consider how stakeholders use the information to make decisions. Figure 1 is an illustration of how a stakeholder may use spatial information based on soil pH and texture to decide on liming rate. For each *decision* made, the *outcome* depends on the possible *states*. The *state* is the state of affairs which our soil information predicts. In this example the stakeholder has a choice of three actions to intervene (apply lime at one of two rates) or not (no lime application). In this example, the stakeholder needs to make a decision about liming at a specific management unit (e.g., farmer’s field) and the decision will be made using data on soil pH and texture. If a stakeholder decides to apply a greater amount of lime when there was no need, the losses attached would be the unnecessary costs (e.g. purchasing of lime and labour) and increased risk of immobilising micronutrients in the soil. When the stakeholder decides not to apply lime when there is a need for it, the loss attached to the decision would be yield loss. The stakeholder may try to be cautious due to uncertainties and decide to apply a moderate amount of lime when there is need for a greater application, the loss attached to this decision would be some yield loss.

In our example of soil pH, the decision whether and how to intervene will reflect the stakeholder’s assessment of the loss attached to each possible *outcome*. The loss is relative to an *outcome*. For example, one might attach a cost in monetary terms to yield loss from failing to lime when necessary, and opportunity cost of unnecessary liming. One may represent the costs as a continuous loss function (Ramsey et al., 2002; Lark and Knights, 2015). For a given state (e.g. optimal liming rate) the loss is the cost of overestimation or underestimation of the soil variable(s) which determine(s) the state. If the loss from overestimation and underestimation by the same amount are equal then the loss function is symmetrical. This is unlikely in general. In the example here, if we underestimate fertiliser required by a certain quantity, we may expect a larger loss, due to yield reduction, than if we overestimate by the same amount. As a result the loss function is asymmetrical. From the perspective of environmental management the asymmetry may go either way– with more severe losses associated with over-application that results in emissions from the field.

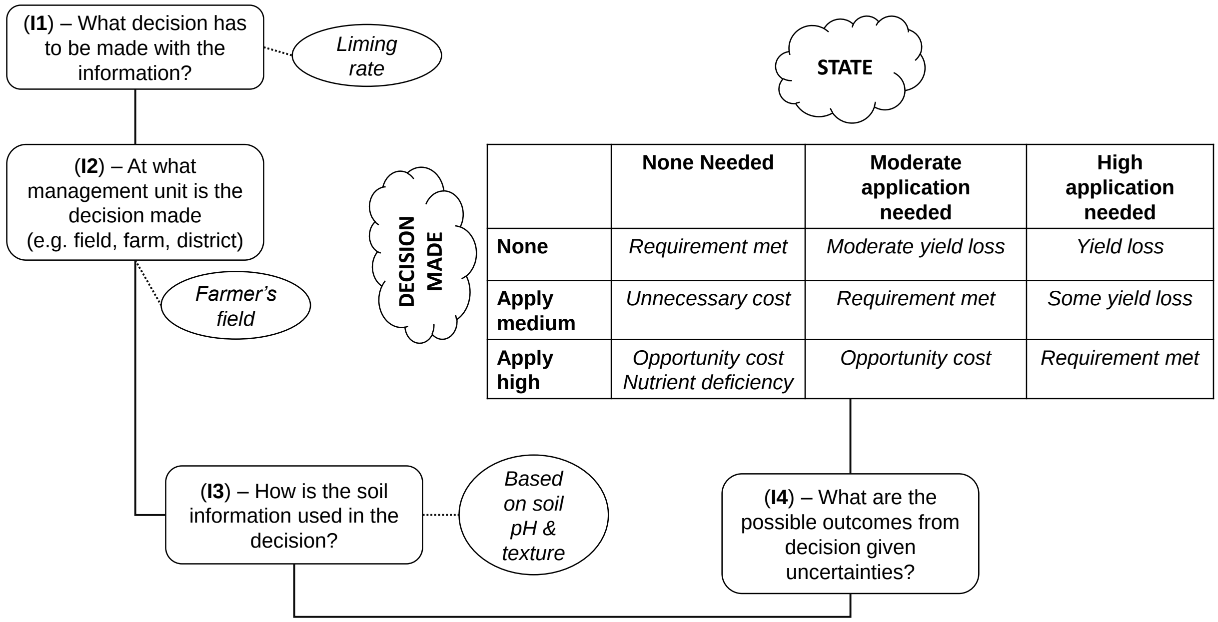


Figure 1. Illustration of a decision process for deciding on liming rate.

2.2 Stakeholders

The considerations described above highlight the importance of considering who the stakeholders are: who carries the risks from the outcomes, and who bears the costs of collecting the information. Three types of stakeholders can be defined: the information user (S1), the sponsor (S2) and the indirect or social client (S3). These three may have different soil information needs. The information user (S1) may include land managers, nutritionists, agronomists, soil scientists, policy-makers, environmental managers, governments and donors e.g. non-governmental organisations. Examples of sponsors (S2) are governments, research organisations or consortia e.g.G-BASE (Johnson and Breward, 2004), GlobalSoilMap.net (ISRIC - World Soil, 2009), GeoNutrition projects (GeoNutrution, 2017) and donors e.g. non-governmental organisations. The third group include farmers and the general public who suffer from the consequences of errors made by the group S1. At times, groups S1, S2 and S3 may have a common interest to a problem and in such a case we propose a "composite stakeholder", a user of information informed about the needs of the social client, and the constraints on the sponsor, and balancing them in a way that is socially and politically acceptable.

$$\text{“Composite” Stakeholders (CS)} \begin{cases} \text{(S1) Who makes the decision? (user)} \\ \text{(S2) Who pays for the survey? (sponsor)} \\ \text{(S3) Who is affected by the outcome? (social client)} \end{cases}$$

The uncertainty in the predictions of the soil variable of interest can usually be reduced by increasing the number of samples taken. Many survey sponsors and users, who make decisions about information, (I1 to I5), which have an impact on sample size, have little knowledge of statistics and so might not be equipped to relate sample size to measures of uncertainty and, in turn, to implications for decision. In most cases their decision are based on financial costs of the surveys. Therefore, the surveyor needs to engage with stakeholders to identify sampling designs that maximise benefits of sampling over the cost by considering the questions (O1 to O6).

$$\text{O} \begin{cases} \text{(O1) What type of survey is appropriate?} \\ \text{(O2) What measure of uncertainty is required?} \\ \text{(O3) How should survey outcomes (including uncertainty) be communicated?} \\ \text{(O4) How many samples should be taken?} \\ \text{(O5) How should the samples be distributed?} \\ \text{(O6) Which covariates should be used in modelling?} \end{cases}$$

In order to address some of the concerns raised by the questions above we need to consider the questions:

$$\text{V} \begin{cases} \text{(V1) Can we characterise the spatial variability of the soil properties of interest?} \\ \text{(V2) Can we characterise other sources of uncertainty? (e.g. analytical lab analysis)} \end{cases}$$

These questions are concerned with how we capture spatial variation (V1) and how uncertainty is quantified (V2). Data and surveys are costly, and therefore rational decisions should be made in this respect. The resources questions (R1to R4) are concerned with budgets and information. R1 is important especially for the survey sponsor (S2) because this is where the decisions on sampling are made.

$$\text{R} \begin{cases} \text{(R1) Is budget fixed or negotiable? i.e. is there an actual decision about sampling?} \\ \text{(R2) Are there legacy sources of information?} \\ \text{(R3) Are there legacy data?} \\ \text{(R4) What covariates are available?} \end{cases}$$

The key questions raised by information needs (I), stakeholder (S), spatial variation and uncertainty (V) and resources (R) are the “base level" questions about a situation where soil information is needed. The uncertainties associated with predictions from partial sampling need to be quantified and communicated effectively to the stakeholders (S1 to S3) who need to make decisions based on their information needs (I1 to I5). First, the value of uncertain information to CS must be quantified as function of uncertainty (U1). Secondly, can the acceptable uncertainty be quantified (U2)? Finally, can the survey effort be linked to tolerable uncertainty measure? (U3).

$$\mathrm{U}\begin{cases}\text{(U1) Can the value of uncertain information to CS be quantified?}\\ \text{(U2) Can acceptable uncertainty be specified quantitatively?}\\ \text{(U3) Can survey effort be linked to the uncertainty measure?}\end{cases}$$

U1 is a complex question which has received a good deal of attention, and this has been summarised by Lark et al. (2022). Lark et al. (2022), Lark and Knights (2015), Ramsey et al. (2002) and Giasson et al. (2000) discuss how value of information theory (Howard, 1966) can be applied to soil information. Lark and Knights (2015) give a simple example in which a loss function for a liming decision based on field-scale estimate by simple random sampling is used to calculate the expected loss of a decision based on an estimate as a function of its standard error (U1). This in turn can be expressed as a function of sample size (U3) given the variance of pH within the field (V1), and so the marginal benefit (reduced expected loss) of an additional sample point may be calculated. In this setting a rational value for tolerable uncertainty (V2) could be the standard error at which the marginal reduction of the expected loss equals the marginal cost of an additional sample (Lark and Knights, 2015; Lark et al., 2022). Criteria for specifying acceptable uncertainty measure (V2) may be conventions based on experience such as purity values for soil maps specified in survey contracts (Western, 1978) or the offset correlation (Lark and Lapworth, 2013) in which survey effort can be limited to a measure of the robustness of the final map to arbitrary variation of the origin of a survey grid (U3) on the basis of the variogram of the target variable (V1).

3 Case study

Malawi is a setting for much activity in the GeoNutrition project, which addresses the recalcitrant challenges of MND in SSA. The GeoNutrition project aims to examine whether better interventions to address MND could be based on spatial information, rather than assuming that the same intervention is required everywhere (Gashu et al., 2020).

The ‘Addressing hidden hunger trials (AHHA)’ trial was conducted in Malawi to test the efficacy, for the alleviation of Se deficiency, of consuming maize flour enriched with Se by agronomic biofortification (Joy et al., 2019). The Se-enriched maize flour, and non-fortified flour, was provided to households in a randomized double-blind design. Comparison of biomarker measurements before and after a period, of 8 weeks, in which this flour was consumed were recorded for one woman of reproductive age and one school-aged child in each household. It was found that, for households receiving the fortified flour, Se status of the individuals improved over the period, but did not change in the control group and that agronomic biofortification is a viable strategy to address Se deficiency (Joy et al., 2022).

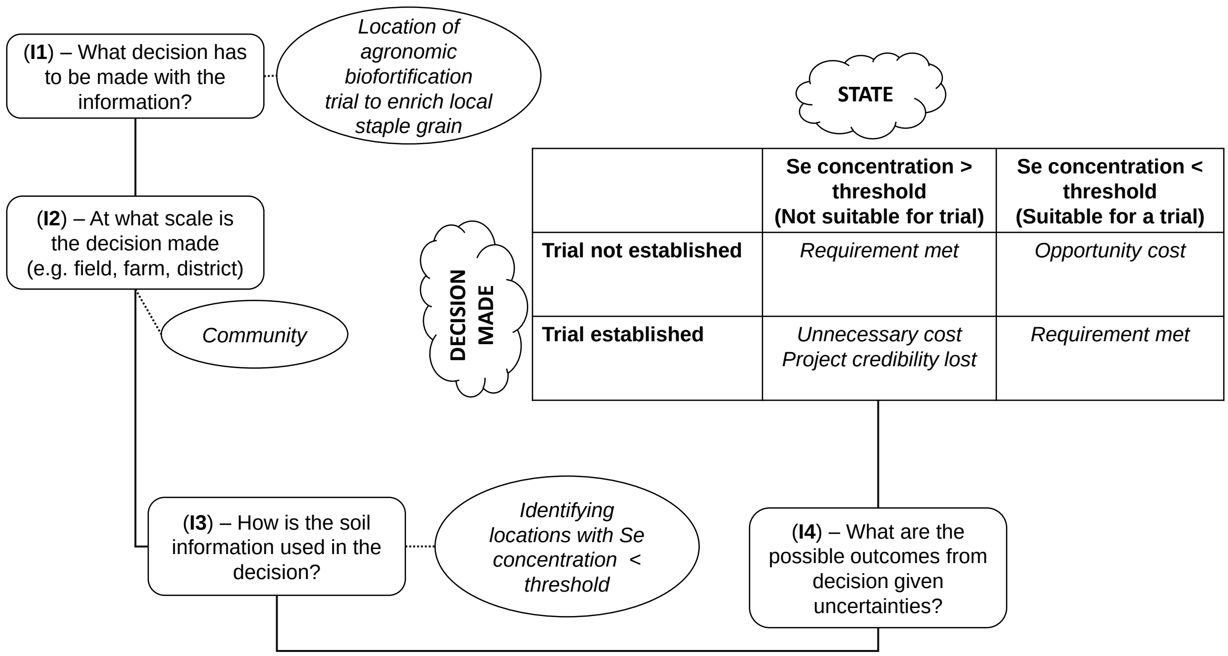


Figure 2. Summary for the decision process for deciding on location of trials, for agronomic biofortification, where concentration of $\mathrm{Se_{sol}}$ is less than a threshold.

In this case study we consider how soil information could be used for the design of a further round of experiments to evaluate the potential of agronomic biofortification as a strategy ‘at scale’ with local agronomic fortification. In the AHHA trial the fortified maize was grown at a single central location. A key question is whether agronomic biofortification can be practiced by farmers, and whether this benefits local communities who then consume the produce. The team managing the experiment, agronomists and soil scientists, public health and nutrition specialists, want to use spatial information to identify potential sites where the concentration of soil soluble Se ($\mathrm{Se_{sol}}$) is small. When a list of experimental sites has been produced, some initial engagement with each local community will be undertaken to explain the project. Informal consent will be obtained to sample soil from local fields to check that the concentrations are small. If they are, then a second more intensive phase of sensitisation will be undertaken leading up to the agronomic component of the trial and the feeding trial. If the proposed site does not have small Se concentrations in the soil then it will be abandoned and an alternative will be examined instead. This means that a ‘false positive’, a site incorrectly identified as having small Se supply from the soil, will not entail losses due to the completion of a trial where the effects of agronomic biofortification are small. However, it will entail costs due to the effort required for the initial community engagement, the completion of soil sampling, and the loss of goodwill and credibility with the community if the trial is not completed there. This might have implications for wider public attitudes to the trial, and further up-scaling of the approach in the future. A decision process for deciding on location of trials, for agronomic biofortification, where concentration of $\mathrm{Se_{sol}}$ is less than a threshold are summarised in Figure 2.

Given this background, we can now consider the question sets in Section 2.1.

I1. The decision is where to locate n community-level trials, given that the objective is to do them on sites which, among other considerations, have small $\mathrm{Se_{sol}}$ concentrations.

I2. The decision is to be made at community level. Conventionally-produced and agronomicaly biofortified crops will be grown separately by recruited farmers in each community, the soils of which have small $\mathrm{Se_{sol}}$ concentrations.

I3. Candidate communities will be those at locations where the predicted $\mathrm{Se_{sol}}$ concentration is less than a threshold. At present there is no accepted threshold to define Se-deficient soil, but agronomists and soil scientists agreed to use the 25^{th} percentile of the measured $\mathrm{Se_{sol}}$ concentration, from GeoNutrition samples (Gashu et al., 2021), as a threshold, denote by $\mathrm{Se_{threshold}}$. Other factors will influence the decision (e.g accessibility), so the project team will weigh up the risk that the site is not suitable because of soil conditions against other factors when short-listing communities for participation.

I4. A community is added to the list for the trial because the predicted $\mathrm{Se_{sol}}$ concentration is below a threshold. The soil will be sampled locally, so if the prediction is found to be correct (at least with respect to the threshold), then the site will be correctly included in the trial, and the process of community engagement discussed above will proceed smoothly. If the soil at the community is found to have Se concentration in excess of the threshold, then the site is not suitable for the trial. The effort already put into community engagement will be largely wasted, the withdrawal of the team may affect their credibility, and that of spatial soil information, in the eyes of the local community. They may be less willing to engage in similar trials in future, and resistant to future attempts to engage them in work to scale up agronomic biofortification practices (or maybe other campaigns to address MND). Uncertainties in the soil information may also mean that eligible sites might not be considered. This could result in opportunity costs for well-positioned locations, but given the size of the threshold this is unlikely to limit the completion of the trial.

I5. The proposed activity is not a survey to undertake mapping or to provide information, but is secondary data analysis. This question is therefore not relevant.

S1. The project team will make the decision, along with local officials.

S2. The sponsor is the Government of Malawi.

S3. In so far as communities are included, at least initially, in the trial but are found to be unsuitable, the project team wastes some resource. They also lose credibility, and this has some impact on the sponsor as well, who are associated with the trial. Local communities who lose faith in the promoters of the study may later choose to exclude themselves from activities, including up-scaling of MND interventions, from which they would have benefited.

O1. This is a secondary analysis, so no new survey decisions have to be made. Spatial predictions of $\mathrm{Se_{sol}}$ concentration are needed effectively at point scale since communities are small relative to the sampled domain (all the country).

O2. A quantitative measure of uncertainty is needed, so that the project team can assess the risk that the site is unsuitable when making this judgement. As there is a threshold specified, the probability that the soil Se at the site is below the threshold would be a useful measure.

O3. Chagumaira et al. (2021) found that a wide range of stakeholders found the probability that the true value of a variable is below or above a threshold is an effective way to communicate uncertainty in spatial information. They found little evidence that different methods to express this probability worked better than others, but methods such as calibrated phrases (Mastrandrea et al., 2010; Lark et al., 2014) or maps of probability with pictographs at different candidate communities might be appropriate.

O4. This is secondary data analysis and this decision is not relevant.

O5. As this is secondary data analysis this decision is not relevant.

O6. The available covariates are surface slope, and topographic index mapping derived from the MERIT Digital Elevation Model of Yamazaki et al. (2017). Downscaled climate data was obtained from CHELSA data set (Karger et al., 2017). Average and variance annual net primary productivity, enhance vegetation index, normalised difference vegetation index and soil adjusted vegetation index were obtained from the MODIS remote sensor satellite (Justice et al., 1998) (see Table S1).

Questions under V and R: probably not relevant to a secondary data analysis.

U1. The costs of a false positive are partly tangible (time and resources wasted) and intangible (loss of goodwill and credibility). Costs of false negatives are harder to evaluate. Given the size of the threshold the project is unlikely to be short of communities. However, if the criterion is too strict we might miss out on communities good for other reasons (accessibility etc). Overall, cost of false positives exceeds that of false negatives.

U2. On the above basis a critical probability might be elicited from a group, following Chagumaira et al. (2022) such that a community is considered for inclusion in the trial of $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < \mathrm{Se}_{\mathrm{threshold}}\right)$ exceeds that critical value. Considerations in U1 suggest that the loss function is asymmetrical with a larger loss from false positives than false negatives. On this basis we would expect that, if the probability is presented as $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < \mathrm{Se}_{\mathrm{threshold}}\right)$, then the critical probability will be in excess of 0.5.

U3. Not relevant for a secondary data analysis.

4 Materials and methods

4.1 Data and study area

Details of soil sampling and laboratory analysis are given by Gashu et al. (2021). Soil soluble Se was extracted in 0.01M $\mathrm{KNO_3}$. The objective of the field sampling in Malawi was to support spatial prediction of soil and crop micronutrient concentration.

The location of sample points were obtained by using the k-means methods as encoded in the spcosa package for the R platform (Walvoort et al., 2010). This method allows one to form a sample which gives good coverage while incorporating the fixed prior points in the sample. There were 820 prior points from the 2015–16 micronutrient survey of Malawi (Phiri et al., 2019), and a further 890 spatial coverage points were added by using spcosa plus a further 190 'close-pair' sample locations. The close-pair samples are required to support spatial modelling–10% of the total samples (Lark and Marchant, 2018). The additional points near the grid nodes (close pairs) are needed to give reliable estimates of variogram parameters and they increase spatial coverage to minimise kriging variances especially at the border of a study region. A total of 1,812 sites of grain and soil samples were taken. However, some sample location had positional uncertainties that were attributed to either poor satellite signal or enumerators not giving the devices enough time to establish the location and were not usable for further analysis. These six samples were removed from the spatial prediction of $\mathrm{Se}_{\mathrm{sol}}$. Of these 1806 sites, 10% of the data was used to create an independent dataset for validation. The 190 close-pair sites were included in the dataset used for prediction and training of the models. From the data with the close-pairs removed, 160 points were selected into the validation dataset using simple random sampling without replacement.

4.2 Linear mixed models for spatial prediction

The best unbiased linear predictors (BLUP) is computed from a linear mixed model (LMM) is an additive combination of one or more fixed-effects and one or more random effects. The independently and identically distributed component is in the model but does not affect the prediction. When the fixed effects are just an unknown constant mean then the BLUP is equivalent to ordinary kriging. The theory of LMM as a geostatistical model for spatial prediction is described in greater detail by Lark et al. (2006). The variance parameters are estimated by maximum likelihood or residual maximum likelihood (REML, Stein, 1999). Both maximum likelihood and REML are based on the assumption that random effects have a joint Gaussian distribution, therefore it is important to study the descriptive statistics of the dataset and do the transformations when necessary. Estimation of variance parameters by REML eliminated bias that results from estimating semivariance computed from a variogram of residuals by using method-of-moments when there are fixed effects added to a constant mean.

Exploratory data analysis, using simple summary statistics and plots (e.g. Q-Q), were done to check whether transformation was needed to make the assumption of normality reasonable. Table 1 shows the summary statistics of $\mathrm{Se}_{\mathrm{sol}}$. Exploratory analysis of the data indicated a possible spatial trend, but this was not pronounced, as indicated on the exploratory variograms. Two LMM were therefore considered as options. First, with a constant mean as the fixed effect, second with fixed effects selected from available coordinates (Table S1), including spatial coordinates.

4.2.1 Constant mean as the fixed effect

The $\mathrm{Se}_{\mathrm{sol}}$ data was strongly skewed, transformation to logarithm reduces this, but the histogram of the data stilled showed marked non-normality and bimodal (see Table 1 and Figure S3). We therefore considered transformation of the data by Gaussian anamorphosis (GA) using Hermite polynomials computed with the anam.fit function for the RGeostats package (MINES ParisTech / ARMINES, 2022) for the R platform (R Core Team, 2021). Variograms were then estimated for the transformed

Table 1. Summary statistics of for soil soluble Se concentration ($\mathrm{Se_{sol}}$), cross-validation errors for ordinary kriging and robust REML E-BLUP, and out-of-bag cross-validation errors for random forest.

	Concentration of $\mathrm{Se_{sol}}$ (µg $\mathrm{kg^{-1}}$)	Ordinary kriging cross-validation errors	robust REML E-BLUP cross-validation errors	Random forest out-of-bag errors
Mean	3.94	0.00	-0.07	-0.00
Median	3.30	0.01	-0.02	-0.05
Standard Deviation	2.99	0.55	0.76	0.50
Minimum	0.18	-2.78	-2.86	-1.90
Maximum	18.8	3.11	2.48	2.41
Skewness	1.29	-0.26	0.00	0.59
Octile skewness	0.24	-0.03	-0.16	0.14

data using the estimates due to Matheron (1962), Dowd (1984) and Cressie and Hawkins (1980). Exponential variogram model were fitted by weighted least squares and the models were validated by cross-validation. Following Lark (2000), we chose the variogram fitted by the different estimators that had a standardised squared prediction error (SSPE) falling within the 95% confidence interval around the expected value of 0.455. Predictions of the GA-transformed scale was obtained by ordinary kriging (OK) at validation sites and at points in a grid across Malawi. The value of $\mathrm{Se_{threshold}}$, 1.49 µg $\mathrm{kg^{-1}}$, was transformed to the GA scale (-0.681), and the probability that the true value is smaller than this was computed assuming a prediction distribution of the mean and variance equal to the OK estimate and kriging variance, respectively. Median-unbiased estimates of $\mathrm{Se_{sol}}$ in units of µg $\mathrm{kg^{-1}}$ were obtained by back transformation.

4.2.2 Fixed effect selected from available covariates

Exploratory analysis with spatial coordinates and environmental covariates (see Table 2) on fixed effects suggests that an assumption of normal random effects was plausible on transformed to natural logarithm, although with some outliers present. For this reason we estimated variance parameters for the LMM by using robust REML following Künsch et al. (2013).

Robust REML algorithm automatically identify outliers within a dataset and the outliers receive small weight when estimating model parameters (Nussbaum et al., 2012). The covariance matrix of the regression coefficients and the variogram parameters are estimated by georob (Papritz and Schwierz, 2021), for the R platform, either by REML or maximum likelihood from the $\mathrm{Se_{sol}}$ dataset and values of the environmental covariates. The estimating equations are robustified by replacing the standardised errors by a bounded function of them and introducing a suitable bias correction terms for Fishers consistency in the Gaussian model (Künsch et al., 2013; Papritz, 2021). The bounded function of the residuals becomes the tuning parameter, c, of robust REML and is used to control the robustness of the procedure. The lower the value of c, the more the outliers are penalised by lower weights. The resulting predictions of $\mathrm{Se_{sol}}$ were on a log-scale, and they needed to be back-transformed to aide interpretation. We used the standard unbiased back-transformations for log-normal kriging procedure (Cressie, 2006) using the lgnpp function of georob package (Papritz and Schwierz, 2021).

Not all available covariates are useful for spatial prediction when using LMM. If all covariates are used for prediction some problems can be encountered. These include risk of propagating error in the regression coefficients when weak covariates are included in a model. Other problems may include overfitting, over-adaptation to the training data hence leading to a lack of generalization and poor predictive performance in new conditions. In order to address some of these challenges Lark (2017) suggested that the problem of variable selection is considered in terms of multiple hypothesis testing. Hence in this study, covariate selection was done by using the method described in Lark (2017) in which false discovery rate is controlled with the α-investment.

In this method of variable selection, a prior ranking of potential predictors of $\mathrm{Se_{sol}}$ was required. The rankings were provided by a panel of eight plant and soil scientists from the University of Nottingham, Rothamsted Research, Lilongwe University of Agriculture and Natural Resources and Addis Ababa University. The rankings were based on a priori expectations of the order of importance based on the processes involved. The rankings from the panel was obtained through group elicitation facilitated by a statistician. The selected order for testing potential predictors for $\mathrm{Se_{sol}}$ obtained in the group are shown in Table 2.

Table 2. Sequence of predictors for $\mathrm{Se_{sol}}$ concentration (environmental covariates) for testing with the α-investment

Order	Environmental Covariate
1	Downscaled mean annual precipitation (BIO12)
2	Downscaled mean annual temperature (BIO1)
3	Slope (SLOPE)
4	Topographic index (TIM)
5	Average enhanced vegetation index (EVI)
6	MODIS band 7 (MB7)
7	MODIS band 2 (MB2)
8	MODIS band 1 (MB1)
9	MODIS band 3 (MB3)

The climate variables (mean precipitation and rainfall) were as the most likely useful predictors because it is expected that rainfall and temperature can enhance the mineralization of organic matter in the soil thereby releasing Se bound in organic compounds into soil solution. The terrain variables, slope and topographic index where considered next. The MODIS enhanced vegetation index (EVI) and bands were also considered because they measure vegetation vigour and health.

In a LMM framework, evidence that the coefficient of a covariate is significantly from zero can be tested by a Wald test

$$\mathrm{W_T} = \frac{\left(\hat{\theta}_1 - \theta_0\right)^2}{\mathrm{Var}(\hat{\theta}_1)}. \tag{1}$$

Where θ_1 and θ_0 are the vectors of all parameters estimated by REML, from fitting the model with additional fixed effect and the simple model, respectively. The statistic is asymptotically distributed χ^2 with one degrees of freedom (Diggle et al., 1994; Drapper and Smith, 1998). In each test, if the p-value did not exceed 0.05 then the predictor was provisionally retained, otherwise it was dropped, and the next predictor was considered. When all the predictors had been considered, the p-values for the Wald test on each were compared to thresholds according to the α-wealth controlling the false discovery rate. The successive hypothesis were tested in the order (1) annual precipitation, (2) annual temperature, (3) slope, (4) topographic index, (5) enhanced vegetation index, (6) MODIS band 7, (7) MODIS band 2, (8) MODIS band 1 and (9) MODIS band 3 (Table 2). The predictors whose p-values were below the thresholds would be used in the final model. The models were sequentially fitted starting with a 'null' hypothesis with the linear spatial trend identified in the exploratory analysis. The were fitted with robust REML with $c = 2$, in order to avoid problems of convergence. Convergence problems often depend on the data and occur when c is low and a numerical solution of the equations would not be found (Papritz, 2021). This occurs when there is low spatial correlation (near pure nugget variogram) and/or poor linear relationship between response and covariates. However, the final model was fitted with a much lower tuning parameter, $c = 1.75$, in order to penalise the outliers in $\mathrm{Se}_{\mathrm{sol}}$ dataset with lower weights.

4.3 Quantile random forest algorithm for spatial prediction

A random forest is an ensemble of tree-structured predictors formed by a collection of classification and regression trees (CART), that depend on the value of a random vector sampled independently, with the same distribution for all the trees in the forest (Breiman, 2001). It is a method often used for performing predictive tasks (e.g., Nussbaum et al., 2018) by combining large number of regression trees by the mean of their predictions. Decision trees are often referred to as a 'greedy algorithm', because each split reduces the residual sum of squares for that particular subgroup, rather than optimising splits to reduce overall residual sum of squares (James et al., 2013). This 'greedy' property tends to over-fit the training data and results in poor predictions (Bramer, 2020). Due to their hierarchical nature decision trees tend to be unstable and have large variance (Hastie et al., 2009), in the sense that large changes in model and prediction following only small changes in the training data. On the other hand, on average tree based predictions tend to be unbiased. Therefore, algorithms like random forest have been developed to balance for the instability of CART, but to still be able to profit from the complexity of interaction-type response-covariate relationships. Quantile random forest is an expansion of random forest that allows for uncertainty quantification for each prediction (Breiman, 2001; Meinshausen, 2006). We used the Boruta algorithm that uses a wrapper approach for variable selection using the Boruta package (Kursa and Rudnicki, 2010). The algorithm creates a shadow attribute dataset consisting of randomly shuffled predictors. A random forest model is fitted including original and shuffled predictors and variable importance is computed. Variables that have on average larger importance than the randomised variables will be used in spatial prediction. The quantile random forest is explained in greater detail by (Meinshausen, 2006). The ranger function of the ranger package (Wright and Ziegler, 2017) was used for fitting the random forest to predict $\mathrm{Se}_{\mathrm{sol}}$ in Malawi. In order to directly compare to robust REML E-BLUP, $\mathrm{Se}_{\mathrm{sol}}$ was also transformed by natural logarithm scale.

4.4 Measures of uncertainty

Prior and posterior measures of uncertainty can be used to quantify uncertainty in spatial predictions. A prior measure of uncertainty results from the prediction process. Kriging variances are example of prior measure of uncertainty for LMM. The kriging variances are the expected square difference between predicted and the observed values and they can be plotted on a map to show areas where additional sampling is required to reduce this uncertainty. The appropriateness of the kriging variances can be assessed by the standardised squared prediction error (SSPE) after internal cross-validation, with leave-one-out or K-fold cross-validation. We used cv.georob function to perform K-fold cross-validation for the robust REML E-BLUP. The model is re-fitted 10 times by robust REML but each time $1/K$th of the data is excluded (Papritz, 2021). After cross-validation, the SSPE is computed by

$$\theta(\mathbf{x}) = \frac{\{z(\mathbf{x}_i) - \tilde{Z}(\mathbf{x}_0)\}^2}{\hat{\sigma}^2_{\mathrm{K}}(\mathbf{x}_0)}, \tag{2}$$

where $\tilde{Z}(\mathbf{x}_0)$ is the kriging prediction of $z(\mathbf{x}_i)$ and $\hat{\sigma}^2_{\mathrm{K}}(\mathbf{x}_0)$ is the kriging variance. The expected value of $\theta(\mathbf{x})$ is 1, this is not a sensitive diagnostic. Assuming that the errors follow a Gaussian distribution $\theta(\mathbf{x})$ is expected to have a χ^2 distribution with one degree of freedom, so that the median value of $\theta(\mathbf{x})$ over all data can be used as a diagnostic (Lark, 2000).

The prior measures of uncertainty for random forests are the out-of-bag mean square error and the quantile regression forest that estimate conditional distribution of the predicted variable. About a third of the samples in the random forest are left out during the bootstrapping of samples. The out-of-bag serve as test sample to assess the prediction accuracy of random forests through computation of universal measures of uncertainty (e.g., mean square error) as with cross-validation. Prediction intervals can be computed from conditional quantiles by using quantile regression forest, a generalisation of random forests by Meinshausen (2006).

The prior measures of uncertainty, kriging variances and conditional quantiles, can be used to compute conditional probabilities that–given the current model– future observations of $\mathrm{Se}_{\mathrm{sol}}$ fall bellow a threshold, $\mathrm{Se}_{\mathrm{threshold}}$ (1.49 µg kg^{-1}). With robust REML E-BLUP, an assumption of normality of the prediction errors, after cross-validation, should be plausible. Exploratory analysis of the kriging errors after computing K-fold cross-validation, showed that the errors could be regarded as a normal random variable. Conditional probabilities also can be obtained from quantile predictions of quantile regression forest without assumption of normality as in the geostatistical approach. Conditional probabilities for a true value exceeding a threshold, $\mathrm{Se}_{\mathrm{threshold}}$, are taken from the full predictive distribution resulting from quantile regression forest (Meinshausen, 2006).

We also used indicator kriging to obtain the conditional probability. The kriging predictions were made by ordinary kriging of a transformed variable, the indicator variable, ω (Webster and Oliver, 2007). The transformation is made by:

$$\omega = \begin{cases} 1 & \text{If } \mathrm{z}(\mathbf{x}) \leq \mathrm{z_t}, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Posterior measures of uncertainty depend on the primary data and are obtained from validation. Validation compares the prediction at a site and the measured value at that site. Several approaches can be used to validate predictions and these include

Jack-knifing/ data splitting and collection of an independent dataset. To assess random forest, OK with GA and robust REML E-BLUP, we validated the predictions with an independent dataset. We computed the universal prediction accuracy measures (e.g., mean error, mean square error, root mean square error). We also computed the mean squared error skill score (MSESS)

$$\mathrm{MSESS} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\{\mathrm{z}(\mathbf{x}_{\mathrm{i}}) - \tilde{\mathrm{Z}}(\mathbf{x_0})\}\right)^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{z}(\mathbf{x}_{\mathrm{i}}) - \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{z}(\mathbf{x}_{\mathrm{i}})\right)^2}, \tag{4}$$

where $\tilde{Z}(\mathbf{x}_0)$ is the prediction and $z(\mathbf{x}_i)$ is the measured value of $\mathrm{Se}_{\mathrm{sol}}$. The MSESS can be interpreted in a number of ways. At the first instance a score of 1 indicates perfect predictions and the root mean square error would be 0. Second, a score of 0 shows that the predictions have the same variance as the data of the validation set and a score less than 0, suggests that the predictions have larger variance than the validation dataset (Wilks, 2011).

Coverage probabilities were estimated from 0.5 to 0.99 for the predictions from cross-validation and of the independent dataset for OK with GA, robust REML E-BLUP and random forest. We used the blakerci function of the PropCIs package for the R platform to compute the 95% confidence interval for each estimated coverage (Blaker, 2000). Chagumaira et al. (2022) showed that a critical probability, P_t, can be elicited from a diverse group of stakeholders (S1 to S3) when provided with maps conditional probabilities of not exceeding a threshold. The P_t is an indication of a stakeholder's judgement when making a decision for an intervention using uncertain spatial information. The elicited P_t can be used to assess the validity of coverage probabilities of prediction intervals for probabilities close to a threshold. In this case study we used a range of notional P_t to assess coverage probabilities of prediction intervals and thereby show proportion of mapped area under the selected critical probability value. This analysis can allow stakeholders to identify locations for setting up field trials.

5 Results

5.1 Linear mixed models for spatial prediction

After transforming $\mathrm{Se}_{\mathrm{sol}}$ by GA, there was no evidence of a spatial trend. Summary statistics for the cross-validation errors for OK with GA are presented in Table 1. The variograms for the transformed $\mathrm{Se}_{\mathrm{sol}}$ are shown in Figure A1. After cross-validation, the variogram estimated using the Matheron estimator has the largest value of median of SSPE of 0.39. The median unbiased back transformed spatial predictions of the concentration $\mathrm{Se}_{\mathrm{sol}}$ in Malawi, by ordinary kriging, are shown in Figure 3a. The map shows that there is higher concentrations of Se in central, northern and southern west parts of Malawi.

After sequential fitting of the models, with robust REML, with ordered predictors for $\mathrm{Se}_{\mathrm{sol}}$ shown in Table 2, enhanced vegetation index (EVI) was selected through the false discovery rate control procedure (Figure 4). The graph (a) shows α-wealth over the sequence of tests and the lower (b) shows the p-values and the corresponding thresholds under the false discovery rate control with the α-investment. The final model fitted with robust REML, with $c = 1.75$, was used for spatial prediction of $\mathrm{Se}_{\mathrm{sol}}$. The maximum likelihood variogram for $\mathrm{Se}_{\mathrm{sol}}$ is presented in Figure A2. The variogram shows strong spatial autocorrelation in the data for $\mathrm{Se}_{\mathrm{sol}}$.

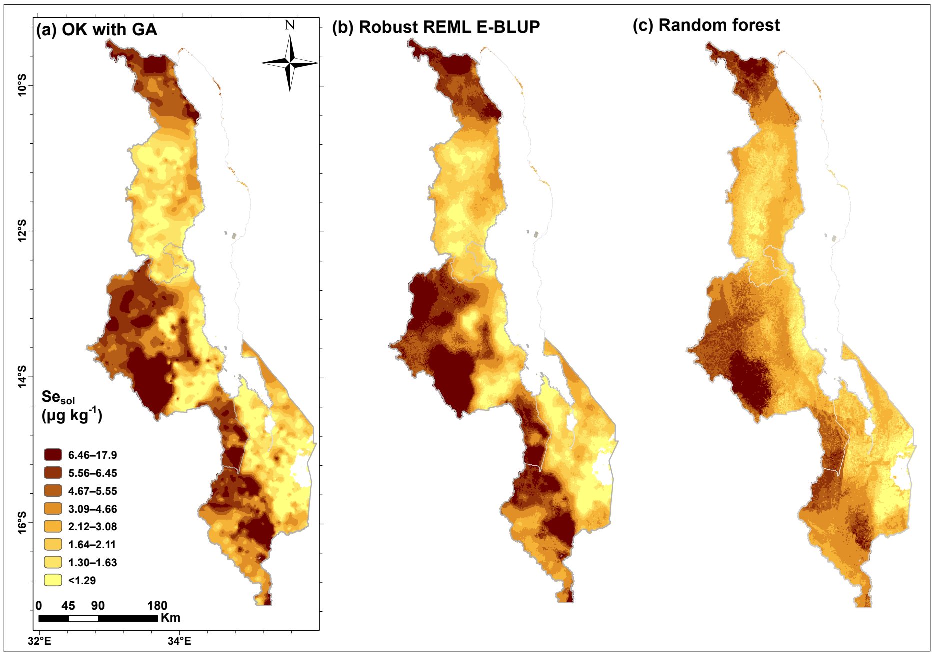


Figure 3. Spatial predictions of $\mathrm{Se_{sol}}$ concentration across study area by: (a) OK with transformed $\mathrm{Se_{sol}}$ (b) robust REML E-BLUP and (c) random forest.

Table 3. Models for $\mathrm{Se_{sol}}$ concentration in Malawi fitted using robust REML.

Predictand	Coefficient				$\breve{R}^2_{adj}$	τ^2	σ^2	ϕ
	β_0	β_1	β_2	β_3				
		Easting	Northing	EVI				
Null model	11.0377	-0.0042	-0.0009			0.1095	0.4087	28.0000
+EVI	11.7980	-0.0045	-0.0010	0.0002	0.025	0.1099	0.3985	28.0000
Final model	11.7847	-0.0045	-0.0010	0.0002		0.1065	0.4026	28.0000

† The symbols β_0 to β_2 are the fixed effects coefficients, β_0 is a constant and β_i is the coefficient for the ith random effect; κ is the smoothness parameter of the correlation function; τ^2 is the nugget variance; σ^2 is variance of the correlated random effect; and ϕ is the distance parameter. $\breve{R}^2_{adj}$ is the difference between σ^2 of the null model and proposed model expressed as a proportion of the variance for the null model.

The variance parameters estimated by robust REML E-BLUP for the null model and the model with EVI as a predictor are shown in Table 3. A small proportion of the spatially correlated variation ($\breve{R}^2_{adj} = 0.025$) is accounted for by inclusion of EVI as a predictor of $\mathrm{Se_{sol}}$.

Summary statistics for the cross-validation errors are shown in Table 1. After cross-validation we computed SSPE, and the median value of $\theta(\mathbf{x})$ was 0.427. This value lies within the 95% confidence interval for the expected value of median under a

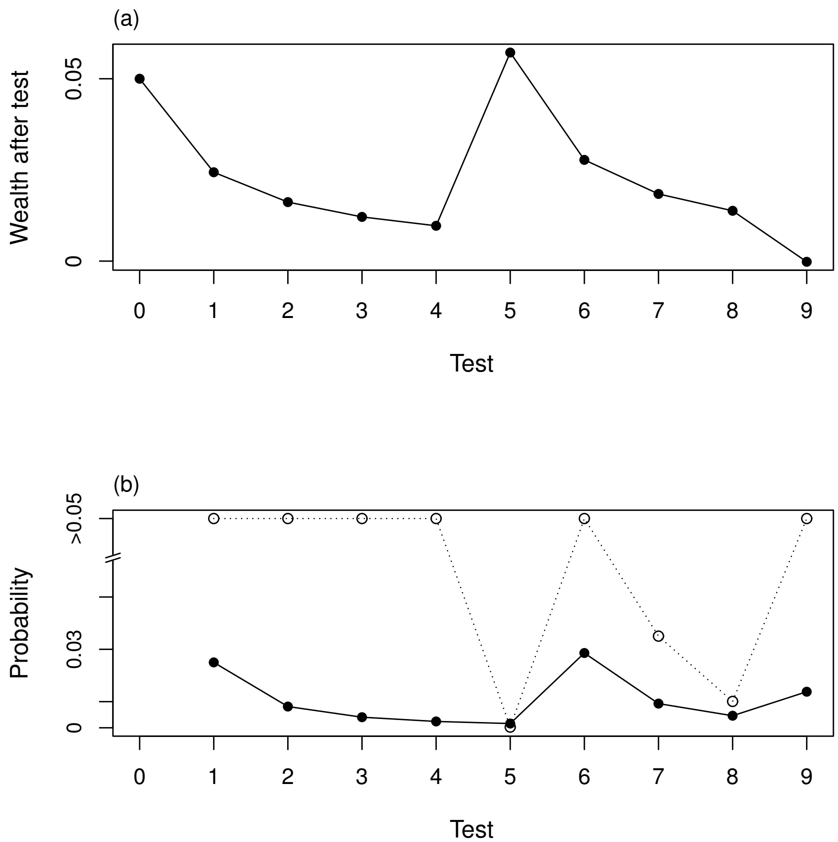


Figure 4. Ordered tests for covariate selection for models for $\mathrm{Se_{sol}}$ fitted by robust REML. The sequence of predictors is as given in Table 1b. The graph (a) shows α-wealth over the sequence of tests and the lower (b) shows the p-values for successive tests (open symbols) and the corresponding threshold values with marginal false discovery rate control.

valid model given the number of observations. The mean unbiased spatial predictions of the concentration $\mathrm{Se_{sol}}$ in Malawi, by robust REML E-BLUP, are shown in Figure 3b.

Figure 5 shows the maps of conditional probability of $\mathrm{Se_{sol}}$ falling bellow $\mathrm{Se_{threshold}}$ (1.49 µg kg^{-1}). Figures 5a to d are the probabilities expressed on a numerical scale, and 5e to f shows the same conditional probabilities with calibrated phrases on the IPCC scale. Figure 5a and e correspond to OK with GA. The Figure 5b and f are for robust REML E-BLUP. Figures 5c and g correspond to the indicator kriging predictions. The maps quantifying the uncertainty would be used with the spatial predictions by the sponsors (S2) and information user (S1) to identify sites where trials can be established, locations with $\mathrm{Se_{sol}}$ < 1.49 µg kg^{-1}.

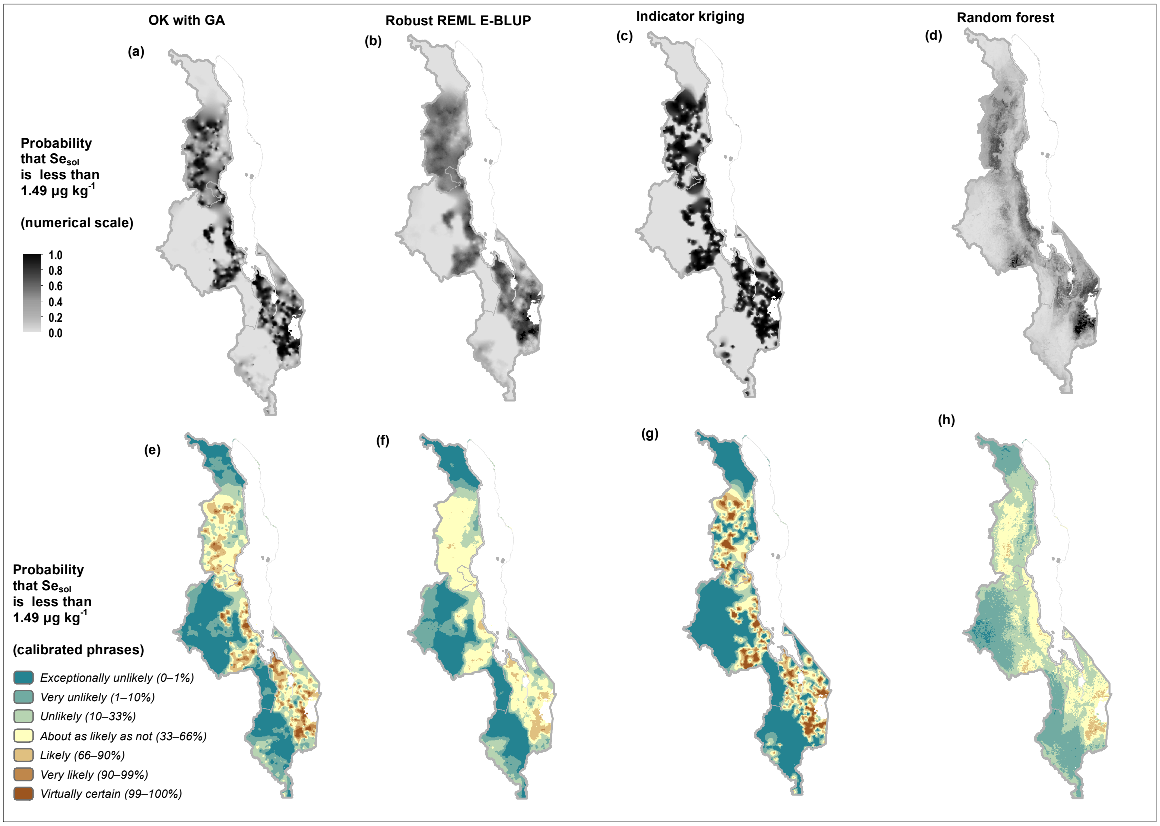


Figure 5. Probability that soil Se concentration does not exceed $\mathrm{Se_{threshold}}$, 1.49 µg $\mathrm{kg^{-1}}$, expressed on a numerical scale (a to d) and according to calibrated phrases (e to h) for spatial predictions by OK with GA (a & e), robust REML E-BLUP (b & f), indicator kriging (e & g) and random forest (d & h).

5.2 Quantile random forest algorithm for spatial prediction

All the available covariates (see Table S1) and the spatial coordinates were used to fit a random forest model. Figure 6 shows the box plots of variable importance of original predictors (green) compared to the minimum, mean and maximum importance of randomly shuffled shadow predictors (dark blue) as computed by the Boruta algorithm from 100 repetitions. No covariate was deemed unimportant because all the variables have larger importance than the randomly shuffled shadow predictors (dark blue). The results also showed that downscaled mean annual precipitation and spatial coordinates were the most important covariates. The spatial coordinates were among the three most important covariates for the random forest algorithm, and this reflected the strong spatial autocorrelation shown the variogram for $\mathrm{Se_{sol}}$ (Figure A2). Therefore, spatial structure in random forest was modelled splitting the area based on north-south and east-west directions.

Summary statistics for the out-of-bag cross-validation are presented in Table 1. Table 4 shows the parameters for the random forest algorithm, the out-of-bag MSE and R^2 were 0.25 ands 63.5%, respectively. The spatial predictions of the random forest

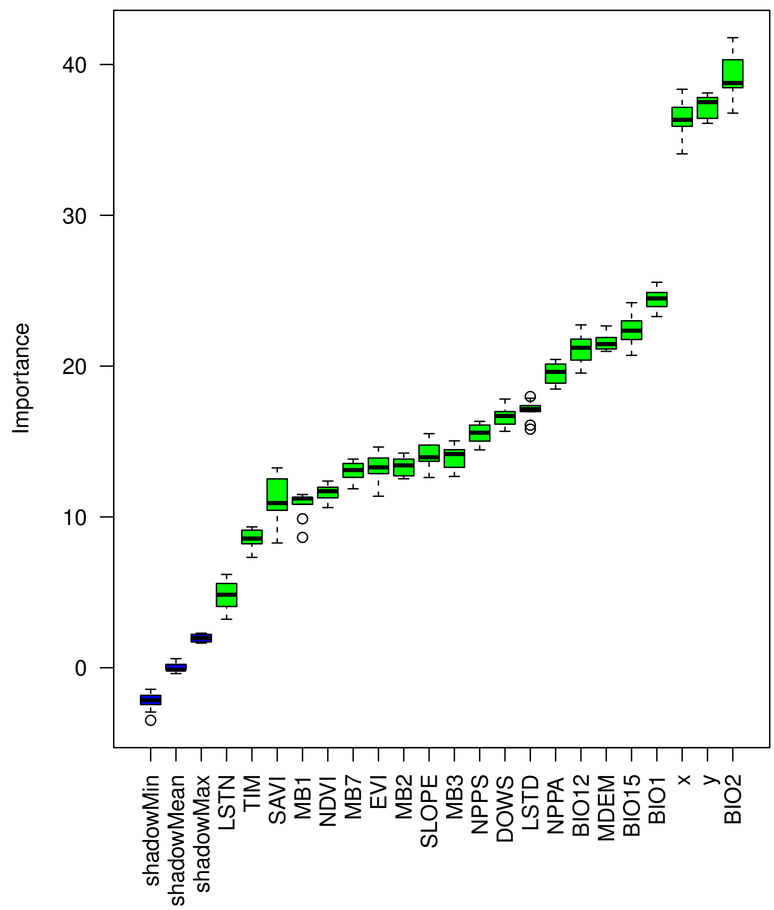


Figure 6. Boxplots of variable importance of original predictors (green) compared to the minimum, mean and maximum importance of randomly shuffled shadow predictors (dark blue) as computed by the Boruta algorithm from 100 repetitions.

Table 4. Parameters of the random forest algorithm for prediction of $\mathrm{Se}_{\mathrm{sol}}$.

Number of trees	Predictors	$m_{try}{}^{\dagger}$	Out-of-Bag MSE	Out-of-Bag R^2
1000	21	4	0.253	0.635

$m_{try}{}^{\dagger}$ number of randomly chosen variables.

are shown in Figure 3c. This map is similar to those produced by the LMMs, there are higher concentrations of Se in central, northern and southern west parts of Malawi. The conditional probability of $\mathrm{Se}_{\mathrm{sol}}$ less than 1.49 µg kg^{-1}, expressed on numerical scale and according to calibrated phrases are presented in Figure 5d and h, respectively.

5.3 How can stakeholder (CS) compare predictions?

The posterior measures of accuracy and precision for spatial prediction, for OK with GA, robust REML E-BLUP and random forest, computed after validation with an independent dataset are presented in Table 5. The mean error for OK is 0.07 on the GA transformed scale and this is close to zero. The mean error for robust REML E-BLUP and quantile random forest are -0.050 and -0.049 log (μg kg^{-1}), respectively. The mean error for both robust REML E-BLUP and random forest are close to zero. Robust REML produces more accurate predictions of the conditional mean than random forest. The MSESS is the only statistic comparable amongst the three spatial prediction methods. The score for the OK with GA is 1.04, and for robust REML E-BLUP and random forest the scores are 0.94 and 0.87, respectively. All the models for spatial prediction performed well.

Table 5. Accuracy of the predictions computed with an independent validation dataset (n=160).

Method	Measure			
	ME	MSE	RMSE	MSESS
Ordinary Kriging	0.07	0.400	0.448	1.04
robust REML E-BLUP	-0.050	0.212	0.326	0.94
Quantile random forest	-0.049	0.216	0.342	0.87

ME–mean error; MSE–mean square error; RMSE–root mean square error, MSESS–mean square error skill score.

Figure 7 shows the coverage probability plots for the OK with GA, robust REML E-BLUP and random forest. Figure 7a to c, shows the coverage probabilities with 95% confidence interval for the cross-validations of OK with GA, robust REML E-BLUP and random forest predictions. All show some deviation from the bisector, with the largest difference seen for the robust REML E-BLUP predictions. Figures 7d to f, show the corresponding coverage probabilities for the validation sites (n=160). The closest agreement is for OK with GA predictions, with the 95% confidence interval included the bisector for all probabilities less than 0.9. The largest deviation is seen for the robust REML E-BLUP. The coverage is less than the specified probability showing that the prediction intervals are to conservative. This may reflect a bias introduced by the tuning parameter or the consistency correction.

Figure 8a shows the proportion of points at which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t based on cross-validation, and (b) shows the proportion of those sites at which the observed $\mathrm{Se}_{\mathrm{sol}}$ meets the condition. The robust REML E-BLUP has the smallest proportion of mapped area at which the $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t. Ordinary kriging has the largest proportion, whilst the proportion of random forest and indicator kriging are similar. Figure 9 shows the proportion of points based on validation with an independent dataset. Note that for both the cross-validation and independent validation the number of points for which the estimated $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds large values goes to zero or small values, so the proportion of the points for which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ may vary strongly, or the be undefined. The random forest have the smallest proportion

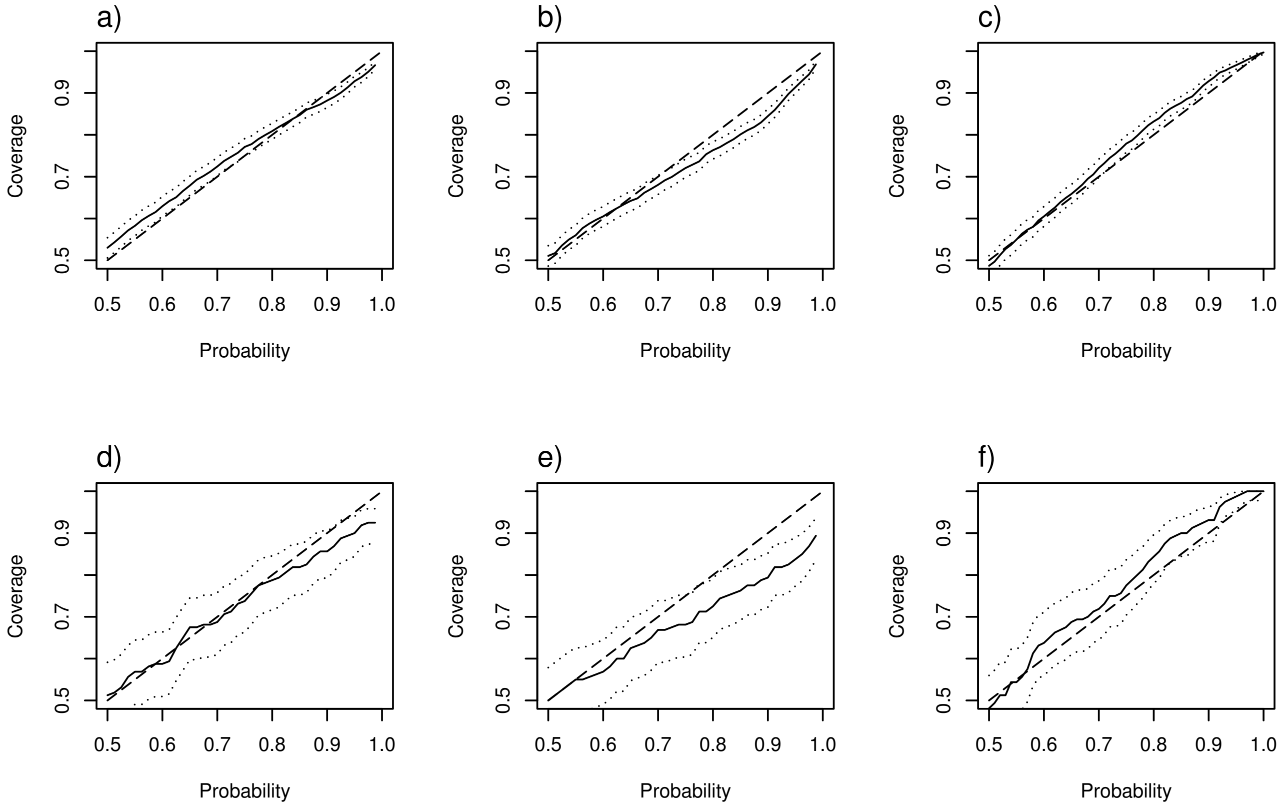


Figure 7. Plot of estimated coverage of prediction intervals from 0.5 to 0.99 at the validation locations with their 95% confidence intervals (dotted lines), for: (a) ordinary kriging after leave-one-out cross-validation, (b) robust REML E-BLUP after K-fold cross-validation, (c) random forest after validation with out-of-bag data, (d) ordinary kriging after validation with an independent dataset, (e) robust REML E-BLUP after validation with an independent dataset, and (f) random forest after validation with an independent dataset.

of points at which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t. The proportion of points for indicator kriging, robust REML E-BLUP and ordinary kriging are similar.

6 Discussion

6.1 Can we meet requirements of the soil information user?

In this study we postulate a set of stakeholders who require soil information to make a decision on where to locate agronomic biofortification trials to address human Se deficiency. Our objective is to appraise three common approaches to digital soil mapping in the light of a decision framework for such stakeholders.

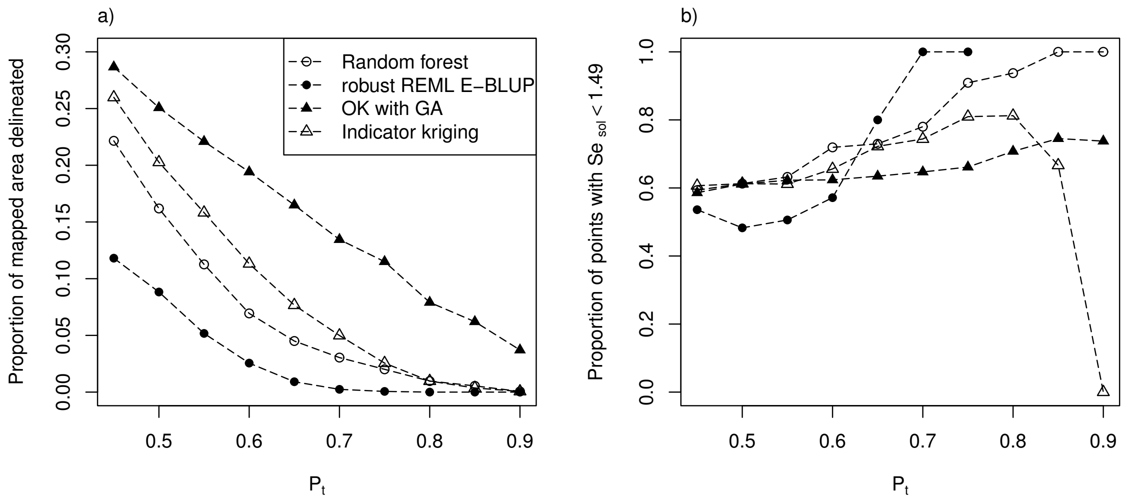


Figure 8. Plots for (a) the proportion of points at which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t based on cross-validation, and (b) shows the proportion of those sites at which the observed soil Se meets the condition.

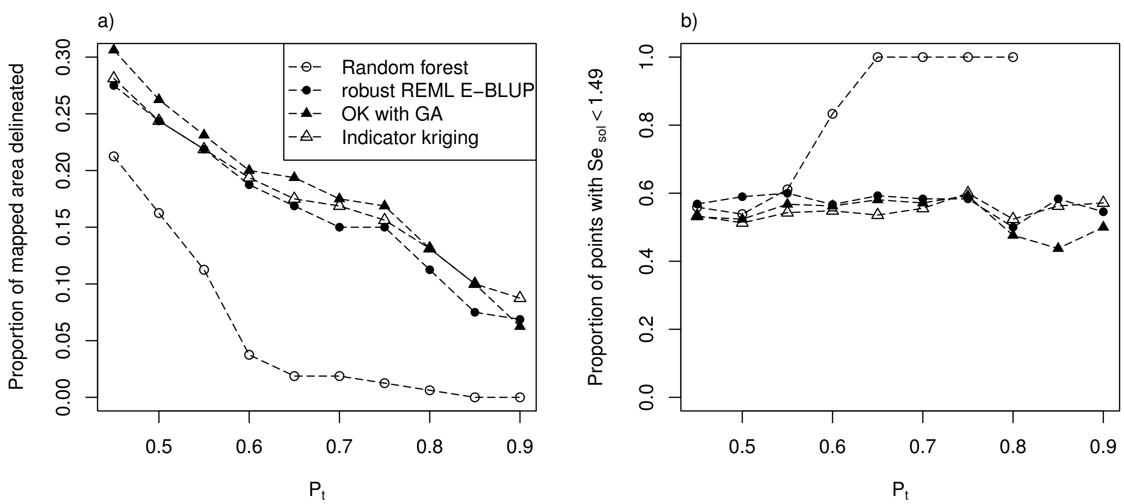


Figure 9. Plots for (a) the proportion of points at which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t based on validation with an independent dataset (n=160), and (b) shows the proportion of those sites at which the observed soil Se meets the condition with independent dataset.

The first observation is that the predictions, as measured by an independent validation set, are more precise for the robust REML E-BLUP predictions from the linear mixed model (smaller mean square error) and have a smaller bias. These differences are not large. However, it is notable that this smaller prediction error variance is achieved by the robust REML E-BLUP with a smaller final covariate set than was used by the random forest, and a model of spatial dependence for the unexplained variation in the target variables.

Second, we note that there is substantial uncertainty in the spatial predictions of $\mathrm{Se}_{\mathrm{sol}}$ by all three methods. Even though the maps of $\mathrm{Se}_{\mathrm{sol}}$ show clear structure, which may encourage the user to assume that it has succeeded in representing variation of this property, a local prediction may have substantial error. That is why the uncertainty of these predictions should be explicitly quantified and presented a way that is accessible to the stakeholders (S1, S2 and S3). The objective is that the stakeholder understands the uncertainty and is equipped to use the information with appropriate caution, and attention to the possible outcomes, when deciding where to locate the agronomic trials to minimise losses. The key challenge is how to characterise the sources of errors with probability distributions (Heuvelink, 2018) and attention should be paid to the assumptions made when modelling uncertainty (Szatmári and Pásztor, 2019).

Geostatistical models and ML methods permit a sophisticated and robust quantification of the uncertainty in spatial information but communicating uncertainty is a challenge. Communicating uncertainty depends on the subject matter and knowledge of the target audience (Milne et al., 2015). Uncertainty in spatial predictions can be quantified by either using general measures (prediction error variances and prediction intervals) or methods based on interpreting probabilities based on exceeding a threshold. Prediction intervals are commonly used in DSM to quantify and communicate the uncertainty of spatial predictions, and this has been applied in many Pedometrics studies. Chagumaira et al. (2021) found that the diverse group of stakeholders (S1,S2, and S3), find methods of communicating uncertainty based on specific interpretation of the uncertainty to be clearer and easier to interpret (e.g.,- the probability that the concentration of a micronutrient in grain does not exceed a nutritionally-significant threshold) than general measures such as prediction intervals. Chagumaira et al. (2022) showed that a further step, from just computing conditional probabilities, should be taken to consider how uncertainty can be communicated to a range of end-users for decision-making.

In our case study we have shown conditional probabilities of not exceeding a threshold can be computed for the spatial predictive methods we used. These conditional probabilities will be used by stakeholders (S1 to S3) to make decisions where to establish the trials. Using the decision process, stakeholders will be provided with information about $\mathrm{Se}_{\mathrm{sol}}$ concentration not exceeding a hypothetical threshold, $1.49\ \mu\mathrm{g\ kg}^{-1}$, to locate sites with inadequate Se supply (I3) in the soil at farm scale (I2). A community would be listed for a trial because the predicted $\mathrm{Se}_{\mathrm{sol}}$ concentration falls below a threshold. Given the uncertainties in the spatial predictions, it might be possible to set up trials at a location where there is sufficient Se supply. Sites for a trial should only be set up if the prediction is found to be correct with respect to the threshold. Stakeholders should be assisted when making this decision. Chagumaira et al. (2022) showed that a critical probability value (P_t), at which a stakeholder would judge an intervention to address MND deficiencies, can be elicited from a diverse group. The P_t at which stakeholders would judge for an intervention reflects the stakeholders judgement of the losses under different outcomes.

The use of P_t to illustrate different proportion of mapped area at which $\mathrm{P}\left(\mathrm{Se}_{\mathrm{sol}} < 1.49\right)$ exceeds different values of P_t, where trials may be established, has been presented in Figures 8 and 9. Stakeholders can set a critical value of their choice considering their different circumstances and information needs. For example, a government research organisation with good and well established linkages with communities, confident that it can engage with minimum disruption if a local sample shows $\mathrm{Se}_{\mathrm{sol}} > 1.49$, can set P_t to 0.8. On the other hand a non-governmental organisation beginning local engagement with a community, that maybe more nervous about project withdrawal, may set P_t to 0.8 and will have a smaller proportion of the mapped area where the concentration of Se_{sol} meets the condition of $\mathrm{Se}_{\mathrm{sol}} < 1.49$.

6.2 How the spatial predictive methods performed?

In our example, we considered the spatial prediction of $\mathrm{Se}_{\mathrm{sol}}$ with OK, robust REML E-BLUP and quantile random forest. Ordinary kriging and robust REML E-BLUP are models for spatial prediction that capture the spatial dependence of soil variation (Webster, 2000; Webster and Oliver, 2007). Whereas, random forest are non-spatial and do not capture spatial dependence of soil variation (Heuvelink and Webster, 2022), although this is the case spatial dependence will be implicitly captured through the input variables. We expected similar results under these models.

In our case study, $\mathrm{Se}_{\mathrm{sol}}$ was transformed by natural logarithms to make the assumption of normality plausible for robust REML E-BLUP. The log transformed $\mathrm{Se}_{\mathrm{sol}}$ variable was used also for spatial prediction with random forest. After the transformation, the data showed traits of being bimodal (see Figure S3). We took a further step to use more robust transformation, Gaussian anamorphosis. The transformed $\mathrm{Se}_{\mathrm{sol}}$ (see Figure S2) was used for spatial prediction with OK.

The cross-validation coverage probabilities are shown in Figure 7a,b and c. Our results shows that after cross-validation of the OK model indicated it to be a good model and produced accurate predictions of the conditional mean. The coverage probabilities of ordinary kriging, however were above and below the nominal coverage at lower and higher probabilities, respectively but with quite small deviance. The coverage probabilities for robust REML nearly follow the nominal coverage at lower probabilities only and then deviate the bisector at larger probabilities. Robust REML E-BLUP produced accurate predictions of the conditional mean and seems to underestimate the error. The random forest, have much better prediction intervals when compared robust REML E-BLUP predictions. The coverage probabilities for random forest deviates from the nominal coverage at higher probabilities. Therefore, the random forest overestimates the uncertainty. Spatial predictions of random forest are less extreme and are central to the mean of the distribution. However, random forest has have the larger error (Table 5).

The coverage probabilities for the independent dataset was wide and nearly followed the bisector line for OK and random forest. However, OK with GA and random forest deviate from the bisector in a opposite directions at larger probabilities. Random forest has uncertainty is better quantified, so the tendency to shrinkage in the prediction is not necessarily a strength. The mapped validations also show that OK with GA is arguably better. The validation with independent dataset also confirm the underestimation of uncertainty by robust REML E-BLUP. The underestimation of uncertainty may be due to the fact we had to use robust methods, which may have down weighted observations in the tails too strongly, such that we do not see the evidence for over-conservative uncertainty quantification by random forest.

6.3 Way forward

This study aimed to address challenges of the future of pedometrics (Wadoux et al., 2021) by analysing a decision process that can be used when mapping micronutrients at a national-scale in sub-Saharan Africa region. In study we aimed to how key users of soil information can be incorporated when designing surveys, mapping, and quantifying uncertainty of the spatial predictions. Many digital soil mapping studies put emphasis of comparing performance of machine learning algorithms and statistical methods of spatial prediction (e.g., Vaysse and Lagacherie, 2017; Szatmári and Pásztor, 2019; Makungwe et al., 2021). It is not enough to only quantify uncertainty and leave it there. This is of little use for key users of soil information who have to make decisions at farm-level, field-scale, regional level, national-level and at policy level. Lark et al. (2022) suggested the need of paying attention to 'decision-focused' measures of uncertainty regardlessly of the method of spatial prediction used. Further steps such as investigating how stakeholders use probabilistic representation of uncertainty is of paramount importance. Chagumaira et al. (2022) showed how discussions centred on how probabilistic representation of uncertain information, with diverse stakeholders can be used to elicit critical probabilities at which they would recommend an intervention.

In the GeoNutrition project, we aim at using spatial information to target areas where specific interventions would be appropriate to efficiently use the scarce financial resources. This is important because most people in countries south of the Sahara (e.g., Ethiopia, Malawi, Zambia, and Zimbabwe) mainly rely on subsistence farming for their food and income. The decision process presented in this paper would be important in addressing some questions raised when addressing micronutrient deficiencies. This decision process may be applied for a different problem, e.g., decisions on sampling, and thorough decision analysis is required when addressing such problems.

7 Conclusions

This study aimed to address challenges of the future of Pedometrics by analysing a decision process that can be used when mapping micronutrient at a national-scale in sub-Saharan Africa region. In study we aimed at showing how key users of soil information can be incorporated when designing surveys, mapping and quantifying uncertainty of the spatial predictions. We have shown how the decision-process for making decisions when using different methods for spatial prediction. The linear mixed models (ordinary kriging and robust REML E-BLUP) underestimate the uncertainty in the spatial predictions of $\mathrm{Se_{sol}}$, whereas the random forest overestimate the uncertainty. However, the decision to which method is better in providing soil information remains difficult. This study has shown the importance of cross-validation and validation of conditional probabilities used when quantifying uncertainty in spatial predictions using a critical probability threshold. This allows stakeholders S1 and S2 to make rationale decisions based on their different circumstances and information needs.

Appendix A

The variance parameters for $\mathrm{Se_{sol}}$ transformed by were estimated from fitting an exponential variogram on an empirical variogram estimated by the method-of-moments (Figure A1, Matheron, 1965). The maximum likelihood variogram functions for the null model (coordinates filtering spatial trend) for $\mathrm{Se_{sol}}$ concentration, and for enhanced vegetation index (EVI) added as a predictor are shown in Figure A2.

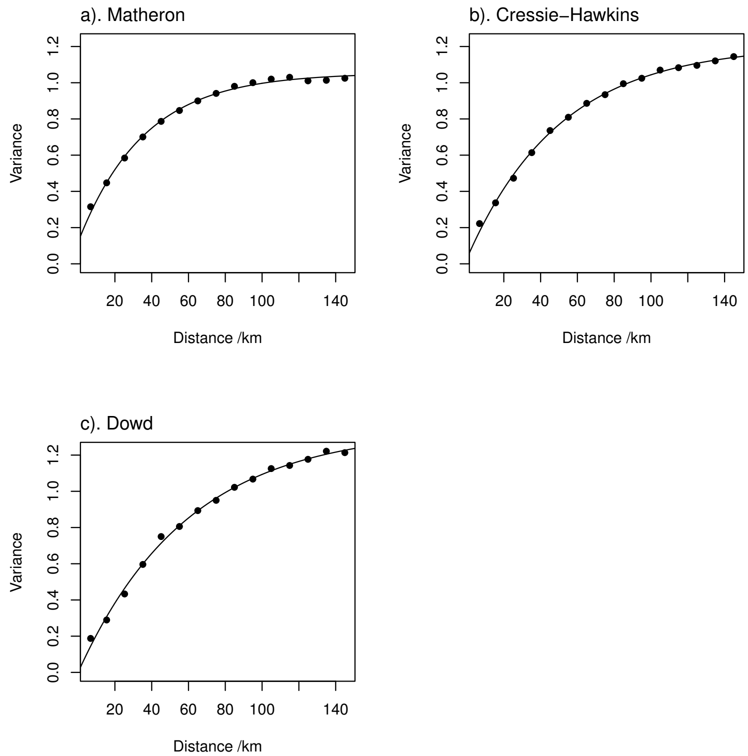


Figure A1. Variogram functions for $\mathrm{Se_{sol}}$ transformed by Gaussian anamorphosis estimated by (a) Matheron (1962) (b) Cressie and Hawkins (1980), and (c) Dowd (1984) estimators

Author contributions. The project was conceived, designed and implemented by CC, RML, AEM and MN. RML, PCN and MRB were responsible for project administration and funding. PCN and JGC supervised the data collection. All authors contributed to the preparation of the article.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. The funders of this paper were not involved in the study design or the collection, management, analysis and interpretation of the data, the writing of the report or the decision to submit the report for publication.

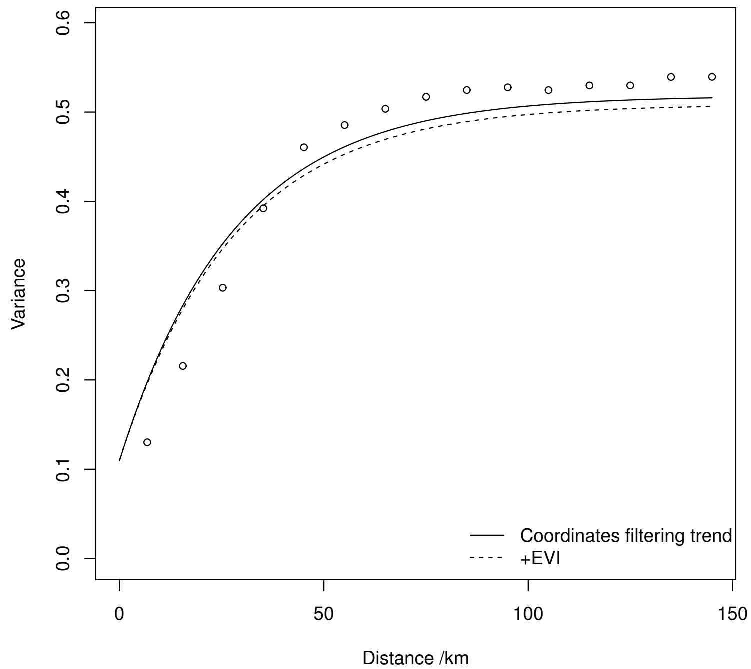


Figure A2. Variogram functions for the null model (coordinates filtering spatial trend) for $\mathrm{Se}_{\mathrm{sol}}$ concentration, and for successive model with selected environmental covariates added as predictors.

Acknowledgements. This work was supported funded by the Nottingham-Rothamsted Future Food Beacon Studentships in International Agricultural Development and GeoNutrition project funded by the Bill & Melinda Gates Foundation (BMGF) [INV-009129]. The authors gratefully acknowledge the contributions made to this research by the participating farmers and field sampling teams from the Department of Agricultural Research Services, and Lilongwe University of Agriculture and Natural Resources.

References

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-De-Forges, A., Morgan, C., Roudier, P., Poggio, L., and Mulder, V.: Impressions of digital soil maps: The good, the not so good, and making them ever better, Geoderma Regional, 20, e00 255, https://doi.org/10.1016/j.geodrs.2020.e00255, 2020.

Belay, A., Joy, E. J. M., Chagumaira, C., Zerfu, D., Ander, E. L., Young, S. D., Bailey, E. H., Lark, R. M., and Broadley, M. R.: Selenium Deficiency Is Widespread and Spatially Dependent in Ethiopia, Nutrients, https://doi.org/10.3390/nu12061565, 2020.

Blaker, H.: Confidence curves and improved exact confidence intervals for discrete distributions, Canadian Journal of Statistics, 28, 783–798, 2000.

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., Mcgrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Broadley, M. R., Lark, R. M., and Nalivata, P. C.: Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi, Scientific Reports, https://doi.org/10.1038/s41598-022-12014-w, 2022.

Bramer, M.: Principles of Data Mining, Springer, London, https://doi.org/10.1007/978-1-4471-7493-6_1, 2020.

Breiman, L.: Random Forests, Machine learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brus, D., De Gruijter, J., and Van Groenigen, J.: Designing spatial coverage samples using the k-means clustering algorithm, Developments in Soil Science, 31, 183–192, 2006.

Burgess, T. M. and Webster, R.: Optimal interpolation and isarithmic mapping of soil properties .1. the semi-variogram and punctual kriging, Journal of Soil Science, 31, 315–331, https://doi.org/10.1111/j.1365-2389.1980.tb02084.x, 1980.

Chagumaira, C., Chimungu, J. G., Gashu, D., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M.: Communicating uncertainties in spatial predictions of grain micronutrient concentration, Geoscience Communication, 4, 245–265, https://doi.org/10.5194/gc-4-245-2021, 2021.

Chagumaira, C., Nalivata, P. C., Chimungu, J. G., Gashu, D., Broadley, M., Milne, A. E., and Lark, M.: Stakeholder interpretation of probabilistic representations of uncertainty in spatial information: an example on the nutritional quality of staple crops, International Journal of Geographical Information Science, https://doi.org/10.1080/13658816.2021.2020278, 2022.

Cressie, N.: Block kriging for lognormal spatial processes, Mathematical Geology, 38, 413–443, 2006.

Cressie, N. and Hawkins, D. M.: Robust estimation of the variogram: I, Journal of the international Association for Mathematical Geology, 12, 115–125, 1980.

de Gruijter, J. J., Brus, D. J., Bierkens, M. F. P., and Knotters, M.: Sampling for natural resource monitoring, Springer, https://doi.org/10.1007/3-540-33161-1, 2006.

Dent, D. and Young, A.: Soil survey and land evaluation., George Allen & Unwin., 1981.

Diggle, P., Liang, K.-Y., and Zeger, S.: Analysis of longitudinal data, Clarendon Press, 1994.

Dowd, P.: The variogram and kriging: robust and resistant estimators, in: Geostatistics for natural resources characterization, pp. 91–106, Springer, 1984.

Drapper, N. R. and Smith, H.: Applied Regression Analysis, Third Edition, John Wiley & Sons, Inc, 1998.

Gashu, D., Lark, R. M., Milne, A. E., Amede, T., Bailey, E. H., Chagumaira, C., Dunham, S. J., Gameda, S., Kumssa, D. B., Mossa, A. W., Walsh, M. G., Wilson, L., Young, S. D., Ander, E. L., Broadley, M. R., Joy, E. J. M., and McGrath, S. P.: Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia, Science of the Total Environment, 733, https://doi.org/10.1016/j.scitotenv.2020.139231, 2020.

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., Towett, E. K., Walsh, M. G., Wilson, L., Young, S. D., and Broadley, M. R.: The nutritional quality of cereals varies geospatially in Ethiopia and Malawi, Nature, 594, 71–76, https://doi.org/10.1038/s41586-021-03559-3, 2021.

GeoNutrution: GeoNutrition Project, www.geonutrition.com/, 2017.

Giasson, E., Van Es, C., Van Wambeke, A., and Bryant, R. B.: Assessing the economic value of soil information using decision analysis techniques, Soil science, 165, 971–978, 2000.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning Second Edition, vol. 27 of *Springer Series in Statistics*, Springer, New York, https://doi.org/0.1007/978-0-387-84858-7_1, 2009.

Heuvelink, G. B. and Webster, R.: Spatial statistics and soil mapping: A blossoming partnership under pressure, Spatial Statistics, p. 100639, https://doi.org/10.1016/j.spasta.2022.100639, 2022.

Heuvelink, G. B. M.: Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling, in: Pedometrics, edited by McBratney, A. B., Minasny, B., and Stockmann, U., pp. 439–461, Springer International Publishing, https://doi.org/10.1007/978-3-319-63439-5_14, 2018.

Howard, R. A.: Information value theory, IEEE Transactions on systems science and cybernetics, 2, 22–26, 1966.

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., and Broadley, M. R.: Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi, Scientific Reports, 3, https://doi.org/10.1038/srep01425, 2013.

ISRIC - World Soil: GlobalSoilMap.Net, www.isric.org/projects/globalsoilmapnet, 2009.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An introduction to statistical learning: with applications in R, Springer, New York, 2013.

Johnson, C. C. and Breward, N.: 2004 G-BASE : Geochemical Baseline Survey of the Environment, Tech. Rep. CR/04/016N, British Geological Survey, Keyworth, 2004.

Joy, E. J., Kalimbira, A. A., Gashu, D., Ferguson, E. L., Sturgess, J., Dangour, A. D., Banda, L., Chiutsi-Phiri, G., Bailey, E. H., Langley-Evans, S. C., et al.: Can selenium deficiency in Malawi be alleviated through consumption of agro-biofortified maize flour? Study protocol for a randomised, double-blind, controlled trial, Trials, 20, 1–9, https://doi.org/10.1186/s13063-019-3894-2, 2019.

Joy, E. J. M., Broadley, M. R., Young, S. D., Black, C. R., Chilimba, A. D. C., Ander, E. L., Barlow, T. S., and Watts, M. J.: Soil type influences crop mineral composition in Malawi, Science of the Total Environment, 505, 587–595, https://doi.org/10.1016/j.scitotenv.2014.10.038, 2015.

Joy, E. J. M., Kalimbira, A. A., Sturgess, J., Banda, L., Chiutsi-Phiri, G., Manase, H., Gondwe, J., Ferguson, E. L., Kalumikiza, Z., Bailey, E. H., Young, S. D., Matandika, L., Mfutso-Bengo, J., Millar, K., Niksic, M., de la Revilla, L. S., Likoswe, B. H., Phuka, J. C., Phiri, F. P., Lark, R. M., Gashu, D., Langley-Evans, S. C., Ander, E. L., Lowe, N. M., Dangour, A. D., Nalivata, P. C., Broadley, M. R., and Allen, E.: Biofortified Maize Improves Selenium Status of Women and Children in a Rural Community in Malawi: Results of the Addressing Hidden Hunger With Agronomy Randomized Controlled Trial, Frontiers in Nutrition, 8, https://doi.org/10.3389/fnut.2021.788096, 2022.

Justice, C. O., Vermote, E., Townshend, J. R., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., et al.: The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research, IEEE transactions on geoscience and remote sensing, 36, 1228–1249, 1998.

Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., Zimmermann, N., Linder, H., and Kessler, M.: Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digital Repository, Scientific Data, 4, 10.1038/sdata.2017.122, 2017.

Künsch, H., Papritz, A. J., Schwierz, C., and Stahel, W. A.: Robust estimation of the external drift and the variogram of spatial data, in: ISI 58th World Statistics Congress of the International Statistical Institute, Eidgenössische Technische Hochschule Zürich, Zürich, https://doi.org/10.3929/ethz-a-009900710, 2013.

Kursa, M. B. and Rudnicki, W. R.: Feature Selection with the Boruta Package, Journal of Statistical Software, 36, 1–13, 2010.

Lark, R. and Lapworth, D.: The offset correlation, a novel quality measure for planning geochemical surveys of the soil by kriging, Geoderma, 197, 27–35, 2013.

Lark, R., Chagumaira, C., and Milne, A.: Decisions, uncertainty and spatial information, Spatial Statistics, p. 100619, https://doi.org/10.1016/j.spasta.2022.100619, 2022.

Lark, R. M.: A comparison of some robust estimators of the variogram for use in soil survey, European Journal of Soil Science, 51, 137–157, https://doi.org/10.1046/j.1365-2389.2000.00280.x, 2000.

Lark, R. M.: Controlling the marginal false discovery rate in inferences from a soil dataset with -investment, European Journal of Soil Science, 68, 221–234, https://doi.org/10.1111/ejss.12413, 2017.

Lark, R. M. and Knights, K. V.: The implicit loss function for errors in soil information, Geoderma, 251-252, 24–32, https://doi.org/10.1016/j.geoderma.2015.03.014, 2015.

Lark, R. M. and Marchant, B. P.: How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters?, Geoderma, 319, 89–99, https://doi.org/10.1016/j.geoderma.2017.12.022, 2018.

Lark, R. M., Cullis, B. R., and Welham, S. J.: On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML, European Journal of Soil Science, 57, 787–799, https://doi.org/10.1111/j.1365-2389.2005.00768.x, 2006.

Lark, R. M., Ander, E. L., Cave, M. R., Knights, K. V., Glennon, M. M., and Scanlon, R. P.: Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland, Geoderma, 226, 64–78, https://doi.org/10.1016/j.geoderma.2014.03.002, 2014.

Ma, T., Brus, D. J., Zhu, A.-X., Zhang, L., and Scholten, T.: Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps, Geoderma, 370, 114 366, https://doi.org/10.1016/j.geoderma.2020.114366, 2020.

Makungwe, M., Chabala, L. M., Chishala, B. H., and Lark, R. M.: Performance of linear mixed models and random forests for spatial prediction of soil pH, Geoderma, 397, https://doi.org/10.1016/j.geoderma.2021.115079, 2021.

Malone, B. P., Odgers, N. P., Stockmann, U., Minasny, B., and McBratney, A. B.: Digital mapping of soil classes and continuous soil properties, in: Pedometrics, edited by McBratney, A. B., Minasny, B., and Stockmann, U., pp. 373–413, Springer, Dordrecht, https://doi.org/10.1007/978-3-319-63439-5_12, 2018.

Mastrandrea, M., Field, C.B.and Stocker, T., Edenhofer, O., Ebi, K., Frame, D., Held, H., Kriegler, E., Mach, K., Matschoss, P., Plattner, G., Yohe, G. W., and Zwiers, F. W.: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties, Tech. rep., Intergovernmental Panel on Climate Change (IPCC), 2010.

Matheron, G.: Traité de géostatistique appliquée, 14, Editions Technip, 1962.

Matheron, G.: Regionalized variables and their estimation: an application of random function theory to the natural sciences, Masson et CIE, Paris, 1965.

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

Meinshausen, N.: Quantile Regression Forests, Journal of Machine Learning Research, 7, 983–999, http://jmlr.org/papers/v7/meinshausen06a.html, 2006.

Milne, A. E., Glendining, M. J., Lark, R. M., Perryman, S. A. M., Gordon, T., and Whitmore, A. P.: Communicating the uncertainty in estimated greenhouse gas emissions from agriculture, Journal of Environmental Management, 160, 139–153, https://doi.org/10.1016/j.jenvman.2015.05.034, 2015.

Minasny, B. and McBratney, A. B.: A conditioned Latin hypercube method for sampling in the presence of ancillary information, Computers and Geosciences, 32, 1378–1388, https://doi.org/10.1016/j.cageo.2005.12.009, 2006.

MINES ParisTech / ARMINES: RGeostats: The Geostatistical R Package, 2022.

Nussbaum, M., Papritz, A. J., Baltensweiler, A., and Walthert, L.: Organic Carbon Stocks of Swiss Forest Soils, Tech. rep., ETH Zurich, Institute of Terrestrial Ecosystems, https://doi.org/10.3929/ethz-a-007555133, 2012.

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

Papritz, A.: Tutorial and Manual for Geostatistical Analyses with the R package georob, https://cran.r-project.org/web/packages/georob/vignettes/georob_vignette.pdf, 2021.

Papritz, A. and Schwierz, C.: georob: Robust Geostatistical Analysis of Spatial Data. R Package version 0.3-14, https://cran.r-project.org/web/packages/georob/index.html, 2021.

Peterson, M.: An Introduction to Decision Theory, Cambridge Introductions to Philosophy, Cambridge University Press, 2 edn., https://doi.org/10.1017/9781316585061, 2017.

Phiri, F. P., Ander, E. L., Bailey, E. H., Chilima, B., Chilimba, A. D. C., Gondwe, J., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Phuka, J. C., Salter, A., Suchdev, P. S., Watts, M. J., Young, S. D., and Broadley, M. R.: The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales, Scientific Reports, 9, https://doi.org/10.1038/s41598-019-43013-z, 2019.

Ramsey, M., Taylor, P., and Lee, J.: Optimized contaminated land investigation at minimal overall cost to achieve fitness-for-purpose, Journal of Environmental Monitoring, 4, 809–814, 2002.

R Core Team: R: A language and environment for statistical computing, https://www.r-project.org/, 2021.

Stein, M. L.: Interpolation of Spatial Data Some theory of Kriging, Springer-Verlag New, 1999.

Szatmári, G. and Pásztor, L.: Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms, Geoderma, 337, 1329–1340, https://doi.org/10.1016/j.geoderma.2018.09.008, 2019.

Vaysse, K. and Lagacherie, P.: Using quantile regression forest to estimate uncertainty of digital soil mapping products, Geoderma, 291, 55–64, https://doi.org/10.1016/j.geoderma.2016.12.017, 2017.

Wadoux, A. M.: Using deep learning for multivariate mapping of soil with quantified uncertainty, Geoderma, 351, 59–70, https://doi.org/10.1016/j.geoderma.2019.05.012, 2019.

Wadoux, A. M.-C., Heuvelink, G. B., Lark, R. M., Lagacherie, P., Bouma, J., Mulder, V. L., Libohova, Z., Yang, L., and McBratney, A. B.: Ten challenges for the future of pedometrics, Geoderma, 401, 115 155, https://doi.org/10.1016/j.geoderma.2021.115155, 2021.

Walvoort, D. J. J., Brus, D. J., and de Gruijter, J. J.: An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means, Computers & Geosciences, 36, 1261–1267, https://doi.org/10.1016/j.cageo.2010.04.005, 2010.

Webster, R.: Quantitative and numerical methods in soil classification and survey., Clarendon Press, 1977.

Webster, R.: Is soil variation random?, Geoderma, 97, 149–163, https://doi.org/10.1016/S0016-7061(00)00036-7, 2000.

Webster, R.: Technological developments for spatial prediction of soil properties, and Danie Krige's influence on it, Journal of the Southern African Institute of Mining and Metallurgy, 115, 165–172, 2015.

Webster, R. and Beckett, P. H. T.: Quality and usefulness of maps., Nature, 219, 680–682, 1968.

Webster, R. and Lark, M. R.: Field Sampling for Environmental Science and Management., Taylor and Francis, 2013.

Webster, R. and Oliver, M. A.: Geostatistics for Natural Environmental Scientists, John Wiley & Sons Chichester, 2nd edn., https://doi.org/10.2136/vzj2002.0321, 2007.

Western, S.: Soil survey contracts and quality control, Clarendon Press, 1978.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic press, https://doi.org/10.1016/B978-0-12-385022-5.00001-4, 2011.

Wright, M. N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, Journal of Statistical Software, 77, 1–17, https://doi.org/10.18637/jss.v077.i01, 2017.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, Geophysical Research Letters, 44, 5844–5853, 2017.

Supplement for: Chagumaira et al 2022 Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework

Chagumaira, C.., Chimungu, J. G., Nalivata, P. C., Broadley, M. R., Nussbaum, M., Milne, A. E., and Lark, R. M.: Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework, *EGUsphere [preprint]*, https://doi.org/10.5194/egusphere-2022-583, 2022.

Supplement

Table S1. Environmental covariates available.

Environmental Covariate	Grid name
Mean annual temperature ($^{\circ}$C $*$ 10, 1979–2013)	BIO1
Mean diurnal range ($^{\circ}$C $*$ 10, 1979–2013)	BIO2
Mean annual precipitation (mm year^{-1}, 1979–2013)	BIO12
Mean rainfall seasonality (1979–2013)	BIO15
Distance to inland water bodies (km)	DOWS
Average night-time land surface temperature ($^{\circ}$C, 2001–2017)	LSTN
Average day-time land surface temperature ($^{\circ}$C, 2001–2017)	LSTD
Average enhanced vegetation index (2000–2016)	EVI
Average MOD13Q1 band 1 reflectance (2000–2016)	MB1
Average MOD13Q1 band 2 reflectance (2000–2016)	MB2
Average MOD13Q1 band 3 reflectance (2000–2016)	MB3
Average MOD13Q1 band 7 reflectance (2000–2016)	MB7
Normalised difference vegetation index (2000–2016)	NDVI
Soil adjusted vegetation index (200-2016)	SAVI
Average annual net primary productivity (kg m^{-2}, 2000–2015)	NPPA
Variance annual net primary productivity (2000-2015)	NPPS
Slope (%)	SLOPE
Topographic index	TIM
Elevation above mean sea level (m)	MDEM

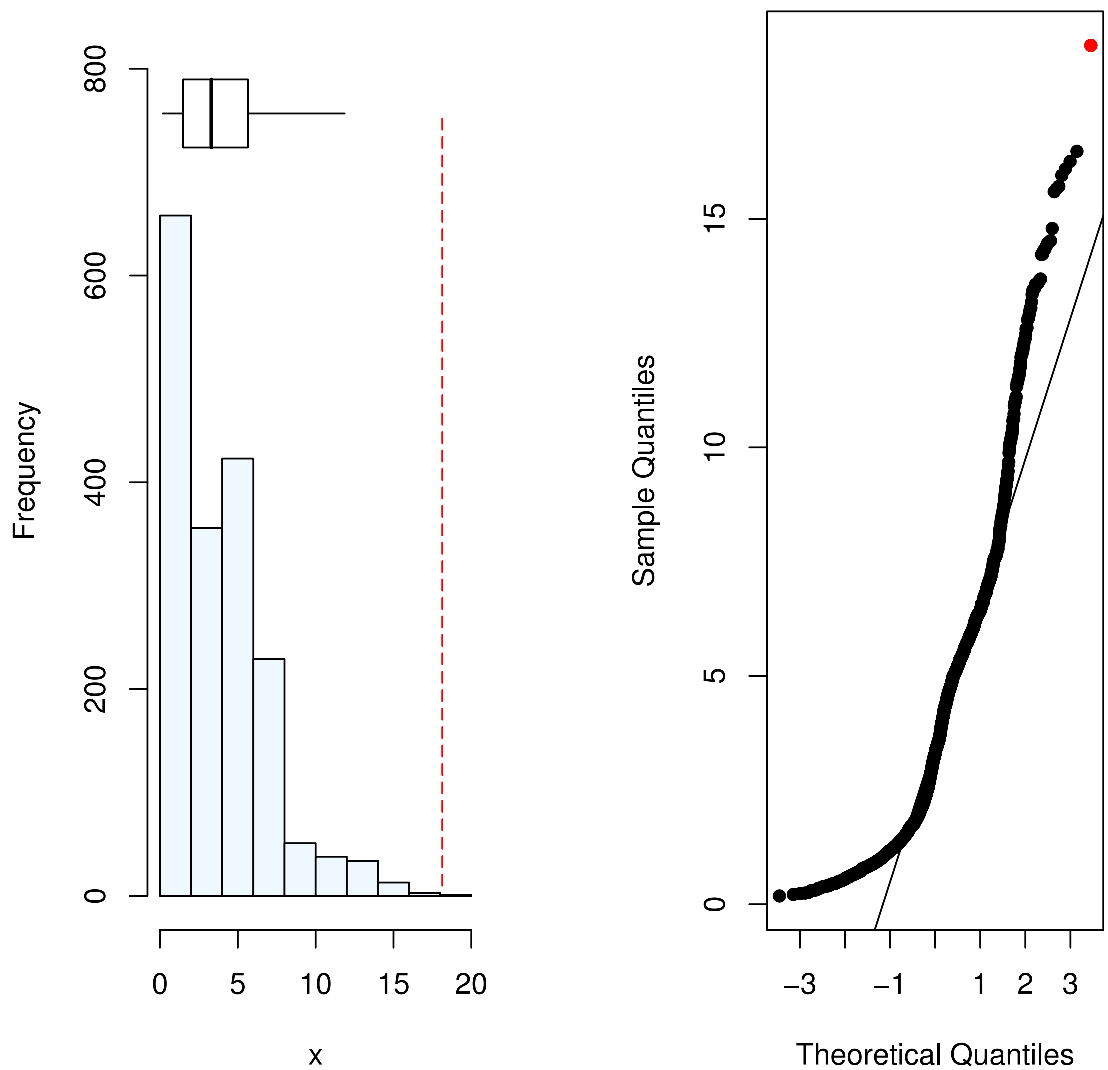


Figure S1. Histogram with boxplot and QQ plot for soil soluble Se (Se_{sol}) concentration in Malawi.

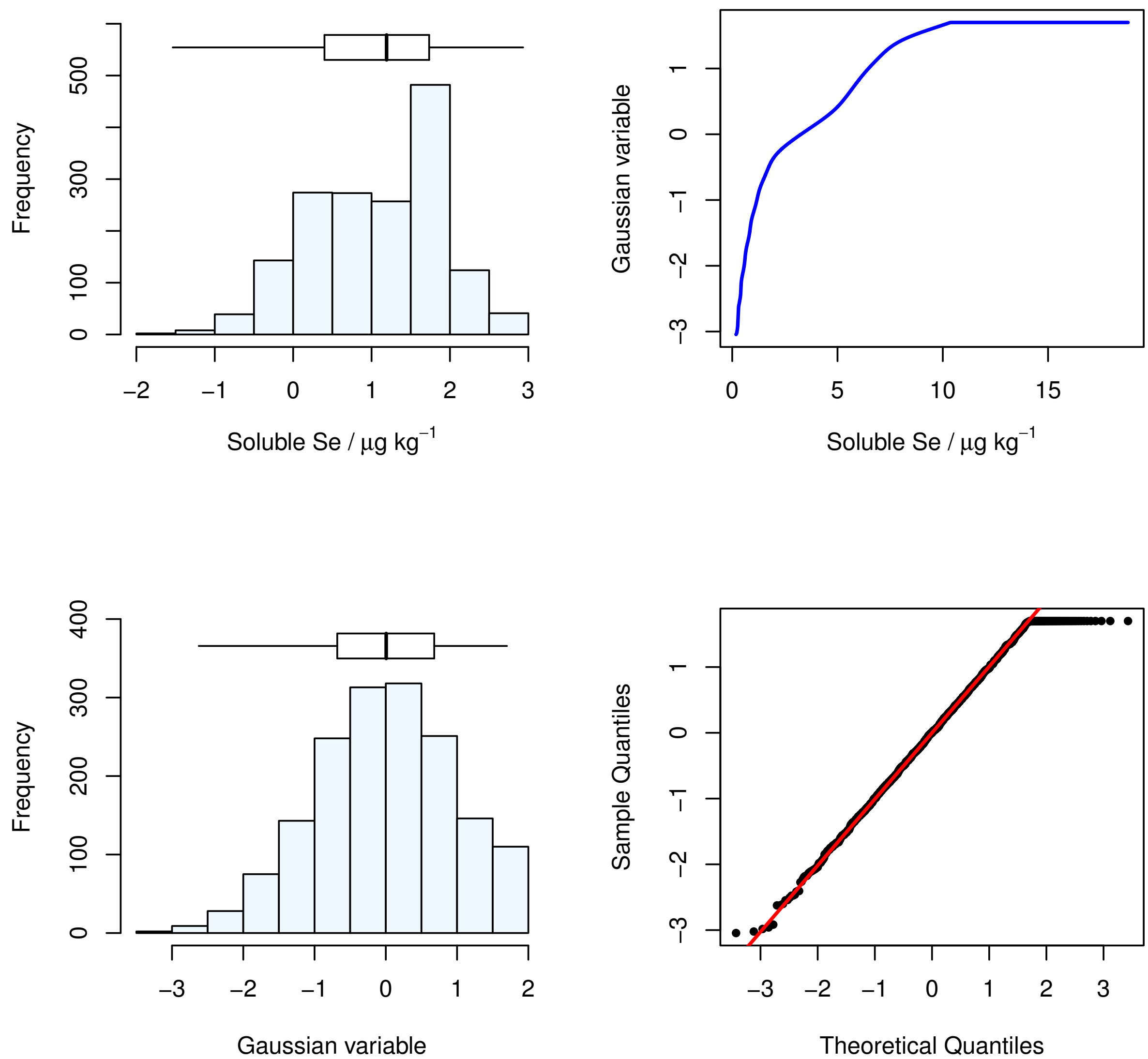


Figure S2. Histogram with boxplot and QQ plot for $\mathrm{Se_{sol}}$ transformed by Gaussian anamorphosis.

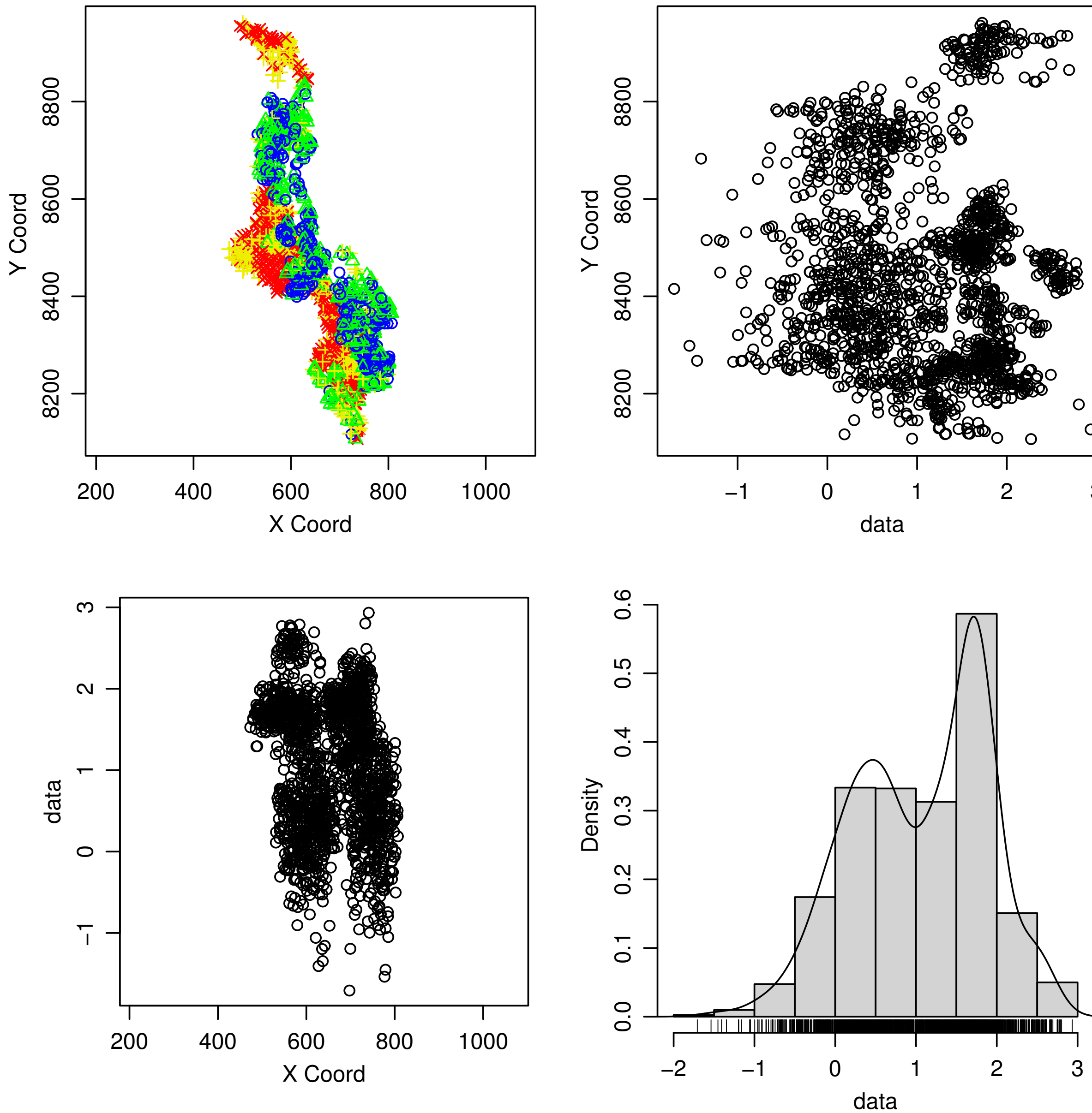
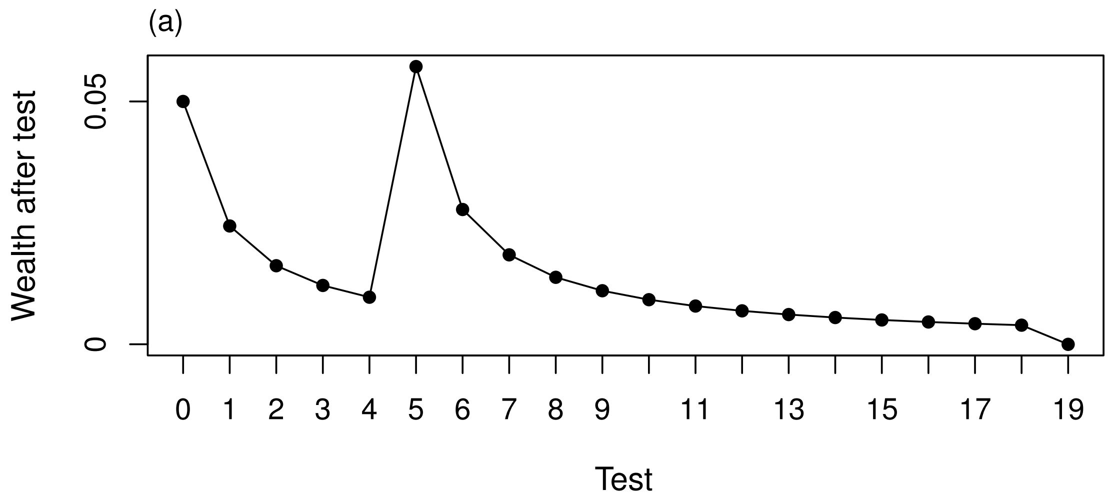


Figure S3. Post plot $\mathrm{Se}_{\mathrm{sol}}$ transformed by natural logarithms. The blue, green, yellow and red symbols in top left panel correspond to 1st, 2nd, 3rd and 4th quartiles.



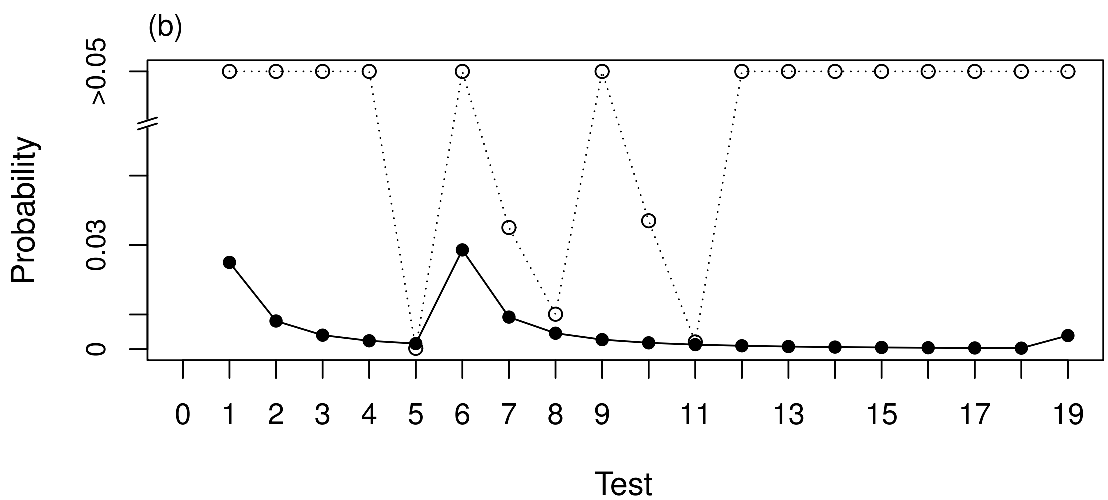


Figure S4. Ordered tests for covariate selection for $\mathrm{Se}_{\mathrm{sol}}$. This for all covariates in Table 2. The graph (a) shows α-wealth over the sequence of tests and the lower (b) shows the p-values for successive tests (open symbols) and the corresponding threshold values with marginal false discovery rate control.

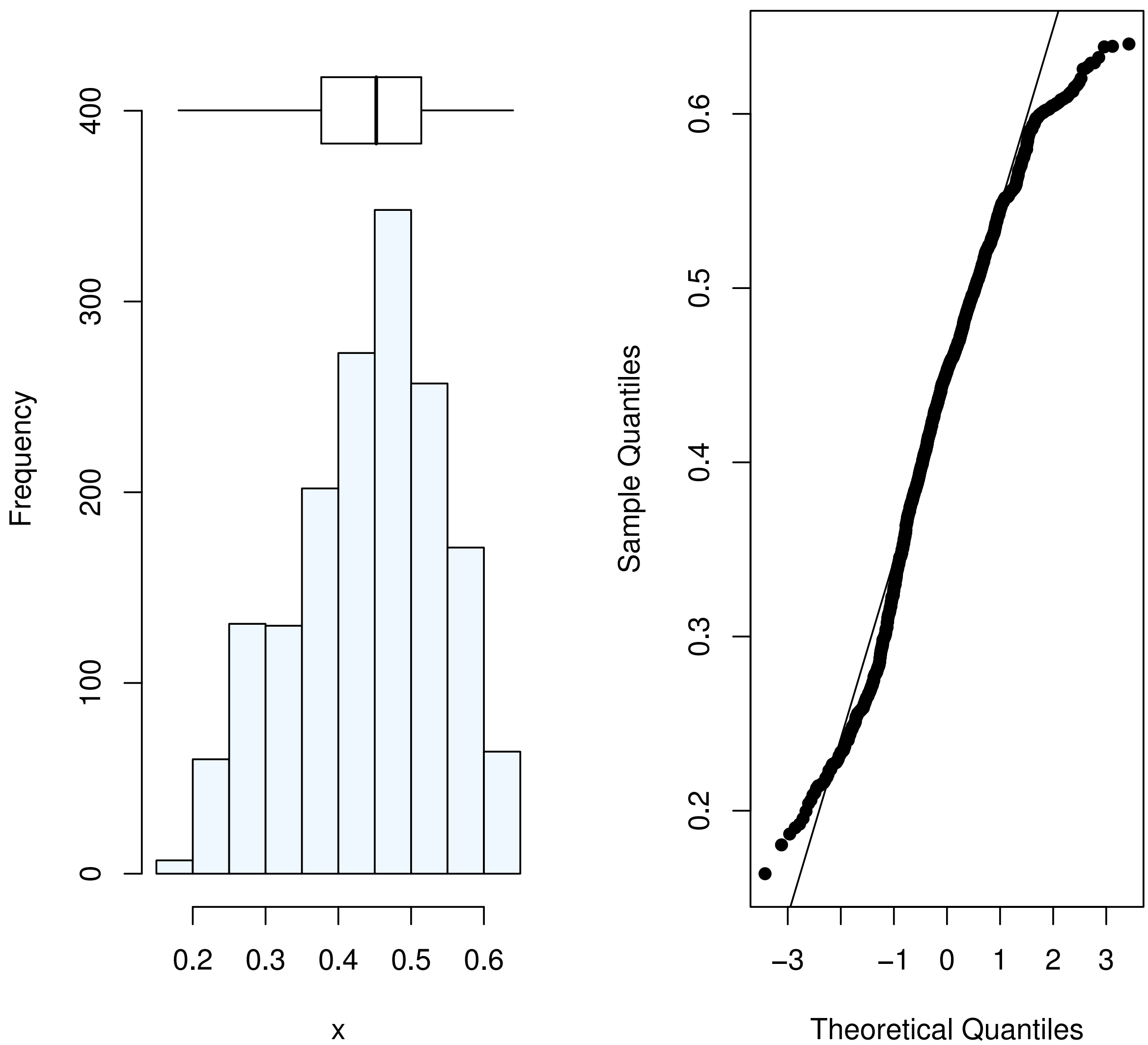


Figure S5. Histogram with boxplot and QQ plot for the prediction error variance from cross validation for the robust REML E-BLUP.

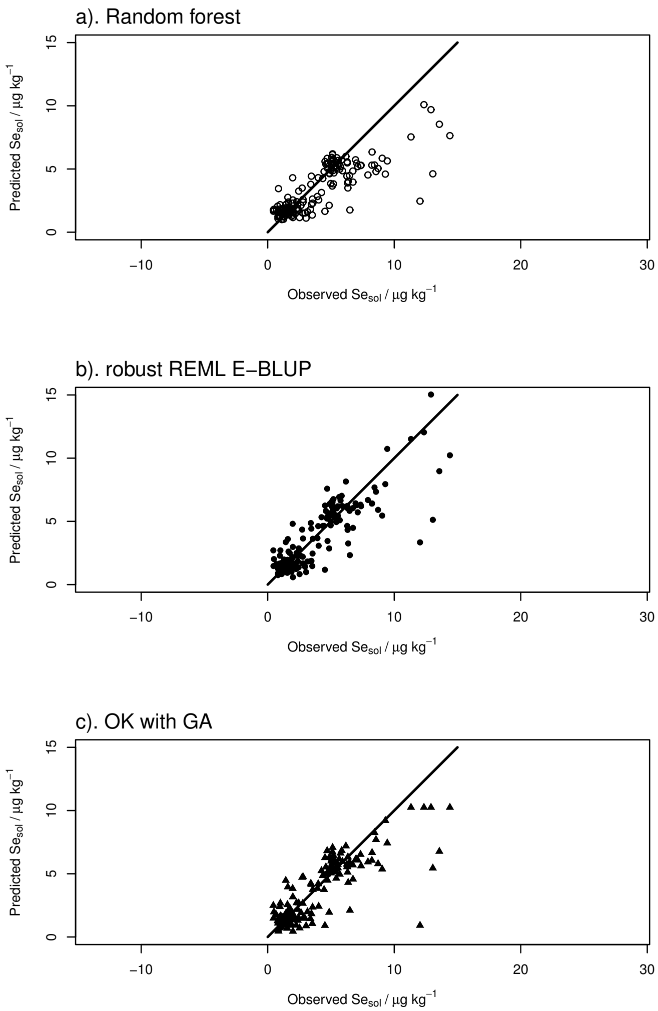


Figure S6. Scatter plot for the predictions of $\mathrm{Se_{sol}}$ (a) random forest, (b) robust REML E-BLUP, and (c) ordinary kriging with Gaussian anamorphosis transformed $\mathrm{Se_{sol}}$.

Chapter 5

Planning a geostatistical survey to map soil and crop properties: eliciting sampling densities

This chapter will be considered submitted for peer review in :

Chagumaira, C., Chimungu, J. G., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M. Planning a geostatistical survey to map soil and crop properties: eliciting sampling densities.

Planning a geostatistical survey to map soil and crop properties: eliciting sampling densities

Christopher Chagumaira[abcd]; Patson C. Nalivata[d], Joseph G. Chimungu[d], Martin R. Broadley[bc], Alice E. Milne[c] and R. Murray Lark[ab]

[a]Future Food Beacon of Excellence, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom; [b]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom; [c]Rothamsted Research, Harpenden, AL5 2JQ, United Kingdom; [d]Crop and Soil Sciences Department, Bunda College, Lilongwe University of Agriculture and Natural Resources, PO BOX 219, Lilongwe, Malawi

Abstract

When planning a survey of soil properties it is necessary to make decisions about the sampling density. Sampling density determines both the quality of predictions and the cost of field work. In this study, four approaches (offset correlation, prediction intervals, conditional probabilities and implicit loss functions), that can be used to assess the implications of uncertainty in spatial predictions using prior information on variability of the target properties, soil pH and selenium concentration in grain, were presented to a diverse group of stakeholders in order to determine an appropriate grid spacing. The background of the stakeholder, i.e. the professional group and frequency of use of statistics in job role, had no influence in the responses selected for each approach. Our results show that there were variations in the selection made by each method. Some were not well understood (conditional probability and implicit loss function). The one which stakeholders favoured, offset correlation, was not directly linked to decision making. Over 70% of the stakeholders specified a correlation of 0.7 or more as a criteria for adequate sampling intensity. The offset correlation will be more useful to stakeholders, with little or no statistical background, who unable to express their requirements of information quality based on other measures of uncertainty.

1 Introduction

When planning a survey of soil properties it is necessary to make decisions about the sampling density (e.g. Ball, 2019). Sampling density determines both the quality of predictions and the cost of field work. If we have a reasonable estimate of variance parameters (i.e. variogram for ordinary kriging) then we can compute kriging variances for different grid spacings and, in principle, select an acceptable one (McBratney et al., 1981). In cases where we do not know the variogram (as we have yet to sample), a variogram from comparable regions can be used instead (Alemu et al., 2022). Alternatively, an approximate variogram can be obtained from a reconnaissance study, allowing for uncertainty, e.g., by a Bayesian approach (Lark et al., 2017). An average variogram or some other generalised model can be extracted from published studies (Paterson et al., 2018), and the variogram can also be elicited from experts with relevant experience (Truong et al., 2013).

The kriging variance at some location depends only on the variogram and the spatial distribution of observations (Webster and Oliver, 2007; Webster and Lark, 2013). As the sampling density increases around a location, then the kriging variance diminishes. Because field and analytical costs increase in parallel with sampling density, the kriging variance, as a measure of the resulting uncertainty, could be used to find an appropriate sample density in which the data user is satisfied that the resulting information is sufficiently precise, but the costs of obtaining it are also considered. However, Chagumaira et al. (2021) found that for many users, the kriging variance is not an accessible measure of uncertainty.

Kriging variances are a direct measure of uncertainty resulting from the prediction process, and they are the variance of the prediction distribution. Because the kriging prediction is unbiased it is the mean of the prediction distribution (Webster and Oliver, 2007). On making the assumption of normality of the kriging errors, prediction intervals can be computed from kriging variances. Prediction intervals reflect the spatial variability of the variable and density of the samples (Webster and Oliver, 2007). They can be visualised on a map. Interpretation of prediction intervals depends on both the width of the interval and its location. However, prediction intervals were not preferred by stakeholders as a method of communicating uncertainty when making decisions, because stakeholders find it easier when uncertainty is tied to particular decision (Chagumaira et al., 2021). Perhaps tying prediction intervals to a particular decision may help with their interpretation.

Due to uncertainty, at a location where the true value indicates that an intervention is needed, the prediction might indicate that it does not. We could consider this as a general, decision related uncertainty measure: if we act at a location where a decision is needed, what is the probability that the prediction will be a false negative? We can compute the probability that the prediction indicates that an intervention is not needed at a site, conditional on the true value at that site's indicating that an intervention is needed. Intervention, in general, might be indicated by exceeding or falling below a threshold. These conditional probabilities are computed from the kriging variance obtained from a variogram. Then the conditional probabilities can then be used to make a decision about soil sampling, by selecting an appropriate grid spacing. Conditional probabilities have not been used in this way before, but there might be a way to tie uncertainty to a specific decision in a way which will help the stakeholder to understand its significance Chagumaira et al. (2021)

A further way to develop the decision-focussed approach to sample planning is to consider the costs of sampling and the costs resulting from uncertainty. This requires the data user's loss function. A loss function expresses the costs incurred resulting from using erroneous information to make a decision for an intervention (Goovaerts, 1997). The precision of the estimate and the loss function determines the expected loss when the estimate is used to make a decision, we can then compute the expected loss for decision for a grid of different samples, and considering the costs of the latter in comparison. It may not be possible to define a loss function prior to making decisions on soil sampling strategy because the cost of the errors are difficult to frame and quantify. However, an implicit loss function, conditional on a logistic model (i.e. a function of sampling effort and statistical information about the estimates of the cost of errors) can be modelled as the loss function that makes a particular decision on sampling effort rational (Lark and Knights, 2015). The logistic model can be obtained from data from a previous survey or a from a comparable region. Lark and Knights (2015) suggested that reflection on the implicit loss function for different sample schemes, or competing projects, might help decision-makers to arrive at loss functions which maybe regarded as plausible.

Decisions on soil sampling can be based on more general measures of uncertainty, that relate to sampling intensity, such as the offset correlation (Lark and Lapworth, 2013). The offset correlation is a measure of the robustness of the resulting map to arbitrary variation in the location of the origin of a fixed regular sampling grid. The offset correlation increases as the uncertainty in the map, attributable to sample density, decreases. It is not directly related to the decision process but dependent on the variogram and the proposed sampling spacing. The offset correlation might be a more intuitive uncertainty measure than prediction intervals and kriging variances. This is because people can more easily grasp and evaluate bounded measures such as the correlation (Hsee, 1998).

In this study we aimed to find out whether diverse groups of stakeholders are able to make decisions on soil and crop sampling strategies, in particular sampling density using soil pH and selenium concentration in grain ($\mathrm{Se}_{\mathrm{grain}}$), with the methods described above. We aimed to address the following questions: (i) *can stakeholders use the different approaches consistently?* (ii) *do the stakeholders have a preference?* and (iii) *does their use/preference depend on their background and experience?*

In the next section of this paper, we describe in detail the test approaches.

2 Theory

2.1 Prediction interval

Some unknown quantity at a location (e.g. soil pH or $\mathrm{Se}_{\mathrm{grain}}$) is characterised by a prediction distribution conditional on the data and statistical model. The kriging variance at the unsampled location, $\mathbf{x}_0$, is defined as

$$\sigma_{\mathrm{K}}^2 = \mathrm{E}[\{Z(\mathbf{x}_0) - \tilde{Z}(\mathbf{x}_0)\}^2], \tag{1}$$

where $\tilde{Z}(\mathbf{x}_0)$ is a prediction of the random variable $Z(\mathbf{x}_0)$. The kriging prediction is a weighted average of the data

$$\tilde{Z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda z(\mathbf{x}_0), \tag{2}$$

where $z(\mathbf{x}_0)$ is the data and λ are the kriging weights (Webster and Oliver, 2007). Then the kriging variance, σ_{K}^2 is given by:

$$\sigma_{\mathrm{K}}^2 = \mathrm{E}[\{Z(\mathbf{x}_0) - \tilde{Z}(\mathbf{x}_0)\}^2]. \tag{3}$$

Cross-validation predictions of the statistical model need to be examined by exploratory analysis of the kriging error, $\varepsilon(\mathbf{x}_0) = \{z(\mathbf{x}_0) - \tilde{Z}(\mathbf{x}_o)\}$ to check if the assumption of the normality holds. The kriging predictor is unbiased and the mean of the errors is zero, and their standard deviation is equal to the kriging standard deviation, σ_{K}, from kriging. Based on this, a 95% prediction intervals can be computed as:

$$\left[\tilde{Z}(\mathbf{x}_0) - 1.96\sigma_{\mathrm{K}}(\mathbf{x}_0), \tilde{Z}(\mathbf{x}_0) + 1.96\sigma_{\mathrm{K}}(\mathbf{x}_0)\right]. \tag{4}$$

The prediction distribution may also be obtained on a block support–for example if predictions are required at the scale of a farm mean or a mean for an administrative region. The same approach holds to the derivation of a prediction interval.

2.2 Conditional probability

We can calculate the joint probability that a location requires an intervention, and that the kriged estimate does not indicate this. If $\tilde{Z}(\mathbf{x}_0)$ is the prediction location of interest, and $z(\mathbf{x}_0)$ the value of the variable at $\mathbf{x}_0$ then $\varepsilon(\mathbf{x}_0)$ is the error of the kriging predictions. The covariance of $z(\mathbf{x}_0)$ and $\varepsilon(\mathbf{x}_0)$ is:

$$\mathrm{Cov}\left[\mathrm{z}(\mathbf{x}_0), \varepsilon(\mathbf{x}_0)\right] = \mathrm{Var}\left[\mathrm{Z}(\mathbf{x}_0)\right] - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}, \tag{5}$$

where $\boldsymbol{\lambda}$ denotes the vector of kriging weights for observations used to make the prediction, and $\mathbf{c}$ denotes the vector of covariances between each of these observations and $Z(\mathbf{x}_0)$. We can therefore, specify the joint distribution of $\{z(\mathbf{x}_0), \varepsilon(\mathbf{x}_0)\}$, assuming a normal random variable and prediction errors. From this it is possible to compute the conditional probability that $\tilde{Z}(\mathbf{x}_0) \geq z_t$ given that $z(\mathbf{x}_0) < z_t$, i.e. the probability, given that an intervention is required at $\mathbf{x}_0$ that, due to error in prediction, the mapped variable does not show this.

2.3 Implicit loss function

The loss function is a function of the error of $\tilde{Z}$, the kriging estimate of Z, as an estimate of the true unknown value, z, $\mathcal{L}(\tilde{Z} - z)$. The loss function is explained in greater detail by Journel (1984), Goovaerts (1997) and Lark and Knights (2015). According to Journel (1984) a general linear loss function is defined as:

$$\begin{aligned} \mathcal{L}(\tilde{Z} - z) &= \alpha_1 |\tilde{Z} - z| \text{ if } \tilde{Z} < z \\ &= \alpha_2 |\tilde{Z} - z| \text{ if } \tilde{Z} \geq z. \end{aligned} \tag{6}$$

The parameters α_1 and α_2 have positive real values. The coefficient α_2 is the loss per unit error of underestimation and α_1 is the loss per unit of error of overestimation. For example, If an intervention is required if z is less than some threshold then α_2 is the cost per unit value of z of an unnecessary intervention. The slopes, α_1 and α_2 define the asymmetry of the loss function. The loss function can be symmetrical, i.e. penalizing overestimation and overestimation equally; or can be asymmetrical because over-and-underestimation have different consequences. The asymmetry of the loss function is the ratio of the loss per unit value by which a quantity is underestimated to the loss per unit value of an overestimation (Lark and Knights, 2015). The asymmetry, a, is obtained by

$$a = \frac{\alpha_2}{\alpha_1}. \tag{7}$$

The loss is independent of the absolute value of z. If the loss function depends only on the estimation error, then z can be set to zero, without loss of generality and the expected loss can be computed as a function of the error variance, and so of the sample size (Lark and Knights, 2015). Increasing sample size reduces the minimum expected loss in so far as it reduced the error variance. Therefore, the cost of obtaining n samples can be measured at which the marginal cost of additional sample point is equal to the reduction in expected loss that single sample achieves (Goovaerts, 1997). However, it maybe difficult to define a loss function prior to making decisions about sampling. The losses may not be easy to quantify, e.g. social costs of failing to intervene, costs of unnecessary interventions, loss of confidence in the decision-making organisation. One sampling campaign does not necessarily map on to one use of the data. How can we consider the value of future use of the information? For example, there will be more costs for errors for underestimating $\mathrm{Se}_{\mathrm{grain}}$ concentration than underestimating soil pH. The implicit loss function aims to help stakeholders to reflect in possible loss functions for the problem in a decision-making setting. The implicit loss function is a loss function that makes a specified sample size, n, a rational choice, given the marginal costs. That is to say, it is the loss function implied by a choice of $\bar{n}$, assuming this is rational.

The implicit loss function is conditional on a logistic model, that expresses the marginal costs of sampling exercise and the conditional distribution of z as a function of effort (Lark and Knights, 2015). The implicit loss function is obtained by finding $\bar{\alpha}_1$ (given asymmetry), such that

$$\breve{\mathcal{L}}\left(\bar{n}-1|\bar{\alpha}_1,\bar{\alpha}_2,\boldsymbol{\phi}\right)-\breve{\mathcal{L}}\left(\bar{n}|\bar{\alpha}_1,\bar{\alpha}_2,\boldsymbol{\phi}\right)=\mathrm{C}(\bar{\mathrm{n}})-\mathrm{C}(\bar{\mathrm{n}}-1), \tag{8}$$

where $\bar{n}$ is the specified number of sample, C(n) is the function that returns the cost of n samples and $\boldsymbol{\phi}$ is a vector of variogram parameters, so kriging variance is a contributor. The asymmetry can be set at different values, or inferred from other elicited opinions of the stakeholder group (Lark and Knights, 2015, e.g.,). Lark and Knights (2015) suggested that a stakeholder group might consider an implicit loss function for different $\bar{n}$ as starting points in the elicitation of a sample size, or compare implicit loss function for different projects given different partitions of a total budget between them. No attempt has been made to elicit opinions from stakeholders on implicit loss function, so we tried it in the current project.

2.4 Offset correlation

The offset correlation is a measure of the correlation that is expected between the kriging predictions, $\tilde{Z}_1(\mathbf{x}_0)$, made from the first square grid, of interval ζ, and predictions, $\tilde{Z}_2(\mathbf{x}_0)$, made from the second grid, a translation of the first grid by $\zeta/2$ in both directions. The offset correlation in described in greater detail by Lark and Lapworth (2013). The correlation of the two kriging predictions can be computed by:

$$\rho_{\tilde{Z}_1,\tilde{Z}_2}=\frac{\mathbf{C}_{\tilde{\mathrm{Z}}_1,\tilde{\mathrm{Z}}_2}(\mathbf{x}_0)}{\sqrt{\sigma^2_{\mathrm{K}_{\tilde{Z}_1}}\sigma^2_{\mathrm{K}_{\tilde{Z}_2}}}}, \tag{9}$$

where $\mathbf{C}_{\tilde{\mathrm{Z}}_1,\tilde{\mathrm{Z}}_2}(\mathbf{x}_0)$ is the covariance $\tilde{Z}_1(\mathbf{x}_0)$ and $\tilde{Z}_2(\mathbf{x}_0)$. $\sigma^2_{\mathrm{K}_{\tilde{Z}_1}}$ and $\sigma^2_{\mathrm{K}_{\tilde{Z}_2}}$ are the kriging variances of the predictions from the first and second grid, respectively. The offset correlation depends on $\mathbf{x}_0$, and is smallest at the location furthest from points on either grid. This minimum offset correlation is used to evaluate predictions from a grid spacing ζ. Offset correlation is bounded on the interval [0,1], which makes it intuitively easy to interpret as an uncertainty measure. The offset correlation increases as the uncertainty in the map, attributable to sample density, decreases. The denser the grid the more consistent the maps and the offset correlation will be 1 if the maps are identical and 0 if they are entirely unrelated to each other.

3 Materials and methods

3.1 Basic approach

We used four methods to assess uncertainty in relation to sampling density, considering the problem of measuring a soil property relevant to crop management: soil pH, and a property of the crop: $\mathrm{Se}_{\mathrm{grain}}$ concentration. We used variograms from a national survey in Malawi for each variable (Gashu et al., 2021) to obtain sampling densities for further notional sampling for an administrative district in Malawi, Rumphi District, with an area of 4,769 km^2. The outputs were presented to participants. The participants considered each method in turn and were asked to select a sampling grid density based on the method. The key questions asked were: *(i) has the method helped you assess the implication of uncertainty in spatial prediction in as far as it is controlled by sampling?* and *(ii) which of these methods was easiest to interpret?*. Finally, the participants were asked to rank the method in terms of ease of use. Evaluation of the test methods were done using an online questionnaire on Microsoft Forms.

The elicitation was conducted online using Zoom Video Communications (2022) in two sessions, 26^{th} and 28^{th} April 2022. There were two sessions in order to accommodate participants from different time zones, and manage the participants in smaller groups to allow for questions and feedback. The invited participants self-identified as (i) agronomist or soil scientist or (ii) public health or nutrition specialists. The participants also self-assessed their statistical/mathematical background and their frequency of use of statistics in their job role (perpetual, regular, occasional use).

We invited professionals working in agriculture, nutrition and health at civic organisations, universities, government departments from Ethiopia, Malawi and wider GeoNutrition sites (United Kingdom, Zambia and Zimbabwe). In total we had 26 participants (18 were agronomist or soil scientist and 8 public health or nutrition specialists). Ethical approval to conduct this study was granted by the University of Nottingham, School of Biosciences Research Ethics Committees (SBREC202122022FEO) and participants gave informed consent to their participation and subsequent use of their responses.

In the exercise, an introductory talk was given to explain the study's objectives. During the talk, we explained the four test methods (offset correlation, prediction intervals, conditional probabilities and implicit loss function) and how they can be used to assess the implications of uncertainty in spatial predictions to determine appropriate sampling grid space for a geostatistical survey. We explained the structure of the questionnaire to the participants. We emphasized to the participants that we were not testing their mathematical/statistical skills and understanding but rather were testing the accessability of the methods.

Evaluation of the test methods was done through a questionnaire, as shown on Figure 1. Using the first four questions, Q1 to Q4, we wanted to find out if the method helped to identify a sampling grid spacing. On Q5, we wanted the participants to assess the test methods in terms of their effectiveness in finding an appropriate grid spacing. We asked the participants to rank these methods in an order of their effectiveness, in their experience, and in terms of finding a level of uncertainty that they were able to tolerate when deciding about a sampling grid spacing. We asked them to put rank 1 as the most effective method and rank 4 the least.

The offset correlation was the first method presented to the participants. This was followed by prediction intervals and conditional probabilities. The implicit loss function was the final method presented to the participants. We followed this order, we started with a measure we thought all our stakeholder would understand and move on to more complex methods.

Number	Question	Response
Q1	We show you here some pairs of example maps of soil pH/Se_{grain}, each pair being based on a different grid spacing, and so with a different offset correlation. We also show scatter plots which illustrate the strength of the correlation. What do you think is the smallest correlation that would be acceptable if one of the maps were to be used to make decisions?	(1) 0.4 (2) 0.5 (3) 0.6 (4) 0.7 (5) 0.8 (6) 0.9
Q2	You are shown different scenarios for the prediction of soil pH/Se_{grain} from different grid spacings, which determine the width of the prediction interval. What is the grid spacing that gives the widest prediction interval that would be acceptable if one of the maps were to be used to make decisions?	(1) Spacing=20km (2) Spacing=40km (3) Spacing=60km (4) Spacing=80km (5) Spacing=100km (6) Spacing=120km
Q3	At some location on the map the true value of soil pH/Se_{grain} indicates that an intervention is required, due to error in prediction there is a non-zero probability that the mapped soil pH/Se_{grain} does not show this, this probability increases with grid spacing as shown on the graph. What grid spacing do you think corresponds to the largest acceptable value of this probability?	(1) Spacing=20km (2) Spacing=40km (3) Spacing=60km (4) Spacing=80km (5) Spacing=100km (6) Spacing=120km
Q4	We have three specified implicit loss functions for predictions Se_{grain} concentration over an area of 4,769 square kilometres (km^2) for a district/administrative region. With the implicit loss function we assume that the sample density is fixed (e.g. on budgetary grounds) and compute the loss function which would make that a rational choice. We then ask does the loss function implied by the decision look sensible?	(1) Spacing=10km (2) Spacing=20km (3) Spacing=40km
Q5	Please rank these methods in an order of their effectiveness, in your experience, in terms of finding a level of uncertainty that you are able to tolerate when deciding about a sampling grid density.	Rank 1 being MOST effective and Rank 4 the least

Figure 1: The list of questions used to elicit stakeholder opinions about the set of methods that can help end-users to assess the implications of uncertainty in spatial prediction in as far as this is controlled by sampling.

3.2 *Test Methods*

3.2.1 Statistical modelling and spatial prediction of grain Se concentration and soil pH

We used the data from a geostatistical survey conducted in Malawi for the GeoNutrition project (Gashu et al., 2021). Field sampling was undertaken to support the spatial prediction of micronutrient concentration in crops and soil across Malawi. Detailed description of soil and crop sampling in Malawi are presented by Gashu et al. (2021) and Botoman et al. (2022), and the full data description is provided by Kumssa et al. (2022).

We undertook exploratory analysis of soil pH and Se_{grain} concentration using QQ plots, histograms and summary statistics to check whether there was need for transformation of the variables for the assumption of normality. The data for Se_{grain} concentration were skewed and it was necessary to transform them to natural logarithms. The variance parameters for both soil pH and Se_{grain} concentration were estimated by residual maximum likelihood using the likfit procedure (Diggle and Ribeiro, 2010) for the R platform (R Core Team, 2022) with a constant mean as the only fixed effect. These variance parameters were used in the subsequent test methods. The thresholds we considered, in this study for the prediction intervals and conditional probabilities were soil pH of 5 and Se_{grain} concentration of 38 µg kg^{-1}. The threshold for soil pH is 5 in Malawi, such that if the pH at a location falls below 5, it would be necessary to apply lime (Chilimba et al., 2013). The threshold Se_{grain} concentration is 38 µg kg^{-1}, such that a serving of 330g of grain flour provides a third of the daily estimated average requirement of Se_{grain} for an adult woman (Chagumaira et al., 2021). The intervention for soil pH was liming, and Se_{grain} was provision of fortified food.

3.2.2 Prediction intervals

Using the variance parameters estimated in Section 3.2.1, we evaluated kriging variances at the centres of cells of square grids of different spacings. We considered minimum and maximum grid spacings of 0.05 and 125km, respectively, with an increment of 0.5km. We then computed the cell-centred block kriging variance the spacings we were considering by block kriging (Webster and Oliver, 2007). We considered different prediction for each variable but the prediction interval was fixed, depending only on grid spacing. The three predictions of soil pH were 4.8, 5.5 and 6.0 and those of Se_{grain} were 20, 55 and 90 µg kg^{-1}. The participants were presented, in a chart, predictions of soil pH and Se_{grain} concentration in relation to a threshold, such that if a prediction falls below z_t an intervention is needed. The chart consisted of (a) box plot of the distribution of the measured variable, (b) a graph of the lower and upper prediction intervals for the prediction for grid spacings from 0 to 120km, and (c) lines indicating the z_t and prediction (see Figure 2a, S7 and S8). From the chart, we asked the participants the grid spacing that gives the widest prediction interval that would be acceptable if the mapped predictions were to be used to make decisions about soil management or interventions to address human Se deficiency.

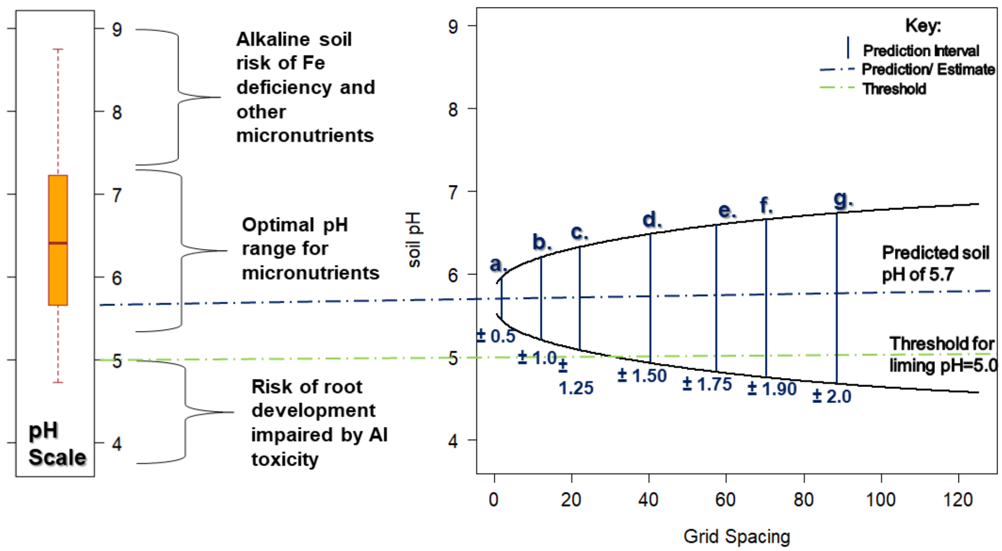


Figure 2a: An example of a chart, for prediction intervals, with prediction of soil pH of 5.7 with prediction intervals and in relation to a threshold of pH = 5.0.

3.2.3 Conditional probability

The conditional probability is a measure of uncertainty in terms of the risk of failing to intervene at some location given that an intervention is needed. We presented the participants with a chart of conditional probabilities plotted against grid spacing is shown on Figure 2b and Figure S9, and this probability increases with grid spacing. The conditional probability is bounded on an interval [0,1]. If the prediction of $\mathrm{Se}_{\mathrm{grain}}$ or pH was below the threshold, z_t, an intervention is needed. We then asked the participant at what grid spacing they thought corresponded to the largest acceptable value of this probability.

3.2.4 Implicit loss functions

In order to compute the implicit loss function, we needed a cost model for Rumphi district. The cost model for was computed from the rate of sampling during the geostatical survey conducted at national-scale in Malawi for the GeoNutrition project (Gashu et al., 2021). We used the function defined in Lark and Knights (2015) to return the costs of n samples over an area A km^2:

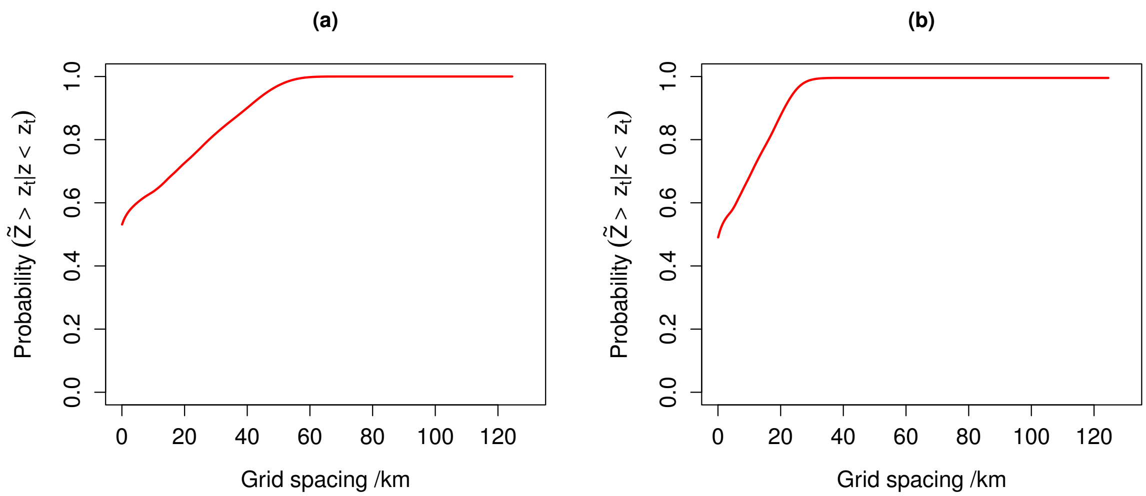


Figure 2b: An example of the chart of conditional probabilities plotted against grid spacing for (a) soil pH and (b) $\mathrm{Se}_{\mathrm{grain}}$ concentration.

$$\mathrm{C}(n) = \omega + \upsilon n + \beta A t_r, \tag{10}$$

where ω are the fixed costs, υ cost of laboratory analysis per unit, and β the field costs per work day per team. The variable t_r is time taken to sample per km^2 at a density of r per km^2. The variable t_r is a function of (i) the total time spent sampling per unit area, (ii) sample density, and (iii) square root of the sampling density. These variables were obtained from extracting from the geostatical survey conducted at national-scale in Malawi for the GeoNutrition project: (i) number of points sampled, (ii) mean time spent travelling per sample (excluding periods of lunch break), (iii) mean time spent at sample site, (iv) total areas sampled that day, and (v) length of sampling day. A detailed description of how the costs were computed is presented in the Supplement.

We fixed the asymmetry ratio as 1.5 following Lark and Knights (2015), implying a bigger loss for overestimation of the variables (i.e. failing to intervene of soil pH or $\mathrm{Se}_{\mathrm{grain}}$ are smaller than prediction). With the implicit loss function we assumed that the sample density is fixed (e.g. on budgetary grounds) and computed the loss function which would make that a rational choice. We presented three specified implicit loss functions for predictions of $\mathrm{Se}_{\mathrm{grain}}$ for Rumphi district, with an area of 4,769 km^2 with sampling densities fixed at 10, 20 and 40km. Figure 2c and Figure S10, shows the implicit loss function for $\mathrm{Se}_{\mathrm{grain}}$. We then asked the participants to identify the loss function implied by the sampling decision that looked more plausible to make decisions about interventions to address human Se deficiency.

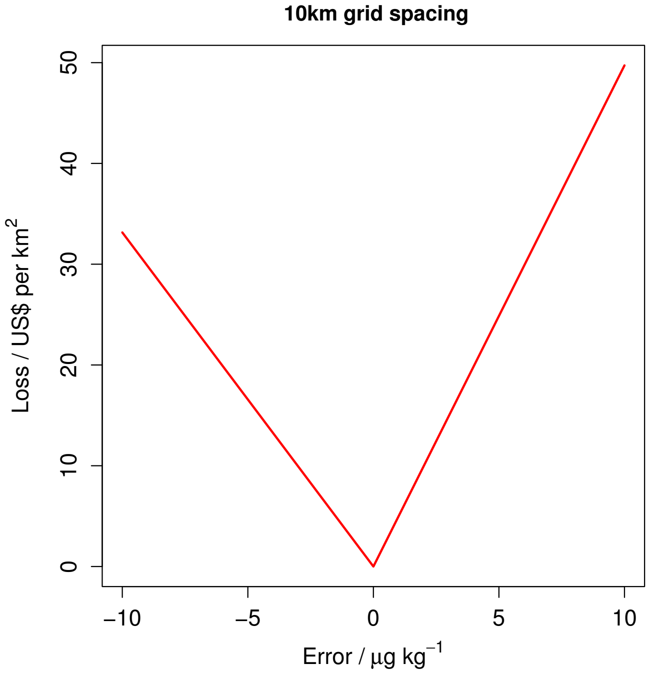


Figure 2c: An example of specified implicit loss functions for predictions of Se_{grain} concentration at a 10km grid spacing.

3.2.5 Offset Correlation

We presented the participants with pairs of example maps of soil pH and Se_{grain} concentration, each pair being based on a different grid spacing, and so with different offset correlation. We also showed scatter plots that illustrated the strength of the correlation. Figure 2d, shows an example of pairs of maps of Se_{grain} concentration and the corresponding scatterplot (see Figure S5 and S6). The correlation plots showed the kriging predictions for soil pH and Se_{grain} concentration predicted with parameters estimated in Section 3.2.1. We asked the participants the smallest offset correlation that would be acceptable if one of the maps were to be used to make decisions based on the soil or grain property.

3.3 Data Analysis

3.3.1 Test methods

The results for Q1 to Q4 were presented as contingency tables. The rows of each table correspond to the response (e.g. the different grid spacings) and, the full table, the columns correspond to the frequency of use of statistics, nested, within professional group and nested within variable used (soil pH or Se_{grain}). Contingency tables allowed us to test the null hypothesis of random association of responses with the different factors in the columns. The expected

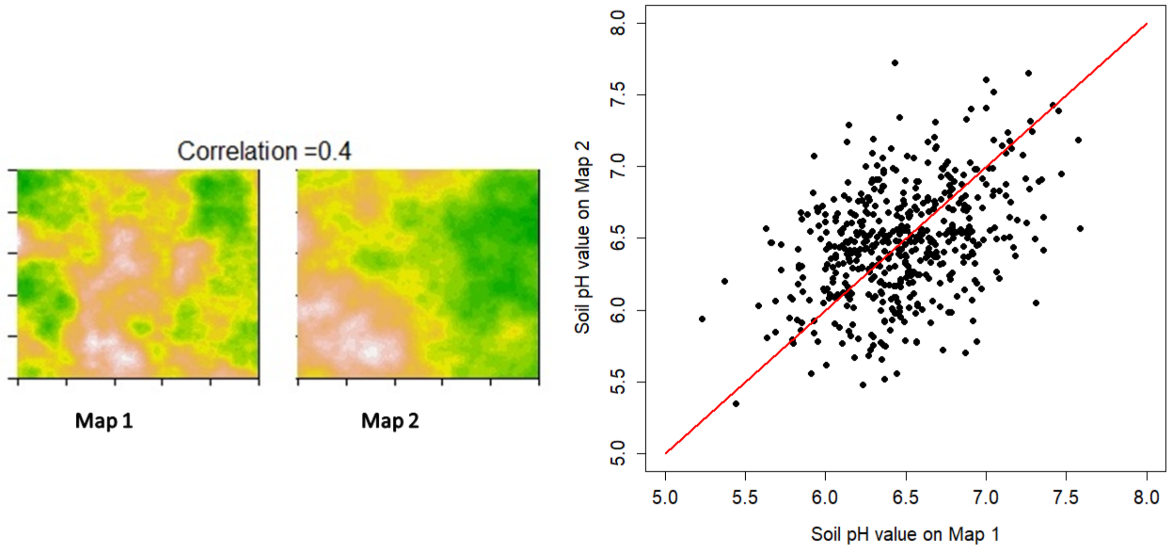


Figure 2d: The pairs of example maps of Se_{grain} concentration and corresponding scatterplot for offset correlation 0.4.

number of responses under the null hypothesis, $e_{i,j}$ in a cell $[i,j]$, is a product of row (n_i) and column (n_j) totals dived by the total number of responses (N), and this the null hypothesis of the contingency table which is equivalent to an additive log-linear model of the table. An alternative to the additive model for the contingency table, is the saturated model that has an extra $(n_r-1)(n_c-1)$ term that allows for interaction amongst the columns and tables of the table. The proportions of observed responses $o_{i,j}$ may differ from $e_{i,j}$ in a cell $[i,j]$ and the likelihood ratio statistic or deviance, L, can be used to provide evidence against the null the null hypothesis. The likelihood ratio statistic is computed by

$$L = 2\sum_{i=1}\sum_{j=1} o_{i,j}\log\frac{o_{i,j}}{e_{i,j}}. \tag{11}$$

where L has an approximate χ^2 distribution under the null hypothesis of random association between the rows and columns of the table, with $(n_r-1)(n_c-1)$ degrees of freedom (Christensen, 1996; Lawal, 2014). We fitted the log-linear models using the loglm from the MASS package (Venables and Ripley, 2002) for the R platform.

A contingency table can be partitioned to evaluate whether there are differences in the responses of the participants based on (i) variable used in the test method, (ii) professional group and (iii) by frequency of use of statistics. In Figure 3, we illustrate how the contingency table can be partitioned. The table can be partition into components corresponding to pooled table and subtables of the full table.

Full table

Deviance = L_f
Response = $O_{i,j}$
degrees of freedom = DF_F = (2−1)×(12−1) = 11

	Soil pH						Se_{grain} concentration					
	Agronomy or soil science (AGS)			Public health or nutrition (PHN)			Agronomy or soil science (AGS)			Public health or nutrition (PHN)		
Response	**Occasionally**	**Regularly**	**Perpetually**	**Occasionally**	**Regularly**	**Perpetually**	**Occasionally**	**Regularly**	**Perpetually**	**Occasionally**	**Regularly**	**Perpetually**
Spacing 1	$O_{1,1}$	$O_{1,2}$	$O_{1,3}$	$O_{1,4}$	$O_{1,5}$	$O_{1,6}$	$O_{1,7}$	$O_{1,8}$	$O_{1,9}$	$O_{1,10}$	$O_{1,11}$	$O_{1,12}$
Spacing 2	$O_{2,1}$	$O_{2,2}$	$O_{2,3}$	$O_{2,4}$	$O_{2,5}$	$O_{2,6}$	$O_{2,7}$	$O_{2,8}$	$O_{2,9}$	$O_{2,10}$	$O_{2,11}$	$O_{2,12}$

Pooled table 1

Pooled table (Professional groups pooled within variable):
Deviance = L_{p1}
degrees of freedom = DF_{P1} = (2−1)×(2−1) = 1

Response	**Soil pH**	**Se_{grain} concentration**
Spacing 1	$O_{1,1}+O_{1,2}+O_{1,3}+O_{1,7}+O_{1,8}+O_{1,9}$	$O_{1,4}+O_{1,5}+O_{1,6}+O_{1,10}+O_{1,11}+O_{1,12}$
Spacing 2	$O_{2,1}+O_{2,2}+O_{2,3}+O_{2,7}+O_{2,8}+O_{2,9}$	$O_{2,4}+O_{2,5}+O_{2,6}+O_{2,10}+O_{2,11}+O_{2,12}$

Subtable 1

Subtable 1 (Soil pH):
Deviance= L_{s1}
degrees of freedom = DF_{S1} = (2−1)×(6−1) = 5

	Soil pH					
	Agronomy or soil science (AGS)			Public health or nutrition (PHN)		
Response	**Occasionally**	**Regularly**	**Perpetually**	**Occasionally**	**Regularly**	**Perpetually**
Spacing 1	$O_{1,1}$	$O_{1,2}$	$O_{1,3}$	$O_{1,4}$	$O_{1,5}$	$O_{1,6}$
Spacing 2	$O_{2,1}$	$O_{2,2}$	$O_{2,3}$	$O_{2,4}$	$O_{2,5}$	$O_{2,6}$

Subtable 2

Subtable 2 (Se_{grain} concentration):
Deviance = L_{s1}
degrees of freedom= DF_{S2} = (2−1)×(6−1) = 5

	Se_{grain} concentration					
	Agronomy or soil science (AGS)			Public health or nutrition (PHN)		
Response	**Occasionally**	**Regularly**	**Perpetually**	**Occasionally**	**Regularly**	**Perpetually**
Spacing 1	$O_{1,7}$	$O_{1,8}$	$O_{1,9}$	$O_{1,10}$	$O_{1,11}$	$O_{1,12}$
Spacing 2	$O_{2,7}$	$O_{2,8}$	$O_{2,9}$	$O_{2,10}$	$O_{2,11}$	$O_{2,12}$

Deviance partition:
$L_f=L_{p1}+L_{s1}+L_{s2}$

Degrees of freedom partition:
$DF_f=DF_{p1}+DF_{s1}+DF_{s2}$

Figure 3: An illustration of how the log-likelihood ratio was used to partition full table into subtables and pooled tables.

The full table in Figure 3, was partition into subtables for soil pH (Subtable 1 in the figure), and Se_{grain} concentration (Subtable 2 in the figure). Then the pooled table completes the partition. The degrees of freedom and deviances for the three table sum to the degrees of freedom and deviance of the full table. Using the contingency table, we could conclude if there are differences in responses for the two variables. The full table in can further be partitioned, in a similar way, by the background of the respondents i.e., professional group and frequency of use of statistics.

In our study, we wanted to find out if the responses recorded by the stakeholders depended on the variable used (soil pH or Se_{grain} concentration), and background of the respondent. We expected the responses to differ. We thought the stakeholders would have different perceptions of the impacts of the uncertainty for soil pH and Se_{grain} concentration. There were more agronomist or soil scientists than public health or nutrition specialists in the meeting, and we expected the priorities of the groups to differ when making interventions for soil pH and Se_{grain} concentration. We also thought the frequency of use of statistics would influence the choice of method used to select an appropriate grid spacing.

We first tested for differences responses recorded for each test method, by the variable used (soil pH or Se_{grain} concentration) using contingency tables. The responses from stakeholders in different professional groups were pooled within the two variables, as illustrated by the Pooled table 1 on Figure 3. This gave us a six (responses) by two (variables) contingency table with 5 degrees of freedom for the questions corresponding to offset correlation, prediction intervals and conditional probabilities (Q1 to Q3). However, for the implicit loss function we did not consider this because we only had a loss function for Se_{grain} concentration.

Second, we considered if the differences in the responses depended on the professional group of the respondent. Finally, we considered whether the frequency of use of statistics in their job role had an impact on the responses recorded by the respondents.

For some questions, we noted differences in the responses when pooled within variable used (soil pH or Se_{grain} concentration) and there was no differences in responses in professional groups and frequency of use of statistics for all questions. We further analysed the pooled tables or separate subtables to examine if the responses where uniformly distributed. The null hypothesis is a random distribution, if this is rejected then the shape of the distribution can be further examined.

3.3.2 Assessment of the method

The responses for the Q5 were tabulated with the methods as the columns and ranks as the rows. The participants ranked their preferred method first. However, in our analysis we reversed the order by assigning a score of 4 for the most preferred method and 1 for the least. We computed the mean ranks, $\bar{r}_i$, for each method for all respondents. We then separated the respondents by their professional group and computed the mean ranks.

Table 1a: The full contingency table for Q1 for offset correlation, showing how many respondents selected the given responses for offset correlation. The table is according to variable used, professional group and frequency of use of statistics in job role. The figures in parentheses are the expected numbers, $e_{i,j}$ a product of row and column totals dived by the total number of responses. The acronyms represent the professional groups (AGS–agronomist or soil scientist; PHN– public health or nutrition specialists), and frequency of use of statistics in job role (Pep– perpetual use of statistics; Occ–Occasional use of statistics and Reg– regular use of statistics).

Response	soil pH						Se_{grain}					
	AGS			PHN			AGS			PHN		
	Pep	Occ	Reg	Pep	Occ	Reg	Pep	Occ	Reg	Pep	Occ	Reg
Offset=0.4	0(0.23)	0(0.31)	1(0.85)	0(0.08)	0(0.31)	0(0.23)	0(0.23)	2(0.31)	1(0.31)	0(0.08)	0(0.31)	0(0.23)
Offset=0.5	0(0.17)	1(0.23)	0(0.63)	0(0.06)	0(0.23)	1(0.17)	0(0.17)	0(0.23)	1(0.23)	0(0.06)	0(0.23)	0(0.17)
Offset=0.6	1(0.40)	0(0.54)	1(1.48)	0(0.13)	1(0.54)	0(0.40)	1(0.40)	0(0.54)	0(0.54)	0(0.13)	1(0.54)	2(0.40)
Offset=0.7	1(0.92)	2(1.23)	2(3.38)	0(0.31)	3(1.23)	2(0.92)	1(0.92)	1(1.23)	3(1.23)	0(0.31)	1(1.23)	0(0.92)
Offset=0.8	0(0.87)	0(1.15)	5(3.17)	1(0.29)	0(1.15)	0(0.87)	0(0.87)	1(1.15)	5(1.15)	1(0.29)	1(1.15)	1(0.87)
Offset=0.9	1(0.40)	1(0.54)	2(1.48)	0(0.13)	0(0.54)	0(0.40)	1(0.40)	0(0.54)	1(0.54)	0(0.13)	1(0.54)	0(0.40)

Finally we separated the respondents by their frequency of use of statistics in their job role. Under a null hypothesis of random ranking for set of k ranks, the expected mean rank is $(k+1)/2$. The evidence against this hypothesis is measured a statistic distributed as $\chi^2(k-1)$:

$$\frac{12n}{k(k+1)} \sum_{i=1}^{k} \left\{ \bar{r}_i - \frac{k+1}{2} \right\}, \tag{12}$$

where n is the total number of rankings (Marden, 1996).

4 Results

4.1 Test methods

4.1.1 Method 1: Offset correlation

The full contingency table for Q1, for offset correlation, is presented as Table 1a. The table shows how many individuals selected the given responses for offset correlation. This table is according to variable used (soil pH vs. Se_{grain}), professional group and frequency of use of statistics. Table 1b shows how many individuals selected a given response to Q1, for offset correlation, when columns are pooled within variable used, soil pH or Se_{grain} concentration. Table 1c shows the pooled counts of the responses for Q1. There were no differences in the responses of the when the columns were pooled by the variable used, soil pH vs. Se_{grain} concentration, $p = 0.656$ (Table 2).

There were no differences in the responses when the columns were also pooled within professional groups ($p = 0.491$) and frequency of use of statistics ($p = 0.595$). Further analysis of the question on offset correlation was based on pooled counts, see Table 1c. There was strong evidence to reject the null hypothesis that the responses are uniformly distributed ($p = 0.003$). Figure 4 shows the responses of how all the participants responded to Q1, for offset correlation.

Table 1b: A subtable showing how many individuals selected a given response to Q1, for offset correlation, when columns are pooled within variable used (soil pH vs. Se_{grain} concentration).

Response	soil pH	Se_{grain}
Offset=0.4	1	3
Offset=0.5	2	1
Offset=0.6	3	4
Offset=0.7	10	6
Offset=0.8	6	9
Offset=0.8	4	3

Table 1c: Pooled responses given to the question on offset correlation.

Response	Pooled counts
Offset=0.4	4
Offset=0.5	3
Offset=0.6	7
Offset=0.7	16
Offset=0.8	15
Offset=0.8	7

Table 2: Analysis of the question on offset correlation, Q1, according to variable used, professional group and frequency of use of statistics.

	Deviance (L^2)	Degrees of freedom	P
Full contingency table analysis			
Full table	54.57	55	0.491
Pooled by variable used (pH v. Se_{grain})	3.29	5	0.656
Pooled by professional group	6.50	5	0.260
Pooled by frequency of use of statistics	8.35	10	0.595
Subtable–pooled counts: variable used			
Soil pH	27.01	25	0.352
Se_{grain}	24.2	25	0.507
Subtable–pooled counts: professional group			
Agronomist or soil scientist	26.25	25	0.394
Public health or nutrition specialist	21.81	25	0.646
Subtable–pooled counts: frequency of use of statistics			
Perpetual use of statistics	8.99	15	0.878
Occasional use of statistics	18.17	15	0.254
Regular use of statistics	19.06	15	0.211
Subtable–pooled counts			
Responses are uniformly distributed	17.69	5	0.003

Most of the respondents selected offset correlation of 0.7 as the smallest offset correlation that would be acceptable if one of the maps were to be used to make decisions based on the soil or grain property. We extracted the grid spacings, for soil pH and Se_{grain}, corresponding to the selected offset correlation of 0.7. The spacings were extracted from a plot of offset correlation against grid spacing obtained from the variance parameters of each variable (see Figure S4). The grid spacing for soil pH is 25km and for Se_{grain} is 12.5 km. The grid spacing corresponding to the offset correlation for each variable were computed from the variogram of the variable.

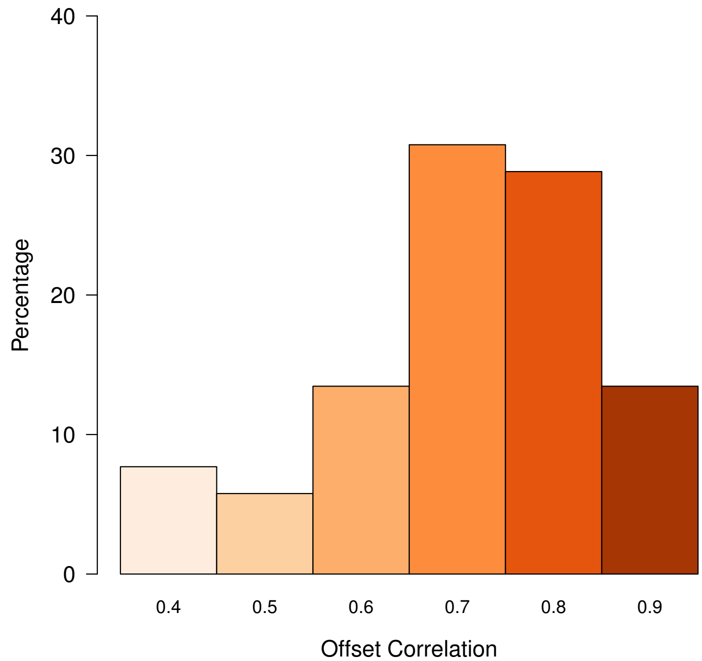


Figure 4: Bar charts showing how the participants responded to the Q1 for offset correlation.

Table 3: Analysis of the question on prediction interval, Q2, according to variable used, professional group and frequency of use of statistics.

	Deviance (L^2)	Degrees of freedom	P
Full contingency table analysis			
Full table	56.0	55	0.437
Pooled by variable used (pH v. Se_{grain})	0.972	5	0.965
Pooled by professional group	4.36	5	0.498
Pooled by frequency of use of statistics	14.5	10	0.152
Subtable–pooled counts: variable used			
Soil pH	23.8	25	0.531
Se_{grain}	31.2	25	0.181
Subtable–pooled counts: professional group			
Agronomist or soil scientist	26.5	25	0.381
Public health or nutrition specialist	25.1	25	0.455
Subtable- pooled counts: frequency of use of statistics			
Perpetual use of statistics	9.68	15	0.840
Occasional use of statistics	16.88	15	0.330
Regular use of statistics	15.08	15	0.450
Subtable- pooled counts			
Responses are uniformly distributed	7.77	5	0.169

4.1.2 Prediction interval

There were no differences in the responses when pooled within the variable used, $p = 0.656$, for prediction intervals (Table 3). We then pooled the responses within the professional groups, and there was no evidence to reject the null hypothesis ($p = 0.498$). Also, there were differences when responses were pooled within frequency of use of statistics, $p = 0.152$. Therefore, further analysis of the question on prediction intervals was based on pooled counts of responses. There was no evidence to reject the null hypothesis that the responses are uniformly distributed ($p = 0.169$). Figure 5 shows the bar charts of how all the participants responded to the Q2 for prediction intervals. For this method, there no clear choice of grid spacing for sampling for soil pH and Se_{grain}.

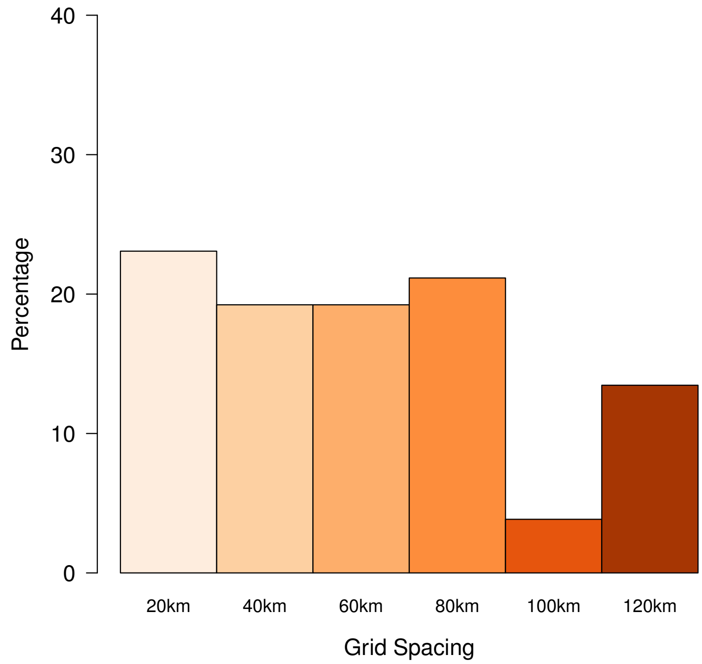


Figure 5: Bar charts showing how the participants responded to the Q2 for prediction intervals.

4.1.3 Conditional probabilities

Table 4 shows the results for partitioning the contingency table for the question on conditional probabilities, Q3. There was strong evidence to reject the null hypothesis when the columns were pooled by variable used, $p \leq 0.001$. Therefore, further analysis was based on separate subtables for soil pH and Se_{grain} concentration. For both variables, there were no differences in the responses when the columns were pooled within professional groups and frequency of use of statistics. For soil pH there was strong evidence to reject the null hypothesis that the responses are uniformly distributed ($p \leq 0.001$). A similar result was found for Se_{grain} concentration ($p \leq 0.001$). The bar charts for the responses for the question on conditional probabilities for soil pH are presented in Figure 6a. The grid spacing chosen by the participants for soil pH is 60km. The responses for Se_{grain} concentration are presented in Figure 6b. The grid spacing selected by the respondents was 40km.

Table 4: Analysis of the question on conditional probabilities, Q3, according to variable used, professional group and frequency of use of statistics.

	Deviance (L^2)	Degrees of freedom	P
Full contingency table analysis			
Full table	60.6	55	0.281
Pooled by variable used (pH v. Se_{grain})	26.7	5	< 0.001
Pooled by professional group	5.32	5	0.378
Pooled by frequency of use of statistics	14.5	10	0.152
Subtable–pooled counts: variable used			
Soil pH	12.1	25	0.986
Se_{grain}	21.8	25	0.647
Soil pH subtable–pooled counts: professional group			
Pooled within professional group	4.48	5	0.483
Agronomist or soil scientist	3.10	10	0.979
Public health or nutrition specialist	4.50	10	0.922
Soil pH subtable–pooled counts: frequency of use of statistics			
Pooled within frequency of use of statistics	0.889	10	1.00
Perpetual use of statistics	4.50	5	0.480
Occasional use of statistics	4.36	5	0.499
Regular use of statistics	2.33	5	0.802
Soil pH subtable–pooled counts			
Responses are uniformly distributed	50.15	5	< 0.001
Se_{grain} subtable–pooled counts: professional group			
Pooled within professional group	4.77	5	0.445
Agronomist or soil scientist	11.0	10	0.361
Public health or nutrition specialist	6.09	10	0.808
Se_{grain} subtable–pooled counts: frequency of use of statistics			
Pooled within frequency of use of statistics	9.55	10	0.481
Perpetual use of statistics	1.73	5	0.886
Occasional use of statistics	5.55	5	0.353
Regular use of statistics	4.99	5	0.417
Se_{grain} subtable–pooled counts			
Responses are uniformly distributed	36.77	5	< 0.001

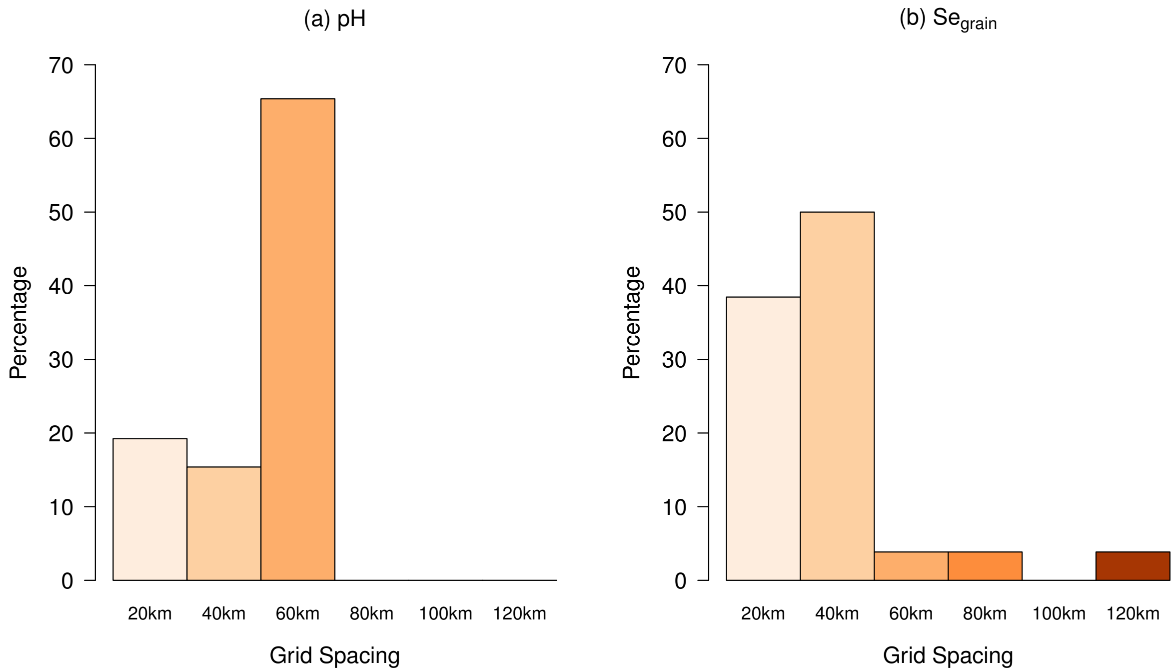


Figure 6: Bar charts showing how all the participants responded to the Q3 for conditional probabilities for (a) soil pH and (b) Se_{grain} concentration.

Table 5: Analysis of the question on implicit loss function, Q4, according to variable used, professional group and frequency of use of statistics.

	Deviance (L^2)	Degrees of freedom	P
Full contingency table analysis			
Full table	8.91	10	0.541
Pooled by professional group	0.49	2	0.781
Pooled by frequency of use of statistics	1.49	4	0.828
Subtable–pooled counts: professional group			
Agronomist or soil scientist	2.33	4	0.676
Public health or nutrition specialist	6.09	4	0.193
Subtable- pooled counts: frequency of use of statistics			
Perpetual use of statistics	1.73	2	0.422
Occasional use of statistics	1.73	2	0.422
Regular use of statistics	3.96	2	0.138
Subtable- pooled counts			
Responses are uniformly distributed	54.00	2	< 0.001

4.1.4 Implicit loss functions

The results for partitioning the contingency table for implicit loss function, Q4, are presented in Table 5. There were no differences in the responses when the columns of the table were pooled within professional groups ($p = 0.781$) and frequency of use of statistics ($p = 0.828$). Further analysis of the question on implicit loss function was based on pooled counts of responses. There was strong evidence to reject the null hypothesis that the responses are uniformly distributed ($p \leq 0.001$). The bar charts for the responses pooled counts for all respondents are shown on Figure 7. The grid spacing chosen by the participants for Se_{grain} concentration is 20km.

4.2 Assessment of the test methods

The question on ranking of the method was analysed in three ways. Firstly, we computed the mean ranks for all participants and tested for the evidence against the null hypothesis of random ranking. There is strong evidence to reject the null hypothesis of random ranking, $p \leq 0.001$ (Table 6). Second, the mean ranks for each professional groups were computed and there was strong evidence to reject the null hypothesis of random ranking ($p \leq 0.001$). Thirdly, we separated the participants according to their frequency of use of statistics in the job role, and computed the mean ranks. There was strong evidence to reject the null hypothesis of random ranking ($p \leq 0.001$).

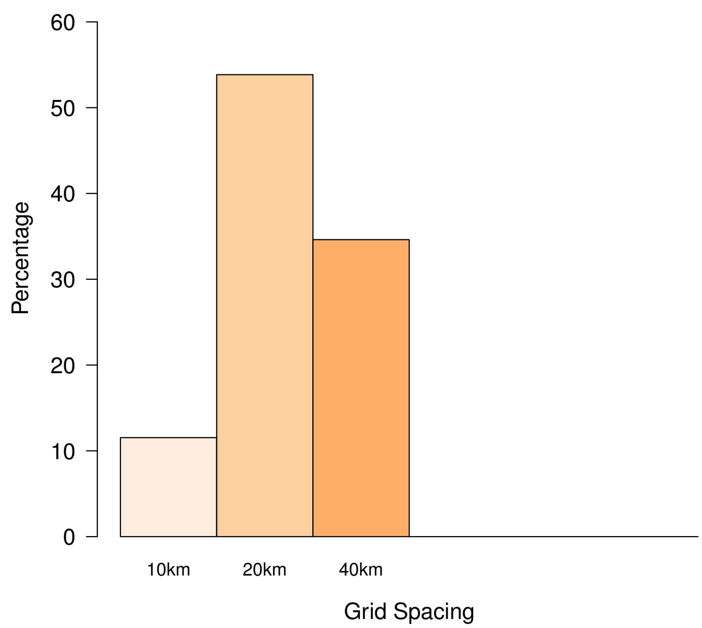


Figure 7: Bar charts showing how all the participants responded to the Q4 for implicit loss function.

Table 6: Analysis of Q6 according to professional group and level of use of statistics in job role

	Test Statistic (X^2)	Degrees of Freedom	P^*
All respondents	61.1	3	< 0.001
Professional group			
Agronomist or soil scientist	49	3	< 0.001
Public health or nutrition specialist	15.6	3	< 0.001
Frequency of use of statistics			
Perpetual user of statistics	34	3	< 0.001
Occasional user of statistics	28.5	3	< 0.001
Regular user of statistics	49.8	3	< 0.001

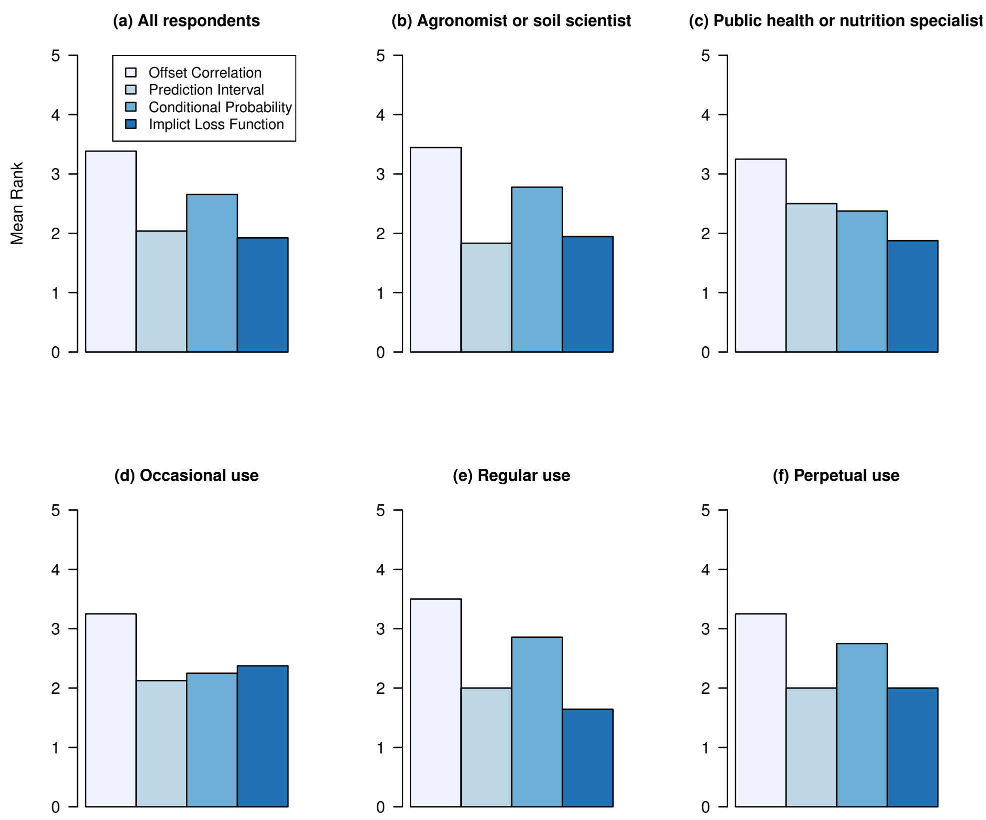


Figure 8: Ranking of test methods in terms on the most effective: (a) by all respondents, professional group: (b) agronomists or soil scientist and (c) public health or nutritionist specialists, and frequency of use of statistics: (d) occasional use, (e) regular use and (f) perpetual use.

The offset correlation was ranked as the most effective by all respondents (Figure 8a) and implicit loss function as the least effective. Both professional groups (i.e. agronomist or soil scientist and public health or nutritionist) ranked offset correlation first but differed in the second and least ranked methods (Figure 8b to 8c). Public health or nutrition specialists ranked second prediction intervals and implicit loss function as the least effective. The agronomist or soil scientist group ranked prediction intervals as the least effective and conditional probabilities as second.

When respondents were separated by their frequency of use of statistics, offset correlation was also ranked first (Figure 8d to 8f). Those who use statistics occasionally, in their job role, ranked the implicit loss function as the second best and the prediction intervals the least. Conditional probabilities were ranked second and implicit loss function as the least effective by those who regularly use statistics in the job role. Those who use statistic at all times, ranked conditional probabilities second. Prediction intervals and implicit loss functions were ranked last.

5 Discussion

In this study, we presented to diverse groups of stakeholders, four methods (offset correlation, prediction intervals, conditional probabilities and implicit loss functions) to support decisions on sampling grid spacing for a geostatistical survey using soil pH and Se_{grain}.

We wanted to find out if the stakeholders had a preference among the approaches presented to them. Offset correlation was ranked first as the method the stakeholders found easy to interpret (see Figure 8), and over 70% of the stakeholders specified a correlation of 0.7 or more as a criteria for adequate sampling intensity. During the feedback session, stakeholders highlighted that they were more familiar with the concept of correlation, with a closed interval of [0,1]. This explains why there more consistent responses under this method. Our results are consistent with findings of Hsee (1998), that relative measures of some quantity (Hsee gives an example of the size of a food serving relative to its container) are more readily evaluated than absolute measures (the size of serving). An easy-to-evaluate attribute, such as the bounded correlation of [0,1], has a greater impact on a person's judgement of utility. Hsee (1998) describe this as the "relation-to-reference" attribute. It is therefore, not surprising that the offset correlation is highly-ranked.

The offset correlation will be more useful for stakeholders who are not able to express their quality requirement for information in terms of quantities such as kriging variance. Furthermore, it is an intuitively meaningful measure of uncertainty, it recognises that spatial variation means that maps interpolated from offset grids will differ but that the more robust the sampling strategy the more consistent they will be. There is a paradox here, however, in that the previous study Chagumaira et al. (2021) showed that interpretation of survey outputs in terms of uncertainty was easiest for stakeholders with measures related directly to a decision made with the information. The offset correlation is a general measure, and the absolute magnitude of uncertainty has greater bearing on a specific decision. Indeed, Lark and Lapworth (2013) proposed the offset correlation particulary with general baseline surveys in mind. There is more research needed to develop sound but accessible ways to engage stakeholders with uncertainty consistent across planning and interpretation.

All the stakeholders ranked conditional probabilities second. Under this method, the stakeholders selected spacings where conditional probabilities was 1.0 or very close, i.e. the prediction equivalent to the overall mean. This suggest that the stakeholders may not have fully understood the method. This finding is consistent with the general view that users of information commonly find probabilities difficult to interpret (Spiegelhalter et al., 2011). Because probabilities are bounded [0,1], the 'relation-to-reference" attribute effect by Hsee (1998) may explain the previous preference for conditional probabilities (Jenkins et al., 2019; Chagumaira et al., 2021), but stakeholders still struggle to interpret them correctly. Perhaps if the problem had been framed in a different way, the stakeholders may have understood this method much better. More work is needed to investigate if framing the conditional probabilities in a different way would improve the judgement of utility of the stakeholders. More examples and more illustration may be needed in order to 'prime' the participants before the exercise.

Prediction intervals were ranked third by all the respondents, but there was no evidence against the null hypothesis of random selection among the available spacings. During a feedback session, the stakeholders cited difficulties of assessing the significance of a given prediction interval given that it can be associated with different prediction values. For very large or small prediction values the uncertainty is immaterial, it is near decision threshold that it becomes important. Similarity, prediction intervals were not highly ranked by stakeholders for communicating uncertainty in maps (Chagumaira et al., 2021). Similar reasons were given the respondents. We expected that prediction intervals to be of greatest value for specific interpretation of particular sites, but would be of limited value for survey planning.

The implicit loss functions was the lowest-ranked method. The group also commented that they had difficulties understanding this method, and most people opted for the central value. Loss functions are not readily accessible. It is difficult to define a loss function because it requires the cost of the errors, and we tried to show stakeholders some consistent approach with some plausible design. The fact that they did not understand the loss functions, shows there is need for more specific examples to help stakeholders think about loss function and their implications. It might help the stakeholders to provide some quantitative information about the costs of the survey, cost associated with intervention campaigns and costs of the impacts on MNDs on a country's gross domestic productivity. A reflection of these would allow the stakeholder to use these implicit assumptions when they were making decisions for selecting a fixed grid spacing for working with (Lark and Knights, 2015). Therefore, more work is need to refine this approach.

The background of the stakeholders, i.e., professional group and frequency of use of statistics, had no influence on their responses for all the methods. However, the background of the stakeholders had an influence on their ranking of the methods in terms of their effectiveness. The offset correlation was ranked as the most effective by all professional groups and by all respondents separated by frequency of use of statistics. For the professional groups there differences in the order of ranking. Prediction intervals were ranked least effective by those respondents who self identified as agronomist or soil scientist, but were ranked second by those in public health or nutrition.

At the begin of the online workshop, we explained each method with the aid of illustrations. After an explanation each method, there was a feedback session to allow the participants opportunities to seek clarity on ambiguous and unfamiliar concepts from the presenters. The participants' questions were answered and explained in different ways by CC, RML and AEM, with the use of illustrations. However, there are limitations with online workshops. Most participants would have the cameras switched off, and the "unconscious" feedback to presenters by observing the reactions of participants could not be noticed as during in-person workshops. The "unconscious" feedback would prompt the presenter to use a different approach to explain unfamiliar concepts and ambiguous terms. Due to internet connectivity, online workshops are timed and there will less time for feedback sessions. In such instances, respondents may seek clarity from the colleagues who have the same interests, resulting in bias (Ball, 2019).

All the methods may give different results for different variable, because they depend on the variogram of the variable in question. There maybe different grid spacings selected for the different variables. A potential problem may exist, if the variables were to be sampled in one survey and what spacing should be used? This is an important question that needs to be addressed when planning for soil and crop sampling. It may be reasonable to opt for the grid spacing for the variable that maybe the hardest to characterise. Another option would to consider some minimum quantile over all variables through a group elicitation.

6 Conclusions

A diverse group of stakeholders was able to make decision on soil and crop sampling strategies based on the four approaches (offset correlation, prediction intervals, conditional probabilities and implicit loss functions) presented to them. The background (professional group and frequency of use of statistics) of the stakeholder had no influence in the responses selected for each approach. There were variations in the selection made by each method. Some were not well understood (conditional probabilities and implicit loss functions). The one which stakeholders favoured, offset correlation, is not directly linked to decision making. The offset correlation will likely be more useful to stakeholders, with little or no statistical background, who unable to express their requirements of information quality based on other measures of uncertainty.

References

Alemu, R., Galew, A., Gashu, D., Tafere, K., Mossa, A. W., Bailey, E., Masters, W. A., Broadley, M. R., and Lark, R. M. (2022). Sub-sampling a large physical soil archive for additional analyses to support spatial mapping; a pre-registered experiment in the Southern Nations, Nationalities, and Peoples Region (SNNPR) of Ethiopia. *Geoderma*.

Ball, H. L. (2019). Conducting online surveys. *Journal of Human Lactation*, 35(3):413–417.

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., Mcgrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Broadley, M. R., Lark, R. M., and Nalivata, P. C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*.

Chagumaira, C., Chimungu, J. G., Gashu, D., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M. (2021). Communicating uncertainties in spatial predictions of grain micronutrient concentration. *Geoscience Communication*, 4(2):245–265.

Chilimba, A. D. C., Kabambe, V. H., Chigowo, M. T., Nyirenda, M., Botoman, L., and Tembo, Y. (2013). Agricultural lime application for improved soil and crop production in Malawi. Technical report, Soil Health Consortium of Malawi (SOHCOM), Malawi.

Christensen, R. (1996). *Log-Linear Models and Logistic Regression*. Springer, Springer-Verlag, New York.

Diggle, P. and Ribeiro, P. J. (2010). *Model-based geostatistics*. Springer Science+Business Media LLC.

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., Towett, E. K., Walsh, M. G., Wilson, L., Young, S. D., and Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, 594(7861):71–76.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.

Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, 11(2):107–121.

Jenkins, S. C., Harris, A. J. L., and Lark, R. M. (2019). When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. *Journal of Risk Research*, 22(5):537–554.

Journel, A. (1984). Mad and conditional quantile estimators. In *Geostatistics for natural resources characterization*, pages 261–270. Springer.

Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., Mcgrath, S. P., Milne, A., Muleya, P., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Haji, I. R., Broadley, M. R., Gashu, D., and Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from geonutrition surveys in ethiopia and malawi. *Scientific Data*, 9(1):1–12.

Lark, R. M., Hamilton, E. M., Kaninga, B., Maseka, K. K., Mutondo, M., Sakala, G. M., and Watts, M. J. (2017). Planning spatial sampling of the soil from an uncertain reconnaissance variogram. *Soil*, 3(4):235–244.

Lark, R. M. and Knights, K. V. (2015). The implicit loss function for errors in soil information. *Geoderma*, 251-252:24–32.

Lark, R. M. and Lapworth, D. J. (2013). The offset correlation, a novel quality measure for planning geochemical surveys of the soil by kriging. *Geoderma*, 197-198:27–35.

Lawal, B. (2014). *Applied statistical methods in agriculture, health and life sciences*. Springer.

Marden, J. I. (1996). *Analyzing and modeling rank data*. CRC Press.

McBratney, A. B., Webster, R., and M, B. T. (1981). The design of optimal sampling schemes for local estimation and mapping of of regionalized variables-i. theory and method. *Computers and Geosciences*, 7(4):331–334.

Paterson, S., McBratney, A., Minasny, B., and Pringle, M. J. (2018). Variograms of soil properties for agricultural and environmental applications. In *Pedometrics*, pages 623–667. Springer.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Spiegelhalter, D., Pearson, M., and Short, I. (2011). Visualizing uncertainty about the future. *Science*, 333:1393–1400.

Truong, P. N., Heuvelink, G. B., and Gosling, J. P. (2013). Web-based tool for expert elicitation of the variogram. *Computers & geosciences*, 51:390–399.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer-Verlag, New York.

Webster, R. and Lark, M. R. (2013). *Field Sampling for Environmental Science and Management.* Taylor and Francis.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Natural Environmental Scientists*. John Wiley & Sons Chichester, 2nd edition.

Zoom Video Communications (2022). *Zoom Video Communications*. Zoom Video Communications, San Jose, California.

Supplement for: Chagumaira et al in prep Planning a geo-statistical survey to map soil and crop properties: eliciting sampling densities

Chagumaira, C., Chimungu, J. G., Nalivata, P. C., Broadley, M. R., Milne, A. E., and Lark, R. M. Planning a geostatistical survey to map soil and crop properties: eliciting sampling densities.

Supplement

1 Theory

1.1 Conditional Probabilities

Calculating the joint probability that a location requires an intervention, and that the kriged estimate does not indicate this

At some location the true value of a property, z, might or might not indicate that an intervention is required. For purposes of this argument we assume that an intervention is required if $z \leq z_t$, a threshold value. We wish to compute the joint probability that a random location (a) requires the intervention (i.e. $z \leq z_t$), and (b) that the prediction, $\tilde{Z}$ indicates otherwise, (i.e. $\tilde{Z} > z_t$). If the kriging error, $z - \tilde{Z}$, were independent of z, then we might consider, assuming normal kriging errors and a known kriging variance, the probability that $\tilde{Z} > z_t$, given a value $Z = z$, $P\left(\tilde{Z} > z_t | z = Z\right)$, and then compute its expected value over the distribution of Z:

$$\int_{-\infty}^{-\infty} P\left(\tilde{Z} > z_t | z = Z\right) f(Z) \mathrm{d}\ Z, \tag{1}$$

where $f(Z)$ denotes the PDF of Z. However, this independence does not hold. The kriging predictor, like any smoothing estimator, is conditionally biased in the sense that the error:

$$\varepsilon_z \;=\; z - \tilde{Z}, \tag{2}$$

is likely to be positive for large z and negative for small z.

We can write the covariance of $z(\mathbf{x}_0)$ and $\varepsilon_z(\mathbf{x}_0)$ at some location $\mathbf{x}_0$ as

$$\mathrm{Cov}\left[z(\mathbf{x}_o), \varepsilon_z(\mathbf{x}_0)\right] \;=\; \mathrm{Var}\left[Z(\mathbf{x}_0)\right] - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}, \tag{3}$$

where $\boldsymbol{\lambda}$ denotes the vector of n_n kriging weights for observations used to make the prediction, and $\mathbf{c}$ denotes the vector of covariances between each of these observations and $Z(\mathbf{x}_0)$. From Eq (2)

$$\tilde{Z} \;=\; z - \varepsilon_z \therefore \tilde{Z} > z_t \;\Leftrightarrow\; z - \varepsilon_z > z_t \;\Leftrightarrow\; \varepsilon_z < z - z_t$$

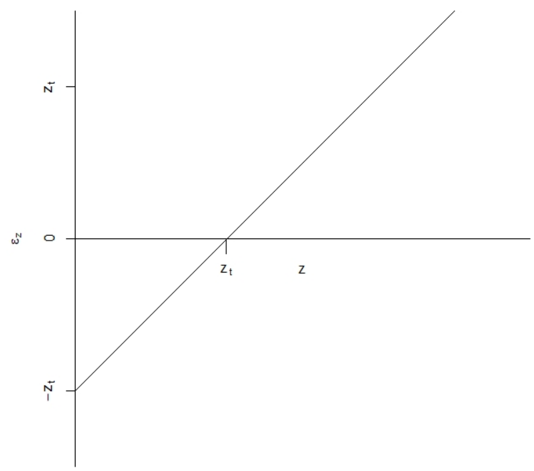


Figure S1: Plot of error (positive or negative) against the true value of z.

Figure S1 shows a plot of error (positive or negative) against the true value of z. The line is the function

$$\varepsilon_z = z - z_t$$

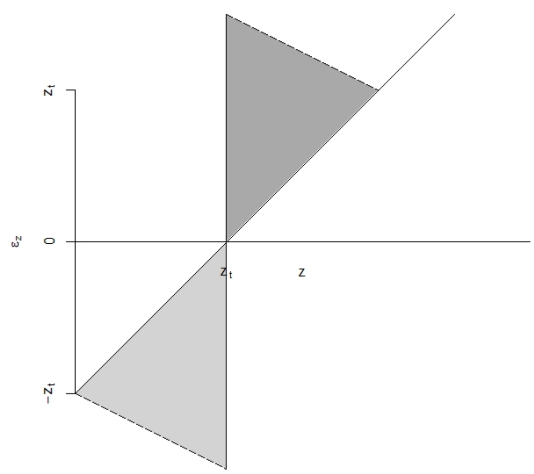


Figure S2: Plot of error against the true value of z.

In Figure S2 the light-grey shaded region, unbounded where the line is

Table S1: Parameters of the joint distribution of Z and ε_z

Mean of Z	Population mean of the variable
Variance of Z	*A priori* variance of the variable, i.e. $c_0 + c_1$.
Mean of ε_z	0, as kriging is unbiased
Variance of ε_z	Kriging variance
Covariance of ε_z and Z	$\mathrm{Var}\left[Z(\mathbf{x}_0)\right] - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c}$

dashed, corresponds to where

$$\begin{aligned} z \leq z_t \\ \text{and} \\ \varepsilon_z < z - z_t, \end{aligned}$$

i.e. to where the intervention is indicated if z is known without error, but $\tilde{Z} > z_t$. The other error condition is that $z > z_t$ and $\tilde{Z} \leq z_t$. This is represented by the dark grey space in Figure 2.
we may therefore, compute the joint probabilities that $z(\mathbf{x}_0) \leq z_t$ and $\varepsilon_z < z - z_t$ by

$$P\left(z(\mathbf{x}_0) < z_t, \varepsilon_z < z(\mathbf{x}_0) - z_t\right) = \int \int f_{z,\varepsilon_z}(z, \varepsilon_z)\mathrm{dz}\ \mathrm{d}\varepsilon_{\mathrm{z}}, \tag{4}$$

where $f_{z,\varepsilon_z}(z, \varepsilon_z)$ is the joint normal distribution of $z(\mathbf{x}_0)$ and ε_z with parameters in Table S1 and the corresponding probability that $z(\mathbf{x}_0) < z_t$ is

$$P\left(z(\mathbf{x}_0) < z_t\right) = \int_{-\infty}^{z_t} f_z(Z)\mathrm{dz}, \tag{5}$$

and the desired conditional probability

$$P\left(\varepsilon_z < z(\mathbf{x}_0) - z_t | z(\mathbf{x}_0) < z_t\right) = \frac{P\left(z(\mathbf{x}_0) < z_t, \varepsilon_z < z(\mathbf{x}_0) - z_t\right)}{P(z(\mathbf{x}_0) - z_t)}. \tag{6}$$

1.2 Implicit loss function

1.2.1 Logistic cost model

In this section we describe how the function defined in Lark and Knights (2015) to return the costs of n samples over an area A km^2:

$$\mathrm{C(n)} = \omega + \mathrm{vn} + \beta \mathrm{At_r}, \tag{7}$$

where ω are the fixed costs, v cost of laboratory analysis per unit, and β the field costs per work day per team. The variable t_r is time taken to sample per km^2 at a density of r per km^2.

Consider a unit area containing the n sample locations. Following Beardwood et al. (1959), the expected distance to travel between sample points can be written as

$$\mathcal{D} \;=\; k\sqrt{n}. \tag{8}$$

If we change the area in which the sample points are distributed to some value A, then the distance travelled is scaled by $\sqrt{A}$ and so

$$\mathcal{D}_A \;=\; k\sqrt{An}, \tag{9}$$

and so we may write the distance travelled to sample n points per unit area as

$$\mathcal{D}_n \;=\; k\sqrt{\frac{n}{A}}. \tag{10}$$

Assuming that the rate of travel is a random variable independent of sample density, we can therefore conclude that the time taken per unit area to travel between sample points is proportional to the square root of sample density

$$\mathcal{T}_t \;=\; \tau_1\sqrt{\frac{n}{A}}. \tag{11}$$

Similarly, assuming that the sampling time is a random variable independent of sample density (time at a sample site), sampling time per unit area is proportional to sample density

$$\mathcal{T}_s \;=\; \tau_2\frac{n}{A}. \tag{12}$$

Given these results, we may propose as a model for total sampling time per unit area

$$\mathcal{T}_o \;=\; \beta_1\sqrt{\frac{n}{A}} + \beta_2\frac{n}{A} + \beta_0 + T + \varepsilon, \tag{13}$$

where β_0 is a constant to allow for fixed time requirements, T is a random effect of mean zero for between-team variation in sampling time and ε is a random effect of mean zero for the between-day (residual) variation.

1.3 Fitting to data

In order to compute the variable t_r, we extracted the required data from the geostatical survey conducted in Malawi for the GeoNutrition project (Gashu et al., 2021). There were 8 teams that collected a total of 1812 sites of soil and crop samples were visited, this is described in detail by Gashu et al. (2021), Botoman et al. (2022) and Kumssa et al. (2022). For each team-day from the GeoNutrition survey of Malawi we have extracted the following:

- Number of points sampled.
- Mean time spent travelling per sample, removing the maximum inter-sample interval each day due to 'lunch break effect'. The units were in minutes.
- Mean time spent at a sample site. The units were in minutes.
- Length of the sampling day. The units were in minutes. The mean value is 331.
- The total area sampled that day. This is defined as the area of the sample domain which is in the Voronoi cell for the day's sample points. Unit were in square kilometres (km^2).

These variables are combined. We then compute the following:

- The total time spent sampling per unit area, $\mathcal{T}_o$ in Eq [13] above, for each team–day.
- Sample density, $\frac{n}{A}$, for each team–day.
- The square root of sample density.

We can then fit a linear mixed model for $\mathcal{T}_o$ in which the fixed effects are $\sqrt{\frac{n}{A}}$ and $\frac{n}{A}$ and in which team is a random effect.

The anova table for the model is as follows

Effect	num DF	denom DF	F-ratio	P
Square root of Sampling density	1	294	347.21	<0.0001
Sampling Density	1	294	9.12	0.0027

This shows significant effects of both powers of sample density.

The estimated model coefficients are as follows

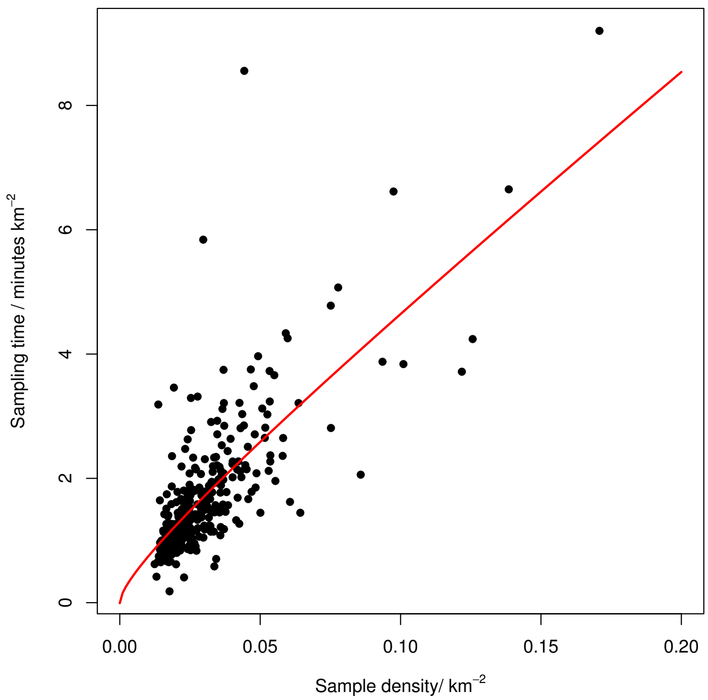


Figure S3: Scatter plot showing the data and fitted model.

Coefficient	Estimate	SE
β_0	-0.007	0.51
β_1	4.08	4.89
β_2	33.6	11.12

The data and fitted model are shown bon Figure S3.

Worked example

Rumphi district: Area 4,769 km^2

Sample size	Sample Density /km^{-2}	Predicted sample effort /min km^{-2}	Total sample effort / team–days*
200	0.0419	2.238	35.6
500	0.1048	4.837	76.9
1000	0.2097	8.907	141.6

*Given total area of Rumphi and assuming a mean sampling day of 331 minutes (as above)

1.4 Offset Correlation

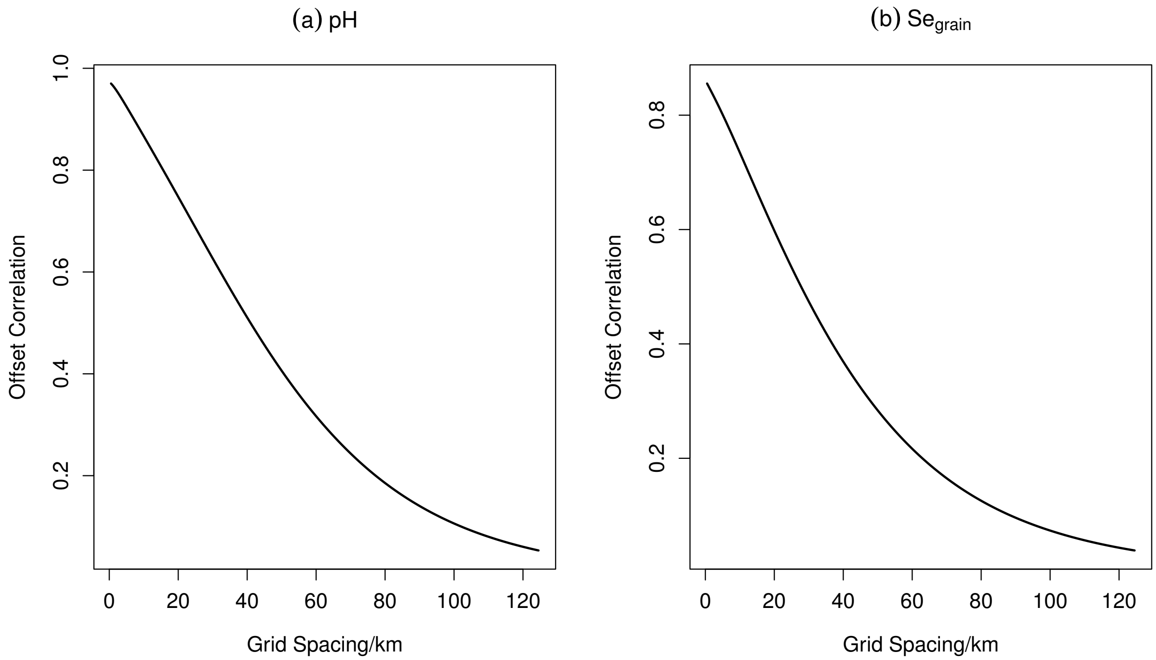


Figure S4: A plot of offset correlation and grid spacing for (a) soil pH and (b) Se_{grain} in Malawi.

2 Test methods: charts presented to the stakeholders

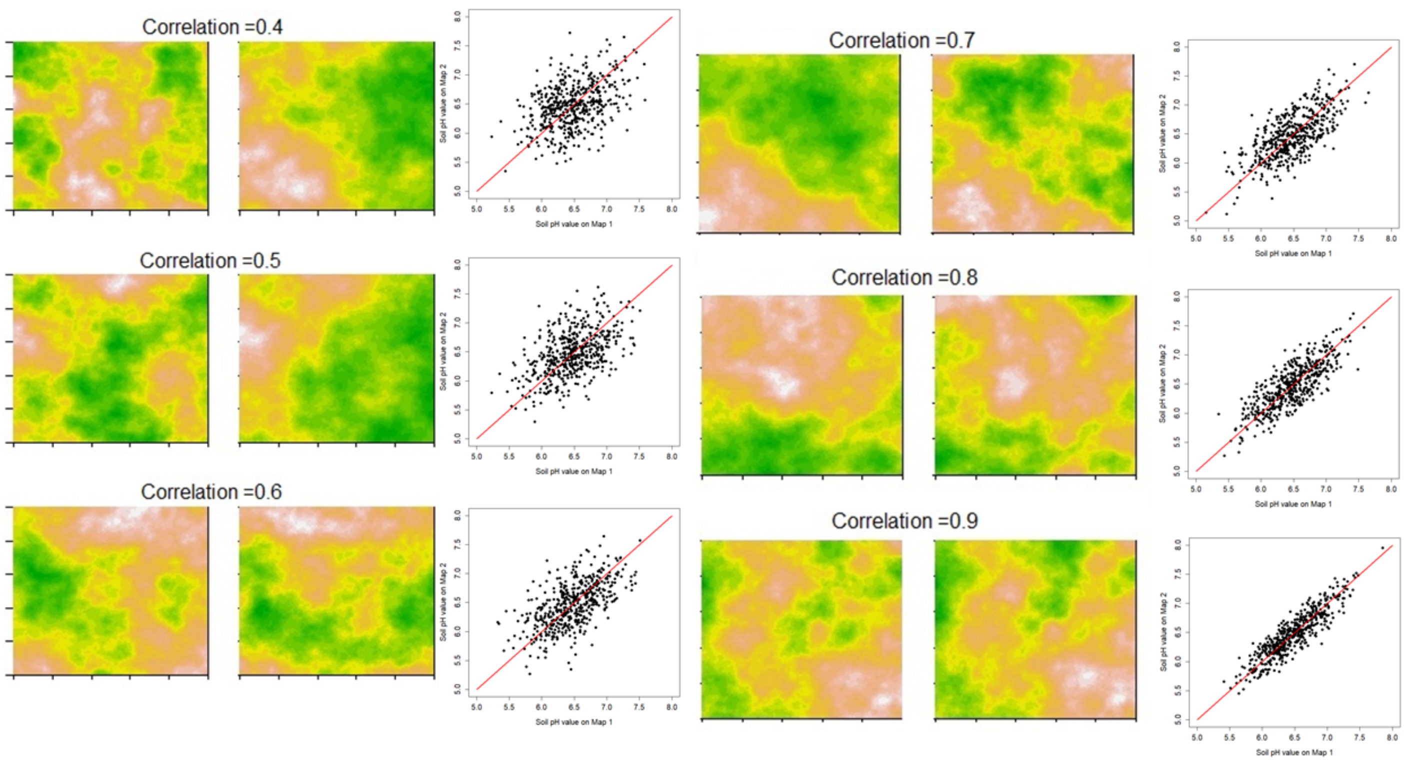


Figure S5: The pairs of example maps of soil pH, each pair being based on a different grid spacing, with a different offset correlation and corresponding scatter plots that illustrated the strength of the correlation.

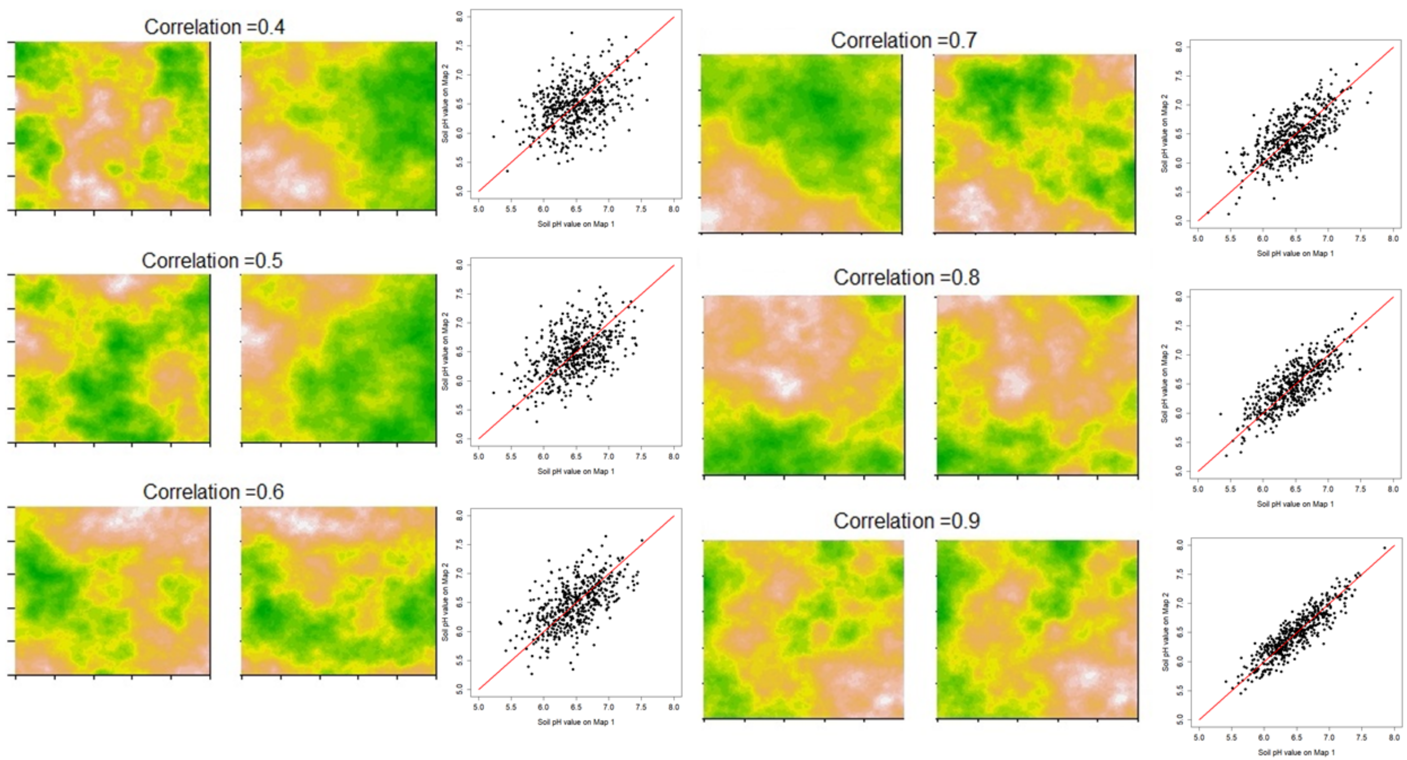


Figure S6: The pairs of example maps of Se concentration in grain, each pair being based on a different grid spacing, with a different offset correlation and corresponding scatter plots that illustrated the strength of the correlation.

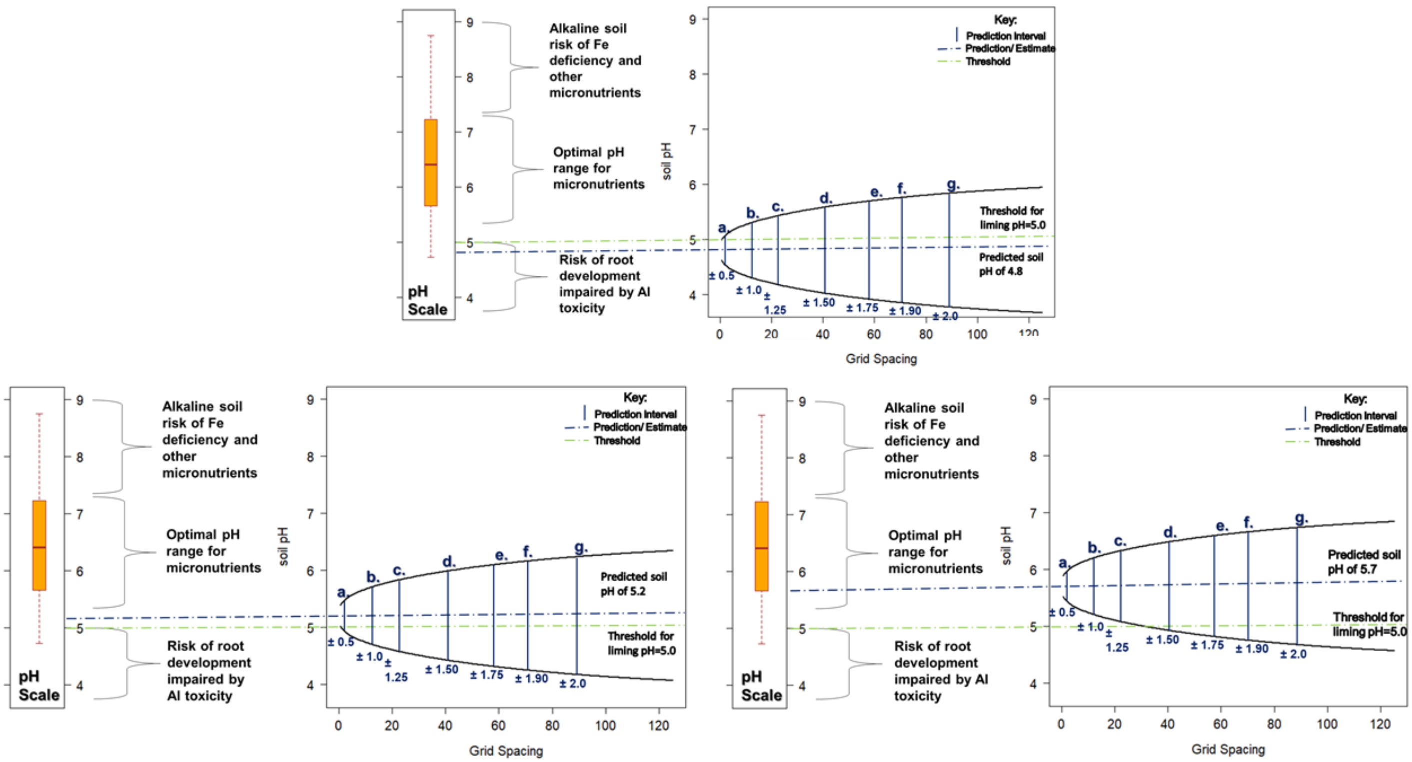


Figure S7: Chart consisting of box plot of the distribution of the soil pH, a graph of the lower and upper prediction intervals for the prediction for grid spacings from 0 to 120 km. With a blue line corresponding to the prediction and the green one for the threshold value.

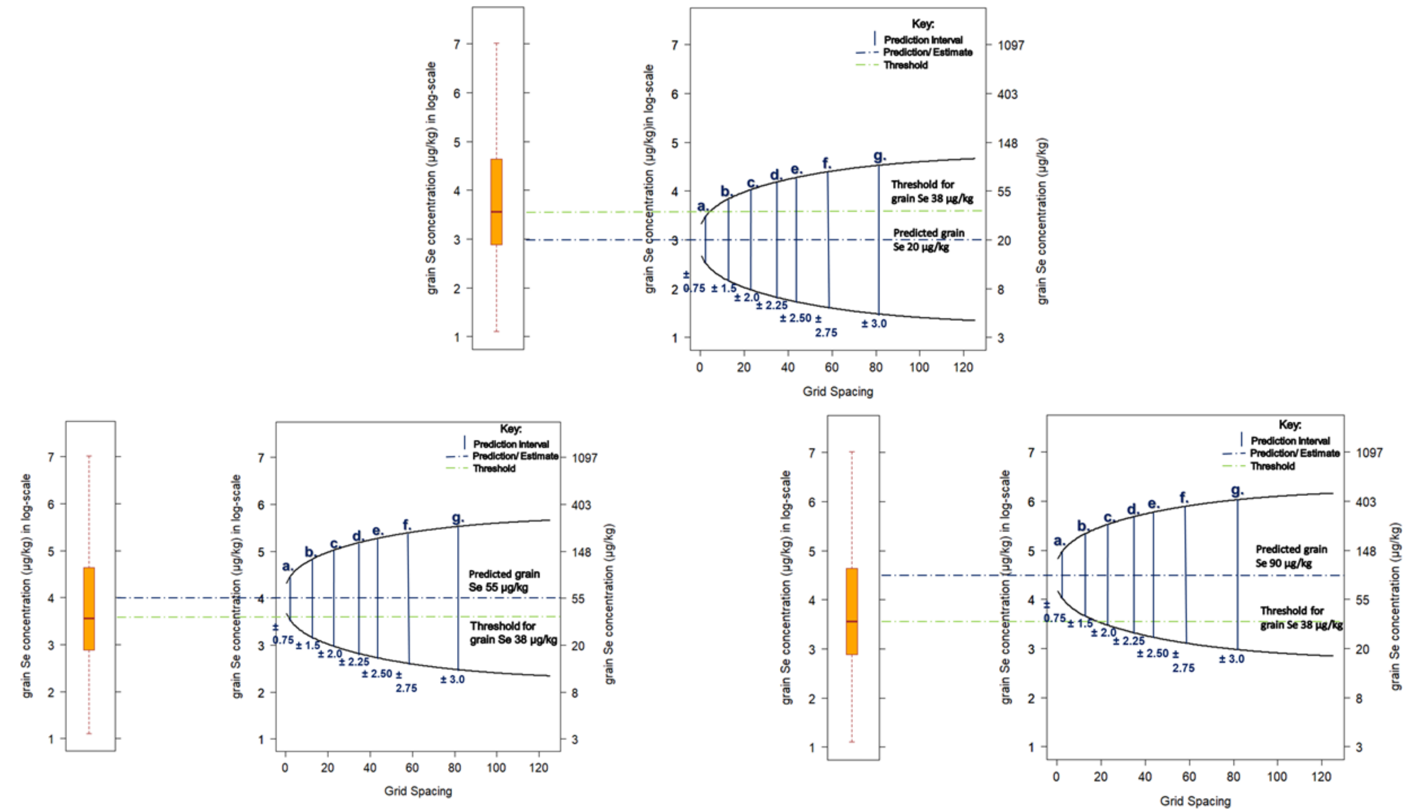


Figure S8: Chart consisting of box plot of the distribution of the Se concentration in grain, a graph of the lower and upper prediction intervals for the prediction for grid spacings from 0 to 120 km. With a blue line corresponding to the prediction and the green one for the threshold value.

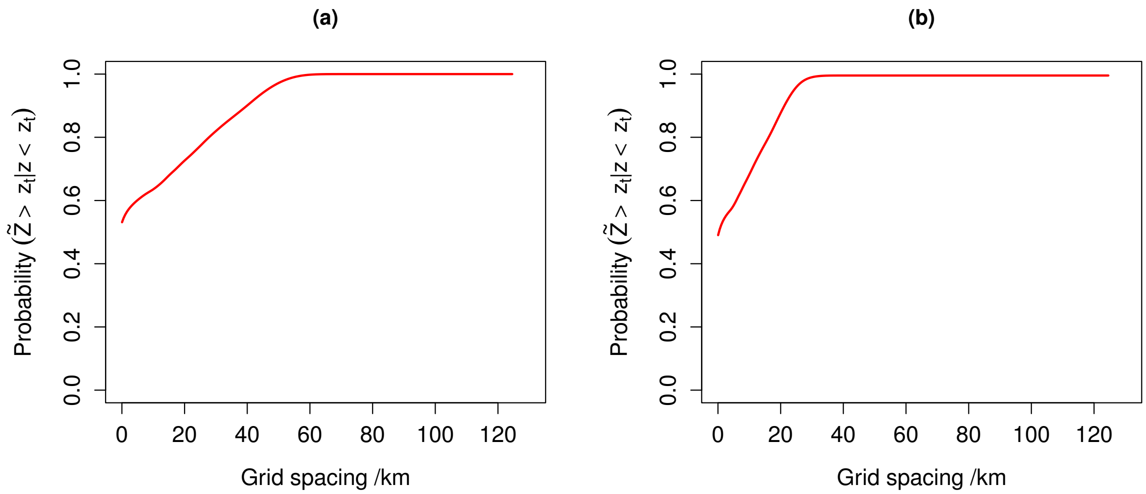


Figure S9: Graph showing the probability, given that an intervention is required at $\mathbf{x}_o$ that, due to error in prediction, the mapped variable does not show this. z_t is the threshold of interest. (a) is for soil pH and (b) for Se_{grain} concentration.

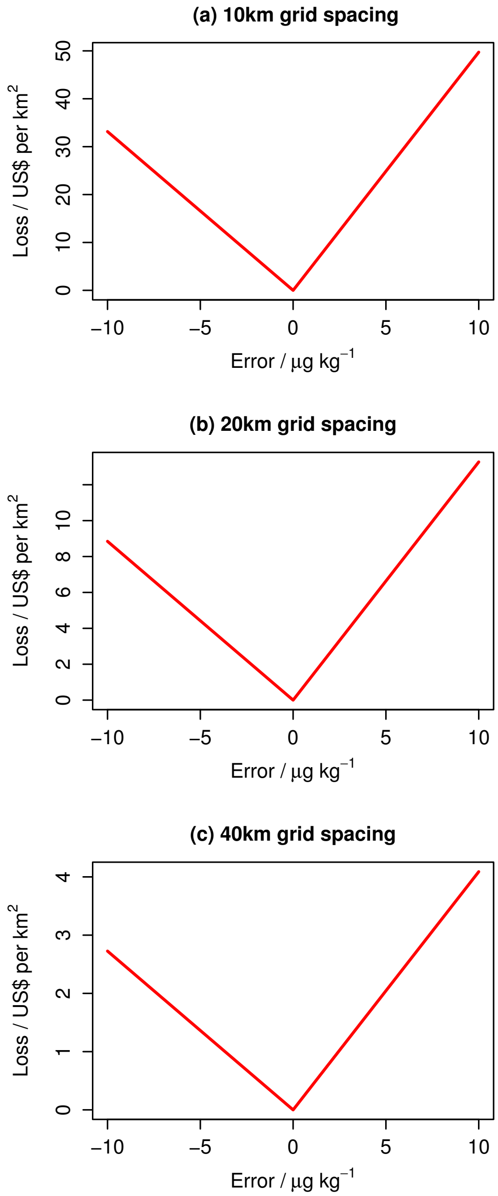


Figure S10: Three specified implicit loss functions for predictions Se concentration in grain an administrative district in Malawi presented to the participants.

References

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 55, pages 299–327. Cambridge University Press.

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., Mcgrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Broadley, M. R., Lark, R. M., and Nalivata, P. C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*.

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., Towett, E. K., Walsh, M. G., Wilson, L., Young, S. D., and Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, 594(7861):71–76.

Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., Muleya, P., McGrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Haji, I. R., Broadley, M. R., Gashu, D., and Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in Ethiopia and Malawi [Dataset].

Lark, R. M. and Knights, K. V. (2015). The implicit loss function for errors in soil information. *Geoderma*, 251-252:24–32.

Chapter 6

General discussion

6.1 The stakeholder perspective on spatial uncertainty and digital soil mapping

Mineral micronutrient deficiencies (MND) are problematic in many countries in sub-Saharan Africa at large. Solutions to address MND exist and these include agronomic biofortification (Broadley et al., 2010; Botoman et al., 2022) and provision of food supplements (Joy et al., 2019, 2022). However, these interventions are costly and should be undertaken at relevant locations. Therefore, spatial information is critical in addressing MND in countries including Ethiopia and Malawi. Stakeholders require spatial information on micronutrient status of soils and crops to design programmes to address micronutrient deficiency in their respective regions. Agronomists working in extension services need to understand the information on whether the concentration of micronutrients falls below a threshold in order to advise farmers at farm level on fertiliser and lime application. Those in public health and nutrition would require spatial on which locations have micronutrients which fall below a threshold in order make decisions on supply of supplements and fortified food.

The spatial information required by stakeholders to make decisions in uncertain and the stakeholders are aware of this. Therefore, the information on uncertainty needs to be understood by its users. However, most end-users feel uneasy with uncertainty and most of them are unsure of how to use and communicate the uncertainty (Arrouays et al., 2020). In this PhD research, one of the main aims was to establish how best stakeholders can be assisted in making decisions when using information about the uncertainty of spatial predictions of concentration of Se in a staple grain. It was important to establish first how best to communicate the uncertainty (Chapter 2). It was found that general-measures of uncertainty, prediction intervals and kriging variance, are of limited value and do not help the end-user to make informed decisions. However, there were more positive responses to threshold-based methods which were found to be clearer by a larger proportion of the participants. The results showed that probabilities, related to some specific interpretation of the spatial information (e.g. nutrient concentration relative to a significant threshold) to make a decision, are preferred by the stakeholders. Calibrated phrases and pictographs added some value to interpretation of probabilities, although there is no strong evidence that they should be preferred to raw probabilities. Verbal interpretations of probabilities are widely used because of the assumption that end-users find probabilities expressed in numerical forms difficult to understand.

The next aspect considered was how can probabilities be presented in a map and used by the stakeholder? Maps with clearly delineated boundaries are often clearer to end-users (Dent and Young, 1981; Malone et al., 2018), and such clear delineation provide interpretable information for making decisions. Perhaps digital soil mapping outputs (DSM) could be presented in this way in order to improve interpretability, consequently help in decision-making but if we simply drew boundaries around areas where the predicted value of a variable, z, is below a threshold value, z_t, uncertainty remains unaccounted for. The results from Chapter 3 in this PhD thesis are a significant step towards addressing some of the challenges of interpretability, and it has been shown that the probability that $z < z_t$ ($P(z < z_t)$) can be presented with clearly delineated boundaries. The delineation of boundaries can be achieved by use of probability isolines for an elicited probability threshold (P_t) to indicate locations which stakeholders believe would require interventions when they consider the uncertain information (see Chapter 3, Figure 5). An intervention would be recommended if the probability that a micronutrient supply falls below a significant threshold exceeds the P_t. This P_t can be elicited from diverse group of stakeholder but attention should be paid to framing.

This PhD research showed how the P_t can be elicited from a diverse group of stakeholders with different data needs, using their preferred method of communicating uncertainty– conditional probabilities, to make decisions on interventions addressing MND in Ethiopia and Malawi, with $\mathrm{Se}_{\mathrm{grain}}$ concentration as an example, in structured experiment (see Chapter 3). The P_t is dependent on the professional group of the stakeholder and how the problem is framed. The results showed that when the problem was framed negatively (probability that nutrient supply is below a significant threshold–in terms of deficiency), there were more consistent responses by either professional group. The stakeholders were more cautious and recommended interventions at much lower probabilities in comparison to when the problem was framed positively (probability that nutrient supply exceeds a significant threshold–in terms of sufficiency).

The elicited P_t represents a complex judgement of losses under errors of omission (fail to intervene where necessary) and commission (intervene where not necessary) in relation to the interpretation of probabilistic information by that particular stakeholder group. In this case study, the elicited P_t was 0.3 (negative framing, i.e., the probability of deficiency). This shows that while stakeholders assign a loss to an unnecessary intervention at a site where $z > z_t$, they assign a larger loss to failure to intervene at a site where $z < z_t$, given the public health costs entailed.

The information requirements of end-users of spatial information are rarely considered in most DSM studies (Wadoux et al., 2021; Lark et al., 2022). It has been suggested that most studies in DSM are 'Quick and Dirty' because they are done without sufficient spatial coverage of observations and do not report uncertainties (Arrouays et al., 2020). The first two results Chapters of this thesis have demonstrated importance of the interaction between stakeholders and shown the value of threshold-based uncertainty measures to support decision-making. Interaction with stakeholders, and communication of the significance of uncertainty is also relevant for the related problem of planning and executing soil and crop surveys. Therefore, it is was important to develop a framework for the planning, execution and evaluation of surveys to address specific requirements of stakeholders. The framework was based on a decision-theory approach to analyse the particular task, to identify the key uncertainties and their implications and so to enable stakeholders to ensure that an approach to survey would meet their needs (Chapter 4). The framework provides a basis of engaging stakeholder to discuss how sampling decisions can be arrived at by considering the relationship of sampling effort, costs and uncertainty. One positive finding in this research is that stakeholders were able to assess the implications of uncertainty in spatial predictions using prior information on variability of the target properties in order to support decisions on a sampling grid spacing. However, more work is required to refine these methods but this is a significant step in addressing the information requirements of end-users

6.2 Further questions on complexity of stakeholder interpretation–consequences and outcomes

Framing, formally irrelevant aspects of how information is presented which may influence interpretation, is an important factor to consider when engaging stakeholders with probabilistic representations of uncertainty. Framing effects could explain the differences in how stakeholders interpreted some question on the interpretative task–Q1 to Q3 (see Chapter 2, Table 3). Q1 was framed positively and there was weak evidence for the differences between the threshold-based methods and general measures for the participants in the Ethiopian elicitation. However, the fact that Q3 was framed negatively might have helped the stakeholders understand this. When comparing Q1 and Q2, there were two confounding effects that could explain the difference, the framing effect and the differences in the magnitude of the predictions. The experimental design was in this elicitation could be improved by considering the effects of framing and this presents scope for further work. The elicitation can be improved by a systematic assessment of statements of uncertainty with predictions of different magnitude. The important question to consider is how to ensure that the end-user understand the information presented to them, in order to get consistent results?

When the question was positively framed (in terms of sufficiency), different values of P_t were obtained for both professional groups. However, there was greater consistency of responses across professional groups when the question was negatively framed, in terms of deficiency, and it led to the stakeholders being conservative. The public health or nutrition group decided to intervene at a much larger probability when compared to their counterparts. The public health or nutrition group decided to intervene at a much larger probability when compared to their counterparts and they might have misunderstood the probabilities because they are accustomed to think about nutrition problems in terms of deficiency. The framing effect was not only evident in this problem, but also on how stakeholders interpreted conditional probabilities in the context of sampling (Chapter 5). The respondents selected the grid spacings where conditional probabilities was 1.0 or very close. These findings shows that probabilities are difficult to interpret and this is evidenced by the framing effects observed. This poses a question on how the effects of framing can be eliminated from the elicitation? One approach would be to accept that framing can not be eliminated but a consistent approach should be used when eliciting P_t from a group of stakeholders. Almashat et al. (2008) suggested the need for preparatory activities that can direct the attention of end-users to implications of errors in both directions to reduce framing effects. Perhaps, an elicitation should be conducted using a consistent framing approach when using threshold-based methods. In the case of MND negative framing resulted in conservative and consistent responses, and this approached should be used consistently.

Threshold-based methods were found to be clear and straightforward, however there was no further evaluation of the correctness of the interpretations of probabilities. Therefore, more work is required to assess the correctness of interpretation of probabilities in order develop sound but accessible ways to engage stakeholders in DSM. Perhaps, pictographs and calibrated phrases may help with this. As for the general-measures, perhaps presenting these methods along with a specific interpretation of spatial information (e.g., whether the prediction is closer to threshold or not) could have yield positive opinions. This suggest more work is required to substantially improve kriging variance and prediction intervals as methods for communicating uncertainty

6.3 Outlook and way forward

Much attention has been paid to quantifying uncertainty in previous DSM studies and there is growing interest communicating the uncertainty to end-users (Heuvelink and Webster, 2022). However, the communication of the uncertainty is not straightforward because it depends on subject matter and background of the stakeholder (Milne et al., 2015). This PhD research is a significant step in addressing some of these concerns of DSM, for example, this thesis has shown the important of having measures of uncertainty that not only communicate clearly the message but as well a basis for stakeholder engagement. The decision-process framework presented in this research can be used in DSM studies to in order to identify the key stakeholders (e.g., sponsors, decision makers and social client) thereby characterize the decisions, states and comes from the spatial information available. This allows for accounting for uncertainty in such a way that risks carried can be identified and possibly mitigated. There is scope for analysing the robustness of this decision process by considering other case studies of the different problems in environmental sciences.

This study will help in better understanding of uncertainties in the data obtained in the GeoNutrition project, thereby facilitating improved use and uptake of that information by decision makers in Ethiopia and Malawi. The GeoNutrition project aims to use spatial information to target areas where specific interventions would be appropriate in order to efficiently use the scarce financial resources. This is important because most people in countries south of the Sahara (e.g. Ethiopia, Malawi, Zambia and Zimbabwe) mainly rely on subsistence farming for their food and income. The decision process presented in this thesis would be important in addressing some questions raised when addressing MND. It is hoped that better decisions will be made on sampling for future surveys in Ethiopia and Malawi, and in other countries which decide to undertake those using better methodologies for national-scale surveys of soil properties or similar environmental variables. This research is not only relevant to Ethiopia and Malawi but to most countries within the sub-Saharan Africa region, because they have similar farming systems. Countries within sub-Saharan Africa may wish to conduct similar surveys such as the GeoNutrition, it is important that the key findings from this study inform the process. Within the framework of the decision-process, stakeholders should be engaged to have discussions centred on uncertainty should be done at the planning phase of the survey. It would be important to engage stakeholders and discuss the key questions relating to information needs (I), stakeholders (S), spatial variation and uncertainty (V) and the resources (R). This will ensure informed decisions are made during sampling, execution and spatial mapping processes of the survey.

Appendices

Publications

List of joint-first authored publications not presented in this thesis:

I made substantial contributions to the GeoNutrition project and have co-authored many publications. Here, I list the two publications, were I am joint-first author and give a short account of my contribution.

Botoman, L., **Chagumaira, C.**, Mossa, A.W., Amede, T., Ander, E.L., Bailey, E.H., Chimungu, J.G., Gameda, S., Gashu, D., Haefele, S.M., Joy, E.J.M., Kumssa, D.B., Ligowe, I.S., McGrath, S.P., Milne, A.E., Munthali, M., Towett, E., Walsh, M.G., Wilson, L., Young, S.D., Broadley, M.R., Lark, R.M., Nalivata, P.C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*, 12. https://doi.org/10.1038/s41598-022-12014-w

Summary

This paper was published in Scientific Reports. The study aimed at understanding soil properties and environmental covariates that affect Zn concentration in maize, the staple crop in Malawi. I contributed mainly on the spatial mapping component of the paper. Using prior ranking of potential covariates for Zn, a formal hypothesis framework described in Section 1.3.1 and Lark (2017) was used for covariate selection. The false discovery rate was controlled with the α-investment. The downscaled mean annual temperature was selected as the covariate that explained a proportion of the spatial variation of grain Zn concentration in Malawi. Within the linear mixed model framework spatial predictions of grain Zn concentration, with downscaled mean annual temperature and spatial coordinates as the fixed effects, were made. The results from this study are important because they provide a basis for further investigations to address dietary Zn deficiencies in Malawi.

Selected co-authored publications from GeoNutrition project work:

- Gashu, D., Lark, R.M., Milne, A.E., Amede, T., Bailey, E.H., **Chagumaira, C**., Dunham, S.J., Gameda, S., Kumssa, D.B., Mossa, A.W., Walsh, M.G., Wilson. L., Young, S. D., Ander, E. L., Broadley, M. R., Joy, E.J.M., McGrath, S.P. (2020). Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia. *Science of the Total Environment*, 733, doi.org/10.1016/j.scitotenv.2020.139231
- Gashu, D., Nalivata, P.C., Amede, T., Ander, E.L., Bailey, E.H., Botoman, L., **Chagumaira, C**., Gameda, S., Haefele, S.M., Hailu, K., Joy, E.J.M., Kalimbira, A.A., Kumssa, D.B., Lark, R.M., Ligowe, I.S., McGrath, S.P., Milne, A.E., Mossa, A.W., Munthali, M., Towett, E.K., Walsh, M.G., Wilson, L., Young, S.D., Broadley, M.R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, 594, 71–76, doi.org/10.1038/s41586-021-03559-3
- Lark. R. M., **Chagumaira, C**., Milne, A. E. (2022). Decisions, uncertainty and spatial information. *Spatial Statistics*, 50, doi.org/10.1016/j.spasta.2022.100619

Bibliography

AfSIS (2015). New cropland and rural settlement maps of Africa. http://africasoils.net/2015/06/07/new-cropland-and-rural-settlement-maps-of-africa/.

Alemu, R., Galew, A., Gashu, D., Tafere, K., Mossa, A. W., Bailey, E., Masters, W. A., Broadley, M. R., and Lark, R. M. (2022). Sub-sampling a large physical soil archive for additional analyses to support spatial mapping; a pre-registered experiment in the Southern Nations, Nationalities, and Peoples Region (SNNPR) of Ethiopia. *Geoderma*.

Almashat, S., Ayotte, B., Edelstein, B., and Margrett, J. (2008). Framing effect debiasing in medical decision making. *Patient Education and Counseling*, 71(1):102–107.

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-De-Forges, A., Morgan, C., Roudier, P., Poggio, L., and Mulder, V. (2020). Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional*, 20:e00255.

Boehmke, B. and Greenwell, B. (2019). *Hands-on machine learning with R*. Chapman and Hall/CRC.

Botoman, L., Chagumaira, C., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Chimungu, J. G., Gameda, S., Gashu, D., Haefele, S. M., Joy, E. J. M., Kumssa, D. B., Ligowe, I. S., Mcgrath, S. P., Milne, A. E., Munthali, M., Towett, E., Walsh, M. G., Wilson, L., Young, S. D., Broadley, M. R., Lark, R. M., and Nalivata, P. C. (2022). Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific Reports*.

Bramer, M. (2020). *Principles of Data Mining*. Springer, London.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Broadley, M. R., Alcock, J., Alford, J., Cartwright, P., Foot, I., Fairweather-Tait, S. J., Hart, D. J., Hurst, R., Knott, P., McGrath, S. P., Meacham, M. C., Norman, K., Mowat, H., Scott, P., Stroud, J. L., Tovey, M., Tucker, M., White, P. J., Young, S. D., and Zhao, F. J. (2010). Selenium biofortification of high-yielding winter wheat (Triticum aestivum L.) by liquid or granular Se fertilisation. *Plant and Soil*, 332(1-2):5–18.

Brown, K. H., Engle-Stone, R., Kagin, J., Rettig, E., and Vosti, S. A. (2015). Use of optimization modeling for selecting national micronutrient intervention strategies: an example based on potential programs for control of vitamin A deficiency in Cameroon. *Food and nutrition bulletin*, 36(3_supplement):S141–S148.

Brus, D. (2019). Sampling for digital soil mapping: A tutorial supported by r scripts. *Geoderma*, 338:464–480.

Brus, D., Kempen, B., and Heuvelink, G. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62(3):394–407.

Burgess, T. M. and Webster, R. (1980). Optimal Interpolation and Isarithmic Mapping of Soil Properties .1. the Semi-variogram and Punctual Kriging. *Journal of Soil Science*, 31(2):315–331.

Burrough, P. A., Beckett, P., and Jarvis, M. (1971). The relation between cost and utility in soil survey(i-iii). *Journal of Soil Science*, 22(3):359–394.

Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., and Broadley, M. R. (2011). Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in malawi. *Scientific Reports*, 1.

Dent, D. and Young, A. (1981). *Soil survey and land evaluation.* George Allen & Unwin.

Fairweather-Tait, S. J., Bao, Y. P., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., and Hurst, R. (2011). Selenium in human health and disease. *Antioxidants and Redox Signaling*, 14(7):1337–1383.

Foster, D. P. and Stine, R. A. (2008). α-investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):429–444.

Gashu, D., Lark, R. M., Milne, A. E., Amede, T., Bailey, E. H., Chagumaira, C., Dunham, S. J., Gameda, S., Kumssa, D. B., Mossa, A. W., Walsh, M. G., Wilson, L., Young, S. D., Ander, E. L., Broadley, M. R., Joy, E. J. M., and McGrath, S. P. (2020). Spatial prediction of the concentration of selenium (Se) in grain across part of Amhara Region, Ethiopia. *Science of the Total Environment*, 733.

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Kalimbira, A. A., Kumssa, D. B., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A. E., Mossa, A. W., Munthali, M., Towett, E. K., Walsh, M. G., Wilson, L., Young, S. D., and Broadley, M. R. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*, 594(7861):71–76.

GeoNutrution (2017). Geonutrition project.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning Second Edition*, volume 27 of *Springer Series in Statistics*. Springer, New York.

Heuvelink, G. B. and Webster, R. (2022). Spatial statistics and soil mapping: A blossoming partnership under pressure. *Spatial Statistics*, page 100639.

Heuvelink, G. B. M. (2018). Uncertainty and uncertainty propagation in soil mapping and modelling. In McBratney, A. B., Minasny, B., and Stockmann, U., editors, *Pedometrics*, pages 439–461. Springer International Publishing.

Hurst, R., Siyame, E. W. P., Young, S. D., Chilimba, A. D. C., Joy, E. J. M., Black, C. R., Ander, E. L., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Stein, A. J., Fairweather-Tait, S. J., Gibson, R. S., Kalimbira, A. A., and Broadley, M. R. (2013). Soil-type influences human selenium status and underlies widespread selenium deficiency risks in malawi. *Scientific Reports*, 3.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer, New York.

Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*. Courier Corporation.

Journel, A. G. and Huijbregts, C. J. (1976). Mining geostatistics.

Joy, E. J., Kalimbira, A. A., Gashu, D., Ferguson, E. L., Sturgess, J., Dangour, A. D., Banda, L., Chiutsi-Phiri, G., Bailey, E. H., Langley-Evans, S. C., et al. (2019). Can selenium deficiency in malawi be alleviated through consumption of agro-biofortified maize flour? study protocol for a randomised, double-blind, controlled trial. *Trials*, 20(1):1–9.

Joy, E. J. M., Ander, E. L., Young, S. D., Black, C. R., Watts, M. J., Chilimba, A. D. C., Chilima, B., Siyame, E. W. P., Kalimbira, A. A., Hurst, R., Fairweather-Tait, S. J., Stein, A. J., Gibson, R. S., White, P. J., and Broadley, M. R. (2014). Dietary mineral supplies in Africa. *Physiologia Plantarum*, 151(3):208–229.

Joy, E. J. M., Broadley, M. R., Young, S. D., Black, C. R., Chilimba, A. D. C., Ander, E. L., Barlow, T. S., and Watts, M. J. (2015). Soil type influences crop mineral composition in Malawi. *Science of the Total Environment*, 505:587–595.

Joy, E. J. M., Kalimbira, A. A., Sturgess, J., Banda, L., Chiutsi-Phiri, G., Manase, H., Gondwe, J., Ferguson, E. L., Kalumikiza, Z., Bailey, E. H., Young, S. D., Matandika, L., Mfutso-Bengo, J., Millar, K., Niksic, M., de la Revilla, L. S., Likoswe, B. H., Phuka, J. C., Phiri, F. P., Lark, R. M., Gashu, D., Langley-Evans, S. C., Ander, E. L., Lowe, N. M., Dangour, A. D., Nalivata, P. C., Broadley, M. R., and Allen, E. (2022). Biofortified maize improves selenium status of women and children in a rural community in malawi: Results of the addressing hidden hunger with agronomy randomized controlled trial. *Frontiers in Nutrition*, 8.

Kerry, R. and Oliver, M. A. (2007). Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, 140(4):383–396.

Künsch, H., Papritz, A. J., Schwierz, C., and Stahel, W. A. (2013). Robust estimation of the external drift and the variogram of spatial data. In *ISI 58th World Statistics Congress of the International Statistical Institute*, Zürich. Eidgenössische Technische Hochschule Zürich.

Lark, R., Chagumaira, C., and Milne, A. (2022). Decisions, uncertainty and spatial information. *Spatial Statistics*, page 100619.

Lark, R. and Cullis, B. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, 55(4):799–813.

Lark, R. M. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, 51(1):137–157.

Lark, R. M. (2017). Controlling the marginal false discovery rate in inferences from a soil dataset with α-investment. *European Journal of Soil Science*, 68(2):221–234.

Lark, R. M., Ander, E. L., Cave, M. R., Knights, K. V., Glennon, M. M., and Scanlon, R. P. (2014). Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland. *Geoderma*, 226:64–78.

Lark, R. M., Ander, L., and Knights, K. (2016). Information for agriculture from regional geochemical surveys: the example of soil pH in the Tellus and Tellus Border data. In Young, M., editor, *Unearthed: impacts of the Tellus surveys of the north of Ireland*, pages 193–204. Royal Irish Academy, Dublin, Ireland.

Lark, R. M., Bishop T. F., A., and Webster, R. (2007). Using expert knowledge with control of false discovery rate to select regressors for prediction of soil properties. *Geoderma*, 138(1-2):65–78.

Lark, R. M., Cullis, B. R., and Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (e-blup) with reml. *European Journal of Soil Science*, 57(6):787–799.

Lark, R. M., Hamilton, E. M., Kaninga, B., Maseka, K. K., Mutondo, M., Sakala, G. M., and Watts, M. J. (2017). Nested sampling and spatial analysis for reconnaissance investigations of soil: an example from agricultural land near mine tailings in zambia. *European Journal of Soil Science*, 68(5):605–620.

Lark, R. M. and Marchant, B. P. (2018). How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters? *Geoderma*, 319:89–99.

Ligowe, I., Phiri, F., Ander, E., Bailey, E., Chilimba, A., Gashu, D., Joy, E., Lark, R., Kabambe, V., Kalimbira, A., et al. (2020). Selenium deficiency risks in sub-saharan african food systems and their geospatial linkages. *Proceedings of the Nutrition Society*, 79(4):457–467.

Malone, B. P., Odgers, N. P., Stockmann, U., Minasny, B., and McBratney, A. B. (2018). Digital mapping of soil classes and continuous soil properties. In McBratney, A. B., Minasny, B., and Stockmann, U., editors, *Pedometrics*, pages 373–413. Springer, Dordrecht.

Marchant, B. P. and Lark, R. M. (2007). Estimation of linear models of coregionalization by residual maximum likelihood. *European Journal of Soil Science*, 58(6):1506–1513.

Matérn, B. (1960). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forst Survey and Other Sampling Investigations*. Esselte.

Matheron, G. (1963). Principles of geostatistics. *Economic Geography*, 58:1246–1266.

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B. (2003). *On digital soil mapping*, volume 117.

McBratney, A. B., Webster, R., and M, B. T. (1981). The design of optimal sampling schemes for local estimation and mapping of of regionalized variables-i. theory and method. *Computers & Geosciences*, 7(4):331–334.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(35):983–999.

Milne, A. E., Glendining, M. J., Lark, R. M., Perryman, S. A. M., Gordon, T., and Whitmore, A. P. (2015). Communicating the uncertainty in estimated greenhouse gas emissions from agriculture. *Journal of Environmental Management*, 160:139–153.

Minasny, B. and McBratney, A. B. (2006). A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32(9):1378–1388.

Minasny, B. and McBratney, A. B. (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma*, 140(4):324–336.

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

Minasny, B., McBratney, A. B., and Lark, R. M. (2008). *Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures*. Digital Soil Mapping with Limited Data.

Nussbaum, M., Papritz, A. J., Baltensweiler, A., and Walthert, L. (2012). Organic Carbon Stocks of Swiss Forest Soils. Technical report, ETH Zurich, Institute of Terrestrial Ecosystems.

Oliver, M. A. (2010). *An Overview of Geostatistics and Precision Agriculture*, pages 1–34. Springer, New York.

Oliver, M. A. and Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer.

Papritz, A. (2021). *Tutorial and Manual for Geostatistical Analyses with the R package georob*.

Rayman, M. P. (2012). Selenium and human health. *The Lancet*, 379(9822):1256–1268.

Royle, J. A. and Nychka, D. (1998). An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers and Geosciences*, 24(5):479–488.

Stein, M. L. (1999). *Interpolation of Spatial Data Some theory of Kriging*. Springer-Verlag New.

Tittonell, P., Vanlauwe, B., Misiko, M., and Giller, K. (2011). Targeting resources within diverse, heterogeneous and dynamic farming systems: towards a 'uniquely African green revolution'. In *Innovations as key to the green revolution in Africa*, pages 747–758. Springer.

Wadoux, A. M.-C., Heuvelink, G. B., Lark, R. M., Lagacherie, P., Bouma, J., Mulder, V. L., Libohova, Z., Yang, L., and McBratney, A. B. (2021). Ten challenges for the future of pedometrics. *Geoderma*, 401:115155.

Wadoux, A. M. J. C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, 351:59–70.

Wadoux, A. M. J. C., Brus Dick, J., and Heuvelink Gerard B., M. (2019). Sampling design optimization for soil mapping with random forest. *Geoderma*, 355:113913.

Walvoort, D. J. J., Brus, D. J., and de Gruijter, J. J. (2010). An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers and Geosciences*, 36(10):1261–1267.

Webster, R. and Lark, M. R. (2013). *Field Sampling for Environmental Science and Management.* Taylor and Francis.

Webster, R. and Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, 43(1):177–192.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Natural Environmental Scientists*. John Wiley & Sons Chichester, 2nd edition.

White, R. E. (2006). *Principles and Practice of Soil Science: The Soil as a Natural Resource*. Blackwell Publishing.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Academic press.

Winther, K. H., Rayman, M. P., Bonnema, S. J., and Hegedus, L. (2020). Selenium in thyroid disorders - essential knowledge for clinicians. *Nature Reviews Endocrinology*, 16(3):165–176.

Witten, I. H., Eibe, F., Hall Mark, A., and Pal Christopher, J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Inc.

World Health Organisation (2004). *Vitamin and mineral requirements in human nutrition : report of a joint FAO/WHO expert consultation*. WHO Publications, second edition.