ENVIRONMENTAL RESEARCH
FOOD SYSTEMS

PURPOSE-LED PUBLISHING™

Part of the ENVIRONMENTAL RESEARCH SERIES

ACCEPTED MANUSCRIPT • OPEN ACCESS

Multiple soil and landscape properties are associated with the spatial variation of selenium concentration in maize grain in Malawi

To cite this article before publication: Patson Cleopus Nalivata *et al* 2025 *Environ. Res.: Food Syst.* in press https://doi.org/10.1088/2976-601X/ae033d

Multiple soil and landscape properties are associated with the spatial variation of selenium concentration in maize grain in Malawi

Patson C Nalivata[1,†], Ivy S Ligowe[1,2,†], Christopher Chagumaira[1,3,4,5,†], Abdul-Wahab Mossa[6,†], Tilahun Amede[7], Louise E Ander[8], Elizabeth H Bailey[3], Lester Botoman[1,9], Joseph G Chimungu[1], Sam Gameda[10], Dawd Gashu[11], Stephan M Haefele[4], Edward J M Joy[4,12], Diriba B Kumssa[3], Steve P McGrath[4], Alice E Milne[4], Moses Munthali[9], Erick Towett[13], Markus G Walsh[14], Lolita Wilson[3], Scott D Young[3], Martin R Broadley[3,4,†] & R Murray Lark[3,4,†]

[1]Lilongwe University of Agriculture and Natural Resources (LUANAR), Bunda College Campus, P.O. Box 219, Lilongwe, Malawi

[2]Mzuzu University, Faculty of Environmental Sciences, Private Bag 201, Luwinga, Mzuzu

[3]School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, United Kingdom

[4]Rothamsted Research, West Common, Harpenden AL5 2JQ, United Kingdom

[5]Agri Food and Biosciences Institute, Newforge Lane, Belfast, BT9 5PX, United Kingdom

[6]ISRIC - World Soil Information, Wageningen, Netherlands

[7]AGRA-Sustainably Growing Africa's Food Systems, Westends Tower, Nairobi 00100, Kenya

[8]Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, United Kingdom

[9]The Department of Agricultural Research Services, P.O. Box 30779, Lilongwe 3, Malawi

[10]International Maize and Wheat Improvement Center (CIMMYT), ILRI Sholla Campus, P.O. Box 5689, Addis Ababa, Ethiopia

[11]Centre for Food Science and Nutrition, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia

[12]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom

[13]World Agroforestry (ICRAF), United Nations Avenue, P.O. Box 30677, Nairobi, Kenya

[14]Africa Soil Information Service, 2423 Pan American Freeway, Albuquerque, New Mexico

[†]These authors contributed equally to this work.

E-mail: pnalivata@luanar.ac.mw

Abstract. Dietary selenium (Se) deficiency is widespread in Malawi, due to the limited supply of Se in the predominantly maize based-food system characterised by low Se concentration. In this study, the aim was to examine the spatial variation of Se in maize grains in Malawi, in relation to soil properties and landscape features. Co-located soil and maize grain samples were collected in a spatially representative survey. Selenium concentration in maize, soil properties, and environmental covariates were determined. Soil and environmental variables were tested as potential predictors of Se concentration in maize. A False Discovery Rate (FDR) control was used within a Linear Mixed Model (LMM) framework. Selenium concentrations in maize ranged from below detection limits (7.69 µg kg^{-1}) to 1852 µg kg^{-1} with mean and median values of 39.1 and 16.8 µg kg^{-1} respectively. The ranges of concentrations of Se fractions in soil were (i) soluble Se 0.181–18.8 µg kg^{-1} with mean and median values of 3.94 and 3.29 mg µg kg^{-1} respectively; (ii) adsorbed Se 0.019–119 µg kg^{-1} with mean and median values of 3.72 and 3.02 µg kg^{-1} respectively; (iii) organically bound Se 9.43–1334 µg kg^{-1} with mean and median values of 123 and 92.3 µg kg^{-1} respectively. A LMM for maize Se concentration was used for which the independent log transformed variables of soil soluble Se, adsorbed Se, oxalate extracted oxides, soluble and exchangeable sulphur had predictive value ($p < 0.01$ in all cases, with FDR controlled at < 0.05). Downscaled mean annual temperature also explained some of the spatial variation in grain Se concentration. Spatial variation of Se in maize showed relationships with soil and environmental variables, which can be used to identify areas most at risk of Se deficiency and thus inform policy responses. However, only a small proportion of the variation was explained indicating more analysis of Se geochemistry in soil may provide more explanatory insights.

Keywords Selenium, Geostatistics, Micronutrients, Maize, Malawi

1. Introduction

Selenium (Se) is a micronutrient essential for human health [1]. Based on dietary Se supply and direct measurement of biomarkers, it is estimated that Se deficiency is present in more than 30% of people living in sub-Saharan Africa (SSA) [2, 3, 4, 6, 5]. In Malawi, estimated Se deficiency prevalence rates of 35.5 % and 62.5 % have been reported, respectively, based on nationally representative surveys of blood plasma Se concentrations from samples collected in 2015 and 2016 [4, 6].

In many SSA food systems, locally grown cereals provide most of the Se in a person's diet [5]. Access to animal-source foods, as a richer source of dietary Se than cereals, are often limited due to their availability and cost [3, 5]. There is considerable variation in the grain Se concentration of cereals, depending on where the crop is grown. For example, in Malawi, the concentration of Se in maize (*Zea mays* L.) grown on Vertisols was reported to be ten-fold larger than the grain concentration from most other soil types in the country [7], and this was linked in subsequent studies to direct evidence of differences in dietary intake [8] and Se status [9] among smallholder communities farming in these areas.

Comprehensive data on grain Se concentration in staple cereal crops have been reported recently for Ethiopia and Malawi, from surveys representing most of their arable land areas [10, 11]. Grain Se concentrations varied by several orders of magnitude within both countries (e.g ranging from below detection limits (7.69 µg kg^{-1}) to 1852 µg kg^{-1} for maize in Malawi). Furthermore, there was strong evidence of spatially correlated variation in grain Se concentration of cereal crops, at distances of more than 100 km, again in both countries. What this means is that, for subsistence farmers, and other rural dwellers dependent primarily on locally grown staples, their location is likely to be the single biggest factor in determining whether they are likely to be at risk of Se deficiency.

Understanding spatial variation of Se concentration in grain is vital for agronomic biofortification interventions aimed at improving dietary intake of Se. Sources of spatially-dependent variation of Se concentration in grains have previously been investigated using surveys and statistical modelling. Gashu et al. 2021 [11] reported that soil pH was positively correlated with grain Se concentration for teff, wheat and maize in Ethiopia and for maize in Malawi. Grain Se concentration was positively correlated with mean annual temperature for teff (*Eragrostis tef* (Zucc.) Trotter) and wheat (*Triticum aestivum* L.) in Ethiopia, and maize in Malawi [11]. In addition, grain Se concentration was negatively correlated with mean annual precipitation for teff, wheat and maize in Ethiopia. In a more detailed analysis of the soils of the Amhara region of Ethiopia, [10] reported that 'soluble' (extracted in 0.01 M KNO_3) and 'adsorbed/exchangeable' (extracted in 0.016 M KH_2PO_4) fractions of soil Se, along with soil pH, were positively correlated with the Se concentration in teff and wheat grain. Whilst these fractions of soil Se are operationally defined, they are considered to represent 'plant-available' fractions.

The sample materials in the survey reported by [12] can be used to determine the same crop and soil Se variables which [11] reported from Amhara. The results from [12] indicate that the survey of Malawi is of sufficient intensity to support a spatial analysis of the joint variation of crop and soil variables. However, we may expect Se to behave differently in the soil, both in terms of retention and uptake by plants, and so a further study to examine the variation of maize grain Se concentration and its joint variation with soil and environmental properties could be expected to yield novel results.

The aim of this study was therefore to investigate whether the 'plant-available' fractions of Se explain the spatially correlated variation of grain Se concentration of maize in Malawi. As with the earlier study in Amhara region [11], we used an approach based on hypothesis-testing to select covariates. To reduce the risk of over-fitting the model we employed False Discovery Rate (FDR) control, while maintaining power to detect useful predictors of the target variable by employing the alpha-investment method of [13]. The α-investment methods uses a ranking of the soil properties and wider environmental/landscape factors that are considered by experts to be most likely, a priori, to influence grain Se concentration. The landscape factors included in this analysis were downscaled precipitation and temperature, terrain index (TIM), slope, and vegetation index.

2. Materials and methods

Most of the materials and methods used in this study are described in greater detail by [10, 11] and [12]. Here we provide a short overview. The final and definitive data set, as published by [14], is the one that was used in the analyses described below.

2.1. Design and field sampling

The sampling domain for Malawi was defined as the raster cells in the European Space Agency Climate Change Initiative map allocated to a land-cover class that included 'cropping' in its designation. The objective of the sampling was to provide adequate spatial coverage to support spatial prediction of the variables of interest. A detailed description of the method used to select sampling points is given by [11].

The primary objective of the sampling was to provide adequate spatial coverage to support spatial prediction of the variables of interest. More detail on the method used to select sample points to achieve this is provided by [11]. Paired soil and grain samples were collected by trained teams from 1812 locations. Only the locations where maize crop was grown (1608 locations ≈ 89 % of total samples), were included in this study.

2.2. Grain and soil analyses

Selenium concentration in grain was determined using inductively coupled plasma mass spectrometry (ICP-MS; iCAPQ, Thermo Fisher Scientific, Bremen, Germany) following acid digestion with 70% HNO_3 (Trace Analysis Grade) in a Multiwave Pro 5000 microwave digestion system (Anton Paar). The following soil properties were determined: soil organic carbon (SOC, dry combustion), effective cation exchange capacity (eCEC) and exchangeable cations (hexamine cobalt trichloride solution), amorphous oxides (AlOx, FeOx, MnOx; ammonium oxalate extraction), Olsen P; pH in water (1:2.5 solid to solution ratio) and pH in 0.01 M $Ca(NO_3)_2$ (1:10 solid to solution ratio), and quasi-total elemental concentration determined by ICP-MS after extraction with Aqua Regia. Different Se fractions in soil were determined by using a 3-step sequential extraction scheme as described in detail in [14]. The scheme is designed to sequentially extract three operationally defined fractions (i) a 'soluble' fraction in 0.01 M KNO_3 (Se_{Sol}), (ii) a 'specifically adsorbed' fraction in 0.016 M KH_2PO_4 (Se_{Ads}), and (iii) an organically bound fraction in 10% tetra methyl ammonium hydroxide (Se_{Org}).

Selenium concentrations in grain fell below the detection limits (average of 7.69 µg kg^{-1}) in 409 samples therefore they were removed from the analysis. Consequently, a total of 1199 samples were included in the analysis in the current study. Values of LODs were calculated for each separate ICP-MS run (n = 6) and samples analysed in each run were compared to the corresponding LOD, and when the concentration of Se in grain was ≤ LOD, the sample was removed from the analysis. The limits of detection for selenium, determined across the six ICP-MS runs, varied from 2.69 to 18.1 µg kg^{-1}, with a mean of 7.69 µg kg^{-1}. For details, please refer to [14]. It is worth noting that we used a different approach in [11], where the average value of LODs, calculated for each ICP-MS run (n = 6), was used.

2.3. Extraction of environmental covariates

We selected environmental covariates which were judged to be possible predictors of grain selenium concentration through their effect on, or status as a proxy for, factors of crop growth and soil conditions. These covariates were the MERIT Digital Elevation Model [15] and derived variables, specifically surface slope and the topographic index which represents the up-slope area that potentially contributed runoff to a point. In addition, we considered climate variables from the CHELSA set [16, 17] (downscaled mean annual temperature and precipitation) and the Enhanced Vegetation Index (EVI) computed from measurement by the MODIS remote sensor satellite [18]. Specifically, we used the average over the period 2000–2016 of the 250-m EVI product (MOD13Q1).

2.4. Data analyses

The data analyses approaches used are described in detail by Botoman et al. 2022 [12]. Summarised they were:

2.4.1. Linear mixed model To identify links between soil properties and Se concentration in grain and to model the spatial variation of Se in maize grain, a Linear Mixed Model (LMM) framework was used. The variable is modelled as a combination of fixed effects (linear functions of soil properties or environmental covariates), a correlated random effect, and an independent and identically distributed (iid) random error (nugget effect). The nugget effect incorporates variation due to measurement error and factors that vary over short distances relative to the spacing of sample points.

2.4.2. Statistical inference and FDR control with α-investment In a LMM framework, the evidence that adding fixed effects to a simpler model achieves a significant improvement by computing the log-ratio statistic:

$$L = 2(\ell_1 - \ell_0), \tag{1}$$

where ℓ_1 and ℓ_0 are the maximised log-likelihoods from fitting the model with the additional fixed effects, and the simpler model without them, respectively [19]. Under the null hypothesis, where the additional fixed effects are not related to the dependent variable, this statistic is asymptotically distributed as chi-square with degrees of freedom equal to the number of additional fixed effects. To avoid the problem of multiple hypothesis testing when evaluating multiple models with different predictors [22] we controlled the FDR over a sequence of tests [20]. This is the expected proportion of rejected null hypotheses which are false rejections, and it can be controlled by various methods [21]. While it is desirable to control FDR when evaluating evidence to include predictors in a final model, this comes at the cost of reduced statistical power to detect informative predictors. The method of α-investment due to [13] can improve the power of testing with FDR control. In this method the threshold value against which the p-value for a new covariate is tested depends on the α-wealth, a quantity which is reduced on acceptance of a null hypothesis increased on rejection, while still controlling FDR.

We used this combination of FDR control with α-investment when modelling grain Se concentration, following [23]. Models were fitted sequentially, first, with a 'null model' with the only fixed effect a spatial trend identified in exploratory analysis of the data. This model was used, rather than a model with a constant mean as the only fixed effect, because the latter model would violate assumptions of second-order stationarity when a spatial trend is pronounced [24]. The null model was fitted by maximum likelihood (ML). The first predictor was then included as a fixed effect, the model refitted and then the log-likelihood ratio statistic (Equation 1) was computed. If the p-value for this test exceeded 0.05 then the predictor was dropped, otherwise it was provisionally retained, and the next predictor was considered. Once all the predictors had been considered the p-values for each were compared to thresholds according to the α-wealth controlling FDR at a target value of 0.05. Those predictors for which the p-values was less than the FDR threshold were retained for inclusion in a final model which was refitted by residual maximum likelihood (REML).

Separate rankings of the soil properties and the environmental covariates as potential predictors of the Se concentration in maize grain were based on a priori understanding of the processes involved, not from data exploration. We used the same ranking of predictors for grain Se which we used previously for analysis of data on grain Se in Ethiopia [10]. These rankings reflected a consensus view of soil and plant scientists on the project team. They were not shown any data on grain Se from Malawi but were shown the correlations among the candidate independent variables (Figure B4). This is because the value of a predictor depends not only on the extent to which it is related to the target variable, but also on its correlation with predictors already included in the model. Such a correlation introduces redundancy. If two predictors are quite strongly correlated then the expert should select just one for inclusion early in the testing sequence. It should be noted that this ranking procedure allows us to improve the statistical power of the overall selection procedure. If the ranking is poor, not representing the real value of the predictors, then the gain in power will be reduced, but the control of the FDR is unaffected.

2.4.3. Exploratory data analysis and model-fitting Summary statistics of the predictor variables were examined along with the octile skewness coefficient [25]. While no assumptions are made about the distribution of independent variables in the LMM, we preferred to avoid using strongly skewed variables as large values in the upper portion of a skewed distribution would be given undue influence Variables for which the absolute value of the conventional skewness coefficient, based on data moments, exceeds 1 are commonly considered for transformation [24]. However, like all statistics based on second or higher-order moments, the skewness coefficient is sensitive to small numbers of outlying observations. For this reason we considered a robust alternative, the octile skewness [25] which takes absolute values larger than 0.2 for a wide range of random variables with a conventional skewness outwith $[-1, 1]$. Those variables with a pronounced octile skew coefficient were transformed to natural logarithms before they were used.

Exploratory spatial analysis of the data was undertaken by creating classified post-plots with the plot.geodata function from the geoR package for the R platform [28, 27]. This reveals evidence of spatial trends in plots, and spatial post-plots of the data, with symbols coded to indicate the quartiles of the data set to which they belong. 'Saturated' exploratory models for grain Se concentration with (i) all soil properties as fixed effects, along with a trend in the eastings identified from the spatial plots, and (ii) all environmental covariates and easting as fixed effects were then fitted by ordinary least squares. The decision as to whether a transformation of the data was needed to justify the assumption that the random variation of the variables is normally distributed was based on exploratory statistics and plots of these residuals as described above.

The parameters of the LMM were estimated by ML or REML, using the likfit function from the geoR library. The ML method allows the most straightforward comparison of models with different fixed effects, necessary for the sequential testing for variable selection. However, REML is preferable for estimation of the random effects parameters, [19, 26]. The variance parameters are the variance of the spatially correlated random variation (σ^2), and a parameter (ϕ) which quantifies how the spatial correlation decays with distance. The smoothness of the spatially correlated random variation is quantified by a parameter (κ) which can be challenging to estimate, so we followed [26] and used a profiling method. This was done for the null model, and the selected value of κ was then used for all others. The final random effects parameter is the variance of an iid component which is uncorrelated at scales resolved by sampling.

Once a set of covariates had been selected a final model was fitted by REML estimation of the LMM parameters. The model was then tested by cross-validation. Each observation was withheld from the data set in turn and predicted from the model and the remaining data. We then computed standardised squared prediction errors (SSPE), the square of the difference between each observation and its cross validation prediction, standardized by the prediction error variance. The mean and median SSPE were computed. For a valid model we expect the mean value to be close to 1 and the median to be close to 0.455 [29].

2.4.4. Spatial prediction Once a LMM was fitted with selected environmental predictors, the Empirical Best Linear Unbiased Predictor (E-BLUP) was computed for each raster cell at which the selected covariates were recorded [19]. This prediction combines a 'regression-type' component, based on the selected covariate(s), and a 'kriging-type' prediction from the random effects. The prediction minimises the expected value of the prediction error variance, which quantifies the uncertainty of the prediction, and which was mapped alongside predictions of grain Se concentration.

Although the prediction error variance quantifies uncertainty, it is not necessarily an accessible measure for all users of information [30]. To illustrate how uncertainty might be communicated, we considered a threshold grain concentration of 38 µg kg^{-1}. Grain with a smaller concentration of Se provides less than one third of the expected average requirement of Se for a woman of reproductive age within a 330-g serving daily intake [30, 31]. On the assumption (checked in the cross-validation procedure) of normally distributed prediction errors, it is possible to compute the probability at a prediction location that the true grain Se concentration is below the threshold. Decision-makers might consider an intervention in these circumstances. Chagumaira et al. (2021) found that decision-makers with varied mathematical experience, and with differing professional background, found such probability maps effective guides to the interpretation of uncertain information. Further, they found [31] that the average threshold probability (that grain Se falls below the threshold), at which the same set of decision-makers favoured intervention was 0.31. This indicates that the information users are generally more concerned to avoid the error of failing to intervene where Se supply is deficient than they are to avoid intervention where it is not necessary. We therefore mapped the probability that grain Se concentration is <38 µg kg^{-1}, and also showed those regions where this probability exceeds 0.31.

3. Results

3.1. Exploratory analysis

Summary statistics of maize grain Se concentration, residuals from the exploratory model, and cross-validation errors are shown in Table 1. Except for pH, which is reported on a logarithmic scale, soil properties (Table 2) are mostly skewed and therefore were transformed to logarithms (natural log). We decided to transform the measurements of grain Se concentration to natural logarithms after inspection of summary plots and statistics for the residuals from the exploratory model (Supplementary Figures B1 and B2).

3.2. Ranking of predictor variables as predictors of Se concentration in grain

The ranking of soil properties and of environmental covariates is presented in Table 3 and was based on the ordering used in [10]. The top three ranked properties were the different operationally defined fractions of 'soluble', 'adsorbed', and 'organically-bound' Se in soil, followed by pH, which were hypothesised to be the most likely predictors of Se concentration in grain. The sum of oxalate-extractable Fe, Mn, and Al oxides was then included, followed by soluble and organic fractions of soil sulphur (S), which will interact with plant Se uptake [33, 32], followed by soil organic carbon (SOC), oxalate-extractable P, and phosphorus buffer index (PBI).

The top three ranked environmental covariates were down-scaled precipitation, down-scaled mean annual temperature, and slope.

Table 1. Summary statistics of Se concentration in grain (n=1603), residuals from fitted exploratory saturated models and cross-validation errors for the E-BLUP with coordinates and downscaled mean annual temperature as fixed effects.

	Concentration Se_{maize} (µg kg^{-1})	Residuals from model with, soil properties as covariates[a]	Residuals from model with, environmental covariates[a]	Cross-validation errors for the E-BLUP
Mean	39.1	0.00	0.00	0.00
Median	16.8	-0.03	-0.04	-0.07
Minimum	-1.85	-4.26	-3.55	-3.93
Maximum	1852	4.61	4.22	4.72
Standard deviation	92.3	1.11	1.00	0.95
Skewness	9.65	0.01	0.32	0.36
Octile skewness	0.509	0.06	0.04	0.12

[a] Residuals from fitting with $\log_e$ transformed maize grain Se.

Table 2. Summary statistics of soil properties proposed as predictors of Se concentration in grain

Variable	Original variables					$\log_e$-transformed				
	Mean	Median	Standard	Skew	Octile skewness	Mean	Median	Standard	Skew	Octile skewness
Se_{Nit}[a] (µg kg^{-1})	3.94	3.30	2.99	1.29	0.23	1.07	1.19	0.83	-0.30	-0.21
Se_{Pho} (µg kg^{-1})	3.72	3.02	3.79	16.6	0.32	1.07	1.10	0.72	-0.67	-0.04
Se_{TMAH} (µg kg^{-1})	123	95.5	104.9	3.12	0.43	4.55	4.53	0.72	0.13	0.04
pH	6.37	6.29	0.69	0.61	0.13					
Oxides (µg kg^{-1})	3853	3238	2625	2.89	0.29	8.09	8.08	0.56	0.33	-0.02
S_{Nit} (mg kg^{-1})	4.41	2.83	15.2	24.3	0.36	1.08	1.04	0.70	0.91	0.01
S_{TMAH} (mg kg-1)	65.46	46.9	76.3	6.23	0.38	3.82	3.85	0.88	-0.70	-0.03
SOC (%)	1.13	0.96	0.68	2.32	0.33	-0.02	-0.04	0.28	0.22	0.03
Oxalate P (mg kg^{-1})	235	155	242	2.35	0.49	4.97	5.05	1.10	-0.80	-0.05
PBI	73.2	57.2	64.3	5.20	0.38	4.08	4.05	0.63	0.29	0.05

[a] The subscripts Nit, Pho and TMAH denote the soluble (nitrate extraction), exchangeable (phosphate extraction) and organic (TMAH extraction) fractions. SOC denotes soil organic carbon. Oxides denotes the sum of oxalate-extractable Fe, Al and Mn oxides. Oxalate P denotes oxalate-extractable P. PBI denotes phosphorus buffer index.

Table 3. Sequence of predictors for maize grain Se concentration (both soil properties and environmental covariates) for testing with the α-investment

Order	Soil property	Order	Environmental covariate
1	Se_{Nit}[a]	1	Downscaled mean annual precipitation
2	Se_{Pho}	2	Downscaled mean annual temperature
3	Se_{TMAH}	3	Slope
4	pH	4	Topographic index
5	Oxides	5	Enhanced vegetation index
6	S_{Nit}	6	MODIS Band 7
7	S_{TMAH}	7	MODIS Band 1
8	SOC	8	MODIS Band 2
9	Oxalate P	9	MODIS Band 3
10	PBI		

[a] The subscripts Nit, Pho and TMAH denote the denote the soluble (nitrate extraction), exchangeable (phosphate extraction) and organic (TMAH extraction) fractions. SOC denotes soil organic carbon. Oxides denotes the sum of oxalate-extractable Fe, Al and Mn oxides. Oxalate P denotes oxalate-extractable P. PBI denotes phosphorus buffer index.

3.3. Model-fitting

The 1st-, 2nd-, 4th, 5th-, 6th-, 8th and 9th-ranked soil properties, Se_{Nit}, Se_{Pho}, pH, oxides, S_{Nit}, SOC and Oxalate P were retained as predictors for grain Se concentration by the FDR criterion (Figure 1a).

The variogram functions for the null model with coordinates filtering spatial trend and models for the selected soil properties, added in succession, are shown in Figure 2. The variogram represents the spatial dependence of the correlated random effect. It

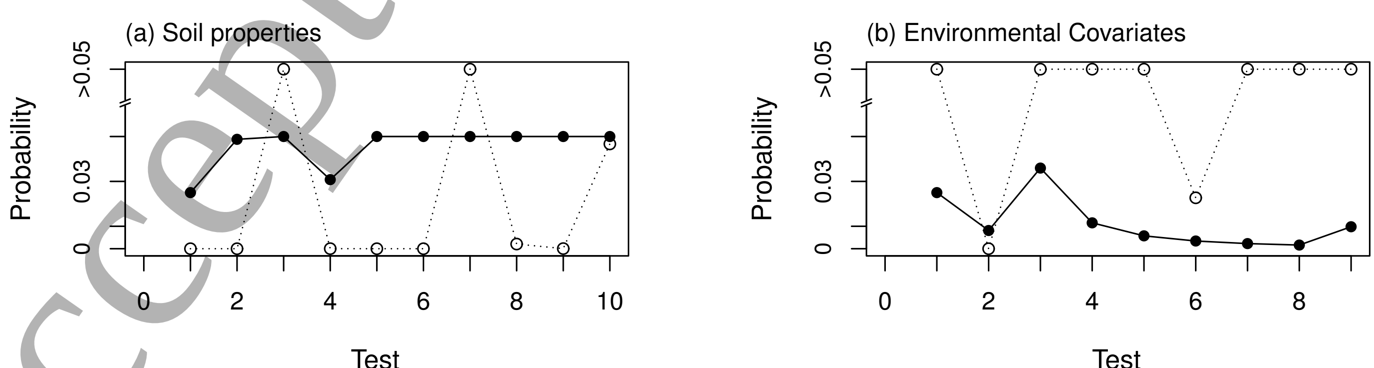

Figure 1. The p-values (open circles) for successive tests on predictors added to the model for maize grain Se concentration from (a) soil properties and (b) environmental covariates. Tests are on addition of variables in the order given in Table 3. The solid circles are the threshold for rejection of each null hypothesis under the FDR control.

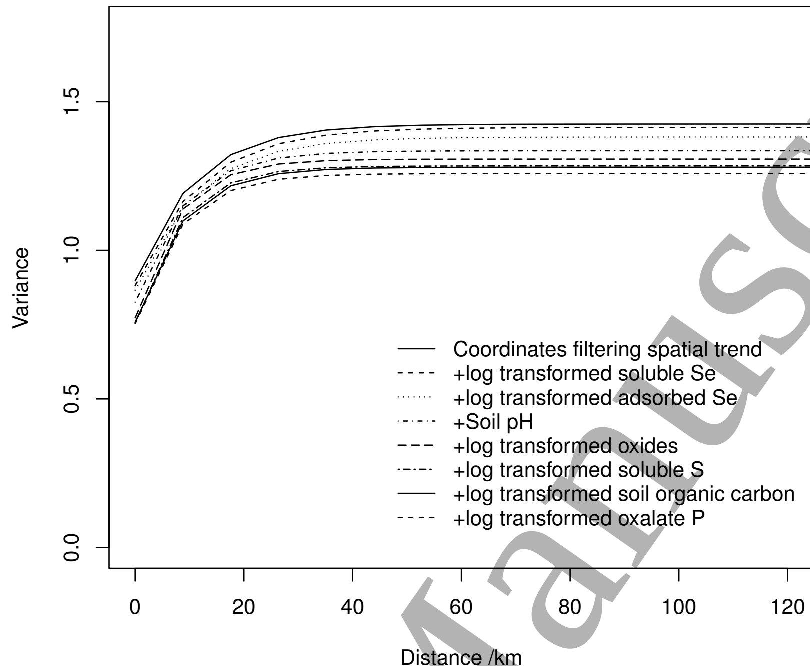

Figure 2. Variogram functions for the null model (coordinates filtering spatial trend) for maize grain Se concentration, and for successive models with selected soil properties added as predictors

is half the expected squared difference between the random components of the target variable at two locations, modelled as a function of the distance between them.

The variance of the iid random effect is the apparent intercept of the variogram, and the function increases to a maximum which is the sum of the variances of both random effects, also called the *a priori* variance. The dependence of the variogram on the separation distance depends on the ϕ and κ parameters. The *a priori* variance is smaller for the random effects of models with predictors added as fixed effects, the reduction representing the information which these terms provide. Here the reduction of this component by adding the selected terms for the final model, expressed as a proportion of the *a priori* variance for the model with spatial coordinates only is small (0.04). The spatially uncorrelated random effect with variance τ^2 accounts for sources of variation spatially dependent at finer scales resolved by sampling, or without any spatial dependence. This will include measurement error. For this reason, it is also useful to compute the adjusted R^2 value for the spatially correlated variation alone, i.e. the reduction in σ^2 on adding predictors as fixed effects to the model expressed as a proportion of this variance component for the null model.

Variogram functions for the null model (coordinates the only fixed effect) and then for successive models in which soil properties were added as predictors, retained with

FDR control, are shown in Figure 2. Although there is strong evidence linking selected soil properties to Se concentration in maize grain, these properties only account for a small fraction of the variation. While soil properties are one source of variation in grain Se, other factors appear to contribute substantially more.

Soil $\mathrm{Se_{Nit}}$ and $\mathrm{Se_{Pho}}$ have a positive coefficient (Table A1), indicating that positive deviations from the spatial trend in grain Se are associated with larger concentrations of soluble and exchangeable Se in the soil. This is plausible and consistent with previous studies, such as that of [44]. The soil pH was also retained in the model, also with a positive coefficient. It should be noted that the sign of the coefficient in a statistical model depends, in part, on the other covariates included. However, here the positive sign makes sense mechanistically. The pH of the soil and redox potential together affect the speciation of Se, with Se(VI), the , more accessible form, increasingly predominant over the less accessible Se(IV) form as $\mathrm{pH} + \mathrm{pE}$ increases, [40]. It has been shown in Malawi [7] that Se uptake into maize grain tends to be larger soils with $\mathrm{pH} > 5$. The negative effect of metal oxides in the model is consistent with findings that these can absorb forms of Se, reducing availability, an effect reduced by soil organic carbon [41]. Soil $\mathrm{S_{Nit}}$ has a negative coefficient which is attributable to the strong competition between selenate and sulphate for transporter sites and hence for plant uptake [34, 35]. Soil organic carbon is included in the predictive model, with a negative sign. This can be attributed to the way in which Se, by substituting for S, can bond with C and O in humic molecules, reducing availability. However, it should be noted that SOC can reduce Se absorption on metal oxides, increasing availability. Oxalate-extractable P has a positive correlation with grain Se in the model. This is not consistent with observations in more acid soil conditions, where Se(IV) dominates in speciation, and there can be competition between Se and P for plant uptake, see [45]. This could be explained through the better development of the root system in maize plants with a good P supply [42, 43], which in turn would improve uptake of Se.

The single environmental covariate selected by the FDR approach was mean annual temperature, with a positive coefficient. Table 4 shows the estimated parameters for this model, relative to the null model with coordinates as the only fixed effects, and Figure 3 shows the variogram functions. There was a larger reduction in the unexplained variation when mean annual temperature was added as a predictor than for the model with soil properties, as shown by the value of $\breve{R}^2_{adj}$ (0.057) for the selected model, after FDR, reported in Table 4. The $\breve{R}^2_{adj}$ for the final model used for spatial prediction is 0.07.

The summary statistics for the cross-validation errors are presented in Table 1 and their exploratory plots are shown in Figure B3. The assumption of normal errors appears plausible.

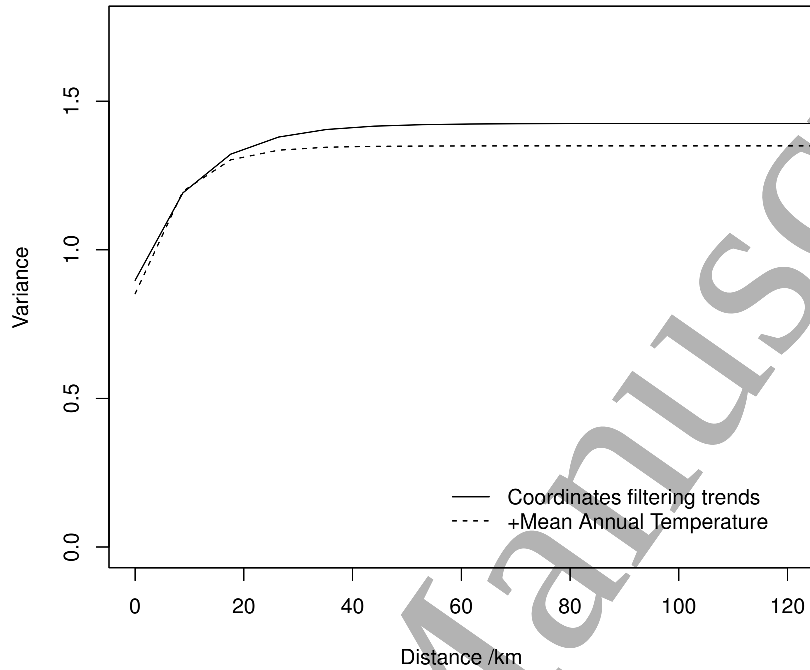

Figure 3. Variogram functions for the null model (coordinates filtering spatial trend) for maize grain Se concentration, and for successive models with selected environmental covariates added as predictors.

Table 4. Parameter estimates for the null model for transformed grain selenium concentration (spatial coordinates the only fixed effects) and a model with selected environmental covariates as additional fixed effects

Predictand	Predictor coefficient				R^2_{adj}	$\breve{R}^2_{\mathrm{adj}}$	κ	τ^2	σ^2	ϕ
	β_0^{a}	β_1	β_2	β_3						
		Easting	Northing	Mean Annual Temperature						
Null model	24.4138	-0.0025	-0.0023				0.5	0.8972	0.5277	10.8
Model after FDR	20.3792	-0.0042	-0.0021	0.0162	0.0529	0.0571	0.5	0.8520	0.4976	7.44
Final model	20.5162	-0.0043	-0.0022	0.0162	0.0446	0.0702	0.5	0.8708	0.4907	8.25

[a] β_0–β_3: fixed effects coefficients β_0 is a constant and β_i is the coefficient for the ith random effect; R^2_{adj}: the difference between the variance of the correlated random effect (σ^2) for the null model and the proposed model expressed as a proportion of that variance for the null model; $\breve{R}^2_{adj}$: the difference between the variance of the correlated random effect (σ^2) for the null model and the proposed model expressed as a proportion of that variance for the null model; κ: smoothness parameter of the Matèrn correlation function; τ^2: variance of the iid random effect (nugget variance); σ^2: variance of the correlated random effect; ϕ:distance parameter of the correlation function.

The mean standardised squared prediction error is 1, but the median is 0.384. This is smaller than expected; the 95% confidence interval for the median under a valid model is [0.40, 0.51], and the kriging variances may be somewhat large, possibly due to outlying observations in the data, so inferences will be conservative in the sense that uncertainty is slightly overestimated.

3.4. Spatial mapping

Spatial mapping of grain Se was done by computing the E-BLUP for the model with mean annual temperature as a fixed effect in addition to spatial covariates. Although these fixed effects left 80% of the variation of grain Se unexplained, the kriging-type component of the E-BLUP, based on the spatial correlation of the random effects, gives an optimal local prediction with quantified uncertainty. There are pronounced spatial patterns in the predicted concentrations of Se in maize at the national scale (Figure 4). There are large concentrations in the Shire River valley in the south of the country and marked east – west trends in the southern and central provinces. Variations in predicted grain Se over small regions where the covariate changes markedly should be interpreted with caution. For example, the predicted concentrations are smaller around the Mulanje Massif in the south east of the country. This may reflect the influence of the mean annual temperature covariate however, it is likely that this reflects primarily the markedly larger concentrations in the hot, low altitude Shire valley than elsewhere in the country, and the prediction of trends associated with shorter-scale topographic variation may be artefacts. Note also that the kriging variances are greater over the Mulanje Massif.

Figure 5a shows the probability that grain Se concentration is less than the threshold of 38 µg kg^{-1}. In Figure 5b, these values are presented using the verbal scale with calibrated phrases proposed by the IPCC [36]. Following [37] the phrases and the probability ranges to which they correspond are both presented. The probability that grain Se concentration is below the threshold is 'Unlikely' or 'Very unlikely' (1–33 %) in the southern part of the Shire valley and near Salima on the south-west shore of Lake Malawi. It is 'Very likely' in much of the north of the country, and near Dedza south of Lilongwe. Over much of the country the probability of being below the threshold is 'As likely as not', i.e., 33–66%. In such areas local decisions on interventions based on predicted grain Se content should probably be based on direct local measurements.

Figure 6 shows the probability the concentration of Se in maize grain falls below 38 µg kg^{-1} annotated with the average probability threshold, $\mathrm{P_t}$, value of 0.31 as the red probability isoline on (a). This average probability threshold can be applied by stakeholder groups in Malawi, those in agronomy/soil science and public health/nutrition. Interventions to address Se deficiencies would be recommended where the probability takes a greater value than the average probability threshold. Figure 6 (b) shows an area of 17,208 km^2 where the probability is above the $\mathrm{P_t}$ 0.31. These are the locations where interventions addressing Se deficiencies should be targeted. Agronomists

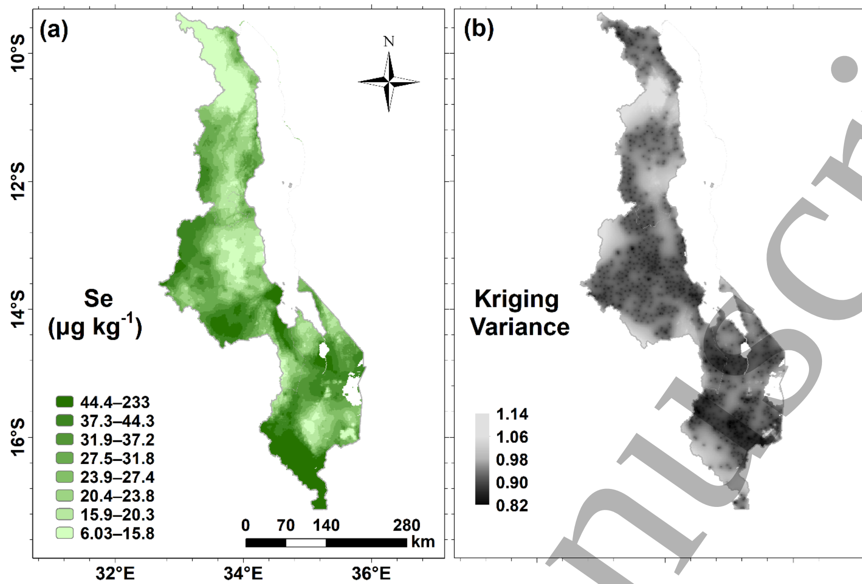

Figure 4. Grain Se concentration in maize across Malawi. (a) E-BLUP predictions, and (b) the prediction error variance (expected squared error) of the E-BLUP.

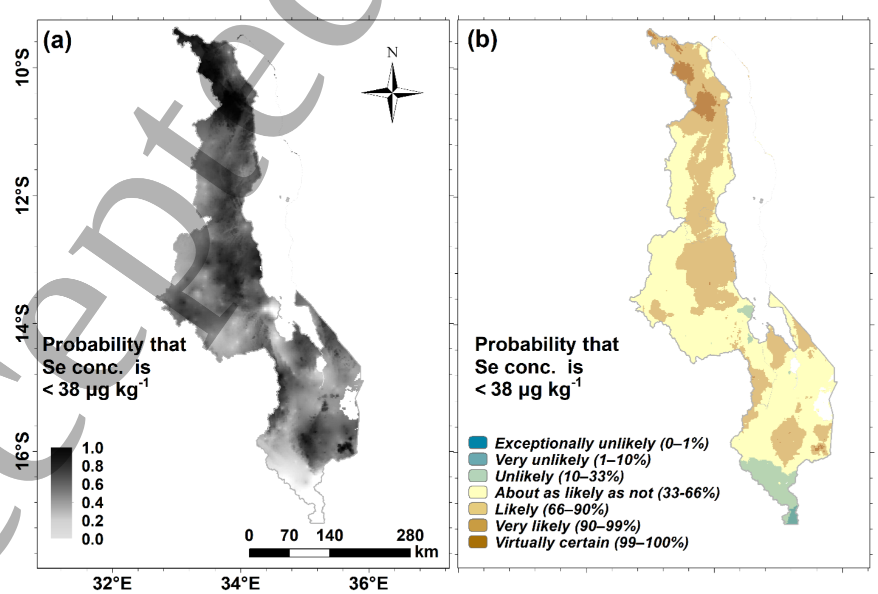

Figure 5. Probability that concentration of Se in maize grain across Malawi is less than 38 µg kg^{-1} expressed on (a) numerical scale, (b) expressed according to 'calibrated phrases'

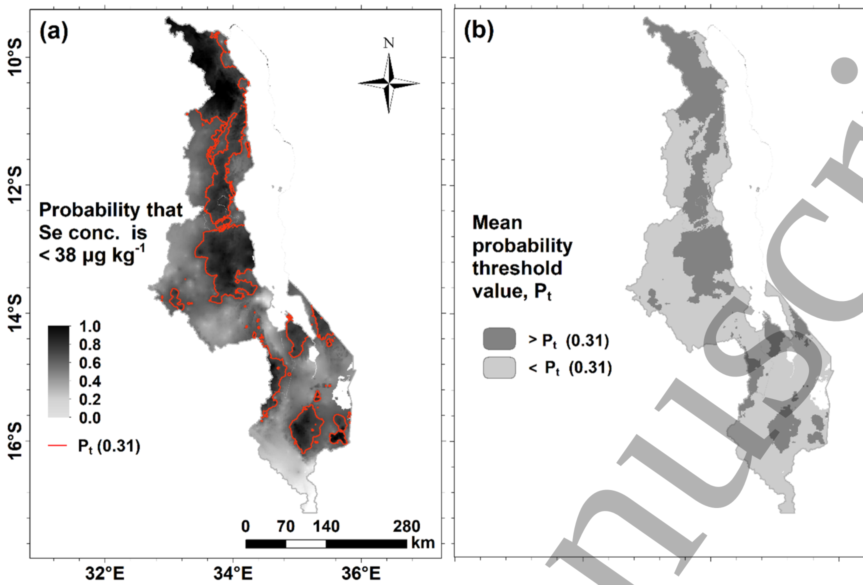

Figure 6. Probability that the concentration of Se in maize grain across Malawi is <38 µg kg^{-1} expressed on a numerical scale. The red line probability isoline on (a) is the mean probability value, P_t , applied by stakeholders' groups (agronomy/soil science and public health/nutrition) in Malawi would to recommend interventions to address Se deficiencies. (b) shows the proportion of the area recommended for interventions by the stakeholder groups.

and soil scientist may advocate for agronomic biofortification as an intervention. Public health and nutrition specialist may recommend provision of Se-fortified food products in those regions with probability above the 0.31.

4. Discussion

If we consider a median grain Se concentration of 16.8 µg kg^{-1} and a reference daily maize intake of 342.8 g capita^{-1} day^{-1} from food balance sheets [2], the typical dietary intake of Se from maize alone in Malawi is 5.76 µg capita^{-1} day^{-1}. This intake represents 10.4% of a Recommended Dietary Allowance (RDA) of 55 µg capita^{-1} day^{-1} for Se. However, grain Se concentrations from the survey ranged from ranging from below detection limits (7.69 µg kg^{-1}) to 1852 µg kg^{-1}. An individual could therefore be consuming 0.353–635 mg capita^{-1} d^{-1} from maize, or 0.64–1150% of the Se RDA, depending upon the location from where this maize is sourced. Location is a critical factor in the likely prevalence of Se deficiency among populations, notably, where a single dietary staple crop dominates and is produced locally [11]. A previous dietary survey in Malawi reported low dietary Se supply. It was estimated that 70% of the population are consuming insufficient Se with an average daily intake range of 27–45 µg capita^{-1} day^{-1} [7, 9, 3], compared to the Se RDA of 55 µg capita^{-1} day^{-1}. Smaller intakes are likely in rural areas and among poorer households who have limited access to more Se-rich food sources such as meat, fish, and vegetables [3].

Geographical differences in Se concentration and intake have been reported previously in Malawi from (i) compositional analysis of dietary intakes in two locations [9], (ii) national-scale dietary surveys linked to food composition data based on convenience sampling [7], (iii) the concentrations of Se in blood plasma and urine as population biomarkers of Se status [4, 38]. The current spatially representative survey of maize grain Se concentration is consistent with data from these earlier studies.

Hurst et al. (2013) designed a cross-sectional study to compare the Se status of women living in locations with contrasting soil types and maize grain Se concentrations. They observed marked differences in the Se status of blood plasma and casual urine. The median value of plasma Se concentration in the Zombwe Extension Planning Area (EPA) was 53.7 µg L^{-1} (ranging between 32.3–78.4 µg L^{-1}; SD = 9.7 µg L^{-1}), which was less than half of the median value, 117 µg L^{-1}, seen in Mikalango (range 82.6–204 µg L^{-1}, SD = 22.5 µg L^{-1}) which was selected because of the local Vertisol soil type used for local crop production. Moreover, Se concentration in casual urine samples in Zombwe EPA ranged between 4.1 and 13.3 µg L^{-1} with a median value of 7.3 µg L^{-1} (SD = 2.0) which was one third that of median value, 25.3 µg L^{-1}, observed in Mikalango EPA (range 12.4–106 µg L^{-1}; SD =18.9 µg L^{-1}). This is consistent with the results of the current survey, which demonstrate low concentration of Se in grain (a light green area in Figure 4) in Northern regions (where Zombwe EPA is located), and high grain Se concentrations (a dark green area in Figure 4) in Southern regions (where Mikalango EPA is located). High erythrocyte Se concentration [39] is consistent with greater plasma Se concentration of people living in areas where Vertisols are prevalent in Malawi.

Chilimba et al. (2011) estimated dietary Se intake in Malawi surveying Se concentrations in maize grain and soil from 88 field sites. They predicted a widespread suboptimal dietary intake and Se deficiency risks in Malawi. They noted spatial variation in Se concentration in maize grain which, in turn, was determined by soil properties, where Se concentration in maize grain was higher by up to 10-fold in crops grown on soils in southern Malawi with high pH (>6.5).

The predictive value of soil factors for maize grain Se concentration was significant, albeit these factors (Se_{Nit}, Se_{Pho}, oxides and S_{Nit}) explained just a small proportion of the random spatial effects (adjusted $R^2 = 0.046$) within the overall model, once the fixed spatial trend effect had been accounted for. However, much of the unexplained variation shows spatial dependence, so the E-BLUP predictions of grain Se concentration are more reliable than predictions based on the fixed effects only.

Finally, it is interesting to compare these findings for grain and soil Se with previous findings, based on the same survey, for Zn [12]. One marked difference is seen in the scale of spatial dependence of the micronutrient concentration in grain. For Zn spatial dependence is seen up to 100 km, but in this study, it was found that variation in grain Se concentration is correlated up to a distance of 40 km. This indicates a finer scale of spatial variation in grain Se than in grain Zn, which would require more intensive survey effort to support local interventions. For both micronutrients, there was evidence for a relationship between measures of the crop-available concentration in the soil and the concentration in grain. However, for both, the proportion of the total variation accounted for by this model was very small, as was the proportion of the spatially dependent variation 0.03 and 0.07 respectively for Zn [12] and 0.04 and 0.05 for Se (Tables 4 and 5).

It is interesting that mean annual temperature was the selected environmental covariate for both micronutrients, suggesting that this variable, or one for which it is a proxy, influences the concentration of both micronutrients in grain. However, the relationship with grain Zn was stronger than for grain Se, the proportions of the total variance, and the spatially correlated variance accounted for by the model with mean annual temperature as a covariate were 0.09 and 0.52 for grain Zn [12], and the corresponding values were 0.03 and 0.21 for Se (Tables 4 and 5). Some of the top ranked covariates such as slope and precipitation were not selected in the final model– one underlying reason is that the variable would be rejected because it is strongly correlated with another one already in the model or was measured with substantial error. This suggests that while such variables may be important at broader scales, their mechanistic role at the scale of our study may be indirect or confounded by other landscape factors. This insight could help refine variable selection in future landscape-level modelling efforts.

In resonance with the soil properties (Tables 2 and 3) modelled (positive coefficient) as related to the grain Se concentration, their mechanistic process can be elucidated. Soil Se_{Nit} and Se_{Pho} are related to grain Se as measures of soil supply. Wang et al. (2020) demonstrated that soluble (nitrate) and adsorbed (phosphate) selenate fractions are the most likely plant-available species; hence, these would correlate strongly with grain uptake. Further, soil pH and redox together affect Se speciation in soil (Se(VI), which is more accessible, predominates over Se(IV), and becomes less accessible, as pE+pH increases (e.g., 40). Se uptake into grain tends to be greater at pH > 7 (e.g., 7). Oxides are also very important in the soil as they absorb forms of Se, while this effect is reduced by SOC (e.g., 41). On the other hand, Sulphate transporters are responsible for the uptake of Se into plants, so larger soil S will result in greater competition with Se for uptake [46]; while some molecules in SOC chelate Se forms in soil, reducing availability. This effect is reduced by increasing pH. Selenium can substitute for sulphur and therefore bond with both carbon and oxygen within humic substances. Concerning Phosphorus, Oxalate-P binds to selenium in the soil. This makes it less accessible for plants to absorb and subsequently reduces the amount available to humans through their diet. The impact varies depending on factors like soil type, pH, and the specific form of selenium involved.

5. Conclusions

This study determined the geospatial variation of Se concentration in maize grain in Malawi and analyses the effects of soil properties and landscape factors in driving the spatially correlated variation of Se concentration in maize grain. Mean annual temperature captured significant variation of Se concentration in maize grain, but substantial variation remained unexplained. However, this variation unaccounted for by the covariate showed spatial dependence, and so Se concentration could be mapped by geostatistical prediction, if grain Se has been measured on a suitable spatial sample design, to provide guidance for designing efficient interventions.

Several soil properties, including measures of available forms of Se were included in the model. For most the sign of the coefficient was consistent with known factors influencing Se availability or uptake by the crop. However, substantial variation in grain Se concentration remains unexplained (over 80%) in the final model. This suggests that the existing approaches to characterise availability of nutrients in soil, using various chemical extractions, have limited value for predicting Se uptake by crop plants. In part, this is because Se availability in soil is determined by dynamic equilibria between the soil solid phases (mineral and organic-bound forms) and soil pore water, and the multiple (and complex) uptake mechanisms by plant roots, whereas chemical extractions only provide a single time-point measurement of micronutrients availability. Measurements on extracted fractions at single time-points cannot capture the dynamics of micronutrients movement between different soil phases, including the capacity of the soil to replenish what is directly available in the soil solution following depletion by root uptake. Furthermore, considering only the Se dynamics of Se in soil leaves aside the multiple processes that drive internal transport and redistribution of Se, including its deposition in the grain during plant growth. Further research is needed to investigate dynamic soil geochemical processes and plant physiology to better inform agronomic biofortification strategies for alleviating Se deficiency in SSA populations.

Acknowledgments

This work was supported primarily by the Bill & Melinda Gates Foundation [INV-009129]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. Rothamsted Research receives strategic funding from the Biotechnology and Biological Sciences Research Council of the United Kingdom (BBSRC); Rothamsted contributors acknowledge support from the Soil to Nutrition [BBS/E/C/000I0310] and the Growing Health [BB/X010953/1] Institute Strategic Programmes. E.L.A.'s contribution is published with the permission of the Director of the British Geological Survey (UKRI).

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.

We thank the participating farmers for their engagement with this study. Field sampling teams from the Department of Agricultural Research Services and Lilongwe University of Agriculture and Natural Resources are gratefully acknowledged. Training support was facilitated by funding from the Royal Society-UK Foreign, Commonwealth & Development Office (FCDO), under project AQ140000, "Strengthening African capacity in soil geochemistry for agriculture and health". Additional mineral analytical support was provided by W. Broadley, K. Davis, P. Muleya, S. Vasquez Reina, S. Dunham, J. Carter and J. Hernandez.

The CHELSA project is acknowledged for making the downscaled climate data available from https://climatedataguide.ucar.edu/. The boundaries, denominations and any other information shown on these maps do not imply any judgment about the legal status of any territory or constitute any official endorsement or acceptance of any boundaries on the part of any Government.

Author contribution

T.A., S.G., D.G., E.L.A., E.J.M.J., S.P.M., E.T., P.C.N., R.M.L., and M.R.B. conceptualised the study; secured and administered the project funding. I.S.L., C.C., A.W.M., E.L.A., D.G., S.H., E.J.M.J., D.B.K., I.S.L., S.P.M., M.G.W., L.W., S.D.Y., E.H.B., L.B., M.M., J.G.C., A.E.M., M.R.B., R.M.L, and P.C.N. contributed to method developments; the field surveys, laboratory analyses and data analyses; data interpretation and supervision. R.M.L and C.C. conducted geostatistical analyses. D.B.K. managed the data. C.C., D.B.K., R.M.L.., A.W.M and M.G.W. developed the data visualisations. P.C.N, I.S.L, C.C, A.W.M., and R.M.L. wrote the primary draft of the paper with editing and reviewing inputs from all of the other authors.

Appendix A.

Table A1. Predictor coefficients for the null model for transformed grain selenium concentration (spatial coordinates the only fixed effects) and a model with selected soil properties as additional fixed effects

Predictand				Predictor coefficient						
		β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
		Easting	Northing	Se_{Nit}	Se_{Pho}	pH	Oxides	S_{Nit}	SOC	Oxalate
Null model	24.4138	-0.0025	-0.0023							
Model after FDR	21.1006	-0.0015	-0.0020	0.0629	0.4567	0.1882	-0.3321	-0.3104	-0.2388	0.1401

[a] β_0–β_8: fixed effects coefficients β_0 is a constant and β_i is the coefficient for the ith random effect.

Table A2. Parameter estimates for the null model for transformed grain selenium concentration (spatial coordinates the only fixed effects) and a model with selected soil properties as additional fixed effects

Predictand	R^2_{adj}	$\breve{R}^2_{\mathrm{adj}}$	κ	τ^2	σ^2	ϕ
Null model			0.5	0.8972	0.5277	10.8
Model after FDR	0.1169	0.0429	0.5	0.7534	0.5051	8.08

[a] R^2_{adj}: the difference between the variance of the correlated random effect (σ^2) for the null model and the proposed model expressed as a proportion of that variance for the null model; $\breve{R}^2_{adj}$: the difference between the variance of the correlated random effect (σ^2) for the null model and the proposed model expressed as a proportion of that variance for the null model; κ: smoothness parameter of the Matèrn correlation function; τ^2: variance of the iid random effect (nugget variance); σ^2: variance of the correlated random effect; ϕ:distance parameter of the correlation function.

Appendix B.

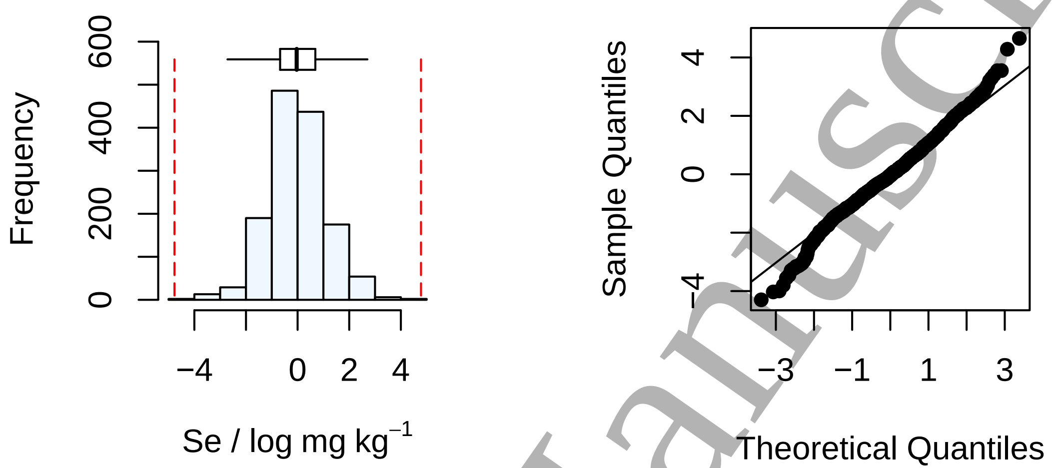

Figure B1. Histogram with boxplot and QQ plot for the residuals from an exploratory fit of the saturated model (soil properties) for grain Se concentration on a log scale.

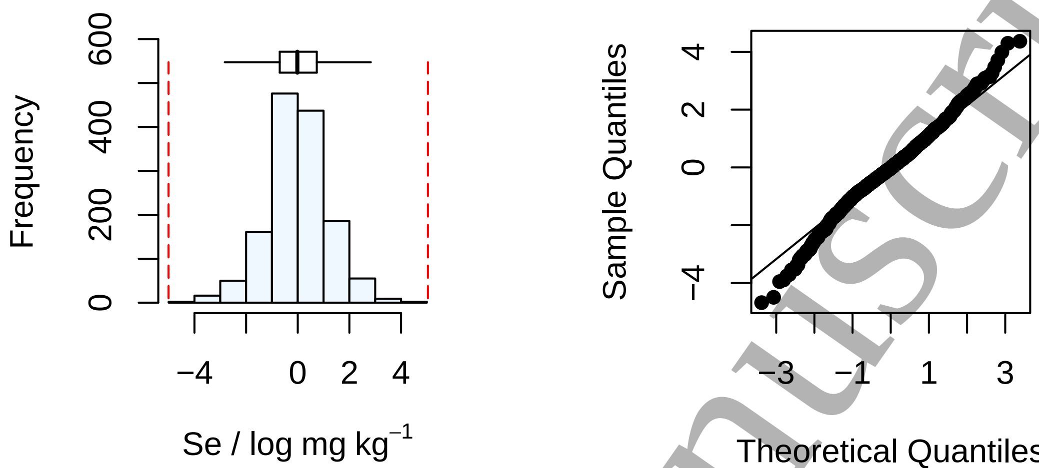

Figure B2. Histogram with boxplot and QQ plot for the residuals from an exploratory fit of the saturated model (environmental covariates) for concentration of Se in grain on a log scale.

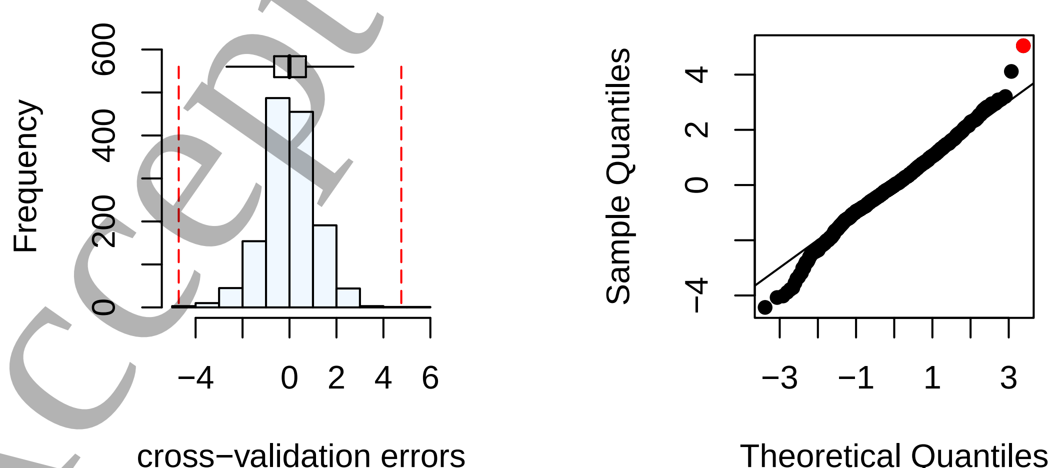

Figure B3. Histogram with boxplot and QQ plot for the cross-validation errors for the E-BLUP

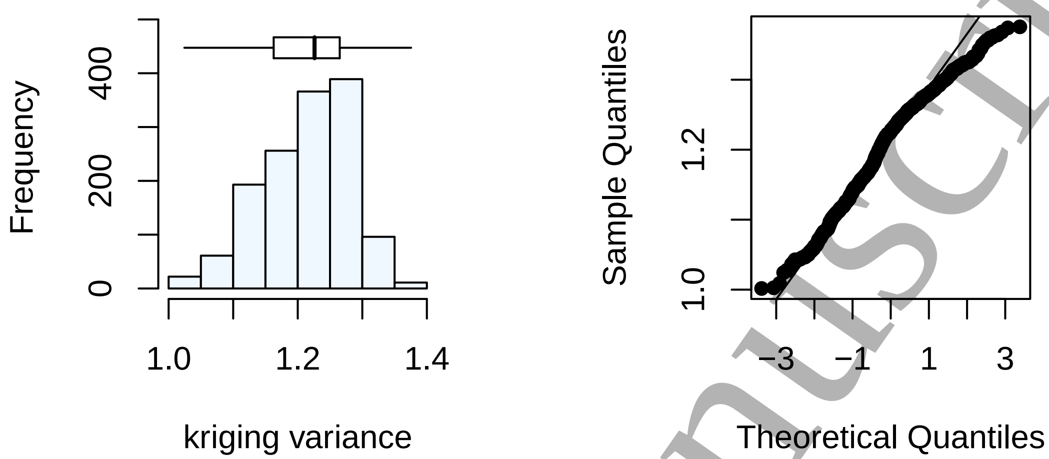

Figure B4. Histogram with boxplot and QQ plot for the kriging variances for the E-BLUP

References

[1] Fairweather-Tait, S. J., Bao, Y. P., Broadley, M. R., Collings, R., Ford, D., Hesketh, J. E., & Hurst, R. (2011) Selenium in Human Health and Disease, *Antioxidants And Redox Signaling*, **14**, 1337–1383.

[2] 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., & Broadley, M. R. (2014) Dietary mineral supplies in Africa, *Physiologia Plantarum*, **151**, 208–229, https://doi.org/10.1111/ppl.12144.

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

[4] 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., & Broadley, M. R (2019) The risk of selenium deficiency in Malawi is large and varies over multiple spatial scales, *Scientific Reports*, **9** 6566 , https://doi.org/10.1038/s41598-019-43013-z

[5] Ligowe, I., Phiri, F., Ander, E., Bailey, E., Chilimba, A., Gashu, D., Joy, E., Lark, R., Kabambe, V., Kalimbira, A. & Others Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proceedings Of The Nutrition Society*. **79**, 457-467 (2020)

[6] Belay, A., Joy, E., Chagumaira, C., Zerfu, D., Ander, E., Young, S., Bailey, E., Lark, R. & Broadley, M (2020) Selenium Deficiency Is Widespread and Spatially Dependent in Ethiopia, *Nutrients* , https://doi.org/10.3390/nu12061565.

[7] Chilimba, A. D. C., Young, S. D., Black, C. R., Rogerson, K. B., Ander, E. L., Watts, M. J., Lammel, J., & Broadley, M. R (2011) Maize grain and soil surveys reveal suboptimal dietary selenium intake is widespread in Malawi, *Scientific Reports*. **1**, https://doi.org/10.1038/srep00072

[8] Siyame, E. W. P., Hurst, R., Wawer, A. A., Young, S. D., Broadley, M. R., Chilimba, A. D. C., Ander, L. E., Watts, M. J., Chilima, B., Gondwe, J., Kang'ombe, D., Kalimbira, A., Fairweather-Tait, S. J., Bailey, K. B., & Gibson, R. S. (2013) A high prevalence of zinc–but not iron-deficiency among women in wural Malawi: a cross-sectional study, *International Journal For Vitamin And Nutrition Research*, **83**, 176–187, https://doi.org/10.1024/0300-9831/a000158

[9] 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., & Broadley, M. R. (2013) Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi, *Scientific Reports*, **3**, https://doi.org/10.1038/srep01425.

[10] 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**, https://doi.org/10.1016/j.scitotenv.2020.139231.

[11] Gashu, D., Nalivata, P.C., Amede, T., Ander, E.L., Bailey, E.H., Botoman, L., Chagumaira, C., Gameda, S., Haefele, S., Hailu, K., Joy, E.J.M., Kalimbira, A., Kumssa, D.B., Lark, R.M., Ligowe, I.S., McGrath, S.P., Milne, A.E., Mossa, A.W., Munthali, M., Towett, E., 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** (7861), 71–76, https://doi.org/10.1038/s41586-021-03559-3.

[12] 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*, https://doi.org/10.1038/s41598-

022-12014-w.
[13] Foster, D. & Stine, R (2008) α-investing: a procedure for sequential control of expected false discoveries, *Journal Of The Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 429–444, https://doi.org/10.1111/j.1467-9868.2007.00643.x.
[14] 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. S., Munthali, M., Towett, E., Walsh, M.G., Wilson, L., Young, S.D, Haji, I., Broadley, M.R., Gashu, D. & Nalivata, P.C (2022) Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in Ethiopia and Malawi. *Scientific Data*, **9**, 1–12, https://doi.org/10.1038/s41597-022-01500-5.
[15] Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S. & Bates, P.D (2017) A high-accuracy map of global terrain elevations, *Geophysical Research Letters*, **44**, 5844–5853.
[16] Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., Zimmermann, N., Linder, H. & Kessler, M (2017) Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digital Repository. *Scientific Data*. **4** (170122), https://doi.org/10.1038/sdata.2017.122
[17] Karger, D., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R., Zimmermann, N., Linder, H. & Kessler, M (2018) Data from: Climatologies at high resolution for the earth's land surface areas, Dryad 2018, http://datadryad.org/stash/dataset/doi:10.5061/dryad.kd1d4
[18] Justice, C., Vermote, E., Townshend, J., Defries, R., Roy, D., Hall, D., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Zhengming, W., Huete, A.R., van Leeuwen, W.,Wolfe, R.E., Giglio, L., Muller, J., Lewis, P. & Barnsley, M.J (1998) The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research, *IEEE Transactions On Geoscience And Remote Sensing*, **36**, 1228–1249.
[19] Lark, R. M., Cullis, B. R., & Welham, S. J. (2006) On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML *European Journal Of Soil Science*. **57**, 787-799, https://doi.org/10.1111/j.1365-2389.2005.00768.x
[20] Benjamini, Y. & Hochberg, Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing *Journal Of The Royal Statistical Society: Series B (Methodological)*, **57**, 289–300 (1995), https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x.
[21] Benjamini, Y. & Yekutieli, D (2001) The Control of the False Discovery Rate in Multiple Testing under Dependency, *The Annals Of Statistics*, **29**, 1165–1188, https://www.jstor.org/stable/2674075
[22] Tukey, J. (1991) The Philosophy of Multiple Comparisons, *Statistical Science*, **6**, 100-116.
[23] 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**, 221–234.
[24] Webster, R. & Oliver, M (2007) *Geostatistics for Natural Environmental Scientists*, John Wiley & Sons Chichester.
[25] Brys, G., Hubert, M. & Struyf, A (2004) A Robust Measure of Skewness *Journal Of Computational And Graphical Statistics* **13**(4), 996–1017, https://doi.org/10.1198/106186004X12632.
[26] Diggle, P. & Ribeiro, P (2010) *Model-based geostatistics*, Springer Science+Business Media LLC.
[27] R Core Team R (2023) *A language and environment for statistical computing*, R Foundation for Statistical Computing, https://www.r-project.org/.
[28] Ribeiro Jr, P., Diggle, P., Christensen, O., Schlather, M., Bivand, R. & Ripley, B (2022) *geoR: Analysis of Geostatistical Data*, https://CRAN.R-project.org/package=geoR, R package version 1.9-2
[29] 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

[30] Chagumaira, C., Chimungu, J.G., Gashu, D., Nalivata, P.C., Broadley, M.R., Milne, A.E. & Lark, R.M. (2021) Communicating uncertainties in spatial predictions of grain micronutrient concentration *Geoscience Communication*, **4**(2), 245–265.

[31] Chagumaira, C., Nalivata, P.N., Chimungu, J.G., Gashu, D., Broadley, M.R., Milne, A.E. & Lark, M.R (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*, **36**(12), 2446–2472, https://doi.org/10.1080/13658816.2021.2020278

[32] Stroud, J.L., Li, H.F., Lopez-Bellido, F.J., Broadley, M.R., Foot, I., Fairweather-Tait, S.J., Hart, D.J., Hurst, R., Knott, P., Mowat, H., Norman, K., Scott, P., Tucker, M., White, P.J, McGrath, S.P. & Zhao, F.J. (2010), Impact of sulphur fertilisation on crop response to selenium fertilisation. *Plant And Soil*, **332**, 31–40

[33] White, P.J., Bowen, H.C., Parmaguru, P., Fritz, M., Spracklen, W.P., Spiby, R.E., Meacham, M.C., Mead, A., Harriman, M., Trueman, L.J., Smith, B.M., Thomas, B. & Broadley, M.R (2004) Interactions between selenium and sulphur nutrition in *Arabidopsis thaliana*, *Journal Of Experimental Botany*. **55**, 1927–1937

[34] Liu, X., Zhao, Z., Duan, B., Hu, C., Zhao, X. & Guo, Z (2015), Effect of applied sulphur on the uptake by wheat of selenium applied as selenite, *Plant And Soil*, **386**, 35–45

[35] Jiang, T., Yu, T., Qi, H., Li, F. & Yang, Z (2022) Analysis of phosphorus and sulfur effect on soil selenium bioavailability based on diffusive gradients in thin films technique and sequential extraction, *Chemosphere*. **302** pp. 134831, https://doi.org/10.1016/j.chemosphere.2022.134831.

[36] Mastrandrea, M., Field, T., Edenhofer, O., Ebi, K., Frame, D., Held, H., Kriegler, E., Mach, K., Matschoss, P., Plattner & Zwiers, F. (2010) *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*, Intergovernmental Panel on Climate Change (IPCC), https://www.ipcc.ch/

[37] Lark, R.M., Ander, E.L, Cave, M., Knights, K., Glennon, M. & Scanlon, R (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

[38] 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. & Broadley, M.R Urine selenium concentration is a useful biomarker for assessing population level selenium status. *Environment International*, **134**, https://10.1016/j.envint.2019.105218

[39] Stefanowicz, F.A., Talwar, D., O'Reilly, D.S., Dickinson, N., Atkinson, J., Hursthouse, A., Rankin, J. & Duncan, A (2013) Erythrocyte selenium concentration as a marker of selenium status, *Clinical Nutrition*, **32**, 837–842, https://www.sciencedirect.com/science/article/pii/S0261561413000290

[40] Somagattu, P., Chinnannan, K., Yammanuru, H., Reddy, U. & Nimmakayala, P. Selenium dynamics in plants: Uptake, transport, toxicity, and sustainable management strategies. *Science Of The Total Environment*. **949** pp. 175033 (2024)

[41] Rong, Q., Chen, J., Zhang, Y., Tan, Z., Wang, W., Sun, C., Guo, X., Zhou, C., Cai, H. & Zhao, X. (2024) The interaction between selenium and other elements in soil and rice roots shaped by straw and straw biochar regulated the enrichment of selenium in rice grain. *Frontiers in Plant Science*. **15**, https://doi.org/10.3389/fpls.2024.1387460

[42] Mollier, A. & Pellerin, S. Maize root system growth and development as influenced by phosphorus deficiency. (1999) *Journal Of Experimental Botany*. **50**, 487-497

[43] Sun, Y., Zhang, J., Li, W., Xu, Z., Wang, S., Zhao, M., Shen, J. & Cheng, L. (2024) Regulation of maize root growth by local phosphorus availability, sucrose metabolism, and partitioning. *Annals Of Botany*. pp. mcae169

[44] Wang, J., Bailey, E., Sanders, H., Izquierdo, M., Crout, N., Shaw, G., Yang, L., Li, H., Wei, B. & Young, S. (2020) Using chemical fractionation and speciation to describe uptake of technetium, iodine and selenium by Agrostis capillaris and Lolium perenne. *Journal Of Environmental*

Radioactivity. **212** pp. 106131

[45] Liu, H., Shi, Z., Li, J., Zhao, P., Qin, S. & Nie, Z. (2018) The impact of phosphorus supply on selenium uptake during hydroponics experiment of winter wheat (Triticum aestivum) in China. *Frontiers In Plant Science*. **9** pp. 373

[46] Shinmachi, F., Buchner, P., Stroud, J., Parmar, S., Zhao, F., McGrath, S. & Hawkesford, M. (2010) Influence of sulfur deficiency on the expression of specific sulfate transporters and the distribution of sulfur, selenium, and molybdenum in wheat. *Plant Physiology*. **153**, 327–336