



OPEN

DATA DESCRIPTOR

Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in Ethiopia and Malawi

D. B. Kumssa^{1,14}, A. W. Mossa^{1,14}, T. Amede², E. L. Ander³, E. H. Bailey¹, L. Botoman^{4,5}, C. Chagumaira^{1,4,6,7}, J. G. Chimungu⁴, K. Davis¹, S. Gameda⁸, S. M. Haefele⁷, K. Hailu^{9,10}, E. J. M. Joy¹¹, R. M. Lark^{1,6}, I. S. Ligowe^{4,5}, S. P. McGrath⁷, A. Milne⁷, P. Muleya¹, M. Munthali⁵, E. Towett¹², M. G. Walsh¹³, L. Wilson¹, S. D. Young¹, I. R. Haji¹, M. R. Broadley^{1,7,15} & D. Gashu^{9,15} & P. C. Nalivata^{4,15}

The dataset comprises primary data for the concentration of 29 mineral micronutrients in cereal grains and up to 84 soil chemistry properties from GeoNutrition project surveys in Ethiopia and Malawi. The work provided insights on geospatial variation in the micronutrient concentration in staple crops, and the potential influencing soil factors. In Ethiopia, sampling was conducted in Amhara, Oromia, and Tigray regions, during the late-2017 and late-2018 harvest seasons. In Malawi, national-scale sampling was conducted during the April–June 2018 harvest season. The concentrations of micronutrients in grain were measured using inductively coupled plasma mass spectrometry (ICP-MS). Soil chemistry properties reported include soil pH; total soil nitrogen; total soil carbon (C); soil organic C; effective cation exchange capacity and exchangeable cations; a three-step sequential extraction scheme for the fractionation of sulfur and selenium; available phosphate; diethylenetriaminepentaacetic acid (DTPA)-extractable trace elements; extractable trace elements using 0.01 M Ca(NO₃)₂ and 0.01 M CaCl₂; and isotopically exchangeable Zn. These data are reported here according to FAIR data principles to enable users to further explore agriculture-nutrition linkages.

Background & Summary

Micronutrients are the vitamins and minerals that our bodies require in small amounts, and which are obtained from the food we eat. Micronutrient deficiencies (MNDs) among people remain a major global concern; more than 2 billion people are likely to be affected worldwide, with greater deficiency risks in sub-Saharan Africa than in most other regions¹. The risk of MNDs in populations can be informed by understanding the supply of micronutrients within food systems. This approach has typically been conducted at a national level, based on

¹School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK.

²International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), ILRI Sholla Campus, P.O. Box 5689, Addis Ababa, Ethiopia.

³Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottinghamshire, NG12 5GG, UK.

⁴Lilongwe University of Agriculture and Natural Resources (LUANAR), Bunda College, P.O. Box 219, Lilongwe, Malawi.

⁵The Department of Agricultural Research Services, P.O. Box 30779, Lilongwe, Malawi.

⁶Future Food Beacon, University of Nottingham, Sutton Bonington Campus, Nottinghamshire, LE12 5RD, UK.

⁷Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK.

⁸International Maize and Wheat Improvement Centre (CIMMYT), ILRI Sholla Campus, P.O. Box 5689, Addis Ababa, Ethiopia.

⁹Centre for Food Science and Nutrition, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia.

¹⁰Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.

¹¹Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK.

¹²World Agroforestry (ICRAF), United Nations Avenue, P.O. Box 30677, Nairobi, Kenya.

¹³Africa Soil Information Service, Selian Agricultural Research Institute, P.O. Box 2704, Arusha, Tanzania.

¹⁴These authors contributed equally: Kumssa DB, Mossa AW.

¹⁵These authors contributed equally: Broadley MR, Gashu D, Nalivata PC. ✉e-mail: martin.broadley@nottingham.ac.uk

Cereal grain	Number of samples	
	Ethiopia	Malawi
Barley	175	0
Finger millet	37	1
Maize	290	1,608
Pearl millet	1	32
Rice	8	54
Sorghum	135	117
Teff	362	0
Triticale	19	0
Wheat	325	0
Total grain-soil pairs	1,352	1,812

Table 1. Type and number of cereal grain samples collected from Ethiopia and Malawi. The total row indicates the total number of cereal grain-soil sample pairs sampled in each country.

Data field name	Element	Data field name	Element
Ag_grain	Silver	Mg_grain	Magnesium
Al_grain	Aluminium	Mn_grain	Manganese
As_grain	Arsenic	Mo_grain	Molybdenum
B_grain	Boron	Ni_grain	Nickel
Ba_grain	Barium	P_grain	Phosphorus
Be_grain	Beryllium	Pb_grain	Lead
Ca_grain	Calcium	Rb_grain	Rubidium
Cd_grain	Cadmium	S_grain	Sulfur
Co_grain	Cobalt	Se_grain	Selenium
Cr_grain	Chromium	Sr_grain	Strontium
Cs_grain	Caesium	Tl_grain	Thallium
Cu_grain	Copper	U_grain	Uranium
Fe_grain	Iron	V_grain	Vanadium
K_grain	Potassium	Zn_grain	Zinc
Li_grain	Lithium		

Table 2. Elemental concentrations in cereal grains reported from Ethiopia and Malawi. All concentrations are in mg kg⁻¹ on dry matter basis.

secondary interpretation of food consumption, expenditure or supply data from household surveys and food balance sheets^{1–5}.

Recent studies from Ethiopia and Malawi have reported substantial variation in the micronutrient concentration of the grains of staple cereal crops at subnational levels, including for calcium (Ca), iron (Fe), selenium (Se) and zinc (Zn)^{6–8}. Some of this variation is spatially correlated at distances of up to several hundred kilometres. What this means is that for people consuming food sourced locally, as is the case for many smallholder farming communities, the location of residence will be a major (sometimes the largest) influencing factor in determining the dietary intake of micronutrients from cereals⁷. Furthermore, for the micronutrient Se, there is strong evidence of linkages between soil and landscape features, cereal grain concentrations, and biomarkers of Se status in people^{9,10}.

Here we report the wider set of primary data for cereal grains and soils from these studies in Ethiopia and Malawi (Tables 1 and 2), with a focus on the data reported in Gashu *et al.*^{6,7}, Mossa *et al.*¹¹, and Botoman *et al.*¹². These data were obtained as part of ongoing work within two ‘GeoNutrition’ projects, funded primarily by the Bill & Melinda Gates Foundation (BMGF) and the UK Government’s Global Challenges Research Fund (GCRF). Soil and grain samples were collected in both countries using spatially balanced sampling designs, along with meta-data, with the informed consent of farmers.

Methods

Research design. *Ethical approval.* The research work that generated these data involved metadata collection using semi-structured questionnaires, and sampling of cereal grain from farmers’ fields or grain stores and soils from the corresponding crop fields with the prior informed consent of the farmers. Farmers who participated in the survey received an information sheet (see Supplementary files 1 and 2) explaining the details of the project, what participation would involve, and how their data would be used, including its eventual release. The work was conducted under ethical approvals from the University of Nottingham, School of Sociology and Social Policy Research Ethics Committee (REC); BIO-1819-001 and BIO-1718-0004 for Ethiopia and Malawi, respectively. These REC approvals were recognized formally by the Directors of Research at Addis Ababa University

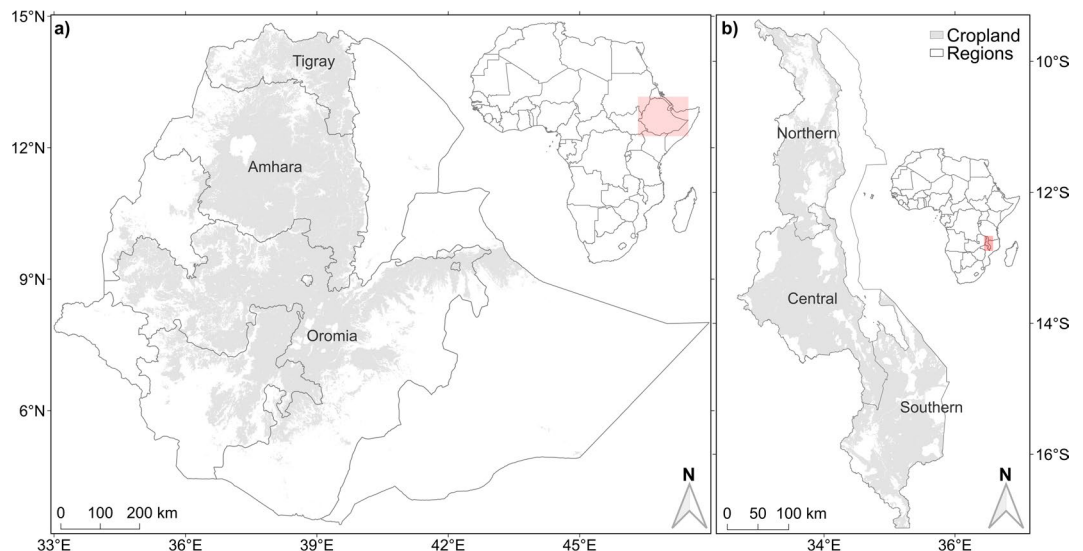


Fig. 1 GeoNutrition cereal grain and soil sampling constrained to areas identified as croplands (shaded grey) in (a) the Amhara, Oromia and Tigray regions of Ethiopia, and (b) Malawi⁷. The (a) Ethiopian and (b) Malawian map insets in the African continental map are shaded pink.

(Ethiopia) and Lilongwe University of Agriculture and Natural Resources (Malawi), who also reviewed the study protocols.

Sampling design. The sampling design is described in full by Gashu *et al.*⁷. The objective of the sampling was to support the evaluation of relationships between crop and soil properties, and spatial mapping of these. For this reason, the basic sample design was selected to achieve spatial coverage of the agreed sample frame. A random subset of the spatial coverage sample was then selected, and an additional close paired sample site was specified for each of these, to support statistical modelling of spatial variation.

Ethiopia's Amhara, Oromia and Tigray regions were sampled (Fig. 1). Target sample frames in Ethiopia were constrained to locations on a 500-m grid (Lambert azimuthal equal-area projection) at which the probability of the land being under crop production had been mapped as ≥ 0.9 ¹³. The sample frame was further constrained to include only those locations on the 500-m grid that fell within 2.5 km of a road available in digital mapping format on OpenStreetMap (OSM)¹⁴. These constraints may introduce possible biases into predictions made at locations outside the designed sample frame, however, it would not otherwise have been possible to visit all the sample locations in the time available. Of a total land area of around 558,500 km² in the three regions of Ethiopia, the total cropland mask represented 354,325 km², of which 220,467 km² was within 2.5 km of a OSM mapped road¹³. Because the sample frame was somewhat fragmented spatially the sample points were selected to achieve spatial balance and spread, the latter denoting spatial coverage¹⁵. A total of 1,825 sample sites were selected this way, and 175 of these were selected at random to be supplemented with a close pair site.

In Malawi, the cropland area was determined from the European Space Agency Climate Change Initiative¹⁶ land-use maps. The agricultural area used was defined as all raster cells that included the category of 'cropland' in their description (Fig. 1). In Malawi, where road access to cropped areas is generally better than in Ethiopia, no constraint to distance from a road was imposed on sample locations. A total of 820 sample sites were selected from the 2015/16 Demographic and Health Survey (DHS) of Malawi (REFS)¹⁷, all the DHS points in the sample frame. A further 890 sites were then selected to complete a spatial coverage survey, with the stratify function from the *spcosa* library in R¹⁸ which can draw a spatial coverage survey conditional on fixed points. An additional 190 close-pair sites were then added to a random subset of the spatial coverage points.

Field data collection. Field data collection was carried out using an open-source georeferenced survey data collection tool, KoBoToolBox (<https://www.kobotoolbox.org/>) and its companion mobile app KoBoCollect via questionnaires (see Supplementary file 3). Data recording, and cereal grain and soil sampling was conducted by teams of enumerators who were trained in standard operating procedures, and participatory risk assessments to safely navigate to target sampling sites. Sample teams used their discretion to exclude sample sites that may have jeopardised their safety (for example, via flooded roads), or created a disproportionate detour for single sample collection in mountainous regions. Field data collected include longitude, latitude, altitude, cereal crop species and source of cereal grain (i.e., standing crop, field stack or store).

Cereal grain and soil sampling. Sampling of cereal grain and soil (Table 1) from farmers' fields in Ethiopia was completed in November 2017–February 2018 for most of the Amhara region and November 2018–February 2019 in all three regions (Fig. 1). Sampling in Malawi was completed in April–June 2018.

At each preselected target sample site, the team would identify the nearest field with a mature cereal crop within a 1-km radius, and take a grain and soil sample, subject to farmer consent. If a field with a standing mature cereal crop was not apparent, that is, the crop had been harvested, or a non-cereal crop had been grown, the team would ask the farmer to identify a field from which a cereal crop had recently been harvested and stored, and from which a sample could be obtained. If sampling was not possible, then the team would either look beyond a 1-km radius for an alternative site, or leave the site without taking samples.

Within a selected field, samples were taken from a 100 m² (0.01 ha) circular plot. This was centred as close as practical to the middle of the field unless this area was unrepresentative due to disease or crop damage. Five subsample points were located (see Extended Data Fig. 1 in Gashu *et al.*⁷). The first point was at the centre of the plot. Two subsample points were then selected at locations on a line through the plot centre along the crop rows, and two more points on a line orthogonal to the first through the plot centre. Where possible, the central sampling location was fixed between crop rows, and the long axis of the sample array (with sample locations at 5.64 m and 4.89 m) was oriented in the direction of crop rows with the short axis perpendicular to the crop rows. A single soil subsample was collected at each of the five subsample points with a Dutch auger with a flight length of 0.15 m and diameter of 0.05 m. The auger was inserted vertically to the depth of one flight and the five subsamples were combined in a single Kraft® paper bag. Where a mature or ripe crop was still standing in the field, grain samples were taken close to each augering position by a different operator, to minimise further contamination by dust and soil. For maize, a single cob was taken at each of the five points. Maize kernels were stripped from around 50% of each cob lengthways and composited into a single sample envelope for each location. For small-grained crops, sufficient stalks were taken so that approximately 20–50% of the sample envelope was filled (dimensions 0.15 m × 0.22 m), with samples placed grain-first into the sample bag and the stalks were twisted off the grain heads and discarded. If a crop was in field stacks, then a subsample, comprising five cobs for maize, or a representative sample for other crops was taken from each available stack, taking material from inside the stack to minimize contamination by dust and soil (see Extended Data Fig. 1 in Gashu *et al.*⁷). If a crop was in a farmer's store, it was considered an already-composited sample and a sample was taken while avoiding grain from the store floor if grain was loosely stored and avoiding grain with visible soil or dust contamination.

Sample management and preparation. Whole-grain samples were air-dried in their sample bags. Each sample was then ground in a domestic stainless-steel coffee grinder, which was wiped clean before use and after each sample with a non-abrasive cloth. All preparation was done away from sources of contamination by soil or by dust. A 20-g subsample of the ground material was then shipped to the University of Nottingham. Soil samples were oven-dried at 40 °C for 24–48 h depending on the moisture content of the soil. Preparation took place in a soil laboratory to avoid cross-contamination with grain samples. Plant material was removed from each soil sample, which was then disaggregated and sieved to pass 2 mm. This material was then coned and quartered to produce subsample splits. A 150-g subsample of soil was poured into a self-seal bag, labelled and shipped to the UK for analysis in the laboratories at Rothamsted Research and the University of Nottingham.

Chemical analysis methods. All grain and soil analyses were conducted in numerical sequence of the sample ID at Rothamsted Research and the University of Nottingham, UK.

Grain analysis. Elemental concentrations (see Figs. 2 and 3 and Supplementary file 4) of grain were determined after microwave digestion of approximately 0.2 g of ground samples with concentrated nitric acid (70% HNO₃, trace analysis grade). Samples collected from the Amhara Region of Ethiopia in 2017 were microwave digested using a Multiwave 3000 48-vessel MF50 rotor (Anton Paar GmbH, Graz, Austria) in 2 mL HNO₃, 1 mL Milli-Q water (18.2 MΩ cm; Fisher Scientific) and 1 mL H₂O₂ at a power of 1400 W, temperature 140 °C, pressure of 2 MPa for 45 min. Samples collected in Malawi and Ethiopia in 2018–2019 were microwave digested in a Multiwave Pro with a 41HVT56 rotor and pressure-activated venting vessels made of modified polytetrafluoroethylene (56-ml 'SMART VENT', Anton Paar). The digestion was achieved using 6 mL of HNO₃, at a power of 1,500 W, with 10 min heating to 140 °C, 20 min holding at 140 °C, and 15 min cooling to 55 °C. After digestion, the samples were made to 15 mL using Milli-Q water, then stored in capped tube at room temperature for ≈ 1 week until the chemical analysis. Prior to analysis by inductively coupled plasma mass spectrometry (ICP-MS; Thermo Fisher Scientific iCAP Q, Thermo Fisher Scientific, Bremen, Germany), samples were further diluted 1:5 with Milli-Q water.

Due to low concentrations of Se in many of the grain samples taken in Malawi, a substantial number of values were below the limit of detection (LOD) when measured at a mass to charge ratio of 78 ($m/z = 78$) on ICP-MS. Consequently, samples collected in Malawi and Ethiopia were re-analysed for Se using QQQ-ICP-MS (iCAP TQ; Thermo Fisher Scientific, Bremen, Germany) with oxygen mass shifting of the Se peak at $m/z = 80$ to $m/z = 96$.

Soil chemical analysis. *Soil characterisation.* Unless otherwise specified, all analysis were performed on samples sieved to < 2 mm. Soil pH was measured in de-ionised water (pH_w) suspension (1:2.5 solid to solution ratio) using a Jenway 3540 meter (Cole-Parmer, Stone, Staffordshire, UK), with a temperature-compensated combination pH electrode. Soil pH was also measured in the 0.01 M Ca(NO₃)₂ (pH_{CaNO}) suspension using a Mettler-Toledo AG pH meter (Mettler-Toledo, Beaumont Leys, Leicester, UK) also with a temperature compensating electrode. Total C and N were determined by dry combustion¹⁹ using a Leco TruMac CN Combustion analyzer (LECO Corporation, St. Joseph, Michigan, USA). Inorganic C was determined by Skalar Primacs Inorganic Carbon Analyser (Skalar Analytical BV, Breda, Netherlands). Estimates of amorphous oxides and poorly crystalline oxides (E_{Oxa}) were determined following ammonium oxalate extraction²⁰. Soil effective cation exchange capacity (eCEC) and exchangeable cations (Na_{Exch}, Mg_{Exch}, K_{Exch}, and Ca_{Exch}) were determined using one-step extraction²¹ with cobalt(III) hexamine chloride solution and analysis by inductively coupled plasma

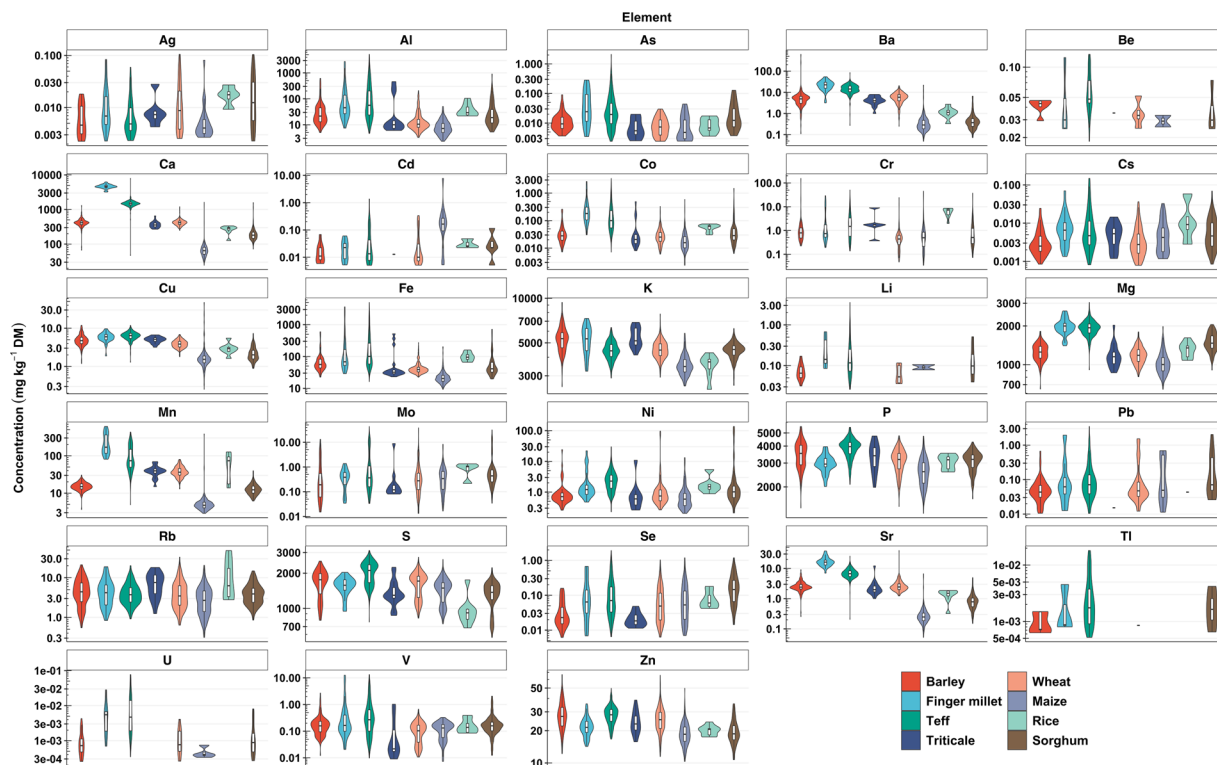


Fig. 2 Combined violin and box-and-whisker plots of the elemental concentration in barley, finger millet, teff, triticale, wheat, maize, rice, and sorghum grains collected from Ethiopia. The middle line in the box represents the median, lower hinge Q1 and upper hinge Q3 of the quartiles, and the ends of the whiskers indicate the highest and lowest concentration values. The y-axis is shown as a logarithmic scale. See Table 1 for the number of samples for each crop which are for those samples greater than the LOD for each analyte. See Table 2 for the names of the elements.

optical emission spectrometry (ICP-OES; Perkin Elmer Life and Analytical, Shelton, USA). Available phosphorus (P_{Olsen}) was determined after extraction with sodium bicarbonate as described by Olsen²². Phosphate buffering index (PBI) was also determined, as an indicator of the soil's ability to control changes in P concentration in the soil solution²³.

Total elemental concentration. Quasi-total concentrations of major and trace elements (E_{Tot}) in the soils were determined after aqua regia extraction²⁴ of finely ground samples using ICP-OES (ICP-OES; PerkinElmer Life and Analytical, Shelton, Connecticut, USA) and ICP-MS.

Extractable elemental concentration. Elements extractable with DTPA (E_{DTPA} ; potentially phytoavailable) were determined by shaking c. 5 g soil with 10 mL of 0.005 M DTPA, 0.1 M triethanolamine (TEA) and 0.01 M CaCl_2 at pH = 7.3 for 2 h on an end-over-end shaker²⁵, followed by centrifugation (3500 rpm), filtration (0.22 μm) and analysis by ICP-MS (iCAP Q; Thermo Fisher Scientific, Bremen, Germany). Soluble major and trace elements ($E_{\text{Sol,Ca}}$; readily available) were determined in the solution phase of soil suspensions in 0.01 M $\text{Ca}(\text{NO}_3)_2$ (1:10 soil: solution ratio) following equilibration for 4 days on an end-over-end shaker. The solutions were isolated by centrifugation and filtration (0.22 μm) prior to elemental analysis by ICP-MS (iCAP Q). For soil samples collected from the Amhara Region of Ethiopia in 2017, soluble major and trace elements (E_{CaCl_2}) were also determined in the solution phase of soil suspension in 0.01 M CaCl_2 (1:10 soil:solution ratio) and equilibration for 2 hours on end-over-end shaker followed by centrifuge and filtration prior to analysis by (ICP-OES; Perkin Elmer Life and Analytical, Shelton, USA). Non-purgeable organic carbon (NPOC) was also determined in 0.01 M CaCl_2 extraction using Chemical UV Oxidation Total Organic Carbon Analyser (Shimadzu Corporation, Japan). Concentrations of the free ion activity of Zn were predicted for these samples (Amhara Region of Ethiopia) using the Windermere Humic Aqueous (WHAM) geochemical model as described in detail in Mossa *et al.*¹¹.

A three-step sequential extraction scheme of sulfur and selenium in soil. This fractionation scheme was adapted from the one used by Mathers *et al.*²⁶ for soil Se and Shetaya *et al.*²⁷ for soil iodine and is intended as a general scheme for extraction of oxy-acid anions. The method aims to sequentially extract (i) a 'soluble' fraction in 0.01 M KNO_3 , (ii) a 'specifically adsorbed' fraction in 0.016 M KH_2PO_4 , and (iii) an organically bound fraction in 10% tetra methyl ammonium hydroxide (TMAH). It is important to note that none of the three fractions are likely to contain the analyte as a single species. For example, the 'soluble' Se fraction will typically include

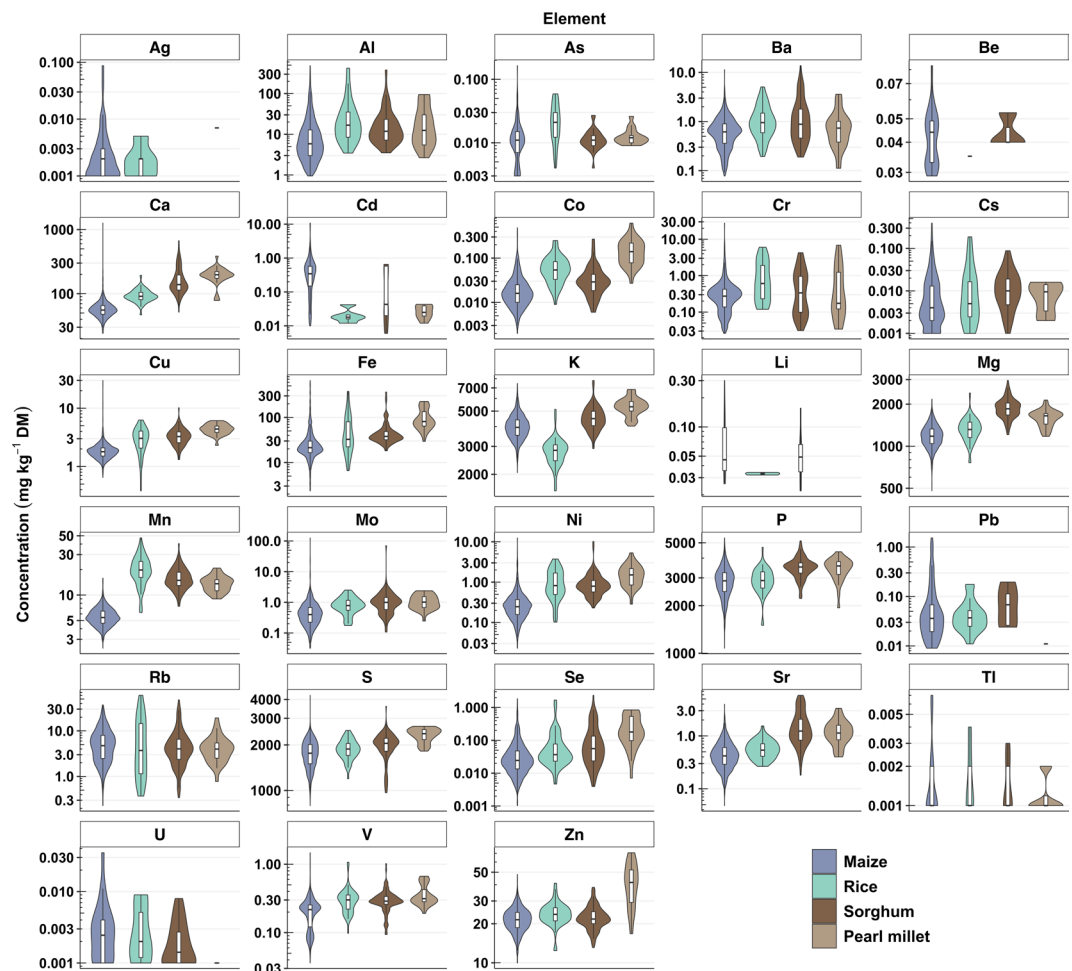


Fig. 3 Combined violin, and box-and-whisker plots of the elemental concentration in maize, rice, sorghum and pearl millet grains collected from Malawi. The middle line in the box represents the median, lower hinge Q1 and upper hinge Q3 of the quartiles, and the ends of the whiskers indicate the highest and lowest concentration values. The y-axis is shown as a logarithmic scale. See Table 1 for the number of samples for each crop which are those samples greater than the LOD for each analyte. See Table 2 for the names of the elements.

selenate, selenite and dissolved organic forms of Se. A mass of dried soil equivalent to ≈ 4.0 g, sieved to < 2 mm, was weighed into a polyethylene centrifuge tube. After adding 20 mL of 0.01 M KNO_3 , the tubes were shaken for 2 h on an end-over-end shaker, then centrifuged at 3500 rpm for 30 min. A volume of 9 mL of supernatant was filtered ($0.22 \mu\text{m}$) using PTFE syringe filters, into tubes containing 1 mL of a mixture of 0.1 M KH_2PO_4 and 10% TMAH to preserve the samples before analysis. After removing the excess supernatant, the centrifuge tubes with wet soil pellets were weighed to account for carry over of 0.01 M KNO_3 extract then 20 mL of 0.016 M KH_2PO_4 was added. The tubes were vortexed to disaggregate the soil pellet and then shaken for 1 h before centrifugation at 3500 rpm for 30 min. A volume of 9 mL of the supernatant was filtered to ($0.22 \mu\text{m}$), into a tube containing 1 mL of 10% TMAH. After removing excess supernatant, the tubes were weighed again before 10 mL of 10% TMAH was added. The tubes were vortexed to disaggregate the pellet, loosely capped and incubated at 70°C for ~ 16 h before centrifugation (3500 rpm for 30 min). Extracts (1 mL) were then diluted with 9 mL of ultrapure MQ water to give a final solution of 1% TMAH. Samples were analysed for S and Se using a QQQ-ICP-MS operated in oxygen cell mode with rhenium (^{187}Re ; $20 \mu\text{g L}^{-1}$) and indium (^{115}In ; $10 \mu\text{g L}^{-1}$) as internal standards to correct for instrumental drift. Sulfur and Se were measured in mass-shift mode following reaction with oxygen to form the analyte ions SO^+ (m/z 32 \rightarrow 48), and SeO^+ (m/z 80 \rightarrow 96).

Zinc isotopic dilution assay. Isotopically exchangeable Zn was determined using the method described in detail in Mossa *et al.*¹¹. Briefly, a mass of 2 g sieved, air-dried soil was equilibrated with 20 mL of 0.01 M $\text{Ca}(\text{NO}_3)_2$ for 24 h. The soil suspension was then spiked with a ^{70}Zn isotopic tracer and further equilibrated for 72 h. To avoid acidification, the pH of the spike solution was adjusted to pH 4.0–4.5 using an ammonium acetate buffer immediately before use. Samples were centrifuged (3500 rpm for 15 minutes), filtered ($0.22 \mu\text{m}$) and the supernatant acidified to 2% HNO_3 . Isotopic analysis was carried out using ICP-MS (iCAP Q) operating in collision cell mode using He for kinetic energy discrimination (KED). Significant and variable interferences (soil derived $^{70}\text{Ge}^+$ and (plasma generated) doubly charged $^{140}\text{Ce}^{++}$) on ^{70}Zn required correction and was achieved by analysing Ge and

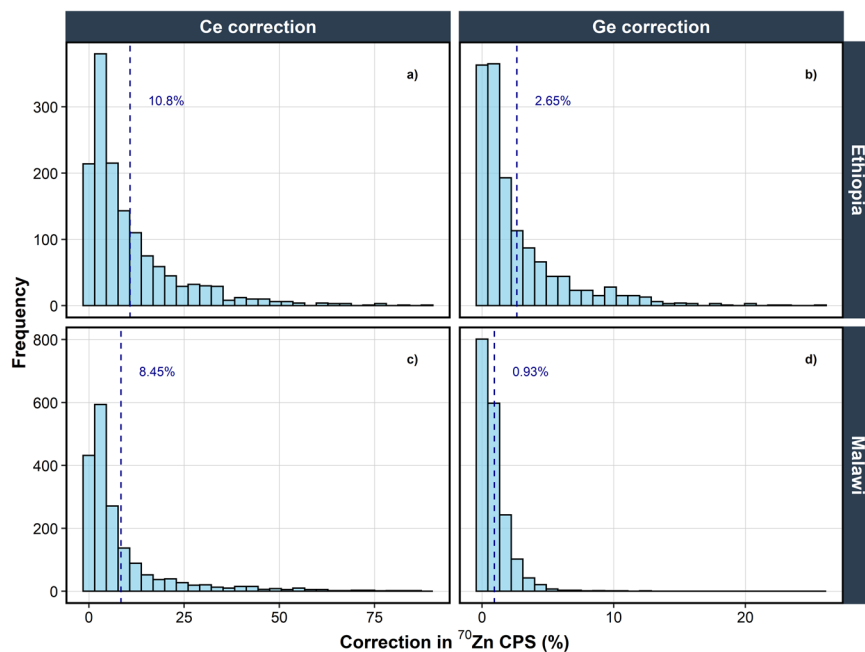


Fig. 4 Histograms showing the percentage correction in ^{70}Zn counts per second (CPS) resulting from cerium (Ce) interferences in samples from (a) Ethiopia and (c) Malawi, and percentage correction resulting from germanium (Ge) interferences in samples from (b) Ethiopia and (d) Malawi. Vertical blue dashed lines represent mean values.

Ce standards alongside the samples and inferring the intensity (count per second, CPS) from the measured CPS ratio 72/70 for Ge standards and 70/140 for Ce standards. The interference from Ge produced a correction for ^{70}Zn CPS ranging from 0.01–25% (median = 0.74%; mean = 1.68%), while the correction resulted from Ce interference ranged from 0.03–88% (median = 4.63%; mean = 9.48%) (Fig. 4).

Data Records

The cereal grain elemental concentration and soil chemistry properties dataset is provided as a set of OpenDocument workbooks and zipped folders containing comma separated value (CSV) files of the worksheets for Ethiopia and for Malawi. The Ethiopian workbook contain six worksheets: ETH_CropSoilData_Raw, ETH_CropSoilData_NA, ETH_Crop_LOD_ByICPRun, CropElements, SoilProperties and Notes. The worksheets in the Malawian workbook are MWI_CropSoilData_Raw, MWI_CropSoilData_NA, Crop_LOD_ByICPRun, CropElements, SoilProperties, and Notes. The individual CSV filename is the same as the worksheet name in the workbooks. The ‘Notes’ worksheets in both Ethiopian and Malawian workbooks and CSV zipped folders provide greater detail about the data structure and contents of each worksheet and fields. It also provides details of how data from different worksheets may be linked using the various identifier fields (e.g., how the LOD for a given cereal crop elemental concentration data record can be linked using the Crop_ICP_Run field). The dataset is accessible currently from the figshare data repository²⁸ at <https://doi.org/10.6084/m9.figshare.15911973>. Figshare uses MD5 checksums when storing a file, which are checked against the file regularly to ensure the file is intact and to verify the integrity of downloads. See Usage Notes before data re-use.

Grain and soil chemistry data. Workbook and spreadsheet names containing these data begin with the three-letter ISO country code for Ethiopia (ETH) and Malawi (MWI), as well as details about what is being stored. The workbook and CSV folders containing the cereal grain and soil properties data for Ethiopia is named **ETH_CropSoilChemData** and for Malawi **MWI_CropSoilChemData**. Due to the complexity in data recording and reporting, we have reported the data in two ways²⁸. In each workbook, the first worksheet (**Country_CropSoilData_Raw**) contains raw data recorded from the analytical equipment and software used to process the data. To avoid blank cells, “NM” (not measured) is used here to indicate where data were not measured. In the second worksheet (**Country_CropSoilData_NA**) data below the LOD, including negative data values, and missing data are replaced by NA (not available).

Ethiopian data. The first 12 fields in the raw and cleaned Ethiopian cereal grain and soil chemistry data worksheets are auxiliary and field data for the cereal grain-soil sample pairs (records). These are described as follows:

1. **FundingSource:** The source of funding to carry out the research that generated these data. **BMGF** = Bill & Melinda Gates Foundation (INV-009129); **GCRF** = Global Challenges Research Fund (BB/P023126/1).
2. **ID:** Data record (row) identification number (ID) for the dataset. These are unique IDs that can be used as a primary key to conduct separate analyses on cereal grain and soil chemistry properties data.

3. **Crop_ICP_Run**: The ICP-MS (Inductively Coupled Plasma Mass Spectrometer) run number to determine the crop elemental LOD for the raw cereal grain dataset using the LOD data in the *ETH_Crop_LOD_By_ICPRun* worksheet.
4. **Latitude**: In decimal degrees, WGS84 datum, coordinate reference system EPSG:4326.
5. **Longitude**: In decimal degrees, WGS84 datum, coordinate reference system EPSG:4326.
6. **Altitude**: The altitude in meters above sea level.
7. **LocationPrecision**: The horizontal and vertical positional accuracy, in meters.
8. **SamplingStart**: Start of sampling date and time stamp, UTC + 3. This is the date and time automatically recorded by the KoBoCollect app when the enumerators start recording cereal grain and soil sample information in the field. The start sampling date for the ID ETH1219 was incorrect and removed.
9. **SamplingEnd**: End of sampling date and time stamp, UTC + 3. This is the date and time automatically recorded by the KoBoCollect app when the enumerators either save or submit the sample metadata recording questionnaire.
10. **Crop**: Type of crop from which cereal grains were sampled.
11. **GrainSource**: Source of cereal grain (i.e., standing crop, field stack or store).
12. **Site**: Whether the site is “main” or “close pair”. Please see Gashu *et al.*⁷ for further details on this.

Malawian data. Identical to the Ethiopian dataset, the first 12 fields in the raw and cleaned Malawian cereal grain and soil chemistry data worksheets are auxiliary and field data for the cereal grain-soil sample pairs (records). The data details which vary from above are as follows:

- **Crop_ICP_Run**: The ICP-MS run number to determine the crop elemental LOD for the raw cereal grain dataset except Se using the data in the *Crop_LOD_ByICPRun* worksheet.
- **Crop_ICP_Run_Se**: The ICP-MS run number to determine the cereal grain Se concentration LOD for the raw cereal grain dataset using the LOD data in the *Crop_LOD_ByICPRun* worksheet.

Cereal grain elemental concentration. In both Ethiopia and Malawi datasets, the cereal grain elemental concentration for 29 elements (columns or data fields) was reported (Table 2). These appear in alphabetical order of element symbol next to the field and auxiliary data described in the sections above. The data field names also contain **_grain** suffix after the elemental symbol.

Soil chemistry. In the Ethiopian and Malawian datasets, 84 and 69 soil chemical properties, respectively were reported (Online-only Table 1). Please refer to the individual country dataset to find out which soil chemical properties were analysed and reported. These are presented in alphabetical order after the cereal grain elemental concentration data columns.

Technical Validation

Grain analysis. Quality control protocols for grain analysis included two operational blanks in each digestion batch and duplicate samples of a certified reference material (CRM) (Wheat flour SRM 1567b, National Institute of Standards and Technology, Gaithersburg, MD, USA) in approximately every fourth digestion batch. An LOD was calculated as 3 times the standard deviation of 10–14 blanks assuming a hypothetical mass of 0.2 g of sample. The data for percentage recoveries of CRM is shown in Table 3.

Soil chemical analysis. Quality control protocols included two operational blanks in each digestion batch and two samples of an internal or external reference material. Ten percent of samples were repeated through both the extraction/digestion and analysis stages, and the results were expected to fall within $\pm 5\%$, if not, the sample batch was repeated apart from results near to the LOD (which occurs with some samples that have low trace element concentrations). All instruments used were drift-corrected during each run as part of the calibration procedure. In each batch of soil pH_w determination, 10% unknowns were repeated and duplicate samples of in-house standard Broadbalk 082 (Rothamsted Research) were included for quality control, and we participated in the Wageningen Evaluating Programmes for Analytical Laboratories (WEPAL) inter-laboratory proficiency testing for soil pH. Total C and N analysis was set up with LECO Soil standard LCRM 502–697, lot 1000, and validated over time using proficiency WEPAL testing. No CRM was available for inorganic C, so in-house standards Summerdells, Broadbalk, Sacrewell (Rothamsted Research) were used in each batch, as well as WEPAL proficiency testing. No CRM was available for E_{Oxa}, and in-house standards Leuven and Woburn (Rothamsted Research) were used. No CRM was available for eCEC, and in-house standard Leuven soil (Rothamsted Research) was used. No CRM was available for P_{Olsen}, and in-house Hoosfield Plot 714 and Plot 444 (Rothamsted Research) were used, as well as WEPAL proficiency testing. No CRM was available for PBI, and in-house standards Leuven and Woburn were used.

For quasi-total concentrations, 10% blanks (extraction solution) were included to check for potential contamination of analytes of interests. Batches were rejected if the blanks showed apparent signals greater than 3x SD of the background signal of the instrument (limit of detection). WEPAL soil reference material ISE 962 were used to check analytical precision of every batch (Online-only Table 2), as well as WEPAL proficiency testing. Comparing average results after blank subtraction for all elements with the results for ISE 962 from WEPAL shows recoveries of $\pm 10\%$ for most of the 20 elements. Exceptions are As, Cd, Na and Ti by ICP-OES, and Cd, Mo and Se by ICP-MS (Online-only Table 3). Note that As, Cd and Pb were determined by ICP-OES only for the GCRF samples from Ethiopia (11 batches) but for all of the other samples in the BMGF project these elements

Country	ICP_Run	Recovery (%) by element															
		Al	As	Ca	Cd	Cu	Fe	K	Mg	Mn	Mo	P	Pb	Rb	Se	V	Zn
ETH_GCRF	1	75.4	78.7	96.3	97.5	87.2	92.3	102.0	95.5	96.2	82.8	96.2	78.9	97.1	98.0	88.7	91.1
	2	80.9	72.7	97.5	84.2	90.4	89.6	100.5	95.1	96.0	100.1	95.9	62.2	94.0	95.0	89.4	89.9
ETH_BMGF	1	37.7	105.1	102.7	233.1	94.8	124.0	105.9	99.3	98.8	96.1	97.0	119.2	96.2	100.0	-10.8	79.3
	2	71.1	54.3	101.9	62.6	92.5	89.7	100.0	93.2	96.0	94.5	98.2	78.2	94.3	97.7	320.9	91.7
	3	71.0	81.1	99.2	123.5	96.5	90.2	100.3	95.3	96.2	669.0	95.2	124.6	96.1	99.6	426.1	91.5
	4	69.6	53.3	105.4	96.6	97.7	98.7	104.8	97.6	99.4	98.4	102.6	42.7	99.9	101.2	190.8	95.2
	5	77.6	72.1	98.9	117.2	86.9	80.0	98.8	92.3	93.2	93.0	95.3	-162.4	93.8	94.3	321.0	85.9
MWI	1	77.7	81.7	106.2	94.1	97.9	118.5	109.1	101.2	99.6	83.5	96.4	1176.9	98.9	101.6	184.8	102.4
	2	77.4	70.0	103.0	94.2	92.9	90.9	107.1	99.5	95.3	106.0	97.5	124.8	95.2	98.3	112.8	92.8
	3	78.2	57.6	104.5	96.2	97.9	200.9	110.8	104.9	99.9	94.6	97.4	119.5	98.6	102.8	102.8	92.2
	4	96.5	615.8	105.4	74.9	96.3	88.7	118.0	109.1	98.9	96.3	102.6	136.0	100.0	101.5	428.8	89.7
	5	85.6	205.3	106.6	-51.0	100.6	94.6	121.7	115.0	103.6	116.4	102.2	-11.5	103.3	103.6	87.2	101.2
	6	80.6	115.1	100.5	123.2	91.3	90.2	113.5	108.2	95.8	92.8	95.5	-101.6	96.7	95.6	178.8	86.7
	7	76.6	215.0	99.6	93.4	98.3	89.9	115.6	109.4	98.2	114.9	98.2	165.3	99.3	99.2	234.8	91.8
	8	66.1	132.4	99.0	97.4	95.2	106.7	116.4	110.4	99.7	99.5	99.7	121.0	98.4	99.2	86.7	88.5

Table 3. The percentage recovery of elements of the certified material (Wheat flour SRM 1567b) used in microwave acid digestion of grain samples. The cereal grains collected from Ethiopia (ETH_GCRF and ETH_BMGF) and Malawi (MWI). ICP Run is the inductively coupled plasma mass spectrometer (ICP-MS) run number.

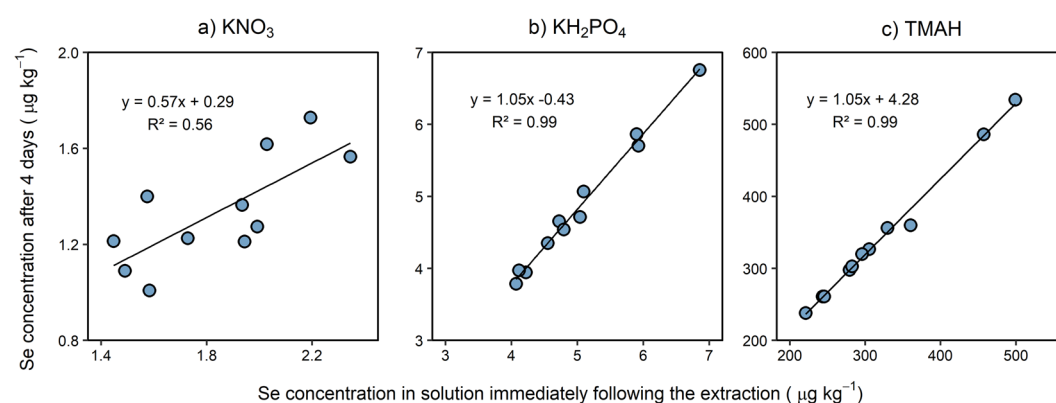


Fig. 5 Selenium concentrations in soil measured immediately following (a) potassium nitrate (KNO_3) extraction; (b) potassium phosphate (KH_2PO_4) extraction; (c) Tetramethylammonium hydroxide (TMAH) extraction and measured again after 4 days of storage of soils from Amhara region, Ethiopia.

were determined by ICP-MS. Analysis by ICP-MS compared to ICP-OES improved both the recovery and variation for As; for Cd the mean value decreased and so did the variation; Pb was similar but with decreased variation. Na recovery by ICP-OES and aqua regia extraction can be incomplete, and the results WEPAL given for Ti are only indicative values. Limits of detection for each of the above methods of soil analysis, except pH, are shown in Online-only Tables 4, 6, 8, 10, 12 to enable users to decide whether results for individual samples are reliable. Results for blanks are also shown in online-only Tables 5, 7, 9, 11, 13 and, except for quasi-total elements, these have not been subtracted from the results in the soil data. However, this information has been provided to enable users to judge the magnitude of the blanks compared to sample results for each method, and to enable them to decide whether they subtract them.

For the three-step sequential extraction scheme of S and Se in soil, a preliminary experiment was conducted to test the temporal stability of the analytes in a 1% TMAH matrix; analytes were measured after the extraction was completed and then again after storage for 4 days. The results showed that the agreement between the two measurements was quite poor for the KNO_3 extraction, especially in case of Se (Fig. 5), whereas there were very good agreements between the two measurements in the case of the KH_2PO_4 and TMAH extractions. Consequently, concentrations in the soluble fractions (0.01 M KNO_3) were preserved in a mixture of 0.01 M KH_2PO_4 and 1% TMAH.

The reproducibility of the isotopic dilution analysis was tested by repeating c. 10% of the samples with precision determined by calculating the relative standard deviation (RSD; %) of the duplicates. The average (error) RSD was 6.34% (sd = 7.68%), and 80% of the repeated samples had values of RSD < 10% (Fig. 6).

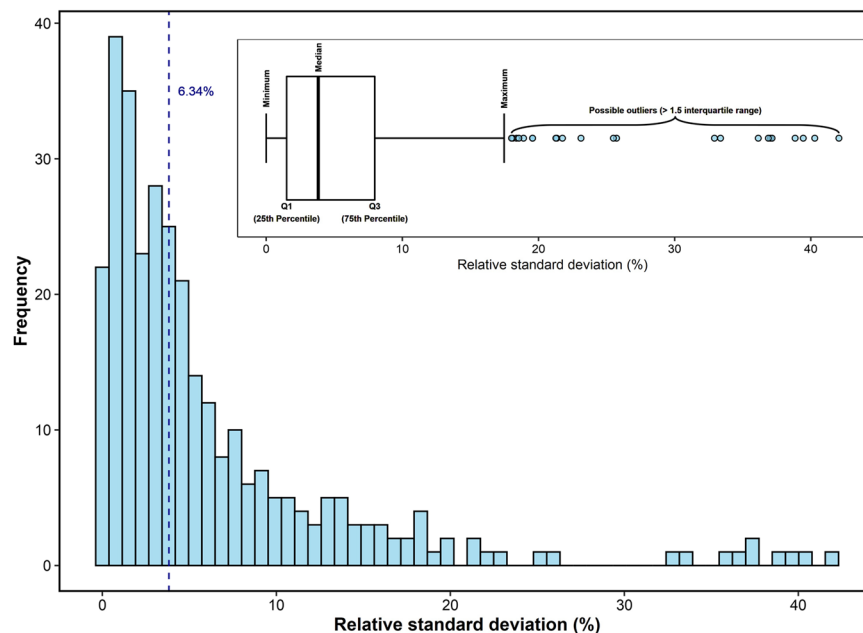


Fig. 6 Frequency of the relative standard deviation (RSD) of duplicate E-value measurements on a subset of soils from Ethiopia and Malawi. Vertical dashed blue line represents the mean value of RSD. The inserted graph shows a whisker-boxplot of the distribution of the RSD of duplicate measurements.

Auxiliary information on the soil samples for the various soil analytical approaches are presented in Supplementary files 5 and 6.

Usage Notes

- Cereal grain elemental concentration data presented in Supplementary File 4 were derived from data which excluded concentration values \leq LOD, including negative values. For some elements, where there are larger numbers of cereal grain samples with concentrations \leq LOD, this can lead to over-estimation of the median concentration values compared with when all data are used. For example, if all data, including those \leq LOD, are used, the median Se concentration in maize in Malawi is $0.0168 \text{ mg kg}^{-1}$ ($n = 1,603$). When data values \leq LOD are excluded, the median Se concentration in maize is $0.02448 \text{ mg kg}^{-1}$ ($n = 1199$). Users wanting to use descriptive data from summary tables should be aware of this, and select data in formats which are appropriate to their purpose. This note also applies to some of the soil chemistry summary table where there are many negative concentrations in the data.
- Users should be aware that some soil properties were not measured in both the GCRF-funded and BMGF-funded rows in the Ethiopian dataset. The following soil chemistry data were not measured in the GCRF-funded records: Cd_DTPA, Co_DTPA, Cu_DTPA, Fe_DTPA, Mn_DTPA, Ni_DTPA and Pb_DTPA. Similarly, the following soil chemistry properties were not measured in BMGF-funded records: Co_CaCl₂, Cu_CaCl, Fe_CaCl, K_CaCl, Mg_CaCl, Mn_CaCl, Mo_CaCl, Na_CaCl, Ni_CaCl, NOPC, P_CaCl, Se_CaCl, Zn_CaCl and Zn_FIA.
- Users wanting to carry out any geospatial analyses using the latitude and longitude fields in these data should ensure that the spatial (horizontal and vertical) accuracy reported meets the requirements of their analyses.

Code availability

There is no specific code developed to access these data. Users can use and process the data in software of their choice.

Received: 6 January 2022; Accepted: 28 June 2022;

Published online: 25 July 2022

References

1. Kumssa, D. B. *et al.* Dietary calcium and zinc deficiency risks are decreasing but remain prevalent. *Sci Rep* **5**, 10974, <https://doi.org/10.1038/srep10974> (2015).
2. Joy, E. J. M. *et al.* Dietary mineral supplies in Malawi: spatial and socioeconomic assessment. *BMC Nutr* **1**, 1–25, <https://doi.org/10.1186/s40795-015-0036-4> (2015).
3. Beal, T., Massiot, E., Arsenault, J. E., Smith, M. R. & Hijmans, R. J. Global trends in dietary micronutrient supplies and estimated prevalence of inadequate intakes. *PLoS One* **12**, e0175554, <https://doi.org/10.1371/journal.pone.0175554> (2017).
4. Nelson, G. *et al.* Income growth and climate change effects on global nutrition security to mid-century. *Nature Sustainability* **1**, 773–781, <https://doi.org/10.1038/s41893-018-0192-z> (2018).

5. Smith, M. R., Micha, R., Golden, C. D., Mozaffarian, D. & Myers, S. S. Global Expanded Nutrient Supply (GENUS) Model: A New Method for Estimating the Global Dietary Supply of Nutrients. *PLoS One* **11**, e0146976, <https://doi.org/10.1371/journal.pone.0146976> (2016).
6. Gashu, D. *et al.* 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).
7. Gashu, D. *et al.* 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).
8. Ligowe, I. S. *et al.* Selenium deficiency risks in sub-Saharan African food systems and their geospatial linkages. *Proc Nutr Soc*, 1–11, <https://doi.org/10.1017/S0029665120006904> (2020).
9. Phiri, F. P. *et al.* 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).
10. Belay, A. *et al.* Selenium deficiency is widespread and spatially dependent in Ethiopia. *Nutrients* **12**, 1565, <https://doi.org/10.3390/nu12061565> (2020).
11. Mossa, A.-W. *et al.* The effect of soil properties on zinc lability and solubility in soils of Ethiopia—an isotopic dilution study. *Soil* **7**, 255–268 (2021).
12. Botoman, L. *et al.* Soil and landscape factors influence geospatial variation in maize grain zinc concentration in Malawi. *Scientific reports* **12**, 7986, <https://doi.org/10.1038/s41598-022-12014-w> (2022).
13. Walsh, M. & Wu, W. *GeoSurvey data prediction workflows*, <https://osf.io/vxc97/> (2020).
14. OpenStreetMap Contributors. *Planet dump*, <https://www.openstreetmap.org>. (2017).
15. Grafström, A. & Schelin, L. How to select representative samples. *Scandinavian Journal of Statistics* **41**, 277–290 (2014).
16. ESA. *Land Cover CCI Data version 2.0.7, land cover data for 2015*, <http://maps.elie.ucl.ac.be/CCI/viewer/download.php> (2016).
17. National Statistical Office (NSO) & The DHS Program (ICF). (NSO & ICF, Zomba, Malawi, and Rockville, Maryland, USA, 2017).
18. Walvoort, D. J., Brus, D. & De Gruijter, J. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers & Geosciences* **36**, 1261–1267 (2010).
19. Tiessen, H., Bettany, J. & Stewart, J. An improved method for the determination of carbon in soils and soil extracts by dry combustion. *Commun Soil Sci Plant Anal* **12**, 211–218 (1981).
20. Schwertmann, U. Differenzierung der Eisenoxide des Bodens durch Extraktion mit Ammoniumoxalat-Lösung. *Zeitschrift für Pflanzenernährung, Düngung, Bodenkunde* **105**, 194–202 (1964).
21. ISO 23470. Soil quality - Determination of effective cation exchange capacity (CEC) and exchangeable cations using a hexaminecobalt (III) chloride solution. (2018).
22. Olsen, S. R. *Estimation of available phosphorus in soils by extraction with sodium bicarbonate*. (US Department of Agriculture, 1954).
23. Burkitt, L., Moody, P., Gourley, C. & Hannah, M. A simple phosphorus buffering index for Australian soils. *Soil Res.* **40**, 497–513 (2002).
24. McGrath, S. P. & Cunliffe, C. H. A simplified method for the extraction of the metals Fe, Zn, Cu, Ni, Cd, Pb, Cr, Co and Mn from soils and sewage sludges. *J Sci Food Agric* **36**, 794–798 (1985).
25. Lindsay, W. L. & Norvell, W. A. Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil Sci Soc Am J* **42**, 421–428 (1978).
26. Mathers, A. W. *et al.* Determining the fate of selenium in wheat biofortification: an isotopically labelled field trial study. *Plant Soil* **420**, 61–77 (2017).
27. Shetaya, W., Young, S., Watts, M., Ander, E. & Bailey, E. Iodine dynamics in soils. *Geochim Cosmochim Acta* **77**, 457–473 (2012).
28. Kumssa, D. B. *et al.* Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in Ethiopia and Malawi. *figshare* <https://doi.org/10.6084/m9.figshare.15911973> (2022).

Acknowledgements

This paper is supported by GeoNutrition projects funded by the Bill & Melinda Gates Foundation (INV-009129) and the UKRI Biotechnology and Biological Sciences Research Council (BBSRC)/Global Challenges Research Fund (GCRF) (BB/P023126/1). The funders were not involved in the study design; the collection, management, analysis, and interpretation of data; the writing of the paper or the decision to submit the paper for publication. 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. We acknowledge the contributions made to this research by the participating farmers and field sampling teams. In Ethiopia, field sampling teams were from the Amhara, Oromia and Tigray 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. Training support for field activities in Malawi 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”. Mineral analytical support was provided by B. Broadley, S. Vasquez Reina, S. Dunham, J. Carter and J. Hernandez with support from BBSRC Institute Strategic Project Soil to Nutrition (BBS/E/C/000I0310). E.L.A.’s contribution is published with the permission of the Director of the British Geological Survey (UKRI).

Author contributions

D.G., P.C.N., T.A., E.L.A., S.G., E.J.M.J., A.A.K., R.M.L., S.P.M., E.K.T. and M.R.B. conceptualized the study, and acquired and administered project funding. D.G., P.C.N., E.L.A., E.H.B., L.B., C.C., S.M.H., K.H., E.J.M.J., D.B.K., R.M.L., I.R.H., P.M., I.S.L., K.D., S.P.M., A.E.M., A.W.M., M.M., M.G.W., L.W., S.D.Y. and M.R.B. contributed to field surveys and laboratory analyses. D.B.K. managed the data; D.B.K., and A.W.M. developed the data visualizations. D.B.K., A.W.M. and M.R.B. wrote the primary draft of the paper with editing and reviewing inputs from other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01500-5>.

Correspondence and requests for materials should be addressed to M.R.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022