The distribution of cadmium in soil and cacao beans in Peru

Evert Thomas [a,*], Rachel Atkinson [a], Diego Zavaleta [a], Carlos Rodriguez [b], Sphyros Lastra [a,c], Fredy Yovera [a,d], Karina Arango [a], Abel Pezo [a], Javier Aguilar [b], Miriam Tames [b], Ana Ramos [b], Wilbert Cruz [c], Roberto Cosme [c], Eduardo Espinoza [d], Carmen Rosa Chavez [e], Brenton Ladd [f]

[a] Bioversity International, Lima, Peru
[b] Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA), Lima, Peru
[c] Instituto Nacional de Innovación Agraria (INIA), Lima, Peru
[d] Cooperativa Agraria Norandino, Piura, Peru
[e] Ministerio de Desarrollo Agrario y Riego del Perú (MIDAGRI), Lima, Peru
[f] Universidad Científica del Sur, Lima, Peru

HIGHLIGHTS

- Cadmium content in cacao exceeds regulatory thresholds from some regions in Peru and limits access to international markets
- We developed nation-wide predictive maps of soil and cacao bean cadmium
- The most important predictors of soil and cacao bean cadmium are geology, precipitation seasonality, rainfall, pH, and geographical location
- Elevated concentrations of soil and cacao bean cadmium are largely restricted to the northern parts of the country
- While in the north of Peru most cacao farmers are impacted, at a national level, <20% will be affected by the current regulations

GRAPHICAL ABSTRACT
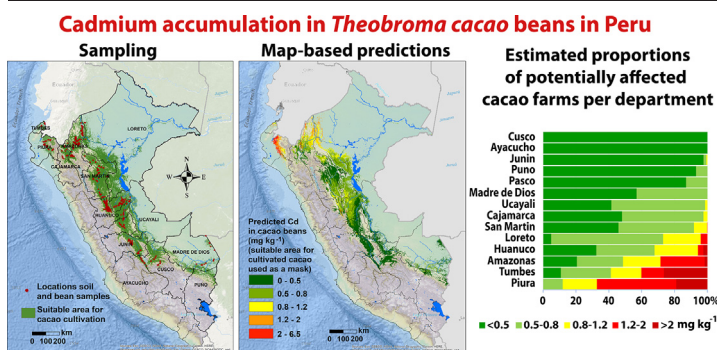
ARTICLE INFO

ABSTRACT

Peru is the eighth largest producer of cacao beans globally, but high cadmium contents are constraining access to international markets which have set upper thresholds for permitted concentrations in chocolate and derivatives. Preliminary data have suggested that high cadmium concentrations in cacao beans are restricted to specific regions in the country, but to date no reliable maps exist of expected cadmium concentrations in soils and cacao beans. Drawing on >2000 representative samples of cacao beans and soils we developed multiple national and regional random forest models to develop predictive maps of cadmium in soil and cacao beans across the area suitable for cacao cultivation. Our model projections show that elevated concentrations of cadmium in cacao soils and beans are largely restricted to the northern parts of the country in the departments of Tumbes, Piura, Amazonas and Loreto, as well as some very localized pockets in the central departments of Huánuco and San Martin. Unsurprisingly, soil cadmium was the by far most important predictor of bean cadmium. Aside from the south-eastern to north-western spatial trend of increasing cadmium values in soils and beans, the most important predictors of both variables in nation-wide models were geology, rainfall seasonality, soil pH and rainfall. At regional level, alluvial deposits and mining operations were also

associated with higher cadmium levels in cacao beans. Based on our predictive map of cadmium in cacao beans we estimate that while at a national level <20 % of cacao farming households might be impacted by the cadmium regulations, in the most affected department of Piura this could be as high as 89 %.

1. Introduction

Cacao production in Latin America currently accounts for some 18 % of global production, and has been on the rise in recent years, for reasons that include its use in development programs to alleviate rural poverty, promote peace in post-conflict regions and replace illicit crops (DEVIDA, 2017; Abbott et al., 2018). However, access to markets is potentially limited due to the presence of cadmium (Cd) in soils which may in turn lead to Cd accumulation in cacao beans. To date this problem appears restricted to Latin America and the Caribbean. The accumulation in beans can result in levels above the thresholds for chocolate and cocoa products set by markets, most importantly the European Union (European Commission regulation (EU) 488/2014), but also the USA state of California (Industrial Agreement Proposition 65 (19/02/2018)). Limits on cadmium levels in chocolate have also been put in place by Australia and New Zealand (Food Standards Code Standard 1.4.1), Indonesia (Indonesian National Standard) and the Russian Federation (SanPin 2.3.2-1078-01). Cd is a non-essential metal which can cause health risks above certain intake levels, and the limits have been set to protect the health of consumers (Meter et al., 2019).

The presence of cadmium in cacao farm soils of Latin-America has been attributed to both geogenic sources such as the weathering of bedrock, alluvial deposition of sediments, and anthropogenic activities including the use of cadmium-contaminated mineral fertilizers (Argüello et al., 2019; Bravo et al., 2021; Meter et al., 2019; Tantalean Pedraza and Huauya Rojas, 2017; Zug et al., 2019). Among the most important edaphic and climatic variables correlated with cadmium accumulation in cacao beans identified to date are soil cadmium levels, soil characteristics such as pH, soil organic matter and micronutrient levels, as well as seasonality of rainfall (Argüello et al., 2019; Gramlich et al., 2018; Scaccabarozzi et al., 2020; Wade et al., 2022).

However, while higher cadmium levels are more pronounced in Central and Latin America than in Africa and Asia, cadmium levels above the regulatory thresholds are not ubiquitous across all cacao cultivation areas (Arévalo-Gardini et al., 2019; Bravo et al., 2021). This is due to the varied geological history of cacao producing regions, and upper water catchments that provide sediments and water to farms downstream (Argüello et al., 2019; Chavez et al., 2015). Because of this variation, mapping the spatial distribution of cadmium in soils and beans is critical to guide decisions on mitigation efforts, to estimate potential impact of the regulatory interventions, and help actors in the value chain comply with regulations. This has already been carried out for Ecuador (Argüello et al., 2019), Colombia (Bravo et al., 2021) and Honduras (Gramlich et al., 2018). In Peru, previous and localized studies indicate that there are areas with high cadmium within the country (Arévalo-Gardini et al., 2019; Remigio, 2014) but to date a complete mapping has not been done.

Peru is an important cradle of cacao diversity (Thomas et al., 2012) and has experienced continued growth of its cacao sector over the past years. It is the eighth largest exporter of cacao globally, with about 70 % of its exports historically going to Europe. It is also the second largest exporter of organic beans, after the Dominican Republic, the main market also being Europe. High Cd concentrations have already led to reductions in cacao bean exports from Peru to European markets, with farmers from the region of Piura reporting income losses of 31 % on average (Villar et al., 2022). Our objective here was to develop predictive maps of the expected cadmium concentration in soils and cacao beans across the environmental niche suitable for cacao cultivation in Peru (Ceccarelli et al., 2021). Drawing on extensive national soil and cacao bean cadmium measurements, we constructed random forest models at ~250 m resolution using climate, geology, terrain, soil and anthropogenic disturbance variables available as spatial data layers to develop national and regional spatial maps of predicted soil and bean cadmium. The random forest method can obtain equally accurate and unbiased predictions as the krigging techniques, used in previous Cd mapping studies (e.g. Argüello et al., 2019), but has the advantage that it does not require any strict statistical assumptions to be met and is more flexible when combining explanatory variables of different types, among others (Hengl et al., 2018). The resulting maps allow for estimation of the number of farmers potentially affected by cadmium limits, can guide the identification of low cadmium areas for cacao cultivation and help prioritize areas for mitigation action.

2. Methodology

2.1. Farm selection and sampling

Cacao is grown widely across Peru, by >100,000 farmers and covering >180,000 ha, mostly located in the regions of Tumbes, Piura, Amazonas, Cajamarca, San Martin, Ucayali, Huánuco, Pasco, Junín and Cusco (MIDAGRI, 2022). To capture the variability of farms in terms of soils and cacao genotypes, we aimed to sample 100–150 cacao trees in each of these regions, with a lesser number in the regions of Madre de Dios and Ayacucho where cacao production covers a much smaller area. The principal cacao cultivation areas in each region were identified based on consultation with local experts. Available spatial information on cadmium in cacao was also used to ensure that hotspots identified previously were included (Arévalo-Gardini et al., 2019; HuamanI-Yupanqui et al., 2012; Remigio, 2014). Samples were collected from at least 5 farms in each cultivation area, with farm choice based on prior coordination with key actors, accessibility, and owner consent. Germplasm banks and clonal gardens were also included in the sampling. Sampling was carried out between April 2018 and September 2019 and a total of 563 farms were visited in 35 provinces of 13 departments of Peru, estimated to represent 90 % of the producing areas. In each farm, once the consent of the owner was given, a brief questionnaire was carried out to understand current and previous crop management (Supplementary material 1). The owner was asked to select trees that represented the diversity of genotypes present and up to a maximum of 6 trees were sampled per farm, resulting in a nationwide total of 2194 trees. We collected leaf material of each tree for posterior genetic characterization, but this information was not included in the current modelling. Each selected tree was given a unique code and marked for future reference and the GPS coordinate was recorded. The sampling was carried out using research permits N° 001 -2021-MIDAGRI-INIA/DGIA and N° 0003-2021-INIA-DGIA.

The method of sampling soil and beans for individual trees was adapted from similar studies carried out in the region (Chavez et al., 2015; Gramlich et al., 2018; Ramtahal et al., 2016; Zug et al., 2019). For soil samples, the leaf litter was cleared under the crown and a soil sample was taken at eight equidistant points 70 cm from the trunk using a soil auger to a depth of 20 cm. The eight samples were combined to form a single composite sample. For bean samples, up to four cacao pods were harvested from each tree. The pods were opened on site and beans stored in a plastic bag. To avoid running lots of samples with cadmium levels below the limit of detection, the Cd concentration from one randomly selected tree from each farm was analysed initially. If the total soil cadmium > 0.05 mg kg^{-1}, or total bean Cd above the limit of quantification limit (0.047 mg kg^{-1}; see below), the remaining samples from the farm were also analysed.

Additionally, 334 composite soil and 484 bean samples taken at a farm level were included in this analysis. Farm-level soil samples were collected using a zig zag methodology with 8 samples per hectare that were

combined to provide one composite soil sample. For beans, 5 to 10 pods were collected to give a total of approximately 1 kg of fresh beans. The reason for using two sampling methodologies is because data were collected under different projects (see funding details) each with their specific objectives. The tree level sampling had the objective of understanding which variables (including genetics) influence variation in cadmium uptake in beans (results will be discussed in a separate manuscript) while the second was to assess the average cadmium concentration at farm gate. Both methodologies took composite samples of soil and beans, the main difference being that the tree-level sampling yielded multiple samples per farm.
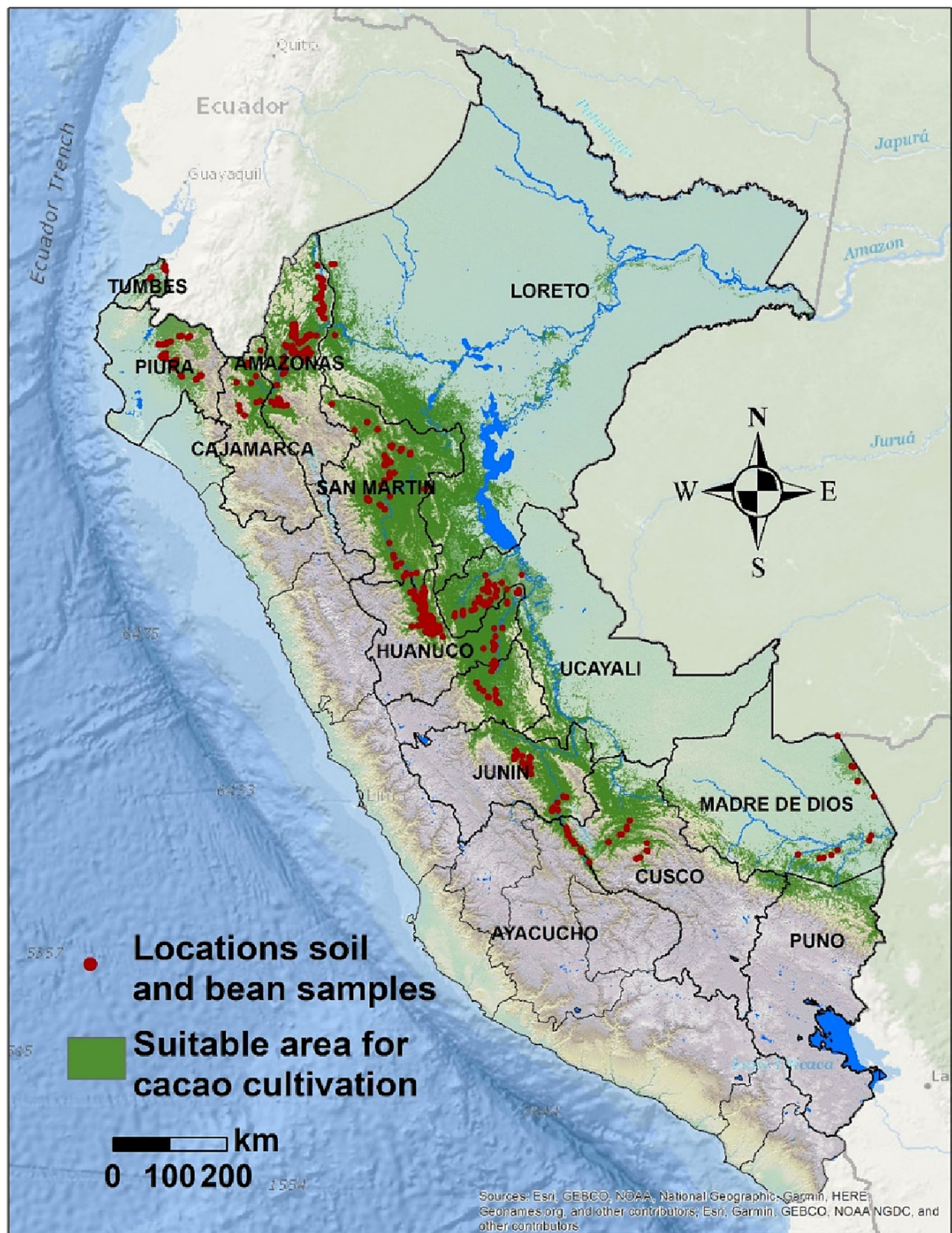

Fig. 1. Sampling locations of soil and cacao bean samples. The area suitable for cacao cultivation is shown in green (*sensu* Ceccarelli et al., 2021).

In sum, we collected 1717 soil and 1534 bean samples from individual trees from 563 farms, in addition to 334 composite soil and 484 bean samples from 484 farms, making a total of 2051 data points for soil Cd and 2018 for bean Cd (Fig. 1). The soil samples were air-dried for 1 to 2 weeks, ground and sieved through a 2 mm mesh to give 400 g of sample. The bean samples were drained and dried in an oven at 70 °C for 48 h on card-board containers until a humidity of about 7 % was reached.

The total Cd concentration in soil was quantified using the analytical services of SGS, a certified multinational laboratory with facilities in Peru. A 300 g sample was requested by the laboratory, who digested a 2.5 g aliquot in agua regia (1:3 HCl:HNO$_3$) for 1.5 h on an open block. Total soil cadmium was quantified by inductively coupled plasma mass spectrometry (ICP-MS) (Nexion 2000C, Perkin Elmer). The laboratory uses a certified reference standard (OREAS 906) with 0.42 ± 0.04 mg kg^{-1} Cd. It carries out a duplicate analysis on 5 % of the samples, and blanks are run every 50 samples. The limit of quantification is 0.01 mg kg^{-1}.

The bean cadmium concentration collected at individual tree level was quantified in the ISO certified laboratory of the Peruvian government agency SENASA (Servicio Nacional de Sanidad Agraria del Peru), which is the competent authority for food safety in Peru. Whole beans were frozen overnight and finely ground in a food mill. 0.5 g of this powder was digested in 6 ml of HNO$_3$ and 2 ml H$_2$O$_2$ for at least 1.5 h at 200 °C prior to analysis by ICP-MS. The laboratory runs blanks and duplicates every 10 samples. Internal controls (bean Cd samples) are also used. The limit of quantification is 0.047 mg kg^{-1}. The composite bean samples taken at farm level were analysed in SGS, following the AOAC Official Method 2013.06, modified in 2019. The limit of detection is 0.01 mg kg^{-1} and limit of quantification 0.03 mg kg^{-1}. Average recuperation is 105 % (80–110 %). Both laboratories took part in an inter laboratory ring test that involved 24 labs from Peru, Ecuador and Colombia running internal reference samples. The results from both laboratories yielded z scores of <2, indicating acceptable levels of accuracy according to ISO 13528 (2015) and ISO 17043 (2010) (Dekeyrel, 2021).

To assess the importance of irrigation water as a source of Cd contamination in cacao plantations, we used DGT passive membranes (Davison and Zhang, 1994) to determine the Cd load of irrigation channels used to irrigate cacao plantations in La Quemazón and Las Lomas, Piura, with contrasting bean Cd concentrations. Cacao beans from La Quemazón and Las Lomas have among the lowest and highest Cd in the region, respectively. One membrane was left in the irrigation channel for one week every month over a full year at each site (a total of 12 membranes per site), and the membranes were analysed using ICP MS following EPA Compendium Method 0.0010 IO-3.1; IO-3.5 (1999) by the certified laboratory TYPSA. The time-averaged concentration of Cd in the DGT membrane (C$_{DGT}$ nmol ml^{-1}) was calculated using the following formula:

$$C_{DGT} = \frac{M\Delta g}{D^{mdl}A_p t}$$

where M (nmol) is the mass of analyte accumulated in the binding layer, calculated from the measured concentration; Δg is the total thickness of the materials in the diffusion layer (0.094 cm); D^{mdl} (cm^2 s^{-1}) is the diffusion coefficient of Cd in the diffusion layer at the deployment temperature; A$_p$ is the physical area of the exposed filter membrane (3.14 cm^2) and; t (s) is the deployment time (Davison and Zhang, 2016). A yearly average of C$_{DGT}$ for each site was calculated from the data.

2.2. Variable selection

One limitation of building spatial predictions of expected cadmium content in soils and cacao beans is that predictor variables need to be available as spatial layers covering the area of interest. For example, soil nutrients such as zinc or manganese have been shown to be useful predictors of cadmium in soil or beans (Argüello et al., 2019; Bravo et al., 2021), but no maps exist of their distribution across the cacao cultivation areas in Peru. We therefore only used available spatial layers of geology, soil, terrain,

geographical position, vegetation, climate, deforestation and the presence of mining. To characterize geology, we used the map developed by INGEMMET (Instituto Geológico Minero y Metalúrgico, Ministry of Energy and Mines, Peru) at a scale of 1:100,000 (https://portal.ingemmet.gob.pe/web/guest/mapa-geologico-nacional). During initial explorations of the map in combination with our cadmium sampling data, we noticed that many high cadmium soils tended to be located on top, or downstream of one geological layer named "continental Quaternary Holocene" (Qh_c). We therefore created a new categorical map identifying both the location of the Qh_c layer and the areas downstream of it within the same watershed. We obtained soil type and seven major edaphic variables, from ISRIC-World Soil Information (Hengl et al., 2017): organic carbon (SOC), pH in H$_2$O (pH), % sand (Sand%), % silt (Silt%), % clay (Clay%), Cation Exchange Capacity (CEC), Bulk density (BLD), Coarse fragments >2 mm (CRF). For the edaphic variables we calculated a weighted mean across 0–5, 5–15, 15–30, 30–60, and 60–100 cm soil depth values to derive a single data value for 0–100 cm. For terrain variables we used altitude, slope, topographical position, terrain ruggedness, and direction of water flow, constructed with the *raster* package for R (Hijmans, 2019). We additionally used the vegetation map developed by the Ministry of Environment of Peru (https://sinia.minam.gob.pe/mapas/mapa-nacional-ecosistemas-peru) to distinguish *terra firme* areas from alluvial plains, based on vegetation types. To characterize climate across the cacao growing areas we used 19 bioclimatic variables obtained from worldclim (Fick and Hijmans, 2017). As anthropogenic disturbance variables we considered on-site deforestation and the presence mining activities in the upstream areas of the watershed as both these activities can lead to increased movement of cadmium from the soil and bed rock and posterior deposition in cacao growing areas. For deforestation, we used the Global Forest Watch maps of deforestation incurred over the last 12 years (Global Forest Watch, 2022), while the map of evidence of mining was constructed based on the 2019 inventory of mining sites by INGEMMET. As there was a trend of increasing cadmium concentrations in both soil and cacao bean samples from south to north and from east to west, we additionally included rasters of longitude and latitude among the predictor values, correspodning to the longitude and latitude of the grid cell centres in decimal degrees. The categorical maps available as shapefiles (geology, vegetation) were rasterized to a spatial resolution of 7.5 arc sec which was the unit area of analysis and of model projection.

To reduce the redundancy among predictor variables, we removed colinear soil, climate and terrain variables based on stepwise calculations of variance inflation factors (VIF), retaining only variables with VIFs <5. The retained continuous variables for model construction were: Longitude (in decimal degrees), mean diurnal temperature range (Temp Range Day; in °C), the maximum temperature of the warmest month (Max Temp; in °C), precipitation seasonality (Rain Season; no unit), precipitation of the warmest quarter (Rain Warm Quart; in mm), precipitation of the coldest quarter (Rain Cold Quart; in mm), the direction of water flow (Water Flow Dir; no unit), the topographical position index (TPI; in m), terrain ruggedness index (TRI; in m), soil clay content (Clay%; in g/kg), soil sand content (Sand%; in g/kg), soil silt content (Silt%; in g/kg), soil cation exchange capacity (CEC; in mmol(c)/kg), soil organic carbon (SOC; in dg/kg), soil pH in water (pH). The categorical explanatory variables were soil type (Soil Type), geological layer (Geology), the presence of continental quaternary Holocene deposits on site or upriver in the watershed (Quatern Geology), vegetation type (Vegetation), number of years since deforestation (Deforest), and presence of mining upriver in the watershed (Mining).

2.3. Statistical analyses

We developed multiple predictive random forest (RF) models of cadmium concentrations in soil and cacao beans using the *cforest* function in the *party* package for R (Strobl et al., 2007). Considering that in nearly all cases individual farms were located in different grid cells, to assess the influence on model performance of multiple observational (tree-level) data per grid cell, versus just one measurement (composite farm-level samples), we developed RF models both for cadmium measurements at tree/farm

level and averaged per grid cell. Furthermore, considering the dramatically different growing conditions in cacao cultivation regions in Peru (eg pH neutral soils and irrigation in the coastal valleys of the north vs acid soils and high rainfall in the Amazon basin), we compared the performance of models constructed at regional and national level. The regions considered were (using spatial extent of departments): the coastal valleys of Tumbes and Piura, the northern Amazon of Cajamarca and Amazonas, the central Amazon of San Martin, Ucayali, Huánuco and Pasco, and the southern Amazon of Junín, Ayacucho, Cusco and Puno.

We cross-validated each of the models based on 15 iterations of model calibrations and testing using spatial blocks with the blockCV package for R (Valavi et al., 2019) which divides a study area (defined by the locations of datapoints) in spatial blocks, while optimizing evenness of the number of datapoints per block. For the national models we divided our study area in 100 km wide squares (blocks), and for the regional models 20 km sized squares, based on approximate median spatial autocorrelation ranges of continuous explanatory raster variables across the respective study regions, as implemented in the blockCV package. Each model calibration was based on the data contained in a random selection of four fifths of the blocks, while the remaining data were used for model testing. Cross-validation with spatial blocks provides a better measure of the predictive power of a model in areas for which no data points are available. To assess model performance, we calculated the root mean squared error (RMSE), a measure of model accuracy, the mean absolute error (MAE) and the R^2 of simple linear models between predicted and observed values.

To quantify uncertainty of model predictions, both in areas with data and areas without data, we calculated the coefficient of variation for each grid cell based on the spatially projected calibration models we developed as part of the 15 cross-validation iterations.

Variable importance values were calculated using the *varimp* function in the *party* package and their variability was quantified using importance values for each of the 15 cross-validation models mentioned above. Importance values were based on the mean decrease in model accuracy and were standardized across runs by dividing by the value of the most important variable.

2.4. Assessing the impact on cacao cultivation areas and farming households

To estimate the proportion of cacao cultivation areas and the number of cacao farming households potentially affected by the regulations limiting Cd in chocolate and cocoa powder, we extracted predicted bean cadmium values from our final projected map for the ~20,000 farms used by Ceccarelli et al. (2021). Based on these values we calculated proportions of farms per department according to five classes of predicted bean Cd (0–0.5; 0.5–0.8; 0.8–1.2; 1.2–2; >2 mg kg^{-1}). Recent publications have proposed 0.6 mg kg^{-1} as an unofficial threshold used by importers (Vanderschueren et al., 2021), but in Peru a threshold of 0.5 mg kg^{-1} is typically used for bulk cacao and 0.8 to even 1.2 for single origin fine flavour cacao. We then used these proportions to estimate the number of cacao farmer households potentially affected based on numbers per department according to the agrarian statistics of 2018 census which reported approximately 84,000 cacao farmers accounting for a total production of 155,500 metric tonnes of cacao beans in Peru (MIDAGRI, 2022).

3. Results

The random forest models calibrated at tree/farm level performed consistently better in predicting observed measurements than the models trained based on average values per grid cell. In comparisons of predicted and observed values RMSE increased 14 to 93 % for soil Cd and 11 to 62 % for bean Cd, respectively. In what follows we therefore focus on the results obtained for the random forest models calibrated at tree/farm level. Principal components analysis identified two main clusters in the data and shows that soil and bean cadmium tend to increase towards the north-western part of the country, an area with lower rainfall and higher rainfall seasonality, and sandy rather than silty soils (Fig. 2). These same variables were among the most important predictors in the nation-wide random forest models for

soil Cd, in addition to the geology of the sampling location and soil pH (Fig. 3). Soil cadmium content was by far the most important predictor variable in the random forest model for bean Cd, in addition to the geographical position (represented by longitude), rainfall seasonality and the climax vegetation type of the sampling location. Both random forest models showed good performance in predicting observed soil and bean Cd (R^2 of 0.77 ± 0.15 and 0.83 ± 0.03 for calibration data, respectively), but the soil Cd models scored considerably lower in spatial block cross validations (R^2 = 0.24 ± 0.11) than the bean Cd models (R^2 = 0.48 ± 0.26).

The performance of random forest models developed on regional basis were variable (R^2 for soil and bean Cd models ranging from 0.36 to 0.87 and 0.22 to 0.85, respectively; Figs. 5 and 6). Interestingly, the most important variables in the models differed between regions. Rainfall seasonality and geology were most important in the north west departments of Piura and Tumbes which have among the highest soil and bean Cd concentrations in the country, and where cacao is exclusively grown under irrigation owing to the very dry and highly seasonal climate. In the northern Amazonian departments of Cajamarca and Amazonas by far the most important predictors for soil Cd were the location of the continental Quaternary Holocene geological layer within each watershed, followed by rainfall and geology more generally. The most important predictors of bean Cd in the region were the climatic variables relating to rainfall and temperature, after soil Cd concentration. In the departments of San Martin, Ucayali, Huánuco and Pasco, where most of the national cacao production is concentrated, geology and location of the continental Quaternary Holocene geological layer within each watershed were the most important predictors of soil Cd, followed by soil type, while for bean Cd these were geology, climatic variables of rainfall and temperature, after soil Cd concentrations. Finally, in the southernmost regions from Junin to Madre de Dios where Cd concentrations in bean and soil are mostly low to very low, soil CEC and pH were the most important predictors of soil Cd. By contrast terrain ruggedness was the by far most important predictor for bean Cd, even before soil Cd, though this was the least performing random forest model of all (Fig. 6).

While the most important predictor variables tended to be similar in the random forest models for both soil and bean cadmium, at a national and regional level, the nature of the relationships between numeric predictors and response variables was more variable. For example, while low rainfall and high soil pH tended to be associated with higher soil Cd concentrations, the opposite was true for bean Cd concentrations in the national-level random forest models (Figs. 3 and 4). Similarly, the functional relations between predictor and response variables often differed between partial dependence plots obtained from national and regional random forest models, as well as between regional models for soil and bean Cd.

As for the categorical predictors, the geological layers with highest and lowest partial dependence scores in the nation-wide random forest model for soil Cd corresponded with the most extreme ones of the regional models. Soil samples collected on top of the continental Quaternary Holocene geological layer (Qh_c) had consistently higher Cd concentrations than samples from other areas.

One observable trend in bean Cd concentrations at a national level is that they tended to be higher in areas close to lakes or rivers, or subjected to alluvial flooding, and on the fluvial deposit geological layers. Furthermore, in the regions of Piura-Tumbes and San Martin-Huánuco-Ucayali-Pasco, cacao beans collected in areas with active mining tended to have higher Cd concentrations than in other areas.

The year average of C_{DGT} Cd in water in Las Lomas, where soil and bean Cd concentrations of our samples were 2.80 ± 0.82 mg kg^{-1} and 7.12 ± 2.70 mg kg^{-1}, respectively, was 0.315 ± 0.267 μg l^{-1}, compared to 0.085 ± 0.108 μg l^{-1} for the La Quemazón area where average soil and bean Cd were 0.54 ± 0.1 mg kg^{-1} and 0.72 ± 0.36 mg kg^{-1}. Considering that approximately 8000m^3 water per hectare is used for irrigating cacao, and mature plantations have approximately 800 cacao trees per hectare, this means an average yearly influx of 3.15 ± 2.67 and 0.85 ± 1.08 mg of bioavailable Cd to the soil from irrigation water per cacao tree in the Las Lomas and Quemazón areas respectively. Cd concentrations in irrigation water showed seasonal patterns, with highest values observed in Las
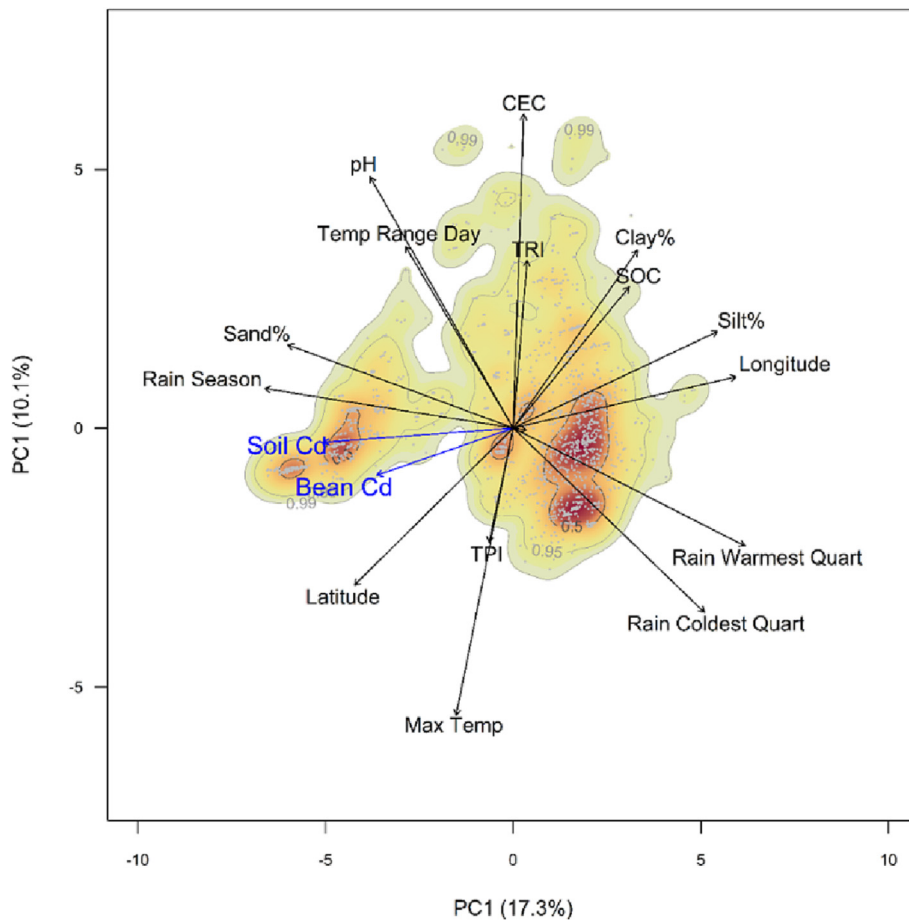
Fig. 2. Biplot of a principal component analysis of soil and bean cadmium and continuous explanatory variables considered here. The colour gradient indicates regions of the highest (dark red) to lowest (light green) densities of data points, through use of Kernel density estimations. Temp Range Day, mean diurnal temperature range; Max Temp, max temperature of the warmest month; Rain Season, precipitation seasonality; Rain Warm Quart, precipitation of the warmest quarter; Rain Cold Quart, precipitation of the coldest quarter; TPI, topographical position index; TRI, terrain ruggedness index; Clay%, soil clay content; Sand%, soil sand content; Silt%, soil silt content; CEC, soil cation exchange capacity; SOC, soil organic carbon; pH, soil pH in water.

Lomas between June and October and in La Quemazón between May and June (Fig. S27).

Spatially explicit predictions of soil and bean Cd by the random forest models showed high consistency for most areas. The coefficient of variation was lower than 0.26 and 0.36 for 95 % of the grid cells for soil and bean Cd respectively (Fig. 7).

Our findings suggest that approximately 16 to 45 % of cacao farming households in Peru might be impacted by the regulation, having concentration of Cd in beans above 0.8 and 0.5 mg kg^{-1}, respectively (Fig. 8). These levels are commonly used as upper threshold by buyers of fine flavour and conventional cacao beans, respectively. Most of these farms are found in the northern parts of the country in the departments of Tumbes, Piura, Amazonas and Loreto (with 59–89 %, 89–100 %, 51–79 %, 27–95 % of farms in each respective region potentially impacted), as well as some very localized pockets in the central departments of Huánuco and San Martin. In terms of the national cacao production in Peru, 9.2 to 41 % might be affected by the Cd restrictions, for the 0.8 and 0.5 ppm thresholds, respectively (Fig. 8).

4. Discussion

Here we present the first maps of Cd in soils and cacao beans at a national scale for Peru. With 2051 data points from 1047 farms, it represents the most comprehensive collection of data for any cadmium mapping exercise to date (Ecuador 159 farms, 560 samples (Argüello et al., 2019), Honduras 55 farms, 110 samples (Gramlich et al., 2018), Colombia 1827 data points, one per farm but only soil analysis (Bravo et al., 2021).

Although our analysis is limited by fewer soil parameters per sample than, for example Argüello et al. (2019), our modelling approach allows for the integration of spatially explicit layers available at the national level which leads to more power in spatial predictions beyond the sampling areas. Further, comparison of the functional relations between predictor and response variables obtained for nation-wide and regional random forest models permits making inferences about their causal or circumstantial nature. Climate, soil and terrain variables with consistent trends in partial dependence curves at different scales and geographies are likely to play a causal role in explaining soil and/or bean cadmium. For example, higher rainfall tended to be associated with lower soil Cd concentrations both at national and regional level, potentially due to higher likelihood of leaching. Variables with contradicting or inconsistent trends might on the other hand point to either regional differences, or the possibility that these variables serve as proxies for other variables not considered in our analysis. For example, inconsistent trends observed for soil texture variables at national and regional levels might reflect the effect that differences in clay minerology may have. This is masked when simply using percent clay in soil texture analysis. Regardless of the underlying mechanisms, the results clearly point to the complexity of predicting Cd uptake across a country with enormous differences in geology, soils, climate, and topography.

4.1. Soil Cd

The most important variable explaining Cd in soil is geology indicating that for cacao, high levels of cadmium in beans is not manmade but based
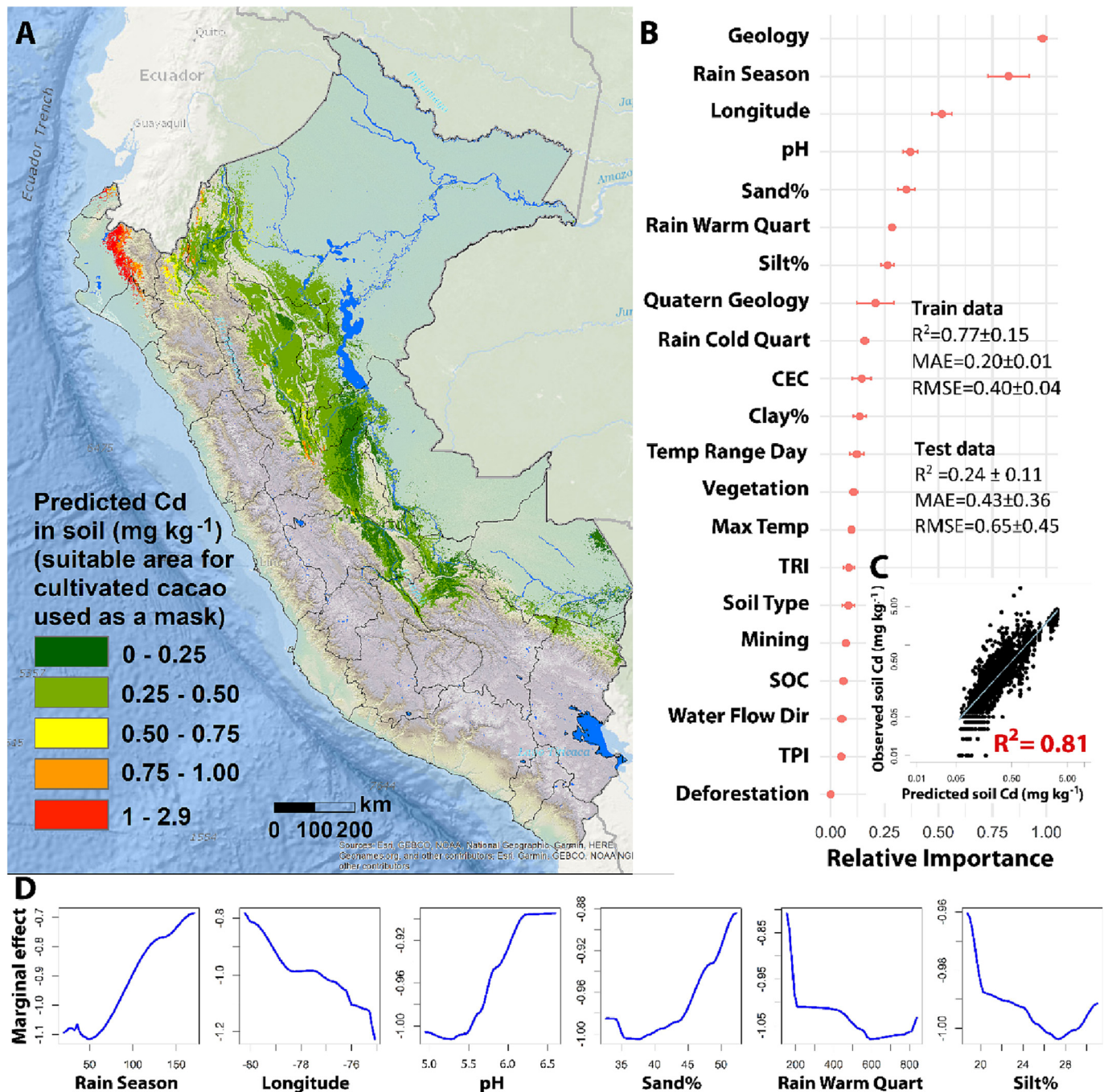
Fig. 3. Predicted soil Cd (mg kg^{-1}) in areas suitable for cacao cultivation (sensu Ceccarelli et al., 2021). The map (A) shows a projected random forest model calibrated with all data points and using the predictors shown in the right-hand-side dot plot, ranked in terms of importance (B). Model performance metrics and variable importances (B) are based on test and train data obtained from 15 cross validations in spatial blocks. A comparison of predicted versus observed soil Cd based on the final random forest models trained using all data for production of map in A yielded an R^2 of 0.81 (C; plotted on log scale axes). Partial dependence curves of the six most important continuous variables are displayed at the bottom (D). For partial dependences of categorical variables please refer to Figs. S1–2; partial dependences curves of all continuous variables are given in Fig. S3.

on natural soil levels. The low usage of fertilizers, even if they are contaminated with cadmium suggests they are not an important input of Cd for farms in Peru (McLaughlin et al., 2021). The importance of the geological formation has been noted previously both in Latin America and the Caribbean (Argüello et al., 2019; Gramlich et al., 2018), and further afield (Birke et al., 2017; Marchant et al., 2010). Our results show that aside from the localized importance of different geological layers, layers that formed during the Quaternary Holocene in continental Peru seem to have a higher likelihood of contributing higher Cd concentrations to soils. These are young layers and may have experienced less leaching. More

broadly, the geological formations associated with elevated soil Cd concentrations in our dataset show a degree of spatial clustering, such that there is a clear gradient of increasing soil Cd concentrations from south to north and east to west. The soils with highest Cd are found in the coastal valleys of Piura and Tumbes in the northwest of the country, an area with very limited and highly seasonal rainfall which could explain why rainfall seasonality is an important predictor of soil Cd models both at national level and for the Tumbes-Piura region. Seasonality of rainfall was an important predictor in other regions too, but in the opposite direction with higher seasonality associated with lower soil Cd. It is therefore likely that seasonality covaries
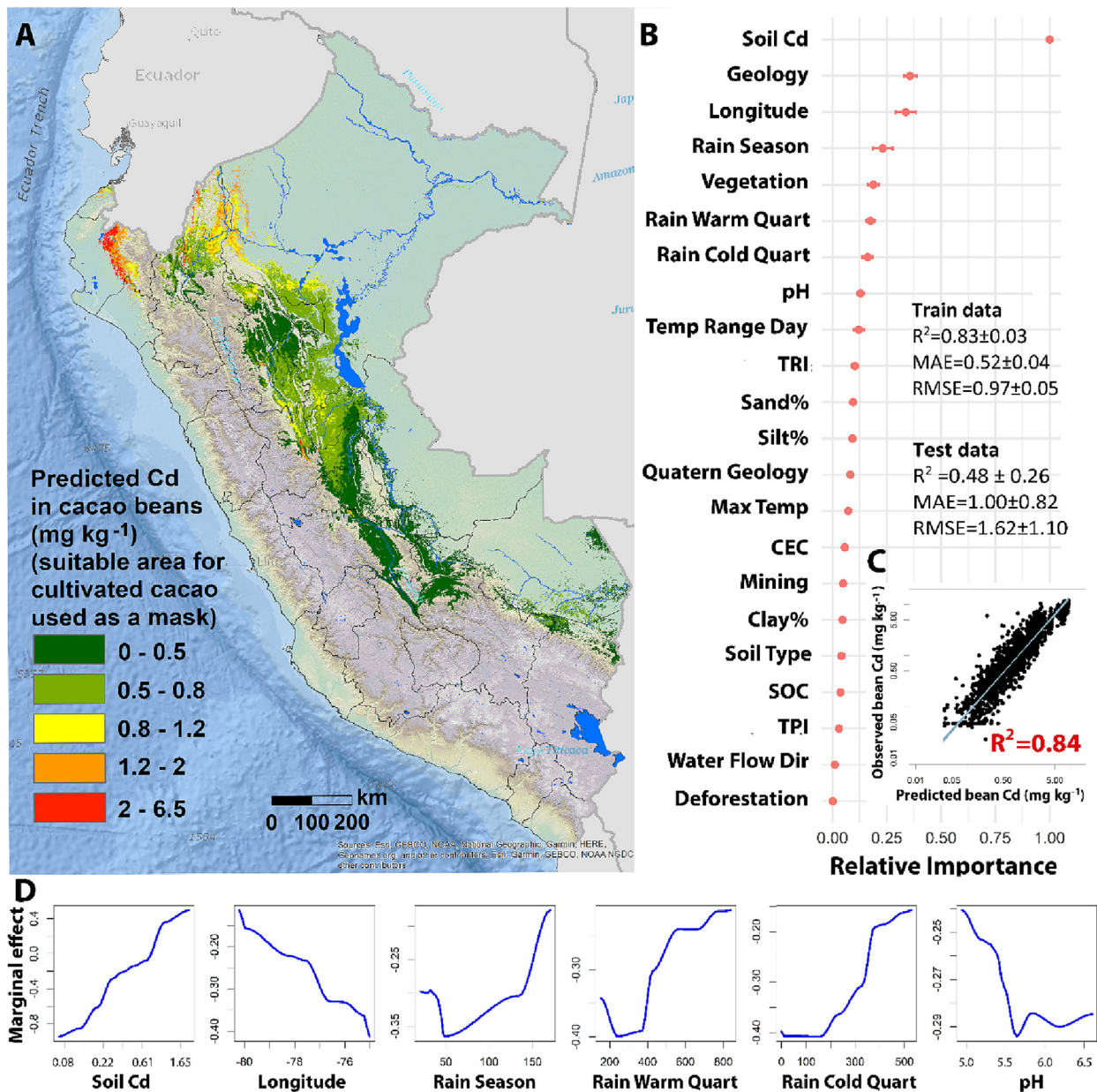
Fig. 4. Predicted Cd in cacao beans in areas suitable for cacao cultivation (sensu Ceccarelli et al., 2021). The map (A) shows a projected random forest model calibrated with all data points and using the predictors shown in the right-hand-side dot plot, ranked in terms of importance (B). Model performance metrics and variable importances (B) are based on test and train data obtained from 15 cross validations in spatial blocks. A comparison of predicted versus observed bean Cd based on the final random forest models trained using all data for production of map in A shows yielded an R^2 of 0.84 (C; plotted on log scale axes). Partial dependence curves of the six most important continuous variables are displayed at the bottom (D). For partial dependences of categorical variables please refer to Figs. S4–5; partial dependences curves of all continuous variables are given in Fig. S6.

with spatial trends in soil type and geology among others. For example, in the highly seasonal department of Cajamarca soil Cd is much lower than in Amazonas department which has a more tropical rainforest climate, in addition to different soil types and underlying geology.

Even though Cd is more available in acidic soils, alkaline soils in our study region are associated with higher Cd. This would appear to be linked to soil formation and weathering, as well as the importance of alluvial soils whose composition reflects their geological origin. Accordingly, the inverse association between rainfall and soil Cd content, both at national and regional level is likely to reflect increased migration of naturally occurring cadmium down the soil profile or through run-off leaching of Cd in soils under stronger rainfall regimes (Kabata-Pendias and Szteke, 2015; Rieuwerts, 2007). The typical neutral to alkaline soils from the Piura-Tumbes region have suffered less leaching than acidic soils typical of the Amazon basin, have a different type of clay (2:1 clays as opposed to 1:1 clays in the Amazon basin, see below) and thus may retain a higher heavy metal composition regardless of their higher soil pH.

The positive relation we found between CEC and total soil Cd both in the nation-wide and regional models, apart from the Piura-Tumbes region where the opposite is found, is expected as soils with a higher cation binding capacity are likely to adsorb Cd more strongly. The opposite relation in the Piura-Tumbes region might be because the lowland soils in this region have a higher percentage of 2:1 clay minerals (e.g. montmorillonite) than tropical soil clays in the other parts of the country. These are rich in cations (especially Magnesium and Sodium), and have a higher pH. The difference in this predictor is thus probably based on soil origin and type rather than anything else. This is discussed in more detail in the next section.
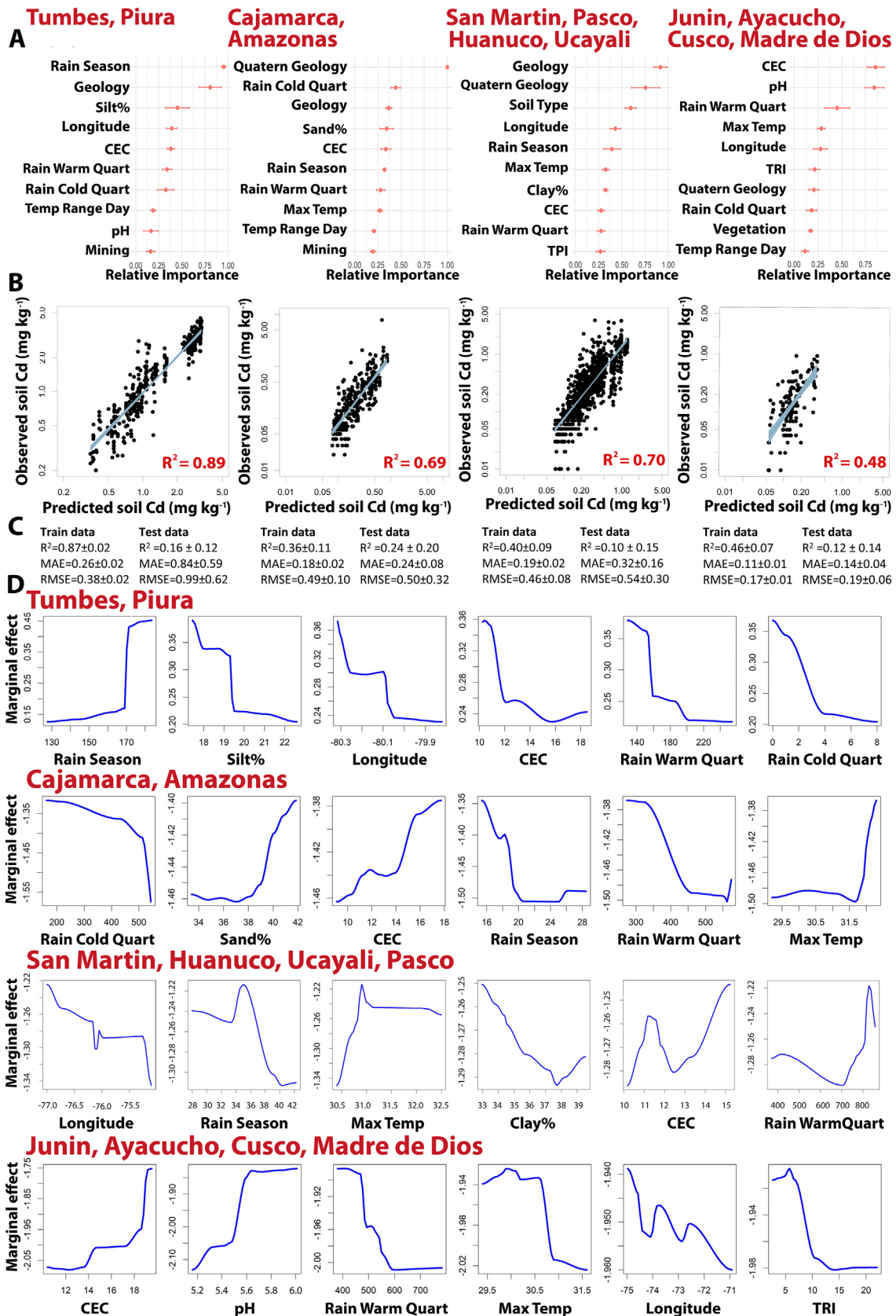
Fig. 5. Variable importance scores (A), comparison of predicted versus observed soil Cd based on final random forest models trained using all data (B; plotted on log scale axes), model performance metrics (C) and partial dependence curves of the six most important continuous variables (D) for each of the four regional random forest models predicting soil Cd (mg kg^{-1}). Variable importance scores and model performance metrics are based on 15 cross validations in spatial blocks. For partial dependences of categorical variables and partial dependence curves of all continuous variables in regional soil Cd models please refer to Figs. S7–17.
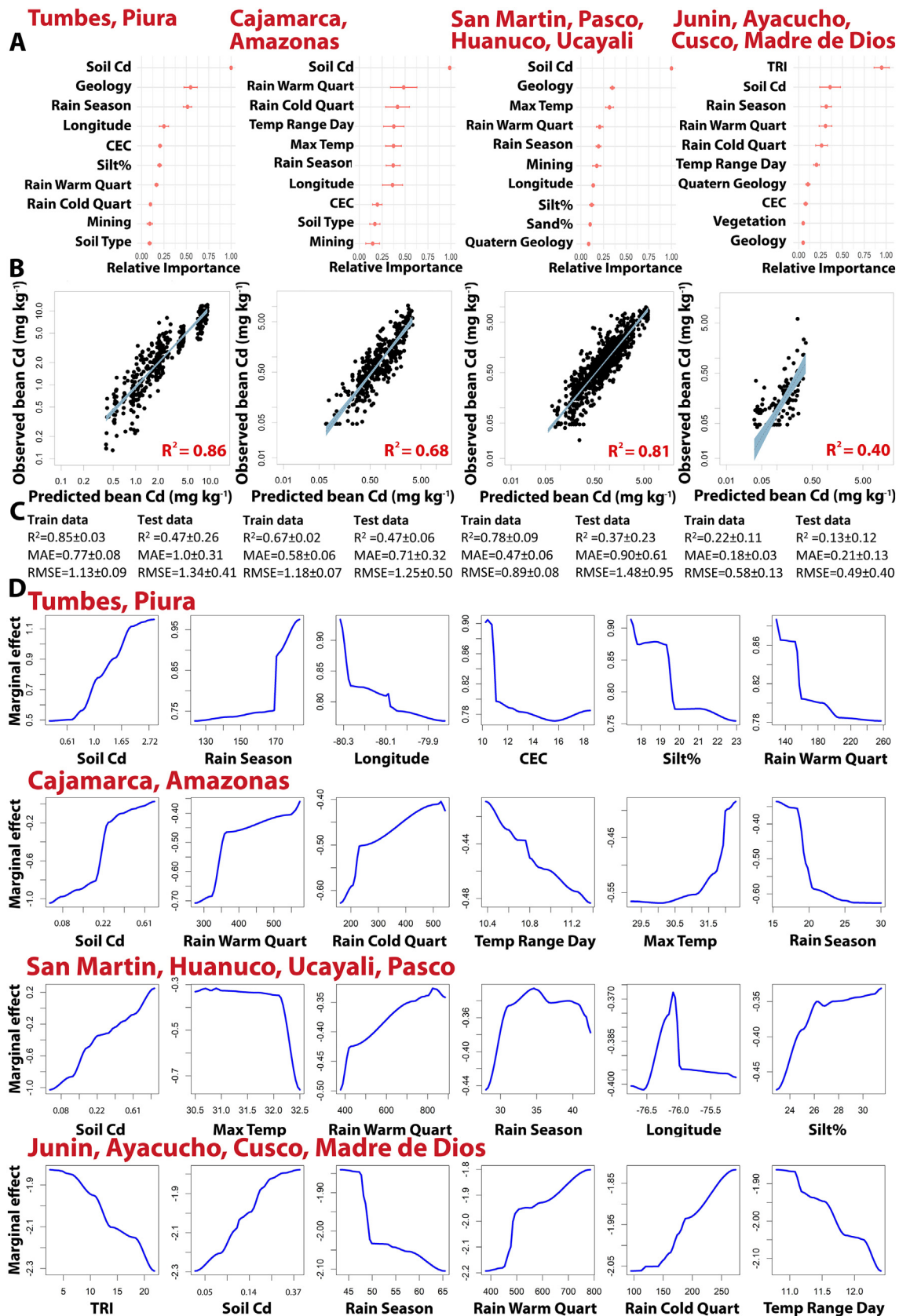
Fig. 6. Variable importance scores (A), comparison of predicted versus observed bean Cd based on final random forest model trained using all data (B; plotted on log scale axes), model performance metrics (C) and partial dependence curves of the six most important continuous variables (D) for each of the four regional random forest models predicting bean Cd (mg kg^{-1}). Variable importance scores and model performance metrics are based on 15 cross validations in sptial blocks. For partial dependences of categorical variables and partial dependence curves of all continuous variables in regional bean Cd models please refer to Figs. S18–26.
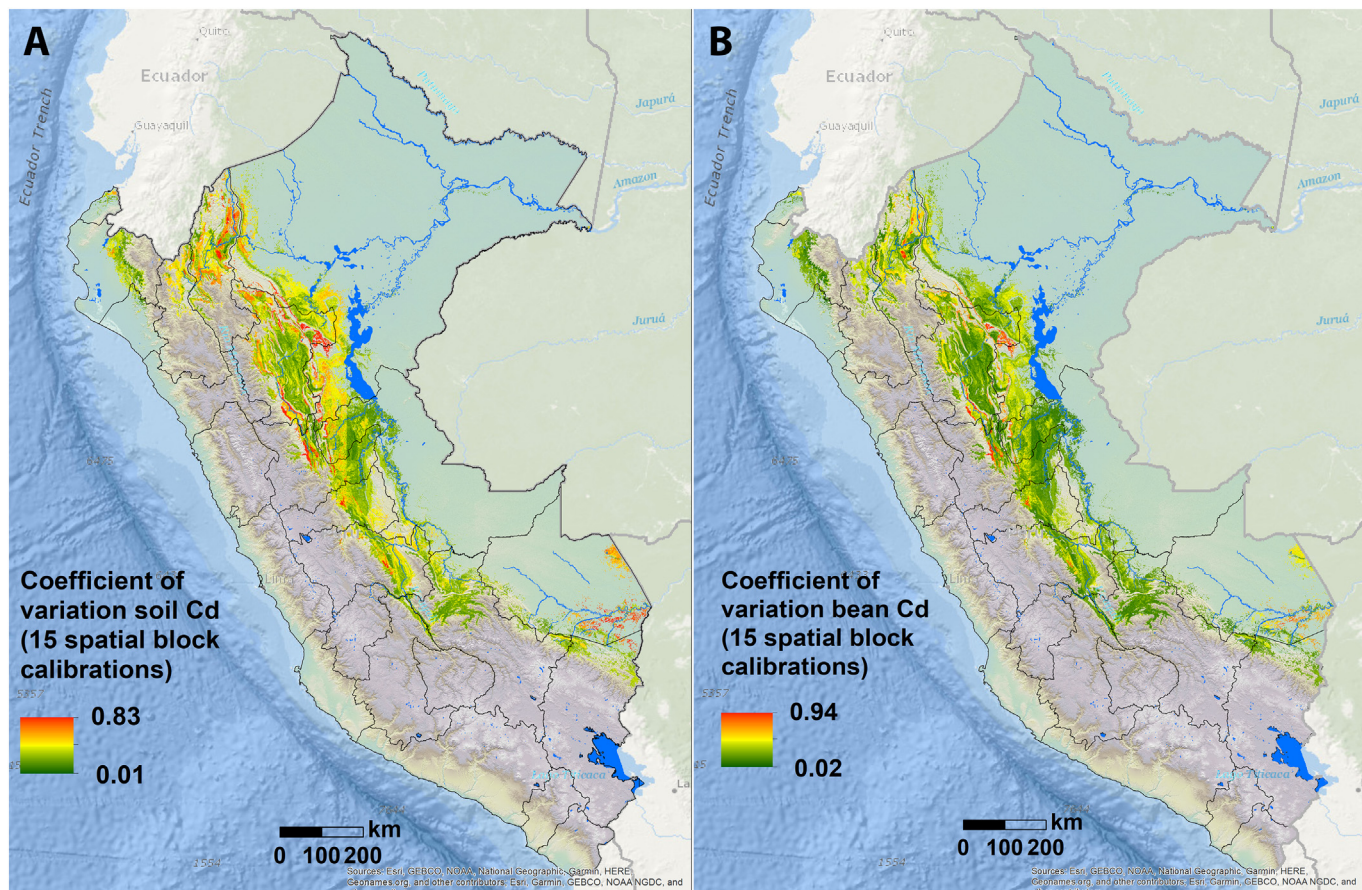
Fig. 7. Uncertainty maps of predicted soil (A) and cacao bean Cd (B), expressed as coefficient of variation calculated based on 15 cross validations in spatial blocks.

4.2. Cacao bean Cd

The availability of Cd to plants in general is complex and depends upon factors such as pH, organic matter content, soil texture and mineralogy, cation exchange capacity, electrical conductivity, macro- and micro-nutrient content and the presence of microorganisms (Adriano, 1986; Correa et al., 2021; McLaughlin et al., 2021; Shahid et al., 2016; Singh et al., 1995). Previous studies tend to agree that key soil properties for cadmium bioavailability and subsequent accumulation in cacao beans include total soil cadmium, pH, organic matter, geological substrate and soil sand content. The effect of clay is inconsistent and factors such as micronutrients and CEC have not been studied in detail yet (Arévalo-Gardini et al., 2019; Argüello et al., 2019; Barraza et al., 2017; Gramlich et al., 2018, 2017).

Unsurprisingly, in this study the most important predictor of bean Cd at a national level is soil Cd, followed by geology, suggesting that the nature of the geological parent material influences both total soil Cd content and its availability. However, bean Cd is not only influenced by the geological layer upon which the tree is growing, but also layers located in upstream areas. Both ancient and more recent alluvial and fluvial deposits, as well as ancient lakes and riverbeds, were associated with higher bean Cd, depending on the nature of the parent material. This suggests that rivers and streams running through areas with high levels of cadmium can deliver not only nutrients but also cadmium and other heavy metals released by weathering of the bed rock or mining to agricultural areas downstream, confirming previous research. Gramlich et al. (2018) suggested that the deposition of sediments from river flooding may be a key source of topsoil
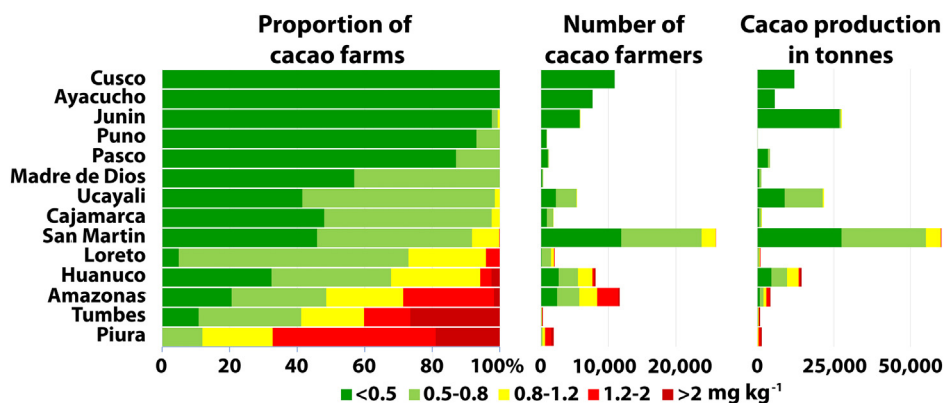

Fig. 8. Proportions of farms within 5 ranges of predicted average cacao bean Cd concentrations (mg kg^{-1}), and the corresponding estimated numbers of farmers and production levels expressed by department in Peru.

cadmium in their study sites in Honduras. In Peru, Llatance et al. (2018) recorded differences in cadmium concentrations in soil samples taken from non-inundated (<0.008 mg kg^{-1}), inundated (0.043 mg kg^{-1}) and semi-inundated soils (0.11 mg kg^{-1}) in which less water is retained but for a longer period of time. A collaborative study in Ecuador led by the French cooperative Ethiquable and the French Research Institute for Development (IRD) similarly found that farms that were regularly flooded by the river had the highest cadmium concentrations in cacao beans (with concentrations reaching 4.3 mg kg^{-1}; Maurice L., pers. com.). In line with this, within the watershed of the Santiago River in the department of Amazonas we found a 3.5-fold difference in Cd translocation factor (Cd bean/Cd soil) comparing farms found in the alluvial plain that are flooded naturally at least every decade with hillside farms that are not.

Water does not have to carry high levels of cadmium to affect plant uptake: several factors such as saline water conditions (high electrical conductivity), and flood-drought cycles can increase the availability of cadmium present in the soil (Singh et al., 1995). While there are no studies looking at the effect of flooding on cadmium accumulation in cacao, in rice this has been studied in detail and there appears to be an interaction between micronutrient, soil type, water content and Cd speciation (de Livera et al., 2011; Rassaei et al., 2020). A flooded soil is anaerobic. Under these conditions, Fe and Mn oxyhydroxides dissolve while metal sulphides precipitate. This results in increased sorption sites for Cd, and precipitation as CdS, in both cases reducing its availability. During drying conditions Fe and Mn oxyhydroxides precipitate, taking them out of solution and resulting in an increased ratio of Cd to Fe and Mn, which, in addition to a decrease in pH (as the sulphides dissolve) results in increased bioavailability of Cd and thus its uptake by the plant.

In the Piura-Tumbes region cacao is grown exclusively under irrigation. Our data show that the Cd load of irrigation water can be substantial but varies depending on the water origin. Leaching of Cd can be exacerbated by mining and land degradation or other operations (Oporto et al., 2007; Smolders et al., 2003; Sun et al., 2010; Yang et al., 2006; Zhai et al., 2008), which could explain why water in the Las Lomas irrigation canal which originates from an active mining area upstream contained four times more Cd than water used for irrigation in La Quemazón which originates from an area with fewer mines and different bedrock. In a study across Ecuador Argüello et al. (2019) similarly found that the bean samples with the highest cadmium concentration (5.28–10.4 mg kg^{-1}) came from a farm in a region with artisanal mining. However, as mentioned above, Cd in irrigation water is only part of the story as Cd bean concentrations in Las Lomas cacao trees were ten times higher than in La Quemazón, despite only a four times higher Cd load in irrigation water. This could be explained by higher levels of soil Cd and by differences in soil characteristics that enhance Cd uptake by the plant.

In accordance with soil Cd, among the remaining most important variables explaining bean Cd are geographical location, rainfall, rainfall seasonality and pH. However, there are some key differences in the direction of the effects. Thus, while there is a positive relation between pH as a predictor and soil Cd, in beans it is negative. This can be explained by the higher solubility of Cd under acidic conditions leading to a higher translocation factor by the cacao plant (McLaughlin et al., 2021). There is a positive relation between rainfall and bean Cd in all regions except for Piura. The underlying explanation may be related to soil pH and CEC. Heavy rains can cause nutrient leaching and result in soil acidification which increases Cd availability and thus uptake. In Piura, with much lower rainfall and generally neutral to alkaline soils, this process is unlikely to happen. The type of clay present is also different: Type 2:1 clays (e.g. montmorillonite) predominate in Piura, and type 1:1 clays (e.g. kaolinite) in the Amazon basin. In the former CEC does not change with pH as it is dependent on the permanent charges of soil particles, while in the latter pH and organic matter content play much more important roles (Solly et al., 2020; Haghiri, 1974)). This means that that 2:1 clays are less prone to acidification under heavy rainfall (Adriano, 1986). It is of note in Piura that there is a negative relation between the predictor soil CEC and Cd in bean. At higher CEC there is higher capacity for soil particle surfaces to retain cations which can lead to a

decrease in cadmium bioavailability. As CEC decreases there is an increased competition between H^+ and Cd^{2+} ions for binding sites which results in cadmium desorption from soil particles into the soil solution.

4.3. Implications and way forward

While the regulatory Cd thresholds are for the final products (chocolate and derivatives), buyers of cacao beans interpret these to inform their purchases. Thus, although the limits in the existing regulations range from 0.3 to 0.8 mg kg^{-1} Cd depending on the product, buyers of beans for chocolate bar production often request concentrations of Cd of <0.8 mg kg^{-1} (Meter et al., 2019). Using this 0.8 mg kg^{-1} threshold, our results show that, although Cd accumulation in cacao beans is of serious concern in some specific locations, affecting >89 % of farms in Piura for example, in general cacao production and cacao farmers in Peru are not expected to suffer significant economic losses due to Cd regulations imposed by the European Union and other markets.

The production of cacao in Peru has two main markets: bulk cocoa and fine flavour cocoa. Bulk cacao is mainly used for the manufacturing of cocoa butter and powder, as well as high-volume mainstream chocolate products and is usually mixed with cocoa from other countries prior to production of the final product. This makes up the majority of Peru's cacao production, focussed in the centre and south of the country (Huánuco, San Martin, Ucayali, Pasco). Bulk cacao is further divided into organic and conventional, with 37 % of Peruvian exports to the EU being organic (9600 tonnes). While the organic bulk cacao market is not origin-specific, it is likely to be highly sensitive to Cd levels. It is interesting to note that the EU has dramatically increased imports of organic cacao from Sierra Leone since 2020, a country with no known issues of Cd accumulation in beans (Profound, 2022). Fine flavour cacao is a niche product where varieties with interesting flavours are sold directly to chocolate makers. Here origin plays a key role in marketing due to unique flavour profiles which are linked to a specific cacao cultivar or varieties, the particular environmental conditions where it is grown, or both. Not only are flavour profiles of interest, but buyers usually require organic and fairtrade certification. Fine flavour cacao in Peru can garner a higher price than conventional bulk cacao, and is thus an important option for small-scale farmers (Tschartntke et al., 2023). To date, the most famous fine flavour cacaos come from Piura, Amazonas and Cusco, but there are many more entering the market (Villar et al., 2022).

As the mapping presented here indicates, the location of a farm is key to predicting levels of cadmium in beans. For bulk cacao, a buyer sources beans across a wide geographical area, so that cadmium thresholds can be met through mixing of the beans from high and low Cd areas. For example, Villar et al. (2022) show that in Huánuco, farmers to date have not reported economic losses due to cadmium levels even though Cd bean concentration is high in farms located in alluvial soils along the Huallaga river. We assume this is due to the mixing of this cacao with beans produced further away from the river. This is despite some buyers requiring maximum concentrations of as low as 0.3 mg kg^{-1}, an interpretation of the regulatory limit for Cd in cocoa powder for final consumption. Hence the 0.5 mg kg^{-1} limit commonly used by bulk cacao buyers is unlikely to impact this market segment significantly. However, it waits to be seen if bulk cacao for the organic market has been affected. In contrast, Piura is home to a highly prized cultivar (cacao Blanco de Piura) which has a unique flavour profile that even varies by valley within the department. It is in high demand by fine flavour cacao buyers and bean to bar chocolate makers and typically fetches prices well above global stock market rates. However, owing to the high bean Cd values in most of Piura, many farmers are no longer able to sell their cacao to those high-profit markets, and are forced to sell their cacao to national markets, resulting in moderate to severe income loss to farmers with limited resources and disintegration of associations (Villar et al., 2022).

The predictive maps we present here are available through the website www.cacacodiversity.org. This allows users to obtain (ranges of) measured values of Cd levels in beans and soil in grid cells where these are available,

and predicted values in other cells within the suitable areas of cultivated cacao calculated using random forest models calibrated on all observational data. For predicted values we provide 95 % confidence intervals calculated from predicted values of each of the 15 cross-validation models in spatial blocks. While the maps have reasonable predictive power, they are limited to the layers available at a national level, which also restricts our understanding of the issue. First the explanatory variables have uncertainty issues of their own (Fick and Hijmans, 2017; Hengl et al., 2017) with unclear implications on soil and bean Cd predictions. Second, the spatial resolution of the explanatory variable rasters (here ~5.4 ha) is generally lower than that of the cacao plantations and hence grid cell values represent an average of the different land uses covered by it. For example, a single grid cell may contain old growth forest and crop land with marked differences in SOC (Duarte-Guardia et al., 2020). Third, the explanatory variables could further be improved through the inclusion of important predictors of bean Cd such as key macro and micronutrients which are currently not available as spatial data layers. It should also be acknowledged that the map and analysis has been carried out at regional scale and we are aware of substantial variation even within a farmer's field. This may be in part due to variation in farm management and genetic makeup of the cacao (Lewis et al., 2018). However, we believe that the maps presented here are an important step in understanding the origin of cadmium accumulation in cacao, and add to the increasing volume of information available to enable better decision making to help actors in the cacao value chain adapt to the regulations for Cd, and thus minimize its impact on small scale rural communities.

CRediT authorship contribution statement

Evert Thomas: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rachel Atkinson:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Diego Zavaleta:** Data curation, Investigation. **Carlos Rodriguez:** Investigation. **Sphyros Lastra:** Data curation, Investigation. **Fredy Yovera:** Investigation. **Karina Arango:** Data curation, Investigation. **Abel Pezo:** Investigation. **Javier Aguilar:** Investigation, Project administration, Resources, Validation. **Miriam Tames:** Investigation, Resources. **Ana Ramos:** Investigation, Resources. **Wilbert Cruz:** Investigation, Project administration, Resources. **Roberto Cosme:** Project administration, Resources. **Eduardo Espinoza:** Investigation. **Carmen Rosa Chavez:** Conceptualization. **Brenton Ladd:** Investigation, Resources, Writing – review & editing.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.163372.

References

Abbott, P., Benjamin, T., Burniske, G., Croft, M., Fenton, M., Colleen, R., Lundy, M., Rodriguez Camayo, F., Wilcox, M., 2018. An Analysis of the Supply Chain of Cacao in Colombia. United States Agency for International Development.

Adriano, D., 1986. Trace Elements in the Terrestrial Environment. 1st ed. Springer, New York.

Arévalo-Gardini, E., Arévalo-Hernández, C.O., Baligar, V.C., He, Z.L., 2019. Heavy metal accumulation in leaves and beans of cacao (*Theobroma cacao* L.) in major cacao growing regions in Peru. Sci. Total Environ. 606, 792–800.

Argüello, David, Chavez, E., Lauryssen, F., Vanderschueren, R., Smolders, E., Montalvo, D., 2019. Soil properties and agronomic factors affecting cadmium concentrations in cacao beans: a nationwide survey in Ecuador. Sci. Total Environ. 649, 120–127. https://doi.org/10.1016/j.scitotenv.2018.08.292.

Barraza, F., Schreck, E., Lévêque, T., Uzu, G., López, F., Ruales, J., Prunier, J., Marquet, A., Maurice, L., 2017. Cadmium bioaccumulation and gastric bioaccessibility in cacao: a field study in areas impacted by oil activities in Ecuador. Environ. Pollut. 229, 950–963. https://doi.org/10.1016/j.envpol.2017.07.080.

Birke, M., Reimann, C., Rauch, U., Ladenberger, A., Demetriades, A., Jähne-Klingberg, F., Oorts, K., Gosar, M., Dinelli, E., Halamić, J., 2017. GEMAS: cadmium distribution and its sources in agricultural and grazing land soil of Europe — original data versus clr-transformed data. J. Geochem. Explor. 173, 13–30. https://doi.org/10.1016/j.gexplo.2016.11.007.

Bravo, D., Leon-Moreno, C., Martínez, C.A., Varón-Ramírez, V.M., Araujo-Carrillo, G.A., Vargas, R., Quiroga-Mateus, R., Zamora, A., Rodríguez, E.A.G., 2021. The first national survey of cadmium in cacao farm soil in Colombia. Agronomy 11, 761. https://doi.org/10.3390/agronomy11040761.

Ceccarelli, V., Fremout, T., Zavaleta, D., Lastra, S., Imán Correa, S., Arévalo-Gardini, E., Rodriguez, C.A., Cruz Hilacondo, W., Thomas, E., 2021. Climate change impact on cultivated and wild cacao in Peru and the search of climate change-tolerant genotypes. Divers. Distrib. 27, 1462–1476. https://doi.org/10.1111/ddi.13294.

Chavez, E., Superior, E., Esp, L., Li, Y., 2015. Concentration of cadmium in cacao beans and its relationship with soil cadmium in southern Ecuador. Sci. Total Environ. 533, 204–214. https://doi.org/10.1016/j.scitotenv.2015.06.106.

Correa, J., Ramírez, R., Ruíz, O., Leiva, E., 2021. Effect of soil characteristics on cadmium absorption and plant growth of Theobroma cacao L. seedlings. J. Sci. Food Agric. 101. https://doi.org/10.1002/jsfa.11192.

Davison, W., Zhang, H., 1994. In situ speciation measurements of trace components in natural waters using thin-film gels. Nature 367, 546–548. https://doi.org/10.1038/367546a0.

Davison, W., Zhang, H., 2016. Principles of measurements in simple solutions. In: Davison, W. (Ed.), Diffusive Gradients in Thin-films for Environmental Measurements. 2021. Proficiency Test on the Determination of Total Cd Concentration in Cacao Samples: Final Report as Part of Project STDF/PG/681 and the Clima-LoCa Project. Cambridge University Press, Cambridge Dekeyrel, J. Leuven, Belgium.

de Livera, J., McLaughlin, M.J., Hettiarachchi, G.M., Kirby, J.K., Beak, D.G., 2011. Cadmium solubility in paddy soils: effects of soil oxidation, metal sulfides and competitive ions. Sci. Total Environ. 409, 1489–1497. https://doi.org/10.1016/j.scitotenv.2010.12.028.

Dekeyrel, J., 2021. Proficiency Test on the Determination of Total Cd Concentration in Cacao Samples: Final Report As part of project STDF/PG/681 and the Clima-LoCa project Leuven, Belgium.

DEVIDA, 2017. Estrategia Nacional de Lucha contra las Drogas 2017–2021. Comisión Nacional para el Desarrollo y Vida sin Drogas, Peru.

Duarte-Guardia, S., Peri, P., Amelung, W., Thomas, E., Borchard, N., Baldi, G., Cowie, A., Ladd, B., 2020. Biophysical and socioeconomic factors influencing soil carbon stocks: a global assessment. Mitig. Adapt. Strateg. Glob. Chang. https://doi.org/10.1007/s11027-020-09926-1.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. https://doi.org/10.1002/joc.5086.

Global Forest Watch (2022) https://www.globalforestwatch.org. Consulted May 2022

Gramlich, A., Tandy, S., Andres, C., Chincheros Paniagua, J., Armengot, L., Schneider, M., Schulin, R., 2017. Cadmium uptake by cocoa trees in agroforestry and monoculture systems under conventional and organic management. Sci. Total Environ. 580, 677–686. https://doi.org/10.1016/j.scitotenv.2016.12.014.

Gramlich, A., Tandy, S., Gauggel, C., López, M., Perla, D., Gonzalez, V., Schulin, R., 2018. Soil cadmium uptake by cocoa in Honduras. Sci. Total Environ. 612, 370–378. https://doi.org/10.1016/j.scitotenv.2017.08.145.

Haghiri, F., 1974. Plant uptake of cadmium as influenced by cation exchange capacity, organic matter, zinc, and soil temperature. J. Environ. Qual. 3, 180–183. https://doi.org/10.2134/jeq1974.00472425000300020021x.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0169748.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.

Hijmans, R.J., 2019. Introduction to the 'raster' Package (Version 2.9-5). pp. 1–26.

HuamanI-Yupanqui, H.A., Huauya-Rojas, M.Á., Mansilla-Minaya, L.G., Florida-Rofner, N., Neira-Trujillo, G.M., 2012. Presencia de metales pesados en cultivo de cacao (Theobroma cacao L.) orgánico. Acta Agron. 61, 339–344.

Kabata-Pendias, A., Szteke, B., 2015. Trace Elements in Abiotic and Biotic Environments. CRC Press.

Lewis, C., Lennon, A.M., Eudoxie, G., Umaharan, P., 2018. Genetic variation in bioaccumulation and partitioning of cadmium in Theobroma cacao L. Sci. Total Environ. 640–641, 696–703. https://doi.org/10.1016/j.scitotenv.2018.05.365.

Llatance, W.O., Gonza Saavedra, C.J., Guzmán Castillo, W., Pariente Mondragón, E., 2018. Bioacumulación de cadmio en el cacao (Theobroma cacao) en la Comunidad Nativa de Pakun, Perú. Rev. For. Perú 33, 63. https://doi.org/10.21704/rfp.v33i1.1156.

Marchant, B., Saby, N., Lark, R., Bellamy, P., Jolivet, C., Arrouays, D., 2010. Robust analysis of soil properties at the national scale: cadmium content of French soils. Eur. J. Soil Sci. 61, 144–152.

McLaughlin, M.J., Smolders, E., Zhao, F.J., Grant, C., Montalvo, D., 2021. Managing cadmium in agricultural systems. Advances in Agronomy. Elsevier, pp. 1–129 https://doi.org/10.1016/bs.agron.2020.10.004.

Meter, A., Atkinson, R., Laliberte, 2019. Cadmium in Cacao From Latin America and the Caribbean. A Review of Research and Potential Mitigation Solutions. CAF, Caracas.

MIDAGRI, 2022. Perfil productivo y competitivo de los principales cultivos del sector.

Oporto, C., Vandecasteele, C., Smolders, E., 2007. Elevated cadmium concentrations in potato tubers due to irrigation with river water contaminated by mining in Potosí, Bolivia. J. Environ. Qual. 36, 1181–1186. https://doi.org/10.2134/jeq2006.0401.

Profound, 2022. Entering the European Market for Organic Cocoa. CBI Ministry of Foreign Affairs, The Netherlands.

Ramtahal, G., Yen, I.C., Bekele, I., Bekele, F., Wilson, L., Maharaj, K., Harrynanan, L., 2016. Relationships between cadmium in tissues of cacao trees and soils in plantations of Trinidad and Tobago. Food Nutr. Sci. 07, 37–43. https://doi.org/10.4236/fns.2016.71005.

Rassaei, F., Hoodaji, M., Abtahi, S.A., 2020. Cadmium speciation as influenced by soil water content and zinc and the studies of kinetic modeling in two soils textural classes. Int. Soil Water Conserv. Res. 8, 286–294. https://doi.org/10.1016/j.iswcr.2020.05.003.

Remigio, J., 2014. Determinacion de procedimientos, interpretacion de resultados de analisis y elaboracion de interrelaciones de los diferentes estudios para determinar la concentracion de cadmio en los granos de cacao. Central Piurana de Cafetaleros CEPICAFE, Piura.

Rieuwerts, J.S., 2007. The mobility and bioavailability of trace metals in tropical soils: a review. Chem. Speciat. Bioavailab. 19, 75–85.

Scaccabarozzi, D., Castillo, L., Aromatisi, A., Milne, L., Castillo, A.B., Muñoz-Rojas, M., 2020. Soil, site, and management factors affecting cadmium concentrations in cacao-growing soils. Agronomy 10. https://doi.org/10.3390/agronomy10060806.

Shahid, M., Dumat, C., Khalid, S., Niazi, N.K., Antunes, P.M.C., 2016. Cadmium bioavailability, uptake, toxicity and detoxification in soil-plant system. In: de Voogt, P. (Ed.), Reviews of Environmental Contamination and Toxicology. Reviews of Environmental Contamination and Toxicologyvol. 241. Springer International Publishing, Cham, pp. 73–137. https://doi.org/10.1007/398_2016_8.

Singh, B.R., Narwal, R.P., Jeng, A.S., Almas, Å., 1995. Crop uptake and extractability of cadmium in soils naturally high in metals at different pH levels. Commun. Soil Sci. Plant Anal. 26, 2123–2142. https://doi.org/10.1080/00103629509369434.

Smolders, A.J.P., Lock, R.A.C., Van der Velde, G., Medina Hoyos, R.I., Roelofs, J.G.M., 2003. Effects of mining activities on heavy metal concentrations in water, sediment, and macroinvertebrates in different reaches of the Pilcomayo River, South America. Arch. Environ. Contam. Toxicol. 44, 0314–0323. https://doi.org/10.1007/s00244-002-2042-1.

Solly, E.F., Weber, V., Zimmermann, S., Walthert, L., Hagedorn, F., Schmidt, M.W.I., 2020. A critical evaluation of the relationship between the effective cation exchange capacity and soil organic carbon content in Swiss forest soils. Front. For. Glob. Change 3.

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinf. 8, 25.

Sun, L.-N., Zhang, Y.-F., He, L.-Y., Chen, Z.-J., Wang, Q.-Y., Qian, M., Sheng, X.-F., 2010. Genetic diversity and characterization of heavy metal-resistant-endophytic bacteria from two copper-tolerant plant species on copper mine wasteland. Bioresour. Technol. 101, 501–509. https://doi.org/10.1016/j.biortech.2009.08.011.

Tantalean Pedraza, E., Huauya Rojas, M.Á., 2017. Distribución del contenido de cadmio en los diferentes órganos del cacao CCN-51 en suelo aluvial y residual en las localidades de Jacintillo y Ramal de Aspuzana. Rev. Investig. Agroproducción Sustentable 1, 69. https://doi.org/10.25127/aps.20172.365.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., van Etten, J., 2012. Present spatial diversity patterns of Theobroma cacao L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. PLoS ONE 7, e47676. https://doi.org/10.1371/journal.pone.0047676.

Tscharntke, T., Ocampo-Ariza, C., Vansynghel, J., Ivañez-Ballesteros, B., Aycart, P., Rodriguez, L., Ramirez, M., Steffan-Dewenter, I., Maas, B., Thomas, E., 2023. Socioecological benefits of fine-flavor cacao in its center of origin. Conserv. Lett. 16, e12936. https://doi.org/10.1111/conl.12936.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: an r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232. https://doi.org/10.1111/2041-210X.13107.

Vanderschueren, R., Argüello, D., Blommaert, H., Montalvo, D., Barraza, F., Maurice, L., Schreck, E., Schulin, R., Lewis, C., Vasquez, J.L., Pathmanathan, U., Chavez, E., Sarret, G., Smolders, E., 2021. Mitigating the level of cadmium in cacao products: reviewing the transfer of cadmium from soil to chocolate bar. Sci. Total Environ. 781, 146779.

Villar, G., Yovera, F., Pezo, A., Thomas, E., Roscioli, F., Sandy da Cruz, R., Jimenez, E., Lopez, A., Aguilar, F., Espinoza, E., Davila, C., Chavez Hurtado, C., Lastra, S., Zavaleta, D., Charry, A., Atkinson, R., 2022. Caracterización socioeconómica de las cadenas de valor de cacao con énfasis en la problemática de cadmio en Piura y Huánuco, Perú. Alianza Bioversity & CIAT, Lima Peru Available at: https://cgspace.cgiar.org/handle/10568/125328.

Wade, J., Ac-Pangan, M., Favoretto, V.R., Taylor, A.J., Engeseth, N., Margenot, A.J., 2022. Drivers of cadmium accumulation in Theobroma cacao L. beans: a quantitative synthesis of soil-plant relationships across the Cacao Belt. PLoS ONE 17, e0261989. https://doi.org/10.1371/journal.pone.0261989.

Yang, Q.W., Lan, C.Y., Wang, H.B., Zhuang, P., Shu, W.S., 2006. Cadmium in soil–rice system and health risk associated with the use of untreated mining wastewater for irrigation in Lechang, China. Agric. Water Manag. 84, 147–152.

Zhai, L., Liao, X., Chen, T., Yan, X., Xie, H., Wu, B., Wang, L., 2008. Regional assessment of cadmium pollution in agricultural lands and the potential health risk related to intensive mining activities: a case study in Chenzhou City, China. J. Environ. Sci. 20, 696–703.

Zug, K.L.M., Huamaní Yupanqui, H.A., Meyberg, F., Cierjacks, J.S., Cierjacks, A., 2019. Cadmium accumulation in Peruvian cacao (Theobroma cacao L.) and opportunities for mitigation. Water Air Soil Pollut. 230, 1–18. https://doi.org/10.1007/s11270-019-4109-x.