Statistical impact of the cadmium forms on element content in cocoa bean in Colombia

Heidy Soledad Rodríguez Albarracín
Universidad Nacional de Colombia

Aquiles Enrique Darghan (aedarghanco@unal.edu.co)
Universidad Nacional de Colombia

Carlos Armando Rivera Moreno
Universidad Nacional de Colombia

Research Article

Keywords: Phytoavailable, speciation, spatial regression, residual phases

Posted Date: June 26th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3016403/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Cocoa is a staple crop for small farmers, has been promoted as an alternative crop to *Erythroxylum coca*. Some soils in the Colombian cocoa zone seem to accumulate high levels of Cd in the beans. To comprehend the relationship between different Cd forms and their phytoavailability in the beans, cocoa-producing farms were sampled located in Yacopi municipality using sequential extraction methods to estimate the forms of Cd present in the soil, its phytoavailability and leaching potential. The exchangeable fractions, carbonate, manganese oxide, organic matter, amorphous iron oxide, crystalline iron oxide and residual fraction were determined. It was observed that the highest Cd concentrations were found in the organic matter, carbonate, Mn oxides and amorphous Fe phases, and the lowest in the exchangeable, crystalline Fe oxide and residual fractions. It was obtained through spatial regression that the Cd fractions that explain the Cd contents in beans correspond to the exchangeable ones, associated with manganese oxides, organic matter, crystalline iron oxides and residual fraction, which were fitted to a spatial error model. Modeling indicated that direct and positive impacts of the Cd content in the bean with the exchangeable and residual fractions, and negative impacts with the organic matter and manganese oxides fractions, which shows that these forms are retained in the colloidal phase and are not translocated to the bean, so that variations in the content of these fractions could contribute to a reduction in the Cd content in the bean.

Introduction

Cacao (*Theobroma cacao*) is a basic crop for the livelihood of more than 8 million smallholder farmers, which has been qualified as a commercial crop alternative to *Erythroxylum coca* (used to manufacture cocaine), helping the conservation and restoration of the tropical rainforest (Wade et al., 2022) and the development of the peace process in Colombia. However, cacao trees (as well as willow and poplar) have a capacity to extract higher amounts of Cd from the soil compared to other plants in general (Gramlich et al. 2018, Barraza et al., 2019), accumulate it in their beans and transfer it to chocolate and other derived products. As reported by Gramlich et al. (2018) and Correa et al. (2021), generally Latin American countries contain higher values of Cd in dried cocoa beans than those in West Africa and since the European Union and Codex Alimentarius (CODEX) have established limits for Cd in chocolate products, cocoa bean exports in Latin America may be limited by the levels of Cd present. For the food industry for example, Contreras et al. (2012) report that Cd cannot exceed 0.5 mg kg\(^{-1}\) and the maximum allowable limits in cereal grains (including chocolate) according to the World Health Organization and the Food and Agriculture Organization of the United Nations (FAO) are 0.1 mg kg\(^{-1}\) (Berg & Licht, 2002). However, the European Union, which imports about half of the world's cocoa production, defined in 2019 a maximum permissible Cd content in cocoa products of 0.1–0.8 mg kg\(^{-1}\), this upper limit being the maximum content in cocoa solids, especially affecting fine cocoa produced in Latin America that has a high solid cocoa content (Wade et al., 2022). In addition, some Colombian soils in the area of the municipality of Yacopí produce cocoa beans with notably high Cd levels (above 3 mg kg\(^{-1}\)) that difficult their commercialization (Rodríguez et al., 2019). Therefore, it is necessary to adequately understand which are the soil attributes...
and/or properties that interfere in the accumulation of heavy metals such as Cd in cocoa beans used for chocolate processing.

Cd, like other metals present in soils, can be found in various chemical forms that interfere with its phytoavailability, because total Cd represents less than 50% of the Cd adsorbed by the plant (Correa et al., 2021, Duplay et al., 2014, Udovic & McBride, 2012). The distribution of an element between various fractions depends on the physical, chemical, mineralogical and biological properties of the soil, so it is important to characterize the Cd concentrations associated with different soil fractions and to recognize those that correspond to bioavailable forms (Martínez & Rivero, 2005), in order to identify the ecological risk associated with Cd contamination (Shahid et al., 2016), since the chemical speciation is key to control the destination as the toxicity of this element in the soil-plant system (Landrot et al., 2012).

Different studies indicate that contaminants such as Cd are initially delayed, but after a period active ingredients in the soil cause it to react and be released back into the environment (Kotresha et al., 2021). For Cd, a susceptibility to periodic and prolonged saturation in soil has been reported (Moghal et al., 2022). In most soils, 99% of total Cd is associated with soil colloids, occurring as cationic (CdHS\(^+\), CdOH\(^+\), CdHCO\(_3\)^\(^+\), CdCl\(^+\)), and anionic (Cd(HS)\(_4\)^\(^-\), Cd(OH)\(_3\)^\(^-\), Cd(OH)\(_4\)^\(^2-\), CdCl\(_3\)^\(^-\)) (Kabata-Pendias & Sadurski, 2004) species. At low pH, Cd in soil solution is predominantly present as Cd\(^{2+}\), CdSO\(_4\) or CdCl\(^+\), whereas in alkaline solutions the less bioavailable species CdHCO\(_3\)^\(^+\), CdCO\(_3\) or CdSO\(_4\) predominate (Sauvé et al., 2000). Being clear that the acidity of the soil solution controls Cd solubility, therefore, in acid soils, solubility is mainly controlled by organic matter and Fe and Mn oxides and hydroxides (Jiao et al., 2012), whereas in alkaline soils with pH above 7.5, Cd is not easily mobilized and is adsorbed and precipitated, as CdCO\(_3\), and possibly Cd\(_3\)(PO\(_4\))\(_2\) (Kabata-Pendias, 2011).

In this soil-plant relationship, Cd does not have a specific cellular entry pathway, it enters the plant through root uptake accidentally transported by specific and non-specific transporters of essential elements such as Fe\(^{2+}\), Ca\(^{2+}\), Zn\(^{2+}\), Cu\(^{2+}\) and Mg\(^{2+}\) (Mendoza-Cózatl et al., 2011, Eller & Bir, 2019, Correa et al., 2021, Shaaria et al., 2021). Calcium (Ca) and Cd compete for the same Ca transport channels (Shaaria et al., 2021), due to their chemical similarity, since Ca has a similar diameter and valence to Cd and therefore a low Ca concentration increases Cd uptake, due to lower Ca competition and ionic activity at root uptake sites (Eller & Bir, 2019).

Estimation of Cd forms in soils has been carried out by individual extractions and sequential extractions. Sequential extraction divides the total Cd content into different operational sets or fractions (Bacon & Davidson, 2008, Rao et al., 2008), which allows the evaluation of the contribution of these different fractions in soil sorption and contributes to the understanding of Cd binding mechanisms (Kotresha et al., 2021), considering that fractionation estimates the association of the metal of interest (Cd) with other chemical species present in the soil, which favors the understanding of its bioavailability and potential leaching (Kotresha et al., 2021, Martínez & Rivero, 2005). In general, this technique provides information on the origin, mode of occurrence, biological and physicochemical availability, mobilization and transport of metals in the soil (Mohammed & Moghal, 2016).
Within these fractions there are soluble forms, such as water soluble, “exchangeable” and acid soluble that have been related to the Cd phytoavailability (Kosolsaksakul et al., 2014, Sungur et al., 2014). The “residual fraction” corresponds to the Cd that is part of the silicates, inherited from or occluded within parent materials; this fraction is considered to have low phytoavailability (Nogueira et al., 2010, Aikpokpodion et al., 2012). For Mohamed et al. (2010), there is a fraction associated with organic matter, which includes binding to fulvic and humic acids which have an affinity to chelate Cd (Bernal et al., 2007). Such a fraction may be of great importance because soils cultivated with cocoa are characterized by an accumulation of organic matter, which can be related to the presence of Cd in the upper layers, mainly due to the decomposition of leaf litter, woody tissues from pruning, dead roots and residues of harvesting (Beer, 1988, Wood & Lass, 2008). Recent research using stable isotopes of Cd has shown that Cd bioaccumulated into the cocoa leaves become the “litter” which falls on the soil below trees and enrich the total Cd and heavy stable isotopes in the surface soil layer (Barraza et al., 2019).

When Cd is associated with anthropogenic activities due to its artificial incorporation into the soil, there are higher contents in the soluble fractions, whether it comes from the parent material, since higher contents can be found in the residual fraction (Rao et al., 2008, Nogueira et al., 2010).

In the present research, the traditional fractionation of Cd in soil samples from cocoa-producing farms was carried out to characterize the chemical forms of cadmium in the soils of Yacopí. Due to the relationship of relatively high concentrations of Cd in the soil and in cocoa plants, it is necessary to define those forms that are highly phytoavailable for cocoa cultivation, because the knowledge of Cd speciation informs soil remediation studies (Bade, 2012).

Therefore, to understand this relationship of the different Cd forms with the contents of the element in cocoa beans, a statistical analysis was developed based on some spatial regression models using the taxonomy of Elhorst (2014), in which case Cd in beans was used as a response variable and the contents of the different fractions were used as explanatory variables. Taking into account that the hyperaccumulative character of Cd presented by cocoa trees may be due to the greater availability of the element in the soil, due to the recycling of Cd from leaf litter and pod husks that reach the soil, which when rapidly decomposed in tropical climates, becomes a source of easily available Cd in the most superficial horizons of the soil, where the fine roots of cocoa trees are located (Gramlich et al. 2018), whose root hairs are the most active ion absorption zones (Song et al., 2016). It is to be expected that the forms associated with organic matter have an important participation in the Cd contents that are translocated to the beans. This modeling strategy based on spatial regression models allows finding positive or negative impacts of the different fractions on Cd accumulation in beans, which could suggest that a change in any of these fractions in the soil could remediate its accumulation in cocoa beans.

**Materials and methods**

**Location**
The study was carried out in the municipality of Yacopí, department of Cundinamarca (Colombia), in four farms with high Cd contents in cocoa beans, plus a control farm, with values below 3 mg kg$^{-1}$ (Fig. 1), selected from the global diagnosis carried out by Soler (2017). The definition of the unit of soils, relief and parental material for each of the farms was taken from the study of IGAC (2000), from which it is observed that farms Y1 and Y2 are located in soil units corresponding to the Typic Udorthents-Typic Dystrudepts association corresponding to mountain soils, with creston type relief and presence of silty-clay sedimentary rocks. Farms 3 and 4 are in the Humic Dystrudepts-Typic Hapludands association, corresponding to mountain soils with crest-like relief and the presence of volcanic ash and sedimentary rocks. Farm 5 is in the Humic Dystrudepts-Typic Hapludands association, associated with valley-type relief with the presence of mixed alluvial and colluvial deposits and volcanic ashes.

**Methodology**

From each farm, 9 to 12 soil and cocoa fruit samples were taken, based on their extension and cultivated area in order to cover the entire geographical coverage in equal proportions. Soil samples collected from 0 to 30 cm depth under the canopy of the selected trees were dried at room temperature and passed through a 2 mm sieve. Total Cd in soil and Cd content in cocoa beans were determined (Rodríguez et al., 2019). Based on these results, soil samples with similar contents in both soil and beans were ranked. A composite sample was prepared from the subsamples taken on each farm, according to the ranges of Cd present in the beans, the total Cd in the soil and their spatial proximity, thus obtaining one sample for each group described in Fig. 1, to perform the sequential extraction of the Cd forms. A total of 26 composite soil samples were prepared for study (Table 1).

Table 1. Grouping of the sampling points for the composite soil samples used in the sequential extraction of Cd forms at a depth of 0-30 cm in the municipality of Yacopí. Each group corresponds to the composite sample formed from the samples marked in the field "Sampling Location", which presented similar Cd contents. The values of total Cd and Cd beans were taken from (Rodríguez et al., 2019).
<table>
<thead>
<tr>
<th>Farm</th>
<th>Sampling Location</th>
<th>Cd (beans) (mg kg(^{-1}))</th>
<th>Total Cd (soil) (mg kg(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>2,3,4,5,6</td>
<td>1.03-17.6</td>
<td>7.6 - 9.3</td>
</tr>
<tr>
<td></td>
<td>1,11</td>
<td>17.8</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>7,9,10,12</td>
<td>15.2-20.0</td>
<td>17.6 - 18.9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>21.8</td>
<td>21.1</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>20.5</td>
<td>38.2</td>
</tr>
<tr>
<td>Y2</td>
<td>2,5,6,10,9</td>
<td>9.7-13.5</td>
<td>1.0 - 2.3</td>
</tr>
<tr>
<td></td>
<td>3,4</td>
<td>1.9-2.6</td>
<td>0.8 - 0.9</td>
</tr>
<tr>
<td></td>
<td>7,11</td>
<td>14.1-18.2</td>
<td>3.0 - 4.0</td>
</tr>
<tr>
<td></td>
<td>1,12</td>
<td>17.1-19.5</td>
<td>5.3 - 8.0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>16.6</td>
<td>12.3</td>
</tr>
<tr>
<td>Y3</td>
<td>6,7</td>
<td>19.2-19.8</td>
<td>24.8 -25.2</td>
</tr>
<tr>
<td></td>
<td>4,5,12</td>
<td>17.9-19.3</td>
<td>14.7 - 19.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>14.9</td>
<td>32.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>20.9</td>
<td>30.7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>30.6</td>
<td>52.7</td>
</tr>
<tr>
<td></td>
<td>10,11</td>
<td>14.3-18.8</td>
<td>32.5 - 42.1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1.02</td>
<td>22.3</td>
</tr>
<tr>
<td>Y4</td>
<td>1,2,3</td>
<td>2.1-10.2</td>
<td>1.6 - 2.6</td>
</tr>
<tr>
<td></td>
<td>4,5,6,7</td>
<td>7.4-17.3</td>
<td>5.2 - 6.1</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>10.2</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Y5</td>
<td>9,10,11,12</td>
<td>1.3-5.1</td>
<td>1.3 - 1.9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.2</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>3,4,5,6</td>
<td>1.0-3.9</td>
<td>0.4 - 1.0</td>
</tr>
<tr>
<td></td>
<td>7,8</td>
<td>1.3-3.9</td>
<td>0.8 - 1.6</td>
</tr>
</tbody>
</table>

**Laboratory methods**
The Cd phases obtained by sequential extraction were identified according to the methodology described by Amacher (1096). The extraction of all phases was carried out with 1.0 g of soil and the Cd content was evaluated by atomic absorption spectrometry using a Perkin Elmer Analyst 300 spectrophotometer at a wavelength of 228.8 nm. To determine the Cd associated with organic matter, extraction with sodium pyrophosphate was used. The sodium pyrophosphate has a higher extraction yield than hydrogen peroxide, since it is a chelating agent that allows the solubilization of organic compounds precipitated by metals (López & Perez, 2013), without affecting the amorphous crystalline oxides of Fe and Al (Herrera et al., 2004, McKeague, 1967) (Table 2).

<table>
<thead>
<tr>
<th>Phase</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchangeable</td>
<td>Magnesium nitrate (Mg (NO₃)₂).</td>
</tr>
<tr>
<td>Carbonates</td>
<td>Sodium acetate (NaOAc).</td>
</tr>
<tr>
<td>Manganese Oxides</td>
<td>Hydroxylammonium chloride (NH₂OH.HCl, 0.1M)</td>
</tr>
<tr>
<td>Organic Matte, Na₄P₂O₇</td>
<td>Sodium pyrophosphate (Na₄P₂O₇)</td>
</tr>
<tr>
<td>Amorphous iron oxide</td>
<td>Hydroxylammonium chloride (NH₂OH.HCl 0.2M + 0.25 M HCl)</td>
</tr>
<tr>
<td>Crystalline Iron Oxide</td>
<td>Ammonium Oxalate (C₂H₈N₂O₄), oxalic acid (C₂H₂O₄), ascorbic acid (C₆H₈O₆)</td>
</tr>
<tr>
<td>Residual</td>
<td>Aqua regia (HCl:HNO₃ – 3:1)</td>
</tr>
</tbody>
</table>

The pH values were taken from the study of (Rodríguez et al., 2019), these values were measured in a soil: distilled water suspension 1:1W:V.

**Data analysis**

Once the chemical information was obtained, the spatial modeling process was conducted in relation to the geographic coordinates of the sampling points, as well as the centroids corresponding to the polygons generated by tessellation (Fig. 1). In total, 26 polygons were obtained for which a matrix of weights associated with all the neighbors. The relationships between Cd concentration in grains (mg kg⁻¹) and the explanatory variables associated with cadmium fractionations (exchangeable, fraction associated with carbonates, manganese oxides, organic matter, amorphous iron oxides, crystalline iron oxides and in the residual phase) were studied. For this purpose, spatial regression models were fitted, initially with nested general spatial modeling (GNS) and finalized with spatial autocorrelation models (SAC) and spatial error model (SEM) (Elhorst, 2014). The models written from the matrix point of view are expressed as shown in the following equations (1) to (3). Equations (2) and (3) can be seen as particular cases of Eq. (1). Based
on the general nested spatial (GNS) model, under certain conditions, the GNS model becomes the spatial error model (SEM) model.

\[
GNS:\quad Y = \lambda WY + \alpha + X\beta_{(1)} + WX\beta_{(2)} + u; \ |u| < 1
\]

\[
u = \rho W u + \epsilon ; \ |\rho| < 1
\]

(Eq. 1)

If \(\lambda = 0\) and \(\beta_{(2)} = 0\), the GNS model is reduced to:

\[
SEM:\quad Y = \alpha + X\beta_{(1)} + u; \ |u| < 1
\]

\[
u = \rho W u + \epsilon ; \ |\rho| < 1
\]

(Eq. 2)

In all of the equations, \(Y\) represented the Cd in beans, \(X\) represents the matrix of explanatory variables associated with Cd fractionation, \(W\) was associated with the weight matrix in relation to the distances from the centroids of the closest neighbor in the polygons generated by tessellation of the soil sampling points, \(\lambda\) represents the spatial autoregressive coefficient, \(\alpha\) corresponded to the intercept, \(\beta_{(1)}\) and \(\beta_{(2)}\) were associated with the parameters linked to the explanatory variables in natural and lagging form respectively, \(u\) and \(\epsilon\) corresponded to the residual vector, the first with spatial autocorrelation given by the structure in the second equation of each model and the second with \(\epsilon \sim \text{iid } N(0,\sigma^2 I)\), where \(I\) is an identity matrix of dimension \(n \times n\).

The choice of the model that best explains the statistical relationship of the experimental data was based on the Akaike information criterion and on the fulfillment of assumptions of normality of residuals (\(\epsilon\)) and fading of spatial dependence in \(u\). In the case that both models met the assumptions, the highest correlation between the observed values of the response in \(Y\) and the values estimated by the model was used as a criterion (Martínez et al., 2009).

Once the spatial regression model has been proposed or selected, it is important to evaluate and interpret the coefficients of the model as has been done in classical regression models when the estimation is OLS (ordinary least squares). However, this relatively simple way in linear models is complicated in some spatial regression models, so it is necessary to study the effect or impact of the explanatory variables in terms of the coefficients of the selected model. The spatial lag of both the response variable and when using a model with lag in the explanatory variables generates coefficients that become confusing for interpretation, so strategies are proposed for the estimation and interpretation of these coefficients, dividing them into direct, indirect and total impacts (Elhorst, 2014). Some models such as the spatial error model (SEM) do not present this difficulty, so the simple interpretation of the coefficients associated with the vector of beta parameters \((b)\) can be used as measures of the impact of the explanatory variables on the response.

**Results**

**Exploratory data analysis**
Table 3 shows the Cd averages for each fraction, for each of the five farms, made up of the different groups of composite samples. The percentage of each fraction was obtained with respect to total Cd.

### Table 3

<table>
<thead>
<tr>
<th>Farm</th>
<th>pH</th>
<th>Exch. (mg kg⁻¹)</th>
<th>Carbonates (mg kg⁻¹)</th>
<th>Mn-oxides (mg kg⁻¹)</th>
<th>OM-P (mg kg⁻¹)</th>
<th>AI-oxides (mg kg⁻¹)</th>
<th>CI-oxide (mg kg⁻¹)</th>
<th>Residual (mg kg⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>5.47</td>
<td>0.899</td>
<td>8.16</td>
<td>4.43</td>
<td>5.96</td>
<td>1.76</td>
<td>0.48</td>
<td>1.29</td>
</tr>
<tr>
<td>Y2</td>
<td>4.61</td>
<td>0.61</td>
<td>1.63</td>
<td>0.66</td>
<td>4.40</td>
<td>0.87</td>
<td>0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>Y3</td>
<td>5.59</td>
<td>1.92</td>
<td>13.3</td>
<td>5.91</td>
<td>7.88</td>
<td>0.63</td>
<td>0.64</td>
<td>1.05</td>
</tr>
<tr>
<td>Y4</td>
<td>5.44</td>
<td>0.27</td>
<td>1.56</td>
<td>1.24</td>
<td>4.79</td>
<td>0.74</td>
<td>0.32</td>
<td>0.65</td>
</tr>
<tr>
<td>Y5</td>
<td>4.07</td>
<td>0.06</td>
<td>0.18</td>
<td>0.87</td>
<td>4.28</td>
<td>1.01</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>

The relationship between the total Cd contents determined with aqua regia and that estimated through the sum of the contents of each fraction, expressed through a linear equation, presents a slope close to the unit and a value of $R^2 = 0.9$ (Coefficient of determination) indicating a good fit with the values found in both methods, giving reliability to the fractioning process used (Fig. 2).

In Yacopí, the highest Cd contents were observed in the phase associated with organic matter in most farms, except for farms Y1 and Y3 (Table 3), where the highest concentrations were in the carbonate phase. This represented 33.1% and 42.5% of the total Cd, respectively, followed by Cd concentrations in the OM fraction. Considerable contents were also observed in the phases associated with Mn and amorphous Fe oxides. These had an average of 10.0% of total Cd, except in the Y3 farm, whose Cd contents associated with Fe oxides contribute less than 6.0%. The exchangeable phase, that of crystalline Fe oxides and the residual fraction, had the lowest percentages (< 5%), like what was found by Chávez et al. (2016) in cocoa soils of southern Ecuador. There, the lowest percentage of total Cd was present in the exchangeable and reducible phases (associated with crystalline Fe oxides), with the highest Cd concentration in the oxidizable fraction, associated with organic matter, followed by carbonate content. In contrast to the present study, this author reports a higher participation of the residual phase (Fig. 3).

Aikpokpodion et al. (2012) evaluated the cocoa crops in Nigeria for Cd content associated with water-soluble, exchangeable, carbonates, Fe-Mn oxides, oxidizable (organic matter) and residual phases, finding in the latter the highest percentage of total Cd, followed by organic matter contents. The cadmium measurements in rock fragments collected in the sampling areas ranged from 0.9 to 26.7 mg kg⁻¹ (Rodríguez et al., 2019). The highest values were found in farms Y1 and Y3 which present high Cd concentrations associated with carbonates. Additionally, the Y3 farm presented the highest Cd values in
carbonates and in the parent material and is in the MPVf soil unit, corresponding to the Humic Dystrudepts - Typic Hapludands association, constituted by clastic clayey rocks with carbonate intercalations in some sectors.

**Spatial regression model**

For the selection of the best model to adjust the relationship between the response associated with the Cd content in cocoa beans with the explanatory variables considered, the models of pure autoregressive spatial regression (PAR), spatial lag model (SLM), spatial error (SEM), spatial Durbin error, spatial autoregressive pooled (SAC) and the usual non-spatial multiple regression model (MRM) (Table 4).

<table>
<thead>
<tr>
<th>Model</th>
<th>$l$</th>
<th>$r$</th>
<th>AIC</th>
<th>$r$</th>
<th>NT</th>
<th>MIT</th>
<th>variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR</td>
<td>0.66</td>
<td></td>
<td>166.45</td>
<td>0.76</td>
<td>0.77</td>
<td>0.75</td>
<td>Cd-a</td>
</tr>
<tr>
<td></td>
<td>(1.5e-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLM</td>
<td>0.69</td>
<td>0.89</td>
<td>156.36</td>
<td>0.46</td>
<td>0.33</td>
<td></td>
<td>Cd-a;F-r;F-ox-Fe-c;F-MO;F-Mn;F-I</td>
</tr>
<tr>
<td></td>
<td>(2.2e-2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEM</td>
<td>0.48</td>
<td>0.89</td>
<td>152.08</td>
<td>0.68</td>
<td>0.93</td>
<td></td>
<td>Cd-a;F-r;F-ox-Fe-c;F-MO;F-Mn;F-I</td>
</tr>
<tr>
<td></td>
<td>(2.1e-3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDEM</td>
<td>0.12</td>
<td>0.90</td>
<td>152.61</td>
<td>0.42</td>
<td>0.56</td>
<td></td>
<td>Cd-a;F-r;F-ox-Fe-c;F-Mn;F-I</td>
</tr>
<tr>
<td></td>
<td>(0.63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAC</td>
<td>-0.47</td>
<td>0.90</td>
<td>153.32</td>
<td>1.00</td>
<td>0.90</td>
<td></td>
<td>Cd-a;F-r;F-ox-Fe-c;F-MO;F-Mn;F-I</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRM</td>
<td></td>
<td></td>
<td>159.53</td>
<td>0.81</td>
<td>0.24</td>
<td>0.49</td>
<td>Cd-a;F-r;F-ox-Fe-c;F-Mn;F-I</td>
</tr>
</tbody>
</table>

The values in parentheses are related to the p-values associated with the autoregressive ($l$) parameters and the spatial autocorrelation coefficient ($r$). The inspection of the Table 4 allows discarding the SAC and SDEM models. All the other models could have been selected by fulfillment of assumptions, normality (NT) of the residuals using the Shapiro-Wilk test and independence of residuals (MIT), using the Moran index with the matrix of weights of all the neighbors. However, by using the Akaike information criterion (AIC) and the correlation coefficient ($r$) between the observed values of the response and those estimated by the model, the SEM model was finally selected. Although the last model labeled as MRM is not spatial, it was equally appropriate in explaining the spatial dependency initially explored before modeling in the response variable, however, it did not obtain the best fit statistics.

Table 5 shows the estimated parameter values of the spatial error model (SEM). This showed the best fit based on the Akaike criterion (153.1) against the explanation of the Cd content in the beans. Another way
of validating the model consists of estimating the Pearson correlation coefficient between the observed values of the response and those estimated by the model, which in the present case turned out to be 0.9, an acceptable value for predictive purposes.

Table 5
Estimated parameters and their significance in the SEM model (Spatial Error Model), for cadmium fractions measured at a depth of 0 to 30 cm from Yacopí municipality.

| Coefficients                  | Estimate | Asymptotic Std. Error | z value | Pr(>|z|) |
|-------------------------------|----------|-----------------------|---------|----------|
| (Intercept)                   | 7.398    | 3.585                 | 2.064   | 0.039    |
| Exchangeable Fraction         | 5.520    | 1.155                 | 4.780   | 1.8e-6   |
| Fraction of Mn oxides         | -1.296   | 0.683                 | -1.897  | 0.058    |
| Fraction of Organic Matter    | -1.748   | 0.854                 | -2.047  | 0.041    |
| Fraction of crystalline Fe oxides | 7.153 | 2.462                 | 2.905   | 3.7e-3   |
| Residual Fraction             | 7.577    | 1.783                 | 4.250   | 2.1e-5   |

The results of the spatial modeling show from a statistical point of view that the variable Cd in cocoa beans is probably explained by the exchangeable fractions, the fraction associated with manganese oxides, organic matter, crystalline iron oxides and the residual fraction, with greater importance in the exchangeable and residual fractions. However, the impact of each explanatory variable on the impact of each explanatory variable on the response can be measured directly from the parameter vector, as shown in Eq. (3).

\[
\frac{\partial E(Y)}{\partial X} = \frac{\partial E \left( \alpha \nu_n + X \beta_{(1)} + (I - \rho W)^{-1} \epsilon \right)}{\partial X} = \frac{\partial (\alpha \nu_n + X \beta_{(1)})}{\partial X} = \frac{\partial (\alpha \nu_n + X b)}{\partial X} = b(eq.3)
\]

where \( E \) is associated with the mathematical expectation operator. The results obtained from Eq. 3 correspond to the impact of the covariates measurable and interpretable directly from Table 5. Since the response is associated with the accumulation of Cd in the beans, it is of great interest to observe negative impacts, which highlight that unit increases in the explanatory variable, as in the case of the organic matter fraction and manganese, are associated with decreases in Cd in beans, in the same magnitude of the specific coefficient associated with the variable, in this case, 1.748 and 1.296, respectively (Table 5).

**Discussion**

Soils cultivated with cocoa are characterized by the accumulation of OM, coming from the decomposition of leaf litter (Wood & Lass, 2008), so it is to be expected that, in these soils, the fractions of Cd associated with OM present high values. According to the study of Rodríguez et al. (2019), the average total carbon
(C) content in the analyzed samples corresponds to 4.28%, however this value corresponds to both organic and inorganic C.

As is well known, Cd is strongly bound to clay mineral surfaces, Fe and Al oxides and humus (Hong et al., 2002). It has also been pointed out that the presence of neutral and/or alkaline pH strengthens the adsorption of Cd by Mn and Fe oxyhydroxides (Tabelin et al., 2018), which is verified in farms Y1, Y3 and Y4, which present the highest pH values (slightly acidic), with the highest contents of amorphous and crystalline Fe oxides (Table 3). In addition, Table 5 corroborates the negative relationship between Cd content in cocoa beans and the fraction associated with Mn oxides, indicating that this form is not translocated to the bean. For Lin et al. (2018) these fractions are relatively stable to rain, snow or acid rain conditions. On the contrary, for Kotresha et al. (2021) these Cd fractions present a medium stability associated with sorption and precipitation processes which is related to the oxido-reductive character of Fe and Mn. For Shahid et al. (2016) these forms of Cd are potentially available since binding to Fe and Mn oxides in the solid phase is reversible, and they can easily pass into the solution phase of the soil. However, in our study, this availability is limited.

In addition, it has been found significant correlations between the Cd in cacao beans and the contents of the elements associated with the residual phase and carbonates that had been measured in the 0 to 5 cm soil depth, with correlation coefficients of 0.57 and 0.7, respectively (Chávez et al., 2016). However, these correlations may be questionable due to the compositional nature of the fractions, in this sense, the current discussion is based primarily on the impacts generated by the selected spatial model (Pawlowsky-Glahn et al., 2015). It is important to point out that regression models such as those analyzed in this study allow understanding the behavior of the fractions in the Cd hyperaccumulator character of cocoa, that is, not only the different fractions are analyzed with the objective of understanding the binding mechanisms of this heavy metal, but also, with the regression model we are getting to answer the question of which fractions contribute to the soil solution and which are being bioavailable for cocoa plants located in the soils of Yacopi.

Several publications have reported that the exchangeable fraction and carbonates are related to the Cd phytoavailability, while the complex in the OM is potentially available (Bacon & Davidson, 2008, Rao et al., 2008, Kosolsaksakul et al., 2014, Sungur et al., 2014), due to its dependence on pH affecting surface charges, as well as that the organic matter can alter Cd speciation from soluble to oxidizable fractions (Mohamed et al., 2010). As noted above, Cd is adsorbed more effectively by the organic matter than the Fe and Mn hydrous oxides at lower soil pH levels characteristic of cocoa soils. Because this fraction is associated with stable humic substances of high molecular weight, which reduces the mobility and release of Cd (Kotresha et al., 2021). This is also reflected in Table 5, with the fraction associated with organic matter having the strongest negative relationship, indicating that these forms, as well as those associated with manganese oxides, are not translocated to the bean and it could be deduced that the increase of this fraction will reduce the contents of the element in the bean.

The carbonates are sensitive to changes in pH, and the Cd associated with them increases their availability in acidic conditions. This is common in soils with cocoa cultivation, because the roots release
low-molecular-weight organic acids (Chávez et al., 2016). The solubility of Cd has a turning point around pH 6.0, above which it is largely held in complexes with organic matter and adsorbed by the mineral surfaces (McBride, 1994, Sullivan et al., 2013). The pH of the soils studied were highly variable, ranging from 3.7 to 7.24, with higher carbonate-associated Cd contents observed at pH between 5 and 7.24. Other authors state that the Cd shows greater mobility in a pH range of 4.5–5.5, while at higher pH values it becomes insoluble in forms of carbonate and phosphate, which could explain why, despite the high Cd content in carbonates, this fraction according to the spatial model of the error does not explain the cadmium contents in the cocoa bean (Kabata-Pendias & Pendias, 2000), taking into account that according to Kotresha et al. (2021), this is one of the most unstable fractions because it is influenced by the variation of environmental conditions.

In general, Cd exists in different chemical forms under different soil pH, presenting both cationic (CdHS⁺, CdOH⁺, CdHCO₃⁺, CdCl⁺) and anionic (Cd(HS)₄²⁻, Cd(OH)₃⁻, Cd(OH)₄²⁻, CdCl₃⁻) species, reason why any pH variation will favor any of its forms, such as the acidic conditions present in the rhizosphere which increase during the respiration process, where carbonic acid dissociates releasing H⁺ and HCO₃⁻, allowing a rapid adsorption of Cd²⁺ and without energy expense, through the epidermal cells of the root, via apoplastic, related to the exchange of H⁺ (Yamaguchi et al., 2011, Kabata-Pendias & Sadurski, 2004). On the other hand, in calcareous soils, Cd²⁺ and CdHCO₃⁺ forms may predominate (Shahid et al., 2016).

Another reported result indicates that Cd released during carbonate weathering complexes more than 10% with organic matter and 30% with Fe and Mn oxides, while only 2.4% leaves to constitute the residual fraction. As noted above, higher levels of Cd in the carbonates were related to higher values of total Cd in the parent material. According to IGAC (2000) mapping, the corresponding crops are located in a soil unit reporting carbonate intercalation, so it could be expected that the residual phase comes from the calcareous parent material, according to (Rao et al., 2008, Nogueira et al., 2010). The Cd fractions associated with carbonates, Fe and Mn oxides and organic matter could be considered relatively active depending on the physical and chemical properties of the soils (He et al., 2015). However, for the present study, it is possible that many of these potentially available forms, such as organic matter and manganese oxides, are retained in the colloidal complex, due to the negative relationship with grain Cd contents (Table 5), or on the contrary enter through the transpiration channel but do not translocate directly to the bean. These negative relationships can be interpreted as a decrease in Cd contents in cocoa beans for each unit increase in the fraction associated with manganese oxides and/or organic matter.

Concentrations of Cd in limestone, higher than those present in the soil, can cause contamination with this element (Alloway et al., 2013), but Cd levels in most limestone sources are lower than found in these soils. The soil carbonate fraction as well as the exchangeable fraction support most of the pollutant in soils enriched with Cd (Akkajit et al., 2013). Although high Cd levels in cocoa litter deposited on the soil surface will likely be a more important source of topsoil Cd than soil amendments applied in cocoa production (Barraza et al., 2019), and these are the Cd fractions that are related to the Cd available to plants (Kosolsaksakul et al., 2014, Sungur et al., 2014). However, not all Cd from leaf litter corresponds only to the organic fraction, because this fraction seems not to be bioavailable in the study area (Table 5). According
to Rao et al. (2008) and Nogueira et al. (2010), when Cd is introduced artificially, high contents are seen in the soluble fractions within which are carbonates. However, the spatial model of the error excluded the carbonate phase in the explanation of the Cd accumulation in cocoa beans and shows the importance of the Cd associated with the exchangeable and residual phases (Table 5).

The residual fraction can be related to the recalcitrant Cd derived from the application of fertilizers (Gabe & Rodella, 1999, Chávez et al., 2016), and according to Soler (2017) applications of products such as dolomite lime with Cd levels between 8.53 to 9.21 mg kg$^{-1}$ have been reported in some cocoa crops located in the department of Cundinamarca. However, the influence of the parent material on the residual fraction could not be ruled out, since some lots with high values in this phase also had high Cd contents in the parent material.

Dolomitic limestone is frequently used to reduce the phytoavailability of heavy metals such as Cd, due to the formation of stronger complexes at increased soil pH. However, its application may lead to increases in Cd content in the soil, since the amounts of the element are variable in this product. Cd enrichment in soils with high carbonate content is due to the relative condensation produced during weathering that generates calcite leaching and/or dolomite and permanence of Cd in residual minerals that are insoluble in acid (Xia et al., 2020). Some forms of Cd are precipitated by the application of limes such as dolomite but can be activated during agricultural activities, since the metabolic activity of the rhizosphere acidifies the medium.

**Conclusion**

The determination of the concentrations of the different fractions of Cd present in the samples collected in the municipality of Yacopí reflected that the highest Cd levels were found in the phase associated with organic matter, carbonate phase, phases associated with Mn oxides, and amorphous Fe, and to a lesser extent the contents were associated with the exchangeable phase, that of crystalline and residual Fe oxides.

The forms of Cd present in the Yacopí soil samples that explain the Cd contents in cocoa beans are adjusted to a spatial error model (SEM), and correspond to the exchangeable fractions, the fraction associated with manganese oxides, organic matter, crystalline iron oxides and the residual fraction, highlighting greater significance in the exchangeable and residual phase, and a negative relationship with the organic matter and manganese oxides phases, which shows that these forms are retained in the colloidal phase and are not translocated to the bean, that is, despite the high Cd contents in the organic matter and in the leaf litter that reaches the soil, these forms of Cd do not increase the Cd contents in the bean. On the contrary, it is possible to suggest, according to this modeling strategy, that a change in the fractions associated with organic matter and manganese oxides could reduce their accumulation in the cocoa beans, considering that an increase of one unit of Cd in these fractions is reflected in a decrease of 1,748 mg kg$^{-1}$ and 1,296 mg kg$^{-1}$, respectively, in the Cd contents in beans.
Although the Cd concentration associated with the carbonate phase was high, it did not explain the Cd concentrations in the cocoa beans, as it was discarded as an explanatory variable in the spatial error model (SEM) because it did not provide evidence against the null impact hypothesis. In contrast, the residual and exchangeable phases, with the lowest contents, showed the largest effects on the cocoa bean Cd model. It is possible that previous dolomite applications and the presence of carbonate intercalations in the parent material make these forms of cadmium available due to the acidification of the medium by the action of the crop rhizosphere and the acidification caused by the decomposition of the leaf litter.

**Declarations**

**Data Availability**

The full dataset generated or analyzed during this study are available from the corresponding author in reasonable request.

**Author contribution**

Heidy Soledad Rodríguez-Albarracín: Conceptualization, methodology, and writing-original draft. Aquiles Enrique Darghan: Formal analysis, methodology supervision, software, and writing-original draft. Carlos Armando Rivera-Moreno: software, review, and editing. All authors reviewed and approved the final manuscript.

**Conflicts of Interest**

The authors declare that there aren’t conflicts of interest regarding the publication of this paper.

**Funding**

No funding was obtained for this study.

**References**


Figures
Location of the groups of samples, tiles and centroids for the sampled farms in Yacopí municipality. The centroids are represented by asterisks preserving the colors by group of samples. The groups respond to the grouping of the samples according to their location and the values obtained for Cd in beans and total Cd measured at a soil depth of 0 to 30 cm described in (Rodríguez et al., 2019).

Figure 2

Scatter plot of the relationship between total Cd estimated by the sum of the fractions and total Cd measured (y) by Aqua Regia extraction (mg kg$^{-1}$) and total Cd (x) by sum of fractions.
Figure 3

Percentages of the Cd fractions in soils from Yacopí municipality.