Journal of Hazardous Materials 167 (2009) 615–624

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Classification and regression trees (CARTs) for modelling the sorption and retention of heavy metals by soil

F.A. Vega [a], J.M. Matías [b], M.L. Andrade [a], M.J. Reigosa [a], E.F. Covelo [a,*]

[a] Departamento de Bioloxía Vexetal e Ciencias do Solo, Universidade de Vigo, Spain
[b] Departamento de Estatística, Universidade de Vigo, Spain

ARTICLE INFO

Article history:
Received 2 October 2008
Received in revised form 2 January 2009
Accepted 6 January 2009
Available online 16 January 2009

Keywords:
CART
Regression trees
Soil
Sorption
Retention
Heavy metal
Soil characteristics

ABSTRACT

The sorption and retention of mixtures of heavy metals by soil is a complex process that depends on both soil properties and competition between metals for sorption sites. In this study, the sorption and retention of mixtures of Cd, Cr, Pb, Cu, Zn and Ni by a representative sample of soils from Galicia (N.W. Spain) was reproduced considerably more precisely by binary decision-tree regression models constructed using the CART algorithm than by linear regression models.

Of the six metals competing for sorption sites in these experiments, Pb, Cu and Cr were sorbed and retained to a greater extent than Cd, Ni and Zn. Non-linear tree regression models constructed with CART fitted the data better than linear models, especially for Cd, Ni and Zn; and with both kinds of model the data for Pb, Cu and Cr were fitted better than those for Cd, Ni and Zn (the difference being much more marked for linear models), suggesting that the influence of soil properties on the sorption and retention of the latter three metals was limited by the preferential binding of the former three.

1. Introduction

Numerous land treatments and other practices, including the application of fertilizer or sewage sludge, the disposal of wastewater on land, and industrial activity, can lead to soils accumulating heavy metal contents substantially in excess of natural levels, with the consequent risk of uptake by plants, pollution of surface or underground waters, and propagation through the food chain [1]. The risk of leaching or uptake by plants depends on the concentration of pollutant in the soil solution, which in turn depends on the sorption–desorption equilibria that govern the partition of pollutant between soil solution and soil solids, soil colloids especially [2,3]. The toxic potential of heavy metals in soil thus depends on soil composition, particularly the amount and type of clay minerals [4–6], organic matter [7,8] and iron and manganese oxides [9–11].

In keeping with the above, in previous work we found that the sorption and desorption of heavy metals by certain soils in Galicia (N.W. Spain) is determined mainly by organic matter, Fe and Mn oxides, and clay and mica contents [12–14]. However, sorption and desorption isotherms have irregular profiles presumably due to competition among metals for sorption sites, and the dependence of sorption and desorption on soil properties is only moderately well represented by linear models [15,16].

A methodology that is gaining favour in an increasingly broad variety of fields for modelling non-linear processes and structures is the use of decision trees, generalizations of the familiar botanical key. When it is a regression model that is needed rather than a classifier, i.e., when the dependent variable Y is a continuous random variable with conditional distribution.

$$Y_{|\boldsymbol{x}} = f(\boldsymbol{x}) + \varepsilon \quad \boldsymbol{x} = (x_1, \ldots, x_n)$$

for some zero-mean random error ε and the problem is to estimate the regression function $f(\boldsymbol{x})$, these methods effectively divide the space $X = \{\boldsymbol{x}\}$, in which the random predictor variables X_i take their values, into a finite number m of disjoint hyper-rectangles D_k that together cover X, and approximate $f(\boldsymbol{x})$ by a piecewise constant function

$$\hat{f}(\boldsymbol{x}) = \Sigma_1^m \hat{c}_k 1_k, \tag{1}$$

where 1_k is the indicator function of D_k ($1_k(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in D_k$, $1_k(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \notin D_k$) and $\hat{c}_k$ is an estimate of the mean of Y in D_k (in practice, the sample mean). The problem is to define the D_k. The regression tree approach (decision-tree regression) does this in successive steps, creating a tree of nested hyper-rectangles $D_k^{(i)}$ (the nodes of the tree), the lowest members of which (the "leaves") are the final D_k. To avoid overfitting the model, this tree may then be "pruned back", a process analogous to backwards elimination of

* Corresponding author. Tel.: +34 986 812630; fax: +34 986 812556.
E-mail address: emmaf@uvigo.es (E.F. Covelo).

0304-3894/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2009.01.016

variables from multiple linear regression models. Compared with other non-parametric regression methods (and with non-linear parametric methods, when the function fitted lacks a firm theoretical basis), decision-tree regression is considered to be especially useful because of the ease with which the resulting tree can be represented graphically, used and interpreted even when there are multiple predictor variables.

In Section 2 below we describe in greater detail the construction of binary regression trees by the CART algorithm (Breiman et al. [17]; see also Ripley [18] and Hastie et al. [19], and for the statistical properties of CART regression estimators, see Devroye et al. [20] and Gyorfi et al. [21]). In Section 3 we describe the use of CART to model data on the sorption and retention of heavy metals by fourteen soils in Galicia (N.W. Spain), and in Section 4 we present and comment on the resulting regression trees, and compare their performance with that of linear regression models. Section 5 concludes.

2. Construction of binary regression trees by cart

The first step in constructing a binary regression tree is to identify the predictor variable x_{j1} and the value a_1 of x_{j1} that minimize the squared error of the sample with respect to the constants $c_1^{(1)}$ in $D_1^{(1)} = \{\boldsymbol{x}|x_{j1} < a_1\}$ and $c_2^{(1)}$ in $D_2^{(1)} = \{\boldsymbol{x}|x_{j1} \geq a_1\}$, where $c_1^{(1)}$ and $c_2^{(1)}$ are the sample means in $D_1^{(1)}$ and $D_2^{(1)}$, respectively. Formally,

$$(j_1, a_1) = \text{argmin}_{i,xi} \Sigma_{\boldsymbol{x}s \in D1}(\bar{y}_s - c_1^{(1)})^2 + \Sigma_{\boldsymbol{x}s \in D2}(\underline{y}_s - c_2^{(1)})^2 \quad (2)$$

where $(\boldsymbol{x}_s, \underline{y}_s)$ are the sample points. This process is repeated for the subsamples contained in the hyper-rectangles $D_1^{(1)}$ and $D_2^{(1)}$, which splits $D_1^{(1)}$ into $D_3^{(2)}$ and $D_4^{(2)}$, and $D_2^{(1)}$ into $D_5^{(2)}$ and $D_6^{(2)}$; and so on. A currently terminal node $D_k^{(i)}$ is made definitively terminal (i.e., is not further split into smaller nodes) if it satisfies some smallness condition, such as containing no more than a certain number of sample points or no more than a certain percentage of sample points.

Without a smallness condition to halt tree growth, the sample would be overfitted, i.e., the random noise would be incorporated in $\hat{f}(\boldsymbol{x})$, the part of the model that is supposed to reflect only meaningful structure. Even with such a condition, overfitting is still a problem. Given a tree T_0 grown as above from a sample $\mathcal{S}_0$, its prediction performance on subsequent independent samples will generally not be as good as that of some subtree T^* formed by retracting some of its branches to internal nodes (which thereby become leaves, the corresponding hyper-rectangles of which are the unions of the leaves terminating the contracted branches of T_0). Briefly, to try to achieve a compromise between good fit and robustness, CART chooses T^* from among a nested sequence $\sum_0 = \sum(T_0, \mathcal{S}_0)$ of subtrees T_j, each of which minimizes, throughout some range of the penalty term weighting parameter α, the function

$$C(T, \alpha; \mathcal{S}_0) = N^{-1} \Sigma_{s \in \mathcal{S}_0}(y_s - \hat{c}_k(s))^2 + \alpha l(T) = S(T, \alpha; \mathcal{S}_0) + \alpha l(T)$$

where N is the number of data points in the sample, $\hat{c}_k(s)$ is the sample mean in $D_k(s)$ (the leaf of T to which sample point s belongs), and $l(T)$ is the size of subtree T (the number of leaves). More specifically, each subtree T_j in Σ_0 minimizes $C(T,\xi; \mathcal{S}_0)$ for all ζ in a unique interval $[\alpha_j,\alpha_{j+1})$, where the α' s form a finite sequence $-\infty = \alpha_0, \ldots, \alpha_r, \alpha_{r+1} = \infty$; and for all ζ in $[\alpha_j,\alpha_{j+1})$, T_j is the smallest subtree of T_0 to minimize $C(T,\zeta; \mathcal{S}_0)$. Given this unique sequence of subtrees (obtained by an algorithm included in the CART program), the one chosen as T^* is the one minimizing an estimate $\hat{s}$ of the predictive error of the tree that is obtained by the v-fold cross-validation process sketched below.

The cross-validation process starts by random division of the sample $\mathcal{S}_0$ into v subsamples, the sizes N_i of which are as equal as possible. Each of these subsamples in turn will be used as a test set $\mathcal{S}_0$ $(i = 1, \ldots, v)$ to test the performance of a tree T_i grown as described above from the remainder of $\mathcal{S}_0$, $\mathcal{S}_0 - \mathcal{S}_i$. In each case, the sequence $\sum_t = \sum(T_i, \mathcal{S}_0 - \mathcal{S}_i)$ is obtained, and the corresponding sequence of regression estimates $\hat{f}(\boldsymbol{x})$ so generated is used to define a piecewise constant error function $s_i(\alpha) = s(T(i,\alpha),\alpha; \mathcal{S}_i)$, where $T(i,\alpha)$ is the member of $\sum(T_i, \mathcal{S}_0 - \mathcal{S}_i)$ that minimizes $C(T,\alpha; \mathcal{S}_0 - \mathcal{S}_i)$ and the mean squared error s_i is calculated using the test set $\mathcal{S}_i$. When this has been done for all i, the weighted average $s^\dagger(\alpha'_j) = N^{-1} \Sigma_i N_i s_i(\alpha'_j)$ is calculated at the points $\alpha'_j = (\alpha_j \alpha_{j+1})^{1/2}$; $\hat{s}$ is defined as the least of these averages; α^* is defined as the largest α'_j such that $s^\dagger(\alpha'_j) \leq \hat{s} + \beta sE$, where the constant β is user-defined and SE is an estimate of the standard error of $s^\dagger$ that is likewise obtained using the data generated from the cross-validation sets $\mathcal{S}_0 - \mathcal{S}_i$ and $\mathcal{S}_i$; and T^* is defined as the member of $\sum(T_0, \mathcal{S}_0)$ that minimizes $C(T,\alpha^*, \mathcal{S}_0)$. The final regression estimate $\hat{f}(\boldsymbol{x})$ is obtained using the whole sample $\mathcal{S}_0$ to calculate the $\hat{c}_k$ corresponding to T^*.

3. Application to heavy metal sorption/retention data

The data used were obtained in previous work [12,13,15,16] on the sorption and retention of six heavy metals (Cd, Cr, Cu, Ni, Pb and Zn) by fourteen soils that were collected in the province of Pontevedra and were representative of the Galician soils on which most crops are grown and on which the above heavy metals are most commonly deposited. For heavy metal sorption and desorption experiments, samples were collected from the surface horizons of all these soils. In each case, six samples were collected using an Eijkelkamp sampler and were transported in polyethylene bags to the laboratory, where they were air dried, passed through a 2-mm-mesh sieve, pooled, and homogenized in a vibratory solid sample homogenizer (a Fritsch Laborette 27). Three subsamples of the homogenized sample were used for soil analyses, and three for sorption/desorption experiments. The soil characteristics determined were particle size distribution, the organomineral fraction, oxides contents (Fe, Mn and Al), effective cation exchange capacity (CEC_e), the percentage of the sub-2-μm fraction constituted by clay, and mineralogy.

The sorption/retention data used here were obtained in experiments in which 12 g of soil was added to 200 mL of a solution containing 100 mg L^{-1} of each of the six metals. In each experiment, cadmium, chromium, copper, nickel, lead and zinc were sorbed from a "sorption solution" containing 100 mg L^{-1} of each metal that had been made up by addition of the metal nitrates to acetate buffer of pH 4.5 (0.02 M acetic acid, 0.02 M sodium acetate), a medium chosen to simulate acid spill conditions. After equilibration by shaking for 24 h at 25 °C in a rotary shaker samples were centrifuged at 1800 × g for 10 min. Metal concentrations in the supernatants were determined by ICP-OES, and the amount of each metal sorbed by the soil sample was calculated by difference and expressed as a percentage of the amount initially present in the sorption solution.

The pellets obtained by centrifugation in the sorption stage were dried at 45 °C, weighed, and resuspended in 200 mL of acetate buffer, after which these suspensions were equilibrated in a rotary shaker for 24 h at 25 °C and centrifuged at 1800 × g for 10 min. Metal concentrations in the supernatants were determined by ICP-OES, and the quantities of metals retained by each soil sample were calculated by difference with respect to the amounts sorbed in the sorption experiments, and were expressed as percentages of the latter.

Each sorption/desorption experiment was performed in triplicate. In each stage of each experiment, following equilibration, pH measurement confirmed that the pH of the medium was still 4.5 [12,13,15,16].

Table 1
Descriptive statistics of the properties of 14 typical Galician soils.

	Mean	Median	Minimum	Maximum	Std. Dev.
Humified O.M. (g kg^{-1})	25.48	22.29	0.95	71.91	20.07
Sand (%)	60.65	62.38	26.71	82.20	15.12
Silt (%)	15.83	14.04	1.67	34.56	11.71
CICe (cmol$_{(+)}$ kg^{-1})	5.22	4.83	0.36	11.22	3.10
Quartz (%)	4.88	4.26	0.84	11.93	3.28
Plagioclase (%)	1.89	0.25	0.00	7.50	2.78
Mica (%)	0.51	0.16	0.00	2.50	0.74
Kaolinite (%)	17.75	14.50	0.00	89.00	21.53
Vermiculite (%)	2.38	2.10	0.00	5.65	1.86
Gibbsite (%)	37.79	38.50	0.00	89.00	22.26
Chlorite (%)	8.73	6.70	0.00	29.75	7.59
Hematite (%)	7.54	1.75	0.00	33.00	10.30
Mn oxides (g kg^{-1})	0.16	0.03	0.00	1.52	0.36
Fe oxides (g kg^{-1})	13.43	8.96	2.47	74.36	16.22
Al oxides (g kg^{-1})	17.74	11.83	3.26	98.21	21.42

Table 2
Descriptive statistics of the capacity of 14 typical Galician soils for sorption and retention of Pb, Cu, Cr, Cd, Ni and Zn in solutions containing all six (percentages of metal sorbed and retained).

	Mean	Median	Minimum	Maximum	Std. Dev.
Sorption					
Cd	21.58	24.28	2.17	37.73	12.18
Cr	31.42	27.79	12.02	81.08	17.12
Cu	47.27	49.96	5.10	82.84	23.17
Ni	18.51	21.73	1.89	36.26	10.39
Pb	71.18	78.23	27.86	92.78	19.61
Zn	19.06	21.40	1.22	38.59	10.95
Retention					
Cd	15.95	16.37	0.00	34.03	11.90
Cr	29.63	27.09	10.96	80.55	17.07
Cu	42.52	45.03	2.09	79.10	23.47
Ni	14.21	14.82	0.00	33.77	10.43
Pb	67.33	74.79	23.64	90.44	20.69
Zn	16.19	16.09	0.00	35.97	11.22

In applying CART to these data (Statistica 7.0), the criterion used for halting the growth of the regression tree was that leaves with less than five data points should not be split, and tree pruning was performed using 10-fold cross-validation and the default value $\beta = 1$.

4. Results

Table 1 summarizes the properties of the soil set used, and Table 2 lists descriptive statistics for the percentage of each metal sorbed by the soils and the percentage of sorbed metal retained during the subsequent desorption phase of the experiments. On average, Pb was the most sorbed and most retained metal, followed at some distance by Cu and, further behind, by Cr.

Cd, Ni and Zn were very similar in their mean sorption and retention behaviour, which was doubtless limited by that of the more readily sorbed metals.

Table 3 lists goodness-of-fit values (coefficients of determination R^2, calculated as usual as $\sum_s(\hat{f}(\boldsymbol{x}_s)-\bar{y})^2/\sum_s(y_s-\bar{y})^2$, where $\bar{y}$ is the mean of the y_s) for the results of using CART to regress sorption and retention data on the set of soil properties or on the soil properties plus the sorption or retention of one of the three most readily sorbed metals, Pb, Cu and Cr The R^2 results achieved by forward stepwise multiple linear regression (LR) on soil properties are also shown for comparison. CART always achieved greater R^2 than LR, and for the less readily sorbed metals the increase was generally by a striking 0.25–0.35, reaching 0.53 in the case of retention of Zn. Even in the case of Cr the R^2 values achieved by LR, 0.899 and 0.895, rose to 0.957 and 0.958, respectively, with CART. Moreover, it should be borne in mind that the performance of LR was favoured by selecting the variables included in the model on the basis of the whole sample rather than by a cross-validation process.

With both LR and CART, R^2 was about 0.97–0.99 for Pb and Cu and much lower for Cd and Zn. The difference was much more pronounced in the case of LR, which reflects the much less linear response of Cd and Zn sorption/retention to soil variables, and the greater ability of CART to predict this non-linear response. With LR, the R^2 values of Cr could be grouped with those of Pb and Cu, and those of Ni with those of Cd and Zn; while with CART, both Cr and Ni grouped with Pb and Cu.

Non-linear tree regression models constructed with CART fitted the data better than linear models, especially for Cd, Ni and Zn; and with both kinds of model the data for Pb, Cu and Cr were fitted better than those for Cd, Ni and Zn (the difference being much more marked for linear models), suggesting that the influence of soil properties on the sorption and retention of the latter three metals was limited by the preferential binding of the former three.

Inclusion of the sorption or retention of Pb, Cu or Cr as a CART predictor variable generally brought about just a small increase in R^2 for Cd and Zn, or a small decrease for Ni. However, larger increases were achieved when Cr was included in the analysis of Zn or Cd sorption and when Cu was included in the analysis of Cd retention, and a larger decrease when Cr was included in the analysis of Cd retention.

Figs. 1–6 show the CART regression trees obtained for the metals using only the soil properties as predictors. Concerning Fig. 1, top (sorption of chromium), the variable with respect to which the first bifurcation of the tree was defined was Fe oxides content, the three soils with the lowest Fe oxides contents all having very high Cr sorption. Soils with higher Fe oxides contents were then segregated on the basis of sand content (which is negatively correlated with clay content); and soils with high sand contents were subsequently separated with respect to clay composition (specifically, on the basis of chlorite, gibbsite and plagioclase contents, in that order). For the retention of chromium (Fig. 1, bottom), the first bifurcation of the tree was again defined with respect to Fe oxides content, but nodes

Table 3
Goodness-of-fit (R^2) of the models fitted by CART and forward stepwise multiple linear regression (LR) to account for the sorption and retention data (percentages of metal sorbed and retained) on the basis of soil properties or, in the case of some CART models, soil properties plus the sorption or retention of one of the metals Pb, Cu and Cr.

	R^2 sorption (%)					R^2 retention (%)				
	LR	CART				LR	CART			
		Soils	+Pb	+Cu	+Cr		Soils	+Pb	+Cu	+Cr
Cd	0.703	0.820	0.840	0.843	0.916	0.556	0.854	0.854	0.947	0.788
Cr	0.899	0.957				0.895	0.958			
Cu	0.970	0.981				0.981	0.991			
Ni	0.688	0.949	0.969	0.949	0.922	0.606	0.947	0.947	0.947	0.918
Pb	0.978	0.993				0.976	0.991			
Zn	0.424	0.774	0.774	0.774	0.887	0.352	0.885	0.885	0.886	0.886

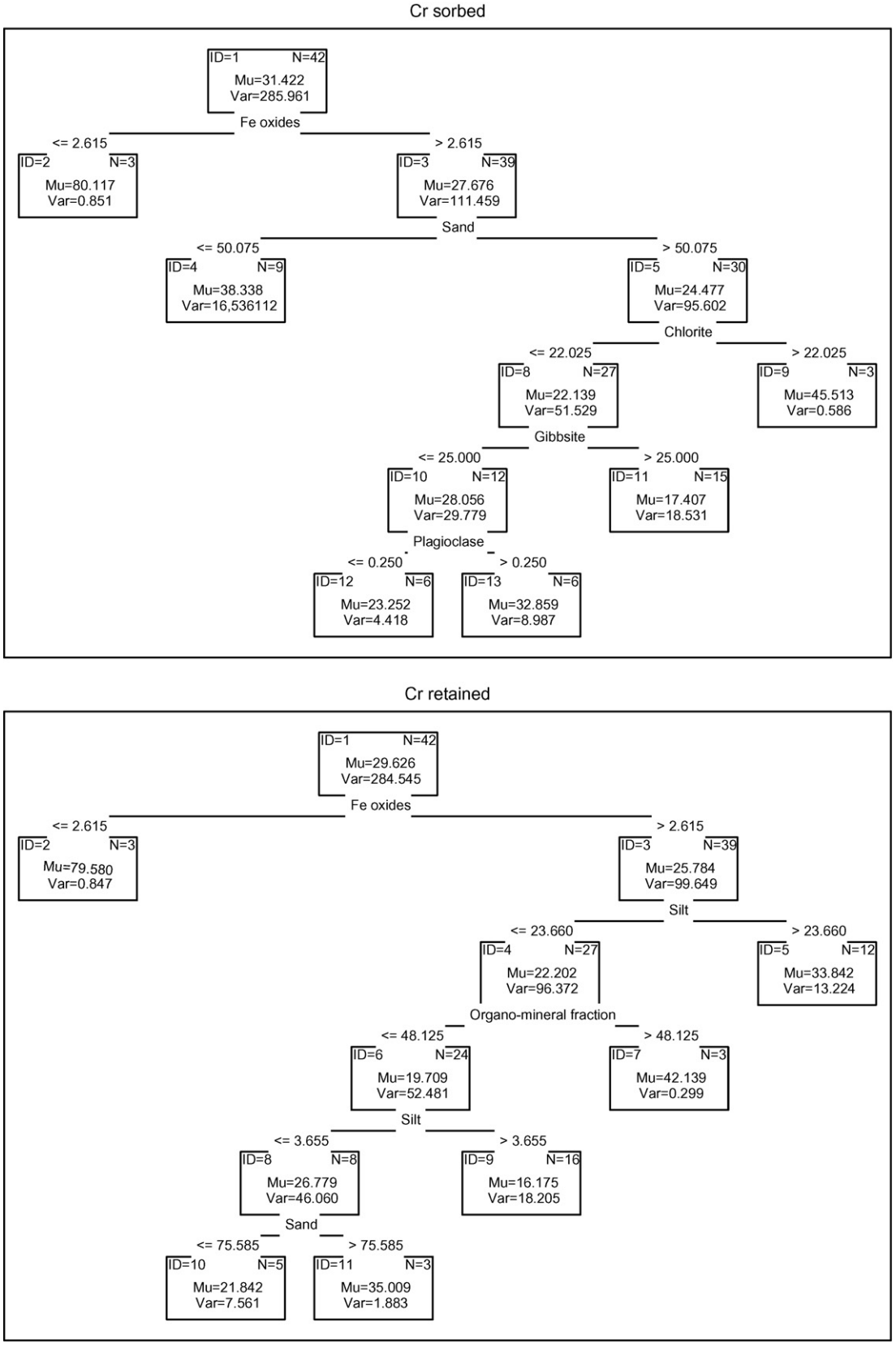

Fig. 1. CART binary regression tree for the sorption (top) and retention (bottom) of Cr by typical Galician soils.

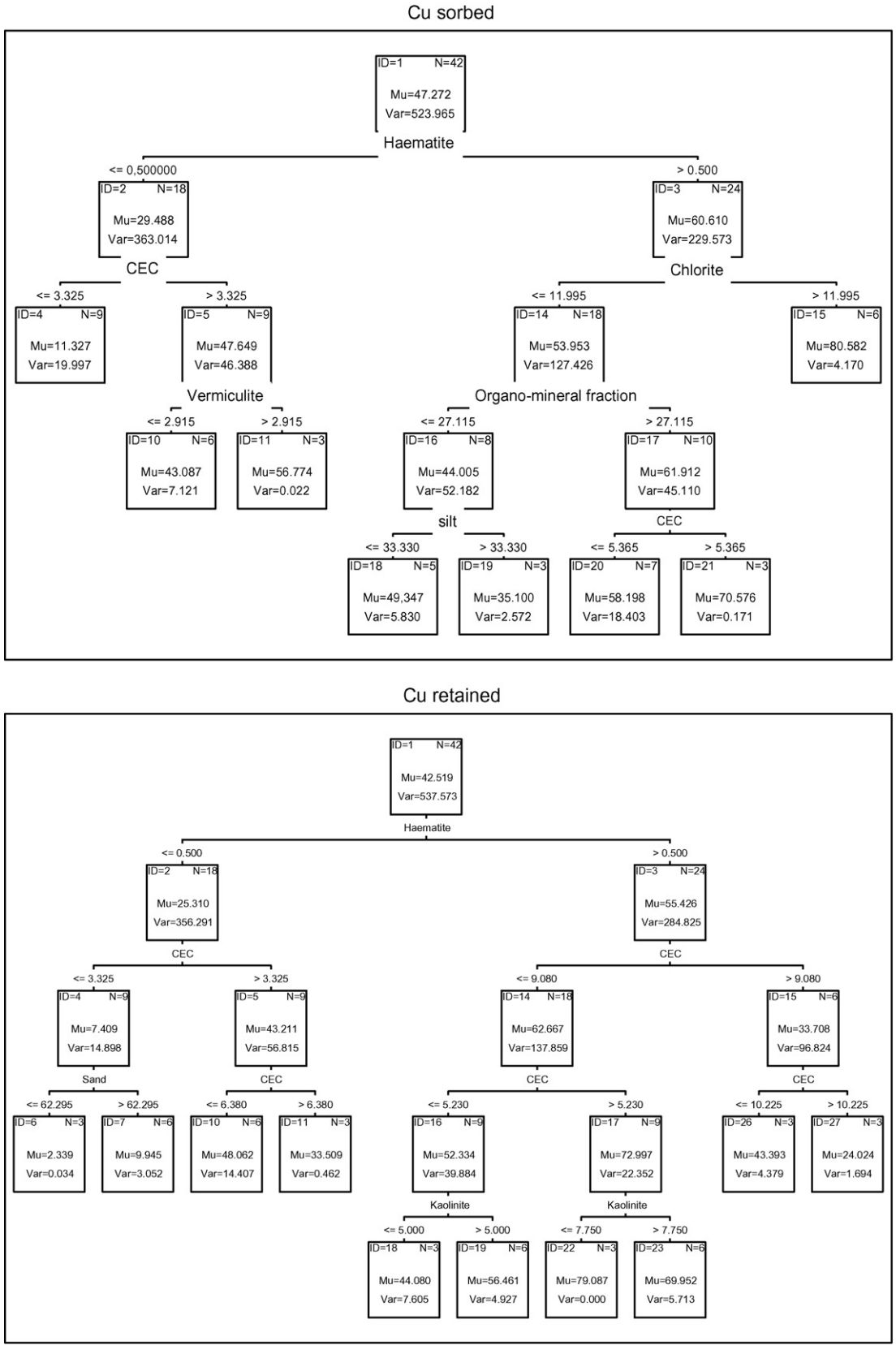

Fig. 2. CART binary regression tree for the sorption (top) and retention (bottom) of Cu by typical Galician soils.

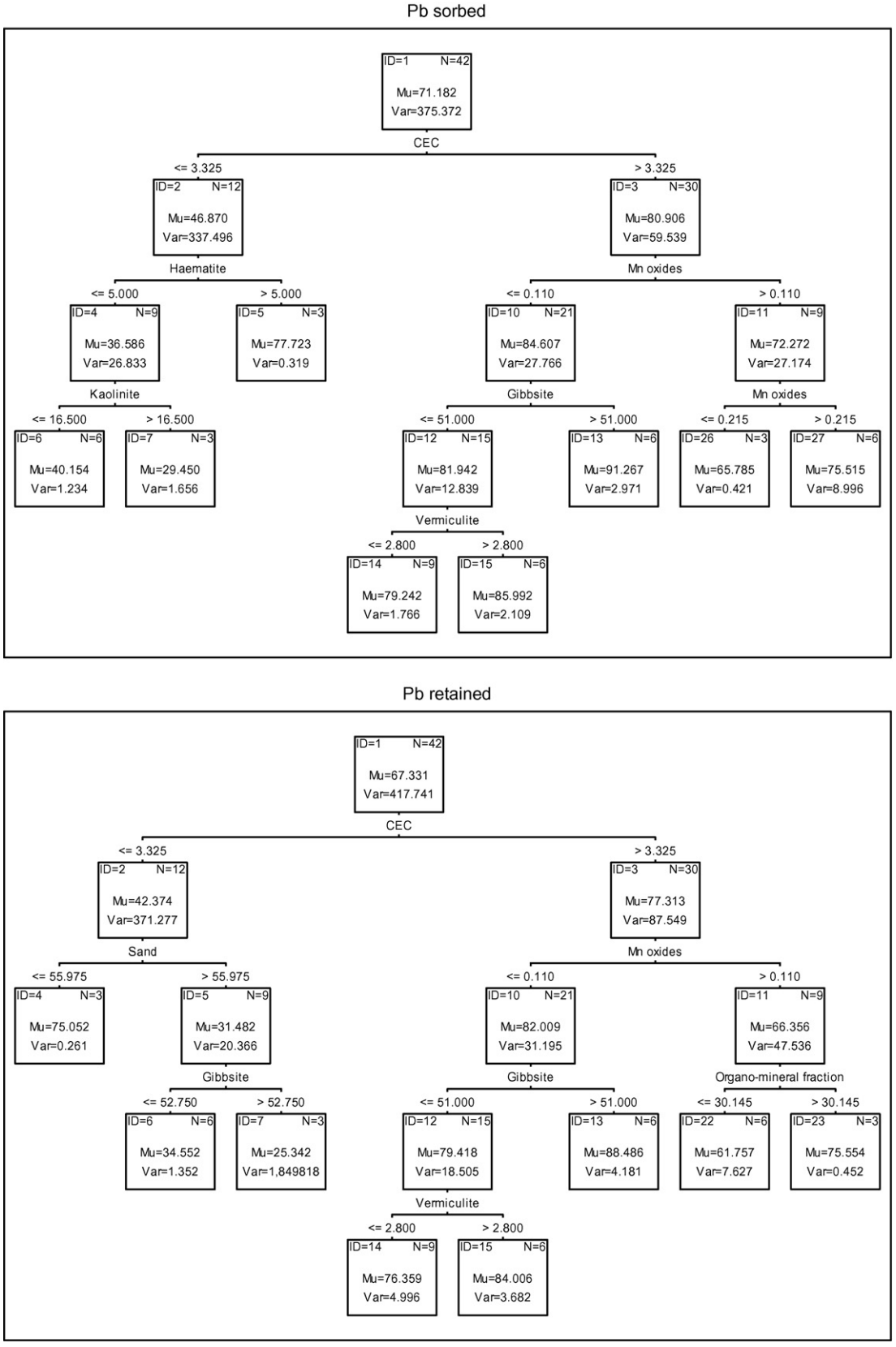

Fig. 3. CART binary regression tree for the sorption (top) and retention (bottom) of Pb by typical Galician soils.

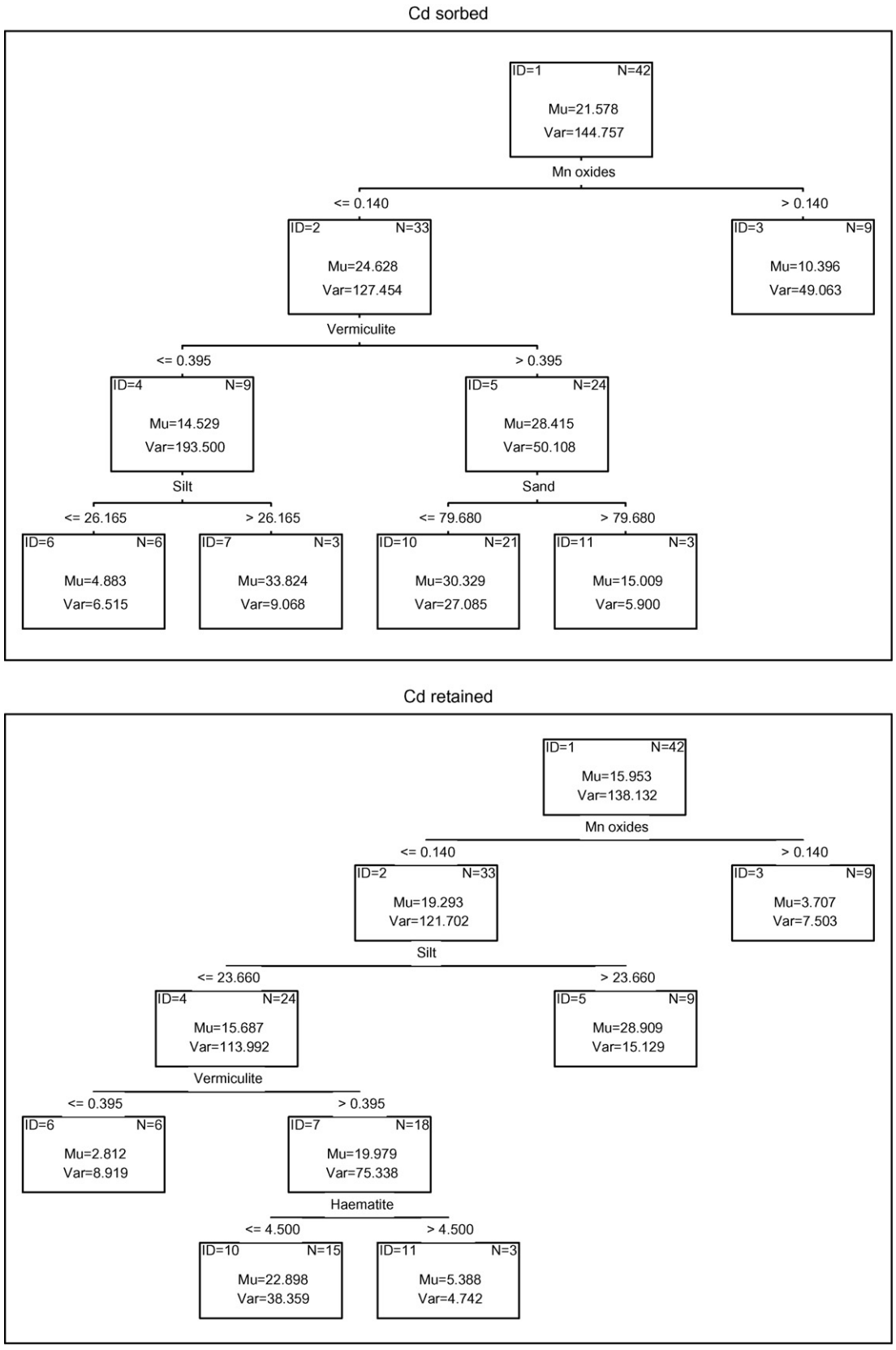

Fig. 4. CART binary regression tree for the sorption (top) and retention (bottom) of Cd by typical Galician soils.

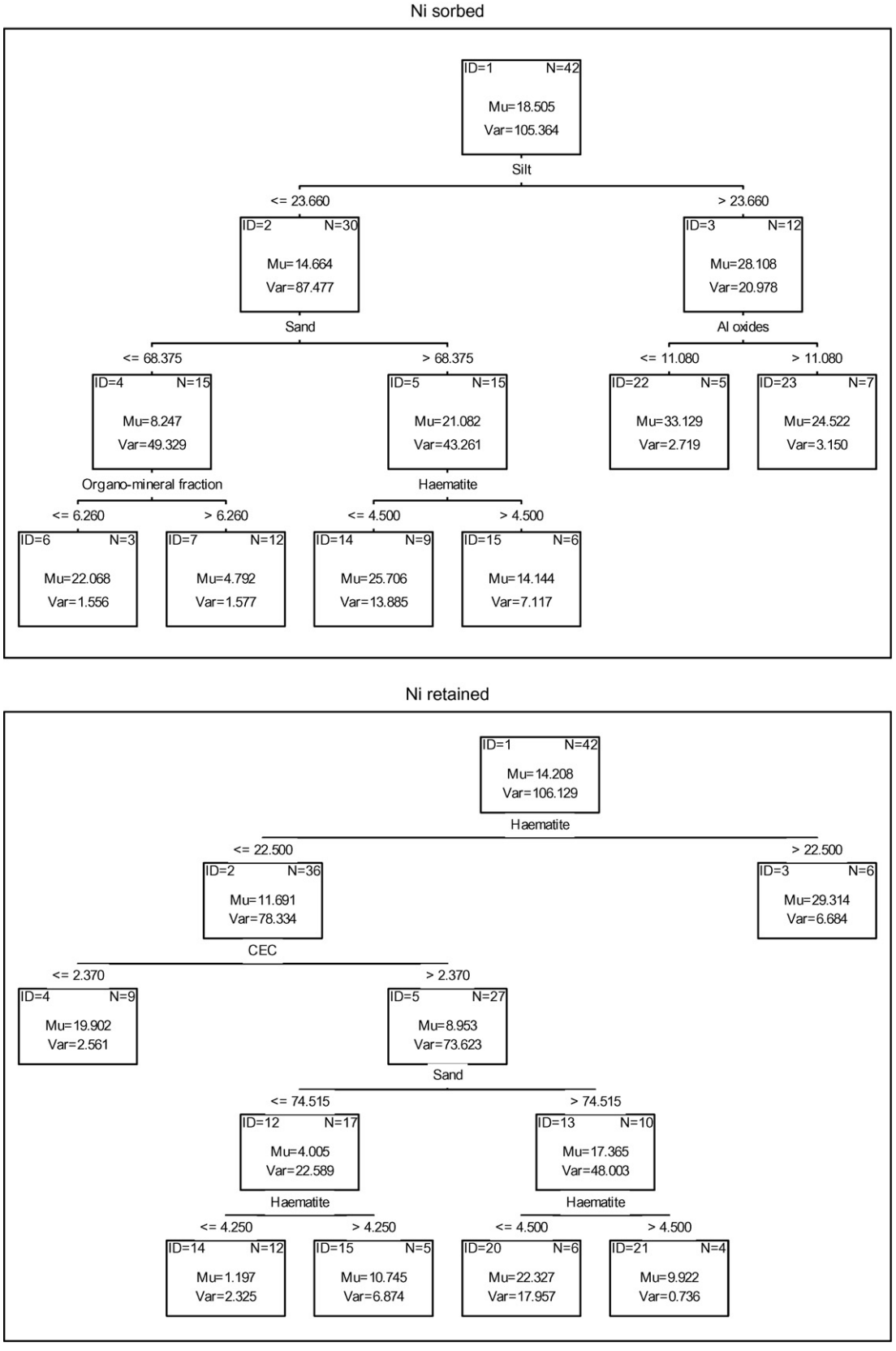

Fig. 5. CART binary regression tree for the sorption (top) and retention (bottom) of Ni by typical Galician soils.

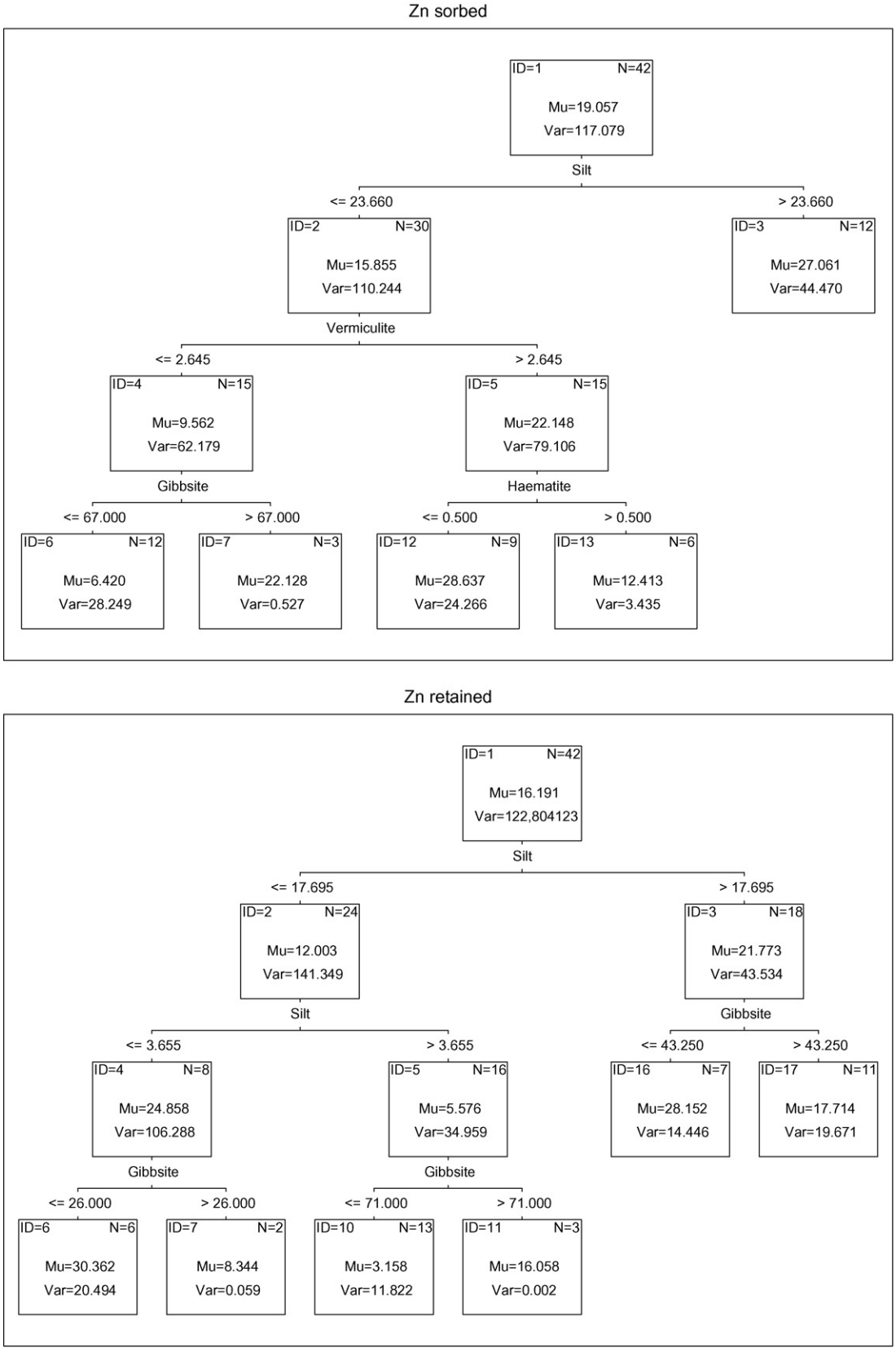

Fig. 6. CART binary regression tree for the sorption (top) and retention (bottom) of Zn by typical Galician soils.

were subsequently split on the basis of silt, sand, and organomineral contents. All the figures can be interpreted the same way; for a more detailed explanation of the practical use of CARTs, see Covelo et al. [22].

For Cu (Fig. 2), the variables haematite content and cation exchange capacity are used to split nodes in both the sorption and retention trees, accompanied by silt, organomineral content, vermiculite and chlorite in the sorption tree, and by sand and kaolinite in the retention tree. For Pb (Fig. 3), CEC_e, Mn oxides, gibbsite and vermiculite split nodes in both trees, accompanied in the sorption tree by haematite and kaolinite, and in the retention tree by sand and organomineral contents. For the less readily sorbed metals, the node-splitting variables common to the sorption and retention trees are Mn oxides, silt and vermiculite for Cd; sand and haematite for Ni; and silt and gibbsite for Zn (Figs. 4–6).

5. Conclusions

The extent to which Galician soils sorb and retain Pb, Cu, Cr, Cd, Ni and Zn in solutions containing all six is modelled in terms of soil properties much better by binary regression trees constructed using the CART algorithm than by stepwise linear regression models. Both kinds of model successfully fitted the data for Pb and Cu, the metals that are most readily sorbed and retained, but CART is clearly superior to linear regression for modelling the behaviour of the other metals, especially the markedly non-linear behaviour of Zn and Cd, which is influenced by competition from the more readily sorbed metals Cr and Cu.

Of the six metals competing for sorption sites in these experiments, Pb, Cu and Cr were sorbed and retained to a greater extent than Cd, Ni and Zn. Non-linear tree regression models constructed with CART fitted the data better than linear models, especially for Cd, Ni and Zn; and with both kinds of model the data for Pb, Cu and Cr were fitted better than those for Cd, Ni and Zn (the difference being much more marked for linear models), suggesting that the influence of soil properties on the sorption and retention of the latter three metals was limited by the preferential binding of the former three.

Acknowledgements

This work was supported by the Spanish Ministry of Education and Science under project CGL2006-01016/BTE and through a grant to J.M. Matías (MTM2005-00820), and by the Xunta de Galicia in partnership with the University of Vigo through a Parga Pondal research contract awarded to E.F. Covelo.

References

[1] U. Förstner, Land contamination by metals: global scope and magnitude of problem, in: H.E. Allen, C.P. Huang, G.W. Bailey, A.R. Bowers (Eds.), Metal Speciation and Contamination of Soil, Lewis, Boca Raton, FL, 1995, pp. 1–33.
[2] R.S. Swift, R.G. McLaren, Micronutrient adsorption by soil and soil colloids, in: G.H. Bolt, M.F. DeBoodt, M.H.B. Hayes, M.B. McBride (Eds.), Interactions at the Soil Colloid-Soil Solution Interface, Part 2, Kluwer Academic Publishers, Dordrecht, Netherlands, 1991, pp. 257–292.
[3] M. Arias, C. Pérez-Novo, F. Osorio, E. López, B. Soto, Adsorption and desorption of copper and zinc in the surface layer of acid soils, J. Colloid Interf. Sci. 288 (1) (2005) 21–29.
[4] K.M. Spark, J.D. Wells, B.B. Johnson, Characterizing trace metal adsorption on kaolinite, Eur. J. Soil Sci. 46 (1995) 633–640.
[5] M. Doula, A. Ioannou, A. Dimirkou, Influence of ionic strength and pH on Cu^{2+} adsorption and on Mg^{2+}, Ca^{2+}, Mn^{2+}, Zn^{2+} release by kaolinite, Agrochym. 43 (1999) 215–222.
[6] M.H. Al-Qunaibit, W.K. Mekhemer, A.A. Zaghloul, The adsorption of Cu(II) ions on bentonite-a kinetic study, J. Colloid Interf. Sci. 283 (2) (2004) 316–321.
[7] R.G. Gerriste, W. van Driel, The relationship between adsorption of trace metals, organic matter and pH in temperate soils, J. Environ. Qual. 13 (1984) 197–204.
[8] L.M. Shuman, Effect of removal of organic matter and iron- or manganese-oxides on zinc adsorption by soil, Soil Sci. 146 (1988) 248–254.
[9] N. Cavallaro, M.B. McBride, Zinc and copper sorption and fixation by an acid soil clay: effect of selective dissolutions, Soil Sci. Soc. Am. J. 48 (1984) 1050–1054.
[10] H.A. Elliott, M.R. Liberati, C.P. Huang, Effects of iron oxide removal on heavy metal sorption by acid subsoils, Water Air Soil Pollut. 27 (1985) 379–389.
[11] R.S. Stahl, B.R. James, Zinc sorption by manganese-oxide-coated sand as a function of pH, Soil Sci. Soc. Am. J. 55 (1991) 1291–12194.
[12] E.F. Covelo, M.L. Andrade, F.A. Vega, Heavy metal adsorption by humic umbrisols selectivity sequences and competitive sorption kinetics, J. Colloid Interf. Sci. 280 (2004) 1–8.
[13] E.F. Covelo, M.L. Andrade, F.A. Vega, Competitive adsorption and desorption of cadmium, chromium, copper, nickel, lead and zinc by humic umbrisols, Comm. Soil Sci. Plant Anal. 35 (19–20) (2004) 2709–2729.
[14] F.A. Vega, E.F. Covelo, M.L. Andrade, Competitive adsorption and desorption of heavy metals in minesoils: influence of minesoil characteristics, J. Colloid Interf. Sci. 298 (2) (2006) 582–592.
[15] E.F. Covelo, F.A. Vega, M.L. Andrade, Simultaneous sorption and desorption of Cd, Cr, Cu, Ni, Pb and Zn in acid soils. I. Selectivity sequences, J. Hazard. Mater. 147 (3) (2007) 852–861.
[16] E.F. Covelo, F.A. Vega, M.L. Andrade, Simultaneous sorption and desorption of Cd, Cr, Cu, Ni, Pb and Zn in acid soils. II. Soil ranking and influence of soil characteristics, J. Hazard. Mater. 47 (3) (2007) 862–870.
[17] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Wadsworth, Belmont, CA, 1984.
[18] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.
[19] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, 2001.
[20] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, New York, 1997.
[21] L. Györfi, M. Kohler, A. Krzyzak, H. Walk, A Distribution Free Theory of Nonparametric Regression, Springer, New York, 2002.
[22] E.F. Covelo, J.M. Matías, F.A. Vega, M.L. Reigosa, M.L. Andrade, A tree regression analysis of factors determining the sorption and retention of heavy metals by soil, Geoderma 147 (1–2) (2008) 75–85.