Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils

Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils

Kang Yu[a,*], Maarten Van Geel[b], Tobias Ceulemans[b], Willem Geerts[b], Miguel Marcos Ramos[c], Cindy Serafim[c], Nadine Sousa[c], Paula M.L. Castro[c], Pierre Kastendeuch[d], Georges Najjar[d], Thierry Ameglio[e], Jérôme Ngao[e], Marc Saudreau[e], Olivier Honnay[b] and Ben Somers[a]

[a]Department of Earth & Environmental Sciences, KU Leuven, 3001, Heverlee, Belgium; E-mails: kang.yu@kuleuven.be; ben.somers@kuleuven.be

[b]Department of Biology, KU Leuven, 3001, Heverlee, Belgium. E-mails: maarten.vangeel@kuleuven.be; tobias.ceulemans@kuleuven.be; willem.geerts@student.kuleuven.be; olivier.honnay@kuleuven.be

[c]Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua Arquiteto Lobão Vital, 172, 4200-374 Porto, Portugal. E-mails: mmramos@porto.ucp.pt; nsousa@porto.ucp.pt; plcastro@porto.ucp.pt

[d]Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie, Strasbourg University, Illkirch, France. E-mails: kasten@unistra.fr; georges.najjar@unistra.fr

[e]Université Clermont Auvergne, INRA, PIAF, F-63000 Clermont Ferrand, France. E-mails: thierry.ameglio@inra.fr; jerome.ngao@inra.fr; marc.saudreau@inra.fr

*Correspondence:
Kang Yu
E-mail: kang.yu@kuleuven.be
Division of Forest, Nature and Landscape
Department of Earth and Environmental Sciences
KU Leuven
Celestijnenlaan 200e - box 2411
3001 Leuven, Belgium

Abstract

Heavy metals in urban soils may impose a threat to public health and may negatively affect urban tree viability. Vegetation spectroscopy techniques applied to bio-indicators bring new opportunities to characterize heavy metal contamination, without being constrained by laborious soil sampling and lab-based sample processing. Here we used *Tilia tomentosa* trees, sampled across three European cities, as bio-indicators i) to investigate the impacts of elevated concentrations of cadmium (Cd) and lead (Pb) on leaf mass per area (LMA), total chlorophyll content (Chl), chlorophyll *a* to *b* ratio (Chla:Chlb) and the maximal PSII photochemical efficiency (Fv/Fm); and ii) to evaluate the feasibility of detecting Cd and Pb contamination using leaf reflectance spectra. For the latter, we used a partial-least-squares discriminant analysis (PLS-DA) to train spectral-based models for the classification of Cd and/or Pb contamination. We show that elevated soil Pb concentrations induced a significant decrease in the LMA and Chl*a*:Chl*b*, with no decrease in Chl. We did not observe pronounced reductions of Fv/Fm due to Cd and Pb contamination. Elevated Cd and Pb concentrations induced contrasting spectral changes in the red-edge (690~740 nm) region, which might be associated with the proportional changes in leaf pigments. PLS-DA models allowed for the classifications of Cd and Pb contamination, with a classification accuracy of 86% (Kappa=0.48) and 83% (Kappa=0.66), respectively. PLS-DA models also allowed for the detection of a collective elevation of soil Cd and Pb, with an accuracy of 66% (Kappa=0.49). This

study demonstrates the potential of using reflectance spectroscopy for biomonitoring of heavy metal contamination in urban soils.

Keywords: soil heavy metal contamination; leaf functional trait; vegetation reflectance spectroscopy; red-edge position; bio-indicator

Capsule

Applying leaf reflectance spectroscopy to urban trees allows for biomonitoring of heavy metal pollution and the classification of pollutants in urban soils.

Introduction

Soil contamination is a widely spread problem across Europe (European Commission, 2006). Among the most frequent soil pollutants are heavy metals such as arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), mercury (Hg), lead (Pb), zinc (Zn), antimony (Sb), cobalt (Co) and nickel (Ni), which accumulate on the soil surface and transfer to deeper soil layers where they can infiltrate into the groundwater (Vince et al., 2014). Plants growing on heavy metal polluted soils passively take up heavy metals, jeopardizing their growth and negatively affecting other organisms feeding on the plants (Panagos et al., 2013; Tóth et al., 2016). Furthermore, elevated concentrations of these heavy metals in agricultural or urban soils endanger food safety and public health (Poggio et al., 2009; Tóth et al., 2016).

Urban soils typically contain elevated concentrations of Cd, Cu, Zn and Pb, originating from anthropogenic activities such as traffic and industrial emissions (Gallagher et al., 2008; Li et al., 2001; Poggio et al., 2009; Pourkhabbaz et al., 2010; Vince et al., 2014). Cd and Pb are the most common heavy metals resulting from road traffic, which is attributed to the historical use of Pb as a gasoline additive (Kovarik, 2005) and Cd accumulation which is mainly due to abrasion of tires (Andersson et al., 2010; Vince et al., 2014). Cd and Pb are toxic for plants, animals and humans (Pandit et al., 2010; Poggio et al., 2009). Cd accumulates in human body and can cause nephropathy, pulmonary lesions and lung cancer after long period of exposure (Poggio et al., 2009). Pb increases blood pressure

and damages liver, kidney and fertility, and most severely it reduces brain functioning and induces hyperactivity and hearing loss in children (Poggio et al., 2009). Therefore, it is vital to detect elevated concentrations of Cd and Pb in urban soils.

Measuring heavy metals is typically based on the collection of soil or road dust samples, which is labor intensive and costly, especially when monitoring heavy metal contamination at larger spatial scales (Wei and Yang, 2010). In European countries, the estimated total annual cost related to monitoring and remediating soil contaminants is 17.3 billion euros (European Commission, 2006), and around 81% of the expenditures is spent on remediation measures (Liedekerke et al., 2014). Consequently, only up to 15% is available to be spent on site investigations (Liedekerke et al., 2014), implying that there is a need for more cost-effective investigation methods to evaluate spatial and temporal heterogeneity of soil pollution. Soil near-infrared (NIR) spectroscopy has been applied for the detection of heavy metals at relatively low cost. However, this method requires intensive soil sampling (Pandit et al., 2010; Shi et al., 2014). Therefore, a spatially explicit characterization of heavy metal contamination at large scales is constrained by the capacity of sampling and sample processing, especially in urban areas characterized by sealed soil surfaces and highly heterogeneous land-use types.

Bio-indicators are living organisms that can be used to assess the quality of the environment (Holt and Miller, 2010; Parmar et al., 2016). Urban vegetation can be used as bio-indicators for monitoring air and soil

79 pollution (Ho, 1990; Khavanin Zadeh et al., 2013; Sawidis et al., 2011).
80 Plants concentrate metal elements in their above ground parts, which are
81 indicative of elevated soil heavy metal concentrations. Furthermore, heavy
82 metals can inhibit plant growth (Giulia et al., 2013; Horler et al., 1980), and
83 decrease chlorophyll content and biomass productivity (Gallagher et al.,
84 2008; Manios et al., 2003). Cd and Pb often limit plant growth by altering
85 leaf internal structures (Giulia et al., 2013; Pourkhabbaz et al., 2010). For
86 instance, Cd can reduce cell wall extensibility and relative water content
87 (Barceló and Poschenrieder, 1990). Pb can reduce not only the leaf
88 expansion but also the total chlorophyll content and efficiency of PSII electron
89 transport (Kastori et al., 1998). Overall, heavy metal toxicity causes multiple
90 direct and indirect effects on various physiological functions and on the
91 morphology of plants (Barceló and Poschenrieder, 1990), reflected in
92 changes of leaf functional traits.

93 Metal induced morphological and physiological changes can further
94 alter vegetation absorbance and reflectance characteristics (Horler et al.,
95 1980). Typically, heavy metal contamination induces most notable changes
96 in the visible and NIR spectral regions, and thus reflectance spectroscopy
97 holds great promise for evaluating the impact of heavy metal contamination
98 on vegetation (Clevers et al., 2004; Kooistra et al., 2004, 2003; Rosso et al.,
99 2005). By applying reflectance spectroscopy to monitoring candidate bio-
100 indicators located at multiple sites in urban areas, researchers have been
101 able to detect polluted sites (Khavanin Zadeh et al., 2013). Previous studies
102 have investigated the effect of individual metals on vegetation spectral

responses, e.g., canopy reflectance in response to manipulated pot-soil Cd changes (Rosso et al., 2005). However, different metals may induce similar or contrasting spectral responses (Amer et al., 2017; Horler et al., 1980; Manios et al., 2003). Some studies have focused on spectral response in specific spectral bands such as the red-edge region (690~740 nm), which has been used to estimate plant chlorophyll variations under stress due to heavy metals (Clevers et al., 2004; Rosso et al., 2005). The red-edge position (REP) is defined as the position generating the maximum slope (inflection point) of the reflectance spectra (or maximum first derivative reflectance) in the red-edge region (Clevers et al., 2004; Horler et al., 1983), and has been found to be negatively related to soil Pb concentration (Clevers et al., 2004; Kooistra et al., 2004). Overall, associating soil heavy metal pollution with a range of plant functional and reflectance characteristics provides a cost-effective method for assessing heavy metal pollutions. However, there is still a lack of vegetation reflectance spectroscopy studies that bio-monitor Cd and Pb contamination across a variety of urban environments, especially for monitoring contamination due to multiple metals.

Here we tested *Tilia tomentosa* as a bio-indicator for elevated soil Cd and Pb concentrations. Selecting 187 study trees cross three European cities (Leuven, Porto and Strasbourg), our objectives were: i) to assess the impacts of elevated concentrations of Cd and Pb on leaf mass per area (LMA), total chlorophyll content (Chl), chlorophyll *a* to *b* ratio (Chla:Chlb) and the maximal PSII photochemical efficiency (Fv/Fm); and ii) to investigate the

feasibility of using leaf reflectance spectroscopy and partial-least-squares discriminant analysis for biomonitoring soil Cd and Pb contamination.

Materials and Methods

Sampling of leaf and soil and heavy metal measurements

We conducted soil and leaf sampling in summer 2017 and randomly selected 19 sites and 187 *T. tomentosa* trees across three medium sized cities (Leuven (Belgium): n = 64; Porto (Portugal), n = 67; Strasbourg (France): n = 56). We randomly selected trees for sampling, and the trunk diameter ranged 5-130 cm. For each tree, we sampled the top soils (0–10 cm) at three random locations surrounding the trunk, and the three locations are mixed for metal measurements. We sampled 15 leaves at three random positions in each tree and stored the leaf samples in a cool box with ice. We performed soil sampling once, while leaf sampling was performed multiple times throughout the growing season for a subset of trees in Leuven and Strasbourg.

Heavy metal concentrations in the soil were measured by digesting 50 mg of dried and sieved soil with 7.5 ml concentrated hydrochloric acid and 2.5 ml concentrated nitric acid. The digested solution was diluted to 10 ml and measured with ICP-OES. For quality control of soil metal analysis, an internal soil standard was run parallel with the soil samples, which deviated less than 5% of the known composition. In this study, we focused on Cd and Pb, as these were the heavy metals that reached the toxicity thresholds (Table 1).

Identification of contamination based on soil heavy metal thresholds

Soil heavy metal contamination levels were identified based on published threshold standards (Tóth et al., 2016) released by the Ministry of the Environment, Finland (MEF, 2007). We grouped the samples into two classes - non-contaminated and contaminated, subjected to individual metals (Table 1). Soil samples and corresponding leaf spectral observations (section 2.3) were grouped into four classes according to Pb contamination following the MEF standard (MEF, 2007). The four classes included class 0 being non-contaminated (Pb < 60 mg/kg), class 1 of low contamination (60 $\leqslant$ Pb < 200 mg/kg), class 2 of medium contamination (200 $\leqslant$ Pb < 750 mg/kg) and class 3 of high contamination (Pb $\geqslant$ 750 mg/kg).

We also defined four contamination classes subjected to both Cd and Pb contamination by re-grouping of the Cd and Pb binary classes (Table S1), i.e., four CdxPb classes including the non-contaminated (class 0), Cd contaminated only (class 1), Pb contaminated only (class 2) as well as when both Cd and Pb are over the thresholds (class 3).

Leaf reflectance and functional traits

Leaf reflectance was measured using an ASD FieldSpec 3 spectroradiometer (ASD Inc., Longmont, CO, USA) connected to a Plant Probe and Leaf Clip Assembly (ASD Inc., Longmont, CO, USA). It allows for reflectance measurement in a spectral range of 350 – 2500nm with a band width of 1 nm. Next, we measured the leaf maximal PSII photochemical efficiency (Fv/Fm, ratio of the variable fluorescence to the maximal fluorescence) using

a chlorophyll fluorescence meter (Handy PEA, Hansatech Instruments Ltd., Pentney, UK), combined with a leaf clip that allows for dark adaption (25 min). Then, we measured the leaf area using a flatbed scanner, followed by oven dry for 3 days, allowing to determine leaf mass per area (LMA). In total, aggregated per tree and sampling time, collected leave samples allowed for further statistical analysis on a sample size of 333 for reflectance and functional traits. The 333 observations of reflectance spectra and functional traits were grouped into their contamination classes subjected to the soil heavy metal contamination classes as defined in the Section 2.2.

A random subset of the leaf samples (n=53) were used to determine the total chlorophyll (Chl) and carotenoid (Car) content. Leaf round discs with a diameter of 28.6 mm were punched from the leaf samples using a paper punch. Chla, Chlb and Car were extracted with a mortar and pestle in 80% acetone and their concentrations determined by measuring the solution absorbance (A) at wavelengths 470, 646.8 and 663.2 nm using a UV-VIS spectrophotometer (Shimadzu 1650 PC, Kyoto, Japan) according to Eqs. (1-3) (Lichtenthaler, 1987).

$$Chla = 12.25 * A_{663.2} - 2.79 * A_{646.8} \#(1)$$

$$Chlb = 21.50 * A_{646.8} - 5.10 * A_{663.2} \#(2)$$

$$Car = \frac{1000 * A_{470} - 1.82 * Chla - 85.02 * Chlb}{198} \#(3)$$

For quality control of chlorophyll analysis, we performed parallel measurements in 12 samples, and the average standard error was lower than 5%.

Spectral and statistical analysis

To highlight the metal-induced spectral variations, we calculated the reflectance relative differences between group means for the contaminated and non-contaminated classes subjected to Cd and Pb contamination. We also applied first derivatives to the reflectance, focusing mainly on the red-edge region, to derive the red-edge inflection point (REIP) and evaluate the metal induced red-edge shifts (Clevers et al., 2004).

Partial least squares (PLS) regression is a multivariate method for relating two data matrices, X and Y, i.e., explanatory and response matrices, by extracting latent variables (components) to model the variations of both matrices (Wold et al., 2001). The PLS regression can reduce high dimensional data (e.g. hyperspectral) to a small number of latent variables which serve as new predictors on which the response variable is regressed (Rosipal and Krämer, 2006). Partial least squares discriminant analysis (PLS-DA) is a variant used when the response variable is categorical. We used PLS-DA for the classification of metal contamination classes. PLS-DA models were applied to four types of data, (i) the original reflectance spectral, and three pre-processed spectral data including (i) first derivative (ii), standard normal variate SNV and (iii) continuum removal (CR) precede applying the PLS-DA models. PLS-DA model calibration was first initiated on the entire dataset for the full spectrum with 10 components. The initial model was

trained using a 10-fold cross-validation with 99 times of permutations, allowing for determination of the optimal number of components and the spectral bands yielding a variable importance in projection (VIP) $\geqslant$ 0.8.

For an independent validation, the entire dataset was randomly split into the training and test subsets, with a sample size being 2/3 (n=215) and 1/3 (n=118) of the total observations (n=333), respectively. The VIP $\geqslant$ 0.8 spectral bands were then used to train and test models on the two subsets, respectively.

PLS-DA Model classification accuracy was evaluated using the overall accuracy (Eq. 4) and kappa coefficient (Eq. 5), as well as for assessing the classification for individual classes using the producer's (Eq. 7) and user's accuracies (Eq. 8),

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \#(4)$$

$$Kappa = \frac{p_a - p_e}{1 - p_e} \#(5)$$

$$p_e = \frac{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)}{(TP + TN + FP + FN)^2} \#(6)$$

$$Producer\ Accuracy = TP/(TP + FP) \#(7)$$

$$User\ Accuracy = TP/(TP + FN) \#(8)$$

where the letters *T* and *F* denote *true* and *false*, respectively, and *P* and *N* denote *positive* and *negative*, respectively, p_a is the actual agreement (identical to accuracy), whereas p_e is the expected agreement by chance (random accuracy) that can be calculated as Eq. (6).

We used linear mixed models to test whether elevated soil heavy metals affect the leaf functional traits. We defined the metal contamination classes, i.e., binary or multi-class, as the fixed effect factor and defined city and sampling site as random effect factors in the mixed models. All analyses were performed in the R programming environment (R Core Team, 2016). The R package 'lme4' (Bates et al., 2015) was used for running the mixed models, and the package 'lsmeans' (Lenth, 2016) was used for post-hoc analysis of pairwise comparisons between the contaminated classes based on Tukey's test. PLS-DA was implemented using the package 'mixOmics' (Rohart et al., 2017).

Results and Discussion

Heavy metal effects on leaf functional traits

Elevated Pb and Cd concentrations had a significant effect on LMA of *T. tomentosa* trees (Table 2). Soil Cd contamination did not induce significant changes in LMA (Fig. 1a), whereas Pb contamination significantly decreased LMA (Fig. 1b). Generally, Cd and Pb stress leads to damages to chloroplasts and thylakoid membranes in plants (Shen et al., 2016; Wu et al., 2014), which often causes reduced leaf growth such as small leaf size and small stomata (Shi and Cai, 2009), as well as thin cuticles of leaf surfaces (Pourkhabbaz et al., 2010). Therefore, elevated Pb concentrations could have reduced leaf thickness and thus decreased LMA. Cd also induces changes in leaf structural properties, while Cd concentrations measured in this study might still be below the threshold that induces significant inhibition of leaf expansion.

Elevated soil Pb induced significant changes in leaf total Chl content, Chl*a* to Chl*b* ratio (Chla:Chlb) and Fv/Fm, whereas Cd and other metals did not yield significant changes (Table 2). Decrease in leaf Chl content is often associated with photoinhibition and reduction of the photosynthetic capacity (Shen et al., 2016). Chla:Chlb decreased significantly along with the increase in soil Pb concentration (Fig. 2), suggesting that Chla was more suppressed compared to Chlb (Nie et al., 2016). Similarly, a significant reduction of Chla:Chlb has been found in *Torreya grandis* (Shen et al., 2016) and *Typha latifolia* plants (Manios et al., 2003) treated with a high concentration of Cd and Pb, suggesting increases in chlorophyll hydrolysis due to the toxic effect (Manios et al., 2003). Results may differ for different plant species, for instance in a greenhouse environment, Horler et al. (1980) observed a significant decrease of Chla:Chlb in pea leaves due to elevated Cd concentrations, but no changes following elevated Pb (Horler et al., 1980).

Cd and Pb contamination induced a decrease in Fv/Fm (Fig. 3a, b), whereas Fv/Fm appeared to be not sensitive to low-level Pb contamination (Fig. 3d), suggesting that Cd and Pb stress may induce photosynthesis inhibition. Similarly, Cd was found to affect Fv/Fm in the wetland plant species *Salicornia virginica* (Rosso et al., 2005) and in the turf grass species *Festuca arundinacea Schreb* (Huang et al., 2017). Generally, the observed decrease in Fv/Fm in plants subjected to Cd/Pb stress is associated with the photoinhibition of PSII, as a result of the overproduction of reactive oxygen species (ROS) (Huang et al., 2017; Shen et al., 2016). However, a significant decrease in Fv/Fm may not always be observable if Cd/Pb concentration does

not exceed a high threshold (Huang et al., 2017; Shen et al., 2016). Giulia et al. (2013) found that a high soil Pb concentration did not decrease Fv/Fm in *Q. ilex* plants, and they argued that these metals may not significantly alter functionality of the photosynthetic apparatus. Similarly, Shi and Cai (2009) reported that Fv/Fm was not affected in peanut plants treated with a high concentration of Cd. Therefore, the effect of heavy metals on Fv/Fm might depend largely on metal type, concentration and plant species.

Mixed models for multi-class CdxPb and Pb contamination showed much more pronounced effects on LMA and Chla:Chlb than on Fv/Fm and leaf total Chl content (Table 3), which suggests that heavy metals induced more structural changes and proportional changes in leaf biochemicals than the quantity changes of individual components. An increase in leaf total Chl content and Fv/Fm was observed at a relative low-level Pb or Cd×Pb contamination (Table 3), suggesting that heavy metals impose complicated effects on photosynthesis and that Cd and Pb may increase the PSII quantum yield within a certain range of low concentrations (Ouyang et al., 2012; Shen et al., 2016).

The effect of soil heavy metals on leaves or the content of heavy metal accumulation in the leaves might be related to the age of trees (Doganlar et al., 2012). To test whether tree age difference affect the observed effects of Cd and Pb on leaf functional traits in this study, we used trunk diameter as a proxy of tree age and added it as an additional random factor in the mixed models (Table S2 and Table S3). Results suggest that the observed effects of

309 Cd and Pb on T. *tomentosa* leaves was not significantly influenced by tree
310 age.

311 *Reflectance and first derivatives in response to heavy metals*
312 Elevated soil Cd concentrations yielded relatively large variations in leaf
313 reflectance centered at the 500, 680 and 720 nm bands (Fig. 4a), whereas
314 elevated Pb yielded large variations at the 550 and 700 nm bands (Fig. 4b).
315 In the red-edge region, Cd had a large effect on reflectance at the red-edge
316 center (~720 nm), whereas Pb had a large effect on reflectance ranging from
317 the red absorption to the beginning of the red-edge bands (680~700 nm).
318 Over the full spectrum, soil Pb contamination induced larger variations
319 (±10%, Fig. 4b) compared to Cd contamination (±5%) (Fig. 4a), which
320 might be attributed to the fact that Pb contamination was severer than Cd in
321 this study. Cd concentration was slightly higher than the threshold (1
322 mg/kg), but was much lower than the 'low guideline' of contamination level
323 (10 mg/kg) at which ecological or health risks present (Tóth et al., 2016).

324 The decrease in the NIR region (750~1400 nm) was associated with
325 elevated Cd and Pb concentrations. This might be attributed partly to the
326 decreased LMA because contaminated trees often have a much thinner outer
327 epidermal layer and thus thinner leaves (Pourkhabbaz et al., 2010), although
328 the effect of Cd on LMA observed in this study was marginal (Fig. 1). Metal-
329 induced decreases in leaf NIR reflectance might be associated mainly with
330 the changes in leaf internal structural properties which decrease the internal

331 light scattering and increase the transmittance of leaves (Horler et al., 1980;
332 Kumar et al., 2001).

333 The first derivative reflectance in the visible-to-NIR bands showed two
334 major peaks centered at 530 and 720 nm (Fig. S1). In the red-edge spectral
335 region, Cd contamination induced a shift of absorbance features towards the
336 shorter wavelengths (Fig. S1a). In contrast, Pb contamination induced a red-
337 edge shift to the longer wavelengths (Fig. S1b). In addition to the red-edge
338 bands, Pb contamination also yielded large variations in the first derivative
339 reflectance at the green bands, suggesting a more pronounced change of the
340 overall shape of reflectance (cf. Fig. 4). As shown in the first derivative
341 reflectance, Pb contamination also induced a shift in the green edges (both
342 sides of the green peak) compared to Cd contamination. This might explain
343 the observed decrease in the Chla:Chlb ratio (Fig. 2), since absorption at the
344 green edge bands is related to Chlb variations (Kumar et al., 2001).

345 The extracted REIP showed contrasting changes in the Cd and Pb
346 contaminated trees, with decreasing and increasing trends, respectively (Fig.
347 S2), which confirms the contrasting effects of Cd and Pb contamination on
348 the red-edge reflectance. Heavy-metal induced REIP changes, or red-edge
349 shifts, have been found to depend to some degree on plant species and
350 sampling sites (Kooistra et al., 2004). Normally, a decreased REIP can be
351 observed when plant stress induces a reduction in leaf total Chl content
352 (Horler et al., 1983). However, here we did not observe obvious Chl
353 reduction associated with Cd or Pb contamination. Therefore, the REIP
354 variations observed here were more likely associated with the proportional

355 changes in the Chla:Chlb ratio, in combination with changes in leaf
356 structures.

357 *PLS-DA model calibration for binary and multi-class classifications*
358 In the binary classifications, the PLS-DA calibration models for Cd-
359 contamination classification yielded a total accuracy of 84.1~86.5% (kappa =
360 0.46~0.49, Table S4). PLS-DA models for Pb contamination yielded a total
361 accuracy of 72.7~77.8% (kappa = 0.46~0.57). For the multi-class
362 classification of CdxPb-mixed contamination, PLS-DA models yielded a total
363 accuracy of 43.2~66.1% (kappa = 0.24~0.49, Table S4). PLS-DA models for
364 the multi-class classification of Pb yielded a total accuracy of 52.0~64.0%
365 (kappa = 0.29~0.43). The best classifications for individual metals are
366 illustrated in confusion-matrix plots (Fig. 5).

367 The best model for Cd correctly classified the Cd class 0 with a
368 producer and use accuracy of 86% and 97%, respectively, and were 77%
369 and 43% for the Cd class 1 (Fig. 5a). The producer and use accuracy for the
370 Pb class 0 were 88% and 67%, respectively, and 77% and 90% for the Pb
371 class 1 (Fig. 5b). The best model for Cd×Pb yielded a relatively low user
372 accuracy in predicting the classes 1 and 3 (Fig. 5c), which however, accounts
373 for a very small proportion of the total observations. The best model for
374 multi-class Pb contamination yielded a relatively high producer accuracy for
375 the classes 0 and 3 (Fig. 5d), with 80% and 100%, respectively. In contrast,
376 the model yielded a higher user accuracy for the classes 0 and 1 than for the
377 classes 2 and 3. The low user accuracy for the Pb classes 2 and 3 was mainly

due to the small sample size of high Pb concentrations, which consists of only 17 and 4 observations for the classes 2 and 3, respectively.

Overall, the high producer accuracy, paired with relatively low user accuracy for a relatively high metal concentration was rather encouraging, since our models slightly tended to overestimate the observed contamination rather than underestimate the elevated contamination. This implies a high probability of detecting the elevated concentrations of soil heavy metals.

PLS-DA model validation using full spectrum and VIP-bands

Compared to model calibration accuracies, model validation based on the full spectrum produced comparable accuracies (Table S5). In binary classifications, models for Pb contamination yielded higher kappa coefficients than the models for Cd contamination. In multi-class classifications, model validation showed improved total accuracies and kappa coefficients (Table S5), suggesting the potential of using calibrated PLS-DA models for detecting elevated soil Cd and Pb concentrations.

Validation of models trained with the VIP ($\geqslant$ 0.8) bands showed slightly improved kappa values and total accuracies compared to the full use of bands (Table S5). The importance of individual spectral bands in the classification is indicated by the VIP scores for individual metals (Fig. 6). Cd contamination yielded relatively high VIP scores at the red-edge (730 nm) and SWIR bands (1300 nm, 1650 nm) compared to Pb contamination, suggesting unique spectral responses to elevated soil Cd in these bands (Fig. 6). Pb contamination yielded higher VIP scores at the green (530 nm) and

401 the beginning of red-edge (700 nm), suggesting that Pb contamination
402 induced more pronounced responses in the visible bands. For the binary
403 classifications, VIP-based PLS-DA models yielded higher accuracies for Pb-
404 contamination classification (kappa = 0.66) than for Cd (kappa= 0.39, Table
405 S5). For multi-class classifications, the VIP-based PLS-DA models yielded
406 comparable accuracies by using a much less amount of bands compared to
407 the use of full spectral bands.

408 Model validation results showed that selecting a set of influential bands
409 (VIP $\geqslant$ 0.8) allowed for maintaining classification accuracy and improving
410 model-use and computational efficiencies. Within a limited number of
411 observations, by randomly dividing independent training and testing subsets
412 of observations, our results suggest that spectrally calibrated PLS-DA models
413 have great potential of applying to future scenarios for monitoring heavy
414 metals.

415 *Comparison between reflectance pre-processing methods*

416 The kappa coefficient is a balanced measure compared to the use of the
417 producer-, user- and total accuracies, especially when the observations in
418 difference classes are highly imbalanced such as in this study. Hence, we
419 evaluated the three spectra-preprocessing methods according to the kappa
420 values. Model calibration and validation both showed that the first derivatives
421 yielded the highest kappa values compared to the use of the original and
422 SNV reflectance data (Table S4 and Table S5).

Using a different number of components might induce some degree of variation in model accuracies, although we used the cross-validation (CV) procedure. In addition to the CV-optimized number of components, model calibration and validation were repeated by using a fixed number of components (Table S6, Table S7 and Table S8). Results showed that the first derivative reflectance yielded the highest kappa coefficients, followed by the CR reflectance and the original reflectance (Fig. S3). The SNV reflectance did not yield improvement compared to the original reflectance data, suggesting that the SNV process may mask subtle spectral responses subjected to individual metals. Overall, PLS-DA models based on the first derivative reflectance produced the best classifications, which also suggests that heavy metals have induced complicated effects on leaf biochemical and structural properties that lead to light absorption changes/shifts over the full spectrum.

First derivative spectra of leaves have been proven to be effective in eliminating background signals and for resolving overlapping spectral features (Demetriades-Shah et al., 1990), which is useful to detect plant stresses or estimate pigment changes (Rundquist et al., 1996; Smith et al., 2004). Also, first derivative reflectance has better discrimination power compared to the original reflectance by characterizing the rate of change of reflectance with respect to wavelengths (Bao et al., 2013; Lassalle et al., 2018; Smith et al., 2004). Typically, derivative analysis may facilitate the detection of changes that might be masked in the original spectra by the presence of plant intrinsic co-variations (Horler et al., 1983). For instance, derivative spectra in the visible region may enable to detect subtle changes

447 in leaf pigment balance associated with physiological disorders or vegetation
448 types (Bandaru et al., 2016; Demetriades-Shah et al., 1990; Pu, 2011).

449 Derivative analysis can be particularly useful for remotely biomonitoring
450 heavy metal using reflectance spectra measured from above the vegetation
451 canopy (Wang et al., 2018). Canopy spectra first derivatives eliminate the
452 additive noises (baseline shifts) induced by illumination instability, canopy
453 structural or soil background influences (Demetriades-Shah et al., 1990;
454 Gnyp et al., 2014; Kochubey and Kazantsev, 2012; Pu, 2011), thereby
455 improving the accuracy for quantification of canopy biochemical or
456 physiological changes (Jin and Wang, 2016; O'Connell et al., 2014).
457 Moreover, PLS modeling further facilitates the use of features of the full
458 derivative spectrum for the characterization of vegetation undergoing
459 changes or stresses.

460 Apparently, PLS-DA models for Pb-contamination classifications
461 exclusively produced higher kappa values than for Cd contamination
462 classifications, across different cases of spectra-preprocessing methods,
463 model calibration (Fig. S3a) and validation (Fig. S3b), as well as when using
464 a subset of VIP-bands (Fig. S3c). This can be attributed to the data
465 imbalance between the Cd- and Pb-contamination levels, which, however,
466 shows a great promise of the proposed approach for spectroscopic detection
467 of elevated soil heavy metals, given that a diverse set of observations are
468 used for model calibration.

Conclusions

This study used *T. Tomentosa* trees growing in three European cities as bio-indicators of soil heavy metal contamination, and evaluated whether tree spectra responses were able to reflect the elevated metal concentrations. Results showed that elevated soil Cd and Pb concentrations led to decrease in the leaf mass per area (LMA) and the chlorophyll *a* to *b* ratio (Chla:Chlb), while no significant reduction in leaf total chlorophyll (Chl) and the maximal PSII photochemical efficiency (Fv/Fm). Soil Pb contamination was severer and showed more pronounced effect on LMA, Fv/Fm, Chl and Chla:Chlb than did the Cd contamination in the studied sites.

Cd and Pb contamination induced specific changes in leaf reflectance and the reflectance first derivatives, particularly in the red-edge spectral region. Partial least squares discriminant analysis (PLS-DA) models calibrated using leaf reflectance showed promise for detecting soil Cd and Pb contamination in urban areas. PLS-DA models based on reflectance first derivatives allowed for the best classification of Cd and Pb contamination. This study shows that elevated soil heavy metals can be monitored by measuring leaf spectra of trees. This holds great potential for mapping urban heavy metal contamination by measuring urban vegetation using high-resolution spectrometers onboard airborne or drone platforms. Future work should investigate whether our findings can be extrapolated to broader scales by using canopy level reflectance data and a diverse set of plant species as bio-indicators. Multi-temporal investigations of the quantitative relationships between the practical content of heavy metals in leaves and reflectance

493 spectroscopic measures are also needed to understand metal translocation
494 from soil to vegetation and for dynamic biomonitoring of heavy metal
495 contamination.

496 Acknowledgements

497 This research was funded through the 2015-2016 BiodivERsA COFUND call
498 for research proposals, with the national funders BelSPo (BE), FWO (BE), FCT
499 (PT) and through the project UID/Multi/50016/2013. We thank Remi
500 Chevalier, the greenery service of the Cities of Leuven, Porto (Câmara
501 Municipal do Porto) and Strasbourg for their assistance in fieldwork and
502 measurements. The authors thank Remi Chevalier and Yasmin Vanbrabant
503 for their assistance in chlorophyll analysis.

504 Declarations of interest

505 None.

506 References

507 Amer, M., Tyler, A., Fouda, T., Hunter, P., Elmetwalli, A., Wilson, C., Mario
508 VALLEJO-MARIN, 2017. Spectral Characteristics for Estimation Heavy
509 Metals Accumulation in Wheat Plants and Grain. Sci. Pap. Ser. Manag.
510 Econ. Eng. Agric. Rural Dev. 17, 47–55.

511 Andersson, M., Ottesen, R.T., Langedal, M., 2010. Geochemistry of urban
512 surface soils — Monitoring in Trondheim, Norway. Geoderma 156,
513 112–118. https://doi.org/10.1016/j.geoderma.2010.02.005

514 Bandaru, V., Daughtry, C.S., Codling, E.E., Hansen, D.J., White-Hansen, S.,
515 Green, C.E., 2016. Evaluating Leaf and Canopy Reflectance of Stressed
516 Rice Plants to Monitor Arsenic Contamination. Int. J. Environ. Res.
517 Public. Health 13. https://doi.org/10.3390/ijerph13060606

518 Bao, J., Chi, M., Benediktsson, J.A., 2013. Spectral Derivative Features for
519 Classification of Hyperspectral Remote Sensing Images: Experimental
520 Evaluation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6, 594–
521 601. https://doi.org/10.1109/JSTARS.2013.2237758

522 Barceló, J., Poschenrieder, C., 1990. Plant water relations as affected by
523 heavy metal stress: A review. J. Plant Nutr. 13, 1–37.
524 https://doi.org/10.1080/01904169009364057

525 Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-
526 Effects Models Using lme4. J. Stat. Softw. 67, 1–48.
527 https://doi.org/10.18637/jss.v067.i01

528 Clevers, J.G.P.W., Kooistra, L., Salas, E.A.L., 2004. Study of heavy metal
529 contamination in river floodplains using the red-edge position in
530 spectroscopic data. Int. J. Remote Sens. 25, 3883–3895.
531 https://doi.org/10.1080/01431160310001654473

532 Demetriades-Shah, T.H., Steven, M.D., Clark, J.A., 1990. High resolution
533 derivative spectra in remote sensing. Remote Sens. Environ. 33, 55–
534 64. https://doi.org/10.1016/0034-4257(90)90055-Q

535 Doganlar, Z.B., Doganlar, O., Erdogan, S., Onal, Y., 2012. Heavy metal
536 pollution and physiological changes in the leaves of some shrub, palm
537 and tree species in urban areas of Adana, Turkey. Chem. Speciat.
538 Bioavailab. 24, 65–78.
539 https://doi.org/10.3184/095422912X13338055043100

540 European Commission, 2006. Commission staff working document -
541 Document accompanying the Communication from the Commission to
542 the Council, The European Parliament, the European Economic and
543 Social Committee and the Committee of the Regions - Thematic
544 Strategy for Soil Protection - Impact assessment of the thematic
545 strategy on soil protection {COM(2006)231 final} {SEC(2006)1165}.
546 Commission of the European Communities, Brussels.

547 Gallagher, F.J., Pechmann, I., Bogden, J.D., Grabosky, J., Weis, P., 2008.
548 Soil metal concentrations and productivity of Betula populifolia (gray
549 birch) as measured by field spectrometry and incremental annual
550 growth in an abandoned urban Brownfield in New Jersey. Environ.
551 Pollut. 156, 699–706. https://doi.org/10.1016/j.envpol.2008.06.013

552 Giulia, M., Lucia, S., Carmen, A., 2013. Heavy metal accumulation in leaves
553 affects physiological performance and litter quality of Quercus ilex L. J.
554 Plant Nutr. Soil Sci. 176, 776–784.
555 https://doi.org/10.1002/jpln.201200053

556 Gnyp, M.L., Miao, Y., Yuan, F., Ustin, S.L., Yu, K., Yao, Y., Huang, S., Bareth,
557 G., 2014. Hyperspectral canopy sensing of paddy rice aboveground
558 biomass at different growth stages. Field Crops Res. 155, 42–55.
559 https://doi.org/10.1016/j.fcr.2013.09.023

560 Ho, Y.B., 1990. Ulva lactuca as bioindicator of metal contamination in
561 intertidal waters in Hong Kong. Hydrobiologia 203, 73–81.
562 https://doi.org/10.1007/BF00005615

563 Holt, E.A., Miller, S.W., 2010. Bioindicators: Using Organisms to Measure
564 Environmental Impacts | Learn Science at Scitable. Nat. Educ. Knowl.
565 3, 8.

566 Horler, D.N.H., Barber, J., Barringer, A.R., 1980. Effects of heavy metals on
567 the absorbance and reflectance spectra of plants. Int. J. Remote Sens.
568 1, 121–136. https://doi.org/10.1080/01431168008547550

569 Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf
570 reflectance. Int. J. Remote Sens. 4, 273–288.
571 https://doi.org/10.1080/01431168308948546

572 Huang, M., Zhu, H., Zhang, J., Tang, D., Han, X., Chen, L., Du, D., Yao, J.,
573 Chen, K., Sun, J., 2017. Toxic effects of cadmium on tall fescue and
574 different responses of the photosynthetic activities in the photosystem
575 electron donor and acceptor sides. Sci. Rep. 7.
576 https://doi.org/10.1038/s41598-017-14718-w

577 Jin, J., Wang, Q., 2016. Hyperspectral indices based on first derivative
578 spectra closely trace canopy transpiration in a desert plant. Ecol.
579 Inform. 35, 1–8. https://doi.org/10.1016/j.ecoinf.2016.06.004

580 Kastori, R., Plesničar, M., Sakač, Z., Panković, D., Arsenijević-Maksimović, I.,
581 1998. Effect of excess lead on sunflower growth and photosynthesis. J.
582 Plant Nutr. 21, 75–85. https://doi.org/10.1080/01904169809365384

583 Khavanin Zadeh, A.R., Veroustraete, F., Buytaert, J.A.N., Dirckx, J., Samson,
584 R., 2013. Assessing urban habitat quality using spectral characteristics
585 of Tilia leaves. Environ. Pollut. 178, 7–14.
586 https://doi.org/10.1016/j.envpol.2013.02.021

587 Kochubey, S.M., Kazantsev, T.A., 2012. Derivative vegetation indices as a
588 new approach in remote sensing of vegetation. Front. Earth Sci. 6,
589 188–195. https://doi.org/10.1007/s11707-012-0325-z

590 Kooistra, L., Leuven, R.S.E.W., Wehrens, R., Nienhuis, P.H., Buydens, L.M.C.,
591 2003. A comparison of methods to relate grass reflectance to soil
592 metal contamination. Int. J. Remote Sens. 24, 4995–5010.
593 https://doi.org/10.1080/0143116031000080769

594 Kooistra, L., Salas, E.A.L., Clevers, J.G.P.W., Wehrens, R., Leuven, R.S.E.W.,
595 Nienhuis, P.H., Buydens, L.M.C., 2004. Exploring field vegetation
596 reflectance as an indicator of soil contamination in river floodplains.
597 Environ. Pollut. 127, 281–290. https://doi.org/10.1016/S0269-
598 7491(03)00266-5

599 Kovarik, W., 2005. Ethyl-leaded Gasoline: How a Classic Occupational
600 Disease Became an International Public Health Disaster. Int. J. Occup.
601 Environ. Health 11, 384–397.
602 https://doi.org/10.1179/oeh.2005.11.4.384

603 Kumar, L., Schmidt, K., Dury, S., Skidmore, A., 2001. Imaging Spectrometry
604 and Vegetation Science, in: van der Meer, F.D., De Jong, S.M. (Eds.),
605 Imaging Spectrometry - Basic Principles and Prospective Applications,

606 Remote Sensing and Digital Image Processing. Springer Netherlands,
607 Dordrecht, pp. 111–155. https://doi.org/10.1007/978-0-306-47578-
608 8_5

609 Lassalle, G., Credoz, A., Hédacq, R., Fabre, S., Dubucq, D., Elger, A., 2018.
610 Assessing Soil Contamination Due to Oil and Gas Production Using
611 Vegetation Hyperspectral Reflectance. Environ. Sci. Technol. 52,
612 1756–1764. https://doi.org/10.1021/acs.est.7b04618

613 Lenth, R.V., 2016. Least-Squares Means: The R Package lsmeans. J. Stat.
614 Softw. 69, 1–33. https://doi.org/10.18637/jss.v069.i01

615 Li, X., Poon, C., Liu, P.S., 2001. Heavy metal contamination of urban soils
616 and street dusts in Hong Kong. Appl. Geochem. 16, 1361–1368.
617 https://doi.org/10.1016/S0883-2927(01)00045-2

618 Lichtenthaler, H.K., 1987. [34] Chlorophylls and carotenoids: Pigments of
619 photosynthetic biomembranes, in: Methods in Enzymology, Plant Cell
620 Membranes. Academic Press, pp. 350–382.
621 https://doi.org/10.1016/0076-6879(87)48036-1

622 Liedekerke, M. van, Prokop, G., Rabl-Berger, S., Kibblewhite, M., Louwagie,
623 G., European Commission, Joint Research Centre, Institute for
624 Environment and Sustainability, 2014. Progress in the management of
625 contaminated sites in Europe. Publications Office of the European
626 Union, Luxembourg.

627 Manios, T., Stentiford, E.I., Millner, P.A., 2003. The effect of heavy metals
628 accumulation on the chlorophyll concentration of Typha latifolia plants,
629 growing in a substrate containing sewage sludge compost and watered
630 with metaliferus water. Ecol. Eng. 20, 65–74.
631 https://doi.org/10.1016/S0925-8574(03)00004-1

632 MEF, 2007. Government Decree on the Assessment of Soil Contamination
633 and Remediation Needs (No. 214/2007). Ministry of the Environment,
634 Finland.

635 Nie, J., Liu, Y., Zeng, G., Zheng, B., Tan, X., Liu, H., Xie, J., Gan, C., Liu, W.,
636 2016. Cadmium accumulation and tolerance of Macleaya cordata: a
637 newly potential plant for sustainable phytoremediation in Cd-
638 contaminated soil. Environ. Sci. Pollut. Res. 23, 10189–10199.
639 https://doi.org/10.1007/s11356-016-6263-7

640 O'Connell, J.L., Byrd, K.B., Kelly, M., 2014. Remotely-Sensed Indicators of N-
641 Related Biomass Allocation in Schoenoplectus acutus. PLoS ONE 9.
642 https://doi.org/10.1371/journal.pone.0090870

643 Ouyang, H., Kong, X., He, W., Qin, N., He, Q., Wang, Y., Wang, R., Xu, F.,
644 2012. Effects of five heavy metals at sub-lethal concentrations on the
645 growth and photosynthesis of Chlorella vulgaris. Chin. Sci. Bull. 57,
646 3363–3370. https://doi.org/10.1007/s11434-012-5366-x

647 Panagos, P., Van Liedekerke, M., Yigini, Y., Montanarella, L., 2013.
648 Contaminated Sites in Europe: Review of the Current Situation Based

649 on Data Collected through a European Network. J. Environ. Public
650 Health 2013, 11. https://doi.org/10.1155/2013/158764

651 Pandit, C.M., Filippelli, G.M., Li, L., 2010. Estimation of heavy-metal
652 contamination in soil using reflectance spectroscopy and partial least-
653 squares regression. Int. J. Remote Sens. 31, 4111–4123.
654 https://doi.org/10.1080/01431160903229200

655 Parmar, T.K., Rawtani, D., Agrawal, Y.K., 2016. Bioindicators: the natural
656 indicator of environmental pollution. Front. Life Sci. 9, 110–118.
657 https://doi.org/10.1080/21553769.2016.1162753

658 Poggio, L., Vrščaj, B., Schulin, R., Hepperle, E., Ajmone Marsan, F., 2009.
659 Metals pollution and human bioaccessibility of topsoils in Grugliasco
660 (Italy). Environ. Pollut. 157, 680–689.
661 https://doi.org/10.1016/j.envpol.2008.08.009

662 Pourkhabbaz, A., Rastin, N., Olbrich, A., Langenfeld-Heyser, R., Polle, A.,
663 2010. Influence of Environmental Pollution on Leaf Properties of Urban
664 Plane Trees, Platanus orientalis L. Bull. Environ. Contam. Toxicol. 85,
665 251–255. https://doi.org/10.1007/s00128-010-0047-4

666 Pu, R., 2011. Detecting and Mapping Invasive Plant Species by Using
667 Hyperspectral Data, in: Thenkabail, P.S., Lyon, J.G., Huete, A. (Eds.),
668 Hyperspectral Remote Sensing of Vegetation. CRC Press, Boca Raton,
669 FL, p. 447.

670 R Core Team, 2016. R: A language and environment for statistical
671 computing. R Foundation for Statistical Computing, Vienna, Austria.

672 Rohart, F., Gautier, B., Singh, A., Cao, K.-A.L., 2017. mixOmics: an R
673 package for 'omics feature selection and multiple data integration.
674 bioRxiv 108597. https://doi.org/10.1101/108597

675 Rosipal, R., Krämer, N., 2006. Overview and Recent Advances in Partial Least
676 Squares, in: Subspace, Latent Structure and Feature Selection,
677 Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp.
678 34–51. https://doi.org/10.1007/11752790_2

679 Rosso, P.H., Pushnik, J.C., Lay, M., Ustin, S.L., 2005. Reflectance properties
680 and physiological responses of Salicornia virginica to heavy metal and
681 petroleum contamination. Environ. Pollut. 137, 241–252.
682 https://doi.org/10.1016/j.envpol.2005.02.025

683 Rundquist, D.C., Han, L., Schalles, J.F., Peake, J.S., 1996. Remote
684 Measurement of Algal Chlorophyll in Surface Waters: The Case for the
685 First Derivative of Reflectance Near 690 nm. Photogramm. Eng.
686 Remote Sens. 62, 195–200.

687 Sawidis, T., Breuste, J., Mitrovic, M., Pavlovic, P., Tsigaridas, K., 2011. Trees
688 as bioindicator of heavy metal pollution in three European cities.
689 Environ. Pollut. 159, 3560–3570.
690 https://doi.org/10.1016/j.envpol.2011.08.008

691 Shen, J., Song, L., Müller, K., Hu, Y., Song, Y., Yu, W., Wang, H., Wu, J.,
692 2016. Magnesium Alleviates Adverse Effects of Lead on Growth,
693 Photosynthesis, and Ultrastructural Alterations of Torreya grandis
694 Seedlings. Front. Plant Sci. 7.
695 https://doi.org/10.3389/fpls.2016.01819

696 Shi, G., Cai, Q., 2009. Leaf plasticity in peanut (Arachis hypogaea L.) in
697 response to heavy metal stress. Environ. Exp. Bot. 67, 112–117.
698 https://doi.org/10.1016/j.envexpbot.2009.02.009

699 Shi, T., Chen, Y., Liu, Y., Wu, G., 2014. Visible and near-infrared reflectance
700 spectroscopy—An alternative for monitoring soil contamination by
701 heavy metals. J. Hazard. Mater. 265, 166–176.
702 https://doi.org/10.1016/j.jhazmat.2013.11.059

703 Smith, K.L., Steven, M.D., Colls, J.J., 2004. Use of hyperspectral derivative
704 ratios in the red-edge region to identify plant stress responses to gas
705 leaks. Remote Sens. Environ. 92, 207–217.
706 https://doi.org/10.1016/j.rse.2004.06.002

707 Tóth, G., Hermann, T., Da Silva, M.R., Montanarella, L., 2016. Heavy metals
708 in agricultural soils of the European Union with implications for food
709 safety. Environ. Int. 88, 299–309.
710 https://doi.org/10.1016/j.envint.2015.12.017

711 Vince, T., Szabó, G., Csoma, Z., Sándor, G., Szabó, S., 2014. The spatial
712 distribution pattern of heavy metal concentrations in urban soils — a
713 study of anthropogenic effects in Berehove, Ukraine. Cent. Eur. J.
714 Geosci. 6, 330–343. https://doi.org/10.2478/s13533-012-0179-7

715 Wang, F., Gao, J., Zha, Y., 2018. Hyperspectral sensing of heavy metals in
716 soil and vegetation: Feasibility and challenges. ISPRS J. Photogramm.
717 Remote Sens. 136, 73–84.
718 https://doi.org/10.1016/j.isprsjprs.2017.12.003

719 Wei, B., Yang, L., 2010. A review of heavy metal contaminations in urban
720 soils, urban road dusts and agricultural soils from China. Microchem. J.
721 94, 99–107. https://doi.org/10.1016/j.microc.2009.09.014

722 Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of
723 chemometrics. Chemom. Intell. Lab. Syst., PLS Methods 58, 109–130.
724 https://doi.org/10.1016/S0169-7439(01)00155-1

725 Wu, M., Wang, P.-Y., Sun, L.-G., Zhang, J.-J., Yu, J., Wang, Y.-W., Chen, G.-
726 X., 2014. Alleviation of cadmium toxicity by cerium in rice seedlings is
727 related to improved photosynthesis, elevated antioxidant enzymes and
728 decreased oxidative stress. Plant Growth Regul. 74, 251–260.
729 https://doi.org/10.1007/s10725-014-9916-x

Figure Captions

Fig. 1. Boxplots with the leaf mass per area (LMA) differences between the binary classes (0 = non-contaminated, 1 = contaminated) of (a) Cd and (b) Pb contamination, as well as among multiple classes of (c) Cd×Pb and (d) Pb contamination. Significance levels are indicated according to the post-hoc Tukey's test of the applied mixed models.

Fig. 2. Boxplots with the leaf chlorophyll a to b ratio (Chla:Chlb) differences between the binary classes (0 = non-contaminated, 1 = contaminated) of (a) Cd and (b) Pb contamination, as well as among multiple classes for (c) Cd×Pb and (d) Pb contamination. Significance levels are indicated according to the post-hoc Tukey's test of the applied mixed models.

Fig. 3. Boxplots show the chlorophyll fluorescence Fv/Fm differences between the binary classes (0 = non-contaminated, 1 = contaminated) of (a) Cd and (b) Pb contamination, as well as among multi-class classifications of (c) Cd×Pb and (d) Pb contamination. Significance levels are indicated according to the post-hoc Tukey's test of the applied mixed models.

Fig. 4. Leaf mean reflectance of the contaminated (1) and non-contaminated (0) trees subjected to (a) Cd and (b) Pb, and their reflectance relative difference ((X1-X0)/X0) between the contaminated and non-contaminated leaves.

Fig. 5. Predicted versus observed classes for (a) Cd binary classification, (b) Pb binary classification, (c) Cd×Pb classification and (d) Pb multi-class classification. Here the first derivative reflectance data were used for (a), (b) and (c), the original reflectance were used for (d). Numbers indicate the confusion matrix of classification.

Fig. 6. The variable importance in projection (VIP) scores for the spectral-based PLS-DA models for binary classification for Cd and Pb contamination, and for multi-class classification of Pb and CdxPb contamination. VIP $\geqslant$ 0.8 highlights the spectral bands contributing significantly to the PLS-DA models.

Tables

Table 1. Measured soil heavy metal content and the threshold values for classification of contamination. Cd and Pb were the major contaminates in this study, and Pb was the only metal that reached the highline and thus Pb contamination was classified into three sub-classes.

Metal	Range (mg/kg)	Threshold (mg/kg)	Number of observations (n)			
			Class 0	Class 1/Pb 1	Pb 2	Pb 3
Cd	0-3.9	1	294	**39**		
Pb*	0-2170.8	60	132	**201**/180	17	4
Co	0-15.9	20	333	0		
Cr	0-120.9	100	327	6		
Cu	0-159.1	100	330	3		
Ni	0-76.8	50	331	2		
Zn	10-265.8	200	329	4		

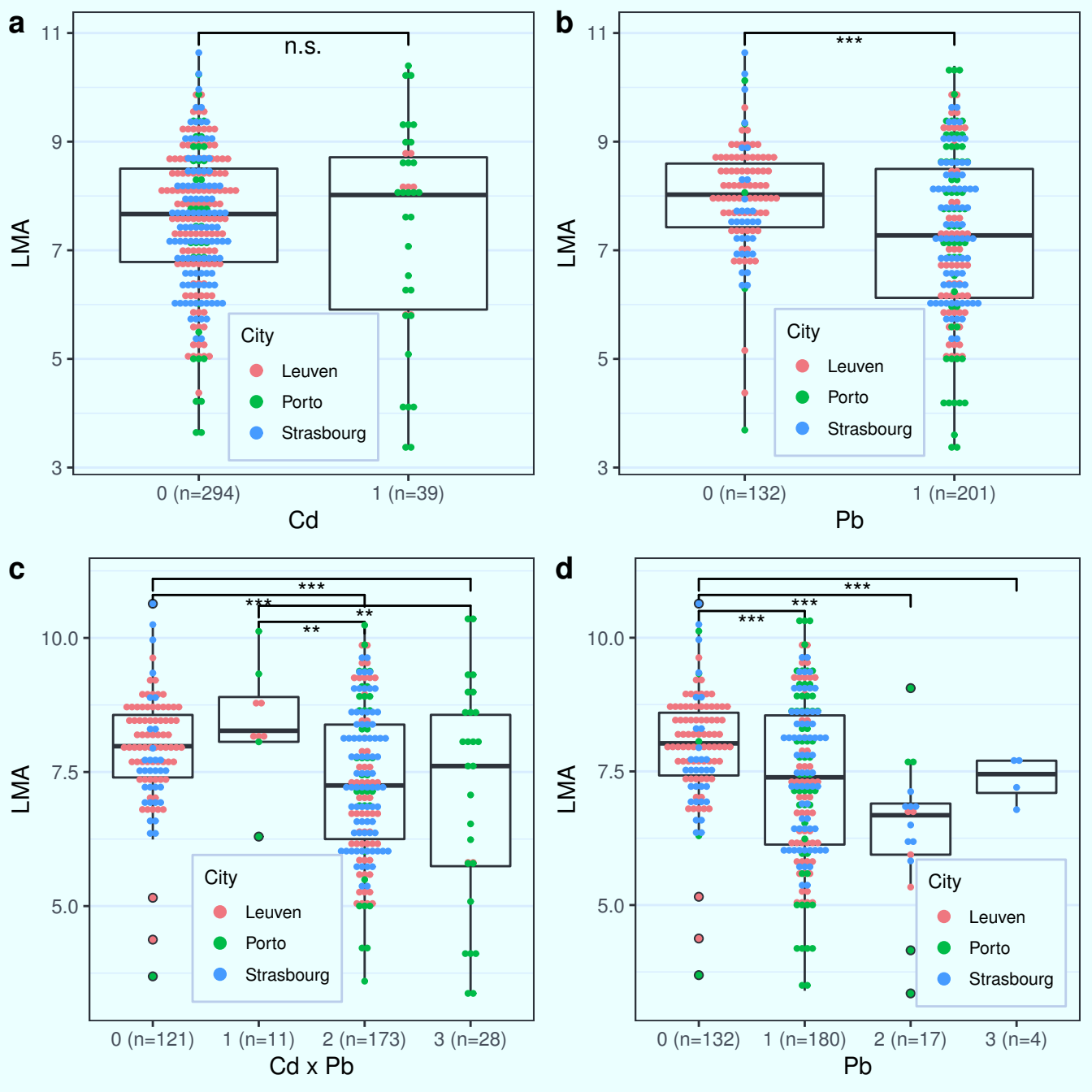
*, Pb contamination sub-levels:
1) Low contamination (60 ≤ Pb < 200 mg/kg);
2) Medium contamination (200 ≤ Pb < 750 mg/kg);
3) High contamination (Pb ≥ 750 mg/kg).

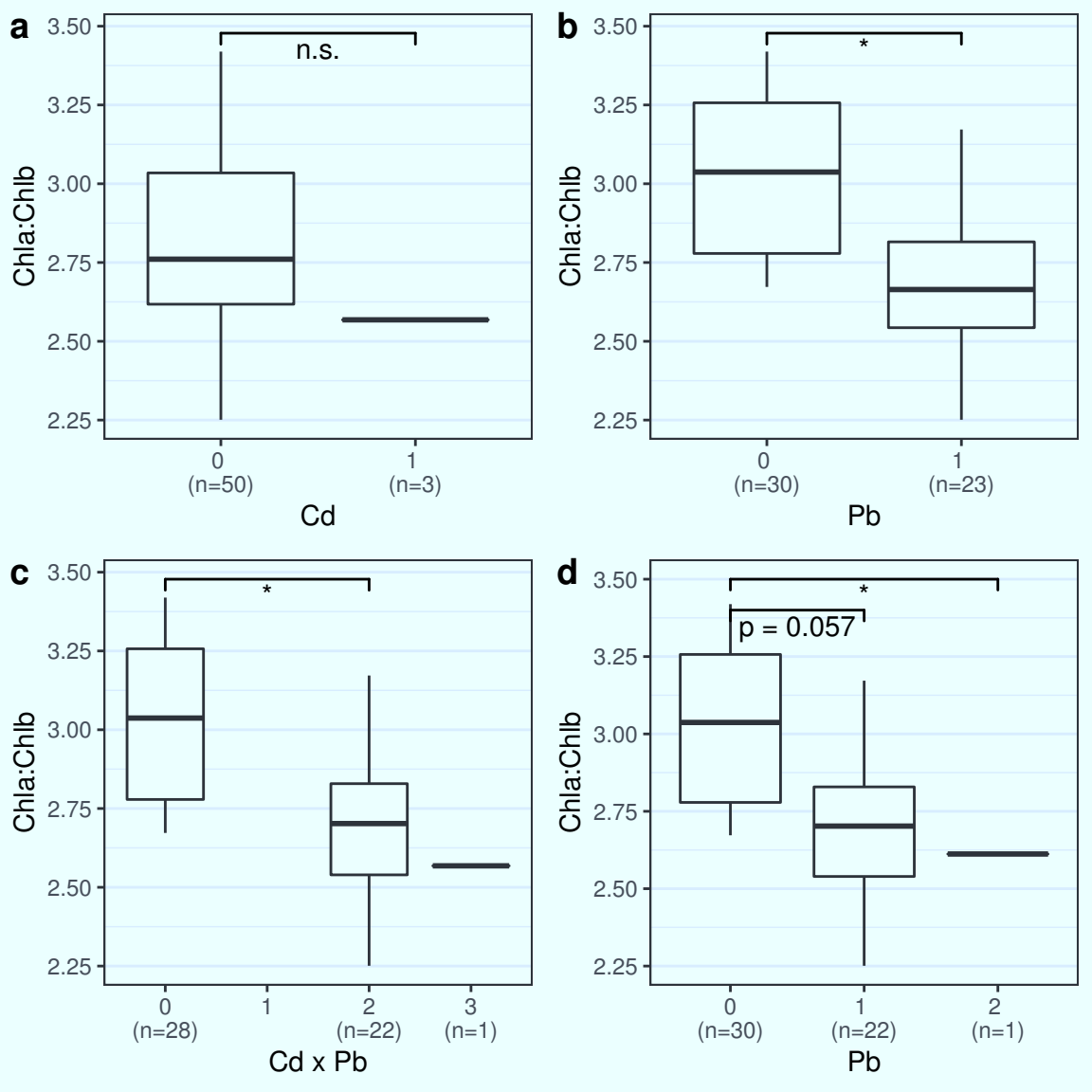
Table 2. Results of mixed models for testing the effect of soil heavy metals on leaf functional traits, including the leaf mass per area (LMA), Fv/Fm, total chlorophyll content (Chl) and Chla:Chlb ratio. Modelle random effects were city and sites. Chlorophyll data were only available for a subset of the samples, where only Cd and Pb reached the thresholds of contamination.

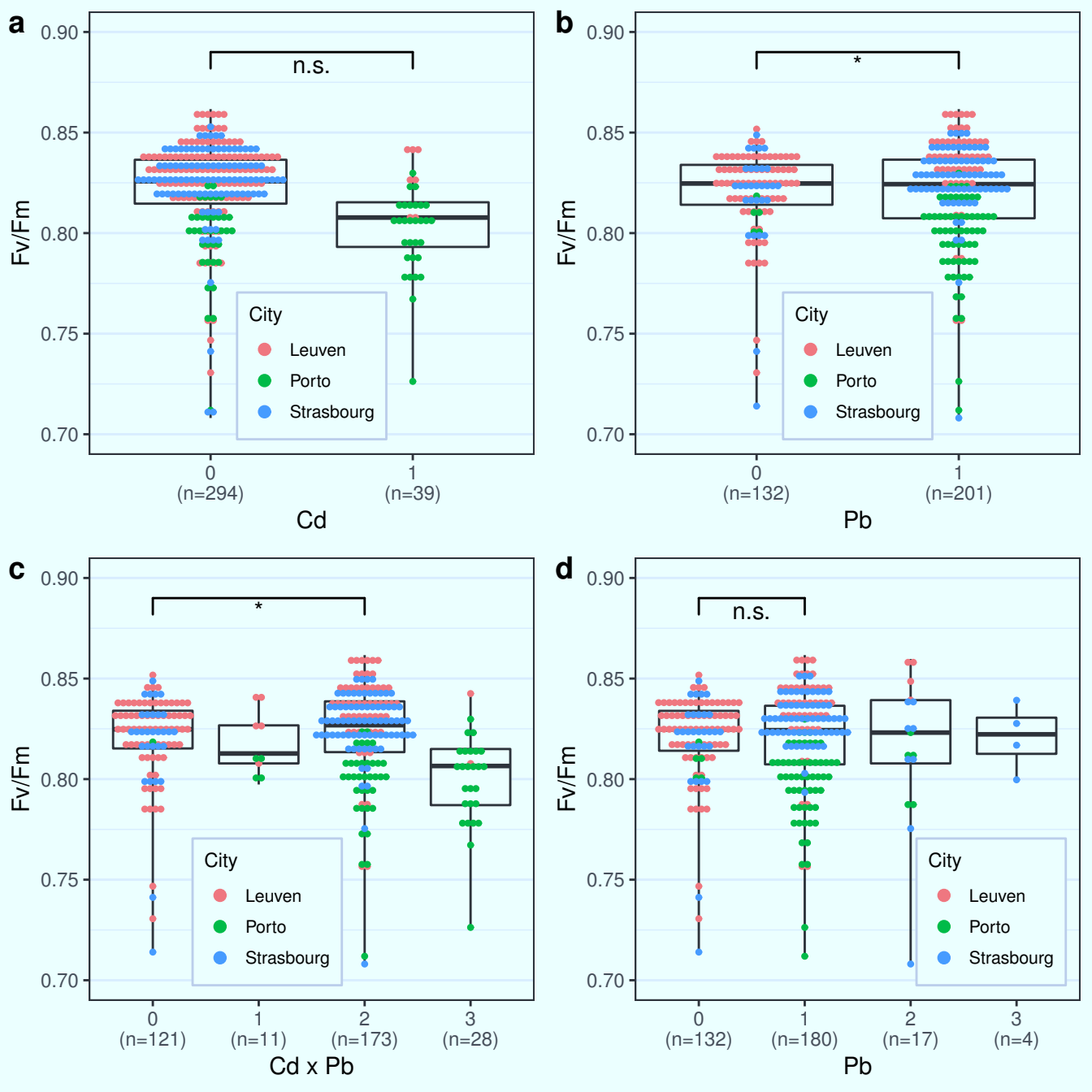
		Mixed Model		Tukey's Test Class 0 – 1	
Trait	Metal	*F*-value	*P*-value	Estimate	*P*-value
LMA	Cd	1.11	0.292	0.249	0.292
	Cr	6.68	**0.010**	-1.319	**0.010**
	Cu	0.16	0.691	0.284	0.691
	Ni	0.23	0.632	0.425	0.632
	Pb	67.08	**<0.001**	1.284	**<0.001**
	Zn	0.70	0.404	0.521	0.404
Fv/Fm	Cd	0.02	0.901	-0.0013	0.901
	Cr	0.01	0.905	0.0027	0.905
	Cu	0.01	0.911	0.0034	0.911
	Ni	0.08	0.772	-0.0109	0.772
	Pb	5.84	**0.016**	-0.0162	**0.016**
	Zn	0.08	0.784	-0.0074	0.784
Chl	Cd	2.31	0.138	18.091	0.138
	Pb	6.78	**0.013**	-9.238	**0.013**
Chla:Chlb	Cd	0.45	0.509	0.181	0.509
	Pb	23.58	**<0.001**	0.331	**<0.001**

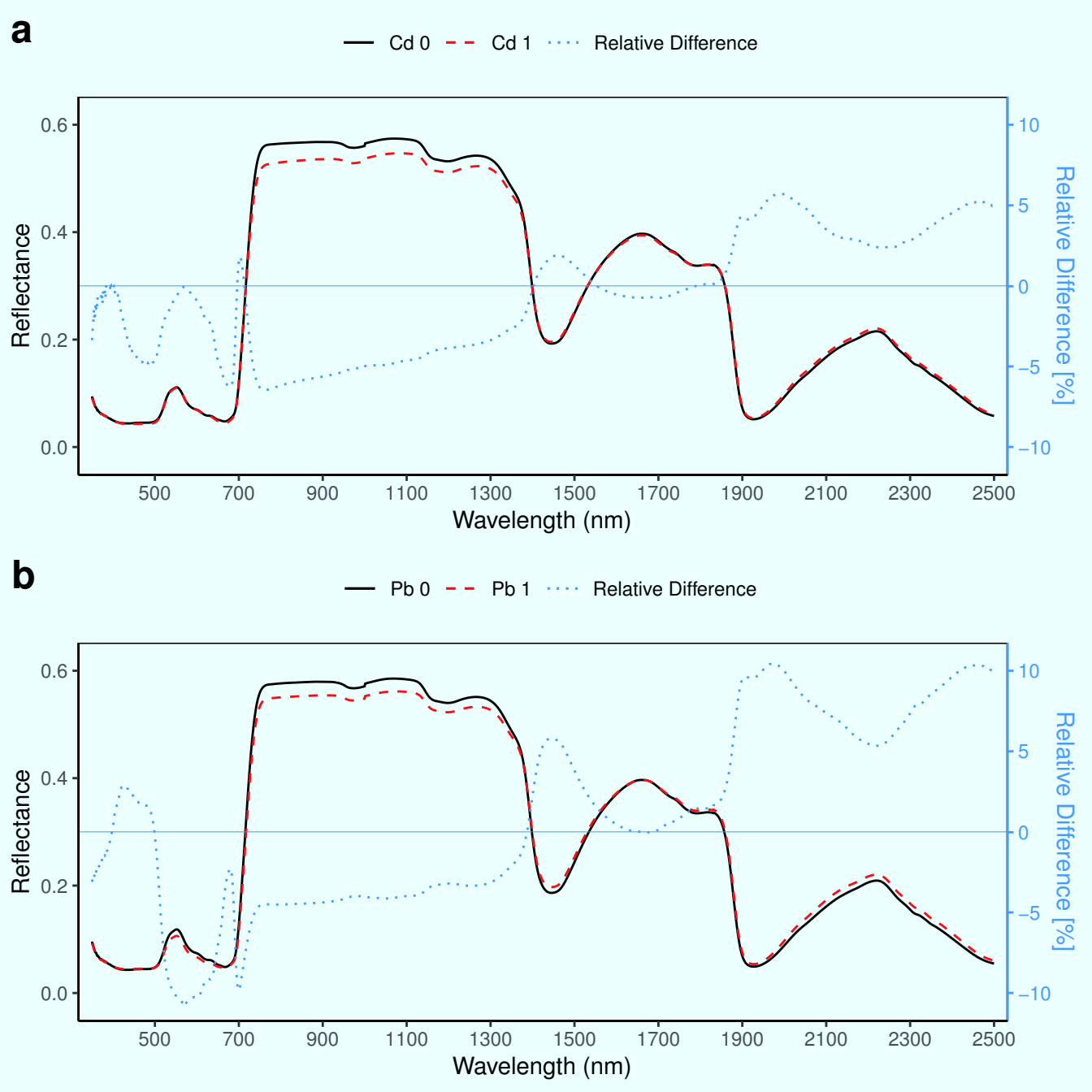
Table 3. Mixed models for testing the effect of multi-level Cd×Pb and Pb contamination on leaf functional traits, including the leaf mass per area (LMA), Fv/Fm, total chlorophyll content (Chl) and Chla:Chlb ratio. The modeled random effects are city and site. Chlorophyll data were only available for a subset of the samples, where only Cd and Pb reached the threshold of contamination.

Mixed model				Tukey's Test											
				Class 0 – 1		Class 0 - 2		Class 0 - 3		Class 1 - 2		Class 1 - 3		Class 2 – 3	
Trait	Metal	*F*-value	*P*-value	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*P*
LMA	Cd×Pb	23.74	**<0.001**	0.025	1.000	1.256	**<0.001**	1.722	**<0.001**	1.231	**0.006**	1.698	**0.001**	0.466	0.246
	Pb	26.29	**<0.001**	1.21	**<0.001**	1.831	**<0.001**	2.444	**<0.001**	0.621	0.104	1.234	0.104	0.613	0.731
Fv/fm	Cd×Pb	2.42	0.066	-0.0211	0.661	-0.0184	**0.044**	-0.0143	0.713	0.0028	0.999	0.0069	0.987	0.0041	0.987
	Pb	2.00	0.113	-0.0168	0.070	-0.0109	0.863	-0.0106	0.980	0.0059	0.971	0.0062	0.996	0.0003	1.000
Chl	Cd×Pb	5.86	**0.006**			-10.275	**0.014**	12.558	0.495					22.834	0.108
	Pb	6.40	**0.004**	-8.108	0.057	-32.96	**0.012**			-24.852	0.070				
Chla:Chlb	Cd×Pb	11.47	**<0.001**			0.332	**<0.001**	0.319	0.319					-0.012	0.998
	Pb	12.01	**<0.001**	0.323	**<0.001**	0.495	0.072			0.172	0.709				


a
n.s.
LMA
11
9
7
5
3
City
Leuven
Porto
Strasbourg
0 (n=294)
1 (n=39)
Cd
b
***
LMA
11
9
7
5
3
City
Leuven
Porto
Strasbourg
0 (n=132)
1 (n=201)
Pb
c
***
***
**
**
LMA
10.0
7.5
5.0
City
Leuven
Porto
Strasbourg
0 (n=121)
1 (n=11)
2 (n=173)
3 (n=28)
Cd x Pb
d
***
***
***
LMA
10.0
7.5
5.0
City
Leuven
Porto
Strasbourg
0 (n=132)
1 (n=180)
2 (n=17)
3 (n=4)
Pb

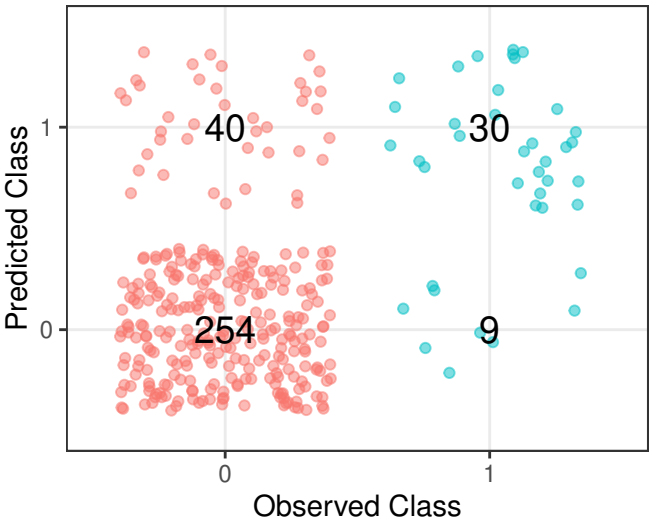

a
b
c
d
Chla:Chlb
3.50
3.25
3.00
2.75
2.50
2.25
n.s.
*
p = 0.057
0
(n=50)
1
(n=3)
Cd
0
(n=30)
1
(n=23)
Pb
0
(n=28)
1
2
(n=22)
3
(n=1)
Cd x Pb
0
(n=30)
1
(n=22)
2
(n=1)
Pb


a
b
c
d
Fv/Fm
0.90
0.85
0.80
0.75
0.70
n.s.
*
City
Leuven
Porto
Strasbourg
0
(n=294)
1
(n=39)
Cd
0
(n=132)
1
(n=201)
Pb
0
(n=121)
1
(n=11)
2
(n=173)
3
(n=28)
Cd x Pb
0
(n=132)
1
(n=180)
2
(n=17)
3
(n=4)
Pb


a
Cd 0
Cd 1
Relative Difference
Reflectance
Relative Difference [%]
Wavelength (nm)
b
Pb 0
Pb 1
Relative Difference
Reflectance
Relative Difference [%]
Wavelength (nm)

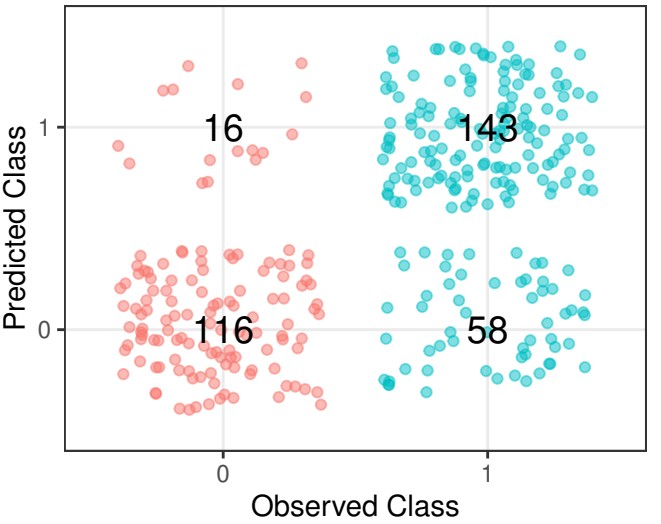
a

Class	ProducerAccuracy	UserAccuracy
0	86.39%	96.58%
1	76.92%	42.86%



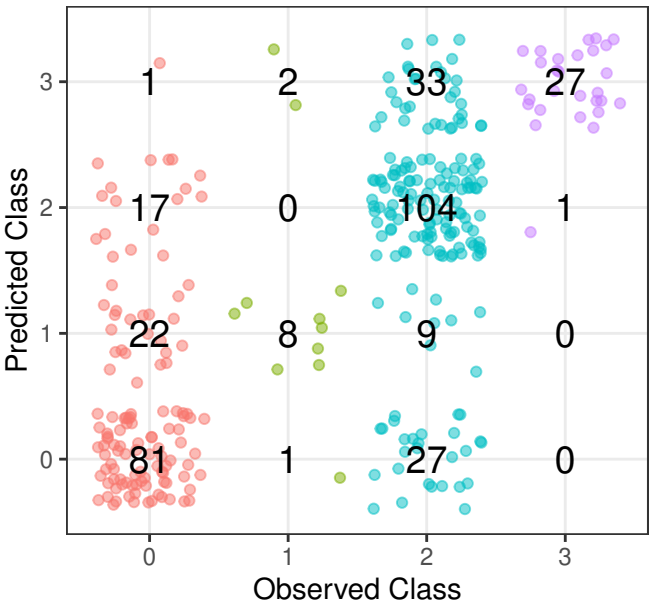
b

Class	ProducerAccuracy	UserAccuracy
0	87.88%	66.67%
1	71.14%	89.94%



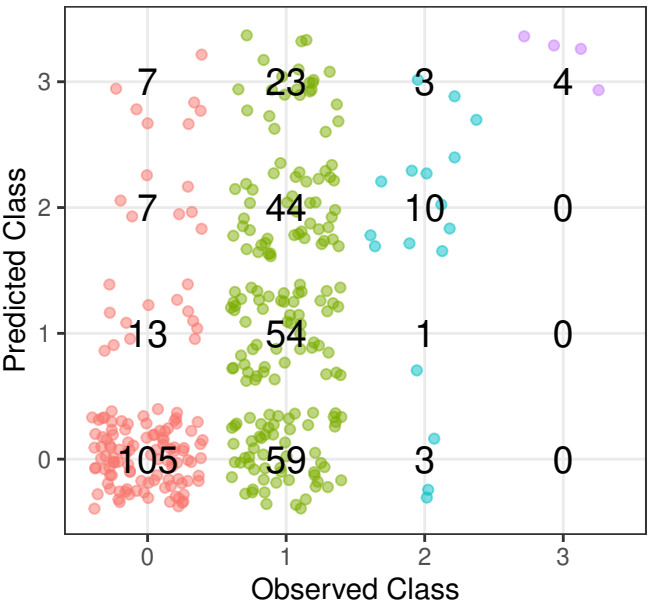
c

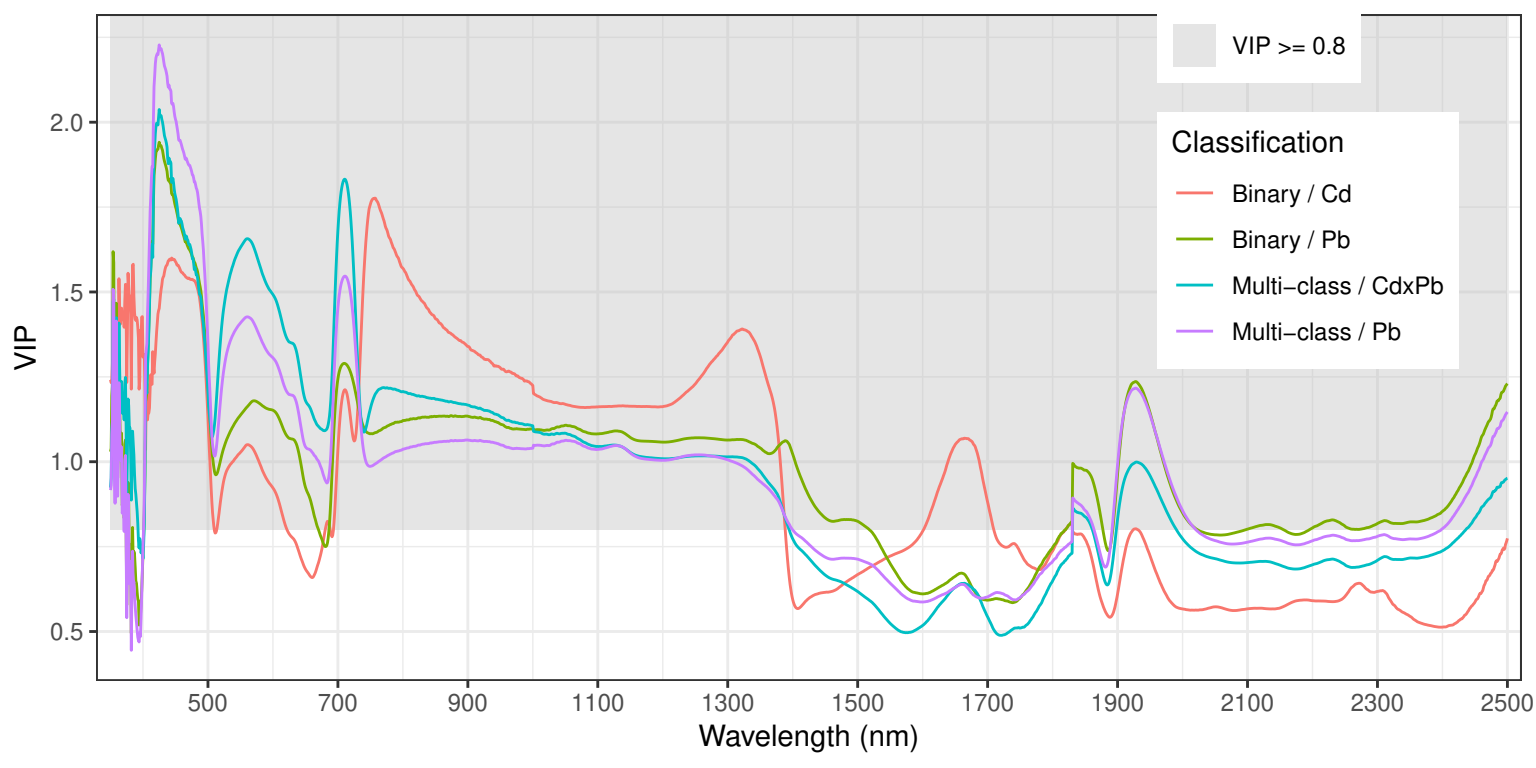
Class	ProducerAccuracy	UserAccuracy
0	66.94%	74.31%
1	72.73%	20.51%
2	60.12%	85.25%
3	96.43%	42.86%



d

Class	ProducerAccuracy	UserAccuracy
0	79.55%	62.87%
1	30.00%	79.41%
2	58.82%	16.39%
3	100.00%	10.81%




VIP >= 0.8
Classification
Binary / Cd
Binary / Pb
Multi−class / CdxPb
Multi−class / Pb
VIP
2.0
1.5
1.0
0.5
500
700
900
1100
1300
1500
1700
1900
2100
2300
2500
Wavelength (nm)

Supplementary Material

Title:

Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils.

Authors:

Kang Yu, Maarten Van Geel, Tobias Ceulemans, Willem Geerts, Miguel Marcos Ramos, Nadine Sousa, Paula M.L. Castro, Pierre Kastendeuch, Georges Najjar, Thierry Ameglio, Jérôme Ngao, Marc Saudreau, Olivier Honnay, Ben Somers.

Table S1. Classification of samples into four contamination levels of heavy metals according to the two major contaminates Cd and Pb in this study.

		Cd		
Metal	Class	0	1	Total
Pb	0	CdxPb 0 (n=121)	CdxPb 1 (n=11)	132
	1	CdxPb 2 (n=173)	CdxPb 3 (n=28)	201
Total		294	39	333

Table S2. Results of mixed models for testing the effect of soil heavy metals on leaf functional traits, including the leaf mass per area (LMA), Fv/Fm, total chlorophyll content (Chl) and Chla:Chlb ratio. Modeled random effects were *City*, *Site* and *Trunk Diameter*. Chlorophyll data were only available for a subset of the samples, where only Cd and Pb reached the thresholds of contamination.

		Mixed Model		Tukey's Test Class 0 – 1	
Trait	Metal	*F*-value	*P*-value	Estimate	*P*-value
LMA	Cd	1.114	0.292	0.249	0.292
	Cr	6.676	**0.01**	-1.319	**0.01**
	Cu	0.158	0.691	0.284	0.691
	Ni	0.23	0.632	0.425	0.632
	Pb	67.079	**<0.001**	1.284	**<0.001**
	Zn	0.699	0.404	0.521	0.404
Fv/Fm	Cd	0.016	0.901	-0.001	0.901
	Cr	0.014	0.905	0.003	0.905
	Cu	0.012	0.911	0.003	0.911
	Ni	0.084	0.772	-0.011	0.772
	Pb	5.842	**0.016**	-0.016	**0.016**
	Zn	0.075	0.784	-0.007	0.784
Chl	Cd	1.859	0.183	17.944	0.183
	Pb	9.32	**0.005**	-13.188	**0.005**
Chla:Chlb	Cd	0.065	0.801	0.071	0.801
	Pb	12.942	**0.001**	0.314	**0.001**

Table S3. Mixed models for testing the effect of multi-level Cd×Pb and Pb contamination on leaf functional traits, including the leaf mass per area (LMA), Fv/Fm, total chlorophyll content (Chl) and Chla:Chlb ratio. The modeled random effects are *City*, *Site* and *Trunk Diameter*. Chlorophyll data were only available for a subset of the samples, where only Cd and Pb reached the threshold of contamination.

	Mixed model			Tukey's Test											
				Class 0 – 1		Class 0 - 2		Class 0 - 3		Class 1 - 2		Class 1 - 3		Class 2 – 3	
Trait	Metal	*F*-value	*P*-value	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*P*
LMA	Cd×Pb	23.741	**<0.001**	0.025	1.000	1.256	**<0.001**	1.722	**<0.001**	1.231	**0.006**	1.698	**0.001**	0.466	0.246
	Pb	26.290	**<0.001**	1.210	**<0.001**	1.831	**<0.001**	2.444	**<0.001**	0.621	0.104	1.234	0.104	0.613	0.731
Fv/fm	Cd×Pb	2.420	0.066	-0.021	0.661	-0.018	**0.044**	-0.014	0.713	0.003	0.999	0.007	0.987	0.004	0.987
	Pb	2.004	0.113	-0.017	0.070	-0.011	0.863	-0.011	0.980	0.006	0.971	0.006	0.996	0.000	1.000
Chl	Cd×Pb	7.103	**0.003**	NA	NA	-14.226	**0.005**	8.608	0.743	NA	NA	NA	NA	22.834	0.132
	Pb	7.659	**0.002**	-12.058	**0.018**	-36.911	**0.009**	NA	NA	-24.852	0.089	NA	NA	NA	NA
Chla:Chlb	Cd×Pb	6.265	**0.006**	NA	NA	0.315	**0.004**	0.297	0.463	NA	NA	NA	NA	-0.018	0.997
	Pb	6.679	**0.004**	0.302	**0.006**	0.483	0.124	NA	NA	0.181	0.729	NA	NA	NA	NA

NA, not applicable.

Table S4. PLS-DA model calibration accuracies for the binary classification of Cd and Pb contamination and for multi-class classification of Cd×Pb interactions and four-level Pb contamination.

Metal	Spectra*	Components	Producer Accuracy	User Accuracy	Accuracy	Kappa
Binary						
Cd	ASD	5	74.36%	45.31%	86.49%	0.49
	ASD.cr	4	82.05%	41.03%	84.08%	0.46
	ASD.d1	2	76.92%	42.86%	85.29%	0.47
	ASD.snv	2	76.92%	42.86%	85.29%	0.47
Pb	ASD	2	64.18%	90.85%	74.47%	0.50
	ASD.cr	5	65.17%	86.18%	72.67%	0.46
	ASD.d1	4	71.14%	89.94%	77.78%	0.56
	ASD.snv	5	64.68%	87.25%	72.97%	0.47
Multi-class						
CdxPb	ASD	4	NA%	NA%	43.24%	0.24
	ASD.cr	8	NA%	NA%	62.16%	0.44
	ASD.d1	5	NA%	NA%	66.07%	0.49
	ASD.snv	4	NA%	NA%	50.15%	0.27
Pb	ASD	3	NA%	NA%	51.95%	0.29
	ASD.cr	8	NA%	NA%	52.85%	0.31
	ASD.d1	8	NA%	NA%	63.96%	0.43
	ASD.snv	6	NA%	NA%	54.05%	0.29

*cr, continuum removal; d1, first derivatives; snv, standard normal variate.

Table S5. Validation of the PLS-DA models calibrated with the full wavebands and VIP ($\geqslant$ 0.8) bands for the binary classification of Cd and Pb contamination and for multi-class classification of CdxPb and Pb contamination. Model training used 2/3 of the observations (n=215), and the remaining 1/3 of the observations (n=118) were used for testing model accuracies.

Metal	Spectra*	Full spectrum					VIP bands				
		Components	Producer Accuracy	User Accuracy	Accuracy	Kappa	Components	Producer Accuracy	User Accuracy	Accuracy	Kappa
Binary											
Cd	ASD	8	13.33%	100.00%	88.98%	0.21	8	13.33%	100.00%	88.98%	0.21
	ASD.cr	3	66.67%	47.62%	86.44%	0.48	3	66.67%	38.46%	82.20%	0.39
	ASD.d1	6	26.67%	80.00%	89.83%	0.36	6	26.67%	80.00%	89.83%	0.36
	ASD.snv	8	13.33%	66.67%	88.14%	0.19	8	6.67%	33.33%	86.44%	0.07
Pb	ASD	8	74.55%	77.36%	77.97%	0.56	8	83.64%	79.31%	82.20%	0.64
	ASD.cr	10	74.55%	83.67%	81.36%	0.62	10	76.36%	85.71%	83.05%	0.66
	ASD.d1	6	80.00%	77.19%	79.66%	0.59	6	89.09%	77.78%	83.05%	0.66
	ASD.snv	10	70.91%	82.98%	79.66%	0.59	10	69.09%	79.17%	77.12%	0.54
Multi-class											
CdxPb	ASD	9	NA%	NA%	68.64%	0.44	9	NA%	NA%	70.34%	0.47
	ASD.cr	5	NA%	NA%	66.10%	0.41	4	NA%	NA%	66.10%	0.40
	ASD.d1	4	NA%	NA%	65.25%	0.45	4	NA%	NA%	66.10%	0.46
	ASD.snv	6	NA%	NA%	68.64%	0.48	6	NA%	NA%	68.64%	0.48
Pb	ASD	9	NA%	NA%	75.42%	0.52	9	NA%	NA%	75.42%	0.52
	ASD.cr	3	NA%	NA%	72.88%	0.49	3	NA%	NA%	74.58%	0.52
	ASD.d1	9	NA%	NA%	78.81%	0.59	9	NA%	NA%	79.66%	0.61
	ASD.snv	10	NA%	NA%	64.41%	0.37	10	NA%	NA%	62.71%	0.35

*cr, continuum removal; d1, first derivatives; snv, standard normal variate.

Table S6. Classification accuracies of PLS-DA model calibration using the original and preprocessed reflectance data with a fixed (n=5) and the optimized number of components determined by cross-validation. Producer and user accuracies are not valid for the cases of absent prediction of individual classes.

				Accuracy			
Classification	Metal	Spectra	Components	Producer	User	Total	Kappa
Binary	Cd	ASD	9	23.08%	69.23%	89.79%	0.31
Binary	Cd	ASD	5	10.26%	57.14%	88.59%	0.14
Binary	Cd	ASD.cr	5	12.82%	71.43%	89.19%	0.19
Binary	Cd	ASD.cr	4	82.05%	41.03%	84.08%	0.46
Binary	Cd	ASD.d1	2	76.92%	42.86%	85.29%	0.47
Binary	Cd	ASD.d1	5	38.46%	68.18%	90.69%	0.44
Binary	Cd	ASD.snv	2	76.92%	42.86%	85.29%	0.47
Binary	Cd	ASD.snv	5	12.82%	62.50%	88.89%	0.18
Binary	Pb	ASD	5	85.07%	84.24%	81.38%	0.61
Binary	Pb	ASD	7	87.06%	86.21%	83.78%	0.66
Binary	Pb	ASD.cr	4	85.57%	85.57%	82.58%	0.64
Binary	Pb	ASD.cr	5	65.17%	86.18%	72.67%	0.46
Binary	Pb	ASD.d1	5	91.54%	88.04%	87.39%	0.73
Binary	Pb	ASD.d1	4	71.14%	89.94%	77.78%	0.56
Binary	Pb	ASD.snv	9	82.59%	87.83%	82.58%	0.64
Binary	Pb	ASD.snv	5	64.68%	87.25%	72.97%	0.47
Multi-class	Cd×Pb	ASD	5	NA%	NA%	68.64%	0.44
Multi-class	Cd×Pb	ASD	4	NA%	NA%	43.24%	0.24
Multi-class	Cd×Pb	ASD.cr	8	NA%	NA%	62.16%	0.44
Multi-class	Cd×Pb	ASD.cr	5	NA%	NA%	71.77%	0.48
Multi-class	Cd×Pb	ASD.d1	9	NA%	NA%	80.78%	0.65
Multi-class	Cd×Pb	ASD.d1	5	NA%	NA%	66.07%	0.49
Multi-class	Cd×Pb	ASD.snv	4	NA%	NA%	50.15%	0.27
Multi-class	Cd×Pb	ASD.snv	5	NA%	NA%	66.37%	0.39
Multi-class	Pb	ASD	5	NA%	NA%	71.19%	0.46
Multi-class	Pb	ASD	6	NA%	NA%	77.48%	0.56
Multi-class	Pb	ASD.cr	8	NA%	NA%	52.85%	0.31
Multi-class	Pb	ASD.cr	5	NA%	NA%	75.38%	0.52
Multi-class	Pb	ASD.d1	8	NA%	NA%	63.96%	0.43
Multi-class	Pb	ASD.d1	5	NA%	NA%	79.28%	0.6
Multi-class	Pb	ASD.snv	6	NA%	NA%	54.05%	0.29
Multi-class	Pb	ASD.snv	5	NA%	NA%	71.77%	0.45

Table S7. Classification accuracies of PLS-DA model validation by dividing the entire dataset to train (n=215) and test (n=118) sub sets and applying onto the original and preprocessed reflectance data with a fixed (5) and the optimized number of components determined by cross-validation. Producer and user accuracies are not valid for the cases of absent prediction of individual classes.

				Accuracy			
Classification	Metal	Spectra	Components	Producer	User	Total	Kappa
Binary	Cd	ASD	8	13.33%	100.00%	88.98%	0.21
Binary	Cd	ASD	5	6.67%	100.00%	88.14%	0.11
Binary	Cd	ASD.cr	3	66.67%	47.62%	86.44%	0.48
Binary	Cd	ASD.cr	5	6.67%	100.00%	88.14%	0.11
Binary	Cd	ASD.d1	6	26.67%	80.00%	89.83%	0.36
Binary	Cd	ASD.d1	5	26.67%	100.00%	90.68%	0.39
Binary	Cd	ASD.snv	8	13.33%	66.67%	88.14%	0.19
Binary	Cd	ASD.snv	5	13.33%	100.00%	88.98%	0.21
Binary	Pb	ASD	8	74.55%	77.36%	77.97%	0.56
Binary	Pb	ASD	5	87.27%	75.00%	80.51%	0.61
Binary	Pb	ASD.cr	10	74.55%	83.67%	81.36%	0.62
Binary	Pb	ASD.cr	5	87.27%	72.73%	78.81%	0.58
Binary	Pb	ASD.d1	6	80.00%	77.19%	79.66%	0.59
Binary	Pb	ASD.d1	5	83.64%	80.70%	83.05%	0.66
Binary	Pb	ASD.snv	10	70.91%	82.98%	79.66%	0.59
Binary	Pb	ASD.snv	5	80.00%	72.13%	76.27%	0.53
Multi-class	Cd×Pb	ASD	9	NA%	NA%	68.64%	0.44
Multi-class	Cd×Pb	ASD	5	NA%	NA%	68.64%	0.44
Multi-class	Cd×Pb	ASD.cr	5	NA%	NA%	66.10%	0.41
Multi-class	Cd×Pb	ASD.cr	4	NA%	NA%	65.25%	0.39
Multi-class	Cd×Pb	ASD.d1	4	NA%	NA%	65.25%	0.45
Multi-class	Cd×Pb	ASD.d1	5	NA%	NA%	70.34%	0.47
Multi-class	Cd×Pb	ASD.snv	6	NA%	NA%	68.64%	0.48
Multi-class	Cd×Pb	ASD.snv	5	NA%	NA%	61.02%	0.32
Multi-class	Pb	ASD	9	NA%	NA%	75.42%	0.52
Multi-class	Pb	ASD	5	NA%	NA%	72.03%	0.48
Multi-class	Pb	ASD.cr	3	NA%	NA%	72.88%	0.49
Multi-class	Pb	ASD.cr	5	NA%	NA%	73.73%	0.51
Multi-class	Pb	ASD.d1	9	NA%	NA%	78.81%	0.59
Multi-class	Pb	ASD.d1	5	NA%	NA%	77.97%	0.58
Multi-class	Pb	ASD.snv	10	NA%	NA%	64.41%	0.37
Multi-class	Pb	ASD.snv	5	NA%	NA%	73.73%	0.5

Table S8. Classification accuracies of VIP-based PLS-DA model validation by dividing the entire dataset to train (n=215) and test (n=118) sub sets and applying onto the original and preprocessed reflectance data with a fixed (5) and the optimized number of components determined by cross-validation. Producer and user accuracies are not valid for the cases of absent prediction of individual classes.

				Accuracy			
Classification	Metal	Spectra	Components	Producer	User	Total	Kappa
Binary	Cd	ASD	5	6.67%	100.00%	88.14%	0.11
Binary	Cd	ASD	8	13.33%	100.00%	88.98%	0.21
Binary	Cd	ASD.cr	5	6.67%	100.00%	88.14%	0.11
Binary	Cd	ASD.cr	3	66.67%	38.46%	82.20%	0.39
Binary	Cd	ASD.d1	5	20.00%	100.00%	89.83%	0.3
Binary	Cd	ASD.d1	6	26.67%	80.00%	89.83%	0.36
Binary	Cd	ASD.snv	5	13.33%	100.00%	88.98%	0.21
Binary	Cd	ASD.snv	8	6.67%	33.33%	86.44%	0.07
Binary	Pb	ASD	5	83.64%	73.02%	77.97%	0.56
Binary	Pb	ASD	8	83.64%	79.31%	82.20%	0.64
Binary	Pb	ASD.cr	5	83.64%	74.19%	78.81%	0.58
Binary	Pb	ASD.cr	10	76.36%	85.71%	83.05%	0.66
Binary	Pb	ASD.d1	5	90.91%	80.65%	85.59%	0.71
Binary	Pb	ASD.d1	6	89.09%	77.78%	83.05%	0.66
Binary	Pb	ASD.snv	5	80.00%	70.97%	75.42%	0.51
Binary	Pb	ASD.snv	10	69.09%	79.17%	77.12%	0.54
Multi-class	Cd×Pb	ASD	5	NA%	NA%	67.80%	0.43
Multi-class	Cd×Pb	ASD	9	NA%	NA%	70.34%	0.47
Multi-class	Cd×Pb	ASD.cr	5	NA%	NA%	66.10%	0.41
Multi-class	Cd×Pb	ASD.cr	4	NA%	NA%	66.10%	0.4
Multi-class	Cd×Pb	ASD.d1	5	NA%	NA%	68.64%	0.44
Multi-class	Cd×Pb	ASD.d1	4	NA%	NA%	66.10%	0.46
Multi-class	Cd×Pb	ASD.snv	5	NA%	NA%	62.71%	0.36
Multi-class	Cd×Pb	ASD.snv	6	NA%	NA%	68.64%	0.48
Multi-class	Pb	ASD	5	NA%	NA%	76.27%	0.55
Multi-class	Pb	ASD	9	NA%	NA%	75.42%	0.52
Multi-class	Pb	ASD.cr	5	NA%	NA%	72.88%	0.49
Multi-class	Pb	ASD.cr	3	NA%	NA%	74.58%	0.52
Multi-class	Pb	ASD.d1	5	NA%	NA%	77.97%	0.58
Multi-class	Pb	ASD.d1	9	NA%	NA%	79.66%	0.61
Multi-class	Pb	ASD.snv	5	NA%	NA%	71.19%	0.46
Multi-class	Pb	ASD.snv	10	NA%	NA%	62.71%	0.35

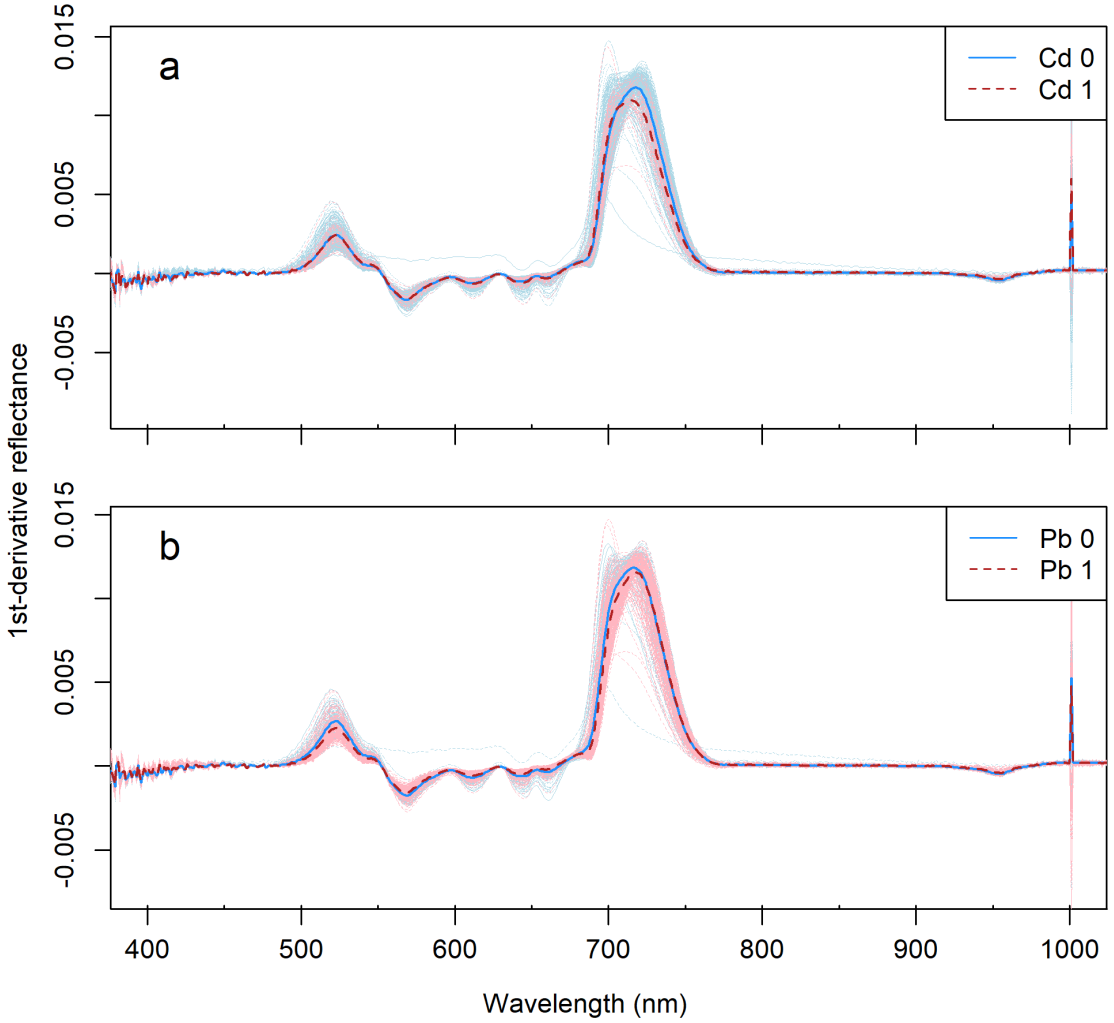


Fig. S1. First derivative reflectance of the leaves of the non-contaminated as compared with the contaminated soils by (a) Cd and (b) Pb.

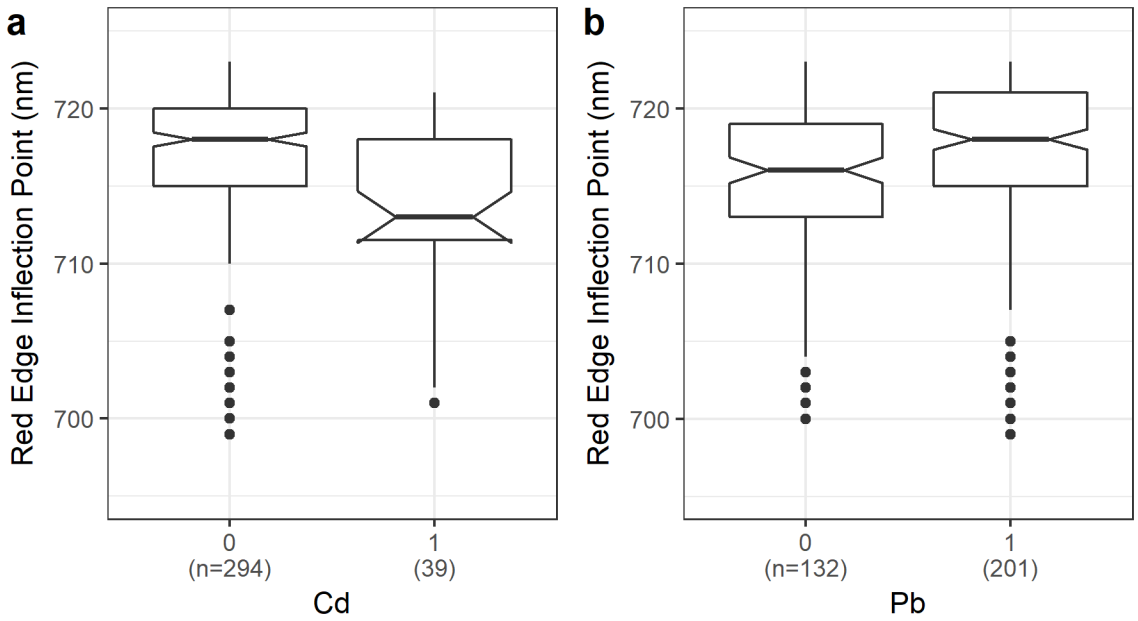


Fig. S2. Red-edge inflection points (REIP) extracted from the first derivative reflectance show the differences between the contaminated and non-contaminated classes of (a) Cd and (b) Pb classifications.

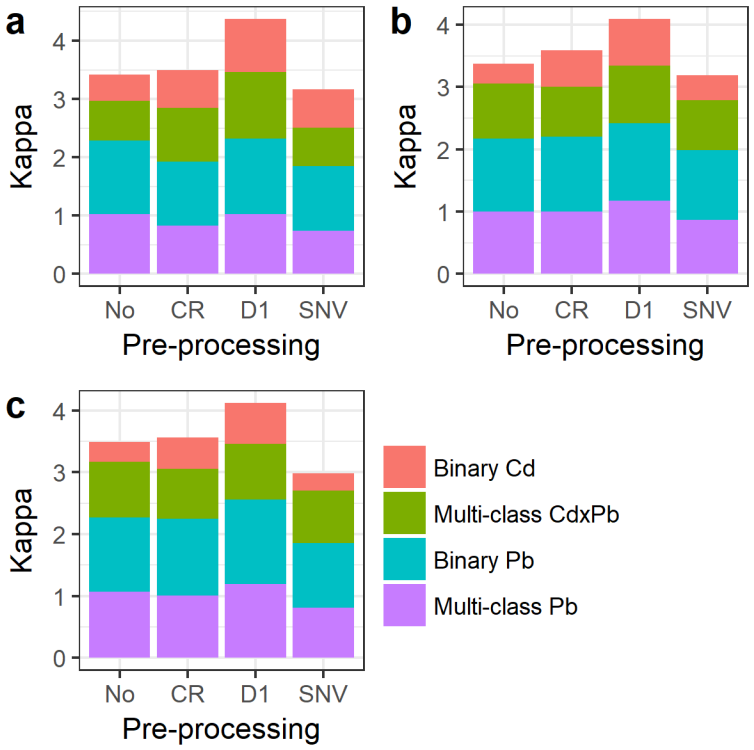


Fig. S3. Kappa coefficients of the PLS-DA classification for model (a) calibration, (b) validation and (c) VIP-based validation, when applied to the original reflectance data (no preprocessing) and three preprocessing types (D1 = first derivatives, SNV = standard normal variate, CR = continuum removal).