Using Community Science to Better Understand Lead Exposure Risks

Matthew J. Dietrich[1,1], John Shukle[1,1], Mark P.S. Krekeler[2,2], Gabriel Filippelli[1,1], and Leah R. Wood[1]

[1]Indiana University - Purdue University Indianapolis
[2]Miami University-Hamilton

November 30, 2022

Abstract

Lead (Pb) is a neurotoxicant that particularly harms young children. Urban environments are often plagued with elevated Pb in soils and dusts, posing a health exposure risk from inhalation and ingestion of these contaminated media. Thus, a better understanding of where to prioritize risk screening and intervention is paramount from a public health perspective. We have synthesized a large national dataset of Pb concentrations in household dusts from across the United States (U.S.), part of a community science initiative called "DustSafe." Using these results, we have developed a straightforward logistic regression model that correctly predicts whether Pb is elevated ($>$ 80 ppm) or low ($<$ 80 ppm) in household dusts 75% of the time. Additionally, our model estimated 18% false negatives for elevated Pb, displaying that there was a low probability of elevated Pb in homes being misclassified. Our model uses only variables of approximate housing age and whether there is peeling paint in the interior of the home, illustrating how a simple and successful Pb predictive model can be generated if researchers ask the right screening questions. Scanning electron microscopy supports a common presence of Pb paint in several dust samples with elevated bulk Pb concentrations, which explains the predictive power of housing age and peeling paint in the model. This model was also implemented into an interactive mobile app that aims to increase community-wide participation with Pb household screening. The app will hopefully provide greater awareness of Pb risks and a highly efficient way to begin mitigation.

Using Community Science to Better Understand Lead Exposure Risks

Matthew Dietrich[1*], John Shukle[1], Mark P.S. Krekeler,[2,3] Leah Wood,[1] Gabriel Filippelli[1,4]

[1]Department of Earth Sciences, Indiana University – Purdue University Indianapolis, Indianapolis, IN, USA

[2]Department of Geology & Environmental Earth Science, Miami University, Oxford, OH 45056, USA

[3]Department of Mathematical and Physical Sciences Miami University-Hamilton, Hamilton, OH 45011, USA

[4]Environmental Resilience Institute, Indiana University, Bloomington, IN, USA

[*]Corresponding Author Email: mjdietri@iu.edu

Abstract

Lead (Pb) is a neurotoxicant that particularly harms young children. Urban environments are often plagued with elevated Pb in soils and dusts, posing a health exposure risk from inhalation and ingestion of these contaminated media. Thus, a better understanding of where to prioritize risk screening and intervention is paramount from a public health perspective. We have synthesized a large national dataset of Pb concentrations in household dusts from across the United States (U.S.), part of a community science initiative called “DustSafe.” Using these results, we have developed a straightforward logistic regression model that correctly predicts whether Pb is elevated (> 80 ppm) or low (< 80 ppm) in household dusts 75% of the time. Additionally, our model estimated 18% false negatives for elevated Pb, displaying that there was a low probability of elevated Pb in homes being misclassified. Our model uses only variables of approximate housing age and whether there is peeling paint in the interior of the home, illustrating how a simple and successful Pb predictive model can be generated if researchers ask the right screening questions. Scanning electron microscopy supports a common presence of Pb paint in several dust samples with elevated bulk Pb concentrations, which explains the predictive power of housing age and peeling paint in the model. This model was also implemented into an interactive mobile app that aims to increase community-wide participation with Pb household screening. The app will hopefully provide greater awareness of Pb risks and a highly efficient way to begin mitigation.

Plain Language Summary

Community science has been gaining traction in many locales throughout the United States, particularly in the field of urban pollution. While this has helped with science education and informing communities of potential hazards and mitigation tools, little has been done to effectively assimilate this information in a useful way to help people in other communities throughout the country. Thus, we utilized a large dataset of household dust samples provided by community scientists across the United States to build a simple predictive model that lets users know if their dust is likely to be high in a toxic metal, lead. Additionally, we built this model into an interactive mobile app that we plan to use as a recruitment tool for usage of lead screening kits. Ultimately, we plan to assess whether this mobile app improves user knowledge of household lead risks and increases participation from start to finish for free lead screening services.

Key Points

- Community science sampling can provide national-level insight
- Mobile apps can be utilized as a lead intervention tool
- Elevated lead in house dust can be reasonably predicted with a simple statistical model and two variables

Key Words

Lead (Pb), Community Science, Predictive Modeling, Pollution Intervention, Pollution Remediation, Scanning Electron Microscopy (SEM)

1. Introduction

Lead (Pb) is a naturally occurring heavy metal neurotoxicant that causes many deleterious effects in humans, even in small quantities (e.g., Assi et al., 2016; Dórea, 2019). It is a biologically non-essential element that is especially detrimental to young children (e.g., Koller et al., 2004). In the United States (U.S.), it has largely been phased out of products, most notably leaded gasoline and paint, but remains in many urban environments as a form of legacy pollution (e.g., Laidlaw et al., 2012). Thus, modern sources of Pb are primarily lead paint in older homes and soil/dusts that contain remnants of both leaded paint and gasoline. Ingestion and inhalation of paint, soil, and dust containing elevated levels of Pb still pose a health risk, particularly for children due to their increased hand-to-mouth behavior (e.g., Ko et al., 2007; Needleman, 2004; Stewart et al., 2014).

Household dust Pb concentrations and loadings have been shown to be strongly related to children's blood Pb levels (BLLs) (e.g., Lanphear et al., 1996; Gulson and Taylor, 2017; Rhoads et al., 1999). Thus, a better understanding of risk factors associated with Pb in household dusts can help predict what homes may have elevated Pb concentrations in dusts, and thus help mitigate Pb exposure and elevated BLLs in children. Predictive modeling of Pb in soil samples with variables such as race and house age has already been shown to be effective in predicting at-risk areas (Obeng-Gyasi et al., 2021), but this has not been attempted with household indoor dust Pb concentrations across a wide geographic area through community-provided samples.

Citizen/community science sampling of environmental media such as soil has been shown to not only aid as an educational tool to those collecting the samples, but also provides important scientific data of inorganic contaminants such as Pb and how they are distributed throughout the environment (e.g., Filippelli et al., 2018; Masri et al., 2021; Ringwald et al., 2021; Taylor et al., 2021). Community science offers a gateway to increased sampling resolution and sampling size, which often cannot be achieved by researchers alone. Thus, we have utilized an ongoing community science project, "DustSafe" (https://www.360dustanalysis.com/), to analyze approximately 434 household dust samples from across the United States (Fig. 1) to determine whether homes at risk for elevated dust Pb can be accurately predicted. While individual variables such as housing age and automobile traffic near homes have been shown to be correlated with indoor dust Pb concentrations (e.g., Meyer et al., 1999; Rasmussen et al., 2011), variables have not been collectively applied in a predictive model across multiple states and cities in the U.S. Additionally, we sought to utilize this predictive model as part of an interactive mobile app to encourage greater community engagement for household Pb screening, which can not only help individuals gain agency in possible Pb mitigation measures, but can also help policymakers and the community at large better understand where/how to focus household Pb intervention efforts. As community science apps have begun to gain traction in fields such as biology and ecology (e.g., https://www.inaturalist.org/ and https://ebird.org/home), and have even helped both community scientists and researchers combat disease vectors such as mosquitoes (Low et al., 2021), we wanted to explore the potential applicability in the realm of household-level environmental pollution.

2. Methods

2.1 *"DustSafe" sampling*

Details of the household dust sampling are provided in Isley et al. (accepted). Briefly, DustSafe was advertised as a program to thousands of households through social media, e-mail, etc. to gain community science participants. Project protocols were approved following ethical review at Indiana University, USA (project #1810831960). Participants completed an online survey (Isley et al., accepted—their SI Text 2) and collected vacuum cleaner dust in a polyethylene bag. Samples were collected from 2019 to present. Once samples were collected by researchers, they were sieved to 250 µm and analyzed for Pb, As, Cd, Cr, Cu, and Zn using X-ray fluorescence spectrometry (XRF). They were dry by virtue of the vacuum sampling and needed no desiccation. NIST 2702 was run periodically as an external standard on the XRF between dust samples, and the arithmetic mean (average) % error for Pb was 14.7% ± 8.6% (n = 9).

Results were reported back to participants following data collection (Example for Pb in Fig. S1), and then plotted on the "Map My Environment" website (www.mapmyenvironment.com) with locations randomly double jittered to protect privacy. This means that the icon for the data point does not appear at the actual sampling location, but rather, it is moved twice randomly within a radius of ~2 city blocks from the actual location: once when the data is first uploaded, and then again each time the map is loaded or refreshed.

2.2 *Data filtering/building of logistic regression model*

The initial dataset (link to data provided in Text S1) of potentially relevant data for this analysis contained 434 samples with matching Pb data (greater than detection limit) from the United States (and three samples from Canada). The most important potential predictive variables of housing age, interior peeling, exterior peeling, and recent renovation were determined by looking for statistically significant differences between questionnaire responses (survey link/details in Isley et al. accepted—their SI Text 2), both through t-tests for binary response variables (Yes/No) and ANOVA tests for multiple categories, specifically for housing age categories (described below). Additionally, we screened for variables based on our global dust data (Isley et al., accepted—their Table 1), looking for variables that may be significant (lower p-values) despite the data being from the global sample set. The data was ultimately filtered down to 342 samples that contained Pb concentrations and questionnaire responses for housing age, interior peeling, exterior peeling, and recent renovation. Because exact housing age is difficult to deduce for many respondents, particularly renters and those who may be surveyed in-person at future community Pb screening events, we classified housing age into categories of Pre-1940, 1940-1959, 1960-1979, 1980-Present, and "Not Sure", so this predictor variable may be more useful/applicable in future surveys.

A logistic regression model was applied using independent potential predictor variables to predict whether an indoor housing dust sample was either ≥80 ppm Pb or <80 ppm Pb. This was used as a conservative cut-off based on California's safe screening level for soils, because we did not collect indoor dust loading data and most other standards used in the U.S. for soil Pb are outdated and likely too high (e.g., the U.S. EPA's 400 ppm residential soil standard; Gailey et al., 2020). Our model was run in RStudio (R Core Team, 2021) using the "glm" function based on the general equation:

$$\log\left[\frac{p}{1-p}\right] = b_0 + b_1 * x_1 + b_2 * x_2 \ldots + b_n * x_n \tag{1}$$

Where p is the probability of an event occurring, b_0 is the intercept, b_n is the regression beta coefficient, and x_n is a given predictor variable.

Each potential independent predictor variable (besides housing age) categorical response of "No," "Yes," and "Not Sure" were reclassified as numeric variables of 0, 1, and 2, respectively, for the model. Housing age categories were reclassified as numeric variables of 0, 1, 2, 3, and 4 for the responses, "1980-Present," "1960-1979," "1940-1959," "Pre-1940," and "Not Sure," respectively.

Our most successful model contained the independent variables of housing age ($p = 0.0002$) and interior peeling paint ($p = 0.008$), which generated the following equation:

$$log\left[\frac{p}{1-p}\right] = 2.1413 - 0.4506\ (Housing) - 1.1535\ (Interior\ Paint\ Peeling) \tag{2}$$

This was based on a random training set of 240 samples from our original 342 samples. We evaluated the model on a random testing dataset of 102 samples from our original 342 samples. All input and output files are freely available on GitHub (link provided in Text S1), as well as the logistic regression model R code.

2.3 Mobile app development

An interactive online web application was developed to implement our predictive model in a simple and straightforward manner (link provided in Text S1). The application was built using the shiny, shinydashboard, shinydashboardPlus, and shinyjs packages in R (Attali, 2020; Chang et al., 2018; Chang and Borges Ribeiro, 2018; Granjon, 2021). Along with providing users with a straightforward interface for answering questions about house age and peeling paint and a custom risk assessment based on the embedded logistic predictive model, the application also provides users with direct links to our mapmyenvironment.com web portal, where they can register for free dust and soil Pb screening. Finally, the application offers background information about the current model version used to make the predictions, and offers direct links to model, data, and application code repositories.

2.4 Scanning electron microscopy (SEM)

A subset of DustSafe household dust samples were prepared on aluminum samples stubs using carbon sticky tab substrates for analysis using a scanning electron microscope (SEM) and energy dispersive X-ray spectroscopy (EDS). EDS lines used to identify Pb specifically include the $L_\alpha = 10.541$ keV (nominally $M_\alpha = 2.342$ keV, $M_\beta = 2.444$ keV). All analyses were conducted at Indiana University-Purdue University Indianapolis with a Zeiss EVO-10 SEM and Bruker XFlash6, 60 mm^2 EDS detector. Backscatter electron (BSE) images were collected at a setting of 15 kV in variable pressure mode. Qualitative elemental composition data (EDS data) were collected at the same conditions.

3. Results & Discussion

3.1 Significant findings between Pb in dust and housing age, vacuum frequency, and peeling paint

Household dust Pb concentrations were significantly higher in homes where there was interior or exterior paint peeling (Fig. 2; Table 1), which is in line with recent global household Pb dust data from the same DustSafe project (Isley et al., accepted). This suggests that leaded paint is still a significant contributor of Pb to dust in many homes. However, it does not exclude outside sources such as soil/street dust that may include Pb from leaded gasoline. For example, indoor dusts have been shown to contain significant Pb sources from outdoor sources such as soils, dust, and industrial pollution as well (e.g., Adgate et al., 1998; Kelepertzis et al., 2020).

Greater housing age has long been known to be associated with increased Pb concentrations in household dusts, such as in Canada and the U.S. (e.g., Rasmussen et al., 2011; Rasmussen et al., 2013; Spalinger et al., 2007). Our results support this, as a moderate positive correlation was seen between housing age and Pb concentration in our samples (Fig. 3A), with more recent housing age categories generally lower in dust Pb as well (Fig. 3B; Table 1). This is most likely due to older homes containing Pb-based paints that can contribute to dust samples, as Pb housing paint was outlawed in the U.S. in 1978 and housing built before 1940 is the most likely to contain Pb paint (e.g., Levin et al., 2021). Furthermore, our global DustSafe dataset also observed a strong increase in Pb house dust concentration with home age (Isley et al., accepted), suggesting that this is a common trend in many countries.

Regular cleaning of homes and the surrounding environment, including measures such as vacuuming, have been shown to effectively lower BLLs in children (e.g., Laidlaw et al., 2017; Rhoads et al., 1999). We also found that those vacuuming more frequently than once a month contained significantly lower concentrations of Pb in their house dust compared to those vacuuming monthly or less (Fig. S2A). However, we did not see any significant differences in Pb house dust concentrations in subcategories where people performed more than monthly vacuuming (Fig. S2B), which corresponds to our general trends in global dust data where increased vacuuming frequency was not associated with Pb dust concentration at all (Isley et al., accepted). Our findings suggest that households that hardly vacuum may be more likely to accumulate Pb-rich larger particles when they do finally vacuum and gather samples, such as Pb-paint chips, which would skew the bulk chemistry Pb concentration to higher values (since we didn't measure loading rates—or the rate of dust deposition). Households that more frequently vacuum may be less likely to sample larger, Pb-rich particles for their DustSafe sample submission.

3.2 Predictive accuracy of logistic regression model

Application of our logistic regression model on a "test" dataset of 102 samples from our original dataset reveals an overall prediction accuracy of 75% when using a probability threshold of 0.8 to determine "high" or "low" Pb. Importantly, only 4 samples out of 102 test samples (4%) were classified as "low" Pb when they were actually a "high" Pb sample, shown in our "confusion matrix" output of sample classifications (Table 2). This implies that from an intervention standpoint our model contains few false negatives, and thus has excellent sensitivity (82%).

3.3 Usefulness and "App"lication of model for household Pb screening

While more sophisticated models can be effective in predicting high risk exposure areas for Pb in soils or dusts (e.g., Obeng-Gyasi et al., 2021), we believe that from a public health intervention standpoint, sometimes a simpler model is better. Because only two independent variables with categorical responses were proven statistically significant in our model and yielded an effective prediction accuracy of 75%, we decided to incorporate our model into a mobile-based app to aid in household Pb screening recruitment efforts (Fig. 4). The goal is to help people understand whether there is an increased chance of elevated Pb in their home based on our model, then give them an opportunity to freely test their home so that they can gain agency in decision-making regarding Pb mitigation. Additionally, we sought to include decision variables of "Not sure" in our app/model for peeling interior paint and the age of the home, because this helps with realistic in-person usage of the app at community events, and many people taking the survey may be renters and unsure of home age. Furthermore, renters are often one of the more likely subgroups of people to contain elevated household Pb in soil or dust (e.g., Masri et al., 2020; Masri et al., 2021) often because of older housing units and less priority from landlords for remediation. Within our model, approximately 28 individuals or 8% were uncertain of their exact home age (Fig. S3). Moving forward, it would be useful to include home ownership in our DustSafe surveys, to understand whether this is correlated to uncertainty in home age and the predictive power this has for elevated dust Pb.

Because our mobile app screening questions are simple, straightforward, and contain only categorical multiple-choice responses, we envision that its usage will be highly effective as a quick screening tool that many in-person events (i.e., community events, schools) can implement to help people know if Pb exposure is a hazard they should be concerned about. Furthermore, because our dataset is based on national-scale data, the mobile app can be utilized in many different locations, further aiding in its "app"licability and versatility as a Pb screening recruitment tool.

3.4 Evidence of Pb paint in dust samples

Through SEM work on several household dust samples that contained elevated bulk Pb concentrations, we were able to identify numerous examples of particles consistent in composition and morphology to Pb paint, ranging from ~10 µm in diameter to >100 µm in diameter (Fig. 5). Our Pb paint chips were similar in composition and morphology to Pb paint analyzed by SEM in Hunt (2016), including several Pb-carbonate paints and the presence of Zn in the paint (Fig. 5). Additionally, the Mg-Al-Si EDS peaks in several paint samples (i.e., Figs. S6, S7, S8) are consistent with montmorillonite, an additive commonly used in Pb-based paint as organo-clays to aid in the suspension of the pigments. This helps explain why the predictor variables of housing age and interior peeling paint were so significant—many household dust samples with elevated concentrations of Pb likely have the Pb predominantly sourced from house paint. However, this does not mean that Pb in house dusts is exclusively from house paint, or that other metals are from exclusively indoor sources. As mentioned earlier, outdoor sources of pollutants can enter homes, such as through dust brought indoors (e.g., Adgate et al., 1998; Kelepertzis et al., 2020), via vectors such as pets, clothing, or shoes. For example, we found clear examples of technogenic Fe-oxide spherules, likely a byproduct of anthropogenic combustion, in house dust samples (Fig. S4). These particles likely came from an outdoor source, such as vehicle exhaust or industrial combustion, as they are similar to Fe-rich spherical particles commonly found in industrial areas from high temperature formation processes (e.g., Dietrich et al., 2019; Gaberšek and Gosar, 2021; Miler and Gosar, 2013; Teran et al., 2020). Furthermore, we found one sample that contains EDS spectra consistent with $PbCrO_4$, or Pb-chromate paint (Fig. 5A), which could have come from yellow-paint inside the home, but may have also been brought in from outdoors where Pb-chromate is often used in traffic paint (e.g., O'Shea et al., 2021).

3.5 Future goals and directions

We based our initial model on predominantly U.S. house dust samples, because of statistically significant differences in bulk metal composition of dusts between other countries (Isley et al., accepted) and there are likely other confounding factors between countries that affect Pb in dusts (i.e., different regulation of Pb paints and Pb gasoline). However, as more data is collected and as we gain a better understanding of what variables predominantly influence Pb in house dust, our model can be applied to additional countries and refined within the U.S. to more accurately differentiate what homes likely contain elevated Pb. A specific area for refinement of the model may lie in spatial data, such as relating zip codes of samples with socioeconomic (i.e., % poverty, racial distribution) and public health data (i.e., blood lead levels) within those zip codes, which may add to the predictive power of our model.

Additionally, this type of simple predictive model usage in a mobile app as an intervention tool can be applied beyond Pb in household dusts, such as to other contaminants of concern in homes like arsenic (As) or radon (Rn). Lastly, community science sampling endeavors should continue to grow, as they are not only a great opportunity for direct household contamination intervention, but also contribute to a greater general understanding of important issues such as Pb pollution and what areas community remediation should be focused in. Scientific information **from the public** is one of the most beneficial ways to **help the public** with pollution remediation and awareness. We have illustrated this with our accessible Pb dust logistic regression model and mobile app, and other recent large-scale community science endeavors have also increased metal pollution mapping and awareness (e.g., Taylor et al., 2021).

We plan to conduct a follow-up study on the effectiveness of this type of simple intervention in engaging participants to have full-cycle involvement, going from initial usage of the mobile app to submittal of samples, to finally opening sample results once generated. Ample examples of citizen science exist with various ways that the engagement does, or does not, provide real, tangible benefits to participants (e.g., Hayhow et al., accepted), but they are typically poorly assessed. One recent example of community science in Australia focused on analyzing garden soil for heavy metals found that 96% of respondents ($n = 361$) would recommend the program to someone else, and 94% said their understanding of heavy metal contaminants in gardens had increased (Taylor et al., 2021). Follow-up surveys from our global DustSafe program found that 39% of participants ($n = 246$) took some remedial action at home, and 94% of participants said the information provided to them was useful (Isley et al, accepted). However, these detailed, large-scale follow-up surveys are often sparse. We hypothesize that this simple app engagement will generate greater “engage to completion” metrics because of simplicity of message. We will therefore develop a follow-up survey once the sample results are generated and returned to users to determine what, if any, impacts the mobile app and corresponding results had on participants’ behavior, including any mitigation steps that they took in response to results.

4. Conclusions

A simple logistic regression model based on real-world samples proved to be effective at identifying homes at risk for higher Pb in household dusts across the United States. Application of the model on a test dataset of 102 samples revealed a 75% classification accuracy of either “high” or “low” Pb in household dust, with the cutoff based on 80 ppm Pb. This illustrates how community science gathered data can provide valuable insight into primary predictor variables for elevated Pb. Additionally, we showed how simplistic, yet effective Pb predictive models can be incorporated into interactive mobile apps such as a Pb screening recruitment tool. Collectively, we hope that modeling efforts such as these and engagement with local communities will aid in Pb exposure prevention and remediation, so that no child grows up with an unnecessarily high risk of Pb exposure.

Acknowledgements

The authors are deeply thankful to the households that supplied dust samples for this work and the lab techs who helped process samples. The MapMyEnvironment program and the related DustSafe sampling effort are partially supported by National Science Foundation Grant ICER-1701132 to G.M.F. and the Environmental Resilience Institute, funded by Indiana University's Prepared for Environmental Change Grand Challenge Initiative. Partial support for this work was provided by NSF-EAR-PF Award #2052589 to M.D. Special thanks to Miguel Cruz for help with SEM instrumentation, Angela Herrmann for input on mobile app development, and Michael O'Shea with input on SEM image interpretation. Lastly, we thank James Montgomery and an anonymous reviewer for their constructive comments that helped improve the manuscript.

Open Research

All data and source code used in this manuscript are freely available on GitHub (https://github.com/dietrimj/Community-Science-Pb-Prediction).

Tables:

Table 1: Summary statistics of household dust Pb concentrations (mg/kg) from significant predictor variables utilized in the logistic regression model. The actual questions for the variables from the questionnaire are provided in Text S2. For "Housing Age," we have included those who did not complete a survey in the "Not Sure" category.

		Mean	*Std Dev*	*Median*	*Max*	*Min*	*(n)*
Total Pb		99	239	32	2328	3	434
Exterior Paint Peeling	**Yes**	131	179	41	815	4	48
	No	80	195	29	1665	3	272
	Not Sure	40	46	28	205	5	23
Interior Paint Peeling	**Yes**	142	175	81	729	7	40
	No	77	188	29	1665	4	302
	Not Sure	35	N/A	35	35	35	1
Housing Age	**Pre-1940**	228	306	134	1665	7	54
	1940-1959	121	221	53	1304	10	33
	1960-1979	78	193	32	1377	6	52
	1980-Present	45	114	24	1205	3	178
	Not Sure	37	44	25	202	5	117

Table 2: Confusion matrix output of logistic regression model from test dataset (n = 102).

	Actual High Pb	**Actual Low Pb**
Predicted High Pb	18	21
Predicted Low Pb	4	59

Figures:

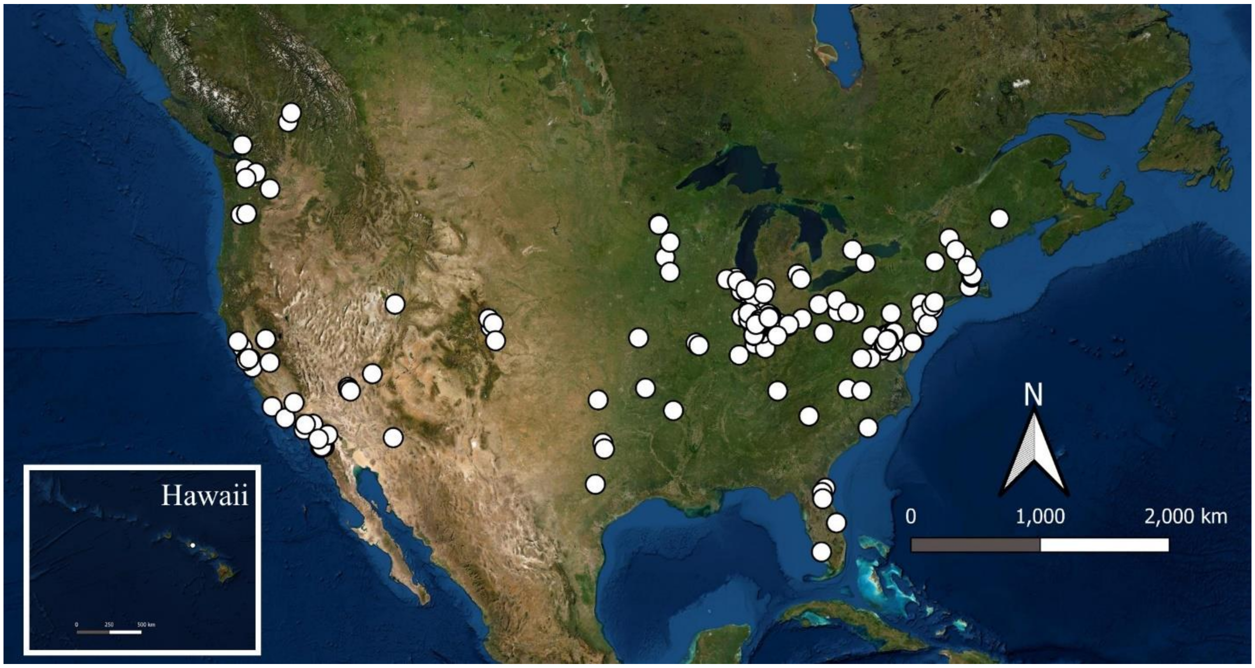


Figure 1 Samples with Pb (and other heavy metal) results reported back to households from the "DustSafe" project in the U.S. and Canada.

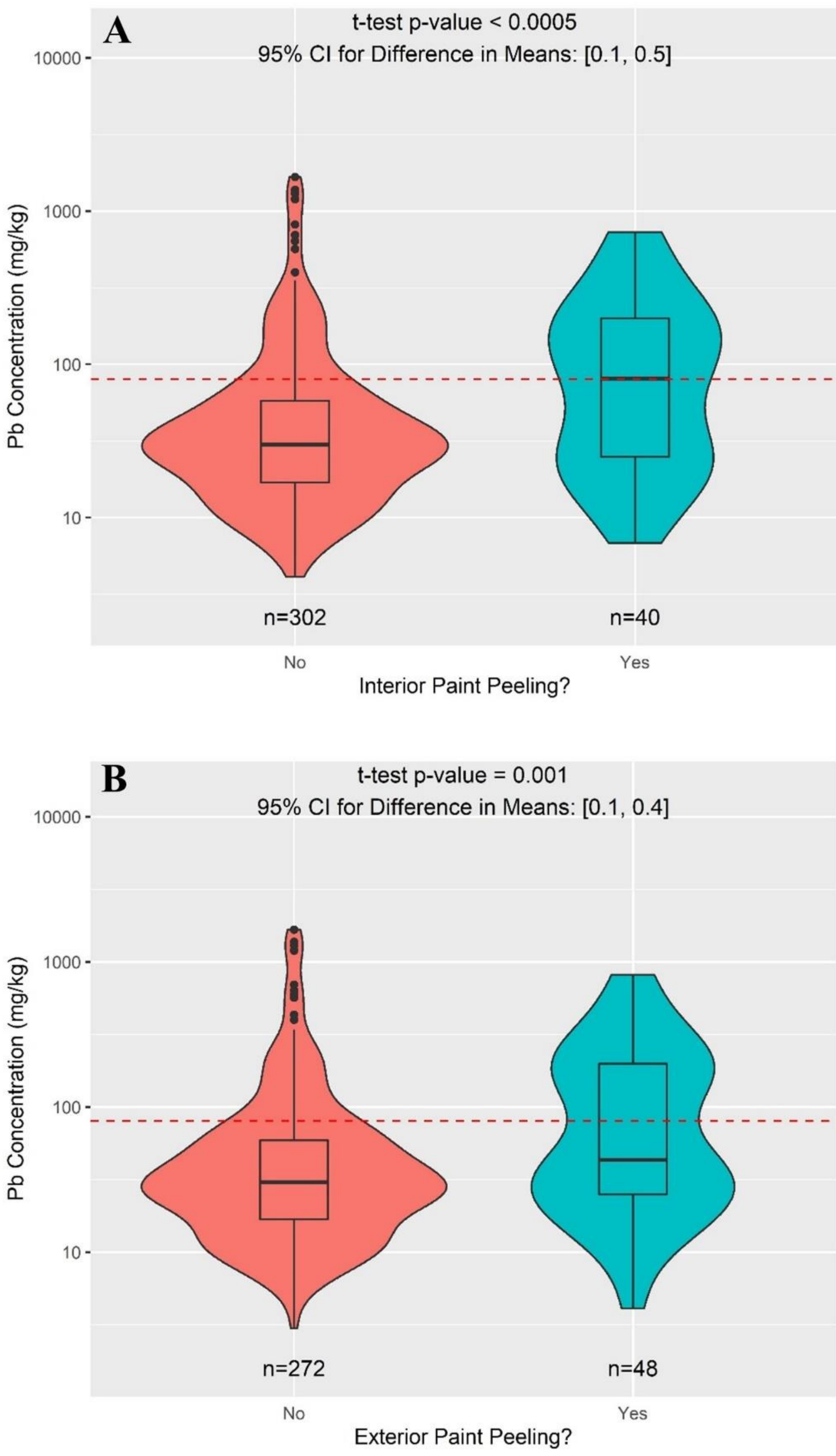


Figure 2 Embedded boxplots within violin plots for both interior (A) and exterior peeling paint (B) questionnaire responses. The boxes represent the interquartile range (IQR) of 25^{th}-75^{th} percentiles of data, the horizontal line is the median, and the whiskers represent 1.5 times the IQR. Two-sample paired t-test results between yes/no responses are also provided. The y-axes are transformed on a $\log_{10}$ scale, and the dashed red lines represent California's safe screening soil Pb level of 80 ppm.

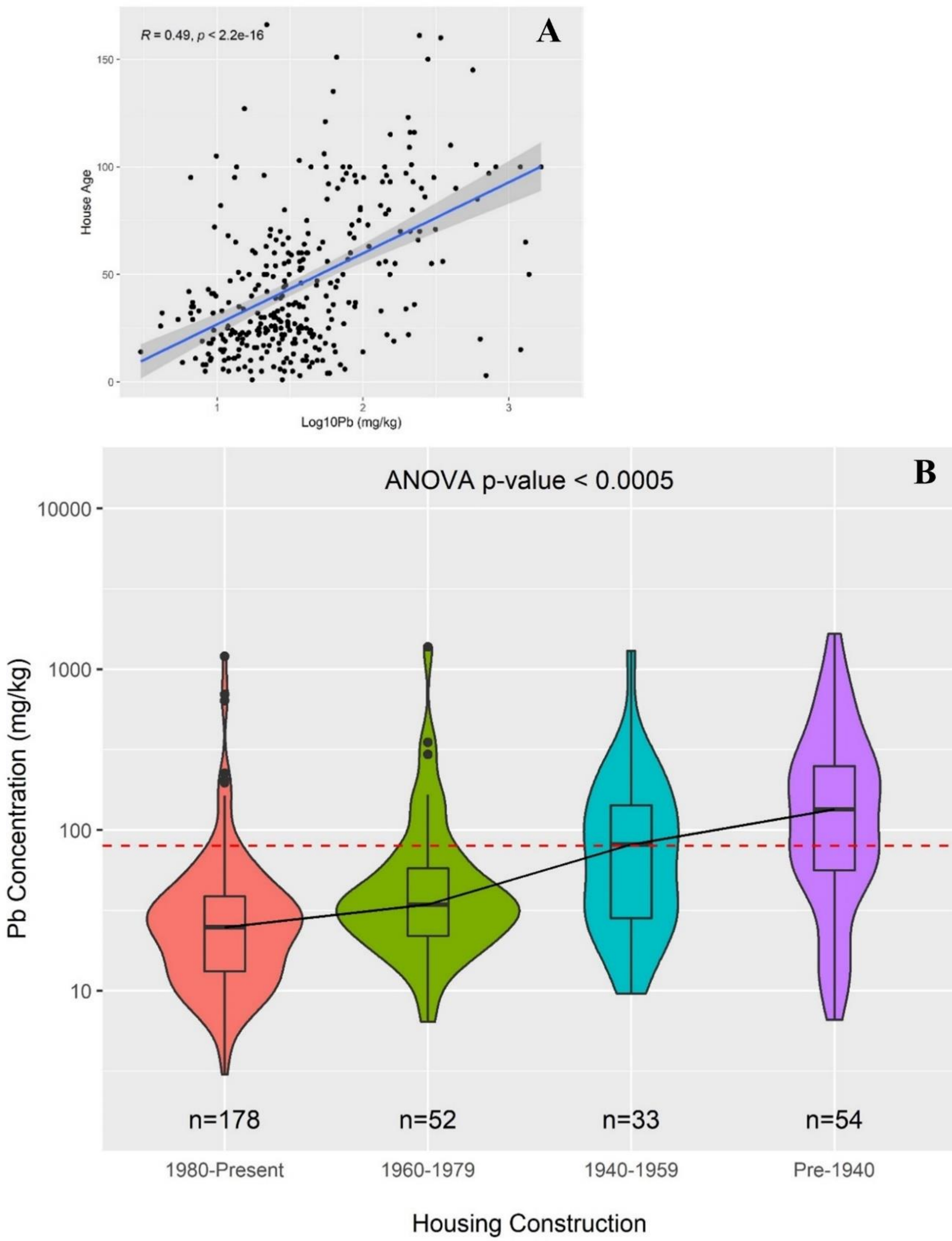


Figure 3 (A) Scatterplot between approximate housing ages and $\log_{10}$ Pb concentrations with the Pearson correlation coefficient and associated p-value provided, as well as a linear regression line in blue with the shaded 95% confidence interval. (B) Embedded boxplots within violin plots for housing age categories used in the predictive model. The boxes represent the interquartile range (IQR) of 25th-75th percentiles of data, the horizontal line is the median (which is connected between housing age categories with a black line), and the whiskers represent 1.5 times the IQR. An analysis of variance (ANOVA) test associated p-value between all housing age categories is provided. The y-axis is transformed on a $\log_{10}$ scale, and the dashed red line represents California's safe screening soil Pb level of 80 ppm.

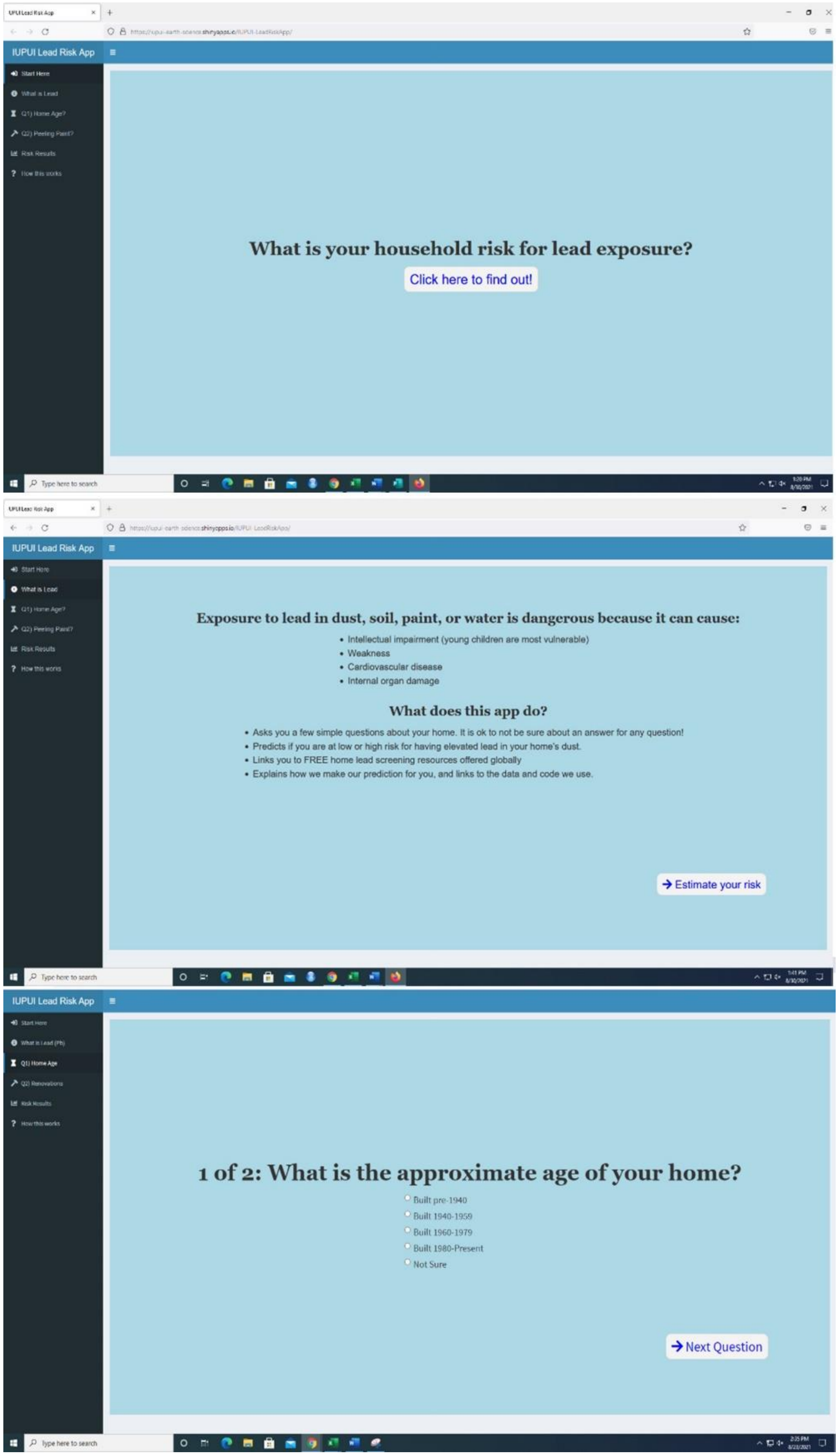


Figure 4 *Screenshots from the beginning of the interactive Pb household dust screening app (https://iupui-earth-science.shinyapps.io/IUPUI-LeadRiskApp/).*

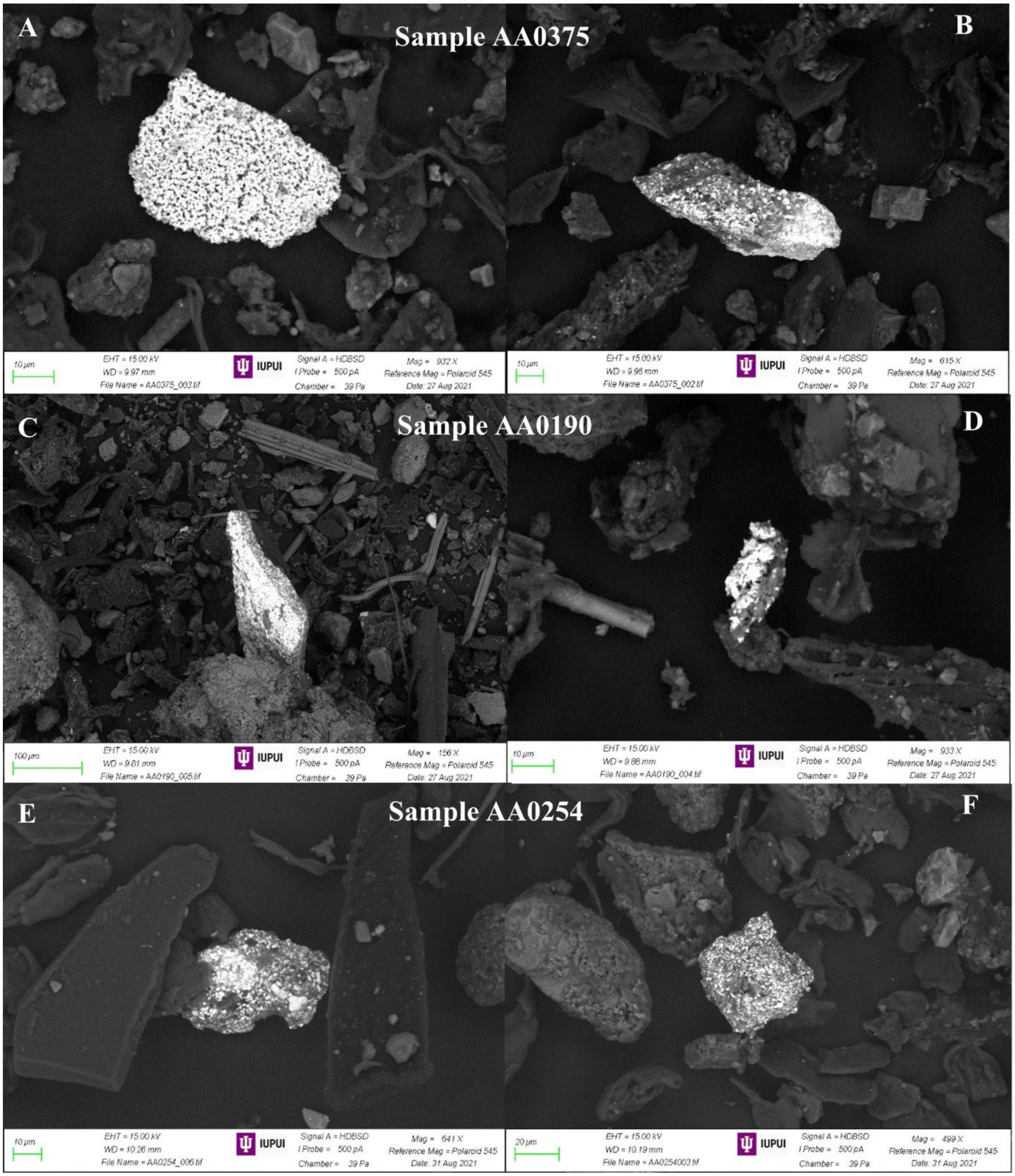


Figure 5 SEM images of particles resembling Pb paint, surrounded by other particulates in various high Pb DustSafe household dust samples (corresponding EDS spectra provided in Supplementary Materials; Figs. S5-S10). Pb paint particles are evident by very high contrast of electron backscatter detection—more so than surrounding particles because of the high atomic number of Pb. Most Pb-bearing particles are angular or jagged, with clear flaky particles on their surface.

References

Adgate, J. L., Rhoads, G. G., & Lioy, P. J. (1998). The use of isotope ratios to apportion sources of lead in Jersey City, NJ, house dust wipe samples. *Science of the Total Environment*, *221*(2-3), 171-180.

Assi, M. A., Hezmee, M. N. M., Abd Wahid Haron, M. Y. M., & Sabri, M. A. R. (2016). The detrimental effects of lead on human and animal health. *Veterinary World*, *9*(6), 660.

Attali,, D. (2020). shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 2.0.0. https://CRAN.R-project.org/package=shinyjs

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

Chang, W. & Borges Ribeiro, B. (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. https://CRAN.R-project.org/package=shinydashboard

Dietrich, M., Wolfe, A., Burke, M., & Krekeler, M. P. (2019). The first pollution investigation of road sediment in Gary, Indiana: Anthropogenic metals and possible health implications for a socioeconomically disadvantaged area. *Environment International*, *128*, 175-192.

Dórea, J. G. (2019). Environmental exposure to low-level lead (Pb) co-occurring with other neurotoxicants in early life and neurodevelopment of children. *Environmental Research*, *177*, 108641.

Filippelli, G. M., Adamic, J., Nichols, D., Shukle, J., & Frix, E. (2018). Mapping the urban lead exposome: a detailed analysis of soil metal concentrations at the household scale using citizen science. *International Journal of Environmental Research and Public Health*, *15*(7), 1531.

Gaberšek, M., & Gosar, M. (2021). Towards a holistic approach to the geochemistry of solid inorganic particles in the urban environment. *Science of the Total Environment*, *763*, 144214.

Gailey, A. D., Schachter, A. E., Egendorf, S. P., & Mielke, H. W. (2020). Quantifying soil contamination and identifying interventions to limit health risks. *Current Problems in Pediatric and Adolescent Health Care*, *50*(1), 100740.

Granjon, D. (2021). shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'. R package version 2.0.2. https://CRAN.R-project.org/package=shinydashboardPlus

Gulson, B., & Taylor, A. (2017). A simple lead dust fall method predicts children's blood lead level: New evidence from Australia. *Environmental Research*, *159*, 76-81.

Hayhow, C., Brabander, D.J., Jim, R., Lively, M., Filippelli, G. (accepted). "We have been the ladder and held the ladder": Evolving GeoHealth models for actionable, community-engaged research. *GeoHealth.*

Hunt, A. (2016). Relative bioaccessibility of Pb-based paint in soil. *Environmental Geochemistry and Health*, *38*(4), 1037-1050.

Isley, C. F., Fry, K.L., Liu, X., Filippelli, G.M., Entwistle, J.A., Martin, A.P., Kah, M., Figueroa, D.M., Shuckle J., Jabeen, K., Famuyiwa, M.A., Wu, L., Sharifi Soltani, N., Doyi, I.N.Y., Argyraki, A., Ho, K.F., Dong, C., Gunkel-Grillon, P., Aelion, C. Marjorie, Taylor, M. P. (accepted). Global analysis of sources and human health risk associated with trace metal contaminants in residential indoor dust. *Environmental Science & Technology*.

Kelepertzis, E., Argyraki, A., Chrastný, V., Botsou, F., Skordas, K., Komárek, M., & Fouskas, A. (2020). Metal (loid) and isotopic tracing of Pb in soils, road and house dusts from the industrial area of Volos (central Greece). *Science of The Total Environment*, *725*, 138300.

Ko, S., Schaefer, P. D., Vicario, C. M., & Binns, H. J. (2007). Relationships of video assessments of touching and mouthing behaviors during outdoor play in urban residential yards to parental perceptions of child behaviors and blood lead levels. *Journal of Exposure Science & Environmental Epidemiology*, *17*(1), 47-57.

Koller, K., Brown, T., Spurgeon, A., & Levy, L. (2004). Recent developments in low-level lead exposure and intellectual impairment in children. *Environmental Health Perspectives*, *112*(9), 987-994.

Laidlaw, M. A., Zahran, S., Mielke, H. W., Taylor, M. P., & Filippelli, G. M. (2012). Re-suspension of lead contaminated urban soil as a dominant source of atmospheric lead in Birmingham, Chicago, Detroit and Pittsburgh, USA. *Atmospheric Environment*, *49*, 302-310.

Laidlaw, M.A., Filippelli, G.M., Brown, S., Paz-Ferreiro, J., Reichman, S.M., Netherway, P., Truskewycz, A., Ball, A.S. and Mielke, H.W. (2017). Case studies and evidence-based approaches to addressing urban soil lead contamination. *Applied Geochemistry*, *83*, 14-30.

Lanphear, B.P., Weitzman, M., Winter, N.L., Eberly, S., Yakir, B., Tanner, M., Emond, M. and Matte, T.D. (1996). Lead-contaminated house dust and urban children's blood lead levels. *American Journal of Public Health*, *86*(10), 1416-1421.

Levin, R., Vieira, C. L. Z., Rosenbaum, M. H., Bischoff, K., Mordarski, D. C., & Brown, M. J. (2021). The urban lead (Pb) burden in humans, animals and the natural environment. *Environmental Research*, 110377.

Low, R., Boger, R., Nelson, P., & Kimura, M. (2021). GLOBE Observer Mosquito Habitat Mapper Citizen Science Data 2017-2020. *GeoHealth*, e2021GH000436.

Masri, S., LeBrón, A., Logue, M., Valencia, E., Ruiz, A., Reyes, A., Lawrence, J.M. and Wu, J. (2020). Social and spatial distribution of soil lead concentrations in the City of Santa Ana, California: Implications for health inequities. *Science of The Total Environment*, *743*, 140764.

Masri, S., LeBrón, A. M., Logue, M. D., Valencia, E., Ruiz, A., Reyes, A., & Wu, J. (2021). Risk assessment of soil heavy metal contamination at the census tract level in the city of Santa Ana, CA: implications for health and environmental justice. *Environmental Science: Processes & Impacts*. 23, 812-830.

Meyer, I., Heinrich, J., & Lippold, U. (1999). Factors affecting lead, cadmium, and arsenic levels in house dust in a smelter town in eastern Germany. *Environmental Research*, *81*(1), 32-44.

Miler, M., & Gosar, M. (2013). Assessment of metal pollution sources by SEM/EDS analysis of solid particles in snow: a case study of Žerjav, Slovenia. *Microscopy and Microanalysis*, *19*(6), 1606-1619.

Needleman, H. (2004). Lead poisoning. *Annual Review of Medicine*, *55*, 209-222.

Obeng-Gyasi, E., Roostaei, J., & Gibson, J. M. (2021). Lead Distribution in Urban Soil in a Medium-Sized City: Household-Scale Analysis. *Environmental Science & Technology*, *55*(6), 3696-3705.

O'Shea, M. J., Vigliaturo, R., Choi, J. K., McKeon, T. P., Krekeler, M. P., & Gieré, R. (2021). Alteration of yellow traffic paint in simulated environmental and biological fluids. *Science of the Total Environment*, *750*, 141202.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rasmussen, P.E., Beauchemin, S., Chénier, M., Levesque, C., MacLean, L.C., Marro, L., Jones-Otazo, H., Petrovic, S., McDonald, L.T. and Gardner, H.D. (2011). Canadian house dust study: lead bioaccessibility and speciation. *Environmental Science & Technology*, *45*(11), 4959-4965.

Rasmussen, P. E., Levesque, C., Chénier, M., Gardner, H. D., Jones-Otazo, H., & Petrovic, S. (2013). Canadian House Dust Study: Population-based concentrations, loads and loading rates of arsenic, cadmium, chromium, copper, nickel, lead, and zinc inside urban homes. *Science of the Total Environment*, *443*, 520-529.

Rhoads, G. G., Ettinger, A. S., Weisel, C. P., Buckley, T. J., Goldman, K. D., Adgate, J., & Lioy, P. J. (1999). The effect of dust lead control on blood lead in toddlers: a randomized trial. *Pediatrics*, *103*(3), 551-555.

Ringwald, P., Chapin, C., Iceman, C., Tighe, M.E., Sisk, M., Peaslee, G.F., Peller, J. and Wells, E.M. (2021). Characterization and within-site variation of environmental metal concentrations around a contaminated site using a community-engaged approach. *Chemosphere*, *272*, 129915.

Spalinger, S. M., von Braun, M. C., Petrosyan, V., & von Lindern, I. H. (2007). Northern Idaho house dust and soil lead levels compared to the Bunker Hill Superfund site. *Environmental Monitoring and Assessment*, *130*(1), 57-72.

Stewart, L. R., Farver, J. R., Gorsevski, P. V., & Miner, J. G. (2014). Spatial prediction of blood lead levels in children in Toledo, OH using fuzzy sets and the site-specific IEUBK model. *Applied Geochemistry*, *45*, 120-129.

Taylor, M.P., Isley, C.F., Fry, K.L., Liu, X., Gillings, M.M., Rouillon, M., Soltani, N.S., Gore, D.B. and Filippelli, G.M. (2021). A citizen science approach to identifying trace metal contamination risks in urban gardens. *Environment International*, *155*, 106582.

Teran, K., Žibret, G., & Fanetti, M. (2020). Impact of urbanization and steel mill emissions on elemental composition of street dust and corresponding particle characterization. *Journal of Hazardous Materials*, *384*, 120963.

GeoHealth

Supporting Information for

Using Community Science to Better Understand Lead Exposure Risks

Matthew Dietrich[1*], John Shukle[1], Mark P.S. Krekeler,[2,3], Leah Wood,[1] Gabriel Filippelli[1,4]

[1]Department of Earth Sciences, Indiana University – Purdue University Indianapolis, Indianapolis, IN, USA

[2]Department of Geology & Environmental Earth Science, Miami University, Oxford, OH 45056, USA

[3]Department of Mathematical and Physical Sciences Miami University-Hamilton, Hamilton, OH 45011, USA

[4]Environmental Resilience Institute, Indiana University, Bloomington, IN, USA

*Corresponding Author Email: mjdietri@iu.edu

Contents of this file

Text S1 and Text S2

Figures S1 to S10

Introduction

Supporting text, including links to the data and code repository for the manuscript, are included below, as well as supporting figures.

Text S1.

Link to GitHub repository of data, R code, mobile app files, and logistic regression model output: https://github.com/dietrimj/Community-Science-Pb-Prediction

Note: The input CSV file has been updated to include a sample that contained survey responses and a Pb value of 80 mg/kg (right at our threshold for determination of high/low Pb). This was added to the testing data set a posteriori, which changed the confusion matrix output of the rmd file from 3 samples misclassified as “Low” Pb when they were really “High” Pb to 4 misclassified samples. This has been updated in Table 2 in the main manuscript.

Text S2.

Questions in the online DustSafe survey for the variables used in the logistic regression model in the manuscript and Table 1.

Approximately what year was your house built in?

____

Does the exterior of your house have any large areas of peeling paint?

Yes/No

Does the interior of your house have any large areas of peeling paint?

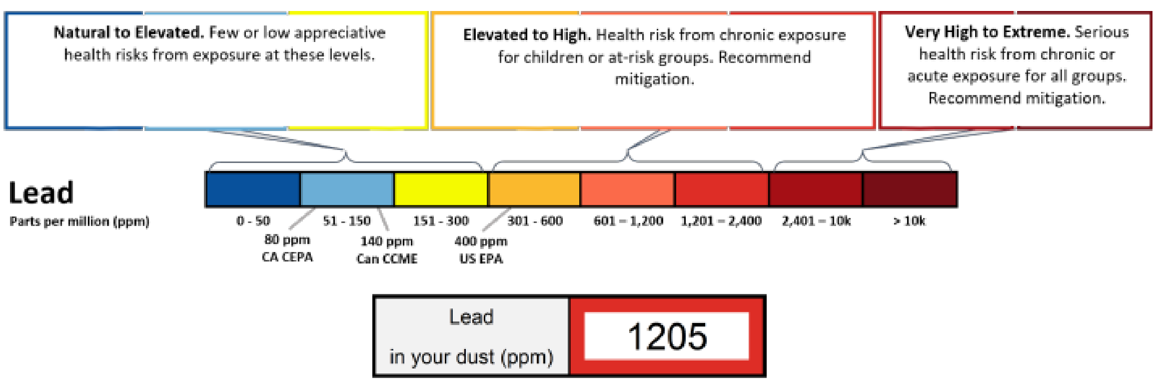
Yes/No

Understanding Contaminant Risk

Lead (Pb)

People often associate lead exposure with contaminated water; however, for most people around the world, **the dominant exposure pathway for lead is through dust**. Lead in house dust comes from a variety of sources, both inside and outside the home. Lead occurs naturally in the environment in small amounts, but leaded gasoline, lead-based paint, agricultural pesticides, and industrial pollution have all artificially increased lead levels in soil, especially near urban centers. Not all urban homes will have high lead in their soil/dust, but if you live in a city, being aware of the risks and getting your soil/dust tested are important first steps. Rural homes may have lead contaminated soil and dust as well. For example, if your home was built before 1980 it may contain some kinds of interior and exterior lead-based paint. Furthermore, rural and suburban homes constructed on land formerly used for agriculture, especially for fruit orchards, may have soil contaminated by lead-based pesticides.

Lead is equally toxic when it is inhaled as a dust or ingested in water. It accumulates in bones and is distributed throughout the blood to most organ systems, including the brain, where it acts as a neurotoxin. The US Agency for Toxic Substance and Disease Registry considers lead so toxic they list **no** acceptable Minimal Risk Level (MRL) for oral (water) or particulate inhalation (dust). Lead-based paint is especially dangerous to young children because of its sweet taste. As older lead-based paint ages and flakes, children may be tempted to eat the flakes because of this sweetness. Even low levels of lead exposure during early childhood can result in lifelong reduced IQ, decreased attention span, decreased impulse control and increased antisocial behavior. People who experience childhood lead poisoning make less money over their lifetime and are more likely to be incarcerated than their peers.



Additional information about lead as an environmental contaminant can be found here:
`https://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=22`

Figure S1: Example of a part of the report issued back to households participating in DustSafe, with lead (Pb) specifically shown as the reported element.

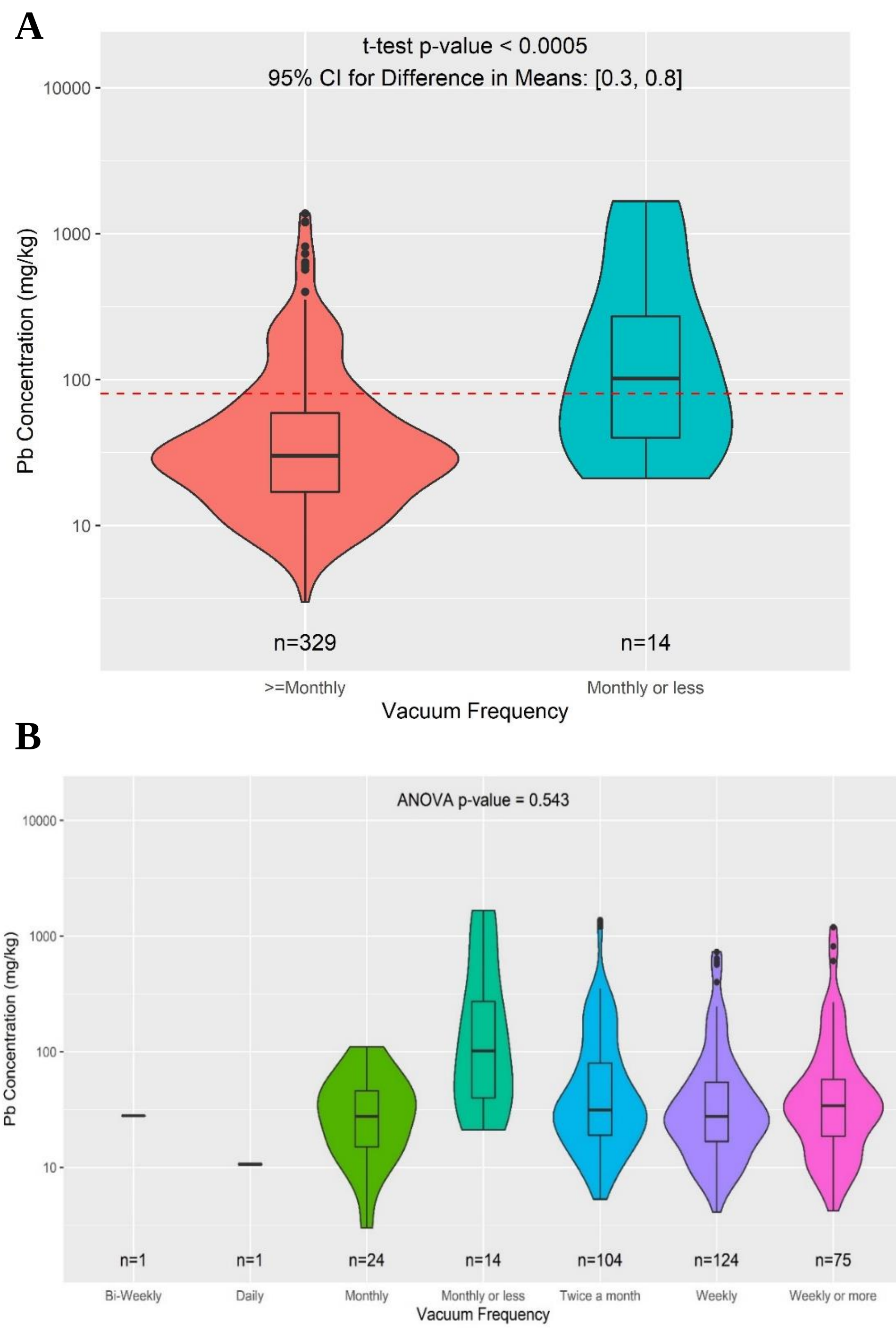


Figure S2: Embedded boxplots within violin plots for both monthly resolution vacuum frequency (A) and weekly resolution vacuum frequency (B). The boxes represent the interquartile range (IQR) of 25th-75th percentiles of data, the horizontal line is the median, and the whiskers represent 1.5 times the IQR. A two-sample paired t-test result between yes/no responses in (A) are also provided, while an ANOVA test p-value is provided in (B). The y-axes are transformed on a $\log_{10}$ scale, and the red dashed line in (A) represents California's safe screening level for soil Pb at 80 ppm.

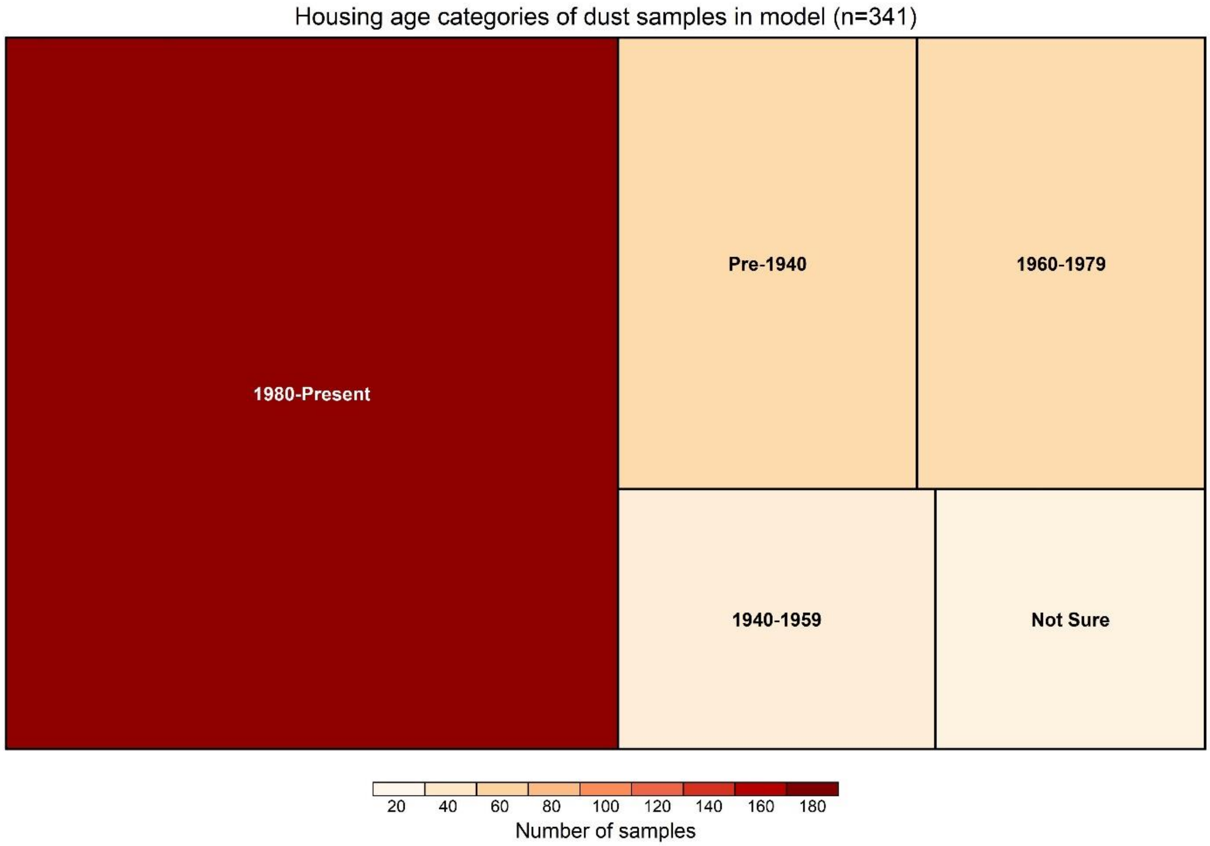


Figure S3: Treemap displaying the proportions (size of rectangles) and raw values (color shading) of housing age category responses from surveys that were included in the logistic regression model.

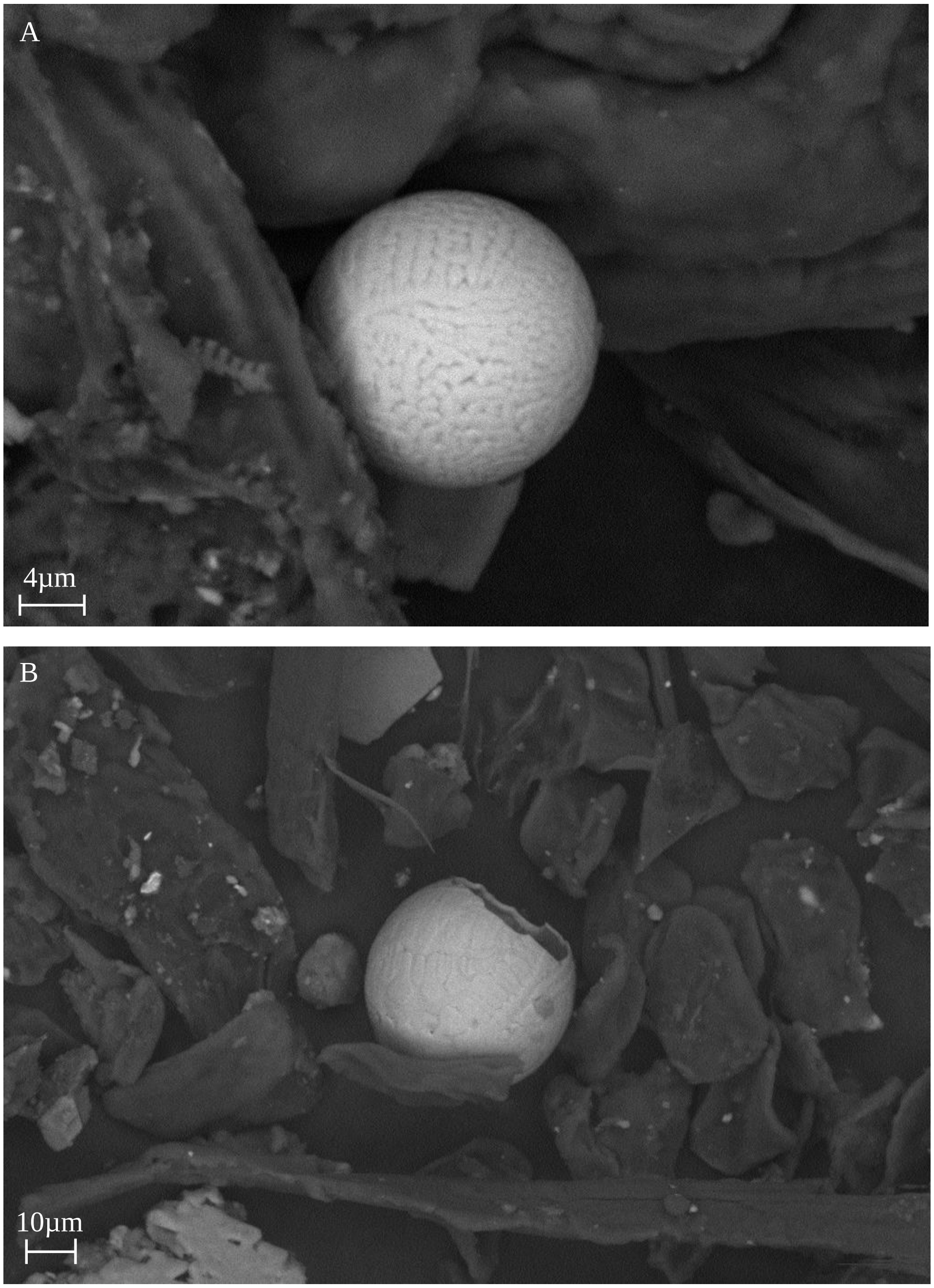


Figure S4: SEM imagery of technogenic Fe-oxide spherules likely of anthropogenic origin found in house dust (A—Sample AA0078, B—Sample AA0254).

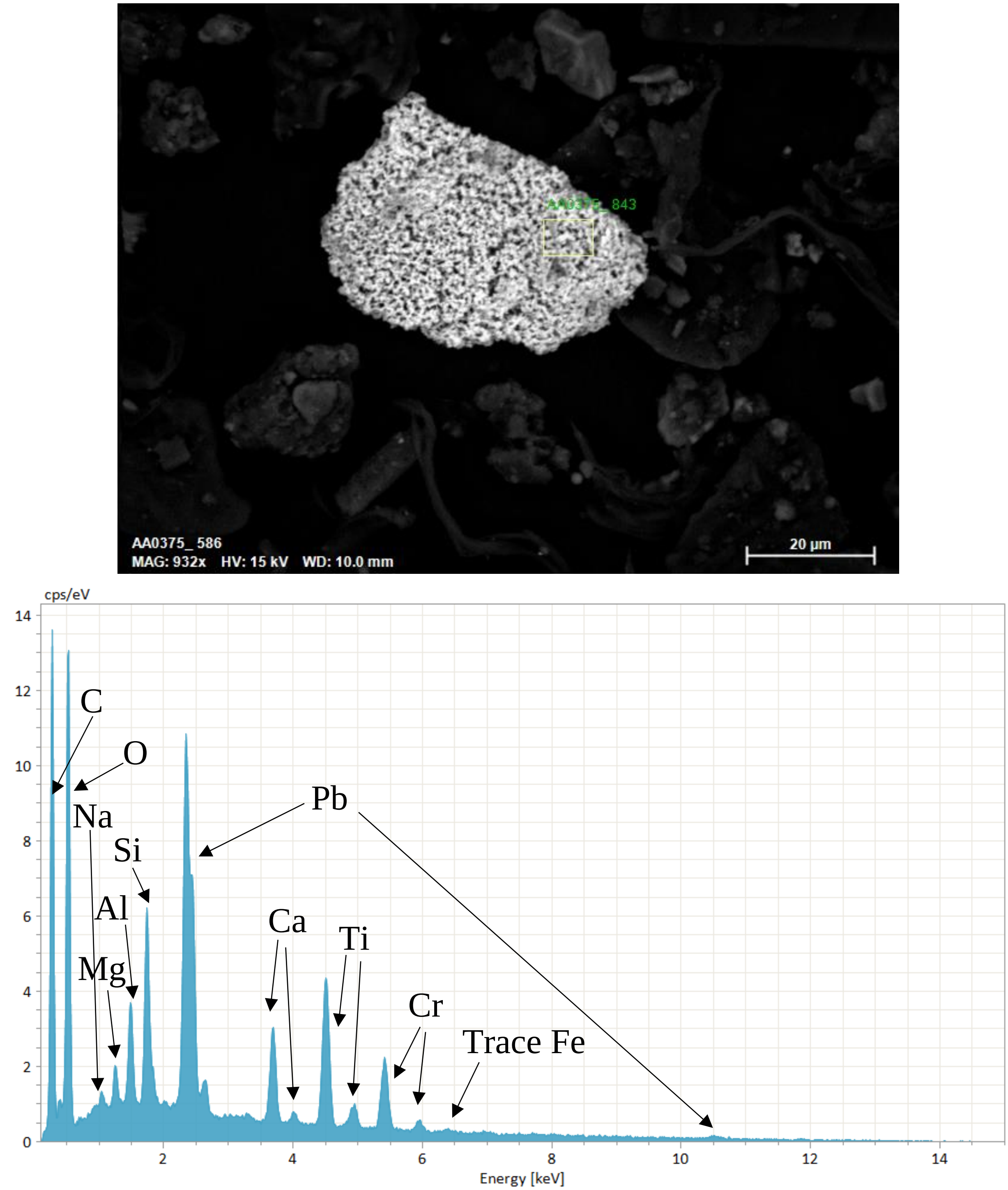
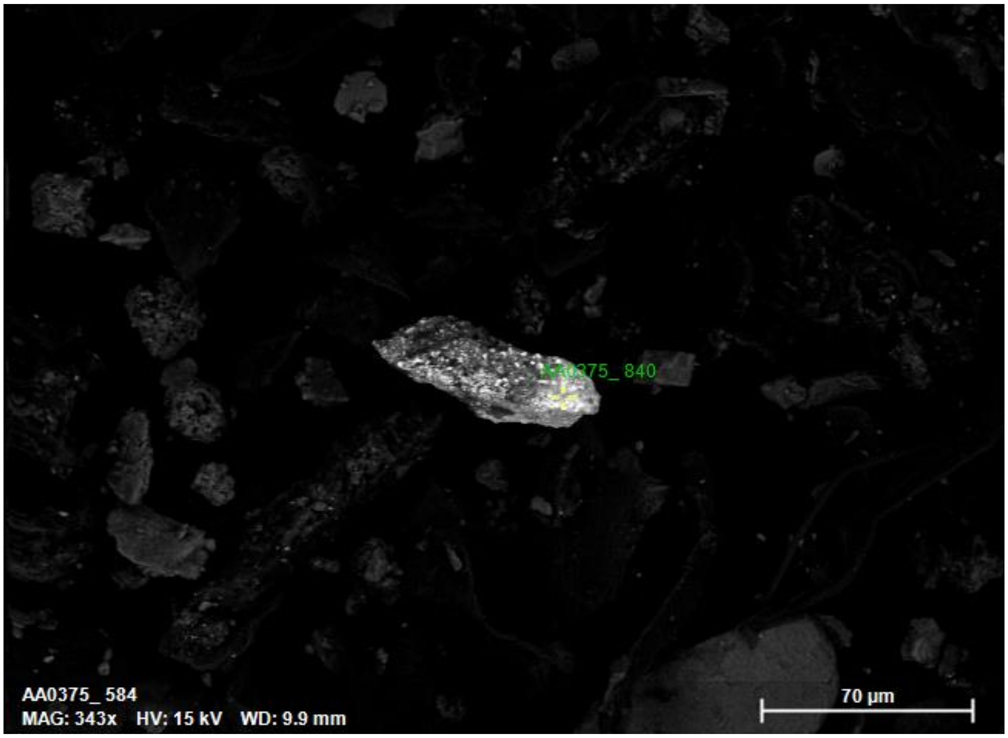


Figure S5: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5A.



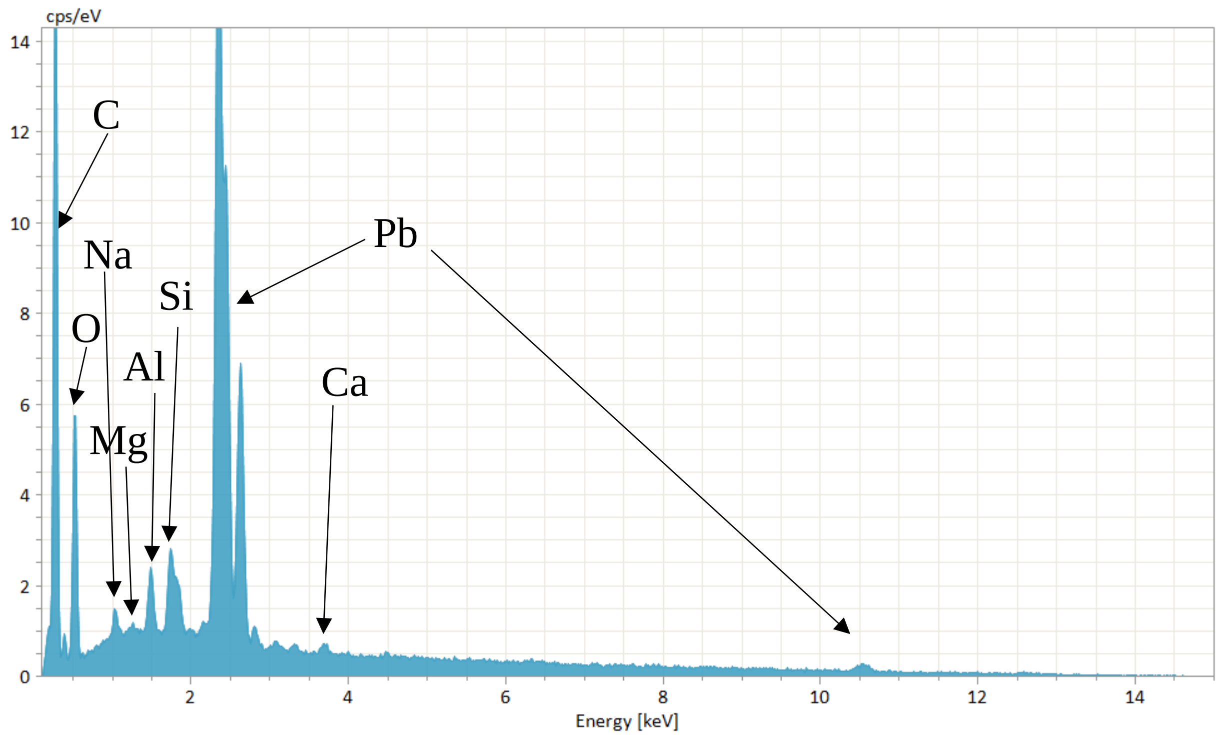
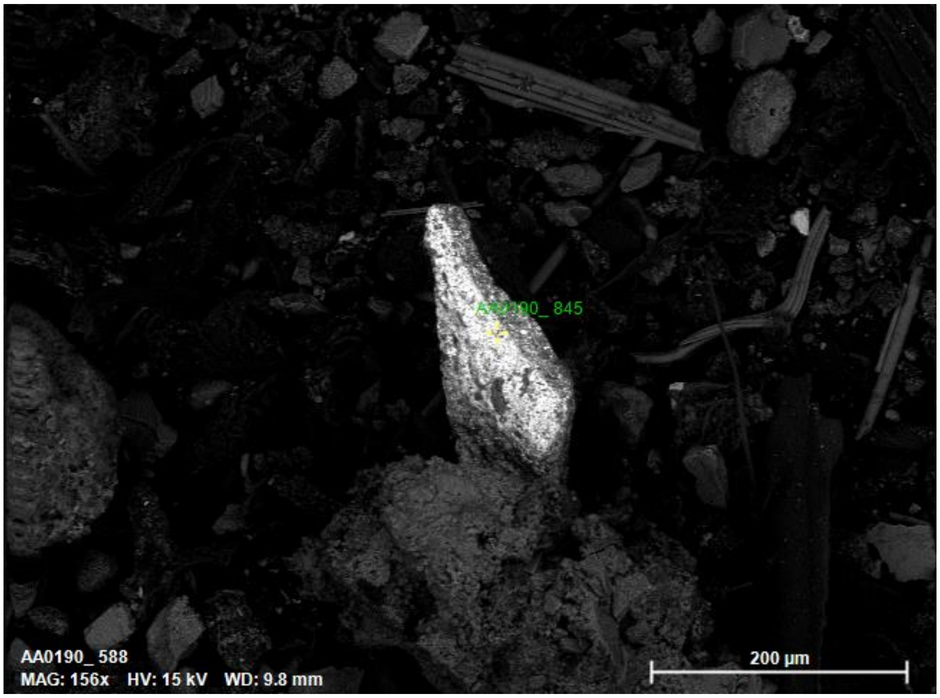


Figure S6: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5B.



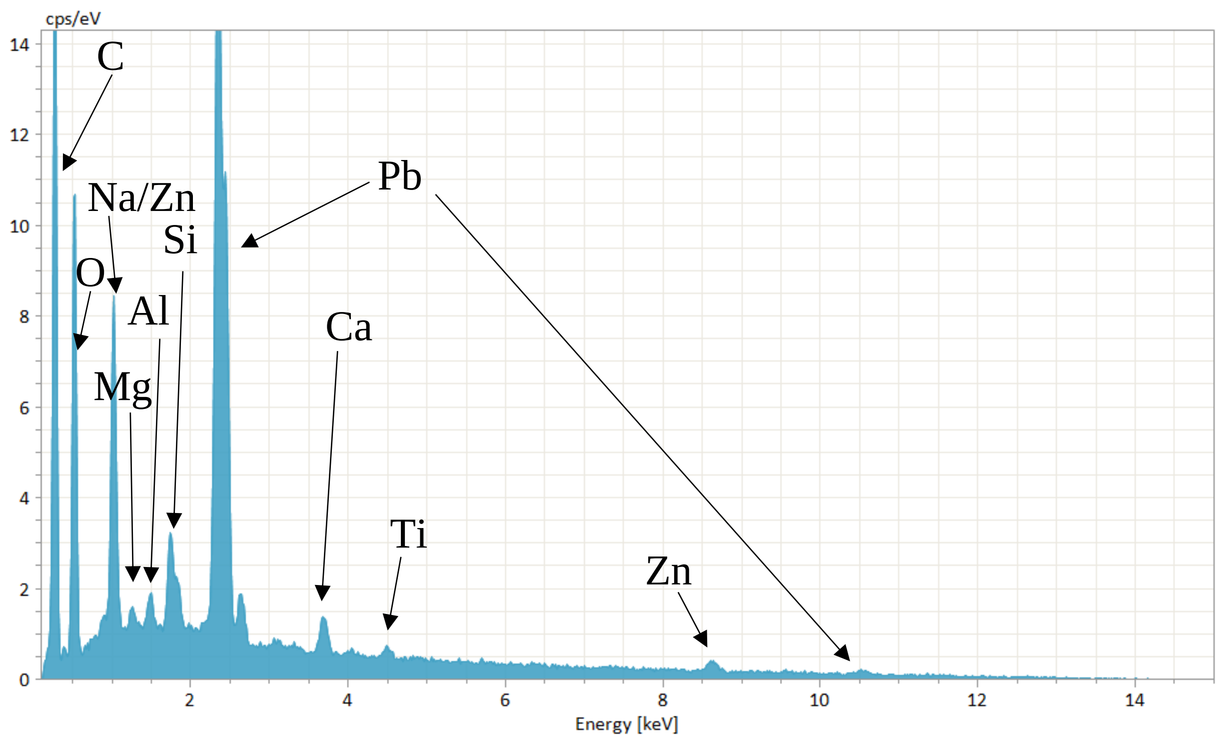
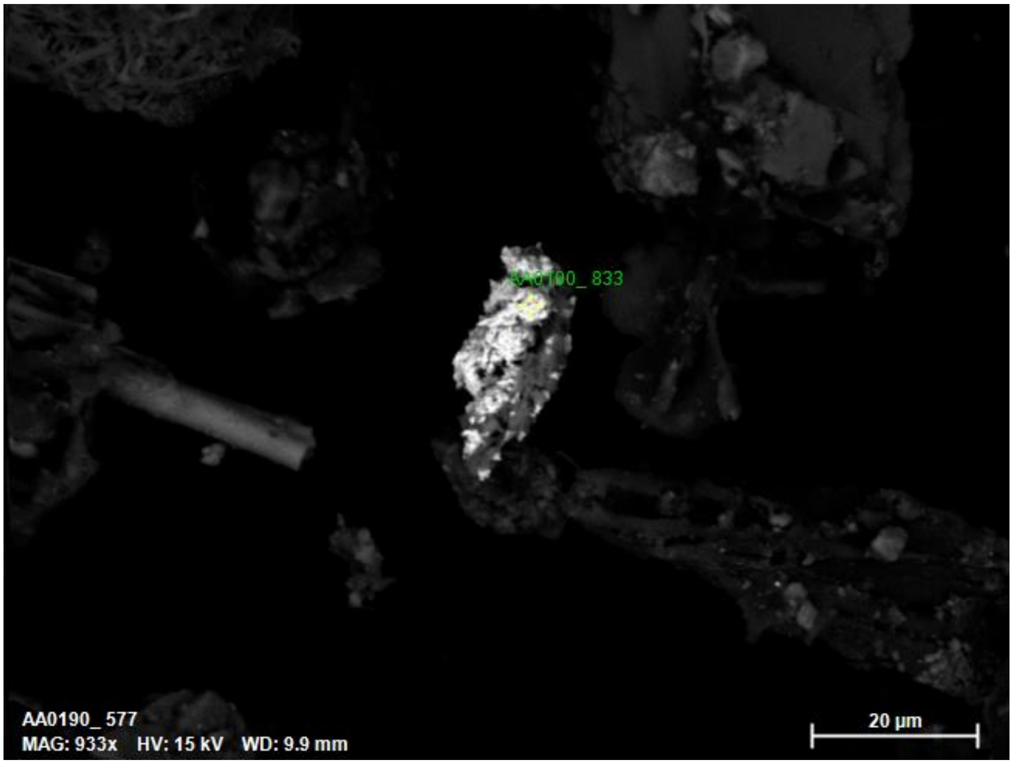


Figure S7: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5C.



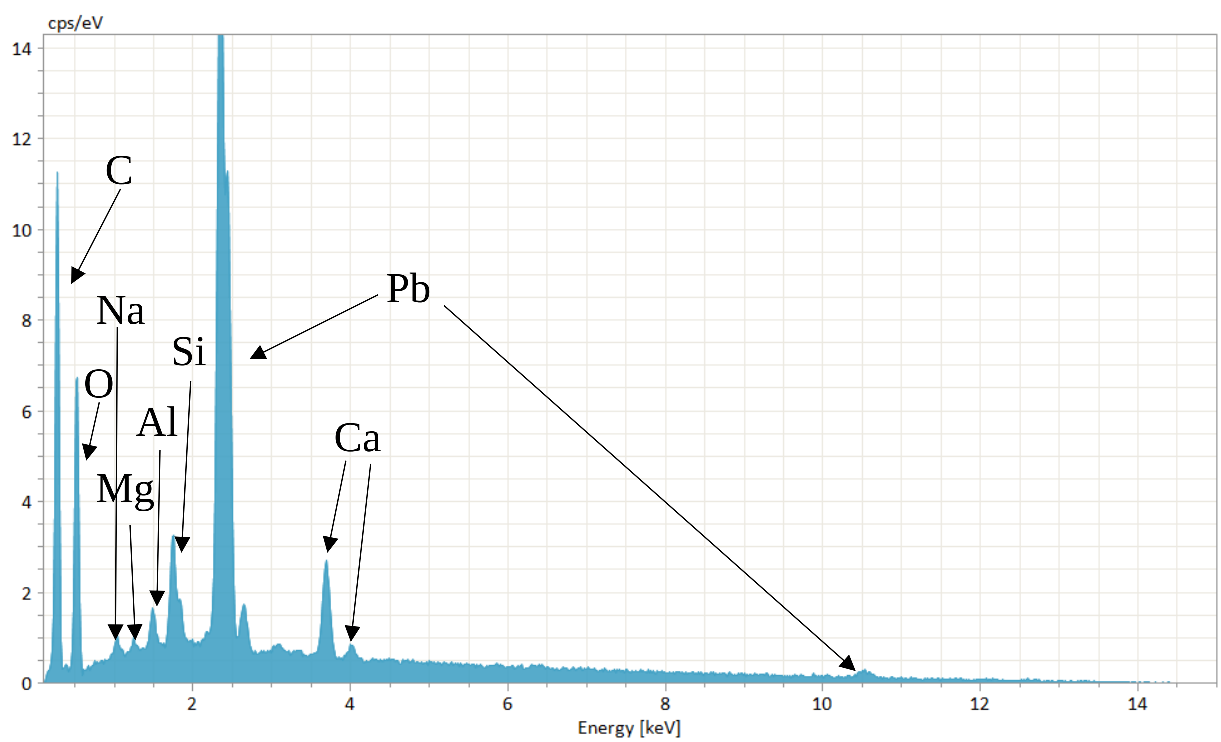
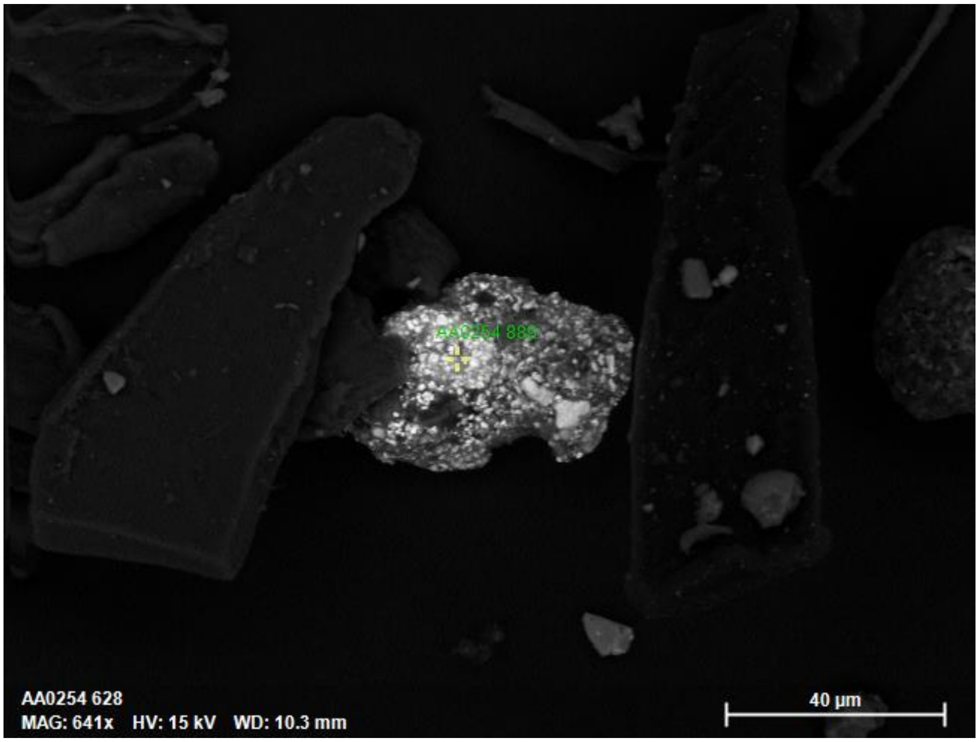


Figure S8: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5D.



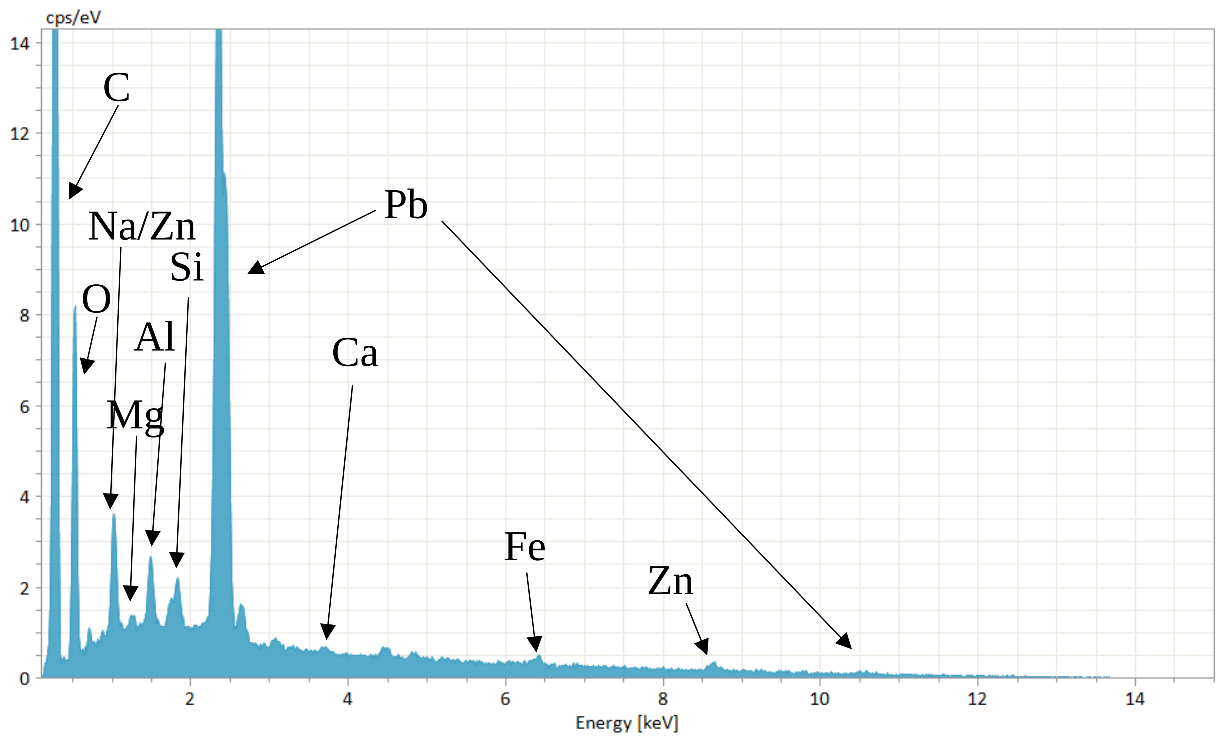
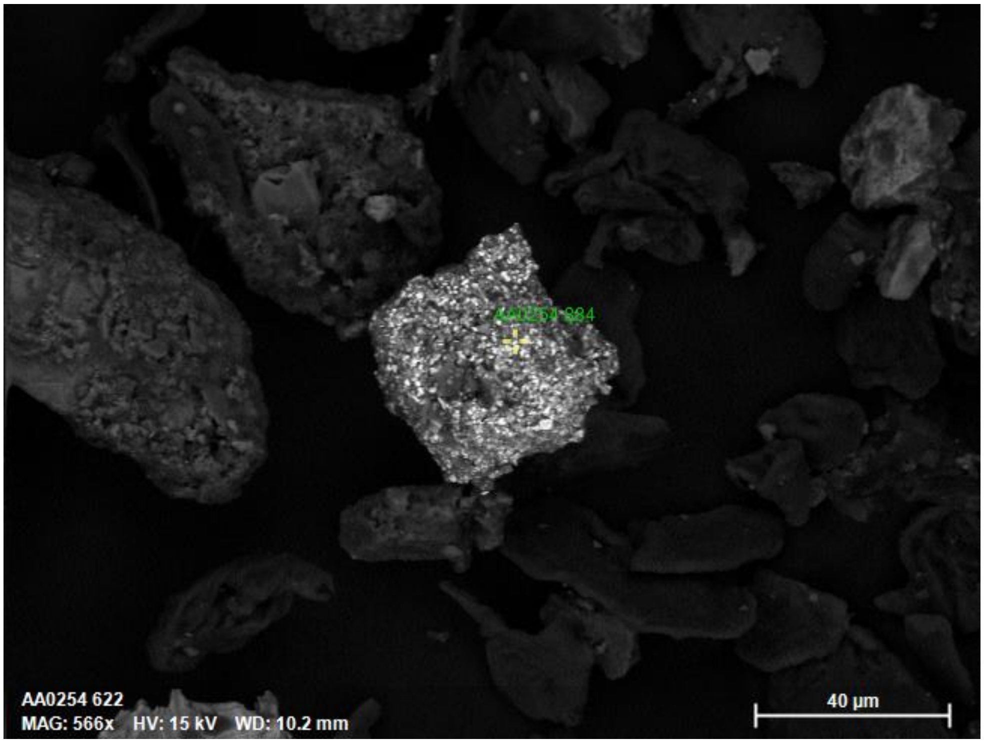


Figure S9: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5E.



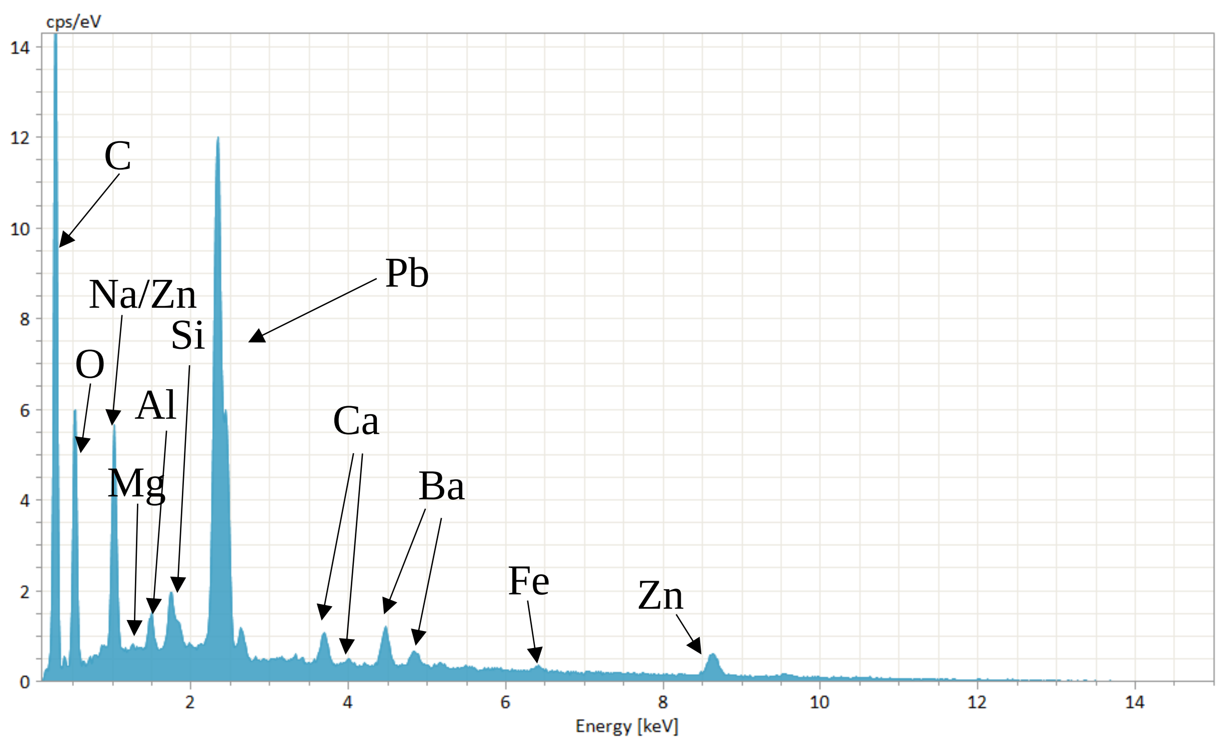


Figure S10: EDS spot point for analysis and the resulting spectra, with major peaks of detected elements labeled. For Pb particle in Fig. 5F. It is noted that the Pb spectra may be complicated by the presence of barite ($BaSO_4$).