

## Article

# A Data-Driven Approach to Assess the Risk of Encountering Hazardous Materials in the Building Stock Based on Environmental Inventories

Pei-Yu Wu <sup>1,2,\*</sup> , Kristina Mjörnell <sup>1,2</sup> , Mikael Mangold <sup>1</sup>, Claes Sandels <sup>1</sup> and Tim Johansson <sup>3</sup>

<sup>1</sup> RISE Research Institutes of Sweden, 412 58 Gothenburg, Sweden; kristina.mjornell@ri.se (K.M.); mikael.mangold@ri.se (M.M.); claes.sandels@ri.se (C.S.)

<sup>2</sup> Department of Building and Environmental Technology, Faculty of Engineering, Lund University, 221 00 Lund, Sweden

<sup>3</sup> Resources, Energy and Infrastructure, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden; tijoh@kth.se

\* Correspondence: pei-yu.wu@ri.se; Tel.: +46-70-462-07-23

**Abstract:** The presence of hazardous materials hinders the circular economy in construction and demolition waste management. However, traditional environmental investigations are costly and time-consuming, and thus lead to limited adoption. To deal with these challenges, the study investigated the possibility of employing registered records as input data to achieve in situ hazardous building materials management at a large scale. Through characterizing the eligible building groups in question, the risk of unexpected cost and delay due to acute abatement could be mitigated. Merging the national building registers and the environmental inventory from renovated and demolished buildings in the City of Gothenburg, a training dataset was created for data validation and statistical operations. Four types of inventories were evaluated to identify the building groups with adequate data size and data quality. The observations' representativeness was described by plotting the distribution of building features between the Gothenburg dataset and the training dataset. Evaluating the missing data and the positive detection rates affirmed that reports and protocols could locate hazardous materials in the building stock. The asbestos and polychlorinated biphenyl (PCB)-containing materials with high positive detection rates were highlighted and discussed. Moreover, the potential inventory types and building groups for future machine learning prediction were delineated through the cross-validation matrix. The novel study contributes to the method development for assessing the risk of residual hazardous materials in buildings.

**Keywords:** hazardous materials; asbestos; PCB; environmental investigation; statistical inference; cross-validation; machine learning pre-processing



**Citation:** Wu, P.-Y.; Mjörnell, K.; Mangold, M.; Sandels, C.; Johansson, T. A Data-Driven Approach to Assess the Risk of Encountering Hazardous Materials in the Building Stock Based on Environmental Inventories.

*Sustainability* **2021**, *13*, 7836.

<https://doi.org/10.3390/su13147836>

Academic Editor: Wann-Ming Wey

Received: 31 May 2021

Accepted: 10 July 2021

Published: 13 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Although a series of bans for the use of hazardous materials in construction have been imposed since the 1970s, an appreciable quantity of contaminated materials remains in the existing building stocks [1]. The frequent presence of asbestos-containing materials [2,3] and PCBs (polychlorinated biphenyls)-containing components [4] is the result of mass production and adoption in the 1920s–1990s [5]. In addition to the negative concerns of human health and the environment, demolished and renovated projects become more expensive and take a longer time if hazardous materials are encountered unexpectedly. The decontamination and abatement cost account for a noticeable amount for waste disposal and working precautions' preparation [6].

Advanced data mining and statistical learning have been accessible to emerging research fields in recent years. Information about the building stock has also been made more available, mainly through governmental open database initiatives in several different areas, i.e., investigation records, project economy, and so on [6]. These two developments

dramatically improve the estimation capability for locating hazardous building materials in demolition and renovation projects. Furthermore, by coupling building stock data and hazardous product registers, predictive detection of in situ hazardous building materials is made possible [7].

The study explores and develops a data-driven approach to assess hazardous materials' detection in the building stock. The importance of effective hazardous materials management is recognized through updating legal requirements and extending criteria for a healthy living environment in building certification [8]. By increasing the quality control and locating the potential in situ hazardous materials, a step toward the circular economy for construction and demolition waste (C&DW) can be realized [9]. For instance, mandatory pre-demolition audits (also called environmental audits or waste audits) are enforced in Austria, Bulgaria, the Czech Republic, Finland, France, Hungary, Luxembourg, The Netherlands, Romania, Spain, and Sweden, whereas optional environmental investigations have been applied to a limited extent in Belgium, Denmark, Germany, Ireland, Italy, Slovakia, and the United Kingdom for 5–10 years [10]. Advantages for pre-demolition audits include improving hazardous construction and demolition waste identification, as well as promoting resource circularity and efficient use of mixed wastes [11]. According to the breakdown of construction and demolition wastes' generation, Sweden presents a higher percentage of hazardous wastes (13%) than the average among European Union (EU) countries (2.5%) owing to its sound separation systems as well as a long tradition of environmental legislation [9]. The use of asbestos and PCB in building materials has been prohibited since the middle of the 1970s in Sweden. Several other EU countries have also achieved advanced progress in establishing waste management systems and databases. For example, pre-demolition audits for the certain scale of the non-residential building are mandatory in Flanders [12]. With an increasing number of emerging databases and extensive documentation, the goal of tracing in situ hazardous building materials through employing data mining on registered records could be attained [13].

Built upon statistics, machine learning, and pattern recognition techniques, data mining enables automatic or semi-automatic exploration of large amounts of data to discover patterns or rules [14]. Owing to these benefits and building data availability, the potential of data-driven built environment management is shown. However, previous literature sheds light on significant challenges for practical implementation, including time-consuming pre-processing obtaining complete digital datasets [15]. Furthermore, limited information regarding the extent of the previous adoption is available, leading to a struggle in designing precautionary abatement policy and decontamination plans [6]. Several researchers attempted to detect asbestos-containing materials through developing new methodologies considering these barriers and knowledge gaps. Employing temporal descriptions of materials in an ontology-based approach; prediction of the presence of asbestos in buildings was explored by Mecharnia et al. [7]. Similarly, statistics were also employed in study inspection reports and online demolition databases to quantify the amounts and abatement costs for asbestos-containing materials in abandoned residential dwellings by Franzblau et al. [6]. Govorko et al. [2,16] developed a mobile application to investigate the types and the conditions of asbestos-containing materials in residential settings.

To realize the Construction 2020 strategy [17] and the Communication on Resource Efficiency Opportunities in the Building Sector [18], the protocols and guidelines for waste audits before demolition and renovation for buildings were established by the EU Commission [19]. The emergence of relevant tools and complementary legislation is expected to improve current practice in waste identification, source separation, and collection [11]. Although Sweden has introduced obligatory environmental audits since 1995, several practical predicaments exist for using the data, i.e., the hardcopies of environmental inventories are kept by several different authorities, a harmonized protocol at the national level is lacking, and no digital query-based database of in situ building materials is available yet [11]. This article attempts to address these challenges by developing a generalizable approach and extending the investigation scope to multiple hazardous materials in the building

stock. The empirical study aims to quantify the risk of residual hazardous materials in the existing buildings and investigate the data quality and quantity of the environmental audits for advanced analysis. Through a case study in the City of Gothenburg, the potential of using environmental inventories to assess the extent of in situ hazardous materials has been explored. The first part of the research involves data assembling and validation, followed by cross-comparison and descriptive statistics of the training dataset. The study results can offer valuable insights into the frequent occurrence of hazardous materials in demolished and renovated buildings and specify building groups where occurrences are more likely. Furthermore, the pilot work lays a good foundation for the subsequent machine learning modeling to predict the presence of hazardous materials. To achieve the research objectives, three research questions are formulated as follows:

RQ1: What is the potential for employing environmental inventories as input data to assess the presence of hazardous materials in the building stock?

RQ2: How representative is the training dataset in relation to the Gothenburg building stock?

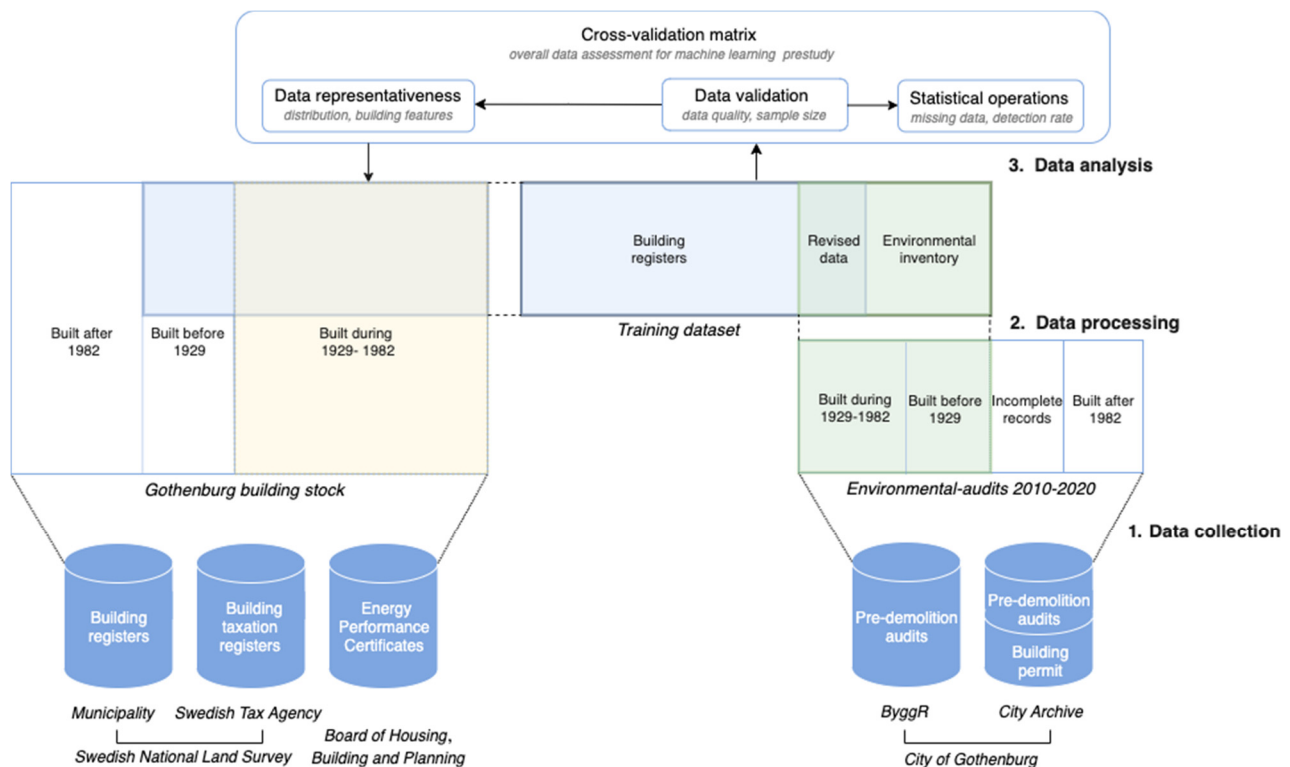
RQ3: How can the risk of encountering hazardous materials in the building stock be assessed?

## 2. Materials and Methods

### 2.1. Study Design

Given that no digital pre-demolition audit dataset exists in Sweden, nor can building material records be found in the national building registers, the study proposed an innovative data coupling method by adding environmental inventories from the field study to the building information database from authorities. A similar data coupling approach has been performed by Wilk et al. [20] and Krówczyńska et al. [21] to study the spatial distribution of asbestos-cement roofing. To assess the potential of using environmental inventories for hazardous materials identification, a training dataset consisting of pre-demolition audits from demolished and renovated buildings in the City of Gothenburg constructed earlier than 1982 and national building registers at a regional scale was created. The developed training dataset compiling the registered records of the environmental investigation from the past decade can be regarded as a pioneered study of sustainable building material management. Data validation for the acquired documents becomes a fundamental step for the future machine learning study that leverages the existing data labels for predicting the potential presence of the remaining hazardous materials.

The study design illustrated in Figure 1 followed the procedure of training dataset creation, processing, and analysis. Firstly, registered-based data were collected from various databases for quality and quantity control. Then, the following data processing included data reformatting, merging, and cleaning. Finally, data analysis was performed in four parts: validating data quality and quantity, evaluating data representativeness with the Gothenburg building stock, assessing missing data and the detection records of each building subgroup, and cross-validating investigation data for risk assessment.



**Figure 1.** A proposed procedure for creating the training dataset comprised of (1) data collection, (2) data processing, and (3) data analysis.

## 2.2. Data Collection

Several data collection relating tasks were performed sequentially to ensure data operability. First and foremost, environmental inventories and national building registers were collected from different authorities. Pre-demolition audits were gathered from the Archive of the City of Gothenburg during the permit application period from 2010 to 2020. Currently, no query-based database of environmental records exists, which allows free search in the text masses. Therefore, the search process was done manually in the building permit register system using the keywords “demolition”, “renovation”, “reconstruction”, “modification”, and “alteration” in the document titles. Extensive document screening was executed to identify the projects with environmental audits in their permit decisions. Thereafter, investigation documents were requested and reformatted into a digital dataset.

Meanwhile, national building registers were received from the Swedish Cadastral and Land Registration Authority, where real estate registers from municipalities and the Swedish Tax Agency are kept. Besides, the Energy Performance Certificates were collected from the Swedish National Board of Housing, Building, and Planning. Merging these national datasets was executed with GIS Feature Manipulation Engine from Safe Software to constitute the comprehensive dataset for the research purpose. This comprehensive dataset comprised registered buildings from the three major economic regions in Sweden: Stockholm, Gothenburg, and Malmö regions. The methods for merging national datasets were developed by Johansson et al. [22] and can add additional auxiliary data for analyses. Then, the data extraction was conducted to retrieve the Gothenburg city dataset for the representativeness study concerning building characteristics. Many-to-many relationships were not included in the larger Gothenburg dataset as register data were at the property level, giving a data loss of approximately 10%. The metadata of the available Gothenburg dataset are appended in Appendix A [23]. Built upon the general building information from the national building registers and the detection records from the environmental inventories, the training dataset, a subset of the Gothenburg building stock, was created

for further data processing and analysis. Many-to-many relationships were examined manually for the observed buildings in the training dataset.

### 2.3. Data Processing

To ensure coherent documentation of environmental inventories and improve the data readability for coding software, a standard procedure was developed and executed iteratively in creating the training dataset. The process consisted of (1) creating a dataset structure by assembling common variables across environmental inventories with the building as an observation unit; (2) checking data eligibility in terms of construction year and investigation completeness; (3) leveling data quality by clustering inventory types and investigator's experiences, then converting the data to pre-defined data types; (4) extracting relevant building registers from the comprehensive datasets using national real estate index and harmonizing updates across datasets; (5) merging and reformatting building registers and environmental inventory to become a training dataset; and (6) revising and manipulating variables of interest through aggregating multiple records to verify data consistency for construction year, renovation year, area, and so on. The final variables of interest and their metadata in the training dataset are presented in Table 1.

**Table 1.** Overview of metadata in the training dataset.

Variable Category	Variable Specification	Data Type	Data Source
<b>Environmental inventories</b>			
Matching keys	National real estate index	String and numerics [index]	Permit registers
Investigation	Address	String	Environmental audits
	Document type	Nominal [report, protocol, control plan, demolition plan]	
	Scope	Ordinal [entire, part]	
	Investigation year	Scale variable [year]	
Hazardous substance	Investigator	String	Environmental audits
	Decontamination	Nominal [asbestos, PCB, NA]	
	Asbestos	Nominal [positive, negative, NA]	
	PCB (polychlorinated biphenyl)	Nominal [positive, negative, NA]	
Hazardous material	CFC (chlorofluorocarbon)	Nominal [positive, negative, NA]	Environmental audits
	Mercury	Nominal [positive, negative, NA]	
	Building components *	Nominal [positive, negative, NA]	
	Building parameter	String [type]	
Building parameter	Class	String [type]	Permit registers
	Usage of the building	String	
	Construction year	Scale variable [year]	
	Renovation year	Scale variable [year]	
	Renovation part	String [the extent of renovation]	
	Building structure **	String [material]	
	Interior areas (BOA)	Scales [m <sup>2</sup> ]	
Floors	Ordinal		
<b>National building registers</b>			
Matching keys	National real estate index	String and numerics [index]	Swedish Tax Agency
Building usage	Address	String	Board of Housing Municipality
	Class code	Nominal	
Building parameter	Category code	Nominal	Swedish Tax Agency
	Construction year	Scale variable [year]	
	Renovation year	Scale variable [year]	
	Complexity	Nominal [complex, non-complex]	
Building area	Ventilation type	Nominal [exhaust, balanced, balanced with heat exchanger, exhaust with heat pump, natural ventilation]	Board of Housing
	Interior areas (BOA) ***	Scales [m <sup>2</sup> ]	
	Floors	Ordinal	

\* Building components imply the building materials that contain primary or secondary contaminants. The former are made of hazardous substances and the latter contain transmitted hazardous substances through external sources. \*\* Building structure describes the building components, including foundation, structure, roof, and façade. \*\*\* BOA and LOA are property valuation measures of usable heated floor area for habitation and non-habitation. BOA and LOA are registered for 90% of the multifamily buildings and 74% of the single-family houses [23].

Necessary data processing was executed to assure uniform data input from heterogeneous data sources. As the data update varied among different authorities, the registered data for variables of interest were compared with the inventory records to determine the actual investigated part. These registered data were used as proxies for filling missing data from the inventories. To assure data alignment for analysis, revised variables were created by prioritizing the inventory data and complementing them with the registered data. If none of the registers contained the information, the variables were labeled as NA. Moreover, irrelevant observations, such as buildings constructed after the ban of asbestos

and PCB in building materials in 1982, as well as the updated registers for reconstructed buildings, were removed. The clean dataset summed up to 402 observations.

The detection results were collected from various environmental investigations and for all types of building usages. Four inventory types were identified based on the document title and the content format: report, protocol, control plan, and demolition plan. Reports contain the most thorough investigation records with test sample results. In comparison, protocols were developed by the municipality with a list of binary questions for the investigated hazardous materials and their amount. Control plans were used for small houses and simple buildings, and hazardous substances were generalized without specifying specific materials. Demolition plans are required documents for demolition permit application, and free text is used to describe the detection of hazardous substances or materials. Considering the various extent of environmental investigations, primary hazardous substances such as asbestos, PCB (polychlorinated biphenyls), CFC (chlorofluorocarbon), and mercury were included in the training dataset. The detection results were documented at two levels in a binary way: hazardous substances and hazardous building materials. Besides, specific building parameters, including construction year, renovation year, detailed usage, area, and the number of the floors, were also noted as data labels. The data quality was controlled through a cross-validation workshop with the research group and a domain expert to affirm the correct interpretation of the inventory documents.

Furthermore, building classes were created by reference to the description of the renovation or demolition permit, primary usage of the building stated in the inventory, as well as building types and building categories from the national building registers. According to the actual investigation area and the past activities, the 402 observations were categorized into ten building classes: single-family house, multifamily house, temporary dwelling, school, office, commercial building, production building, industrial building, warehouse, and other/infrastructure. Determining the building class can facilitate clustering the buildings with similar scale and construction tectonic. The categorization of the inventory types and the building classes is fundamental to structure the data subgroups for comparative analysis.

#### 2.4. Data Analysis

Through conducting statistical operations on the 402 observations in the training dataset, data representativeness and risk assessment of encountering hazardous building materials were addressed. Python's built-in library and interactive modules such as Pandas, Matplotlib, and Seaborn were employed for the explorative data analysis. The training dataset's representativeness was evaluated by comparing the building parameters' mean values with the same parameters in the Gothenburg dataset 1929–1982. Furthermore, the underlying correlations between the positive detection rates and different clustering subgroups were ascertained, i.e., inventory type, building class, construction year, and area. The descriptive statistical results provided an overall picture to assess the positive detection rates of residual hazardous materials in the building stock. To finalize the statistical results and set the scene for the future machine learning study, a cross-validation matrix evaluating the data quality and quantity was created.

Based on the cross-validation matrix, the data subgroup for each building class and investigated materials were assigned an assessment score. The assessment scores were created following (1) below. First of all, the investigation results for each building class were transformed into dummy variables, and four inventory types were given different weights based on the level of detail. The weights from high to low in decile points were assigned to the report, the protocol, the control plan, and the demolition plan, respectively. For each hazardous material in a given building class, the number of observations for various inventory types was multiplied by individual inventory weight, then the results were summed and divided by the number of observations. After evaluating the observation quality on an individual basis, a data number threshold was introduced to assess missing values for each subgroup. For example, if the data size was more than 30 observations,



denoted as 1, between 15 and 30 observations were marked as 0.5, otherwise they were 0. Taking data size into account allowed assessing whether the observation number was large enough for generating useful statistical results. Cross-validating the individual observation quality and missing data in each subgroup, we can distinguish whether the detection results were reliable through adopting a data boundary. In the end, the findings were summarized and indicated the data subgroups that were found to be promising for machine learning pre-processing.

$$y = \frac{(Ir \times r + Ip \times np + Ic \times nc + Id \times nd)}{n} * K \quad (1)$$

$y$  = Assessment score.

$I$  = Inventory type for weighting the individual observation.  $I = 1$  if is the report ( $r$ ),  $I = 0.75$  if is the protocol ( $p$ ),  $I = 0.5$  if is the control plan ( $c$ ), and  $I = 0.25$  if is the demolition plan ( $d$ ).

$n$  = Number of the observations in the studied subgroup.

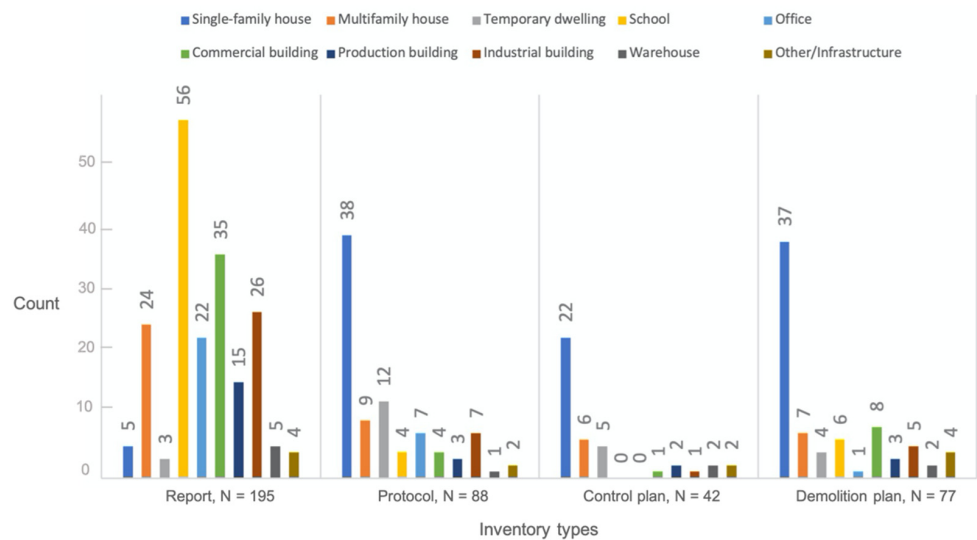
$K$  = Number of the observations enough for statistical operation.  $K = 1$  if  $n \geq 30$ ,  $K = 0.5$  if  $15 \leq n < 30$ ,  $K = 0$  if  $n < 15$ .

### 3. Results and Discussions

The results and discussions are structured in five parts: evaluating data quality and size, data representativeness, statistical operations, cross-validation matrix, and method replicability. Examining data quality and data size facilitated identifying subgroups appropriate for data analysis. By comparing the Gothenburg dataset and the training dataset, the distribution of building parameters was displayed to show data representativeness. Subsequently, the positive detection rates were highlighted with missing values for hazardous substances and materials through clustering with different parameters, such as inventory type, building class, construction year, and area. To summarize the previous analysis results and minimize the possible errors involving heterogeneous data, a cross-validation matrix was created as an indicator based on data quality and quantity for investigated materials in each building class. Finally, a short discussion regarding the method replicability to other contexts and the relation to previous research were discussed at the end of the section.

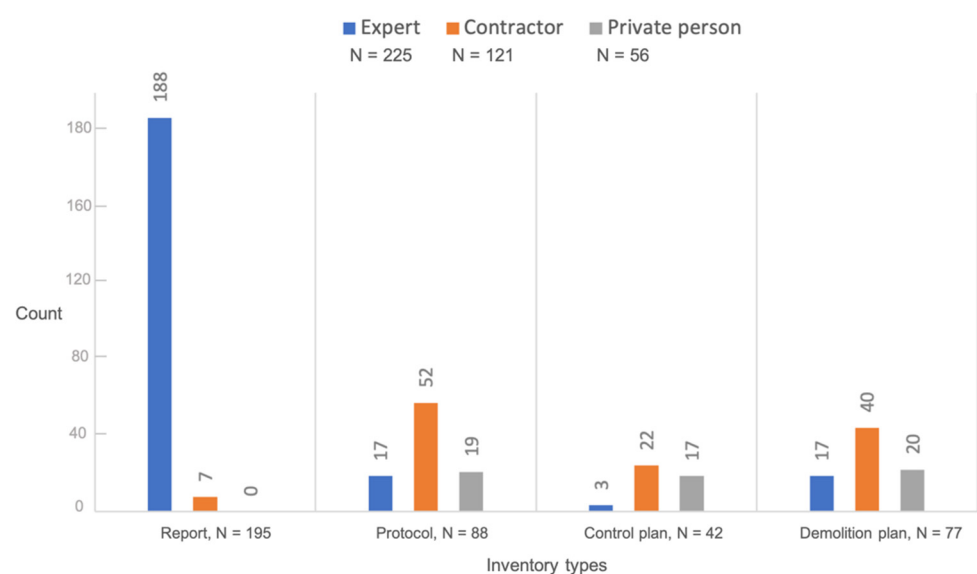
#### 3.1. Evaluating Data Quality and Size

In Figure 2, the inventory types were ranked in descending order according to investigations' comprehensiveness and documentation details. The high-detailed levels of environmental inventories specified the presence of hazardous substances and containing materials in semi-uniform formats, such as report and protocol, constituted 48.5% and 21.6% of the investigations, respectively. They described whether missing data is the result of investigations that have not been done or because investigation materials are not in place. Field sampling of hazardous materials was usually executed by hazardous waste experts, lowering the risk of mislabeling caused by visual distinguishment. Conversely, hazardous materials' information is occasionally missing in simple investigations such as control plans or demolition plans. These information sources are only useful when estimating contamination at the building level. Data granularity became visible by clustering inventory types and building classes. High data granular building groups came from reports mainly by hazardous waste experts and contractors for large-scale and complex buildings. These included schools (13.9%), commercial buildings (8.7%), industrial buildings (6.5%), multifamily houses (6.0%), and offices (5.5%). Owing to the risk of polluted operations, industrial and production buildings, in most cases, require thorough environmental investigations by legislation. On the contrary, pre-demolition audits for single-family houses and temporary buildings mainly consisted of protocols, control plans, and demolition plans performed by contractors or private persons.



**Figure 2.** Clustering the building classes according to the data sources in the training dataset through aggregating the number of inventories (N = 402).

An overview of the experience level of the investigators across the environmental inventories helped evaluate the data source quality, as shown in Figure 3. Around 56.0% of the observations in the training dataset were performed by hazardous waste experts who are skilled in doing complicated environmental investigations. These environment consultants were primarily involved in drafting environmental reports and field sampling for the buildings obliged to pre-demolition audits. Another one-third of the observations came from contractors, such as demolition companies or waste handling companies. They are responsible for the demolition work and permit application, and thus skillful in making protocols, demolition plans, and control plans. The rest, 13.9% of the inventories, were done by private persons. They could be the building owners who only do the ocular inspections for a part of buildings. Considering the proportion of the observations carried out by hazardous waste experts or contractors, the reliability of the data is high.



**Figure 3.** Overview of the experience level of the environmental investigators across the source data (N = 402).

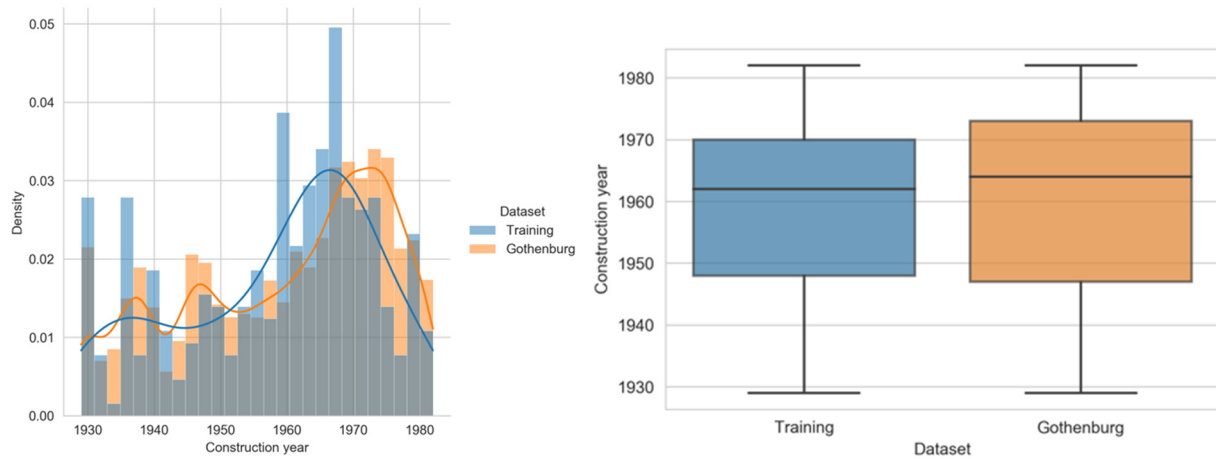


### 3.2. Data Representativeness

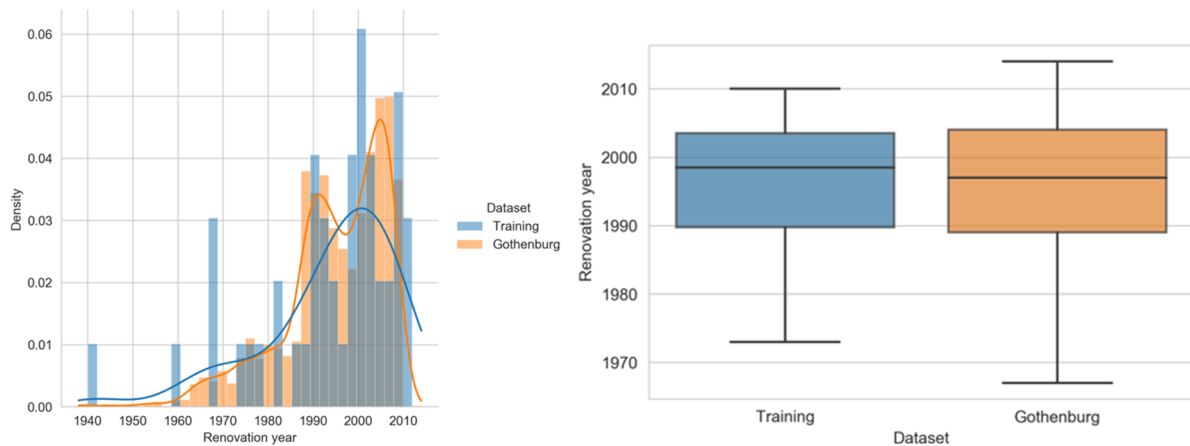
To test the feasibility of building machine learning prediction models from the training dataset for future studies, data representativeness in terms of similarity of building parameters needs to be controlled for. Representativeness of the training dataset was addressed by comparing the distribution and the building parameters with the entire building stock in the City of Gothenburg. Evaluating the variances between datasets enabled us to identify interest groups based on indicative variables, including construction year, renovation year, area range, and the number of floors. Gothenburg building stock data were retrieved from the national building registers that contained 157,301 buildings. 100,635 buildings were older than 1982, when the construction industry's use of asbestos products was banned. Building class of the observations in the comprehensive dataset was classified according to municipality data using 1–99 indexing. Aggregation of building built before 1982 showed that most of Gothenburg's old buildings were residential buildings. The rest were unspecified buildings, school buildings, industrial buildings, production buildings, and commercial buildings. The lack of registers for temporary buildings, offices, and warehouses may be categorized as unspecific buildings.

According to Figure 4, the training dataset ( $N = 336$ ) represents around 2.2% of the Gothenburg building stock constructed from 1929 to 1982 ( $N = 14,996$ ). The period was chosen for consistency as the earliest building registers traced back to 1929. The density plots were used to balance the unequal numbers of observations before comparing their distribution. Density normalization scales the bars for the individual dataset, thus the areas sum up to 1 [24]. Then, boxplots were created to illustrate the quartile of both datasets. They are used to display data variation in statistical sampling [25]. Figure 4A showed that more than half of the buildings were built between 1950 and 1970 in both datasets, corresponding to the periods of the two massive construction activities in Sweden, the People's Home [26] and the Million Homes Programme [27] eras. A majority of the renovation activities ( $\geq 70\%$ ) in both datasets took place during 1990–2005, based on Figure 4B. According to Figure 4C, living area measurements of Gothenburg buildings were between 101 and 1000 m<sup>2</sup>, whereas the area in the training dataset was either for buildings larger than 1500 m<sup>2</sup> or smaller than 100 m<sup>2</sup>. An interpretation is that buildings with environmental investigations are larger complicated buildings or smaller complementary buildings. Furthermore, the difference between the training dataset and the Gothenburg dataset in the number of floors indicates that low-level buildings were more commonly demolished or renovated for other use purposes, as presented in Figure 4D.

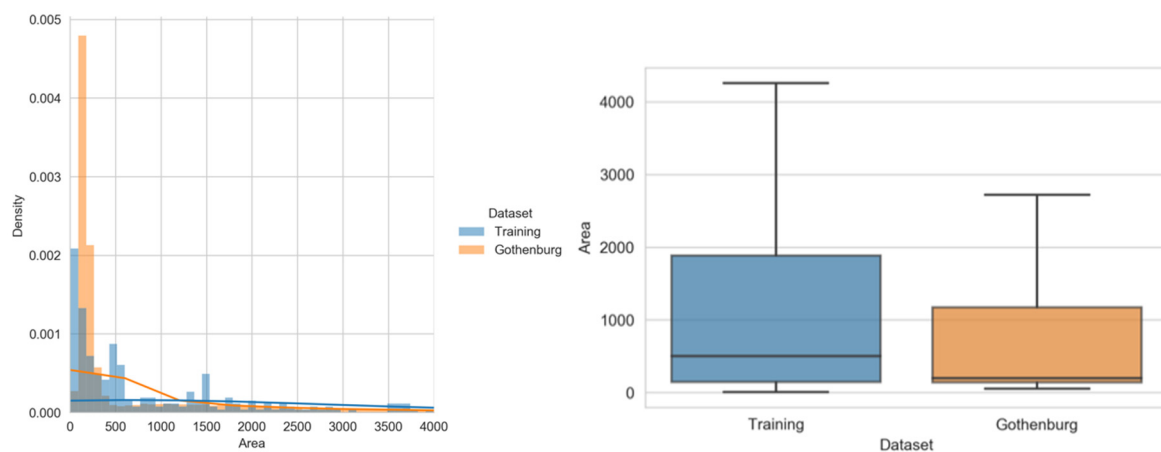
Building parameters for each building class in the Gothenburg dataset and the training dataset were compared to comprehend the building subgroups' underlying characteristics, as presented in Table 2. The distribution of the building class was calculated by dividing the number of observations in each subgroup by the total number of observations in the dataset. Buildings in the city center are often mixed residential and commercial buildings with commercial zones on the lower floors. If these two building classes in the training dataset are summed, the amount will be comparable to the Gothenburg dataset. Moreover, school buildings were more frequently renovated with the removal of hazardous materials, resulting in more environmental inventories than other building classes. One reason for oversampling could be that multiple environmental investigations were executed for the individual buildings in the school complexes, leading to an over representative data size. The differences in the mean area and the mean floor of the school buildings could also be understood from an aggregation level perspective, where Gothenburg registers took properties into account rather than buildings. Lastly, industrial buildings and production buildings accounted for a few numbers, and the buildings in the training dataset were older than the corresponding Gothenburg subgroups. A reasonable doubt could be the concern of contaminated activities. In most cases, the City Environment Administration requested environmental audits for the industrial buildings before demolition, resulting in more environmental inventories than other building classes.



(A) Construction year: The construction year of the observations in the training dataset tends to be earlier than in the Gothenburg dataset, yet the two datasets show a similar distribution pattern. The median construction year for the training dataset is around 1962, while that for the Gothenburg dataset is around 1964.

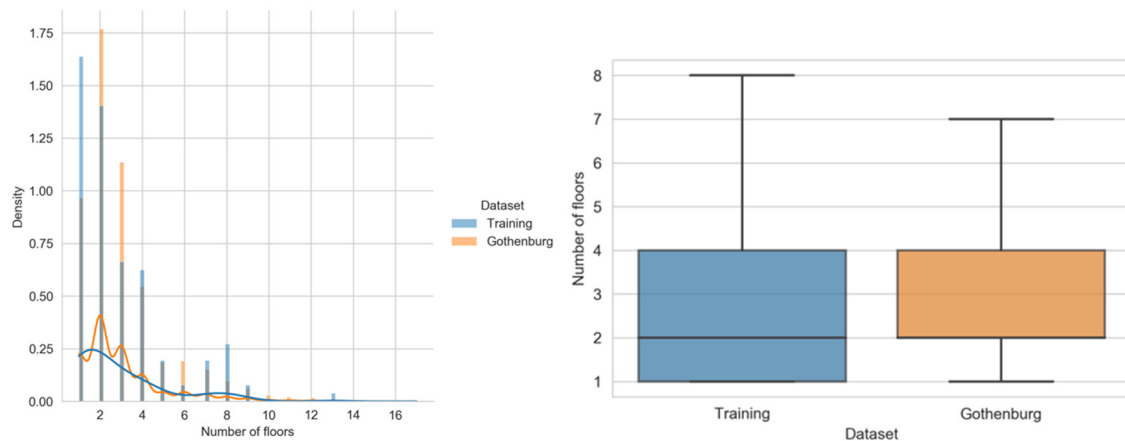


(B) Renovation year: The distribution of the renovation year between the two datasets is similar. There are two renovation peaks in the Gothenburg dataset in 1990 and 2005, while the training dataset has only one renovation peak in 2000. The median renovation year for datasets is around 1998.



(C) Area: The median area for the training dataset is around 500 m<sup>2</sup>, while that of the Gothenburg dataset is approximately 200 m<sup>2</sup>.

Figure 4. Cont.



(D) Number of floors: While most of the buildings in the training dataset are in the range of 1–4 floors, the Gothenburg dataset has the most buildings between 2 and 4 floors.

**Figure 4.** Comparison between the Gothenburg subset and the training dataset within 1929–1982 by plotting normalized density distribution (on the left) and boxplot (on the right) across (A) construction year, (B) renovation year, (C) area, and (D) the level of floors.

**Table 2.** Overview of building parameters for each building class in the Gothenburg dataset and the training dataset in 1929–1982 (numbers in bold show similar building characteristics between the two datasets).

Building Class	Dataset	Distribution [%]	Mean Year Built	Mean Year Renovated	Mean Area [m <sup>2</sup> ]	Mean Floor [N]
Single-family house	Gothenburg	64.6%	1962	1989	171	2
	Training	23.2%	1958	1980	129	1
Multifamily house	Gothenburg	24.6%	1953	<b>1999</b>	<b>3110</b>	<b>4</b>
	Training	12.2%	1961	<b>1998</b>	<b>3324</b>	<b>4</b>
Temporary dwelling	Gothenburg	NA	NA	NA	NA	NA
	Training	6.50%	1954	1971	45	1
School	Gothenburg	2.80%	<b>1966</b>	2002	2352	<b>2</b>
	Training	17.9%	<b>1968</b>	2010	1497	<b>2</b>
Office	Gothenburg	NA	NA	NA	NA	NA
	Training	7.70%	1963	1995	5146	4
Commercial building	Gothenburg	0.40%	<b>1963</b>	<b>2006</b>	6625	<b>3</b>
	Training	11.6%	<b>1962</b>	<b>2001</b>	5611	<b>3</b>
Production building	Gothenburg	0.3%	1968	1998	6410	2
	Training	5.1%	1956	2001	4728	3
Industrial building	Gothenburg	0.5%	1964	1995	<b>5005</b>	<b>2</b>
	Training	11.0%	1951	2008	<b>4680</b>	<b>2</b>
Warehouse	Gothenburg	NA	NA	NA	NA	NA
	Training	2.7%	1968	NA	2688	1
Other/Infrastructure	Gothenburg	1.7%	1962	2001	4262	3
	Training	2.1%	1964	NA	2822	2
Unlabeled *	Gothenburg	5.0%	NA	NA	NA	NA

\* The unlabeled building class implied that the buildings were assigned to “unspecific” building types in the national building registers. Thus, it is not possible to determine its building class in the study.

### 3.3. Statistical Operations

In Table 3, the positive detection rates and the amount of missing data in each inventory type were described. By reviewing the amount of missing data (Appendix B), a better understanding of the usefulness of detection results could be developed. A positive detection rate showed the detection of hazardous materials, which was calculated by dividing the number of positive results by the total number of observations, excluding the missing data. The results showed that different inventory types had their strengths in detecting hazardous substances and materials. Large numbers of missing data ( $\geq 90\%$ )

were presented in control plans and demolition plans as simple inventories lacked information about hazardous materials. They often only show the positive detection of one sought construction part; thus, these environmental inventories were mainly conducted to remove the specific asbestos-containing material. The inclusion of simple inventories into a more extensive dataset could lead to the risk of a skewed dataset. Detailed inventories, on the other hand, contained comprehensive detection records of hazardous materials. However, the detection records of certain building materials were not always available as they were not included in the protocol, such as asbestos-containing switchboards and joints, PCB-containing door closers, and cables.

**Table 3.** Positive detection rates and missing data of hazardous materials in each environmental inventory type (numbers in bold contain more than 30 observations).

Substance and Material	Inventory							
	Detailed Report N = 195		Protocol N = 88		Simple Control Plan N = 42		Demolition Plan N = 77	
	Rate	NA	Rate	NA	Rate	NA	Rate	NA
<b>Asbestos</b>	<b>0.84</b>	<b>6%</b>	<b>0.51</b>	<b>9%</b>	<b>0.47</b>	<b>14%</b>	<b>0.70</b>	<b>27%</b>
Pipe insulation	<b>0.56</b>	<b>34%</b>	<b>0.19</b>	<b>23%</b>	0.78	79%	0.75	90%
Valves	<b>0.44</b>	<b>79%</b>	<b>0.10</b>	<b>55%</b>	NA	100%	0.67	96%
Door/windows insulation	<b>0.59</b>	<b>34%</b>	<b>0.03</b>	<b>31%</b>	NA	100%	0.50	97%
Cement panel board	<b>0.61</b>	<b>67%</b>	<b>0.22</b>	<b>48%</b>	0.67	93%	0.71	91%
Tile/clinker	<b>0.39</b>	<b>27%</b>	<b>0.17</b>	<b>20%</b>	1.00	93%	0.60	94%
Carpet glue	<b>0.31</b>	<b>42%</b>	<b>0.11</b>	<b>30%</b>	NA	100%	0.40	94%
Floor mat	<b>0.40</b>	<b>31%</b>	0.43	92%	1.00	95%	0.75	95%
Ventilation channel	<b>0.55</b>	<b>47%</b>	0.83	93%	1.00	98%	1.00	94%
Switchboard	<b>0.17</b>	<b>66%</b>	NA	100%	NA	100%	NA	100%
Joint	<b>0.40</b>	<b>70%</b>	NA	100%	NA	100%	NA	100%
Others	<b>0.75</b>	<b>69%</b>	0.50	93%	0.25	90%	0.50	92%
<b>PCB</b>	<b>0.63</b>	<b>8%</b>	<b>0.49</b>	<b>10%</b>	<b>0.19</b>	<b>26%</b>	<b>0.51</b>	<b>52%</b>
Joint/sealant	<b>0.66</b>	<b>38%</b>	<b>0.41</b>	<b>23%</b>	0.18	74%	0.50	92%
Insulation windows	<b>0.20</b>	<b>43%</b>	<b>0.16</b>	<b>23%</b>	0.50	95%	0.67	96%
Capacitors in lamp/burner	<b>0.66</b>	<b>41%</b>	<b>0.44</b>	<b>12%</b>	0.11	79%	0.83	84%
Acrylic flooring	<b>0.06</b>	<b>58%</b>	<b>0</b>	<b>23%</b>	0.10	76%	0	97%
Door closer	0.72	91%	NA	100%	NA	100%	NA	100%
Cable with PCB-oil	0.28	91%	NA	100%	NA	100%	NA	100%
Others	0.52	87%	0	97%	0.33	86%	0.50	97%
<b>CFC</b>	<b>0.79</b>	<b>34%</b>	<b>0.60</b>	<b>12%</b>	<b>0.62</b>	<b>19%</b>	<b>0.50</b>	<b>56%</b>
Fridge/freezer	<b>0.79</b>	<b>50%</b>	<b>0.39</b>	<b>15%</b>	0.76	50%	0.81	79%
Building insulation	<b>0.21</b>	<b>63%</b>	<b>0.33</b>	<b>20%</b>	0.40	76%	0.33	96%
Cooling unit	<b>0.53</b>	<b>54%</b>	<b>0.19</b>	<b>24%</b>	0.40	88%	0.75	95%
Rolling gate	0.40	92%	NA	98%	NA	100%	NA	100%
Others	0.36	94%	0.67	97%	0.40	88%	0.67	96%
<b>Mercury</b>	<b>0.99</b>	<b>11%</b>	<b>0.72</b>	<b>3%</b>	<b>0.55</b>	<b>26%</b>	<b>0.76</b>	<b>35%</b>
Lighting tube	<b>0.99</b>	<b>16%</b>	<b>0.61</b>	<b>6%</b>	0.74	55%	<b>0.97</b>	<b>56%</b>
Relay/switch	<b>0.19</b>	<b>73%</b>	<b>0.26</b>	<b>22%</b>	0.69	69%	0.71	91%
Level monitor/sensor	<b>0.42</b>	<b>66%</b>	<b>0.04</b>	<b>23%</b>	NA	100%	0.67	96%
Thermometer	<b>0.45</b>	<b>59%</b>	<b>0.09</b>	<b>22%</b>	1.00	93%	0.80	94%
Thermostat	<b>0.16</b>	<b>77%</b>	<b>0.13</b>	<b>22%</b>	NA	100%	NA	99%
Water lock/drain line	<b>0.08</b>	<b>73%</b>	<b>0.17</b>	<b>22%</b>	NA	100%	0.50	97%
Low energy lamp	<b>0.97</b>	<b>67%</b>	0.83	93%	1.00	98%	1.00	92%
Doorbell	NA	96%	0.50	98%	NA	100%	1.00	99%
Others	<b>0.97</b>	<b>68%</b>	1.00	88%	0.50	90%	0.89	88%

Note: Positive detection rate = Number of Positives/(Total number of observations—Number of NA).

Asbestos showed high positive detection rates in reports (0.84), demolition plans (0.70), and protocols (0.51), especially in pipe insulation, door or window insulation, cement panel boards, and ventilation channels. PCB generally had lower positive detection rates compared with asbestos in building materials. The results from the reports and the protocols indicated a higher risk of encountering PCB-containing joints/sealants and capacitors in fluorescent lamps/burners than other potential PCB-containing materials. Furthermore, CFC and mercury occurred frequently in buildings with positive detection rates higher than 0.50 in all inventories. CFC-containing materials were primarily located in freezers/fridges. However, the positive detection rates for building insulations and cooling units were not in agreement between reports and protocols. Investigations of CFC-containing materials in reports showed a high positive detection rate at refrigerations,

whereas building insulation has a higher positive detection rate in protocols. The positive detection rate of mercury was the highest across inventory types. Mercury-containing materials were primarily found in lighting tubes. Positive detection rates of mercury-containing level monitors or sensors and thermometers were also high in reports, while investigations of protocols reported a high positive detection rate at relay or switches. The significant differences in the positive detection rates and the reported frequency across inventory types can be attributed to multiple reasons, including the building features of each subgroup's observations, purposes for environmental audits, and investigators' experience levels. Hence, further data clustering was performed to analyze positive detection rates and missing data in the subgroups of building classes (Table 4), construction year range (Appendix C), and area range (Appendix D).

**Table 4.** The positive detection rates and missing data of hazardous materials in building classes (numbers in bold contain more than 30 observations).

Substance and Material	Building Class									
	Multifamily House N = 46		School N = 66		Office N = 30		Commercial Building N = 48		Industrial Building N = 39	
	Rate	NA	Rate	NA	Rate	NA	Rate	NA	Rate	NA
<b>Asbestos</b>	<b>0.93</b>	4%	<b>0.89</b>	5%	0.71	7%	<b>0.74</b>	2%	<b>0.61</b>	5%
Pipe insulation	<b>0.81</b>	<b>30%</b>	<b>0.24</b>	<b>36%</b>	0.55	33%	0.56	44%	0.62	46%
Valves	0.59	63%	0.25	82%	0.14	77%	0.22	81%	0.25	79%
Door/windows insulation	0.62	43%	<b>0.41</b>	<b>44%</b>	0.52	30%	0.56	44%	0.57	41%
Cement panel board	0.78	50%	0.35	74%	0.38	57%	0.47	65%	0.33	77%
Tile/clinker	<b>0.63</b>	<b>35%</b>	<b>0.40</b>	<b>21%</b>	0.39	23%	<b>0.29</b>	<b>35%</b>	0.23	44%
Carpet glue	0.54	48%	<b>0.23</b>	<b>33%</b>	0.46	57%	0.17	40%	0.22	54%
Floor mat	0.71	54%	<b>0.35</b>	<b>27%</b>	0.44	40%	0.26	44%	0.54	67%
Ventilation channel	0.88	65%	<b>0.57</b>	<b>44%</b>	0.43	53%	0.68	54%	0.31	67%
Switchboard	0.33	93%	<b>0.06</b>	<b>52%</b>	0.20	83%	0.10	79%	0.50	74%
Joint	0.87	67%	0.31	80%	0.20	67%	0.17	75%	0.50	90%
Others	0.77	72%	0.70	70%	0.62	73%	0.78	81%	0.80	74%
<b>PCB</b>	<b>0.57</b>	<b>24%</b>	<b>0.61</b>	<b>8%</b>	0.70	10%	0.57	17%	<b>0.65</b>	<b>8%</b>
Joint/sealant	<b>0.57</b>	<b>35%</b>	<b>0.59</b>	<b>30%</b>	0.78	40%	0.54	42%	0.65	49%
Insulation windows	0.41	63%	<b>0.18</b>	<b>41%</b>	0.20	33%	0.19	46%	0.22	54%
Capacitors in lamp/burner	0.58	59%	<b>0.56</b>	<b>35%</b>	0.55	27%	0.52	52%	0.86	28%
Acrylic flooring	0.06	65%	<b>0.06</b>	<b>50%</b>	NA	60%	NA	56%	0.06	59%
Door closer	NA	100%	0.57	89%	0.50	93%	0.75	92%	1.00	97%
Cable with PCB-oil	NA	98%	NA	92%	0.20	83%	0.60	90%	1.00	97%
Others	1.00	91%	0.50	91%	0.33	80%	0.17	88%	0.67	92%
<b>CFC</b>	0.62	48%	<b>0.70</b>	<b>35%</b>	0.74	23%	0.83	38%	<b>0.74</b>	<b>15%</b>
Fridge/freezer	0.67	54%	<b>0.68</b>	<b>42%</b>	0.45	63%	0.74	52%	0.74	41%
Building insulation	0.36	70%	<b>0.19</b>	<b>53%</b>	0.23	57%	0.30	58%	0.20	62%
Cooling unit	0.38	65%	<b>0.28</b>	<b>52%</b>	0.74	37%	0.47	60%	0.45	49%
Rolling gate	NA	100%	0.25	94%	NA	97%	NA	96%	0.60	87%
Others	1.00	98%	NA	95%	1.00	97%	NA	96%	0.50	95%
<b>Mercury</b>	<b>0.86</b>	<b>20%</b>	<b>0.98</b>	<b>11%</b>	<b>0.90</b>	<b>0%</b>	<b>0.95</b>	<b>19%</b>	<b>0.95</b>	<b>0%</b>
Lighting tube	0.79	37%	<b>0.98</b>	<b>12%</b>	<b>0.87</b>	<b>0%</b>	<b>1.00</b>	<b>25%</b>	<b>0.97</b>	<b>5%</b>
Relay/switch	0.27	67%	0.04	59%	0.46	57%	0.27	77%	0.42	69%
Level monitor/sensor	0.10	78%	<b>0.37</b>	<b>47%</b>	0.27	50%	0.31	73%	0.57	64%
Thermometer	0.56	61%	<b>0.25</b>	<b>45%</b>	0.44	47%	0.47	69%	0.38	67%
Thermostat	0.22	80%	NA	59%	0.38	57%	NA	83%	0.50	79%
Water lock/drain line	0.10	78%	NA	62%	0.36	63%	NA	71%	0.11	77%
Low energy lamp	1.00	91%	1.00	56%	0.89	70%	1.00	81%	1.00	82%
Doorbell	NA	100%	NA	95%	0.33	90%	NA	98%	NA	100%
Others	1.00	91%	1.00	67%	0.92	60%	1.00	81%	1.00	64%

Based on the conclusion of Figure 2, building classes with high data quality and adequate data size were selected for further data analysis. These included multifamily houses, schools, offices, commercial buildings, and industrial buildings. Table 4 describes the data size and the positive detection rate of hazardous substances and materials for each building class. A threshold value of 30 valid observations was set to enhance the certainty of the results, underlined in Table 4. Asbestos was identified predominately in multifamily houses with a positive detection rate of 0.93. In multifamily houses, asbestos-containing ventilation channels, joints, and pipe insulation were encountered. For school buildings and commercial buildings, ventilation channels contained a higher risk of asbestos. In contrast, asbestos-containing pipe insulations and door or window insulation were common in offices and industrial buildings. PCB positive detection rates were generally lower than

asbestos across building classes. Office buildings had the highest PCB positive detection rate (0.70), mainly from PCB-containing joints/sealants. On the contrary, industrial buildings had an outstanding PCB detection value in capacitors in lamps/burners. For the rest of the building classes, the positive detection rates of PCB-containing joints/sealants and capacitors in lamps/burners showed similar patterns.

Concerning the exposure to environmental hazardous substances, mercury had a more frequent presence than CFC. The high positive detection rate of mercury was primarily due to the massive adoption of mercury-containing lighting tubes. Mercury-containing thermometers were also commonly used in multifamily houses and office buildings. On the other hand, CFC was found frequently in commercial buildings. CFC-containing fridges or freezers were the primary reason for high positive detection rates in multifamily houses, school buildings, commercial buildings, and industrial buildings. CFC-containing cooling unit was the main attribute in office buildings. Overall, the patterns of detecting hazardous materials in each building class appear to be reasonable considering the usage of the building and their average construction year. Our results are to some extent in agreement with the experience-based expert knowledge regarding frequent occurrences of hazardous materials in certain building classes. However, it has been challenging to cross-compare the positive detection rate of a specific building material among other building classes given varied data sizes. To determine the generalization potential of the results in relation to the regional building stock, incorporating more valid data of the studied buildings classes into the subsequent analysis will be essential.

### 3.4. Cross-Validation Matrix

The cross-validation matrix was developed as a tool to evaluate the results' reliability. It helps set a boundary for hypothesis formulation by considering the data quality and quantity on hand. Table 5 presents the overall assessment scores for each building class calculated by the cross-validation matrix. The assessment score for the individual hazardous substance and material differed significantly among building class subgroups, indicating high variation of data quality and data size. The assessment score in 0 or NA implies that the investigation records had low reference values owing to their data source primarily from simple inventories or have an insufficient data size. Despite that the total data size was high, high-quality detection records from reports or protocols were few. The lack of detection records in hazardous material levels led to the high number of missing values, thus the overall assessment score ranked low. The data sizes of temporary dwellings, production buildings, warehouses, and others were inadequate or lacking investigation records for certain building materials. Hence, their assessment scores were the lowest and should be excluded in the subsequent machine learning modeling.

**Table 5.** The overview of the assessment scores for each building class based on data quality and data size (numbers in bold are higher than 80 scores).

Substance and Material	Class									
	C1 *	C2 *	C3 *	C4 *	C5 *	C6 *	C7 *	C8 *	C9 *	C10 *
	N = 102	N = 46	N = 24	N = 66	N = 30	N = 48	N = 23	N = 39	N = 10	N = 12
Asbestos	55	79	32	<b>94</b>	46	<b>86</b>	40	<b>84</b>	0	0
Pipe insulation	67	<b>88</b>	0	<b>99</b>	48	47	0	44	0	0
Valves	36	44	0	0	0	0	0	0	0	NA
Door/windows insulation	75	46	0	<b>99</b>	48	48	0	46	0	0
Cement panel board	34	45	0	50	0	42	0	0	0	0
Tile/clinker	73	<b>91</b>	0	<b>98</b>	48	<b>97</b>	0	44	0	0
Carpet glue	74	45	0	<b>98</b>	0	47	0	45	0	0
Floor mat	0	48	0	<b>98</b>	48	50	0	0	0	0
Ventilation channel	0	44	NA	<b>100</b>	0	47	0	0	0	0
Switchboard	0	0	NA	<b>100</b>	0	0	0	0	NA	0
Joint	0	50	NA	0	0	0	0	0	0	NA
Others	0	0	0	48	0	0	0	0	0	0



Table 5. Cont.

Substance and Material	Class									
	C1 *	C2 *	C3 *	C4 *	C5 *	C6 *	C7 *	C8 *	C9 *	C10 *
	N = 102	N = 46	N = 24	N = 66	N = 30	N = 48	N = 23	N = 39	N = 10	N = 12
<b>PCB</b>	58	84	32	92	47	92	42	90	0	0
Joint/sealant	71	90	0	96	47	48	0	42	0	0
Insulation windows	74	44	0	97	48	48	0	45	0	0
Capacitors in lamp/burner	70	42	35	96	46	46	0	44	0	0
Acrylic flooring	71	40	0	98	0	48	0	42	0	0
Door closer	0	NA	0	0	0	0	0	0	NA	NA
Cable with PCB-oil	0	0	NA	0	0	0	NA	0	NA	NA
Others	0	0	0	0	0	0	0	0	0	NA
<b>CFC</b>	58	39	34	94	48	88	0	85	0	0
Fridge/freezer	64	40	0	95	0	46	0	42	0	0
Building insulation	71	0	0	99	0	48	0	41	0	0
Cooling unit	71	38	0	99	48	48	0	45	0	0
Rolling gate	0	NA	0	0	0	0	0	0	0	NA
Others	0	0	0	0	0	0	0	0	0	0
<b>Mercury</b>	56	81	32	96	92	83	40	85	0	0
Lighting tube	62	39	34	97	92	85	42	86	0	0
Relay/switch	67	36	0	49	0	0	0	0	0	0
Level monitor/sensor	74	0	0	99	46	0	0	0	0	0
Thermometer	73	42	0	97	46	48	0	0	0	0
Thermostat	74	0	0	49	0	0	0	0	0	0
Water lock/drain line	73	0	0	49	0	0	0	0	0	0
Low energy lamp	0	0	NA	48	0	0	0	0	0	0
Doorbell	0	NA	NA	0	0	0	NA	NA	0	NA
Others	0	0	0	47	0	0	0	0	0	0

\* Building class C1: single-family house; C2: multifamily house; C3: temporary dwelling; C4: school; C5: office; C6: commercial building; C7: production building; C8: industrial building; C9: warehouse; C10: other/infrastructure.

Based on the cross-validation matrix, a ranking coupling building class and hazardous materials in descending order, were presented in Table 6. The ranking of the cross-validation results can not only guide further data collection procedures, but also show the limitations of the current environmental audits. The school buildings reached the highest score in most hazardous substances and materials investigations, implying that their detection records were reliable. The fact may be because their inventory data come mainly from reports with extensive investigation records. For commercial buildings and industrial buildings, the assessment scores on the hazardous substance level, asbestos-containing tile or clinker, and mercury-containing lighting tubes were high as well. Yet, the rest of the hazardous materials in these two groups had low scores. While PCB, PCB-containing joints, mercury, and asbestos-containing pipe insulation and tiles or clinkers had high reference values in multifamily houses, only the detection records of mercury and mercury-containing lighting tubes were high in office buildings. The fact that PCB was better surveyed in multifamily houses than other building classes could be explained by their conventional construction method of concrete elements and bricks with many sealants. Therefore, their detection results of the potential PCB-containing joints or sealants and asbestos-containing tiles or clinkers showed fewer missing values in the investigation records.

**Table 6.** Ranking the hazardous substance and material detection records in building class based on the cross-validation results; only the assessment scores higher than 80 were listed.

Rank	Class	Substance	Hazardous Material	Score	NA	N
1	4—School	Asbestos	Ventilation channel	100	44%	66
1	4—School	Asbestos	Switchboard	100	52%	66
2	4—School	Asbestos	Pipe insulation	99	36%	66
2	4—School	Asbestos	Door/windows insulation	99	44%	66
2	4—School	CFC	CFC—Building insulation	99	53%	66
2	4—School	CFC	CFC—Cooling unit	99	52%	66
2	4—School	Mercury	Level monitor/sensor	99	47%	66
3	4—School	Asbestos	Tile/clinker	98	21%	66
3	4—School	Asbestos	Carpet glue	98	33%	66
3	4—School	Asbestos	Floor mat	98	27%	66
3	4—School	PCB	Acrylic flooring	98	50%	66
4	4—School	PCB	Insulation windows	97	41%	66
4	4—School	Mercury	Lighting tube	97	12%	66
4	4—School	Mercury	Thermometer	97	45%	66
4	6—Commercial building	Asbestos	Tile/clinker	97	35%	48
5	4—School	PCB	Joint/sealant	96	30%	66
5	4—School	PCB	Capacitors in lamp	96	35%	66
5	4—School	Mercury		96	11%	66
6	4—School	CFC	Fridge/freezer	95	42%	66
7	4—School	Asbestos		94	5%	66
7	4—School	CFC		94	35%	66
8	4—School	PCB		92	8%	66
8	6—Commercial building	PCB		92	17%	48
8	5—Office	Mercury		92	0%	30
8	5—Office	Mercury	Lighting tube	92	0%	30
9	2—Multifamily house	Asbestos	Tile/clinker	91	35%	46
10	8—Industrial building	PCB		90	8%	39
10	2—Multifamily house	PCB	Joint/sealant	90	35%	46
11	6—Commercial building	CFC		88	38%	48
11	2—Multifamily house	Asbestos	Pipe insulation	88	30%	46
12	6—Commercial building	Asbestos		86	2%	48
12	8—Industrial building	Mercury	Lighting tube	86	5%	39
13	8—Industrial building	CFC		85	15%	39
13	8—Industrial building	Mercury		85	0%	39
13	6—Commercial building	Mercury	Lighting tube	85	25%	48
14	8—Industrial building	Asbestos		84	5%	39
14	2—Multifamily house	PCB		84	24%	46
15	6—Commercial building	Mercury		83	19%	48
16	2—Multifamily house	Mercury		81	20%	46

### 3.5. Method Replicability

Nowadays, obligatory environmental audits and open databases for building registers, the method's key data input, are already available in many EU countries. For example, the Waste Register database in Estonia, the Integrated Waste Management System, and the Asbestos Database in Poland [10]. Thus, these countries have developed an established waste management practice and are in the position to adopt the proposed approach to estimate the residual hazardous material stock. As for the countries with the limited implementation of environmental audits concerning building size and function, as described in the introduction, it will be beneficial to create a harmonized environmental protocol for auditing hazardous materials in an online database [11]. Having a uniform digital dataset template can reduce the risk of information asymmetry and save processing time for data compiling, making data queries for environmental information much more effective.

Building registered data have become an essential source to describe building stock. However, data uncertainties are required to be addressed to assure accurate analytic results [23]. Previous efforts validated the EPC databases using stepwise regression models [23] and deploying a data quality assurance method [28]. Compared with the broad applications of EPC data [28], the environmental audits data remain relatively unexplored. Moreover, a large number of uninvestigated building components and undetermined

assessment results, as well as the varied extent of environmental investigation execution between building classes, fail to be exposed. The study systematically examines the quality and content of environmental inventories, and based on that, evaluates their usability to enrich the building databases by adding specific environmental information. Referencing the EU data validation levels [29], a standard procedure to transform the field data into an organized, reliable dataset was exhibited. By showing the limitations and the possibilities of the environmental audits data, we hope to encourage more research in the application domain of safe construction and waste management.

The metabolism of the residual hazardous materials in the building stock is slow and the risk of exposure exists in every stage of the building life cycle. Previous literature showed that exceeding PCB concentration in the indoor air from volatilization during the building operation phase [30] and the airborne asbestos emission from the emergency demolitions [3] place an indispensable requirement for a long-term preventive maintenance strategy. To facilitate the abatement policy of in situ hazardous materials, extensive studies were conducted at the urban inventory level and the specific building class level. A stocks and flows model for asbestos was developed to determine common types of asbestos-containing materials [31]. Surprisingly, it was found that cement sheeting and waterpipe accounted for 90% of the asbestos consumption in Australia. Another study in the Australian residential buildings also showed high positive detection rates of asbestos backing board to the electrical meter boxes and asbestos eaves [2]. Overall, asbestos-containing materials presented in 82.3% of the sampled houses. Similarly, asbestos-containing materials were found in around 95% of the abandoned residential homes in Detroit, especially in flooring, roofing, siding, and duct insulation [6]. These findings aligned with our results as the positive detection rate of asbestos in multifamily houses was 93%, and the high-risk materials were pipe insulation and cement panel board.

The investigations of building-related PCBs also showed a similar trend. An extensive survey in Switzerland verified that 48% of the buildings built between 1950 and 1980 contained PCBs [32]. In the same period, 46% of the school buildings in the USA were constructed [4]. Implementing the engineering controls to mitigate PCB diffusion from joint sealants and building caulk in American school buildings indicated a strong need for immediate actions [33,34]. In a citywide building sampling study, the positive detection rate of PCB-containing sealants was found at 14% in Toronto [30]. In particular, a high density of PCBs existed in commercial and electricity-intensive buildings [35]. Compared with brick or glass buildings, concrete buildings had a higher tendency to be contaminated by PCBs [30]. The results from our study agree with their findings. High detection rates of PCB-containing joints or sealants were noticed in concrete-built office and industrial buildings. Overall, the frequent presence of asbestos and PCB-containing materials worldwide reveals a necessity to develop an effective identification approach for general buildings. The method proposed in the study presents a data-driven solution to evaluate the risk of encountering hazardous materials in a high detail.

This study confirmed the potential to assess hazardous materials in regional building stock by using multiple registered records. Combining the environmental information from numerous registered data sources made it possible to systematically estimate the risk of hazardous substance occurrence in the building stock [20]. High data completeness of the detailed inventories, such as protocols and reports, enables a thorough analysis of hazardous material types in various building classes. Although the final output is highly dependent on data availability and quality, it is a rather cost-effective approach to trace the existing hazardous materials. In addition, the method's generability can be replicated in other countries and help prioritize decontamination plans before demolition or extensive retrofit.

#### 4. Conclusions

This paper studied the feasibility of tracing hazardous materials in the existing building stock based on multi-sourcing registered records. The opportunities and challenges

of using environmental audits to guide the risk assessment of hazardous materials were explored. By associating national building registers and environmental inventories, a training dataset was created to verify the experienced-based expert knowledge. Around 65% of the training dataset's observations comprised reports and protocols with high data granularity. Asbestos was the most frequently investigated substance, following by mercury, PCB, and CFC. The extent of environmental investigations of each building class varies, depending on building complexity and ownership. Most of the observations in reports were schools, commercial buildings, industrial buildings, multifamily houses, and office buildings, whereas in protocols, control plans, and demolition plans, single-family houses were most common. By validating data size and quality, building groups appropriate for the statistical operations were determined. Furthermore, comparing the distribution of building parameters between the training dataset and the Gothenburg dataset helps to understand the data representativeness, which involves the viability of constructing prediction models from the training dataset. The risk of encountering hazardous materials was assessed by evaluating the positive detection rates and the amount of missing data. Through clustering data subgroups such as inventory types, building classes, construction year, and area range, different perspectives for evaluating positive detection rates and missing data were presented. The results indicated high positive detection rates of asbestos-containing materials in multifamily buildings and prevalent PCB-containing materials in industrial buildings. High number of missing values for hazardous materials in single-family houses, production buildings, and warehouses were highlighted to improve the current environmental investigation practices.

The explorative approach of delineating quality environmental data demonstrates a general workflow for studying in situ hazardous building materials' management. The novel method is cost-effective in identifying general occurrence patterns of hazardous building materials and can be used to complement traditional environmental investigations. The findings from the cross-validation matrix showed that the potential data subgroups for machine learning modelling were school buildings, commercial buildings, multifamily houses, and industrial buildings at hazardous substance and material levels. Future research is planned to include more observations from the abovementioned building classes to increase the prediction accuracy and conduct cross-verification when constructing machine learning models. The developed data-driven method and the structure of the training dataset proposed in the study are replicable in the countries accessible to the environmental-audit records and general building information.

**Author Contributions:** K.M. acquired the research fund and was responsible for project management. M.M. initiated the contact with the Gothenburg City Archives for arranging the data collection process, requested building registers from the authorities, helped structure the tables in the method and results and discussions sections, and revised the manuscript. T.J. assisted with merging building registers. P.-Y.W. screened the relevant documents, retrieved the information, compiled them into a dataset, drafted the manuscript and figures, and conducted the manuscript revision. All authors participated in cross-validation workshops, results discussion, and manuscript revision. C.S. conceived the idea of the cross-validation matrix in the results and discussions section. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Swedish Foundation for Strategic Research (SSF), grant number FID18-0021.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The environmental inventories collected in the study are confidential and regulated by the Gothenburg city archive. The national building registers acquired from different authorities were requested for the specific research purpose. Therefore, they are not available online or in the MDPI Research database.

**Acknowledgments:** The work is part of the research project Prediction of hazardous materials in buildings using AI and is supported by RISE Research Institutes of Sweden. Special thanks are sent to Theresa Salhammar and Anna-Sara Berg in the Gothenburg City Archives for the support of searching environmental inventories in the building permit documents.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

The metadata of the Gothenburg dataset from Swedish EPC, Swedish Real Estate Taxation register, and Municipal cadastral register.

**Table A1.** Swedish Energy Performance Certificate (EPC) data overview.

Value Category	Data Specification	Measurement Type
1. Matching, keys, and sorting	National real estate number and index, Address, EPC index	
2. Building characteristics	Building age Complexity Shared walls with other buildings Recognition of heritage value	Scale variable [Year] Binary [Complex, non-complex] Ordinal [Detached, semi-attached, attached] Nominal
3. Building usage	National registration of building usage type code. Detailed usage of building	Nominal Share of building used for the 12 most common types
4. Building area	Heated floor area Floors Stair cases Number of apartments Floors below ground Heated garage space	Scale [m <sup>2</sup> ] Ordinal Ordinal Ordinal Ordinal Scale [m <sup>2</sup> ]
5. Heating	Energy usage for heating divided in 13 energy sources Tic box for measurement type Period of measurement	Scales [kWh/year] Nominal [Measured, Distributed] Interval [Year and Month]
6. Household electricity and water	Cooling energy usage Tap water heat usage Electricity usage divided in: domestic, shared, and non-domestic usage	All measurements in the category include Scale [kWh/year] and Nominal [Measured, Distributed]
7. Ventilation	Ventilation type Ventilation control conducted	Nominal [Exhaust, balanced, balanced with heat exchanger, Exhaust with heat pump, natural ventilation] Nominal [Yes, No, Partially]
8. Recommended energy usage reducing measures	Tic box for 28 common energy usage reducing measures Estimated decreased energy usage Estimated cost per saved kWh	Nominal Scale [kWh/year] Scale [SEK/kWh]

**Table A2.** Swedish Real Estate Taxation register overview.

Value Category	Data Specification	Measurement Type
1. Matching, keys, and sorting	Coordinates, National real estate number and index, Address	
2. Building ages	Building age Value year (Swedish tax agency's proxy for renovation state) Latest renovation year	Scale variable [Year] Scale variable [Year] Scale variable [Year]
3. Building taxation value	Building taxation value Rent level	Scale variable [SEK] Scale variable [SEK]
4. Building area	Building size (BOA)	Scales [m <sup>2</sup> ]

**Table A3.** Municipal cadastral register overview.

Value Category	Data Specification	Measurement Type
1. Matching, keys, and sorting	Coordinates, National real estate number and index, Address	
2. Building ages	Building age	Scale variable [Year]
3. Building usage	Simple classification	Nominal [1–7]
	Detailed classification	Nominal [1–99]
4. Building area	Building size (BOA)	Scales [m <sup>2</sup> ]
5. Building status	Changes Demolished or burned down	Nominal [Building permit applied for, Under construction, Existing building, Changed information] Binary
6. Legal status	Land ownership	Binary
	Legal land restrictions	Binary

## Appendix B

**Table A4.** Percentage of missing data of the hazardous materials in each building class.

Substance and Material	Class									
	C1 * N = 102	C2 * N = 46	C3 * N = 24	C4 * N = 66	C5 * N = 30	C6 * N = 48	C7 * N = 23	C8 * N = 39	C9 * N = 10	C10 * N = 12
<b>Asbestos</b>	19%	4%	17%	5%	7%	2%	22%	5%	30%	33%
Pipe insulation	58%	30%	62%	36%	33%	44%	52%	46%	80%	58%
Valves	80%	63%	79%	82%	77%	81%	87%	79%	80%	100%
Door/windows insulation	70%	43%	62%	44%	30%	44%	57%	41%	70%	75%
Cement panel	75%	50%	67%	74%	57%	65%	87%	77%	70%	92%
Tile/clinker	61%	35%	67%	21%	23%	35%	70%	44%	80%	67%
Carpet glue	68%	48%	67%	33%	57%	40%	83%	54%	80%	67%
Floor mat	91%	54%	96%	27%	40%	44%	70%	67%	90%	75%
Ventilation channel	93%	65%	100%	44%	53%	54%	83%	67%	90%	83%
Switchboard	97%	93%	100%	52%	83%	79%	91%	74%	100%	92%
Joint	99%	67%	100%	80%	67%	75%	91%	90%	90%	100%
Others	91%	72%	96%	70%	73%	81%	83%	74%	90%	92%
<b>PCB</b>	34%	24%	21%	8%	10%	17%	4%	8%	10%	25%
Joint/sealant	64%	35%	50%	30%	40%	42%	70%	49%	70%	67%
Insulation windows	67%	63%	58%	41%	33%	46%	48%	54%	80%	67%
Capacitors in lamp/burner	59%	59%	38%	35%	27%	52%	43%	28%	70%	67%
Acrylic flooring	64%	65%	54%	50%	60%	56%	74%	59%	80%	50%
Door closer	99%	100%	96%	89%	93%	92%	91%	97%	100%	100%
Cable with PCB-oil	99%	98%	100%	92%	83%	90%	100%	97%	100%	100%
Others	93%	91%	96%	91%	80%	88%	91%	92%	90%	100%
<b>CFC</b>	25%	48%	21%	35%	23%	38%	48%	15%	40%	50%
Fridge/freezer	44%	54%	42%	42%	63%	52%	52%	41%	70%	50%
Building insulation	62%	70%	46%	53%	57%	58%	87%	62%	80%	67%
Cooling unit	66%	65%	54%	52%	37%	60%	74%	49%	80%	67%
Rolling gate	98%	100%	96%	94%	97%	96%	96%	87%	90%	100%
Others	93%	98%	96%	95%	97%	96%	91%	95%	90%	92%
<b>Mercury</b>	26%	20%	12%	11%	0%	19%	13%	0%	10%	25%
Lighting tube	45%	37%	29%	12%	0%	25%	17%	5%	20%	25%
Relay/switch	61%	67%	50%	59%	57%	77%	78%	69%	80%	67%
Level monitor/sensor	71%	78%	54%	47%	50%	73%	83%	64%	80%	83%
Thermometer	68%	61%	50%	45%	47%	69%	70%	67%	70%	75%
Thermostat	70%	80%	54%	59%	57%	83%	87%	79%	80%	83%
Water lock/drain line	69%	78%	50%	62%	63%	71%	78%	77%	80%	83%
Low energy lamp	94%	91%	100%	56%	70%	81%	70%	82%	50%	83%
Doorbell	97%	100%	100%	95%	90%	98%	100%	100%	90%	100%
Others	92%	91%	75%	67%	60%	81%	74%	64%	70%	83%

\* Building class C1: Single-family house; C2: Multifamily house; C3: Temporary dwelling; C4: School; C5: Office; C6: Commercial building; C7: Production building; C8: Industrial building; C9: Warehouse; C10: Other/Infrastructure.



## Appendix C

**Table A5.** The positive detection rates and missing data of hazardous materials in construction year clusters (numbers in bold contain than 30 observations).

Substance and Material	Construction Year Group									
	1930–1939 N = 41		1940–1949 N = 37		1950–1959 N = 43		1960–1969 N = 112		1970–1979 N = 69	
	Rate	NA	Rate	NA	Rate	NA	Rate	NA	Rate	NA
<b>Asbestos</b>	<b>0.76</b>	<b>10%</b>	<b>0.66</b>	<b>14%</b>	<b>0.86</b>	<b>16%</b>	<b>0.75</b>	<b>6%</b>	<b>0.69</b>	<b>12%</b>
Pipe insulation	0.54	41%	0.47	49%	0.79	44%	<b>0.42</b>	<b>34%</b>	<b>0.27</b>	<b>52%</b>
Valves	0.25	71%	0.31	65%	0.50	86%	0.43	75%	0	86%
Door/windows insulation	0.54	32%	0.31	57%	0.75	63%	<b>0.45</b>	<b>43%</b>	<b>0.33</b>	<b>48%</b>
Cement panel board	0.36	66%	0.50	68%	0.55	74%	<b>0.49</b>	<b>67%</b>	0.59	58%
Tile/clinker	0.32	54%	0.21	49%	0.18	49%	<b>0.47</b>	<b>30%</b>	<b>0.26</b>	<b>45%</b>
Carpet glue	0.19	61%	0.12	57%	0.39	58%	<b>0.37</b>	<b>44%</b>	<b>0.03</b>	<b>55%</b>
Floor mat	0.57	83%	0.21	62%	0.32	56%	<b>0.62</b>	<b>54%</b>	<b>0.19</b>	<b>54%</b>
Ventilation channel	0.70	76%	0.44	76%	0.81	63%	<b>0.63</b>	<b>62%</b>	0.37	61%
Switchboard	0.50	85%	0.33	84%	0	86%	0.21	79%	0	77%
Joint	0	90%	0.17	84%	0.25	81%	0.74	79%	0.22	87%
Others	1.00	78%	0.57	81%	1.00	84%	0.75	75%	0.40	78%
<b>PCB</b>	<b>0.61</b>	<b>20%</b>	0.48	27%	<b>0.58</b>	<b>28%</b>	<b>0.44</b>	<b>13%</b>	<b>0.63</b>	<b>14%</b>
Joint/sealant	0.11	56%	0	57%	0.11	58%	<b>0.29</b>	<b>31%</b>	<b>0.28</b>	<b>43%</b>
Insulation windows	0.11	54%	0.14	62%	0.24	60%	<b>0.18</b>	<b>49%</b>	<b>0.22</b>	<b>48%</b>
Capacitors in lamp/burner	0.59	34%	0.50	51%	0.56	58%	<b>0.39</b>	<b>47%</b>	<b>0.68</b>	<b>36%</b>
Acrylic flooring	0.06	59%	0	59%	0.12	60%	<b>0.02</b>	<b>56%</b>	0.04	59%
Door closer	0	98%	1.00	89%	1.00	88%	1.00	98%	0.67	96%
Cable with PCB-oil	0.50	90%	0	97%	1.00	98%	0.20	96%	0.33	96%
Others	0.50	90%	0.25	89%	0.33	93%	0.69	86%	0	96%
<b>CFC</b>	<b>0.78</b>	<b>22%</b>	0.67	27%	0.72	42%	<b>0.68</b>	<b>28%</b>	<b>0.60</b>	<b>35%</b>
Fridge/freezer	0.27	34%	0.21	46%	0.33	56%	<b>0.29</b>	<b>46%</b>	0.35	57%
Building insulation	0.81	63%	0.60	62%	0.63	65%	<b>0.64</b>	<b>56%</b>	0.60	62%
Cooling unit	0.57	49%	0.50	57%	0.50	63%	<b>0.29</b>	<b>54%</b>	0.48	61%
Rolling gate	0	98%	0.50	89%	1.00	93%	0	99%	0.25	94%
Others	0.75	90%	0.33	92%	0.67	93%	0.33	97%	0.50	97%
<b>Mercury</b>	<b>0.91</b>	<b>15%</b>	<b>0.73</b>	<b>19%</b>	0.83	33%	<b>0.78</b>	<b>9%</b>	<b>0.95</b>	<b>10%</b>
Lighting tube	<b>0.88</b>	<b>22%</b>	0.84	32%	0.88	42%	<b>0.79</b>	<b>21%</b>	<b>0.97</b>	<b>13%</b>
Relay/switch	0.46	68%	0.36	70%	0.19	63%	<b>0.22</b>	<b>60%</b>	0.21	72%
Level monitor/sensor	0.27	63%	0.11	76%	0.10	77%	<b>0.28</b>	<b>58%</b>	0.38	70%
Thermometer	0.40	63%	0.20	73%	0.27	74%	<b>0.41</b>	<b>52%</b>	0.30	57%
Thermostat	0.10	76%	0.12	78%	0.11	79%	<b>0.14</b>	<b>67%</b>	0.06	74%
Water lock/drain line	0.33	63%	0.10	73%	0.10	77%	<b>0.10</b>	<b>65%</b>	0.05	72%
Low energy lamp	1.00	80%	0.83	84%	1.00	86%	1.00	84%	1.00	67%
Doorbell	0	98%	0	97%	0	98%	1.00	98%	0	99%
Others	1.00	85%	0.78	76%	1.00	81%	<b>0.97</b>	<b>73%</b>	0.87	78%

## Appendix D

**Table A6.** The positive detection rates and missing data of hazardous materials in area clusters (numbers in bold contain more than 30 observations).

Substance and Material	Area Range (m <sup>2</sup> )									
	–100 N = 74		101–1000 N = 143		1001–2000 N = 52		2001–3000 N = 24		3000– N = 67	
	Rate	NA	Rate	NA	Rate	NA	Rate	NA	Rate	NA
<b>Asbestos</b>	<b>0.38</b>	<b>18%</b>	<b>0.73</b>	<b>12%</b>	<b>0.82</b>	<b>4%</b>	0.87	4%	<b>0.85</b>	<b>7%</b>
Pipe insulation	0.04	62%	<b>0.37</b>	<b>43%</b>	<b>0.68</b>	<b>40%</b>	0.64	42%	<b>0.76</b>	<b>37%</b>
Valves	0.17	76%	0.20	86%	0.64	73%	0.29	71%	0.32	72%
Door/windows insulation	0	65%	<b>0.35</b>	<b>50%</b>	0.55	44%	0.42	50%	<b>0.79</b>	<b>42%</b>
Cement panel board	0.26	69%	<b>0.43</b>	<b>76%</b>	0.70	62%	0.78	63%	0.52	60%
Tile/clinker	0.07	62%	<b>0.34</b>	<b>43%</b>	<b>0.39</b>	<b>40%</b>	0.43	42%	<b>0.44</b>	<b>28%</b>
Carpet glue	0.04	64%	<b>0.17</b>	<b>55%</b>	<b>0.50</b>	<b>42%</b>	0.20	58%	<b>0.38</b>	<b>45%</b>
Floor mat	0.60	93%	<b>0.30</b>	<b>63%</b>	<b>0.56</b>	<b>38%</b>	0.50	58%	<b>0.44</b>	<b>46%</b>
Ventilation channel	0.50	97%	<b>0.51</b>	<b>70%</b>	0.67	48%	0.60	58%	<b>0.59</b>	<b>52%</b>
Switchboard	1.00	97%	<b>0.15</b>	<b>77%</b>	0	85%	0.20	79%	0.21	79%
Joint	NA	100%	0.31	91%	0.53	71%	0.67	88%	0.38	64%
Others	0.20	93%	0.76	80%	0.75	69%	0.75	83%	0.79	79%

Table A6. Cont.

Substance and Material	Area Range (m <sup>2</sup> )									
	−100 N = 74		101–1000 N = 143		1001–2000 N = 52		2001–3000 N = 24		3000– N = 67	
	Rate	NA	Rate	NA	Rate	NA	Rate	NA	Rate	NA
<b>PCB</b>	<b>0.34</b>	<b>32%</b>	<b>0.47</b>	<b>16%</b>	<b>0.43</b>	<b>19%</b>	0.78	25%	<b>0.78</b>	<b>10%</b>
Joint/sealant	0	58%	0.17	50%	0.14	44%	0.09	54%	0.37	27%
Insulation windows	0.04	62%	0.15	50%	0.06	67%	0.36	54%	0.26	43%
Capacitors in lamp/burner	0.39	55%	0.51	42%	0.43	60%	0.64	42%	0.78	45%
Acrylic flooring	0	58%	0.03	56%	0.07	71%	0	63%	0.07	55%
Door closer	NA	100%	1.00	98%	1.00	90%	NA	100%	0.67	91%
Cable with PCB-oil	NA	100%	0.25	97%	0	94%	1.00	96%	0.43	90%
Others	0.50	97%	0.45	92%	1.00	98%	0.50	92%	0.36	79%
<b>CFC</b>	<b>0.53</b>	<b>26%</b>	<b>0.71</b>	<b>26%</b>	<b>0.71</b>	<b>40%</b>	0.64	54%	<b>0.80</b>	<b>34%</b>
Fridge/freezer	0.49	42%	0.70	41%	0.65	50%	0.56	63%	0.77	61%
Building insulation	0.30	55%	0.28	57%	0.18	67%	0.25	67%	0.27	67%
Cooling unit	0.20	59%	0.35	55%	0.59	67%	0.22	63%	0.66	52%
Rolling gate	NA	100%	0.20	97%	0.50	92%	NA	100%	0.75	94%
Others	0	99%	0.62	94%	0	96%	1.00	96%	0.33	96%
<b>Mercury</b>	<b>0.55</b>	<b>26%</b>	<b>0.88</b>	<b>12%</b>	<b>0.93</b>	<b>17%</b>	0.90	17%	<b>0.91</b>	<b>13%</b>
Lighting tube	0.62	43%	0.89	20%	0.91	35%	0.85	17%	0.94	19%
Relay/switch	0.25	57%	0.33	62%	0.25	62%	0.14	71%	0.50	79%
Level monitor/sensor	0	62%	0.33	64%	0.14	73%	0.29	71%	0.48	63%
Thermometer	0.03	61%	0.29	61%	0.48	56%	0.40	58%	0.58	61%
Thermostat	0	62%	0.16	69%	0	77%	0	79%	0.50	79%
Water lock/drain line	0.14	61%	0.16	66%	0	79%	0	79%	0.28	73%
Low energy lamp	1.00	99%	0.97	75%	0.80	81%	1.00	83%	1.00	73%
Doorbell	NA	100%	0.67	98%	0	96%	NA	100%	0	96%
Others	0.67	96%	0.94	75%	0.90	81%	1.00	79%	0.94	75%

## References

- Wahlström, M.; Teittinen, T.; Kaartinen, T.; van Liesbet, C. *Hazardous Substances in Construction Products and Materials: PARADE. Best Practices for Pre-Demolition AUDITS Ensuring High Quality RAW Materials*; VTT Technical Research Centre of Finland: Esbo, Finland, 2019.
- Govorko, M.; Fritschi, L.; Reid, A. Using a mobile phone app to identify and assess remaining stocks of in situ asbestos in australian residential settings. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4922. [[CrossRef](#)] [[PubMed](#)]
- Neitzel, R.L.; Saylor, S.K.; Demond, A.H.; d'Arcy, H.; Garabrant, D.H.; Franzblau, A. Measurement of asbestos emissions associated with demolition of abandoned residential dwellings. *Sci. Total Environ.* **2020**, *722*, 137891. [[CrossRef](#)] [[PubMed](#)]
- Brown, K.W.; Minegishi, T.; Cummiskey, C.C.; Fragala, M.A.; Hartman, R.; MacIntosh, D.L. PCB remediation in schools: A review. *Environ. Sci. Pollut. Res.* **2016**, *23*, 1986–1997. [[CrossRef](#)] [[PubMed](#)]
- Laurie Kazan-Allen Chronology of Asbestos Bans and Restrictions. Available online: [http://www.ibasecretariat.org/chron\\_ban\\_list.php](http://www.ibasecretariat.org/chron_ban_list.php) (accessed on 17 January 2021).
- Franzblau, A.; Demond, A.H.; Saylor, S.K.; D'Arcy, H.; Neitzel, R.L. Asbestos-containing materials in abandoned residential dwellings in Detroit. *Sci. Total Environ.* **2020**, *714*, 136580. [[CrossRef](#)] [[PubMed](#)]
- Mecharnia, T.; Khelifa, L.C.; Pernelle, N.; Hamdi, F. An approach toward a prediction of the presence of asbestos in buildings based on incomplete temporal descriptions of marketed products. In Proceedings of the 10th International Conference on Knowledge Capture, Marina Del Rey, CA, USA, 19–21 November 2019; pp. 239–242.
- Kim, J.T.; Yu, C.W.F. Hazardous materials in buildings. *Indoor Built Environ.* **2014**, *23*, 44–61. [[CrossRef](#)]
- Wahlström, M.; Bergmans, J.; Teittinen, T.; Bachér, J.; Smeets, A.; Paduart, A. *Construction and Demolition Waste: Challenges and Opportunities in a Circular Economy*; Eionet Report—ETC/WMG2020/1; European Environment Agency European Topic Centre on Waste and Materials in a Green Economy: Mol, Belgium, 2020.
- Deloitte. *Study on Resource Efficient Use of Mixed Wastes, Improving Management of Construction and Demolition Waste - Final Report*; Prepared for the European Commission, DG ENV, Deloitte: Nantes, France, 2017.
- Wahlström, M.; Zu Castell-Rüdenhausen, M.; Hradil, P.; Smith, K.H.; Oberender, A.; Ahlm, M.; Götbring, J.; Hansen, J.B. *Improving Quality of Construction & Demolition Waste-Requirements for Pre-Demolition Audit*; Nordic Council of Ministers: Copenhagen, Denmark, 2019.
- Bergmans, J.; Dierckx, P.; Broos, K. Semi-selective demolition: Current demolition practices in Flanders. In Proceedings of the HISER Conference, Delft, The Netherlands, 21–23 June 2017.
- Wu, P.-Y.; Mjörnell, K.; Sandels, C.; Mangold, M. Machine Learning in Hazardous Building Material Management: Research Status and Applications. *Recent Prog. Mater.* **2021**, *3*. [[CrossRef](#)]

14. Kantardzic, M. *Data Mining: Concepts, Models, Methods, and Algorithms*, 3rd ed.; Wiley-IEEE Press: Hoboken, NJ, USA, 2019; ISBN 978-1-119-51604-0.
15. Hong, T.; Wang, Z.; Luo, X.; Zhang, W. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build.* **2020**, *212*, 109831. [[CrossRef](#)]
16. Govorko, M.H.; Fritschi, L.; White, J.; Reid, A. Identifying Asbestos-Containing Materials in Homes: Design and Development of the ACM Check Mobile Phone App. *JMIR Form. Res.* **2017**, *1*, e7. [[CrossRef](#)]
17. Commission, E. EUR-Lex—52012DC0433—EN—EUR-Lex. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52012DC0433> (accessed on 16 March 2021).
18. Commission, E. *Committee and the Committee of the Regions on Resource Efficiency Opportunities in the Building Sector*; European Commission: Brussels, Belgium, 2014.
19. ECORYS. *EU Construction & Demolition Waste Management Protocol*; European Commission: Brussels, Belgium, 2016.
20. Wilk, E.; Krówczyńska, M.; Zagajewski, B. Modelling the spatial distribution of asbestos-cement products in Poland with the use of the random forest algorithm. *Sustainability* **2019**, *11*, 4355. [[CrossRef](#)]
21. Krówczyńska, M.; Raczko, E.; Staniszevska, N.; Wilk, E. Asbestos-cement roofing identification using remote sensing and convolutional neural networks (CNNs). *Remote Sens.* **2020**, *12*, 408. [[CrossRef](#)]
22. Johansson, T.; Olofsson, T.; Mangold, M. Development of an energy atlas for renovation of the multifamily building stock in Sweden. *Appl. Energy* **2017**, *203*, 723–736. [[CrossRef](#)]
23. Mangold, M.; Österbring, M.; Wallbaum, H. Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock. *Energy Build.* **2015**, *102*, 328–336. [[CrossRef](#)]
24. Visualizing Distributions of Data—Seaborn 0.11.1 Documentation. Available online: <https://seaborn.pydata.org/tutorial/distributions.html> (accessed on 22 June 2021).
25. Box Plot—Wikipedia. Available online: [https://en.wikipedia.org/wiki/Box\\_plot](https://en.wikipedia.org/wiki/Box_plot) (accessed on 22 June 2021).
26. Nylander, O. *Svensk Bostad 1850–2000*; Studentlitteratur: Lund, Sweden, 2013.
27. Hall, T.; Vidén, S. The million homes programme: A review of the great Swedish planning project. *Plan. Perspect.* **2005**, *20*, 301–328. [[CrossRef](#)]
28. Pasichnyi, O.; Wallin, J.; Levihn, F.; Shahrokni, H.; Kordas, O. Energy performance certificates—New opportunities for data-enabled urban energy policy instruments? *Energy Policy* **2019**, *127*, 486–499. [[CrossRef](#)]
29. Simon, A. *Definition of Validation Levels and Other Related Concepts*; Eurostat: Luxembourg, 2013.
30. Robson, M.; Melymuk, L.; Csiszar, S.A.; Giang, A.; Diamond, M.L.; Helm, P.A. Continuing sources of PCBs: The significance of building sealants. *Environ. Int.* **2010**, *36*, 506–513. [[CrossRef](#)]
31. Donovan, S.; Pickin, J. An Australian stocks and flows model for asbestos. *Waste Manag. Res.* **2016**, *34*, 1081–1088. [[CrossRef](#)]
32. Kohler, M.; Tresp, J.; Zennegg, M.; Seiler, C.; Minder-Kohler, S.; Beck, M.; Lienemann, P.; Wegmann, L.; Schmid, P. Joint sealants: An overlooked diffuse source of polychlorinated biphenyls in buildings. *Environ. Sci. Technol.* **2005**, *39*, 1967–1973. [[CrossRef](#)] [[PubMed](#)]
33. MacIntosh, D.L.; Minegishi, T.; Fragala, M.A.; Allen, J.G.; Coghlan, K.M.; Stewart, J.H.; McCarthy, J.F. Mitigation of building-related polychlorinated biphenyls in indoor air of a school. *Environ. Health A Glob. Access Sci. Source* **2012**, *11*, 24. [[CrossRef](#)] [[PubMed](#)]
34. Herrick, R.F.; Stewart, J.H.; Allen, J.G. Review of PCBs in US schools: A brief history, an estimate of the number of impacted schools, and an approach for evaluating indoor air samples. *Environ. Sci. Pollut. Res.* **2016**, *23*, 1975–1985. [[CrossRef](#)] [[PubMed](#)]
35. Diamond, M.L.; Melymuk, L.; Csiszar, S.A.; Robson, M. Estimation of PCB stocks, emissions, and urban fate: Will our policies reduce concentrations and exposure? *Environ. Sci. Technol.* **2010**, *44*, 2777–2783. [[CrossRef](#)] [[PubMed](#)]