




Article

Modelling the Spatial Distribution of Asbestos—Cement Products in Poland with the Use of the Random Forest Algorithm

Ewa Wilk , Małgorzata Krówczyńska  and Bogdan Zagajewski 

Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, Chair of Geomatics and Information Systems, University of Warsaw, 00-927 Warsaw, Poland

* Correspondence: ewa.wilk@uw.edu.pl (E.W.); mkrowczynska@uw.edu.pl (M.K.); Tel.: +48-225-520-654 (E.W.)

Received: 21 June 2019; Accepted: 7 August 2019; Published: 12 August 2019



Abstract: The unique set of physical and chemical properties of asbestos has led to its many industrial applications worldwide, of which roofing and facades constitute approximately 80% of currently used asbestos-containing products. Since asbestos-containing products are harmful to human health, their use and production have been banned in many countries. To date, no research has been undertaken to estimate the total amount of asbestos–cement products used at the country level in relation to regions or other administrative units. The objective of this paper is to present a possible new solution for developing the spatial distribution of asbestos–cement products used across the country by applying the supervised machine learning algorithm, i.e., Random Forest. Based on the results of a physical inventory taken on asbestos–cement products with the use of aerial imagery, and the application of selected features, considering the socio-economic situation of Poland, i.e., population, buildings, public finance, housing economy and municipal infrastructure, wages, salaries and social security benefits, agricultural census, entities of the national economy, labor market, environment protection, area of built-up surfaces, historical belonging to annexations, and data on asbestos manufacturing plants, best Random Forest models were computed. The selection of important variables was made in the R v.3.1.0 program and supported by the Boruta algorithm. The prediction of the amount of asbestos–cement products used in communes was executed in the randomForest package. An algorithm explaining 75.85% of the variance was subsequently used to prepare the prediction map of the spatial distribution of the amount of asbestos–cement products used in Poland. The total amount was estimated at 710,278,645 m² (7.8 million tons). Since the best model used data on built-up surfaces which are available for the whole of Europe, it is worth considering the use of the developed method in other European countries, as well as to assess the environmental risk of asbestos exposure to humans.

Keywords: Asbestos; waste; spatial distribution of asbestos–cement products; random forest algorithm; estimation of the asbestos–cement products used

1. Introduction

Asbestos is a mineral which, due to its physical and chemical properties, has found wide application in many areas of industry and economy [1]. Traditionally, asbestos fibers were mainly used in the construction industry, in which asbestos fibers were used to manufacture asbestos–cement products, constituting over 80% of the products used, which were mainly used in roofing materials and façade claddings [2]. Due to the pathogenic effect of asbestos, in 1997, the ban on the production of asbestos-containing products was introduced in Poland; their use in an environmentally and human health safe manner is allowed until the end of 2032 [3]. Exposure to asbestos causes a wide range of

diseases, such as asbestosis, as well as cancers such as malignant mesothelioma and lung cancer [4]. The risk assessment of asbestos-related diseases requires detailed data on the quantity and where the asbestos–cement products are used [5–7]. Statistical models are used to assess the potential risk of asbestos-related diseases, e.g., pleural mesothelioma [8] or lung cancer [9], based on the total amount of asbestos fibers used in the production per country population. Due to the lack of data on the quantity and the spatial distribution of asbestos–cement products used, research is undertaken to assess the risk of developing asbestos-related diseases on the basis of asbestos consumption indices in production per country's population [10,11]. There is a lack of detailed data on the number of products used, spatially broken down into administrative units, which could significantly affect the obtained results [12]. In the National Program for Asbestos Abatement in Poland [3], it was estimated that approximately 14.5 million tons of asbestos-containing products are in use. These estimations were based on the approximation of data on asbestos fibers imports. However, there were no attempts undertaken to present the spatial distribution of the aforementioned products in relation to communes. The objective for this paper is to present a new and promising solution for developing the spatial distribution of the asbestos–cement products used across the country by the application of a machine learning algorithm, i.e., Random Forest. The motivation behind this research is to support the decision makers who are interested in public issues concerning asbestos abatement safety for humans and the environment, as well as to support them with the overall estimation of asbestos–cement products in use in Poland to assess the environmental risk of asbestos exposure to humans.

Related Work

Machine learning systems are the basis of intelligent data analysis, in which the idea of supervised learning is based on the observed dataset and the relation binding the input vector with the initial vector, i.e., explanatory variables [13]. Breiman [14] has defined Random Forest as a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. Over the past few years, the Random Forest algorithm has become a popular and widely used nonparametric regression tool in many fields of science for the estimation of the quantitative changes in phenomena, e.g., in medicine [15], environmental research [16] or economics [17], and is also considered as one of the best methods of prediction [16,18]. Among the advantages of using the high accuracy of classification of Random Forest in comparison with other used algorithms, its efficiency and effectiveness of operation with large databases, ability to maintain accuracy in the absence of data and selection of variables important in the classification process are mentioned [14]. To date, no research was undertaken on the spatial distribution of asbestos–cement products in use at the country level. The Random Forest algorithm is used in quantitative studies of various phenomena, e.g., the number of disease cases [19], the spatial distribution and spread of fires [20], the amount of seabed biomass [21], forest biodiversity and the spatial extent of vegetation [22] or animal population estimation [23]. Furthermore, a survey conducted on spatial distribution with the Random Forest method showed that the Random Forest model provides good predictions of site-specific yields [24]. The results obtained by other researchers were the premise for undertaking the research on the estimation of the amount of asbestos–cement products with the use of the Random Forest method, which is a good predictor of the scale of phenomena and spatial modelling.

2. Materials and Methods

2.1. Data Collection

The first stage of the undertaken research was to collect data: field data related to the physical counting of asbestos–cement products in use, statistical data on the socio-economic situation of communes (the lowest level of administrative division in Poland), land use and historical conditions referring to annexations, and information on asbestos manufacturing plants (Figure 1). Physical

inventory taking of asbestos–cement products was performed in 160 communes across the country with the use of orthophotomap printouts in the scale of 1:2500, ensuring that all types of communes are represented from all provinces. The following data were acquired: localization, asbestos–cement products (type, quality and amount), and buildings features (roof slope, type by function). The results of fieldwork have been implemented to the geospatial database, where the roofing area was calculated on the basis of collected data.

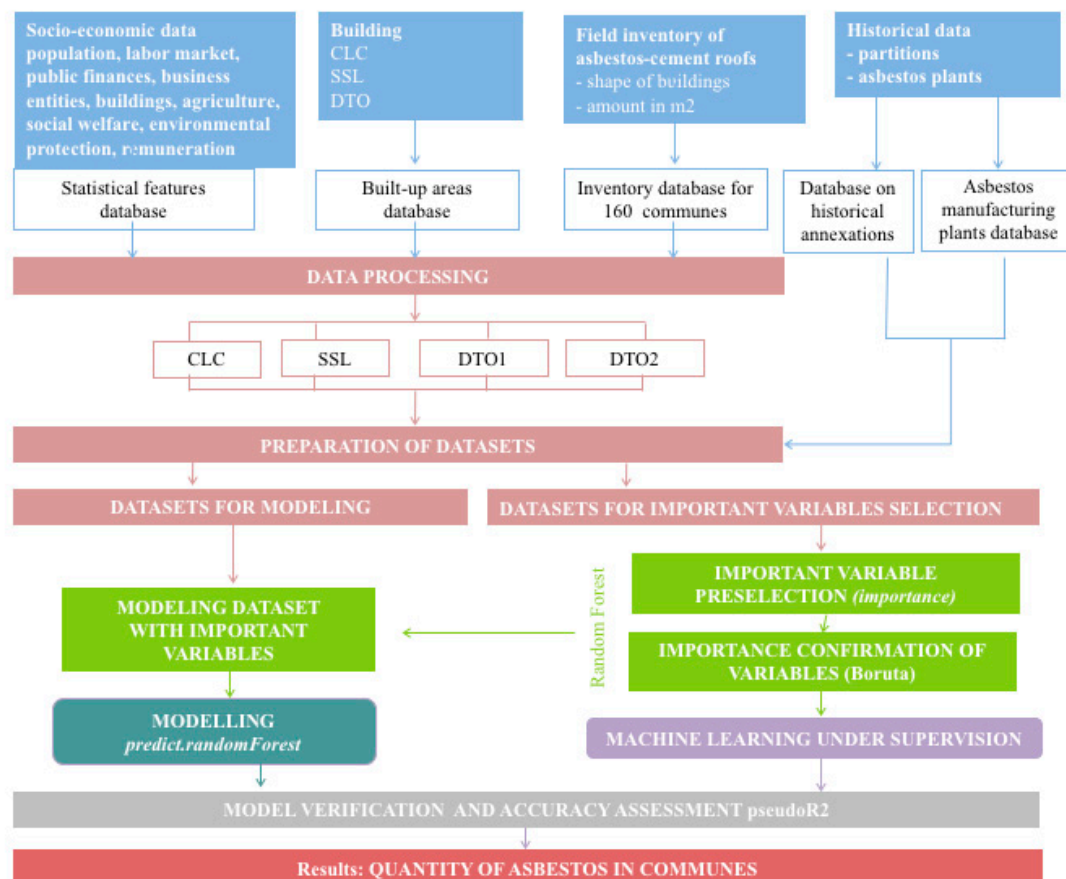


Figure 1. Flow chart of work.

To determine the socio-economic situation of the communes, data from the Local Data Bank, collected by the Central Statistical Office of Poland as part of the annual public statistics research program were used [25]. A total of 192 variables were collected for each commune, including the following categories of data: population, buildings, public finance, housing economy and municipal infrastructure, wages, salaries and social security benefits, agricultural census, entities of the national economy, labor market, and environmental protection. Features acquired for all 2478 communes in Poland were collected in a structured database, which was the source of data for the preparation of data sets, i.e., the sets for modelling and sets for the selection of significant variables.

In order to determine the area covered by buildings in each commune, data on built-up areas acquired from the Corine Land Cover Project [26], Soil Sealing Layer Project [27] and the Database of Topographic Object [28] were processed. A total of 4 datasets were prepared based on Corine Land Cover data (CLC), Soil Sealing Layer data (SSL) and Database of Topographic Object (DTO1 and DTO2, 2 sets based on different levels of detail). ArcGIS 10.5 software was used to separate built-up areas for each commune. This stage of work resulted in the development of a built-up area database for 4 datasets.

Information regarding historical belonging to annexations and, therefore, affiliation of particular parts of the territory of Poland to the partitioning countries was obtained from the Mosaic project, implemented by the Max Planck Society, and the Institute for Demographic Research in Rostock (Germany). Historical census data from different European countries are collected and harmonized, and as a result, databases on historical administrative boundaries in Europe are developed [29]. Data on the administrative division of Germany for 1910 and 1930 [30], and Austro-Hungary for 1910 [31], were used. A layer of historical affiliation to the partitions of communes in the current administrative borders of Poland was developed and constituted an input database in the modelling process. It was processed with ArcGIS 10.5 software.

The preliminary results of the research undertaken have indicated that the location of manufacturing plants that used asbestos in production has a significant impact on the estimation of the amount of asbestos–cement products used [32]. Data regarding asbestos manufacturing plants were obtained through surveys, individual interviews and on-site visits [33]. Using the ArcGIS10.5 software, all gathered data were then designed as a geospatial database, in which the following data were collected: the name of the plant, its location, types of manufactured products, the amount of production, data on the quantity of production and subsequently, the clean-up of the asbestos plant.

2.2. Data Processing

The data processing stage involved the recalculation of variables characterizing buildings acquired from the Local Data Bank in relation to the area of built-up surfaces, which were gathered in the built-up surface database. Variables on the amount of asbestos–cement products, collected in the database of field inventory, were also recalculated on the same basis. All calculations were made in the R v.3.1.0 program [34]. Other features were not subject to additional processing operations. As a result, 4 datasets were obtained, of which separate datasets of explanatory variables were developed, and the corresponding 4 sets of explained variables, i.e., the number of inventoried asbestos–cement products expressed as m² per 1 ha of built-up area.

The purpose of developing 4 datasets was to eliminate variables that did not carry any information. The analysis was performed in the caret package (Classification and Regression Training) v.6.0.37, containing a set of functions that streamline the process of developing predictive models, including division and preliminary data processing [35]. As a result of calculations made, a set of 41 explanatory variables was obtained, which was used in further research. The selection of important variables, i.e., a subset of the set of all pre-selected explanatory variables that will be used to develop the model, is an important stage in the construction of machine learning algorithms. The selection of important variables enables the reduction of the dimensionality of data and the removal of non-important variables, which increases the accuracy of learning and improves the results of the algorithm [15]. Important variables were selected separately for each of the 4 datasets for supervised learning, each learning set consisting of 160 lines, corresponding to communes in which field inventory was conducted and 42 columns, i.e., an explained variable, which is the amount of asbestos–cement product in 160 communes expressed as m² per 1 ha of built-up surface, and 41 explanatory variables.

2.3. Selection of Important Variables

The selection of important variables was made in the R v.3.1.0 program [34]. The research was carried out with the Random Forest algorithm [14]. The *importance* function was used to determine the importance of variables [36]. For each of the 4 datasets, calculations were made for the optimal number of regression tree iterations of 500 (*ntree* = 500) and the default number of variables in each node of regression, amounting to M/3 (*mtry* = 13). Measures of the significance of the variables were mean decrease accuracy (% IncMSE) and mean decrease Gini (IncNodePurity). The results obtained by the application of the *Random Forest importance* function were then verified with the use of the *Boruta* algorithm [37]. The advantage of using the Boruta algorithm is that the level of falsely detected

important variables is low; on average, less than one false-significant variable is chosen for each set [38,39].

2.4. Supervised Machine Learning

The Random Forest algorithm from the *randomForest* package [40] was used for the supervised machine learning process, separately for 4 datasets. Each regression tree is constructed using a different dataset in a learning sample, in which 2/3 of variables from the dataset are used for the learning process, and the remaining 1/3 of variables are used to test the quality of the model [18]. The *pseudo-R²* parameter is the measure of the quality of the developed algorithm [40] (1):

$$\text{pseudoR}^2 = 1 - (\text{MSE}/\text{Var}(Y)) \quad (1)$$

where: MSE—average square error,

Var—variance, and

Y—set of variables.

Calculations were made for the optimal number of regression trees, 500 (*ntree* = 500) and the optimal number of variables in each node for regression was calculated using the *tune.RF* function, allowing for the selection of the optimal number of nodes in each tree.

2.5. Random Forest Prediction

The prediction of the amount of asbestos–cement products used in communes was executed in the *randomForest* package using the *predict* function, allowing for the determination of the value of the explained variable based on observations belonging to the learning set [36]. The use of the response option in the formulae of the predict function enabled the calculation of data estimated for a given class based on the majority voting in the learning set [36]. The results were obtained for each of the 4 datasets separately and were expressed as the amount of asbestos–cement products (m²) per unit area of built-up surface (ha). The model with the highest value of the *pseudo-R²* parameter was then the subject of further research. Additional analysis included the number of inhabitants and the area of the built-up surfaces.

2.6. Accuracy Assessment

The first stage of the accuracy assessment was the comparison of the *pseudo-R²* parameter values as a measure of the quality of the model for the four tested datasets. Then, the results obtained in the process of modelling were compared with fieldwork data acquired in 160 communes. The result of the model fit measurement was elucidated and is expressed in the value of the determination coefficient *R²*.

2.7. Spatial Autocorrelation

Using spatial statistics, the amount of asbestos–cement products per capita in communes were tested. Moran's I autocorrelation coefficient was used to measure the correlation between neighboring observations in communes, calculated with the following formula, (2) [41,42]:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2}, \quad i \neq j \quad (2)$$

where *n* represents the number of study areas (communes), *w_{ij}* represents the weight matrix of links between object *i* and object *j* (kg of asbestos–cement products per capita in *i* or *j* commune), *x_i*, *x_j* are variable values in *i* and *j* spatial units (kg of asbestos–cement products per capita) and \bar{x} is the arithmetic mean of the variable for all units.

3. Results

Modelling of the spatial distribution of the quantity of asbestos–cement products used in communes was performed for four datasets. The *pseudo-R*² parameter of the developed models ranged from 54.05 to 75.85%, whereas the quantity of asbestos–cement products derived from the developed models amounted to 7.4 to 9.3 million tons (Table 1).

The verification of the developed model was based on the comparison of the obtained values of the *pseudo-R*² parameter for four tested datasets. The highest value of the *pseudo-R*² parameter (75.85%) was obtained for the SSL dataset. Due to this fact, the SSL dataset modelling results were then of further interest and were compared with data from the field inventory taken in 160 communes. The coefficient of determination *R*² reached 94% (Figure 2).

Table 1. Summary of the *pseudo-R*² parameter in four datasets in relation to the estimated amount of asbestos–cement products in use.

Dataset	Pseudo- <i>R</i> ² (%)	The Amount of Asbestos–Cement Products (tons)
Corine Land Cover (CLC)	54.05	9,271,099
Database of Topographic Object (DTO) 2	68.80	7,579,424
Database of Topographic Object (DTO) 1	70.39	7,398,852
Soil Sealing Layer (SSL)	75.85	7,813,065

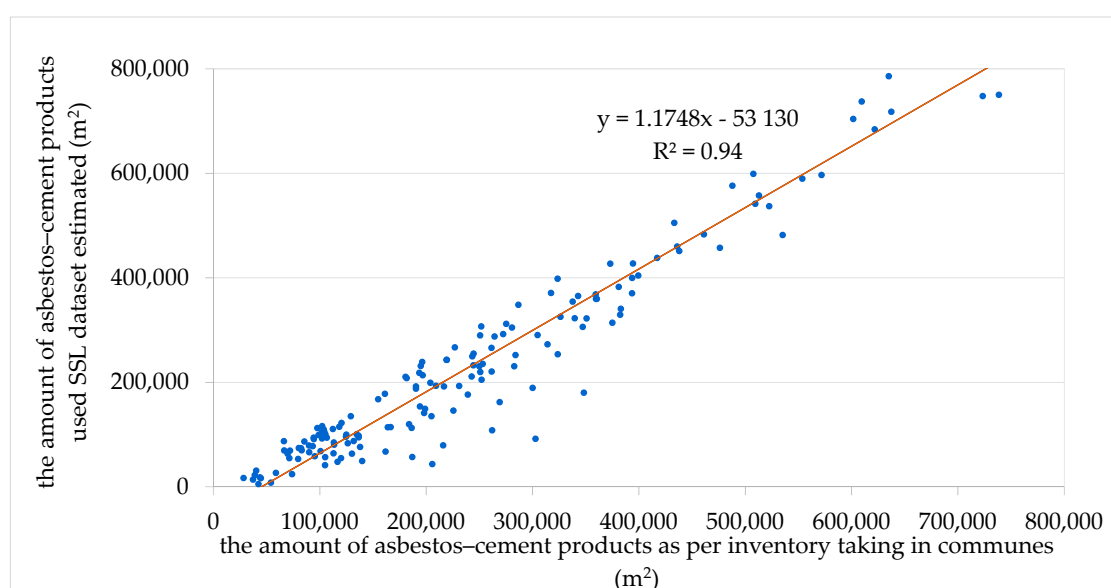


Figure 2. SSL dataset modelling the amount of asbestos–cement products in use, fit to the data derived from field inventory taking.

It has been estimated that 710,278,645 m² of asbestos–cement products are used in all communes in Poland, which gives 7.8 million tons (Table 1). The largest amount of asbestos–cement products, i.e., 18% of the total estimated amount, is in Mazowieckie, followed by Lubelskie (12%), and Łódzkie and Wielkopolskie (each with a 9% share; Figure 3).



Figure 3. SSL dataset modelling results in communes for all provinces (tons).

The results of the estimation of the amount of asbestos–cement products by the type of commune indicate that 62% of these products are used in rural communes, 26% in urban–rural communes, and the remaining 12% in urban communes. On average, the per capita use of the amount of asbestos–cement products amounts to 202 kg per inhabitant of Poland. This indicator is the highest in Lubelskie—423 kg per person—followed by Podlaskie—416 kg per person—and Świętokrzyskie—387 kg per person (Figure 4). In Łódzkie, Mazowieckie, Podkarpackie and Wielkopolskie, it is over 200 kg, and in Kujawsko-Pomorskie, Opolskie, Małopolskie, Warmińsko-Mazurskie, Pomorskie, Lubuskie and Zachodniopomorskie, it is over 100 kg. Less than 100 kg per inhabitant is used in Śląskie and Dolnośląskie.

On average, per hectare of built-up surface, there are nine tons of asbestos–cement products in use. The largest number of asbestos products per 1 ha of built-up surface is in Lubelskie and is slightly over 18 tons (Figure 5). More than 10 tons per hectare of built-up area is used in Świętokrzyskie, Podlaskie, Łódzkie, Mazowieckie, Podkarpackie, Łódzkie and Małopolskie. In Kujawsko-Pomorskie, the indicator has values over the national average. The smallest number of products per unit of built-up area is used in Śląskie, Dolnośląskie, Zachodniopomorskie and Lubuskie (5 tons).

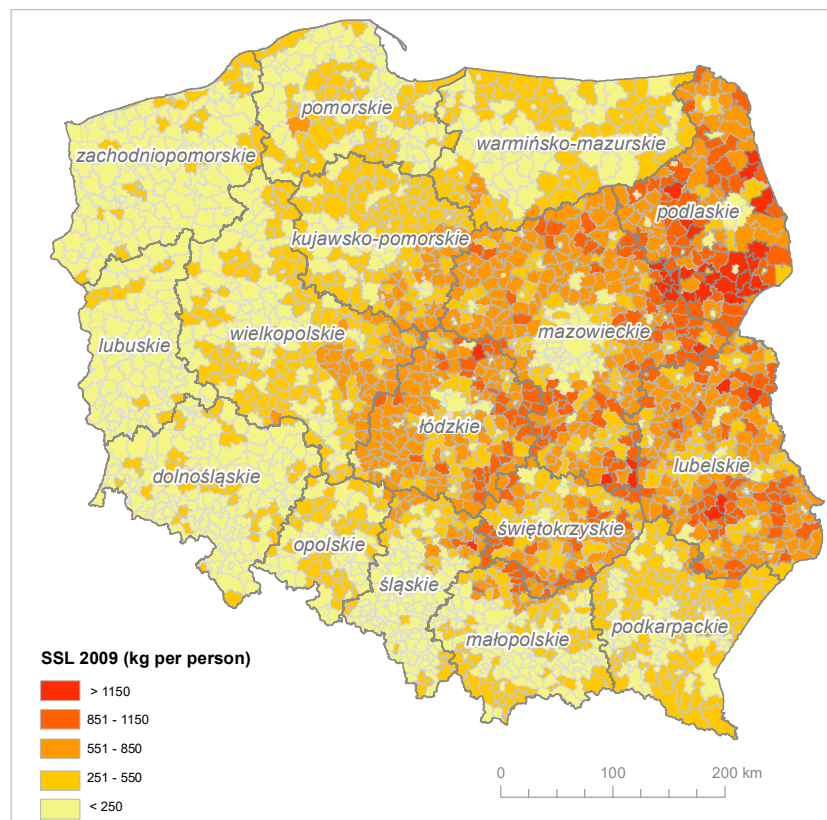


Figure 4. SSL dataset modelling results in communes per capita (kg per capita).

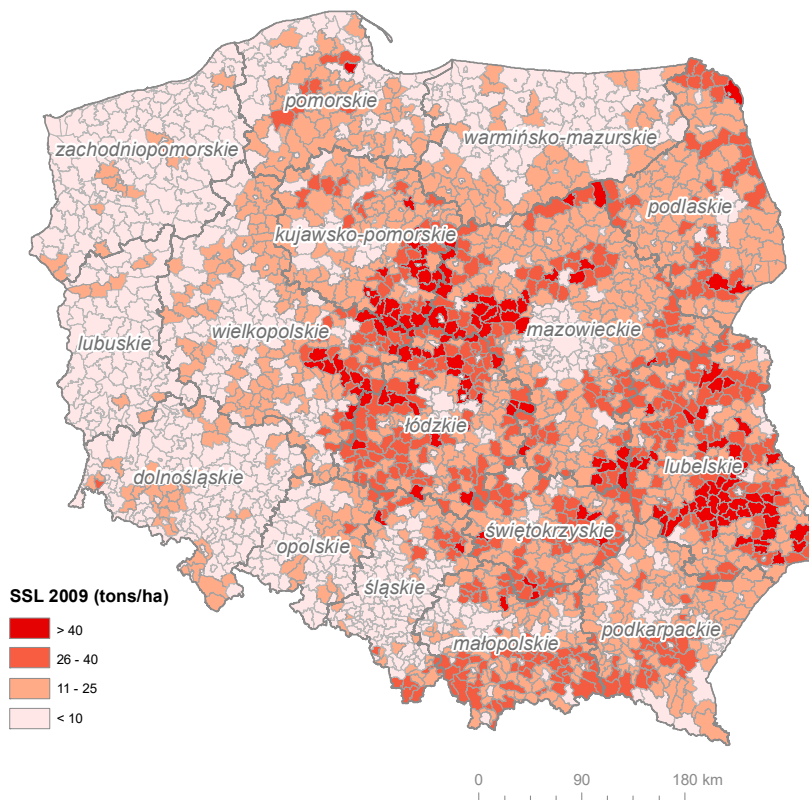


Figure 5. SSL dataset modelling results in communes per built-up surface area (tons per ha).

There is a statistically significant positive autocorrelation of the amount of asbestos cement products used per capita (Table 2).

Table 2. Global Moran's I statistics for the amount of asbestos–cement products per capita.

	<i>Moran I</i>	<i>Variance (I)</i>	<i>Z-score</i>
kg of asbestos–cement products per capita	0.569	0.00018	41.697

Communes with a high amount of asbestos–cement products per person surrounded by communes with a similar number of asbestos products per capita are located in Central and Eastern parts of Poland (Łódzkie, Świętokrzyskie, Mazowieckie, Podlaskie and Lubelskie). In Śląskie, communes with a small amount of asbestos–cement products per person were noted. This is due to the fact that in Śląskie, there exists a concentration of agglomerations, which causes the amount of asbestos–cement products per capita to be quite low (Figure 6).

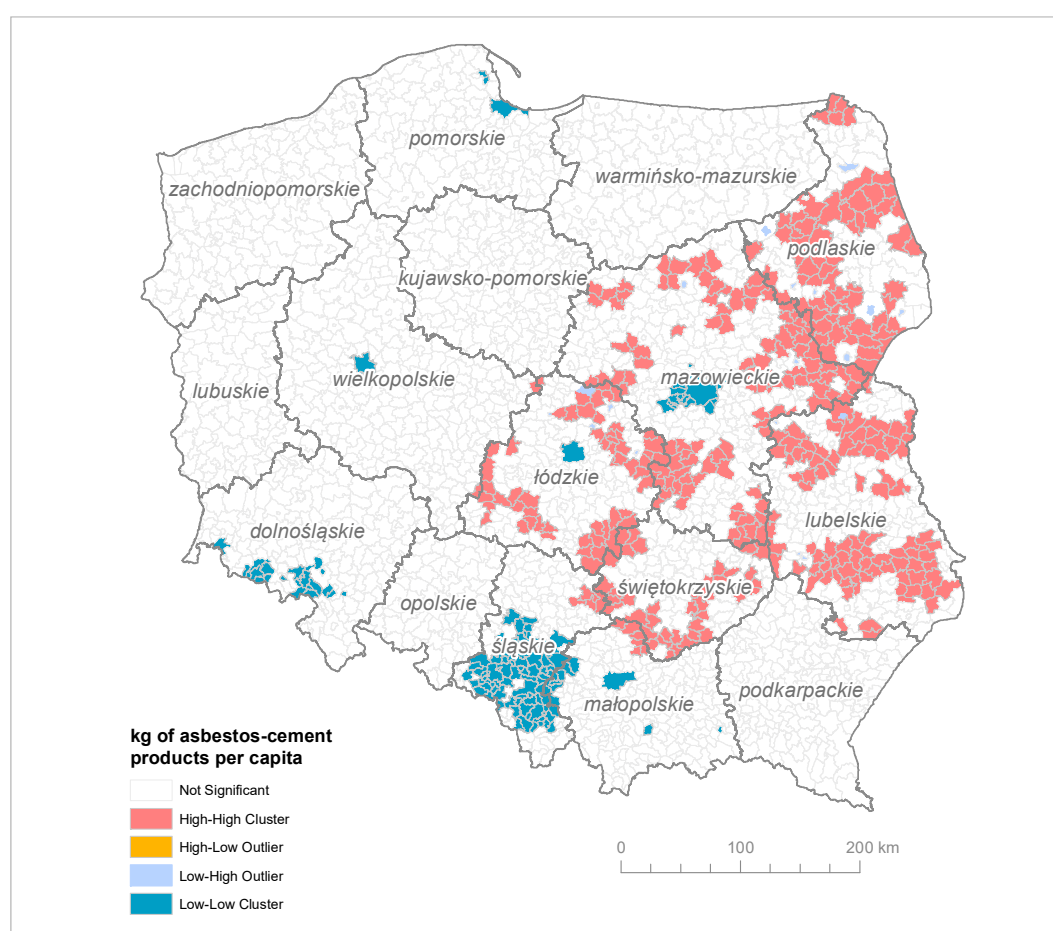


Figure 6. Local autocorrelation of the amount of asbestos–cement products per capita (kg per capita).

4. Discussion and Conclusions

Veall and Zimmermann [43] stated that in the assessment of prediction models, there is no threshold value of the $pseudo-R^2$ parameter above which it would be possible to determine the quality of the model developed with the Random Forest algorithm. Various $pseudo-R^2$ values are considered as acceptable to explain changes depending on the type and nature of the phenomenon studied. Pierce et al. [20], in the modelling of fires in California, considered a $pseudo-R^2$ value of

55%–68% as a reasonable level to explain the perceived changes. Using the same method to model the spatial patterns of the spread of fires in the Mediterranean area, Oliveira et al. [44] obtained a result of 93–95%. Leutner et al. [22], in the modelling of mountain forest biodiversity in Germany, obtained a *pseudo-R*² result in the range of 26–55%, depending on the type of vegetation in different layers of the forest. Wei et al. [21], in forecasting the amount of seabed biomass, obtained a *pseudo-R*² result ranging from 63 to 88%, depending on the group of organisms studied (bacteria, meiofauna, macrofauna and megafauna). In medical science, values of the *pseudo-R*² parameter over 80% are obtained, but the sets of genes for which the parameter values are achieved in the range of 16 to 26.4% are also analyzed [43].

The results of modelling the amount of asbestos–cement products used were then compared with data collected in the Asbestos Database [3]. Since 2012, the Asbestos Database has provided a register of products containing asbestos, both used and removed. There is a legal obligation to report on the amount of used and removed asbestos products. As of the end of 2016, data on 5.5 million tons of asbestos–cement products in use were collected in the Asbestos Database, i.e., 2.3 million tons less than the amount estimated in the developed model (Table 3).

Table 3. The comparison of the quantity of asbestos–cement products estimated in the SSL model and Asbestos Database.

Province	SSL Estimation Model (tons)	Asbestos Database (tons)	Difference (tons)
Mazowieckie	1,390,000	1,040,811	349,189
Lubelskie	921,011	850,436	70,575
Łódzkie	705,605	531,553	174,052
Wielkopolskie	691,099	511,729	179,370
Małopolskie	566,924	286,771	280,153
Podlaskie	500,264	404,172	96,092
Świętokrzyskie	495,805	378,583	117,222
Podkarpackie	458,703	245,378	213,325
Śląskie	456,096	239,193	216,903
Kujawsko-pomorskie	416,870	375,490	41,380
Dolnośląskie	272,783	114,198	158,585
Pomorskie	264,171	166,753	97,418
Warmińsko-mazurskie	216,392	165,013	51,379
Zachodniopomorskie	174,848	110,603	64,245
Opolskie	172,737	61,936	110,801
Lubuskie	109,760	65,171	44,589
Total	7,813,065	5,547,790	2,265,278

Although the percentage of communes fulfilling the obligation to keep a register in the Asbestos Database is high, there are no mechanisms for checking the completeness of data [45,46]. In previous research undertaken by Wilk et al. [47], it was estimated that 8.2 million tons of asbestos–cement products are used in Poland, i.e., approximately 0.4 million tons less than in the analyzed SSL dataset, and the *pseudo-R*² parameter reached a value of 72.9%. In previous research, modelling was executed using all variables that had been verified as important, regardless of the calculated value of the coefficient *Z*, which is the significance of variables. In the present research, the selection of variables relevant for four datasets was carried out using only variables that were confirmed as important in the modelling process, which has contributed to an increase in the fit of the model to 75.85%. Nevertheless, similarly to the previously obtained result, the highest value of the *pseudo-R*² parameter was achieved for the SSL dataset, and the amount of asbestos–cement products calculated on its basis was estimated at approximately 7.8 million tons.

Due to the fact that the highest value of the *pseudo-R*² parameter was reached for the SSL dataset regarding built-up areas, which is being developed for the whole of Europe as part of the Copernicus Land Monitoring Service project [27], it will be possible to apply the algorithm in other European countries. A necessary condition is to derive data from the SSL project and compile it in relation

to the administrative units of a given country, and to conduct a field inventory of asbestos–cement products, which is a time-consuming and laborious task, but essential to obtain training set data for the application of the Random Forest algorithm. Then, it is necessary to verify the obtained results with the test set. The collection of the explanatory variables is the next stage, i.e., to gather statistical data on the social and economic situations of administrative units, collected by public statistics, in order to determine the importance of the variables. Special circumstances that have an impact on the potential development of the built-up areas also need to be considered, such as if, historically, they belonged to former annexations in Poland. At the same time, it should be emphasized that data on the amount of asbestos–cement products used contribute to the estimation of environmental asbestos exposure and the study on modelling the incidence of asbestos-related diseases based on the number of cases and the amount of asbestos–cement products in use [48].

Author Contributions: Conceptualization and methodology, all authors; software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, visualization, E.W. and M.K.; writing—review and editing, all authors; supervision, project administration, funding acquisition, E.W. and M.K.

Funding: The publishing costs were financed from the theme No. 500-D119-12-1190000 awarded by the Polish Ministry of Science and Higher Education.

Acknowledgments: Parts of this analysis was funded by the University of Warsaw and the WGS84 Polska Sp. z o.o. company. We thank staff members of both units. A special thanks go to anonymous Reviewers who, pointing to the weak parts of the manuscript, contributed to its improvement. We are grateful to Mrs. Maria Liu (Assistant Editor) for help in the article publication procedure.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Virta, R. Worldwide Asbestos Supply and Consumption Trends from 1900 through 2003, U.S. Geological Survey Circular 1298. 2006. Available online: <http://pubs.usgs.gov/circ/2006/1298/c1298.pdf> (accessed on 20 March 2017).
2. Krówczyńska, M.; Wilk, E. Asbestos Exposure and the Mesothelioma Incidence in Poland. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1741. [CrossRef] [PubMed]
3. *Programme for Asbestos Abatement in Poland 2009–2032*; Ministry of Economy: Warsaw, Poland, 2010. Available online: <https://www.gov.pl/web/przedsiębiorczosc-technologia/usuwanie-azbestu> (accessed on 11 April 2019).
4. Rake, C.; Gilham, C.; Hatch, J.; Darnton, A.; Hodgson, J.; Peto, J. Occupational, domestic and environmental mesothelioma risks in the British population: A case control study. *Br. J. Cancer* **2009**, *100*, 1175–1183. [CrossRef] [PubMed]
5. Krówczyńska, M.; Wilk, E. Aerial imagery and geographic information systems used in the asbestos removal process in Poland. In *Proceedings of the 33th EARSeL Symposium Towards Horizon 2020: Earth Observation and Social Perspectives*; Lasaponara, R., Masini, N., Biscione, M., Eds.; Matera, Italy, 2013; pp. 823–828.
6. Fiumi, L.; Tocci, S.; Meoni, C. Remote sensing and GIS for land use planning: An application for mapping asbestos–cement roofing in Tiburtina, Rome, Italy. *Int. J. Remote Sens. Geosci.* **2014**, *3*, 1–9.
7. Krówczyńska, M.; Wilk, E. Geoazbest serwis do monitorowania procesu usuwania wyrobów azbestowych. *Rocz. Geomatyki* **2016**, *14*, 477–486.
8. Peto, J.; Hodgson, J.T.; Matthews, F.E.; Jones, J.R. Continuing increase in mesothelioma mortality in Britain. *Lancet* **1995**, *345*, 535–539. [CrossRef]
9. Nelson, H.H.; Kelsey, K.T. The molecular epidemiology of asbestos and tobacco in lung cancer. *Oncogene* **2002**, *21*, 7284–7288. [CrossRef] [PubMed]
10. Van der Borre, L.; Deboosere, P. Asbestos in Belgium: An underestimated health risk. The evolution of mesothelioma mortality rates (1969–2009). *Int. J. Occup. Environ. Health* **2014**, *20*, 134–140. [CrossRef] [PubMed]
11. Krówczyńska, M.; Wilk, E. Spatial analysis of the exposure to asbestos and health care in Poland in 2004–2013. *Geospat. Health* **2018**, *13*. [CrossRef] [PubMed]

12. Pan, X.L.; Day, H.W.; Wang, W.; Beckett, L.A.; Schenker, M.B. Residential Proximity to Naturally Occurring Asbestos and Mesothelioma Risk in California. *Am. J. Respir. Crit. Care Med.* **2005**, *172*, 1019–1025. [[CrossRef](#)] [[PubMed](#)]
13. Koronacki, J.; Ćwik, J. *Statystyczne Systemy Uczące Sie*; Akademicka Oficyna Wydawnicza EXIT: Warszawa, Poland, 2015; p. 327.
14. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
15. Strobl, C.; Boulesteix, A.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [[CrossRef](#)] [[PubMed](#)]
16. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
17. Ballings, M.; Poel, D.V.D.; Hespeels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst. Appl.* **2015**, *42*, 7046–7056. [[CrossRef](#)]
18. Breiman, L.; Cutler, A. Random Forests. 2007. Available online: <https://www.stat.berkeley.edu/~breiman/RandomForests/> (accessed on 15 April 2017).
19. Lehmann, C.; Koenig, T.; Jelic, V.; Prichep, L.; John, R.E.; Wahlund, L.O.; Dodge, Y.; Dierks, T. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J. Neurosci. Methods* **2007**, *161*, 342–350. [[CrossRef](#)] [[PubMed](#)]
20. Pierce, A.D.; Farris, C.A.; Taylor, A.H. Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. *For. Ecol. Manag.* **2012**, *279*, 77–89. [[CrossRef](#)]
21. Wei, C.; Rowe, G.T.; Escobar-Briones, E.; Boetius, A.; Soltwedel, T.; Caley, J.M.; Soliman, Y.; Huettmann, F.; Qu, F.; Yu, Z.; et al. Global Patterns and Predictions of Seafloor Biomass using Random Forests. *PLoS ONE* **2010**, *5*, e15323. [[CrossRef](#)]
22. Leutner, B.F.; Reineking, B.; Muller, J.; Bachmann, M.; Beierkuhnlein, C.; Dech, S.; Wegmann, M. Modelling Forest α -Diversity and Floristic Composition—On the Added Value of LiDAR plus Hyperspectral Remote Sensing. *Remote Sens.* **2012**, *4*, 2818–2845. [[CrossRef](#)]
23. Obidziński, A.; Pabjanek, P.; Medrzycki, P. Determinants of badger *Meles meles* location in Białowieża Primeval Forest, northeastern Poland. *Wildl. Biol.* **2013**, *19*, 48–68. [[CrossRef](#)]
24. Vincenzi, S.; Zucchetta, M.; Franzoi, P.; Pellizzato, M.; Pranovi, F.; De Leo, G.A.; Torricelli, P. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecol. Model.* **2011**, *222*, 1471–1478. [[CrossRef](#)]
25. Local Data Bank. Available online: <https://bdl.stat.gov.pl/> (accessed on 17 July 2017).
26. Corine Land Cover. 2006. Available online: <http://clc.gios.gov.pl/> (accessed on 15 March 2014).
27. SSL. Copernicus Land Monitoring Service. 2009. Available online: <http://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/imperviousness-2009/view> (accessed on 29 March 2014).
28. DTO. Database of Topographic Objects BDOT10k. 2011 Head Office of Geodesy and Cartography, Warsaw, Poland. Available online: <http://www.gugik.gov.pl/projekty/gbdot/produkty> (accessed on 5 January 2017).
29. Szoltysek, M.; Gruber, S. Mosaic: Recovering surviving census records and reconstructing the familial history of Europe. *Hist. Fam.* **2016**, *21*, 38–60. [[CrossRef](#)]
30. MPIDR Population History GIS Collection. Max Planck Institute for Demographic Research & Chair for Geodesy and Geoinformatics. In *Grundriß der Deutschen Verwaltungsgeschichte*; Hubatsch, W., Klein, T., Eds.; University of Rostock: Rostock, Germany, 1975.
31. The Mosaic Project MPIDR Population History GIS Collection. Max Planck Institute for Demographic Research & Chair for Geodesy and Geoinformatics, University of Rostock, Rostock. Available online: <https://censusmosaic.demog.berkeley.edu/data/historical-gis-files> (accessed on 17 July 2017).
32. Wilk, E.; Króczyńska, M.; Pabjanek, P. Determinants influencing the amount of asbestos-cement roofing in Poland. *Misc. Geogr.* **2015**, *19*, 82–86. [[CrossRef](#)]
33. Wilk, E.; Króczyńska, M.; Zagajewski, B. Asbestos manufacturing plants in Poland. *Misc. Geogr.* **2014**, *18*, 53–58. [[CrossRef](#)]
34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
35. Kuhn, M. A Short Introduction to the Caret Package. 2016. Available online: <https://CRAN.R-project.org/package=caret> (accessed on 28 June 2017).

36. Liaw, A. Package ‘Randomforest’. Breiman and Cutler’s Random Forests for Classification and Regression. 2015. Available online: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> (accessed on 13 March 2017).
37. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
38. Kursa, M.B.; Jankowski, A.; Rudnicki, W.R. Boruta—A system for feature selection. *Fundam. Inform.* **2010**, *101*, 271–285.
39. Rudnicki, W.R.; Wrzesień, M.; Paja, W. All relevant feature selection methods and applications. *Stud. Comput. Intell.* **2015**, *584*, 11–28.
40. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
41. Moran, P.A.P. Notes on Continuous Stochastic Phenomena. *Biometrika* **1950**, *37*, 17–23. [[CrossRef](#)] [[PubMed](#)]
42. Li, H.; Calder, C.A.; Cressie, N. Beyond Moran’s I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geogr. Anal.* **2007**, *39*, 357–375. [[CrossRef](#)]
43. Veall, M.R.; Zimmermann, K.F. Pseudo-R² Measures for Some Common Limited Dependent Variable Models. *J. Econ. Surv.* **1996**, *10*, 241–259. [[CrossRef](#)]
44. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [[CrossRef](#)]
45. Pang, H.; Lin, A.; Holford, M.; Enerson, B.E.; Lu, B.; Lawton, M.P.; Floyd, E.; Zhao, H. Pathway analysis using random forest classification and regression. *Bioinformatics* **2006**, *22*, 2028–2036. [[CrossRef](#)] [[PubMed](#)]
46. Krówczyńska, M.; Wilk, E.; Zagajewski, B. The Electronic Spatial Information System—Tools for the monitoring of asbestos in Poland. *Misc. Geogr.* **2014**, *18*, 59–64. [[CrossRef](#)]
47. Wilk, E.; Krówczyńska, M.; Pabjanek, P.; Mędrzycki, P. Estimation of the amount of asbestos–cement roofing in Poland. *Waste Manag. Res.* **2017**, *35*, 491–499. [[CrossRef](#)] [[PubMed](#)]
48. Krówczyńska, M.; Wilk, E. Environmental and Occupational Exposure to Asbestos as a Result of Consumption and Use in Poland. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2611. [[CrossRef](#)] [[PubMed](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).