UCLouvain

AGRO LOUVAIN

Faculté des bioingénieurs

Geospatial Modeling of Heavy Metals Concentrations in Urban Road Dust

Anderlecht, Brussels-Capital Region

Author : Benoit de Saint-Hubert
Promoters : Patrick Bogaert
Michel Verbanck
Readers : Pierre Defourny
Emmanuel Hanert

Academic year : 2022-2023
Bioengineer : Agronomical sciences

Acknowledgements

I would like to take this opportunity to express my gratitude to the many individuals who have contributed to the completion of this master thesis.

I am deeply grateful to my promoters, Professor P. Bogaert for his guidance and constant support during the entire master thesis, and Professor M. Verbanck for his trust and for welcoming me in his research team. Their help and advices were instrumental in the realization of this master thesis.

I would like to extend my thanks to G. Diélie for sharing with me his expertise for his guidance during the laboratory analyses. I would also like to thank A. Briffault for sharing with me his research and findings, allowing me to continue what he started.

I would like to thank Professor P. Defourny and Professor E. Hanert for accepting to be part of my jury and review my master thesis.

Finally, I would like to express my appreciation to my family and friends for their supporting and encouraging presence during this master thesis. I would also like to thank them for being an essential source of motivation during my entire cursus.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

AGVIF : Adjusted generalized variance inflation factor

AIC : Akaike Information Criterion

AM : Anderlecht municipality

APCS : absolute principal component scores

BCR : Brussels-Capital Region

BM : Brussels Mobility

CA : Clustering analysis

CF : contamination factor

EF : enrichment factor

FA : Factor analysis

HM(s) : heavy metal(s)

IBSA : Belgium National Institute for Statistics

IRM : Royal Meteorological Institute

MLR : multivariate linear regression

MUSTI : Strategic multimodal mobility model

PAH(s) : polycyclic aromatic hydrocarbon(s)

PCA : principal component analysis

PCR : principal component regression

PCS : principal component scores

PM : particulate matter

PM2.5 : particulate matter ($\varnothing < 2.5\mu m$)

PM10 : particulate matter ($\varnothing < 10\mu m$)

PMF : positive matrix factorization

RD : road dust

URBIS : IT service of the BCR

1 Introduction

Urban pollution is a pressing global concern due to its detrimental effects on the environment and the potential health risks it poses to the inhabitants. In urban areas, atmospheric pollution is a major form of environmental pollution, mainly related to fine and coarse particulate matter (PM2.5 and PM10, respectively). Particulate matter (PM) is a complex atmospheric mixture of particles that can cause respiratory and cardiovascular damages, due to their small size [1, 2]. A diverse range of pollutants such as Heavy metals(HMs) and polycyclic aromatic hydrocarbons (PAHs) can be adsorbed onto or incorporated into PM [3, 4]. The presence of these pollutants in PM is an additional concern for human health due to their toxicity (HMs [5, 6] and PAHs [7, 8]). In urban areas, an important proportion of PM in the atmosphere is originating from road dust (RD) resuspension [9].

RD is a complex mixture of solid mineral and organic particles of various sizes originating from a variety of sources and contaminated by adsorbed pollutants (also HMS and PAHs) [10]. Accordingly, RD also raises health concerns since it can contaminate people through inhalation, ingestion or dermal contact [11, 12]. HMs presence in the finest fraction of RD is due to dry and wet deposition of contaminated PM [13, 14] and direct deposition, mainly associated to vehicular metallic body parts wear and tear [15, 16]. The impermeable surface associated to the pavement of the urban areas intensifies the accumulation of RD [17]. RD is also a major concern for the pollution of urban water environment, as stormwater runoffs can cause the transfer of RD to water bodies [18, 19]. The resuspension (i.e., transfer of the fine fraction to the atmosphere) of contaminated RD can occur due to turbulences generated by wind and vehicular traffic. This phenomenon is intensified within street canyons due to the modification of the speed and direction of the wind caused by the peculiar geometry of these streets (i.e., tall buildings on both side of the street forming a narrow space) [20–22]. These local conditions can also concentrate PM and HMs in RD. For all these reasons, the study of RD and their contamination in HMs is of major scientific interest.

In order to better understand this pollution, descriptive approaches can be used, even with a relatively limited amount of samples. Such approaches are generally oriented towards the identification and apportionment (i.e., quantification of the contribution) of pollution sources of RD and its contaminants (e.g., [23, 24]). They can also assess the RD level of contamination by using pollution indices (e.g., geoaccumulation index, contamination factor, enrichment factor) to compare the observed pollutants concentrations with background values (i.e., approximate concentrations of a theoretical or local unpolluted sediment). These indices have been used to assess a variety of sediments including RD (e.g., [25–27]). This study will present results for one of these indices. In recent studies, additional indices are used to assess the risks for human health (e.g., hazard index, hazard quotient, carcinogenic risk index) (e.g., [17, 26]).

On the other hand, modeling approaches, mainly oriented toward the study of the spatial distribution of HMs concentrations in RD, are generally requiring a consequent number of samples. The modeling is mainly done by using the spatial autocorrelation of the HMs concentrations in RD (e.g., [23, 26, 28, 29]) and more rarely by using information related to the surroundings and characteristics of the sampled areas (e.g., [30]) even if this approach is used to model other atmospheric pollutants (e.g., [31, 32]). Our study aims to provide a detailed approach to identify the most relevant variables toward modeling the spatial distribution HMs concentrations in RD in urban areas. Resulting maps could prove useful to the local authorities by providing practical insights for the implementation mitigation measures (e.g., informed decisions concerning street sweeping).

2 State-of-the-art

2.1 Sources of pollution

2.1.1 Identified Sources

In urban areas, Cadmium (Cd), Chromium (Cr), Copper (Cu), Nickel (Ni), Lead (Pb) and Zinc (Zn) presence in RD originates from diverse sources [38], mainly of anthropogenic nature [34]. Table 1 presents the main anthropogenic sources for these six HMs, based on a literature review. An important note is that the contributions of those sources largely varies between different areas. Traffic is a main source contributing to the pollution of RD in HMs due to the wear, tear and abrasion of the of vehicles components on the road network (brake pads, tires, vehicle body, combustion engine) [15–17, 26, 33, 34] and vehicles exhausts emissions [15, 26, 34, 35]. The concentrations of HMs in RD are influenced by various human activities that can be categorized as industrial, commercial, administrative, and residential. These activities can be spatially linked to their respective Land use delimitations. Consequently, these different activities have been identified as sources due to their contribution to RD contamination in HMs is distinct [15, 29, 33, 36]. Other known sources are the erosion of the road coatings, buildings materials, and traffic infrastructures as well as the waste produced by their construction/renovation [16, 17, 34, 37].

Table 1 : Main anthropogenic sources of HMs in RD. "Vehicles components" corresponds to vehicles components wear, tear and abrasion. "Infrastructures" corresponds to the construction/renovation works associated to roads, buildings and traffic infrastructures as well as their erosion.

Source	Cd	Cr	Cu	Ni	Pb	Zn
Vehicles components	[15–17, 26, 33]	[17, 33]	[15–17, 33, 34]	[17, 33, 34]	[15, 33, 34]	[15, 16, 33, 34]
Exhausts emissions	-	[15]	[34, 35]	[15, 26]	[26, 34]	[15, 26, 34]
Land use	-	[15, 33, 36]	[15, 29, 33, 36]	[33, 36]	[36]	[29]
Infrastructures	[17, 34, 37]	[16, 17]	[16, 17, 37]	[16, 17]	[16, 37]	[16, 34, 37]

2.1.2 Source Identification and Apportionment

The identification of the sources can be done by different approaches. It can be based on ANOVA I and simple average comparison for categorical variables or by Pearson correlation analysis for continuous variables (e.g., [17, 23, 29, 36]). However, these methods are failing to provide the contribution of those sources to the contamination of RD, due to the one-to-one relationship between the sources and the HMs concentrations.

In order to estimate the contribution of the sources, studies can use different approaches often including dimensionality reduction on the HMs concentrations in the samples. These approach are using algorithms such as Factor Analysis (FA) (e.g., [23, 34]), Principal Components Analysis (PCA) (e.g., [24, 28]) or Positive Matrix Factorization (PMF) [15, 26, 38] to regroup HMs by similarity and identify underlying structures in the data[1]. These underlying structures are generally called factors, virtual variables that can be, with prior assumptions, viewed as underlying variables (in this case, sources of contamination). By comparing the patterns accross all HMs studied and by gathering information from previous literature, these underlying variables can be linked with known sources. Further steps, such as a multivariate analysis based on these factors (e.g., PCA-MLR receptor models) might be needed to estimate the contributions of the sources to the HMs concentrations (e.g., [16, 40]).

2.2 Pollution indices

In order to assess the relative contamination of a sediment (e.g., RD, soil, aquatic sediment) in an element (e.g., HMs, PAHs), a wide array of indices (e.g., Geoaccumulation index, Enrichment Factor, Contamination Factor) has been created. In recent literature, these indices have been used to characterize the contamination of RD in HMs (e.g., [17, 36, 41]). For the majority of these indices, the calculations

[1]Pearson correlation and Clustering Analyses (CA) can also be used to compare and regroup HMs by similar behaviors (e.g., [23, 24, 39]).

revolve mainly around the relation between the average concentration of the element in the sediment and a geochemical background value for the same element. The choice of this background has a strong impact on the quality of the results as it is supposed to represent an uncontaminated sediment. These values can be defined by sampling a local soil unaffected by human activities or by using a reference background (e.g., the average HMs content of the upper continental crust, lower continental crust or topsoil horizons worldwide) [42, 43].

2.3 Geospatial Modeling

The most classical approach to represent the spatial distribution of the HMs concentrations in RD is the use of geospatial modeling techniques based on the spatial autocorrelation of the variable of interest (e.g., kriging, inverse distance weighting). These methods are widely used in environmental sciences and do not need other information than the HM concentrations in the RD sample and

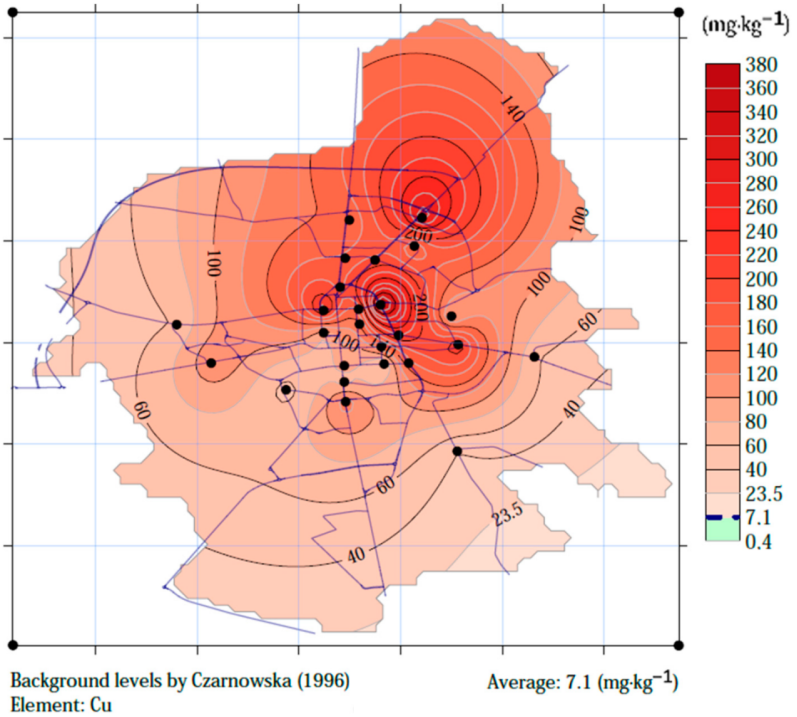
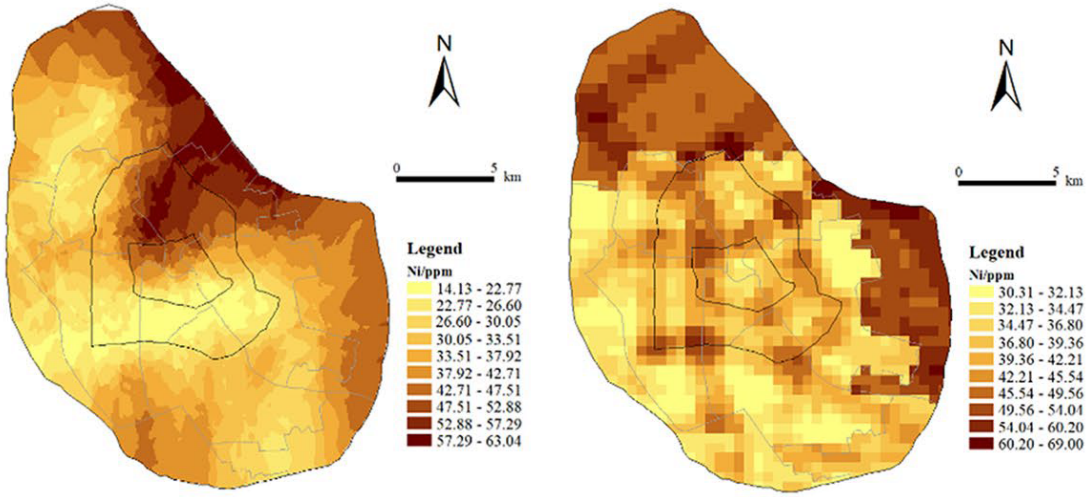


Figure 1 : Cu concentrations in road dust mapped over the city of Suwałki (Poland) by using Ordinary Kriging [23].



(a) by using Ordinary Kriging. **(b)** by using Multivariate Linear Regression.

Figure 2 : Ni concentrations in road dusts mapped over the city of Tianjin (China) [30].

the location of these samples to predict HM concentrations over the whole study area (e.g., [23, 26, 28, 29]). This approach allows the identification of pollution hot spots for the study area. However, the results cannot be extrapolated to a larger area without additional sampling. For the sake of illustration, Figures 1 and 2a present the mapped concentrations obtained using Ordinary Kriging (i.e., the most commonly used technique) for two case studies [23, 30]. Valid concerns about the use of this technique is that estimations outside the road network might be difficult to interpret and that isotropy is wrongly assumed [32]. In other words, the spatial autocorrelation of between samples locations might not be the same in all directions.

To address the concern of accurately modeling street-related pollution in urban areas, [32] propose the use of multivariate analysis as an alternative to ordinary kriging. The key feature of this alternative method is its ability to restrict the mapping of the pollution to the boundaries of the road network, thus improving the clarity of the resulting maps without values assigned to locations with no roads. The results of this approach for this case study are presented in Figure 3 using Geographically Weighted Regression (GWR), an extension of Multivariate Linear Regression (MLR) models allowing the coefficients to vary locally. The results for

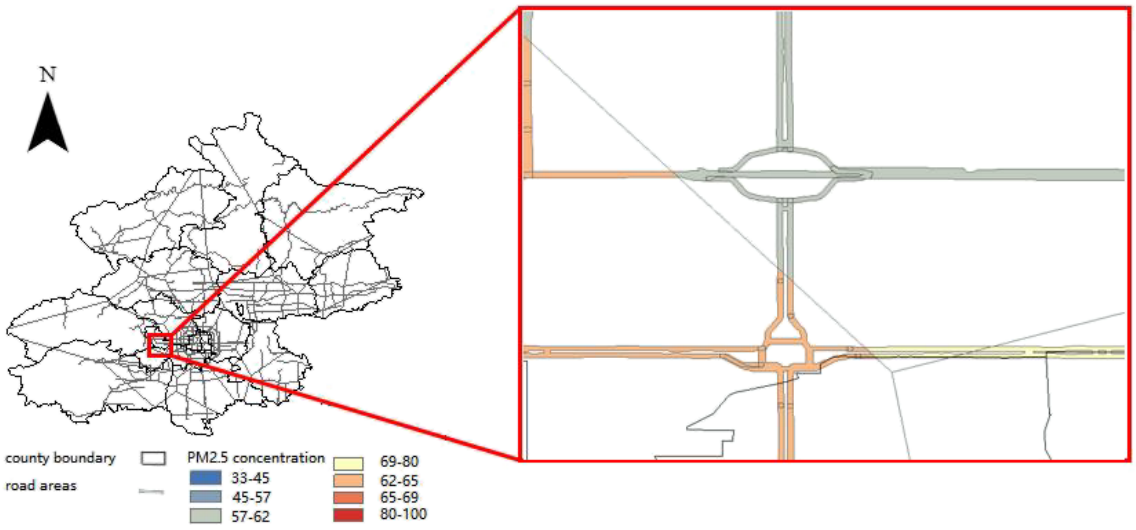


Figure 3 : Average PM2.5 atmospheric concentrations mapped over the of city of Beijing (China) by using Geographically Weighted Regression [32].

a case study using MLR are presented in Figure 2b. It is worth noting that the MLR approach used in this study did not restrict predictions to the road network, allowing a comparison with results obtained by using Ordinary Kriging presented in Figure 2 and revealing distinct patterns yielded by these differentapproaches.

Principal Components Regression (PCR)[2] is an extension of MLR models using Principal Components Scores (PCS) obtained trough PCA (on the independent variables) as predictors. The main advantage of this method is reduction of collinearity problems between predictors. Even if the PCR approach has the potential to be employed for the geospatial modeling of HMs concentrations in RD, it has only been used for HMs source apportionment as of now (e.g., [44]).

2.4 Proxies and relevant variables

Proxies (i.e., proxy variables) are extensively used in this study. They are variables that are spatially distributed and that have an explanatory power with respect to the observed HMs concentrations but that are not necessarily identified as sources of pollution (for all information about the sources see Section 2.1.1). However,

[2]Not equivalent to PCA-MLR models mentioned in Section 2.1.2 as the dimensionality reduction is applied to a set of independent variables and not the HMs concentrations themselves.

they are expected to be correlated to the sources of pollutions. A variety of such variables have been directly correlated to the accumulation of HMs concentrations in RD in previous literature (often to indirectly quantify a source). Other variables have been identified by simple observation in previous studies or can be reasonably expected to be proxies for known sources of pollution.

Data related vehicular traffic are expected to be relevant, as vehicles parts wear, tear and abrasion as well as exhausts emissions are known sources of pollution. For example, higher road classes are showing higher HMs concentrations in RD as they are expected to be subjected to higher traffic intensity [45, 46]. Other data related to the specificities road network are expected to bring information about the variability of HMs concentrations in RD (e.g., crossroads, speed limit) [47]. It is also notable that different road coatings types (e.g., asphalt, concrete) have been reported to contribute differently to the HMs concentration in RD [48]. As it has already been mentioned, the erosion and weathering of roads, infrastructures and buildings materials are known sources of pollution [16, 17, 34, 37]. Further information about the conditions and spatial distribution of these structures might be relevant for modeling.

Data that can be associated to different levels of human activities, which are correlated to HMs concentrations in RD are expected to be relevant for modeling. For example further information about industrial and commercial activities can be related to a distinct spatial distribution of the HMs concentrations [15, 29, 33, 36]. A proxy that might combine correlation to the spatial distribution of traffic and distinct levels of human activities might be the distance from the center of the urban area. It can be related to a decrease of the "urban characteristics" when moving away from the center of a city corresponding to diminishing HMs concentrations in RD [34]. Another formulation is that this proxy can be related to areas with different levels of urbanization [45] expected to show distinct HMs concentrations in RD further away from the center of the area.

As it has been defined earlier, identifying influential variables for modeling RD contamination in HMs is not equivalent to source apportionment in the sense that influential variables are possibly non-causal (even if correlated to variation in HMs concentrations in RD) but useful to make predictions. Accordingly, variables related to known impacting phenomenons are also falling into the category of influential variables (see Section 1). In this study, for the purpose of simplification, all influential variables will be denoted as proxies in the analyses.

Accordingly, all phenomenons related to the transfer of HMs in RD from other form of pollution are worthy of interest for modeling. Atmospheric pollutants concentrations might prove relevant as atmospheric deposition and resuspension can be correlated to HMs concentrations in RD [13, 49]. Another variable related to atmospheric pollution is the Street Canyon Ratio, a structural variable defining the geometry of a street (i.e., street height/width). If the ratio is high enough, the local conditions can concentrate HMs concentrations in RD by modifying the speed and direction of the wind [20–22]. The ratio in itself can be viewed as a proxy as it is only partial information concerning the street canyon effect. Meteorological data, especially average precipitation values for the week preceding the sampling are also expected to play an important role in wet deposition and water runoffs [19]. Data associated to mitigation strategies are also to be considered as they are associated to a direct interaction with the RD (i.e., removal).

3 Objectives

The main objective of this master thesis is to identify useful variables to create models predicting the geospatial distribution of the road dust (RD) contamination in heavy metals (HMs) in an urban environment. The articulation of a modeling approach and the creation of informative maps are secondary objectives of this master thesis. This approach was primarily built around the use of statistical linear modeling techniques. The maps produced will be of potential interest for local urban actors such as the cleaning services, the local authorities and the public services to help decision-making in regard to mitigation measures.

Multivariate Linear Regression was selected over other geospatial models presented in Section 2.3 for their ability to identify relevant proxies with clear associated significance and effects. Failing to communicate this would lower the scientific contribution of this study. Furthermore, other models would either fail to produce predictions restricted to the road network either fail to be potentially extended to the whole region. These two limitations would be detrimental to the implementation of mitigation strategies for the BCR. Therefore, this model was deemed the most appropriate for our study.

All statistical results were produced using the R software [50]. All processing and visual results concerning geospatial data were produced using the QGIS software [51]. All geospatial data that were needed for our study were available in open access and were provided by Belgian public services. The results presented in this master thesis were summarized and disseminated in a publication [52].

4 Materials and Methods

4.1 Study Area

The Brussels-Capital Region (BCR) is one of the three regions of Belgium. It is located in the center of the country and is composed of 19 municipalities. All these municipalities are forming a single urban agglomeration, making the BCR the most densely populated region of Belgium with 7528 inhabitants/km^2. It has a population of 1.2 million people and an area of 161.4 km^2. Green spaces (e.g., park, forests) represent 14% of the total area of the BCR. A large part of these green areas are listed as "Natura 2000" sites.

The BCR is the economic capital of Belgium and his zone of economic influence as a city exceeds the boundary of the region and extends over the two other regions[3]. This affects the transport network, the associated infrastructures and the daily traffic. The BCR is serviced by the STIB-MIVB (busses, trams, subways) and SNCB (trains) but also connected toward the Flemish region by De Lijn (busses) and Walloon region by the TEC (busses). The heavy traffic and the subsequent need for this complex mobility network are partly explained by the commuters (accounting for one out of two workers in the BCR and representing approximately three hundred sixty thousand people)[4].

The BCR industrial past also shaped the city and his transport infrastructures such as the railway and the Brussels canal. Nowadays, few industrial enterprises remain and represent only 3% of the PIB of the region. The BCR mainly hosts Small and Medium-sized Enterprises (SMEs) and tertiary sector industries (shops, hotels). The Brussels harbor is one of the few industrial enterprises still active today and currently employing 12.000 peoples [55].

The BCR is also a major administrative center of Belgium as it hosts his own regional

[3]North : Flemish region, South : Walloon region
[4]The majority of figures provided in this section are provided by the IBSA [53, 54].

government and parliament but also the federal government and parliament, the Flemish government and the French Community government, along with the Europeans Institutions and the NATO administration.

The Anderlecht municipality (AM) is one of the 19 municipalities of the BCR. For this study, the AM was chosen among all the other municipalities as it represent a wide variety of conditions that are typically found in the BCR (e.g., all road classes and nearly all land uses are represented). Furthermore, this municipality represents 10% of both the area and population of the BCR.

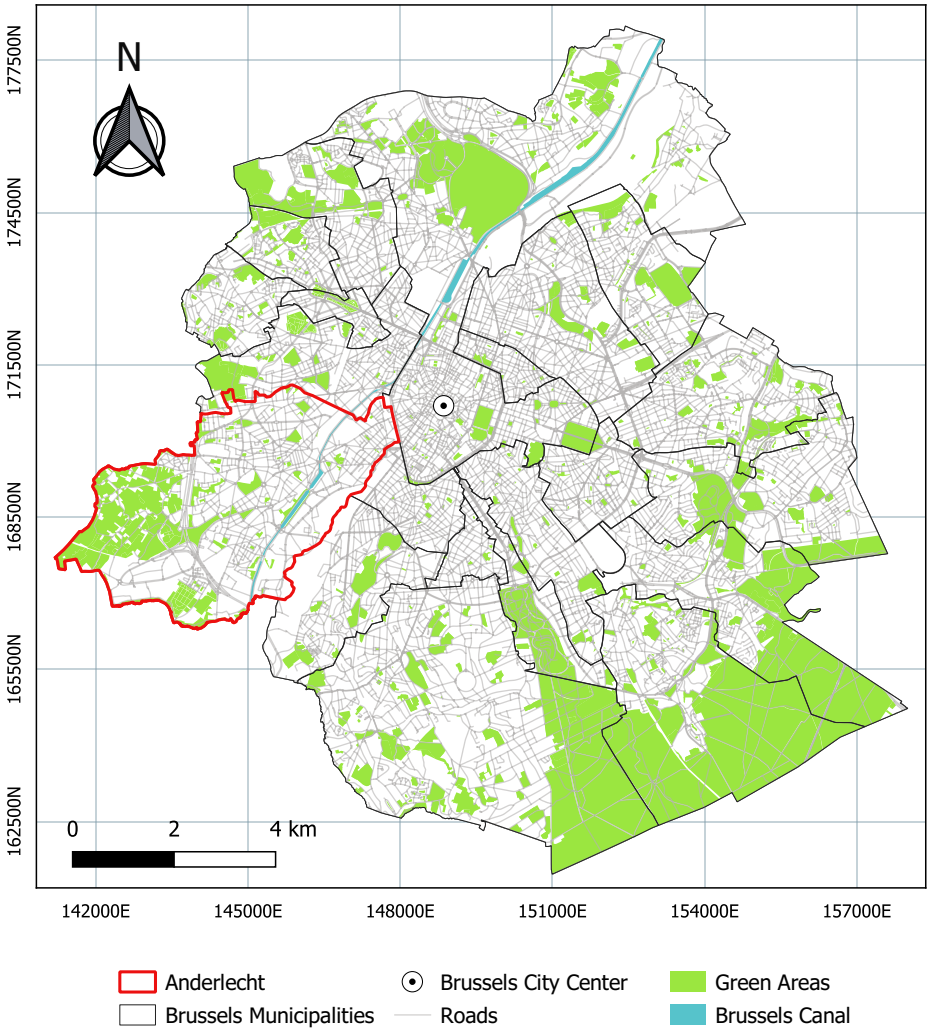


Figure 4 : Map of the Brussels-Capital Region (BCR). Data source : URBIS.

4.2 Sampling plan

The first sampling campaign took place in 2019. It was followed by two smaller campaigns in 2020 and 2021. These campaigns had to be conducted between April and September[5] in order to get dry samples usable for analysis. Accordingly, the sampling had to take place at least 24 hours after a raining event.

The major campaign took place in 2019 and was the main input for the identification of the relevant proxies and for the prediction. However, these proxies were not known prior to the study. Therefore, designing a balanced experimental sampling plan (i.e., a sampling that would equally account for the relevant variables) in regard to those relevant proxies was not a realistic possibility. However, the sampling locations were chosen in order to mitigate the possible imbalances by spreading them relatively homogeneously over the whole municipality and also by taking into account the diversity of road types and urban surroundings. The 2020 and 2021 campaigns also contributed to the modeling but were mainly conducted in order to assess the spatial dependence of the regression residuals at a short distance.

The 2019, 2020 and 2021 campaigns allowed us to collect 100, 20 and 8 samples, respectively. The locations of these samples are displayed on Figure 4.2. For 2019, 2 samples of the first year had to be discarded as they showed abnormally high concentrations values. These high values could be explained by their peculiar location (i.e., inside a tunnel and near a railroad). The 2020 and 2021 sampling campaigns accounted for two crossroads (2x4 points), three small transects (3x4 points) and a bigger transect along the canal (1x8 points). The crossroads were added to the main models as 2 points (averaged HMs concentrations) instead of 8 for the main analysis.

[5]The driest periods of the year in Belgium.

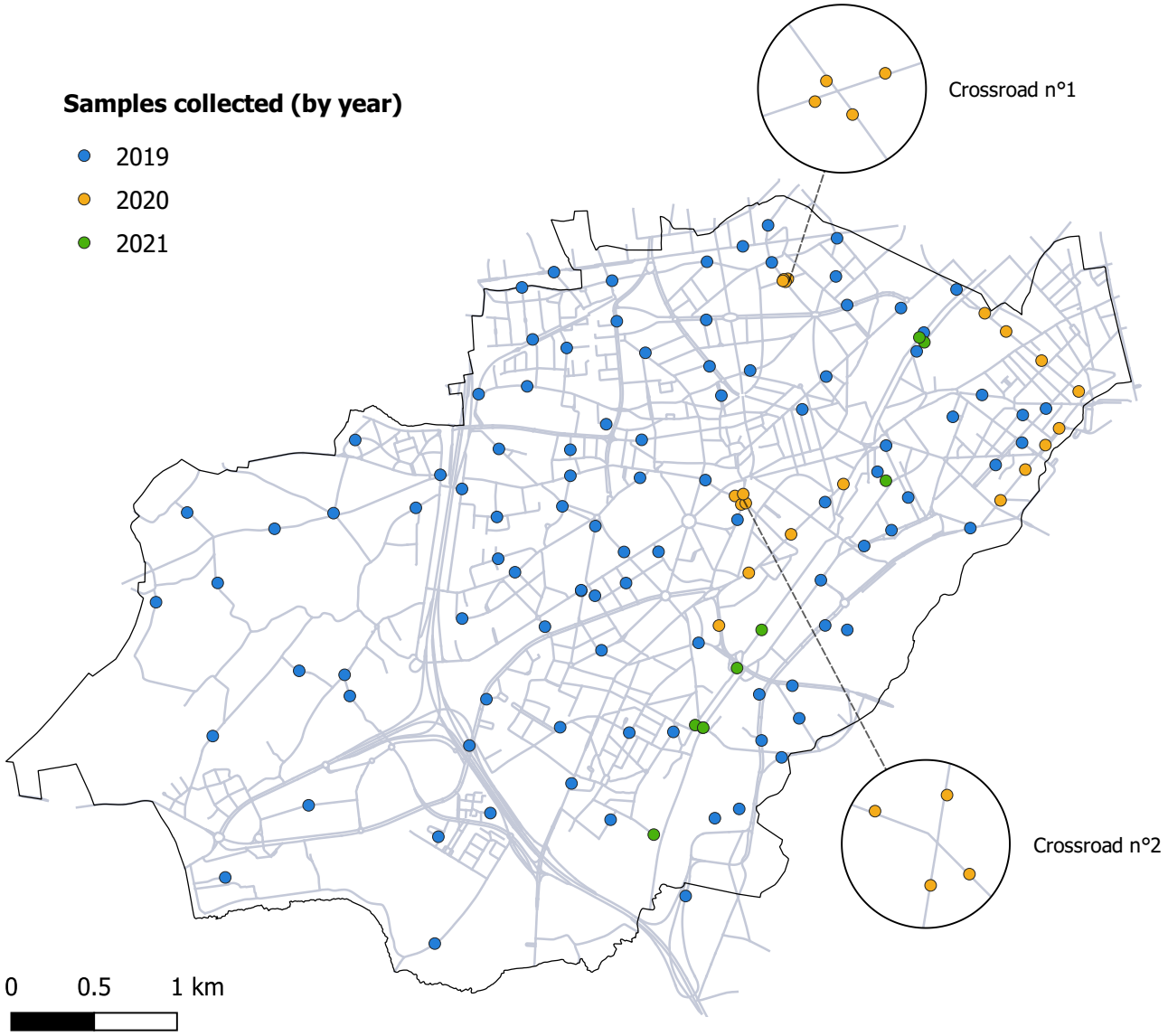
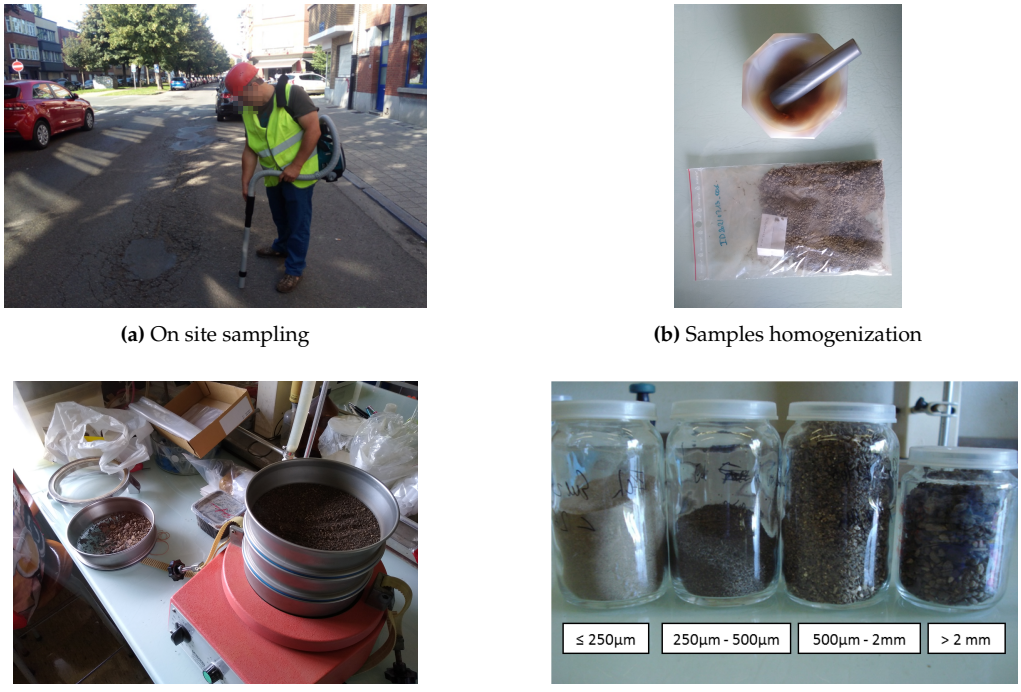


Figure 5 : Locations of the samples in the Anderlecht municipality, with the initial 100 samples of 2019 (in blue), the 20 additional samples of 2020 (in orange) and the 8 additional samples of 2021 (in green). The 8 locations sampled in two crossroads are highlighted in the zoomed parts of the figure.

4.3 Sampling and Extraction methods

At all sampled locations, RD was collected over an area of 6 $[m^2]$ using an industrial-grade MAKITA vacuum cleaner[6]. The samples were dried overnight at 50°C in a dry oven. The dried matter was homogenized and then sieved to obtain different size fractions. The physical fractioning was made using a control sieve shaker[7].

The finest fraction of the samples ($\varnothing < 250[\mu m]$) was analyzed for the six metals concentrations : 1 $[g]$ of the fraction was extracted in 10 $[ml]$ of aqua regia solution[8] for two hours at a temperature above 300°C. After this hot extraction, the mixture was filtered[9]. This operation was repeated three times for each RD sample. The filter was rinsed and the extraction and rinsing liquids were added to Milli-Q quality water in a 50 $[ml]$ volumetric flask.



(a) On site sampling

(b) Samples homogenization

(c) Physical fractioning

(d) Fractions obtained

Figure 6 : Collection and pre-processing of the RD samples. (photographs provided by G. Diélie)

[6]18V Li-ion 100 mbar DCL 501Z
[7]Analysette by FRITSCH
[8]Mixture of nitric acid (HNO_3) and hydrochloric acid (HCl) with a molar ratio of 1:3
[9]Filter n°1 by WHATMAN

(a) Extraction **(b)** 50ml solutions **(c)** ICP-OES

Figure 7 : Extraction of the finest RD fractions and concentration measurements. (photographs provided by G. Diélie)

The solutions were then injected in the Inductively Coupled Plasma with an Optical Emission Spectrometry[10] (ICP-OES) and the various HMs concentrations (in $[\mu g/g]$) were obtained for the three solutions of each sample. Those measures were averaged to obtain a single concentration value for a given sampled location.

4.4 List of proxies

A large dataset of proxies was assembled using data from various sources. The detailed list of proxies is provided in Table 2. The upper part of the table refers to the continuous proxies and the lower part to the categorical ones (all proxies displayed were either cited by name or easily associated to concepts mentioned in Section 2.4). Road network and land use data were provided by the IT service of the region (URBIS), atmospheric concentrations of several pollutants and infrastructural factors were provided by Brussels Environment (BE), socio-economic and infrastructural factors as well as parking related data were provided by the Belgium National Institute for Statistics (IBSA), traffic related measures were provided by Brussels Mobility (BM), meteorological data were provided the Royal Meteorological Institute (IRM) and street-sweeping related data were provided by the Anderlecht municipality (AM). Road coating categorization and visual assessment of its condition were made during the collection of the samples.

[10]vista pro by VARIAN

Table 2 : List of the available proxies. First part of the table refers to continuous-valued proxies, whereas second part refers to categorical-valued ones. "a" type proxies are the ones related to traffic. "b" type proxies are the ones related to roads, buildings and infrastructures. "c" type proxies are the one related to human activities, mainly industrial and commercial. "d" type proxies are related to impacting phenomenons.

Proxy	Details	Unit	Provider	type
Distance to BCR center	center of BCR: Great Market place	km	URBIS	a, c
Crossroad distance	closest street node	m	URBIS	a
Black carbon	peak time, bulk time	$\mu g/m^3$	BE	d
Atmospheric pollutants	Cd, Cr, Cu, PM_{10}, $PM_{2.5}$	$\mu g/m^3$	BE	d
Street canyon index	average building height/street width	-	BE	d
Population density	by district	km^{-2}	IBSA	c
Road occupation rate	by district, 8-9 am and 17-18 am	-	ISBA	a
Roadside parking Occupation rate	by district, 5-7 am and 10-12 am	-	ISBA	a
Waterproof surface ratio	by district	-	ISBA	d
Buildings older than 1961	by district, ratio	-	IBSA	b
Number of shops	by district, per 1000 people	-	IBSA	c
Employment rate	by district	-	IBSA	c
Sedentary rate	by district	-	IBSA	c
Number of vehicles (MUSTI)	daily, rush hours	-	BM	a
Distance to shopping area	closest shopping area	m	BM	c
Average & total precipitation	week before sampling	mm	IRM	d
Proxy	Classes	Unit	Provider	type
Road coating type	asphalt, pavement, mixture	-	samples	b
Road condition	new, good, degraded	-	samples	b
Speed Limit	30 km/h, 50 km/h, 70 km/h	-	URBIS	a
Land use	green area, densely populated, industrial, residential, mixed	-	URBIS	c
Road hierarchy	metropolitan road, main & interdistrict road, district collector & district road	-	BM	a
Regular Street Sweeping	yes, no	-	AM	d
Street Manager	municipality, region	-	AM	d

4.5 Statistical analyses

In this section, the linear models used for geospatial modeling and the precise procedure for the selection of the most relevant proxies will be presented. A source apportionment method and a pollution index for contamination assessment will also be briefly detailed.

4.5.1 Simple Linear Regression

When statistically relating a single continuous y variable (i.e., the concentration of a given HM) and a single continuous x variable (i.e., a continuous proxy for our study), the most parsimonious model is the **simple linear regression** where the expected relationship between y and x is given by

$$E[y] = \beta_0 + \beta_1 x \tag{1}$$

where β_0 and β_1 are the intercept and slope that needs to be estimated from the data. These estimated values $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are easily estimated using an ordinary least squares approach, and the best predicted value of y (in the mean square sense) is given by

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x \tag{2}$$

where $\widehat{y}$ is the estimated value of y for a given x.

4.5.2 ANOVA I

If y is still a continuous variable but x is now a categorical one, the simplest statistical model is the ANOVA I. This statistical model can be seen as a special case of linear regression where the independent variable is a set of binary indicators

representing the different categories with[11]

$$E[y] = \beta_0 + \sum_{k=1} \beta_k x_k \tag{3}$$

where β_0 is the average value for a chosen reference category, β_k is the difference between the average value for the k^{th} category and the reference category, while x_k is a binary variable that take the value 1 if the observation belongs to the j-th category and 0 otherwise. The parameters of this model can also be estimated through an ordinary least squares methods so that

$$\widehat{y} = \widehat{\beta}_0 + \sum_{k=1} \widehat{\beta}_k x_k \tag{4}$$

where $\widehat{y}$ is the estimated value of y for a given k category of x.

4.5.3 Mutivariate Linear Regression

When a single continuous variable y needs to be related to several variables of continuous and categorical types, the simplest model is the multivariate linear regression (MLR). This model can be viewed as a combination of the two precedent models. The relationship between y and the several variables is given by

$$E[y] = \beta_0 + \sum_{i=1} \beta_i x_i + \sum_{j=1} \sum_{k=1} \beta_{jk} x_{jk} \tag{5}$$

where β_0 is the theoretical intercept, β_i are the slopes that needs to be estimated for the continuous variables, β_{jk} are the differences between the average value for the k^{th} category and the corresponding reference. The intercept represents the expected value of the dependent variable y when all the continuous independent

[11]This formulation is unorthodox but facilitates the analogy with linear regressions and eases the presentation of the next model.

variables are set to 0 and all the categorical independent variables are set to their reference level. The parameters of this model can also be estimated through an ordinary least squares methods with the formula

$$\widehat{y} = \widehat{\beta}_0 + \sum_{i=1} \widehat{\beta}_i x_i + \sum_{j=1} \sum_{k=1} \widehat{\beta}_{jk} x_{jk} \tag{6}$$

where $\widehat{y}$ is the estimated value of y for a set of x.

4.5.4 Coefficient of determination

The most classical metric to assess of the performance of linear models is the coefficient of determination (R^2) as given by

$$R^2 = 1 - \frac{RSS}{TSS} \tag{7}$$

where RSS is the residual sum of squares and TSS is the total sum of squares. This metric ranges between 0 and 1 and is quantifying (in percents) the part of the variance explained by the model. However, this metric is known to increase with the number of independent variables added to the model. In order to avoid this effect, the adjusted coefficient of determination (R_a^2) is to be preferred with

$$R_a^2 = 1 - \frac{RSS \ / \ (n - k - 1)}{TSS \ / \ (n - 1)} \tag{8}$$

where again RSS is the residual sum of squares, TSS is the total sum of squares, n is the number of samples and k is the number of parameters. Its values are always lower or equal to the R^2 for the same model. The adjustment provided is particularly important when comparing models with different numbers of independent variables (and thus different number of parameters).

4.5.5 Two-step Screening Method

Identifying the most relevant variables for modeling is done through a procedure involving two screenings steps. The first screening step aim is to assess the significance of the proxies when taken individually but also eliminate a certain amount of proxies that need to be discarded for practical reasons. The main reason for this is that some variables cannot be interpolated for all the road segments in the municipality, which is a problem for the mapping part of the project. This step allows us to verify if these variables are relevant for predicting HMs concentrations. In this pursuit, a single model is created for each proxy and HM (simple linear regression for continuous variables, ANOVA I for categorical variables) and their p-value is extracted. The average p-value across all HMs for each variable is used to rank their relevance for prediction.

The second screening step aim is to select the best subset of proxies that will be used in the MLR models (one per HM). This part of the procedure is not a simple as including the most significant proxies of the first step in the models. The inclusion of multiple proxies in the model might highlight significance not appearing in the simple models. The selection in itself is done by using a forward stepwise selection procedure. In other words, the best model is found by adding one by one the most relevant proxy (i.e. the proxy minimizing the residual sum of square to the greatest extent) left in the subset of candidate. This additive process stops when the quality of the model is not improving substantially by the addition of a new proxy. The metric chosen to assess this quality is the Akaike information criterion (AIC) [56] with

$$AIC = n \ \ln(RSS \ / \ n) + 2(k + 1) + C \tag{9}$$

where n is the sample size, k is the number of parameters and RSS is the residual sum of square of the model, while C is a constant that can be ignored for the

sake of comparing models fitted on the same data. This metric keeps track of two important qualities of the model. Namely, its performance (low prediction variance) and its relative simplicity (low number of variables used to make the prediction). When adding any proxy left in the dataset increases the AIC instead of lowering it (i.e., the cost of adding a variable is greater than the benefit of the associated residual sum of square decrease), we consider that the model has included all the relevant proxies in the tested dataset. This part of the procedure is applied with the *ols_step_forward_aic()* function from the olsrr package of the R software [57].

In addition to this automated procedure, we look at the adjusted generalized variance inflation factor[12](AGVIF) of each proxy in a model with

$$AGVIF_i = (GVIF_i)^{1/2df} \tag{10}$$

where df is the degree of freedom associated to the i^{th} independent variable and $GVIF_i$ given by

$$GVIF_i = \frac{det(\mathbf{A}) \times det(\mathbf{B})}{det(\mathbf{C})} \tag{11}$$

where A is the correlation matrix of the i^{th} proxy, B is the correlation matrix of the other proxies present in the model and C is the correlation matrix for all the proxies in the model. The AGVIF allows us to identify multicollinearity between the different predictors. If the value is greater than 2.236 for a proxy, it is an indication of multicollinearity and it might be necessary to discard it. The reason why this is a major concern is that the MLR models require the predictors to be independent in order to produce meaningful results.

[12]This metric is derived from the variance inflation factor (VIF). More details on AGVIF can be found in [58, 59].

4.5.6 Variograms

Besides the use of the adjusted coefficient of determination that allows us to assess which part of the variance can be explained from the proxy at hand, another important aspect is linked to the property of the regression residuals. Ideally, if there is no more information that could help to improve the model, one would expect that these residuals are not spatially correlated. This would correspond to the situation where these residuals are pure noise (i.e, the part of the variability that cannot be explained due to the impact of various and unpredictable random effects). On the contrary, if a spatial correlation does exist between these residuals, it is the indication that a relevant and spatially structured proxy might have been omitted in the model (although this proxy cannot be identified based on this evidence). The spatial dependence can be expressed by the theoretical variogram $\gamma(\|h\|)$, with

$$\gamma(\|h\|) = \frac{1}{2} Var[\varepsilon(x+h) - \varepsilon(x)] \tag{12}$$

where $\gamma(\|h\|)$ is the variogram value for a distance h between two locations x and $x+h$, while $\varepsilon(x)$ and $\varepsilon(x+h)$ are the regression residuals at these locations. The variogram can be empirically estimated by a moment-based approach with binning. In other words, the pairs of points at hand are placed in bins (i.e., groups) corresponding to their separation distance h with a tolerance δ and, for each bin, $\gamma(\|h\|)$ is estimated by

$$\hat{\gamma}(\|h \pm \delta\|) = \frac{1}{2N(h \pm \delta)} \sum_{i=1}^{N(h \pm \delta)} [\varepsilon(x_i) - \varepsilon(x_i + h)]^2 \tag{13}$$

where $N(h \pm \delta)$ is the number of pairs in the bin centered around h, while $\varepsilon(x_i)$ and $\varepsilon(x_i + h)$ are the regression residuals for the i^{th} pair of points in this bin. This approach allows for a precise estimation of the spatial dependence when

the points are not located on a regular grid (i.e., when the pairs at hand are not automatically grouped, since the distances h between them are all different). The next step toward the assessment of the spatial dependence is to approximate the empirical variograms obtained with a parametric model. In our study, $\gamma(\|h\|)$ was approximated for each HM by an exponential variogram model combined with a nugget effect (i.e., discontinuity at the origin[13]) given by

$$\gamma(\|h\|) = \alpha_1 \delta_{(\|h\| \neq 0)} + \alpha_2 (1 - \exp(-3\|h\|/r)) \tag{14}$$

where α_1 is the parameter associated to the uncorrelated part of the variance of the residuals and $\delta_{(\|h\| \neq 0)}$ is the so-called nugget effect while α_2 is the parameter corresponding to the correlated part of the variance of the residuals and r is the range of this correlation. Accordingly, $(\alpha_1 + \alpha_2)$ is the total variance of the residuals. From these numbers we can define $\alpha_2/(\alpha_1 + \alpha_2)$ as the part of the variance attributed to spatially correlated variations.

4.6 APCS-MLR receptor model

In order to provide the contribution of the sources (including natural ones) of an HM in RD, an ACPS-MLR receptor model can be used. More details on the exact procedure and PCA can be found in [16]. After the dimensionality reduction, the Principal components scores (PCS) obtained are transformed into absolute principal components scores (APCS) used in a MLR model with the HM concentrations as the dependent variables given by

$$E[y] = \beta_0 + \sum_{i=1} \mathrm{APCS}_i \, \beta_i \tag{15}$$

that can be viewed as a simplified version of the equation 5 with The APCS as the

[13]It can be viewed as an intercept

independent continuous variables. The mean contribution (in %) of each factor is given by

$$x^{th} \text{ factor contribution} = \frac{\overline{\text{APCS}_x}\,\widehat{\beta}_x}{\sum_{i=1}\overline{\text{APCS}_i}\,\widehat{\beta}_i} \tag{16}$$

where $\overline{\text{APCS}_x}\,\widehat{\beta}_x$ is the average contribution of the x^{th} factor to the HM concentrations (in $\mu g/g$) and $\sum_{i=1}\overline{\text{APCS}_i}\,\widehat{\beta}_i$ is the average contributions of all factors combined to the HM concentrations (i.e. equal to $\overline{y}$).

4.7 Pollution Indices

4.7.1 Contamination Factor

The Contamination Factor (CF) was proposed by [25] in order to assess the pollution of an aquatic sediment in a single element. This index has been used in recent studies to assess contamination of RD in HMs [17, 36, 60]. The CF is defined as

$$CF = C(x)_{sample} \,/\, C(x)_{background} \tag{17}$$

where $C(x)_{sample}$ is the concentration of the heavy metal in the sampled sediment and $C(x)_{background}$ is the concentration of the heavy metal in the sediment before industrialization (i.e, the background). This is an alternate way to look at the pollution in the samples rather than the absolute concentrations values. Table 3 displays the classes proposed by [25] to interpret the contamination level of the sediment.

Table 3 : Pollution assessment using the contamination factor by [25].

Range	Contamination assessment
$CF \leq 1$	Low
$1 \leq CF < 3$	Moderate
$3 \leq CF < 6$	Considerate
$6 < CF$	Very high

4.7.2 Enrichment Factor

The Enrichment Factor (EF) is also a commonly used pollution index to assess contamination of RD in HMs [41, 61], with

$$EF = \frac{C(x)_{sample}}{C(y)_{sample}} \Big/ \frac{C(x)_{background}}{C(y)_{background}}$$

where $C(y)_{sample}$ and $C(y)_{background}$ are the concentrations value of a normalizing element (i.e., an element identified as unaffected by anthropogenic activities) in the sample and the background, respectively. This normalizing element can be chosen among a variety of elements such as iron (Fe), aluminum (Al), calcium (Ca), titanium (Ti), scandium (Sc) and manganese (Mn) [42]. The enrichment factor can be seen as an adjusted version of the contamination factor, allowing for a better characterization of the pollution caused by human activities. Unfortunately, as we do not have at hand the concentrations of a normalizing element, we are no able to present EF results for our study.

5 Results and Discussion

5.1 Exploratory Analysis

The main statistics of the collected samples concentrations (in mg/kg) for all HMs studied are presented in Table 4. The first part of the table ($n = 98$) corresponds to the 2019 campaign, while the second part ($n = 28$) corresponds to the 2020-2021 campaigns. The mean concentrations observed for the 2020-2021 campaigns are strictly higher than those of 2019 campaign. This does not indicate an increase of the concentrations over the years as the values and categories of the proxies are not comparable to those of 2019 at these sampled locations (e.g., on the average the new points are closer to the center of the BCR). In order to avoid seasonal effects that could prevent comparisons, the 2019, 2020 and 2021 campaigns took place on the same period of the year (see Section 4.2).

Table 4 : Main statistics for the n concentrations (in mg/kg) of the various HMs collected during the 2019 campaign (first part) and the 2020-2021 campaigns (second part).

2019 campaign ($n = 98$)						
	Cd	Cr	Cu	Ni	Pb	Zn
Min.	0.20	2.48	5.23	5.09	4.16	21.43
Median	2.69	82.07	159.41	41.25	75.46	330.99
Mean	2.66	94.98	210.74	47.83	117.39	353.55
Max.	6.71	473.24	1076.80	218.22	1671.75	1511.78
2020-2021 campaigns ($n = 28$)						
	Cd	Cr	Cu	Ni	Pb	Zn
Min.	1.71	49.58	46.68	26.18	47.24	163.32
Median	6.44	148.55	304.90	88.20	166.69	456.37
Mean	5.96	158.34	360.52	85.13	202.03	478.67
Max.	8.46	288.60	994.91	171.69	624.37	1018.95

As seen from the diagonal of Figure 8, the log-transformed HMs concentrations distributions are all relatively close to normality and moderately to highly correlated between HMs (They are all statistically significant ; p-values ≤ 0.01). From now on, the log-concentrations will be kept for the statistical analyses and modeling and back-transformed on the maps for clarity. These observations are consistent with other studies [17, 23, 36]. As expected, the highest correlation (r=0.92) is observed between Cr and Ni concentrations as both those HMs are found in stainless steel, an alloy used in many vehicles accessories and a well known source of contamination through their wear and tear (see Section 2.1). Cd is the HM showing the weakest correlations with other HMs. This can be explained by the low concentrations of Cd and the associated higher analytical uncertainty compared to other HMs.

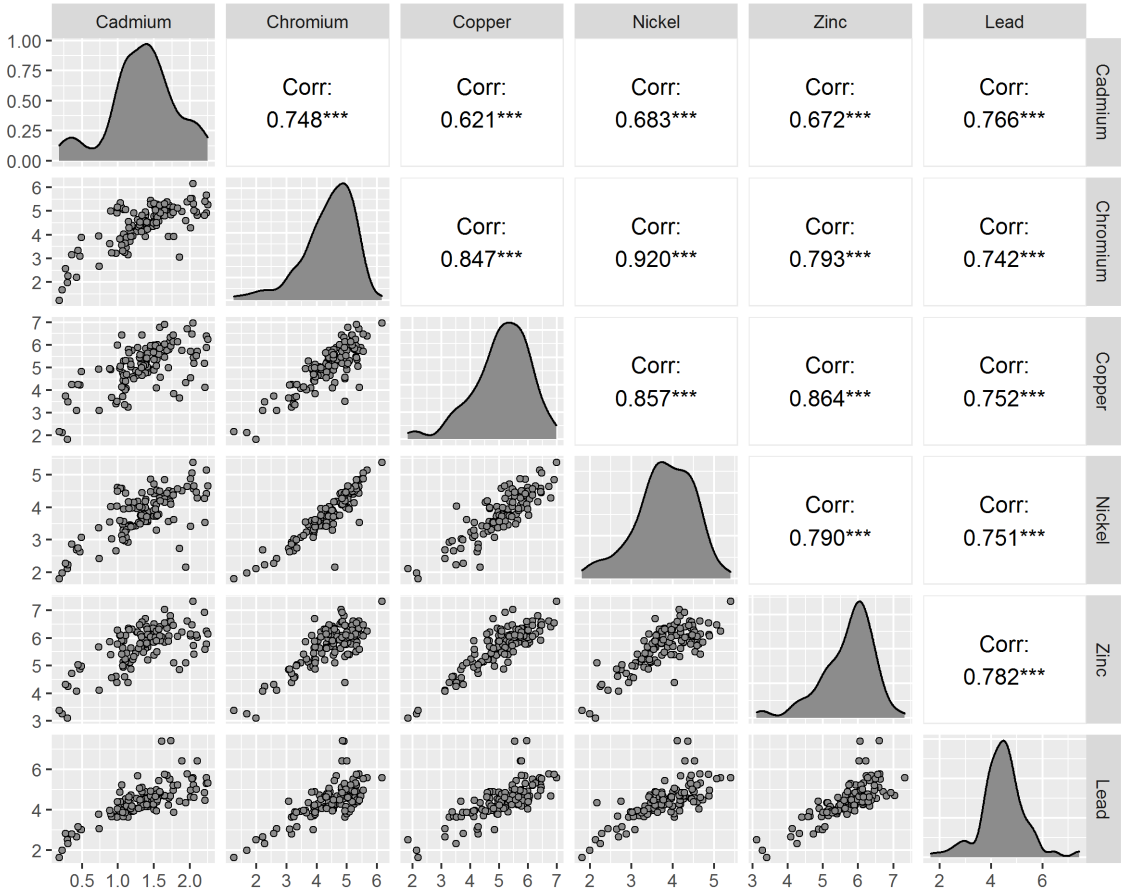


Figure 8 : distribution for the 126 samples in HMs concentrations (in natural logarithm), along with the corresponding scatter plots and Pearson correlation coefficients for all pairs.

5.2 Source apportionment with an ACPS-MLR receptor model

The suitability of the HMs concentration was assessed by Barlett's sphericity and Keyer-Meyer-Olkin tests. They confirmed the suitability of the data for a factor analysis [41]. The results of the APCS-MLR approach, displayed on Table 5 indicate that factors 1, 2 and 3 contribute (in different proportions) to all HMs. Factors 4 and 5 demonstrate notable contributions only to Pb and Cu, respectively. No notable contribution was found for factor 6.

Linking the different factors to known sources is a difficult task, even based on previous studies, since the contributions varies between different areas and countries. Additional studies in the BCR and possibly other Belgian cities might be needed to interpret those results with more confidence. However, these results are coherent with the fact that the six HMs studied are sharing common main anthropogenic sources (See Section 2.1.1). This is also coherent with the exploratory analysis, as the profile of Cadmium, exhibits a distinct contributions profile (68.7 % contribution from factor 3), while Ni and Cr are displaying similar contributions profile across the first 3 factors. It is also worth noting that factor 2 makes substantial contributions to Cu and Zn (33.2 % and 61.7 %, respectively), which are two strongly correlated HMs (see Section 5.1).

Table 5 : ACPS-MLR receptor model for our study. The values corresponds to the estimated contribution of a factor to the HMs concentrations (only the contributions over 10 % are displayed).

	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6
Cd	16.2 %	12.0 %	68.7 %	-	-	-
Cr	54.2 %	18.1 %	26.3 %	-	-	-
Cu	37.9 %	33.2 %	17.2 %	-	10.4 %	-
Ni	51.0 %	19.5 %	21.8 %	-	-	-
Pb	15.9 %	2.7 %	26.7 %	36.6 %	-	-
Zn	17.2 %	61.7 %	16.6 %	-	-	-

5.3 Pollution Assessment with Contamination Factor

Table 6 presents the frequency of samples that belongs to the different CF classes for each HM (See Table 3 for more details on the classes). Since we do not have local background values at hand, the upper continental crust (UCC) average composition is used as reference. The values for Cd, Cr, Cu, Ni, Pb and Zn are respectively 0.09, 92, 28, 47, 17 and 67 $\mu g/g$ [62]. It is important to interpret these results with caution as the choice of a background does not account for local variations. Most samples in our study are considered very highly contaminated in Cd. Cu, Pb and Zn are also highly contaminated for a significant number of samples according to the results. Cr and Ni distributions across the different CF classes are also relatively similar. The [Cd > Cu, Pb, Zn > Cr, Ni] hierarchy based on pollution indices has been observed for other urban areas [26, 63] with a similar high contamination of RD in Cd despite the lower concentrations in comparison to other HMs (Cd background value : 0.09 $[\mu g/g]$; average Cd concentration in 2019 samples : 2.67 $[\mu g/g]$). However, this hierarchy is not shared for all studied urban areas [24, 61]. Similarly to source apportionment, further studies in the BCR and other Belgian cities could facilitate the interpretation of those results.

Table 6 : RD pollution assessment results with the classes proposed by [25] for CF.

	Low	Moderate	Considerate	Very High
Cd	-	1.70 %	4.20 %	94.2 %
Cr	48.3 %	50.0 %	1.70 %	-
Cu	5.80 %	15.8 %	25.8 %	52.5 %
Ni	51.7 %	45.8 %	2.50 %	-
Pb	5.80 %	18.3 %	37.5 %	38.3 %
Zn	4.20 %	20.8 %	32.5 %	42.5 %

5.4 First screening step : Individual assessment of the Proxies

Table 7 presents the outcomes of the first screening step where each proxy is assessed individually based on simple linear regression and ANOVA I (see section 4.5.5). The table is divided in three different sections corresponding to the various levels of information completeness we have at hand for these proxies. The first section include all proxies whose values are known at any arbitrary location in the municipality (i.e., all sampled and prediction data points). The second section consists of all proxies for which we lack information for certain locations as the corresponding layer of geospatial data does not cover all the road segments. The third section comprises two proxies that are only known at the sampled locations as they were collected during the sampling itself. The six first columns present the p-value associated to the corresponding proxy and HM. The second-to-last column presents the p-value averaged over all HMs for a given proxy (based on the values in the first six columns). In each section, the proxies are ranked according to their averaged p-value. The last column corresponds to the number of sampled locations for which we know the value of the proxy (for the 2019 campaign).

The lack of information for the proxies in the second and third sections will prevent these proxies to be passed to the next step of the screening procedure. The decision to exclude these proxies originates from the fact that the estimation of the MLR models coefficients would require an alternate methodology to an ordinary least squares approach (which assumes the absence of missing values for the independent variables). However, Even with these alternative approaches, the estimation of the coefficients could be inaccurate and lead to a misrepresentation of the relationship between the proxies and the HMs concentrations. Conveniently, the majority of these proxies were not found significant ($\overline{p_v} > 0.05$) except for "Employment rate" and "Speed limit". However, the information loss is limited for "Employment rate" as it is strongly correlated to "Roadside parking occupation (10-12 am)" (r=-0.85), a proxy that will be used in the second step of the screening procedure. Speed limit on the other hand, is not very well correlated to the proxies

Table 7 : p-values for the simple linear regression models (continuous proxies) and ANOVA I models (categorical proxies) that relate the concentration of each HM with each proxy.

Proxy	Cd	Cr	Cu	Ni	Pb	Zn	$\overline{p_v}$	n
Distance to BCR center	4.2e-5	2.3e-9	1.5e-9	1.3e-10	1.2e-8	2.3e-13	7.0e-6	98
Roadside parking occupation (5-7am)	5.4e-4	5.9e-5	1.5e-6	4.6e-5	1.4e-4	3.7e-7	1.3e-4	98
Land use	9.3e-4	4.3e-5	5.6e-9	9.0e-5	5.4e-6	8.7e-9	1.8e-4	98
Waterproof surface ratio	3.9e-3	2.2e-6	1.0e-8	6.6e-7	9.9e-6	1.4e-10	6.5e-4	98
Distance to shopping area	0.022	1.4e-5	5.7e-8	2.4e-5	4.0e-3	2.9e-7	4.0e-3	98
Black carbon (peak time)	0.025	2.9e-3	9.7e-5	5.0e-4	1.4e-3	2.3e-3	5.0e-3	98
Street canyon ratio	0.037	1.2e-3	1.5e-3	6.1e-4	3.3e-3	2.3e-4	7.0e-3	98
Black carbon (bulk time)	0.050	3.7e-3	1.2e-4	6.0e-4	1.4e-3	2.4e-3	0.010	98
Population density	0.323	2.8e-3	1.1e-3	7.3e-3	0.036	1.9e-3	0.062	98
Road hierarchy	0.051	1.9e-3	1.5e-3	3.8e-3	0.612	0.016	0.114	98
Roadside parking occupation (10-12am)	0.939	0.075	8.3e-3	0.042	0.292	8.5e-3	0.227	98
Regular Street Sweeping	0.445	0.065	0.114	0.073	0.714	0.031	0.240	98
Average precipitation	0.691	0.739	0.294	0.173	0.117	0.106	0.353	98
Crossroad distance	0.694	0.634	0.039	0.369	0.624	0.033	0.399	98
Buildings older than 1961	0.191	0.702	0.575	0.672	0.421	0.625	0.531	98
Employment rate	5.1e-4	4.14e-3	9.8e-4	3.4e-3	1.2e-4	4.0e-4	1.6e-3	78
Speed limit	1.6e-3	0.013	6.7e-3	0.055	0.026	7.7e-3	0.018	95
Road occupation rate	0.048	0.093	5.4e-3	0.039	0.119	0.013	0.053	44
Street Manager	0.058	0.020	0.016	6.6e-3	0.306	0.026	0.072	93
Number of shops	0.161	0.028	0.153	0.027	0.131	0.020	0.087	67
Atmospheric pollutants (Cu)	0.149	0.168	0.281	0.067	0.024	0.022	0.118	87
Atmospheric pollutants (Cr)	0.318	0.264	0.437	0.132	0.061	0.038	0.208	87
Atmospheric pollutants (Ni)	0.251	0.295	0.275	0.247	0.185	0.047	0.217	87
Road occupation rate (17-18 am)	0.376	0.069	0.117	0.267	0.444	0.220	0.249	96
Sedentary rate	0.952	0.326	7.0e-3	0.221	0.891	0.041	0.406	78
Road occupation rate (8-9 am)	0.616	0.121	0.246	0.321	0.836	0.460	0.433	96
Road coating	0.138	7.4e-3	8.8e-3	0.021	0.134	0.099	0.068	98
Road condition	0.327	0.868	0.329	0.408	0.012	0.125	0.346	98

of the first section but not found to be highly significant either ($\overline{p_v} > 0.01$) and will thus be neglected.

Only seven proxies of the first section are found significant ($\overline{p_v} \leq 0.05$). Despite the lack of individual significance, the other proxies of this section will be passed to the next step of the screening procedure. This decision is motivated by the possibility of a potential joint significance to be found for those proxies in the MLR models (i.e., proxies becoming significant when considered collectively).

Note that all previous statements are based on the average p-values across the six

HMs. This is motivated by the fact that all HMs are correlated to each others (see Section 5.1) and are sharing common sources of pollution (see Section 2.1 and 5.2). Accordingly, we expect the MLR models, even if selected individually, to share common proxies and the coefficients to exhibit consistent signs across all HMs. The interpretations of the results will be approached from a global perspective, focusing on the overall patterns rather than attempting to individually explain each HM.

5.5 Annual variation assessment

In the case where the annual variability is considered negligible, incorporating the additional data from the 2020-2021 campaigns is expected to enhance the accuracy of the MLR models and the estimation of the coefficient. In order investigate the possibility of aggregating all the samples for analysis, MLR models can be estimated using the samples of the 2019 campaign ($n = 98$). Subsequently, to the observed values found in the samples of the 2020-2021 campaigns ($n = 28$) can be compared to the prediction interval of these models. On the average 95% of the predictions samples for the 2020-2021 campaigns should lie within the 95% prediction interval that are computed using the 2019 campaign. We need to verify that these observed proportions (of observations lying within the 95% prediction interval) are falling within acceptable ranges.

It can be done by formalizing the problem as a binomial distribution with a finite number of trials (n=28) where a measure falling inside the prediction confidence interval specified by the corresponding MLR is considered a success of the trial (m) and the opposite, a failure of the trial. Accordingly, the frequency ($f = m/n$) is an estimation of the theoretic 95% prediction proportion. Based on this value, a 99% Clopper-Pearson confidence interval for a binomial proportion is calculated.

As displayed on 8, all proportions falls within the confidence interval for every HMs except Cd. This supports the idea that the annual variability can be neglected (for

Table 8 : Pearson-Cobbler binomial proportion confidence intervals for each HMs (p_a and and p_b are the upper and lower bounds of the confidence interval, respectively). The proportion corresponds to the prediction success of the intermediary MLR models.

	Cd	Cr	Cu	Ni	Pb	Zn
m	21	28	27	27	28	25
f	0.75	1	0.96	0.96	1	0.89
p_a	0.49	0.83	0.76	0.76	0.83	0.66
p_b	0.92	1.00	1.00	1.00	1.00	0.99

our approach at least, since the prediction intervals are defined by the intermediary MLR models). The moderate result for Cd might be related to the analytical uncertainty previously mentioned. However, treating this HM distinctly from the others would complicate the global interpretations. Accordingly, we decide to re-estimate the models with the data for the three campaigns pooled together, while keeping in mind this assumption when interpreting further results for Cd.

5.6 Second Screening step : MLR models

The second step of the screening procedure corresponds to the determination of the most relevant proxies in the candidate subset provided by in the first step of the screening. This done by building model with a forward selection procedure (see section 4.5.5). Table 9 displays the p-values associated with a given proxy when it is included alongside other relevant proxies in the corresponding model. As expected, a side by side comparison of the models reveals the significance of the same proxies for modeling multiple HMs. This Table also presents the R_a^2 values associated to each MLR model. These values are around 0.5 for each HM. In other words, the models are explaining approximately 50% of the variability of the HMs concentrations in RD.

A global assessment of the relevance of proxies can be made by considering the number of models in which they appear. By this metric, "Distance to BCR center" and "Land use" are the most relevant proxies for the modeling. "Road

Table 9 : Selected proxies ($p_v \leq 0.05$) for the MLR models, with their corresponding p-values, and R_a^2 for simple and complex models.

Proxy	Cd	Cr	Cu	Ni	Pb	Zn
Distance to BCR center	2.7e-2	5.4e-3	-	1.3e-8	1.7e-5	2.6e-7
Land use	5.0e-2	9.0e-3	4.2e-5	-	1.8e-2	8.0e-4
Road hierarchy	1.0e-2	1.0e-3	5.4e-4	3.6e-3	-	-
Roadside parking occupation (5-7am)	4.2e-5	2.5e-4	-	4.4e-2	4.9e-3	-
Street canyon ratio	-	-	2.8e-2	4.9e-3	1.6e-2	-
Population density	3.5e-2	6.0e-4	-	-	-	-
Roadside parking occupation (10-12am)	1.9e-2	-	-	-	-	-
Black carbon (peak time)	-	-	-	-	1.7e-2	-
R_a^2 (models given above)	0.45	0.50	0.51	0.48	0.48	0.49
R_a^2 (with higher-order terms)	0.45	0.54	0.54	0.54	0.55	0.53

hierarchy", "Roadside parking occupation (5-7am)" and "Street canyon ratio" are also quite relevant. "Population density", "Roadside parking occupation (10-12am)" and "Black carbon (peak time)" were also selected by the automated procedure. However, they appear in a limited number of models.

While "Population density" correlation with HMs concentrations has been documented by [36, 64], the limited relevance of this proxy in the models is unclear. This might be explained by the presence of the "Distance to BCR center" in the models, as these two proxies are moderately correlated (r=-0.55). However, the moderate correlations between some proxies is not sufficient to discard one of them from the models. It is important to address a possible redundancy beyond the scope of a simple Pearson correlation analysis, as verifying the independence of a proxy has to be done with respect to the proxies individually (i.e., collinearity), but also in regard to the collective set of proxies within the same model (i.e., multicollinearity). This second problem is addressed by calculating the AGVIF for each proxy in a model (see equation 10) and making sure that none of these values are greater than 2.236, as this would be a multicollinearity concern possibly impacting the quality of the models. Accordingly, no proxy is discarded as no values are found to be greater than this threshold.

Another valid concern regarding the quality of the models is the omission of combined effects of two proxies (interaction) or the presence of a curvature in the relationship between a continuous proxy and the concentrations (rather than a purely linear). To address these possible oversights, it is recommended to include higher-order terms (i.e., interaction and quadratic terms) and to assess their relevance. In our analysis, 11 terms were found significant ($\overline{p_v} \leq 0.05$). As shown in Table 9, their inclusion in the models leads to a slight increase of the R_a^2 for all HMs except for Cd for which no higher terms were found significant. However, the decision was made to present the models without these higher-order terms as they each appear for a limited number of HMs and would complicate the interpretations from a global perspective.

Table 10 presents the estimated coefficients associated to the models. The first observation that can be made concerning the coefficients is that they are all consistent across all HMs (the sign is either positive or negative for the models). This eases the interpretation and allow for a global description of the proxies effects. An important reminder is that the coefficient for the different categories corresponds to the difference with a reference level. Accordingly, the intercept should be seen as the estimated log-concentration of a data point with all continuous proxy values set to 0 and all categorical ones set to their reference level. This intercept value will not be interpreted further as it is representing a purely virtual value that cannot reasonably be linked to a real location.

The inverse relationship between "Distance to BCR center" and the HMs concentrations (i.e, negative coefficients) implies a positive correlation between the proximity to the center of the urban area and higher concentrations, which aligns the observations made by [34, 45]. This can be explained by the higher vehicular traffic and level of human activity expected for a location close to the center of an urban area. As said earlier, the limited relevance of the population density might be caused by a possible interference between those two proxies. However, when included in the models, the positive relationship between population density and

Table 10 : Values of the coefficients for the MLR models.

Coefficients	Cd	Cr	Cu	Ni	Pb	Zn
Intercept	1.541	4.708	3.412	4.862	5.460	5.963
Distance to BCR center [km]	-0.096	-0.208	-	-0.271	-0.324	-0.235
Land use : Densely populated	0.086	0.688	1.442	-	0.623	0.783
Land use : Industrial	0.128	0.951	1.568	-	0.348	0.858
Land use : Mixed	0.334	0.836	1.463	-	0.733	0.882
Land use : Residential	0.044	0.522	0.815	-	0.162	0.456
Road hierarchy : Metropolitan (A1)	0.279	0.475	0.947	0.439	-	-
Road hierarchy : Main (A2), Interdistrict (A3)	0.205	0.500	0.504	0.335	-	-
Roadside parking occupation (5-7 am)	-0.011	-0.017	-	-0.005	-0.011	-
Street canyon index	-	-	0.749	0.497	0.549	-
Population density [km^{-2}]	2.36e-5	6.99e-5	-	-	-	-
Roadside parking occupation (10-12 am)	7.22e-3	-	-	-	-	-
Black carbon (peak time) [$\mu g/m^3$]	-	-	-	-	0.100	-

HMs concentration is in agreement with the observation of [36, 64].

The "Land use" coefficients are the distinct concentrations with the reference level of the proxy ("Green areas"). Based on this, we can see that the relevance of this proxy lies in the difference between mainly lies in the difference of concentrations between "Green areas" and the other categories (the coefficient of "Green areas" can be viewed as 0 for the sake of comparison). The order of impact of the categories on HMs concentrations from highest to lowest can be presented as follows : [Industrial, Mixed areas > Densely populated > Residential > Green]. The only exception is for Cu as "Densely populated areas" exhibits a higher coefficient than "Industrial areas". This hierarchy aligns with the expectation that industrial and commercial activities ("Mixed areas" represents commercial and public service areas) exert a greater influence on the concentrations compared to residential areas, this is in agreement with the results of [29]. Between densely populated and residential areas, it is logical to assume that a more populated area is subjected to higher vehicular traffic density and level of human activity. "Green areas" lower impact on the HMs concentrations could be explained by the mitigating impact of

vegetation on atmospheric deposition [65]. The absence of a significantly greater effect attributed to "Industrial areas" was expected as few industrial enterprises are still in activity in the BCR (see Section 4.1).

The "Street canyon ratio" coefficient positive sign is consistent with expectation that streets with a high canyon ratio value are expected to show higher HMs concentrations (see Section 1). Another proxy related to atmospheric deposition is "Black Carbon (peak time)" which corresponds to measurements of this pollutant in the atmosphere. The relevance of this proxy is limited to Pb but the correlation between these two pollutants concentrations in atmospheric deposition might be explained by the findings of [4] concerning the inclusion of Pb within Black Carbon aggregate in atmospheric deposition.

"Roadside parking occupation (5-7am)" coefficient displays a negative sign. This might be explained by the fact that locations with a high parking occupation before working hours are typically associated to residential areas that are less affected by commuter traffic and consequently less polluted in HMs. On the other hand, "Roadside parking occupation (10-12am) exhibits a positive effect on the concentrations of Cd. This result remains unclear and should be interpreted with caution as it appears relevant for Cd only.

The coefficients associated with "Road hierarchy" categories reports higher concentrations for higher class roads. The reference category for this proxy is the "district collector and district road (A4, A5)" class. This is in agreement with the results of [45] as it is expected that higher class roads are associated with a heavier traffic intensity [66]. The grouping of categories is done in order to dispose of a sufficient number of observations for each class. The difference between "A1" and "A2, A3" roads with the reference is clear, but the difference between those two categories is less clear. For Cr the estimated effect is even higher for "A2,A3" than for "A1". The hierarchy based on concentrations values can be described as : [A1, A2, A3 > A4, A5].

5.7 Spatial dependence of the regression residuals

In order to assess the spatial dependence of the MLR residuals, empiric variograms were build and modeled for all HMs as described in section 4.5.6. The results are presented graphically in Figure 9 for Cr and Cu[14]. It can be observed that the residuals are spatially correlated for an approximate range of 700 $[m]$. However, the part of the spatially correlated variance (reported in Table 11) is uneven between HMs, with a strong spatial correlation observed for Cr and a moderate to low correlation for the other HMs. These results are hinting the possible omission of one or several relevant spatially correlated proxies (i.e., proxies with an apparent spatial structure) in the MLR models. These results also reveals that the concentrations of HMs in RD are exhibiting large variations over short distances. This effect is difficult to observe and quantify at the scale of our study due to the limited amounts of samples close to each other and the low spatial resolution of some proxies.

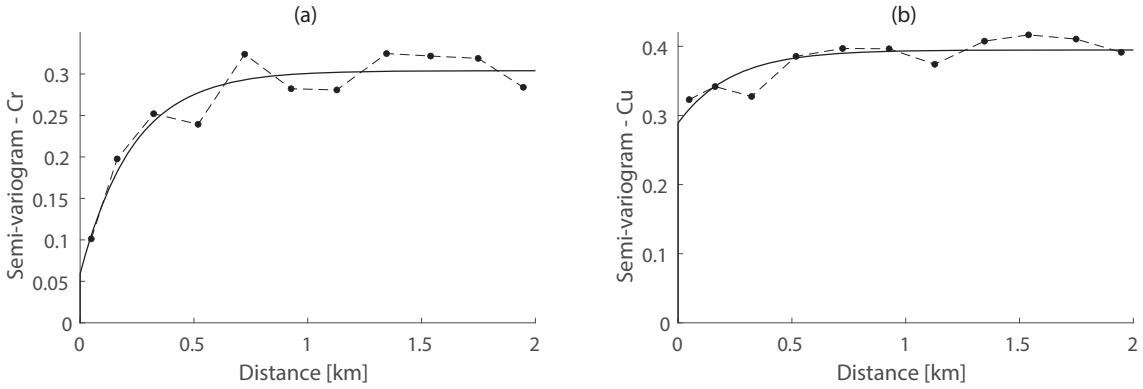


Figure 9 : Variograms of the regression residuals for Cr (left) and Cu (right). Dashed lines correspond to the empirical variograms (see equation 13) while plain lines correspond to the parametric exponential models (see equation 14).

[14]The results are not presented graphically for the other HMs as they show comparable behaviors to Cr and Cu residuals.

Table 11 : Spatially correlated part of the variance of the regression residuals.

	Cd	Cr	Cu	Ni	Pb	Zn
$\alpha_2 \,/\, (\alpha_1 + \alpha_2)$	0.24	0.81	0.26	0.63	0.53	0.38

Another observation that can be made about the residuals is the fact that they are positively correlated across all HMs (results not presented here). Accordingly, it is expected that the inclusion of an omitted relevant proxy would influence the concentrations of the various HMs similarly (i.e. exhibit MLR coefficients with consistent signs across all HMs), as it was observed for the proxies at hand in section 5.6.

5.8 Mapping

By applying the various MLR models on the entire AM, we were able to map the predicted HMs concentrations in RD for the majority of the road segments[15]. The part of the road network with no associated prediction corresponds with to locations with "Highway (A0)" road hierarchy category and the "Transport and port activity" Land use category, as we do not dispose of samples for road segments representing those categories.

For all HMs, the concentrations are showing a gradient of increasing concentrations towards the center of the BCR, situated East from the municipality (see Figure 4). This is particularly visible for Zn (Figure 15), as it is one of the two proxies included in his model. This observation is also true in a certain measure for Cu (Figure 12), that is not including the Distance from the BCR center in his model. Accordingly, the HMs concentrations are displaying relatively concentric patterns to the around the center of the BCR. The deviations from this concentric pattern can be attributed to remaining proxies. Street canyon ratio and Parking occupation might add to the complexity of the patterns, especially in the Eastern part of the municipality. However, The effect of these proxies is difficult to see on the maps

[15]Prediction value for a road segment is calculated for his centroid.

compared to other proxies. It is noteworthy to say that location with higher Street canyon ratio are generally found in the Eastern part and absent in the Western part of the municipality. It is also notable that the road network displays a pattern of higher concentrations for the main axes (this pattern is absent for Zn and unclear for Pb, both not including the road hierarchy in their model).

For all HMs, a relatively homogeneous pattern of lower concentrations can be observed in the Western part of the municipality. This can be explained by the already mentioned concentric pattern attributed to the Distance from the BCR center but also by the local Land use (mainly "Green areas" and "Residential areas") and Road hierarchy (mainly "district collector and district road"). However, a small region South of the Western part of the municipality displays higher concentrations values for several HMs, disrupting the homogeneous pattern. This also can also be explained by the prevailing local Land use (mainly "Industrial areas") and Road hierarchy (mainly "Main and Interdistrict road"). This pattern is less visible for the HMs that do not include both Land use and road hierarchy in their model (i.e., Ni, Pb, Zn).

The patterns associated to distinct Land use alone are not clearly visible except for Zn due to the low complexity of the model. Accordingly, the deviations from the mainly concentric pattern can be attributed to distinct Land use. A small cluster of roads in the North-eastern part of the municipality displays lowers concentrations values corresponding to "Residential areas" and another small cluster displays higher concentrations values corresponding to "Mixed areas" in the South-Eastern part of the municipality.

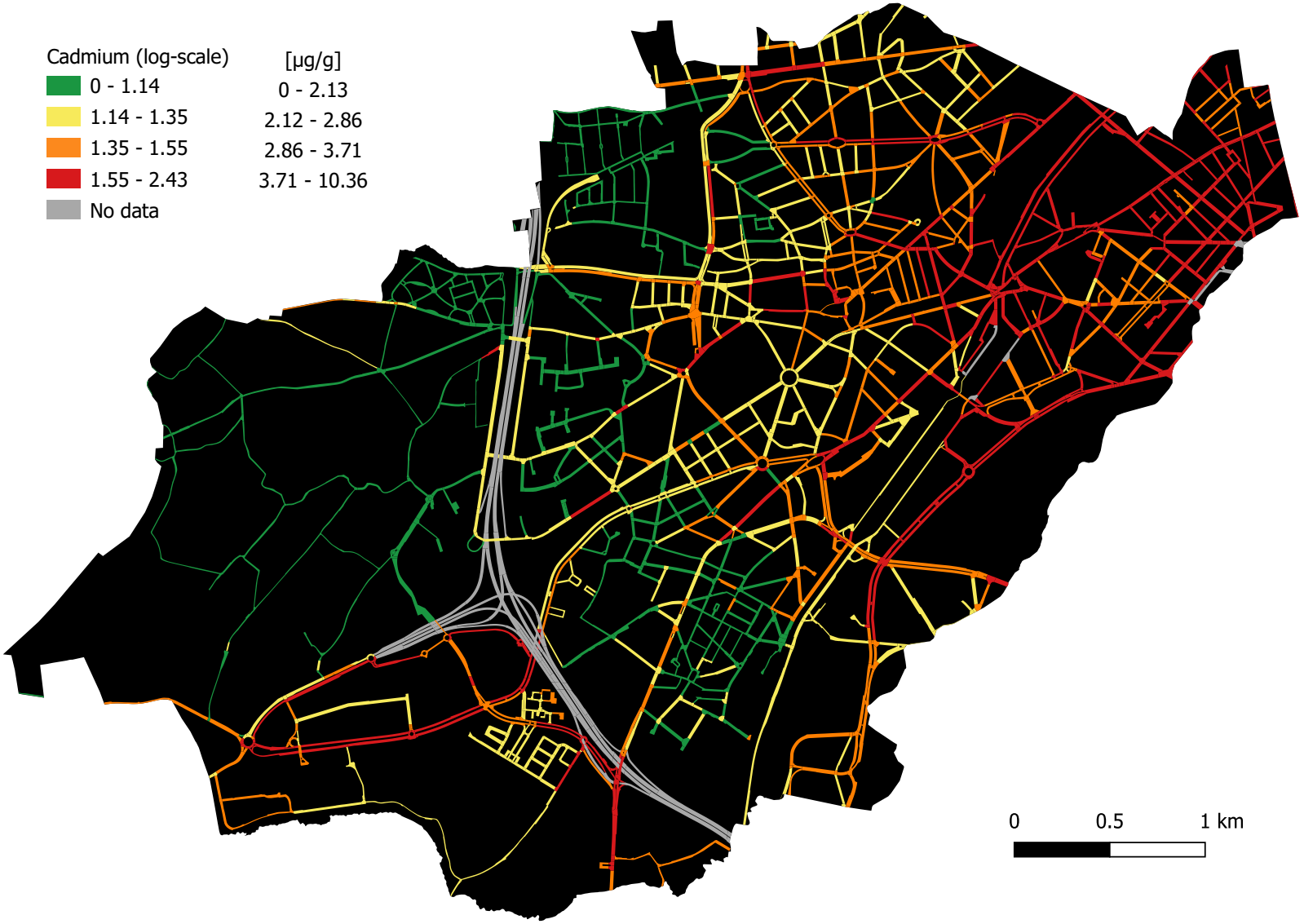


Figure 10 : Estimation with corresponding MLR of the Cd concentrations in RD for each road segment of the AM.

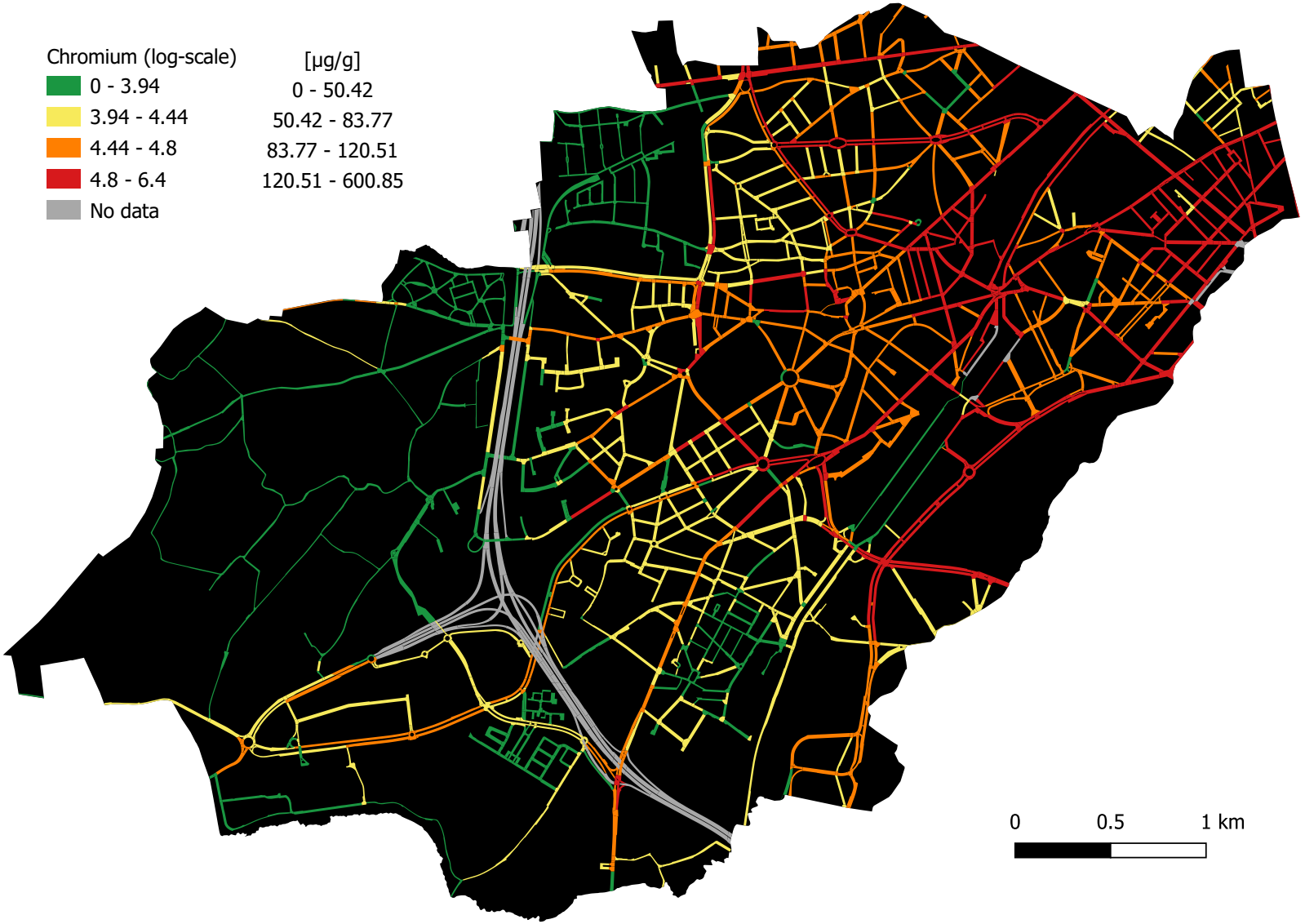


Figure 11 : Estimation with corresponding MLR of the Cr concentrations in RD for each road segment of the AM.

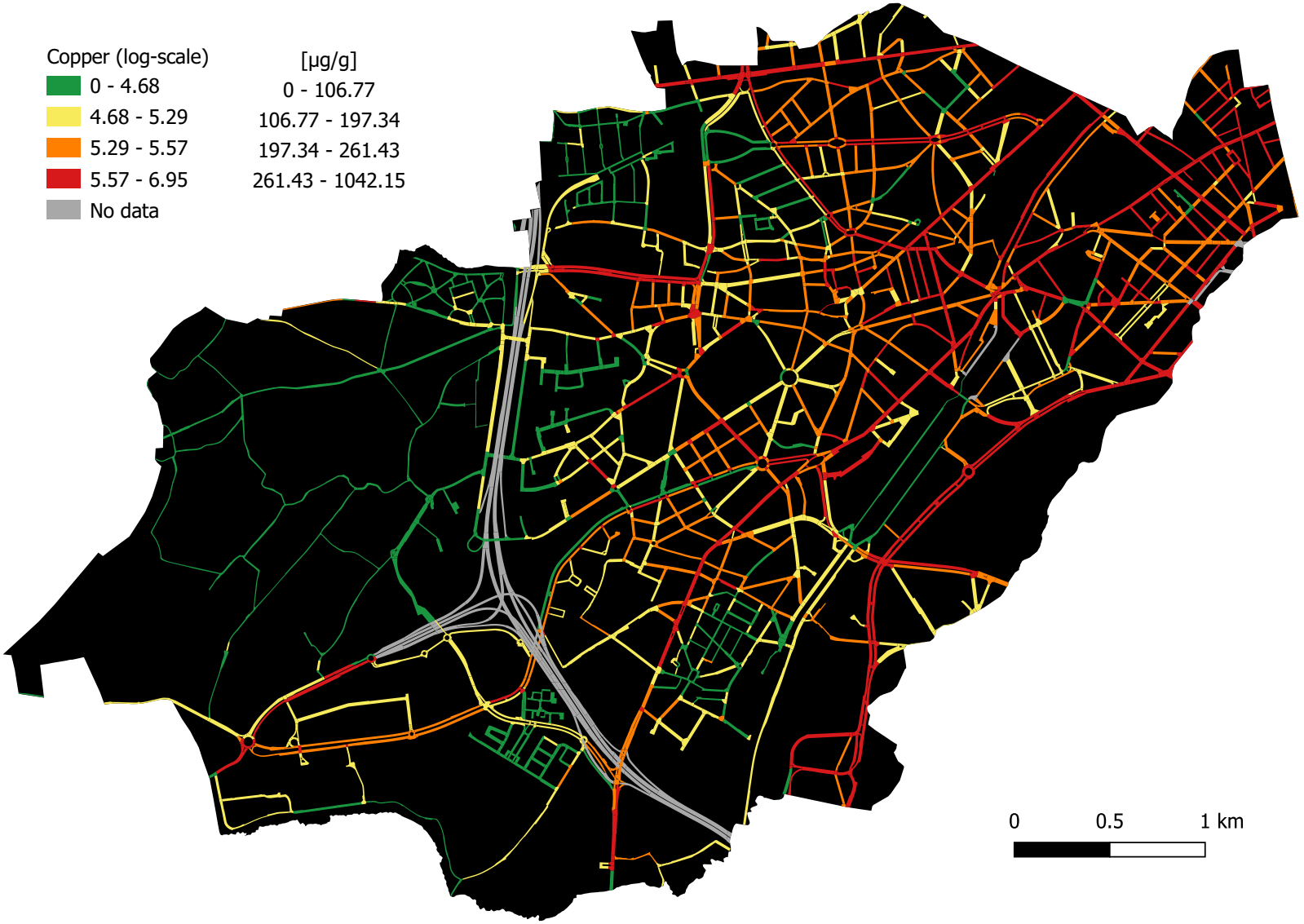


Figure 12 : Estimation with corresponding MLR of the Cu concentrations in RD for each road segment of the AM.

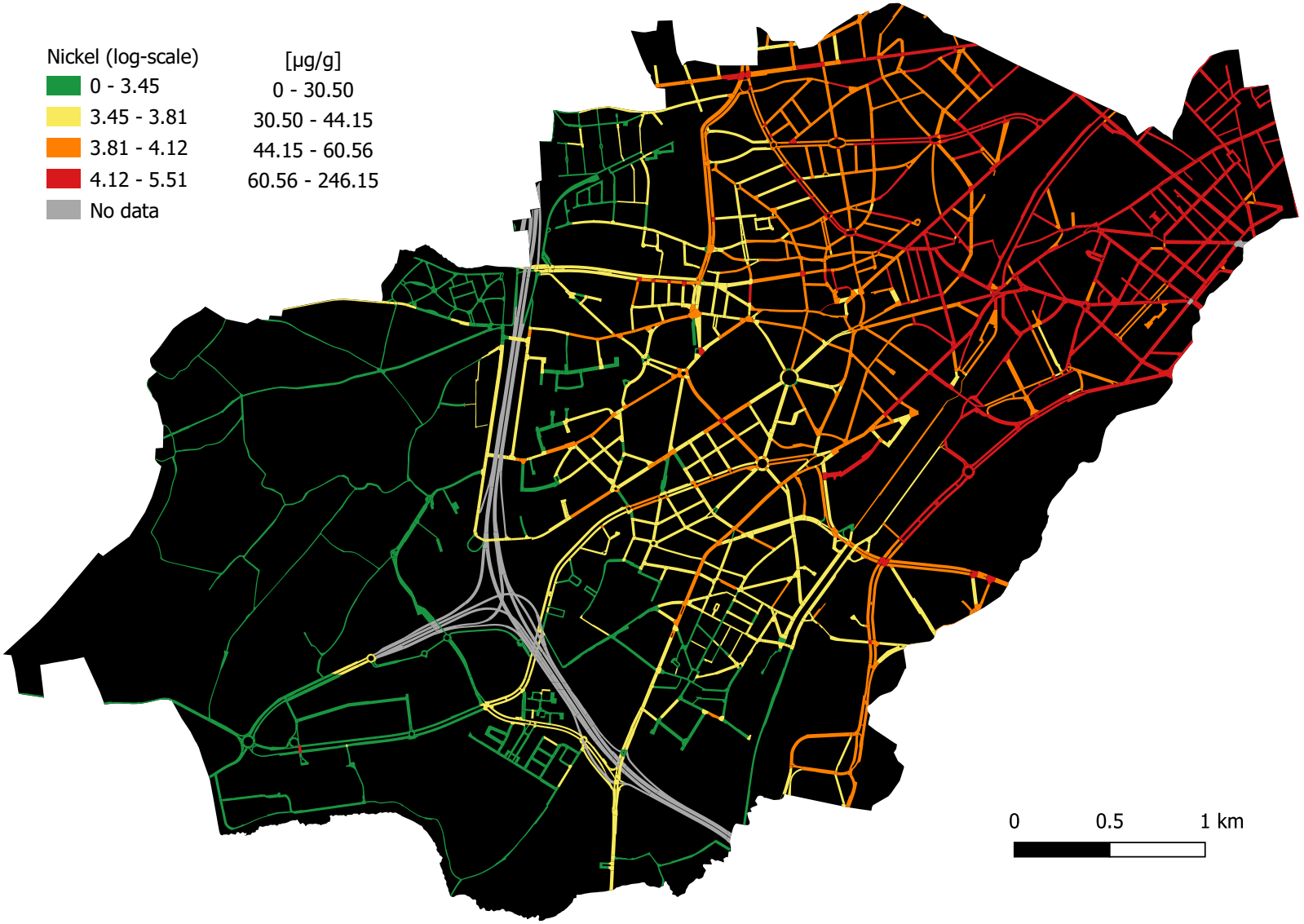


Figure 13 : Estimation with corresponding MLR of the Ni concentrations in RD for each road segment of the AM.

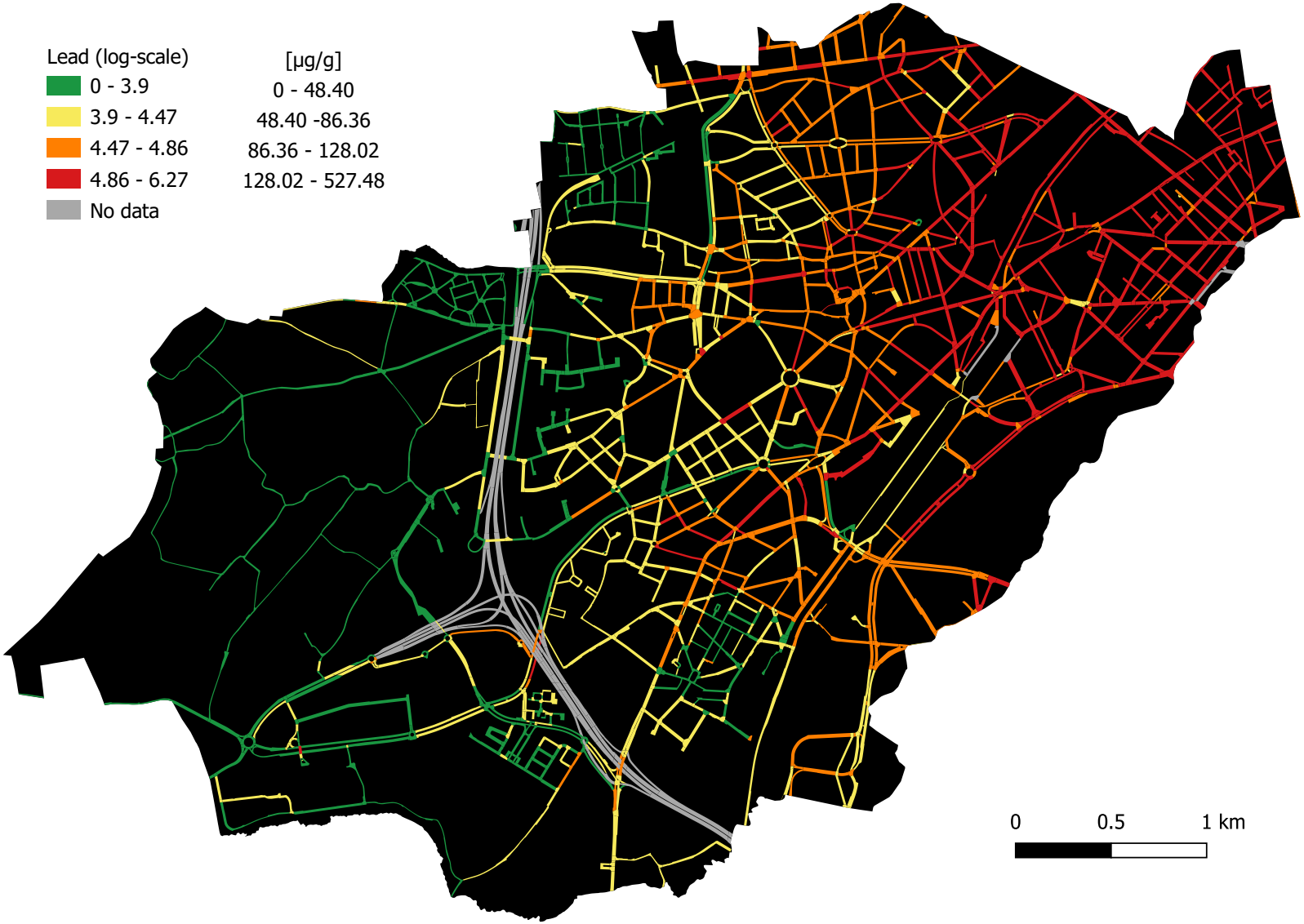


Figure 14 : Estimation with corresponding MLR of the Pb concentrations in RD for each road segment of the AM.

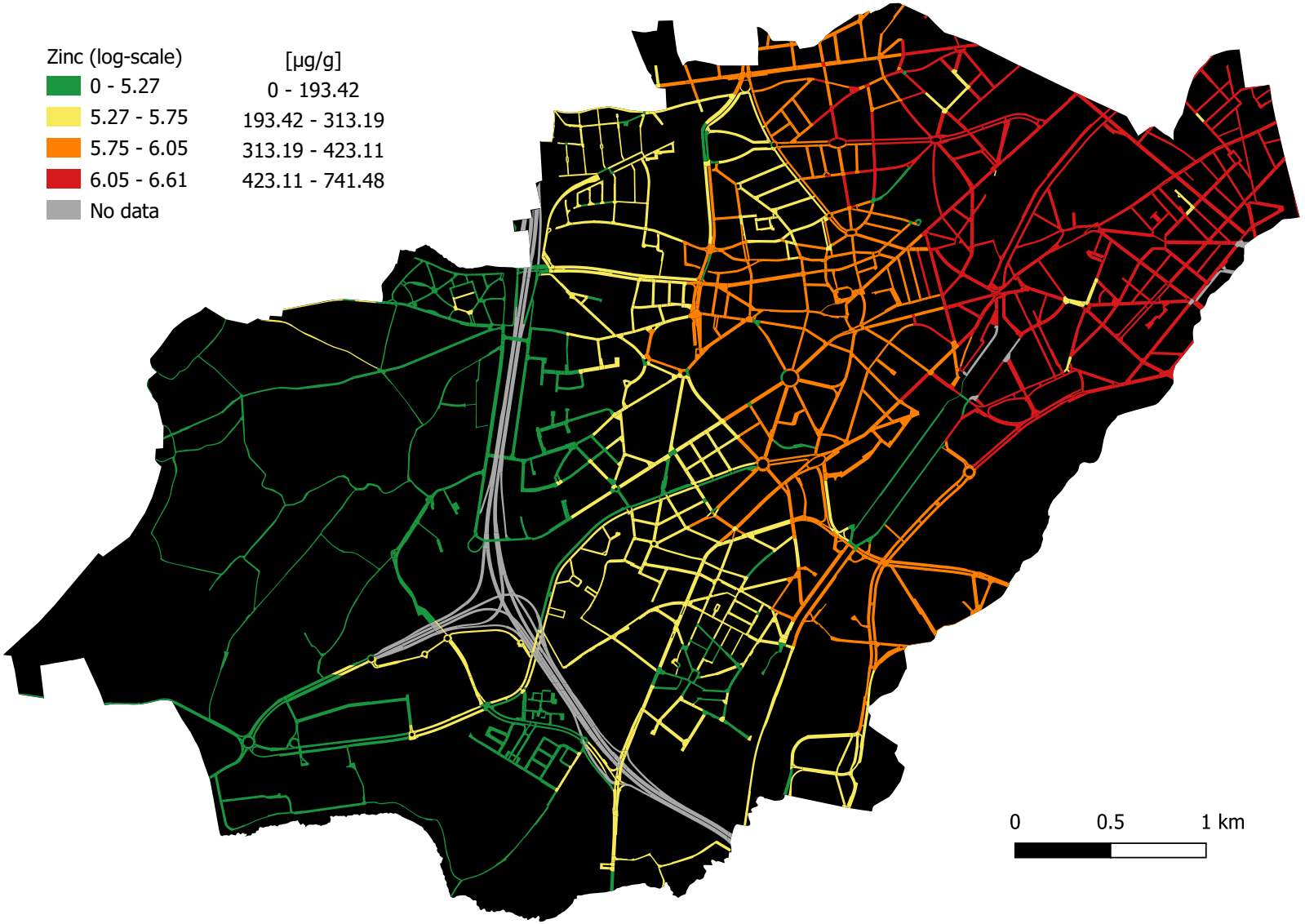


Figure 15 : Estimation with corresponding MLR of the Zn concentrations in RD for each road segment of the AM.

6 Conclusion

This master thesis has investigated the spatial distribution of HMs concentrations in RD in the municipality of Anderlecht. More precisely, the concentrations of Cd, Cr, Cu, Ni, Pb and Zn in the finest fraction of RD ($\varnothing < 250\ \mu m$) were modeled by MLR models selected with a stepwise forward selection procedure with a variety of candidate proxies and used to predict HMs concentrations for all the road segments of the AM. The most relevant proxies were : "Distance from the center of the BCR", "Land use", "Road hierarchy" and "Roadside parking occupation". The MLR models displayed moderate performances (with R_a^2 values around 0.5) but are coherent in a side-by-side comparison (multiple significance of the proxies and consistent coefficient signs across all HMs). A spatial dependence analysis of the regression residuals revealed that relevant proxies might have been omitted from our analyses.

Despite the moderate performances of the models, we believe that the resulting maps provide useful information for the eventual implementation of mitigation strategies (such as targeted street sweepings for identified hot spots) by local authorities and public services of the BCR. We believe that our predictions could be extended to the whole region with a small targeted sampling campaign for the missing information about specific Land uses categories (i.e., "Transport and port activities", "Administrative areas"). In order to verify the possibility of extending the predictions to the BCR, a second small campaign could be conducted for locations in other municipalities[16].

In the first step of the screening procedure, numerous proxies were assessed individually to predict the concentrations for each HM. This could prove useful, especially for investigating a specific HM. A focused approach of a single HM across all his "environmental compartments" (i.e., RD, PM, soil, aquatic sediment) might prove useful to quantify the transfer processes and bring a broader perspective on

[16]The annual variability associated to new campaigns should not be a problem as it was deemed negligible for our approach.

urban pollution, taking into account the complexity of urban pollution dynamics.

Based on a technique used in previous studies, a source apportionment analysis was conducted. While the results are unclear, it seems to confirm that the studied HMs share common contamination sources in RD for the BCR. This reinforces the idea that the modeling of the studied HMs concentrations should be done with a holistic approach. In that perspective, the possibility be to use a single model (potentially different from the linear approaches presented here) to predict the concentrations rather than six separated models should be investigated. There is an important distinction between this potential approach and the prediction of the aggregated HMs concentrations, as the latter would disregard the individual characteristics and contribution of the HMs, particularly those present in significantly lower concentrations, such as Cd.

Another important consideration concerning the source apportionment is that the determination of the precise contributions of the natural sources of HMs in RD might help us to define meaningful background values for the BCR, which in turn strengthens the assessment of pollution levels for further studies. The analysis of supplementary HMs, notably unaffected by anthropogenic sources might help in the identification of the natural sources by providing clearer sources contributions patterns. They could, in parallel, be used as normalizing elements in Enrichment Factor calculations for pollution assessment.

The geospatial modeling might be improved in regard to the precision of the predictions with a larger number of samples and the use of more complex models (e.g., artificial intelligence, machine learning) allowed by this large number of data. However, the benefits brought by more precise predictions might be relatively low in comparison to the economic costs and additional workload that the chemical analyses would involve. Improving the approach without altering it too much seems a more reasonable expectation. Taking into account the spatial dependence of the HMs concentrations in the models with consideration for the complexity of

the urban environment might improve the results without the need of additional proxies (relevant proxies such as "Distance from the BCR center" might to become less relevant as a result). However, the modeling would be restricted to the sampled area. On the other hand, the improvement of the spatial resolution of certain proxies we used could also improve the results. However, at the moment, the division and categorizations of the road network, Land use areas and statistical districts seems more motivated by administrative considerations than environmental concerns. An improvement in this sense would require an effort from the BCR local authorities to revise or provide alternate geospatial information to this effect in their open access databases. This should be done with consideration for the broader scientific community requirements for geospatial data.

In conclusion, we believe that our approach is worth presenting as it provides a step-by-step procedure (with adequate verifications for the potential problems) for the geospatial modeling of urban pollutants that could potentially be applied to other urban areas.

References

[1] J. O. Anderson, J. G. Thundiyil, and A. Stolbach. Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health. In: *Journal of Medical Toxicology* 8.2 (2012), pp. 166–175. DOI: 10.1007/s13181-011-0203-1.

[2] US EPA. *Health and Environmental Effects of Particulate Matter (PM)*. URL: https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm.

[3] H. Li, X. Qian, and Q. Wang. Heavy Metals in Atmospheric Particulate Matter: A Comprehensive Understanding Is Needed for Monitoring and Risk Mitigation. In: *Environmental Science & Technology* 47.23 (2013), pp. 13210–13211. DOI: 10.1021/es404751a.

[4] P.-K. Lee, S. Yu, H. J. Chang, H. Y. Cho, M.-J. Kang, and B.-G. Chae. Lead chromate detected as a source of atmospheric Pb and Cr (VI) pollution. In: *Scientific Reports* 6.1 (2016), p. 36088. DOI: 10.1038/srep36088.

[5] K. Cai, C. Li, and S. Na. Spatial Distribution, Pollution Source, and Health Risk Assessment of Heavy Metals in Atmospheric Depositions: A Case Study from the Sustainable City of Shijiazhuang, China. In: *Atmosphere* 10.4 (2019), p. 222. DOI: 10.3390/atmos10040222.

[6] Y. Du, B. Gao, H. Zhou, X. Ju, H. Hao, and S. Yin. Health Risk Assessment of Heavy Metals in Road Dusts in Urban Parks of Beijing, China. In: *Procedia Environmental Sciences* 18 (2013), pp. 299–309. DOI: 10.1016/j.proenv.2013.04.039.

[7] A. Guilbert, K. De Cremer, B. Heene, C. Demoury, R. Aerts, P. Declerck, O. Brasseur, and A. Van Nieuwenhuyse. Personal exposure to traffic-related air pollutants and relationships with respiratory symptoms and oxidative stress: A pilot cross-sectional study among urban green space workers. In: *Science of The Total Environment* 649 (2019), pp. 620–628. DOI: 10.1016/j.scitotenv.2018.08.338.

[8] G. Wang, Y. Wang, W. Yin, T. Xu, C. Hu, J. Cheng, J. Hou, Z. He, and J. Yuan. Seasonal exposure to PM2.5-bound polycyclic aromatic hydrocarbons and estimated lifetime risk of cancer: A pilot study. In: *Science of The Total Environment* 702 (2020), p. 135056. DOI: 10.1016/j.scitotenv.2019.135056.

[9] S. Chen, X. Zhang, J. Lin, J. Huang, D. Zhao, T. Yuan, K. Huang, Y. Luo, Z. Jia, Z. Zang, Y. Qiu, and L. Xie. Fugitive Road Dust $PM_{2.5}$ Emissions and Their Potential Health Impacts. In: *Environmental Science & Technology* 53.14 (2019), pp. 8455–8465. DOI: 10.1021/acs.est.9b00666.

[10] R. K. Khan and M. A. Strand. Road dust and its effect on human health: a literature review. In: *Epidemiology and Health* 40 (2018), e2018013. DOI: 10.4178/epih.e2018013.

[11] M. Gabarrón, A. Faz, and J. A. Acosta. Soil or Dust for Health Risk Assessment Studies in Urban Environment. In: *Archives of Environmental Contamination and Toxicology* 73.3 (2017), pp. 442–455. DOI: 10.1007/s00244-017-0413-x.

[12] C. C. Kaonga, I. B. M. Kosamu, and W. R. Utembe. A Review of Metal Levels in Urban Dust, Their Methods of Determination, and Risk Assessment. In: *Atmosphere* 12.7 (2021), p. 891. DOI: 10.3390/atmos12070891.

[13] L. Zhang, M. Gao, J. Cui, F. Yang, H. Wang, C. Fu, and Y. Huang. Wet Deposition of Trace Metals at a Typical Urban Site in Southwestern China: Fluxes, Sources and Contributions to Aquatic Environments. In: *Sustainability* 10.2 (2017), p. 69. DOI: 10.3390/su10010069.

[14] A. Christoforidis and N. Stamatis. Heavy metal contamination in street dust and roadside soil along the major national road in Kavala's region, Greece. In: *Geoderma* 151.3-4 (2009), pp. 257–263. DOI: 10.1016/j.geoderma.2009.04.016.

[15] C. Men, R. Liu, Q. Wang, L. Guo, Y. Miao, and Z. Shen. Uncertainty analysis in source apportionment of heavy metals in road dust based on positive matrix factorization model and geographic information system. In: *Science of The Total Environment* 652 (2019), pp. 27–39. DOI: 10.1016/j.scitotenv.2018.10.212.

[16] Y. Ma, S. Mummullage, B. Wijesiri, P. Egodawatta, J. McGree, G. A. Ayoko, and A. Goonetilleke. Source quantification and risk assessment as a foundation for risk management of metals in urban road deposited solids. In: *Journal of Hazardous Materials* 408 (2021), p. 124912. DOI: 10.1016/j.jhazmat.2020.124912.

[17] H. Wang, C. Shen, Y. Kang, Q. Deng, and X. Lin. Spatial distribution of pollution characteristics and human health risk assessment of exposure to heavy elements in road dust from different functional areas of Zhengzhou, China. In: *Environmental Science and Pollution Research* 27.21 (2020), pp. 26650–26667. DOI: 10.1007/s11356-020-08942-7.

[18] H.-M. Hwang, M. J. Fiala, D. Park, and T. L. Wade. Review of pollutants in urban road dust and stormwater runoff: part 1. Heavy metals released from vehicles. In: *International Journal of Urban Sciences* 20.3 (2016), pp. 334–360. DOI: 10.1080/12265934.2016.1193041.

[19] H. Zhao, J. Zhao, C. Yin, and X. Li. Index models to evaluate the potential metal pollution contribution from washoff of road-deposited sediment. In: *Water Research* 59 (2014), pp. 71–79. DOI: 10.1016/j.watres.2014.04.012.

[20] O. Brasseur, P. Declerck, B. Heene, and P. Vanderstraeten. Modelling Black Carbon concentrations in two busy street canyons in Brussels using CANS BC. In: *Atmospheric Environment* 101 (2015), pp. 72–81. DOI: 10.1016/j.atmosenv.2014.10.049.

[21] F. De Nicola, F. Murena, M. A. Costagliola, A. Alfani, D. Baldantoni, M. V. Prati, L. Sessa, V. Spagnuolo, and S. Giordano. A multi-approach monitoring of particulate matter, metals and PAHs in an urban street canyon. In: *Environmental Science and Pollution Research* 20.7 (2013), pp. 4969–4979. DOI: 10.1007/s11356-012-1456-1.

[22] L. Pirjola, T. Lähde, J. Niemi, A. Kousa, T. Rönkkö, P. Karjalainen, J. Keskinen, A. Frey, and R. Hillamo. Spatial and temporal characterization of traffic emissions in urban microenvironments with a mobile laboratory. In: *Atmospheric Environment* 63 (2012), pp. 156–167. DOI: 10.1016/j.atmosenv.2012.09.022.

[23] M. Skorbiłowicz, Ł. Trybułowski, and E. Skorbiłowicz. Spatial Distribution and Pollution Level of Heavy Metals in Street Dust of the City of Suwałki (Poland). In: *International Journal of Environmental Research and Public Health* 20.6 (2023), p. 4687. DOI: 10.3390/ijerph20064687.

[24] P. V. Suryawanshi, B. S. Rajaram, A. D. Bhanarkar, and C. V. Chalapati Rao. Determining heavy metal contamination of road dust in Delhi, India. In: *Atmósfera* (2016). DOI: 10.20937/ATM.2016.29.03.04.

[25] L. Hakanson. An ecological risk index for aquatic pollution control.a sedimentological approach. In: *Water Research* 14.8 (1980), pp. 975–1001. DOI: 10.1016/0043-1354(80)90143-8.

[26] C. Men, R. Liu, F. Xu, Q. Wang, L. Guo, and Z. Shen. Pollution characteristics, risk assessment, and source apportionment of heavy metals in road dust in Beijing, China. In: *Science of The Total Environment* 612 (2018), pp. 138–147. DOI: 10.1016/j.scitotenv.2017.08.123.

[27] H. Khademi, M. Gabarrón, A. Abbaspour, S. Martínez-Martínez, A. Faz, and J. A. Acosta. Environmental impact assessment of industrial activities on heavy metals distribution in street dust and soil. In: *Chemosphere* 217 (2019), pp. 695–705. DOI: 10.1016/j.chemosphere.2018.11.045.

[28] B. D. Škrbić, M. Buljovčić, G. Jovanović, and I. Antić. Seasonal, spatial variations and risk assessment of heavy elements in street dust from Novi Sad, Serbia. In: *Chemosphere* 205 (2018), pp. 452–462. DOI: 10.1016/j.chemosphere.2018.04.124.

[29] A. Aguilera, C. Armendariz, P. Quintana, F. García-Oliva, and F. Bautista. Influence of Land Use and Road Typeon the Elemental Composition of Urban Dustin a Mexican Metropolitan Area. In: *Polish Journal of Environmental Studies* 28.3 (2019), pp. 1535–1547. DOI: 10.15244/pjoes/90358.

[30] X. Yuan, T. An, B. Hu, and J. Zhou. Analysis of spatial distribution characteristics and main influencing factors of heavy metals in road dust of Tianjin based on land use regression models. In: *Environmental Science and Pollution Research* 30.1 (2023), pp. 837–848. DOI: 10.1007/s11356-022-22151-4.

[31] D. J. Briggs, S. Collins, P. Elliott, P. Fischer, S. Kingham, E. Lebret, K. Pryl, H. Van Reeuwijk, K. Smallbone, and A. Van Der Veen. Mapping urban air pollution using GIS: a regression-based approach. In: *International Journal of Geographical Information Science* 11.7 (1997), pp. 699–718. DOI: 10.1080/136588197242158.

[32] L. Hu. Spatiotemporal Analysis Method of Urban Environmental Factors Along Streets Constrained by Road Network. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-3/W1-2022 (2022), pp. 13–18. DOI: 10.5194/isprs-archives-XLVIII-3-W1-2022-13-2022.

[33] D. Vlasov, N. Kosheleva, and N. Kasimov. Spatial distribution and sources of potentially toxic elements in road dust and its PM10 fraction of Moscow megacity. In: *Science of The Total Environment* 761 (2021), p. 143267. DOI: 10.1016/j.scitotenv.2020.143267.

[34] E. de Miguel, J. F. Llamas, E. Chacón, T. Berg, S. Larssen, O. Røyset, and M. Vadset. Origin and patterns of distribution of trace elements in street dust: Unleaded petrol and urban lead. In: *Atmospheric Environment* 31.17 (1997), pp. 2733–2740. DOI: 10.1016/S1352-2310(97)00101-5.

[35] J. Gunawardena, A. M. Ziyath, P. Egodawatta, G. A. Ayoko, and A. Goonetilleke. Sources and transport pathways of common heavy metals to urban road surfaces. In: *Ecological Engineering* 77 (2015), pp. 98–102. DOI: 10.1016/j.ecoleng.2015.01.023.

[36] J. M. Trujillo-González, M. A. Torres-Mora, S. Keesstra, E. C. Brevik, and R. Jiménez-Ballesta. Heavy metal accumulation related to population density in road dust samples taken from urban sites under different land uses. In: *Science of The Total Environment* 553 (2016), pp. 636–642. DOI: 10.1016/j.scitotenv.2016.02.101.

[37] S. Yang, J. Liu, X. Bi, Y. Ning, S. Qiao, Q. Yu, and J. Zhang. Risks related to heavy metal pollution in urban construction dust fall of fast-developing Chinese cities. In: *Ecotoxicology and Environmental Safety* 197 (2020), p. 110628. DOI: 10.1016/j.ecoenv.2020.110628.

[38] M. Faisal, Z. Wu, H. Wang, Z. Hussain, and C. Shen. Geochemical Mapping, Risk Assessment, and Source Identification of Heavy Metals in Road Dust Using Positive Matrix Factorization (PMF). In: *Atmosphere* 12.5 (2021), p. 614. DOI: 10.3390/atmos12050614.

[39] C. Huang, L. Zhang, J. Meng, Y. Yu, J. Qi, P. Shen, X. Li, P. Ding, M. Chen, and G. Hu. Characteristics, source apportionment and health risk assessment of heavy metals in urban road dust of the Pearl River Delta, South China. In: *Ecotoxicology and Environmental Safety* 236 (2022), p. 113490. DOI: 10.1016/j.ecoenv.2022.113490.

[40] H. Pan, X. Lu, and K. Lei. A comprehensive analysis of heavy metals in urban road dust of Xi'an, China: Contamination, source apportionment and spatial distribution. In: *Science of The Total Environment* 609 (2017), pp. 1361–1369. DOI: 10.1016/j.scitotenv.2017.08.004.

[41] L. Chen, H. Zhang, M. Ding, A. T. Devlin, P. Wang, M. Nie, and K. Xie. Exploration of the variations and relationships between trace metal enrichment in dust and ecological risks associated with rapid urban expansion. In: *Ecotoxicology and Environmental Safety* 212 (2021), p. 111944. DOI: 10.1016/j.ecoenv.2021.111944.

[42] J. B. Kowalska, R. Mazurek, M. Gąsiorek, and T. Zaleski. Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–A review. In: *Environmental Geochemistry and Health* 40.6 (2018), pp. 2395–2420. DOI: 10.1007/s10653-018-0106-z.

[43] M. H. Kabir, M. H. Rashid, Q. Wang, W. Wang, S. Lu, and S. Yonemochi. Determination of Heavy Metal Contamination and Pollution Indices of Roadside Dust in Dhaka City, Bangladesh. In: *Processes* 9.10 (2021), p. 1732. DOI: 10.3390/pr9101732.

[44] G.-L. Shi, Y.-C. Feng, F. Zeng, X. Li, Y.-F. Zhang, Y.-Q. Wang, and T. Zhu. Use of a Nonnegative Constrained Principal Component Regression Chemical Mass Balance Model to Study the Contributions of Nearly Collinear Sources. In: *Environmental Science & Technology* 43.23 (2009), pp. 8867–8873. DOI: 10.1021/es902785c.

[45] E. Apeagyei, M. S. Bank, and J. D. Spengler. Distribution of heavy metals in road dust along an urban-rural gradient in Massachusetts. In: *Atmospheric Environment* 45.13 (2011), pp. 2310–2323. DOI: 10.1016/j.atmosenv.2010.11.015.

[46] B. Wei, F. Jiang, X. Li, and S. Mu. Spatial distribution and contamination assessment of heavy metals in urban road dusts from Urumqi, NW China. In: *Microchemical Journal* 93.2 (2009), pp. 147–152. DOI: 10.1016/j.microc.2009.06.001.

[47] S. K. Pal and R. Roy. Evaluating heavy metals emission' pattern on road influenced by urban road layout. In: *Transportation Research Interdisciplinary Perspectives* 10 (2021), p. 100362. DOI: 10.1016/j.trip.2021.100362.

[48] M. Fiala and H.-M. Hwang. Influence of Highway Pavement on Metals in Road Dust: a Case Study in Houston, Texas. In: *Water, Air, & Soil Pollution* 232.5 (2021), p. 185. DOI: 10.1007/s11270-021-05139-7.

[49] A. Liu, Y. Ma, J. M. Gunawardena, P. Egodawatta, G. A. Ayoko, and A. Goonetilleke. Heavy metals transport pathways: The importance of atmospheric pollution contributing to stormwater pollution. In: *Ecotoxicology and Environmental Safety* 164 (2018), pp. 696–703. DOI: 10.1016/j.ecoenv.2018.08.072.

[50] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2023. URL: https://www.R-project.org/.

[51] QGIS Development Team. *QGIS Geographic Information System*. 2023. URL: https://www.qgis.org.

[52] P. Bogaert, G. Diélie, A. Briffault, B. De Saint-Hubert, and M. A. Verbanck. Identifying proxies and mapping heavy metals concentrations in city road dusts: A case study in the Brussels-Capital Region, Belgium. In: *Heliyon* 9.2 (2023), e13312. DOI: 10.1016/j.heliyon.2023.e13312.

[53] IBSA. *Chiffres-clés de la Région bruxelloise*. URL: https://ibsa.brussels/chiffres/chiffres-cles-de-la-region.

[54] IBSA. *Anderlecht - chiffres-clés*. URL: https://ibsa.brussels/chiffres/chiffres-cles-par-commune/anderlecht.

[55] Région de Bruxelles-Capitale. *Port de Bruxelles - Région bruxelloise*. URL: https://be.brussels/a-propos-de-la-region/les-organismes-regionaux/port-de-bruxelles?set_language=fr.

[56] H. Akaike. A new look at the statistical model identification. In: *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

[57] A. Hebbali. *olsrr: Tools for Building OLS Regression Models*. 2020. URL: https://CRAN.R-project.org/package=olsrr.

[58] J. Fox and G. Monette. Generalized Collinearity Diagnostics. In: *Journal of the American Statistical Association* 87.417 (1992), pp. 178–183. DOI: 10.1080/01621459.1992.10475190.

[59] N. M. Schroeder and A. Panebianco. Sociability strongly affects the behavioural responses of wild guanacos to drones. In: *Scientific Reports* 11.1 (2021), p. 20901. DOI: 10.1038/s41598-021-00234-5.

[60] T. T. Duong and B.-K. Lee. Determining contamination level of heavy metals in road dust from busy traffic areas with different characteristics. In: *Journal of Environmental Management* 92.3 (2011), pp. 554–562. DOI: 10.1016/j.jenvman.2010.09.010.

[61] S. N. Taghavi, H. Kamani, M. H. Dehghani, R. Nabizadeh, N. Afshari, and A. H. Mahvi. Assessment of Heavy Metals in Street Dusts of Tehran Using Enrichment Factor and Geo-Accumulation Index: A. In: *Health Scope* In Press.In Press (2019). DOI: 10.5812/jhealthscope.57879.

[62] R. L. Rudnick and S. Gao. "Chapter 3 : Composition of the Continental Crust". In: *Treatise on geochemistry*. Ed. by H. D. Holland and K. K. Turekian. 1st ed. Amsterdam ; Boston: Elsevier/Pergamon, 2004. URL: https://www.geol.umd.edu/~rudnick/PDF/Rudnick_Gao_Treatise.pdf.

[63] B. Wei and L. Yang. A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China. In: *Microchemical Journal* 94.2 (2010), pp. 99–107. DOI: 10.1016/j.microc.2009.09.014.

[64] J. Acosta, M. Gabarrón, A. Faz, S. Martínez-Martínez, R. Zornoza, and J. Arocena. Influence of population density on the concentration and speciation of metals in the soil and street dust from urban areas. In: *Chemosphere* 134 (2015), pp. 328–337. DOI: 10.1016/j.chemosphere.2015.04.038.

[65] B. Kończak, M. Cempa, Ł. Pierzchała, and M. Deska. Assessment of the ability of roadside vegetation to remove particulate matter from the urban air. In: *Environmental Pollution* 268 (2021), p. 115465. DOI: 10.1016/j.envpol.2020.115465.

[66] C. Brandeleer, T. Ermans, M. Hubert, I. Janssens, P. Lannoy, C. Loir, and P. Vanderstraeten. *Le partage de l'espace public en région de Bruxelles-Capitale*. URL: https://mobilite-mobiliteit.brussels/sites/default/files/le_partage_de_lespace_public.pdf.

All web pages are accessible at the date of submission of this master thesis on the 12/06/2023.

Geospatial Modeling of Heavy Metals Concentrations in Urban Road Dust

Benoit de Saint-Hubert

The aim of this master thesis is to perform an in-depth analysis of the spatial distribution of heavy metals (HMs) concentrations in road dust (RD) in a municipality of the Brussels Capital-Region (BCR) and identify the most relevant proxies (i.e., proxy variables) for geospatial modeling. For this goal, 128 samples of RD were collected in the Municipality of Anderlecht (AM) over 3 consecutive years (2019-2021). The concentrations of Cd, Cr, Cu, Ni, Pb and Zn in the finest fraction of these samples (ø < 250 μm) were determined by ICP-OES. Based on a literature review, several continuous and categorical proxies were collected to build a geospatial dataset at the scale of the BCR. For each heavy metal, the best candidate Multivariate Linear Regression (MLR) model was selected by a forward stepwise procedure. Their ability to predict HMs concentrations remains limited with R^2_a values around 0.5 for all models. The most relevant proxies were : "Distance from the center of the BCR", "Land use", "Road hierarchy" and "Roadside parking occupation". Investigation of the models residuals trough a spatial dependence analysis revealed the potential omission of one or multiple proxies in our analyses. These MLR models were used afterwards to predict HMs concentrations for all the road segments of the AM. Despite the moderate performances of the models, we believe that the resulting maps provide useful information for decision-making regarding mitigation measures for the local authorities and public services. We also believe that our predictions could be extended to the whole BCR and that our approach is reproducible for other urban areas. The results presented in this master thesis were summarized and disseminated in a publication.

Keywords : Urban Pollution – Heavy Metals – Road Dust – Geospatial Modeling – Multivariate Linear Regression – Brussels-Capital Region

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**