



# Spatial, infrastructural and consumer characteristics underlying spatial variability in residential energy and water consumption in Amsterdam

Ilse M. Voskamp<sup>a,b,\*</sup>, Megan N. Visscher<sup>c,1</sup>, Corné Vreugdenhil<sup>c</sup>, Ron J.A. Van Lammeren<sup>c</sup>, Nora B. Sutton<sup>a</sup>

<sup>a</sup> Environmental Technology, Wageningen University & Research, P.O. Box 17, 6700 AA, Wageningen, the Netherlands

<sup>b</sup> Landscape Architecture Group, Wageningen University & Research, P.O. Box 47, 6700 AA, Wageningen, the Netherlands

<sup>c</sup> Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, P.O. Box 47, 6700 AA, Wageningen, the Netherlands

## ARTICLE INFO

### Keywords:

Urban metabolism  
Residential consumption  
Linear regression  
Sustainable resource management

## ABSTRACT

To design effective strategies for sustainable urban resource management, it is essential to understand which urban characteristics underlie consumption patterns. We used multiple linear regression analyses to examine sixteen factors on their explanatory power for spatial variation in residential electricity, gas and water consumption in Amsterdam. Four models per resource were used, based on distinct spatial units aggregating high-resolution data: neighbourhoods, districts, 100 m squares and 500 m squares. We found twelve explanatory variables for spatial variability in consumption in total and nine or ten per resource. The number and relative importance of explanatory variables varies with the spatial units used. Overall, neighbourhood models explain variance in consumption data best (adjusted  $R^2 = 0.88, 0.86, 0.74$ ). *Income level* and *building type* stand out for having high relative importance (top 4) in all four models for two of the three resources; *migration history* shows an important correlation with water consumption, which was not described hitherto. We conclude that explanatory variables for resource consumption are sensitive to size and shape of spatial units used. We recommend to use future high resolution studies for different resources of interest to determine which spatial and temporal resolutions of analysis can support urban planners and designers in formulating context-specific interventions.

## 1. Introduction

With more than half of the world's population dwelling in cities (UN, 2019), urban areas are facing unprecedented challenges regarding efficient resource consumption and related pollution issues (Agudelo-Vera, Mels, Keesman, & Rijnaarts, 2011; Pulido Barrera, Rosales Carreón, & de Boer, 2018). At the same time, climate change is expected to further increase urban resource demands, including increased seasonal water demand and demands for building heating and cooling (Ciancio, Salata, Falasca, Curci, & Golasi, 2020; Rasifaghihi, Li, & Haghighat, 2020). These impacts are not only expected to differ among cities, given their different geographic locations and climate conditions, but also within cities due to microclimate differences (Chen, Han, & Vries, 2020). Hence, urbanization, climate change and the interlinked processes of

change pose a complex challenge to urban planning and decision making (McPhearson, Haase, Kabisch, & Gren, 2016).

To formulate sustainability strategies and affiliated interventions on consumption reduction, urban planners and designers need detailed insight into urban resource flows (Kennedy, Pincetl, & Bunje, 2011; Pulido Barrera et al., 2018; Voskamp et al., 2018). To inform such strategies and interventions, it is particularly essential to comprehend not only where resource consumption takes place, but also how the specific conditions and urban characteristics at hand affect this consumption (Bettignies et al., 2019; Broto, Allen, & Rapoport, 2012; Cui, 2018; Jayarathna et al., 2017). After all, cities are typically characterised by the heterogeneity of their demographic, socio-economic, environmental and cultural characteristics that are forming complex mechanisms which underlie differences within and amongst cities (Li &

\* Corresponding author at: P.O. Box 17, 6700 AA, Wageningen, the Netherlands.

E-mail address: [ilse.voskamp@wur.nl](mailto:ilse.voskamp@wur.nl) (I.M. Voskamp).

<sup>1</sup> These authors contributed equally to this publication.

Kwan, 2018; Pickett et al., 2017). Scholars agree that it is essential to acquire better understanding of the factors and mechanisms underlying urban resource demands (Cui, 2018; Rosales Carreón & Worrell, 2018; Villarín, 2019; Voskamp, Sutton, Stremke, & Rijnaarts, 2020; Zhang, Yang, & Yu, 2015). A fine level of spatial detail must be adopted in studies aiming to provide this understanding, for only then can the spatial variation of urban and resource characteristics within the city be captured (Beloin-Saint-Pierre et al., 2017; Li & Kwan, 2018; Pincetl, Bunje, & Holmes, 2012; Pistoni & Bonin, 2017).

Hitherto, various factors have been described to underlie variability in residential energy and water demand. Studies have substantiated correlations between consumption and climate and weather conditions (Meng et al., 2020; Rasifaghihi et al., 2020), demographic and economic factors (Bettignies et al., 2019) and urban and architectural morphological characteristics (Chen, Han, & Vries, 2020; Fox, McIntosh, & Jeffrey, 2009; You & Kim, 2018). Others revealed a combination of climate, socio-economic and morphological characteristics as explanatory variables for energy (Chen et al., 2020a; Kennedy et al., 2015; Wiedenhofer, Lenzen, & Steinberger, 2013) and water consumption (Chang, Bonnette, Stoker, Crow-Miller, & Wentz, 2017; Jayarathna et al., 2017; Stoker & Rothfeder, 2014; Villarín, 2019). To date, however, few studies consider a long-list of potential factors of influence. They account only for one or a small number of independent variables instead, with *household size* and *income level* being frequently studied (Bich-Ngoc & Teller, 2018; Wiedenhofer et al., 2013). Still, it is known that consumption patterns cannot be explained by individual factors but rather by a large number of interconnected factors that have direct and indirect effects (Bich-Ngoc & Teller, 2018; Voskamp et al., 2020; You & Kim, 2018). Moreover, studies that consider both water and energy demand are rare (e.g. Pincetl et al., 2014) and those assessing correlations at different levels of spatial resolution too (e.g. Chang et al., 2017; Ghavidelfar, Shamseldin, & Melville, 2017; Kennedy et al., 2015). Studies that look into spatial variability of consumption using more detailed data aggregation levels than city level and that rely upon metered data are still uncommon (Bettignies et al., 2019; Bich-Ngoc & Teller, 2018; Voskamp et al., 2020). Yet, multi-level, high resolution case studies can provide insight in the scale- and city-dependent behaviour of explanatory variables for consumption, about which still little is known (Bettignies et al., 2019; Chang et al., 2017).

In this study we address the need for high resolution studies that provide insight in the factors underlying spatial variability in energy and water consumption at different spatial scales. We use multiple linear regression analysis to examine a large set of factors as explanatory variables for the spatial variation in residential energy (electricity, natural gas) and water consumption within the city of Amsterdam. We examined the relative importance of the explanatory variables for each resource in four multiple linear regression models, each relying upon another spatial data aggregation level.

## 2. Theory

This research builds upon a systematic review of the factors underlying spatial and temporal variations in energy and water consumption to ensure a comprehensive consideration of potential explanatory variables for spatial variability in consumption (Voskamp et al., 2020). The authors reveal not only a long-list of factors, but they also conceptualize their coherence and categorize them in line with the work of Dijst et al. (2018) and Poças Ribeiro, Harmsen, Rosales Carreón, and Worrell (2019). Three main categories of factors are differentiated: I) the *needs* of consumers, II) *facilitators/constraints* that increase or decrease the probability or degree to which resource consuming activities are

triggered, comprising consumer, resource, spatial and infrastructural characteristics<sup>2</sup> and III) *drivers*, which encompass large scale developments in societal context. For more theoretical background on these three sorts of factors and their effect on consumption, we refer to Dijst et al. (2018), Poças Ribeiro et al. (2019) and Voskamp et al. (2020). Of these sorts of factors, only spatial, infrastructural and consumer characteristics (*facilitators/constraints*) have the feasibility to explain spatial variation in residential energy and water consumption patterns within a city. *Drivers* and *resource characteristics* can be assumed homogeneous at city-level, and consumer *needs* are disregarded because no spatially explicit data are available on these type of traits of individual consumers, such as individual aspirations.

Accordingly, spatial, infrastructural and consumer characteristics are assessed in this study as potential explanatory variables for spatial variation in residential energy and water consumption, given their potentially facilitating or constraining effect on consumption (Table 1). The factors listed in Table 1 comprise those listed in the aforementioned review and additional spatial, infrastructural and consumer characteristics derived from other reviews and case studies. Voskamp et al. (2020) namely used strict inclusion criteria for their review of urban metabolism (UM) literature, excluding for example studies based on survey data, because critical data quality assessment is central to their systematic approach to reviewing literature. So, although their categorization is comprehensive, the authors stress that other fields of literature should be consulted to complement their list of factors and to also assess these factors on their influence in future case studies. We selected the following additional factors for our case study: *property value*, *home ownership*, *rental housing type*, *education level*, *age*, and *migration history* (Baiocchi, Creutzig, Minx, If, & Pichler, 2015; Bich-Ngoc & Teller, 2018; Chang et al., 2017; House-Peters & Chang, 2011; Lenzen, Dey, & Foran, 2004; Pincetl et al., 2014; Porse, Derenski, Gustafson, Elizabeth, & Pincetl, 2016; Wiedenhofer et al., 2013).

## 3. Methods and materials

### 3.1. Study area and levels of investigation

This research was conducted using data of Amsterdam municipality, the capital city of the Netherlands. We investigated the explanatory variables for spatial variation in water and energy consumption in this study area using four spatial aggregation levels to identify whether results are affected by the resolution of observation. These four spatial aggregation levels are given by the data provider and are aggregated for privacy reasons as it concerns privacy-sensitive information on individual household level. Spatial aggregations using artificial or administrative boundaries are prone to the Modifiable Areal Unit Problem (MAUP), which indicates that variation in results occurs when the same analysis is applied to the same data but using different aggregation units (Openshaw, 1984).<sup>3</sup> To limit the possibility of MAUP negatively affecting the results, all four spatial aggregations are taken into account and results compared with each other to draw conclusions independent of the spatial aggregation. Two units are based on administrative

<sup>2</sup> We make use of the definitions by Voskamp et al. (2020): “Spatial characteristics describe the characteristics of the built and non-built surfaces within the urban landscape and reflect the form and functions of these spaces. These characteristics range from very generic city level indicators such as population density, to specific building characteristics such as building size. ... Infrastructural characteristics refer to the characteristics of the physical networks and their system components that are providing goods and services to a city, for instance systems for transport, resource supply and telecommunications. Infrastructural facilitators/constraints thus include specifications of the city networks as well as of specific system components.”

<sup>3</sup> For more details on this approach to MAUP, see for example Nelson and Brewer (2017).

**Table 1**

Overview of potential explanatory variables for spatial variability in energy and water consumption.

Type of urban characteristic	Factor	Potential facilitating / constraining effect	Type of correlation			Reference
			Electricity	Natural gas	Water	
Spatial	Population density	Lower consumption per household with a higher population density	–	–	–	1
	Amount of tree canopy cover	Lower energy usage for heating and cooling with a higher presence of productive tree canopy cover	–	–	n.d.	1
	Garden type	Higher water consumption with larger gardens and higher presence of grass	n.d.	n.d.	+	1
	Building size	Higher energy and water consumption with larger building size	+	+	+	1
	Building type	Lower usage with a higher share of multiple family buildings (e.g. apartment buildings), at the expense of the share of single family dwellings	–	–	–	1
	Building age	Lower natural gas usage and higher electricity and water usage with more recent building year	+	–	+	7, 8
	Property value	Higher water consumption with higher tax assessed values	n.d.	n.d.	+	4
	Home ownership	Higher energy consumption with a higher share of owner-occupied houses, at the expense of the share of rental houses	+	+	n.d.	8
	Rental housing type	Different energy usage levels with a higher share of rental housing owned by housing associations			n.d.	8
	Presence of supply network	Higher electricity consumption and lower natural gas consumption with electric heating or a district heating network present	+	–	n.d.	1
Infrastructural	Presence of a pool	Higher electricity and water consumption with a private swimming pool present	+	n.d.	+	1
	Home appliances	Higher electricity consumption with more home appliances (e.g. air conditioning) present	+	n.d.	n.d.	1
	Sanitary fittings	Lower water consumption with high efficiency fittings (e.g. low-flow taps and showers) present	n.d.	n.d.	–	1
	Income level	Higher consumption with a higher income level	+	+	+	1
Consumer	Lifestyle	Higher electricity consumption with a high-income lifestyle	+	n.d.	n.d.	1
	Education level	Different energy and water consumption levels with a different education level				3, 5, 6
	Age	Different energy and water consumption levels with a different age structure in the population				3, 5, 6, 9
	Migration history	Different consumption levels with a different ethnic community, indicated by their migration history				2, 3, 5
	Household size	Lower per capita and higher household consumption levels with larger household size	+	+	+	1

Note: ‘+’ indicates that literature describes a positive relationship, meaning that an increase in the factor is correlated with a higher consumption level; ‘–’ indicates that literature describes a negative relationship meaning that an increase in the factor is correlated with a decrease in consumption; ‘n.d.’ indicates that this relationship is not described; and when the cell is blank, literature does not present conclusive supporting evidence of the type of correlation. References: 1: Voskamp et al. (2020); 2: Baiocchi et al. (2015); 3: Bich-Ngoc and Teller (2018); 4: Chang et al. (2017); 5: House-Peters and Chang (2011); 6: Lenzen et al. (2004); 7: Pincetl et al. (2014); 8: Porse et al. (2016); 9: Wiedenhofer et al. (2013).

divisions within the city, neighbourhoods and districts, and two based on a grid tessellation, 100 m and 500 m squares (Fig. 1). These spatial units correspond to the data aggregation levels used by the Dutch Bureau of Statistics to publish data (CBS, 2016a, 2016b). Overall, Amsterdam municipality consists of 480 neighbourhoods with an average size of 21 ha (range: 2–1012 ha), 99 districts with an average size of 198 ha (range: 17–2899 ha), 6711 squares of 1 ha and 993 squares of 2.5 ha (Fig. 1).

### 3.2. Data collection and preparation

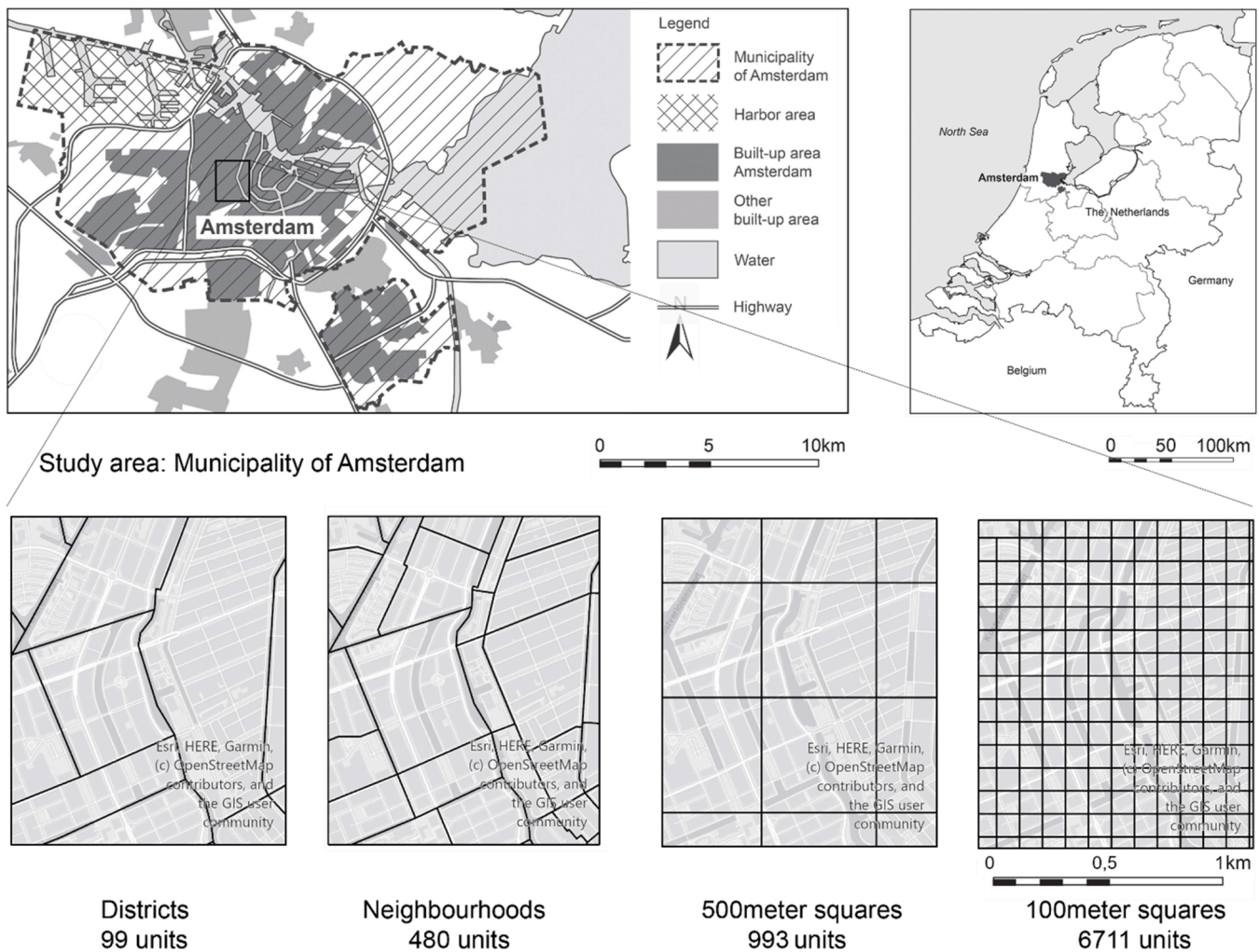
Open datasets were the main source of data for the residential consumption data and data representing the potential explanatory variables. Electricity and natural gas consumption data for 2016 are available as open data through the Dutch Central Bureau of Statistics, whilst the water usage data were provided by the drinking water provider Waternet for 2014 (Supplementary Material Tables S1 and S2). Acquired consumption data consists of aggregated yearly household meter readings, aggregated by the four levels described in 3.1. The water consumption dataset partially contains estimates by Waternet because 30 % of the households in Amsterdam had no water meter at the time. We acquired data for the nineteen factors listed in Table 1 from various Dutch open data sets on each data aggregation level. No data were obtained for *presence of a pool*, *home appliances* and *sanitary fitting*. Therefore, these factors were not considered in our analysis. For the remaining sixteen factors, we used the available data to formulate a

long-list of variables that could represent these factors, including 59 variables for the administrative units and 29 and 33 for the 100 m and 500 m-grids (Supplementary Material Table S3).

### 3.3. Evaluation of collinearity among explanatory factors

Subsequently we evaluated collinearity within the long-list of variables, to select appropriate variables for all factors and decide whether different factors must be observed together in the regression analysis. Collinearity is present when two or more of the independent variables that ought to explain a dependent variable are highly correlated to each other (Dormann et al., 2013). This can be problematic, as correlated independent variables can interfere with each other, resulting in incorrect selection of significant variables for the multiple linear regression. We used the Pearson coefficient ( $r$ ) to detect the degree of linear relationship between two independent variables, with  $r = 0$  signifying no relationship and  $r = 1$  a perfect correlation. Because various scholars draw diverging conclusions on a threshold value for when correlation among explanatory variables becomes harmful, varying roughly between 0.7 and 0.8 (Dormann et al., 2013; Pavelescu, 2014), we choose a threshold of  $r = 0.75$  for collinearity. If collinearity was present, we eliminated the variable that has the highest correlation with all other variables. In case of collinearity between two variables representing different factors, the variable represented at all four spatial units was selected. Such was the case for the following pairs of collinear factors (see Table 3), where the latter one of each pair was selected: *income level* and *value of the house*, and *education level* and *migration*





**Fig. 1.** The study area Amsterdam municipality and its location in the Netherlands (top) and the four spatial aggregation levels adopted in this study illustrated for the same surface area in Amsterdam (bottom).

*history.* After the collinearity evaluation, we also checked whether the other assumptions of multiple linear regression<sup>4</sup> were not severely distorted as this could lead to bias in relationship estimation and unreliable significance tests (Williams, Grajales, & Kurkiewicz, 2013).

### 3.4. Multiple linear regression analyses

We developed a multiple linear regression model at each of the four spatial aggregation levels for all three resources, resulting in twelve models in R (R Core Team, 2017). We performed backward stepwise elimination with a threshold significance value of  $p = 0.05$  to identify significant explanatory variables. This procedure eliminates the least significant independent variables from the model in a stepwise manner, until all independent variables meet the threshold value (Hastie, Tibshirani, & Friedman, 2009). We then compared the results per resource

to check for potential MAUP effects. If we identified major difference in regression results across the models, such as a positive relationship using one spatial aggregation level and a negative using another, we considered the relationship affected by MAUP. Such conflicting relationships over the different spatial levels were assumed to be the result of MAUP effects, because an inconsistent direction of the relationship over the spatial levels shows a dependence for spatial aggregation. After this evaluation, we created multiple regression models at each level of spatial aggregation with only the variables that were found significant and un-affected by MAUP. We calculated for each independent variable in all models its relative importance (*relaimpo* package v2.2-3, Groemping & Lehrkamp, 2018), expressing how much the variable contributes to the overall model (Grömping, 2006).

## 4. Results

### 4.1. Findings across models

Our research gives an overview of explanatory variables for residential electricity, natural gas and water consumption in Amsterdam, using four different spatial data aggregation levels (Tables 2A–2C). A comparison of the model performance (second row, Tables 2A–2C) shows that the electricity models have the highest fit (adjusted  $R^2 = 0.58 < 0.88$ ) and the water models the lowest (adjusted  $R^2 = 0.46 < 0.74$ ). The 100 m-grid natural gas model has the lowest fit

<sup>4</sup> To test the assumption for multiple linear regression (Poole & O'Farrell, 1971), in addition to collinearity evaluation, we created scatterplots of the independent variables against the dependent variable to evaluate whether a linear relationship between the dependent and independent variables could be assumed. To test for homoscedasticity, meaning that the random disturbance is the same across all values of the independent variable; the residuals versus fitted values and scale-location were plotted. For the assumption of multivariate normality, a q-q plot was created and autocorrelation in the residuals was tested by inspecting the residual plot.

**Table 2A**

Explanatory variables for spatial variability in residential electricity consumption in Amsterdam.

Electricity			Districts		Neighbourhoods		500m-grid		100m-grid	
Adjusted R <sup>2</sup> per model for each spatial unit			0.80		0.88		0.78		0.58	
Factor	Indicator variable	Correlation	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance
Building size	Percentage of houses with living space area above 80 m <sup>2</sup>	Positive	<i>Not significant</i>		1	0.22	<i>NA</i>		<i>NA</i>	
Income level	Property value	Positive	3	0.17	2	0.16	2	0.16	2	0.17
Building type	Percentage of multiple family homes	Negative	1	0.40	3	0.15	1	0.24	3	0.16
Household size	Household size	Positive	<i>Not significant</i>		4	0.14	3	0.16	1	0.37
Rental housing type	Housing associations	Negative	4	0.14	5	0.10	5	0.13	4	0.10
Age	Percentage of people between 25 and 44 years old	Negative	<i>Not significant</i>		6	0.07	4	0.14	5	0.10
Population density	Population density	Negative	2	0.21	7	0.07	6	0.08	8	0.02
Garden type	Garden size	Positive	<i>Not significant</i>		8	0.05	8	0.04	7	0.03
Migration history	Percentage of people with non-western migration history <sup>1</sup>	Positive	6	0.04	9	0.03	7	0.05	6	0.05
Presence of supply network	Percentage of houses connected to district heating	Positive	5	0.05	10	0.02	<i>NA</i>		<i>NA</i>	

**Table 2B**

Explanatory variables for spatial variability in residential natural gas consumption in Amsterdam.

Natural gas			Districts		Neighbourhoods		500m-grid		100m-grid	
Adjusted R <sup>2</sup> per model for each spatial unit			0.90		0.86		0.65		0.42	
Factor	Indicator variable	Correlation	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance
Presence of supply network	Percentage of houses connected to district heating	Negative	1	0.27	1	0.33	<i>NA</i>		<i>NA</i>	
Income level	Property value	Positive	4	0.15	2	0.29	3	0.22	2	0.34
Garden type	Garden size	Positive	5	0.05	3	0.11	4	0.08	3	0.06
Building age	Percentage of houses with building year after 2000 <sup>2</sup>	Negative	1	0.27	4	0.11	1	0.37	1	0.53
Building type	Percentage of multiple family homes	Negative	<i>Not significant</i>		5	0.09	2	0.27	7	0.01
Age	Percentage of people between 25 and 64 years old	Negative	6	0.04	6	0.04	6	0.02	5	0.02
Household size	Household size	Positive	7	0.04	7	0.03	5	0.04	8	0.01
Population density	Population density	Negative	<i>Not significant</i>		<i>Not significant</i>		<i>Not significant</i>		4	0.03
Home ownership	Percentage of owner-occupied housing	Negative	3	0.17	<i>Not significant</i>		<i>Not significant</i>		6	0.01

**Table 2C**

Explanatory variables for spatial variability in residential water consumption in Amsterdam.

Water			Districts		Neighbourhoods		500m-grid		100m-grid	
Adjusted R <sup>2</sup> per model for each spatial unit			0.68		0.74		0.46		0.57	
Factor	Indicator variable	Correlation	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance	Rank	Relative importance
Household size	Household size	Positive	<i>Not significant</i>		1	0.30	<i>Not significant</i>		2	0.36
Migration history	Percentage of people with non-western migration history <sup>1</sup>	Positive	1	0.32	2	0.23	3	0.18	1	0.41
Building type	Percentage of multiple family homes	Negative	3	0.22	3	0.17	1	0.40	3	0.07
Building size	Percentage of houses with living space area above 80 m <sup>2</sup>	Positive	<i>Not significant</i>		4	0.13	<i>NA</i>		<i>NA</i>	
Age	Percentage of people between 25 and 44 years old	Negative	2	0.27	5	0.12	2	0.25	5	0.04
Rental housing type	Percentage of rental housing owned by housing associations	Negative	5	0.06	6	0.04	5	0.04	6	0.04
Building age	Percentage of houses with building year after 2000 <sup>2</sup>	Negative	6	0.06	7	0.01	6	0.03	7	0.01
Income level	Property value	Positive	<i>Not significant</i>		<i>Not significant</i>		7	0.01	4	0.05
Garden type	Garden size	Positive	4	0.07	<i>Not significant</i>		4	0.10	<i>Not significant</i>	

<sup>1</sup> Non-western migration history comprises first and second generation migration from Africa, Latin-America Asia and Turkey (CBS, 2014, 2016a, 2016b).<sup>2</sup> For the grids this is a building year after 2005.

of all models (adjusted  $R^2 = 0.42$ ). Findings furthermore indicate that the models using administrative divisions show a better fit than those using grids. The variance in electricity and water consumption is best explained with the neighbourhood models, (adjusted  $R^2 = 0.88$  and  $0.74$ ), whilst for natural gas the district model outperforms the neighbourhood model (adjusted  $R^2$  values  $0.90$  versus  $0.86$ ). In the electricity findings, the 500 m-grid model stands out for having a comparable fit to the district model.

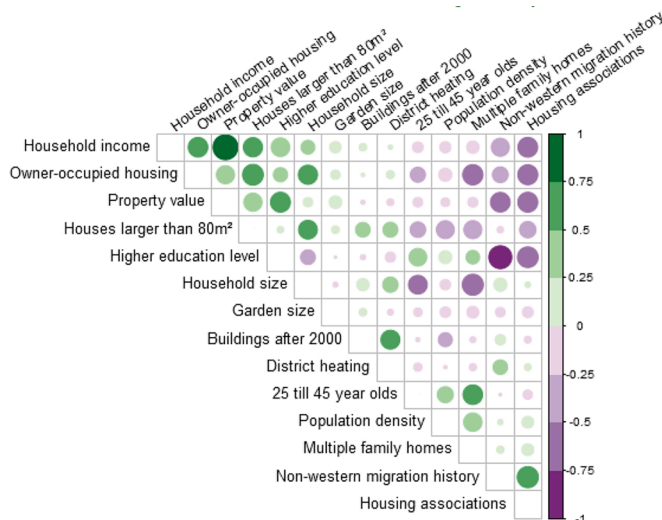
Overall, twelve independent variables are found to be correlated with residential energy and water consumption in Amsterdam, having significance at the 95 % confidence level. The results of the multiple linear regression analyses (Tables 2A–2C) show that five of these factors are explanatory variables for the consumption of all three resources: *income level*, *household size*, *age*, *building type* and *garden type*. In addition to these, *property value* can be considered a factor of importance for residential energy and water consumption in Amsterdam. This factor was not considered as variable in the models, and thus does not show in Tables 2A–2C, because it was found to be strongly collinear with average *income level* (depending on indicator variables,  $r = 0.817 < 0.919$ ) (Table 3).<sup>5</sup> We also find that electricity and water consumption have the explanatory variables *building size*, *rental housing type* and *migration history* in common. The high collinearity of *migration history* with *education level* ( $r = 0.777$ ) (Table 3) excluded *education level* from the regression models; this does however also imply a relationship between *education level* and electricity and water consumption. *Population density* shows as explanatory variable for electricity and natural gas consumption, *building age* for natural gas and water consumption and *home ownership* is a unique explanatory variable for natural gas consumption.

#### 4.2. Explanatory variables for residential electricity consumption

Using four multi linear regression models, we found a total of ten explanatory variables for electricity consumption in Amsterdam

**Table 3**

Cross-correlations between the twelve explanatory variables found and the factors collinear to those.



Note: Green indicates a positive correlation and purple a negative one; the size and shade of the circle show the strength of the correlation.

Values  $0.75 < 1$  indicate collinearity; collinear factors were excluded from our regression analyses.

<sup>5</sup> Average property value was employed to represent the factor income level in this study, because data was present for all four spatial units, whereas this was not the case for the indicator variables for income level.

(Table 2A). *Income level*, *building type*, *rental housing type*, *population density* and *migration history* are found significant for all four models. The factors *income level* and *building type* and *household size* are of high importance for all four models, ranking in the top 3 for all models (Table 2A). A higher *income level* is positively correlated with electricity consumption, whereas the presence of more multiple family homes has a negative correlation. Although *building size* ranks first for neighbourhoods, this factor is not significant for districts and could not be assessed for the grid models. *Household size* is not found significant for the district model, albeit a significant explanatory variable in the other electricity models. In the 100 m-grid model *household size* is the most important explanatory variable. Likewise, *percentage of people between 25 and 44 years old* and *garden size* are not significant for the district model, whereas they are significant in the other three models. Finally, *population density* has a high relative importance in the districts model ( $0.21$ ), whilst it has considerably lower importance in the other three models.

#### 4.3. Explanatory variables for residential natural gas consumption

We found nine explanatory variables for residential consumption of natural gas in Amsterdam (Table 2B). Five of these are significant with all spatial units used: *income level*, *garden type*, *building age*, *age*, and *household size*. *Building age* is the most important variable in all models except the neighbourhood model, indicating that lower residential gas consumption levels are found in areas that have a higher share of dwellings built after 2000. For the district model, this variable is of equal importance to *presence of district heating* (relative importance  $0.27$ ); no data was available for the grids. At neighbourhood level, the share of homes connected to district heating is found to have an even stronger negative correlation with natural gas consumption (relative importance  $0.33$ , ranking first). *Income level* exhibits an important, positive relationship with natural gas consumption in all models and a similar, but less important correlation is found for *garden type*. Whereas *percentage of multiple family homes* is not significant for the districts model, it is significant for the other three natural gas models and is even of high importance for the 500 m-grid model ( $0.27$ ). The *share of owner-occupied houses*, on the other hand, ranks only high (number 3) in the district model. Finally, *population density* is only significant for the 100 m-grid model.

#### 4.4. Explanatory variables for residential water consumption

We found also nine explanatory variables for Amsterdam's residential water consumption, including five that are significant for all models: *migration history*, *building type*, *age*, *rental housing type* and *building age* (Table 2C). A higher *share of people with a non-western migration history* and a lower *share of multiple family homes* are consistently correlated with higher water consumption levels. Both *migration history* and *building type* consistently rank in the top 3. Meanwhile, *the share of people between the age of 25 and 44* negatively correlated with residential water consumption for all models. This was the second most important variable for the district and 500 m-grid models with a relative importance of  $0.27$  and  $0.25$ . Although average *household size* in an area is an explanatory variable with very high importance for the neighbourhoods and 100 m-grid water models, ranking first and second respectively, this factor is not found significant for the other two models. *Building size* could only be assessed for the models using administrative units, and is found only significant for the neighbourhood model. Finally, *income level* and the average *garden size* are both found significantly correlated to residential water consumption for two of the four models.

## 5. Discussion

### 5.1. Findings across spatial units used

Our findings show that spatial variability in energy and water consumption levels is better explained by models using administrative units



than by the grid models. Overall the model fit is highest for the neighbourhood models and lowest for the 100 m-grid models (Tables 2A–2C). Partly these differences in performance reflect that less data was available for the grid models and that data availability has affected the 100 m-grid models most, owing to privacy restrictions. The performance of the grid models is probably especially affected by the lack of data on *building size* and *presence of supply network* aggregated in square units, considering that these factors have very high importance in the neighbourhood models. Another explanation for the differences in model fit is the different geometry and size of the spatial units used in the models. Neighbourhoods and districts have been shaped historically and follow urban form and natural boundaries like water or roads, whereas the generally smaller square units ignore these “natural” boundaries. Hence, the square units aggregate data with relatively higher variance in spatial and infrastructural characteristics. This effect of unit shape especially shows when comparing 500 m-grid and neighbourhood results, as these models had comparable percentages of observations with missing values. The 500 m-grid models have an adjusted  $R^2$  of 0.78, 0.65 and 0.46, whereas this is 0.88, 0.86 and 0.74 for the neighbourhood models. It is recommended for future studies to explore further whether models using homogenous aggregates of spatial and infrastructural characteristics explain variance in consumption levels better. This would be a valuable insight, because urban form based units resonate with the urban planning and design field.

Our results confirm earlier research showing that explanatory variables for resource consumption are sensitive to the spatial units used. So far this was shown only in studies comparing two different levels of spatial resolution, in the case of water for census area versus household or census block level (Chang et al., 2017; Ghavidelfar et al., 2017) and in the case of energy for city versus census area level (Bettignies et al., 2019; Kennedy et al., 2015). By considering four levels of spatial aggregation, we show that it matters at which level of detail within the city

spatial variation of urban and resource characteristics is studied. Our results show that the number and relative importance of explanatory variables varies with the spatial units used. Findings are particularly distinctive for the three districts models, which make use of the largest spatial units and thus represent the lowest spatial resolution in this study (see Fig. 1). It is for example striking that *household size* is not significant for the electricity and water district models. This factor ranks high for the other electricity and water models and the effect of *household size* on water and energy consumption is substantiated in many studies (e.g. House-Peters & Chang, 2011; Lenzen et al., 2004; Weisz & Steinberger, 2010). It also stands out that *home ownership* was found to have a much greater importance for the natural gas district model than the other models; the same applies to *population density* for explaining electricity consumption. Kennedy et al. (2015) suggest that *population density* is a relevant factor to explain variation in electricity consumption at a lower resolution of analysis, such as at district or city level, whilst other factors are appropriate explanatory variables at a more detailed level instead. Indeed, our results seem to show that the significance of explanatory variables for consumption is dependent on the spatial unit used. To get a better understanding of the scaling impact, we recommend future high-resolution studies to also use different spatial aggregation levels when assessing factor-consumption relationships. Thereby, the modifiable areal unit problem (MAUP) should be considered and units should be included that correspond with the field of decision-makers, planners and designers.

## 5.2. Generic and Amsterdam specific explanatory variables

The majority of factors mentioned in the literature were found to be significant explanatory variables, with effects that are in line with previous studies (Table 4). If we compare our results with evidence from literature, our finding for *migration history* and *age* stand out. It is striking

**Table 4**  
Summary of the identified relationships, classified according to their relative importance.

Type of urban characteristic	Factor	Explanatory variable	Relative importance and direction of correlation		
			Electricity	Natural gas	Water
Spatial	Population density	Number of people per km <sup>2</sup>	Moderate negative	Low negative	
	Amount of tree canopy cover	Number of trees per 100m <sup>2</sup>			
	Garden type	Average size of the garden per household (m <sup>2</sup> )	Low positive	Moderate positive	Low positive
	Building size	Share of houses with a living space area above 80m <sup>2</sup> (%)	High positive		Moderate positive
	Building type	Share of multiple family homes (%)	High negative	Moderate negative	High negative
	Building age	Share of houses with built after 2000 (%) <sup>1</sup>		High negative	Low negative
	Property value	Collinear with income level			
	Home ownership	Share of houses that is owner-occupied (%)		Low negative	
	Rental housing type	Share of rental houses owned by housing associations (%)	Moderate negative		Low negative
	Presence of supply network	Share of houses connected to district heating (%)	Low positive	High negative	
Infrastructural	Presence of a pool	Not assessed due to lack of data			
	Home appliances	Not assessed due to lack of data			
	Sanitary fittings	Not assessed due to lack of data			
	Income level	Average property value (€)	High positive	High positive	Low positive
	Lifestyle	Average number of passenger cars per household			
	Education level	Collinear with migration history			
Consumer	Age	Share of people between 25 and 44 years old (%) <sup>2</sup>	Moderate negative	Low negative	Moderate negative
	Migration history	The share of people with a non-western migration history (%) <sup>3</sup>	Low positive		High positive
	Household size	Average number of people in a household	High positive	Low positive	High positive

Note: the classification in this table follows from the number of models for which the factor was significant, the relative importance scores and model fit (adjusted  $R^2$ ) for each level (Tables 2A–2C).

<sup>1</sup> For the 100 and 500m-grids this is the year 2005.

<sup>2</sup> For natural gas this is the share of people between 25 and 64 years old.

<sup>3</sup> Western migration history is defined by the Dutch Bureau of Statistics as first and second generation migration from Europe (excluding Turkey), North-America, Oceania, Indonesia or Japan.

that that *migration history* is found to be an explanatory variable of high importance for water consumption, although this factor is not often considered in residential water demand studies thus far and no clear connection between this factor and water consumption has been described (Bich-Ngoc & Teller, 2018; House-Peters & Chang, 2011). Given the collinearity between *migration history* and *education level* (Table 3), the latter consumer characteristic could also be explanatory. However, also for *education level* there is no agreement in literature on its potential effect. Explanations for the influence of these two consumer characteristics may be found in differences in environmental awareness, behavioural aspects, practices and attitudes towards water use (Bich-Ngoc & Teller, 2018). It is also notable that we found a moderately negative correlation between *the share of people between 25 and 44 years old* and electricity and water consumption. Previous studies do not agree on the effect of consumer age on water consumption and little is described in the context of electricity consumption (Bich-Ngoc & Teller, 2018; Wiedenhöfer et al., 2013). We assume our results reflect that less time spent at home results in lower consumption, considering that the abovementioned age group is of working age.

Also our findings for *building age*, *home ownership* and *lifestyle* are not in line with what is described in literature. Instead, these results seem to indicate that different cities have different relationships between factors of influence and consumption, as found previously (Bettignies et al., 2019). For example, our findings do not confirm a relationship between *building age* and electricity consumption, as Pincetl et al. (2014) found for Los Angeles. Whereas in L.A. newer buildings have more electricity consuming appliances installed like air conditioning, air conditioning is uncommon in the Netherlands. Likewise, we did not confirm the positive correlation between *house ownership* and electricity consumption and gas consumption found for L.A. (Porse et al., 2016). Yet, it should be noted that in 2016 almost 70 % of all houses in Amsterdam comprised rentals, and over 60 % of those rental houses was owned by housing associations (CBS, 2016a). Indeed, our findings do show that in the case of Amsterdam it is relevant to distinguish *rental housing type*, with significantly lower electricity and water consumption levels in areas with a higher share of rental houses owned by housing associations. Contrary to earlier findings, *lifestyle* was also not found to be a variable of significance. However, *lifestyle* was assessed with the variable '*number of passenger cars per household*'. Earlier studies that mention car ownership as indicator for *lifestyle* are from North America and Australia (Hall, 2011; Wiedenhöfer et al., 2013). In these contexts, cities have a distinctively different, generally more car-dependent urban form than Europe and car ownership may be a useful indicator to distinguish between automobile, transit and walking urban fabric or urban and sub-urban lifestyle (Thomson & Newman, 2018; Weisz & Steinberger, 2010; Wiedenhöfer et al., 2013). Our findings thus reflect that car ownership is not an appropriate indicator for Amsterdam, which has a dense cycling and public transit network. Further research should show to what extent our results are specific to Amsterdam, and to what extent they are applicable to other Northern European or Dutch cities.

### 5.3. High resolution data for accurate explanation of variation

In this research we used multiple linear regression to assess the significance of the explanatory factors for residential electricity, natural gas and water consumption. The assumptions for linear regression were met for most resources and spatial aggregation levels. However, the assumption that the random disturbance is the same across all values of the independent variable, i.e. homoscedasticity, was not fully met for the natural gas grid models. This is probably because data on the *presence of district heating* was not available for the grids, whilst this was found the most important factor for explaining spatial variation in natural gas consumption with the neighbourhood and district models. Consequently, the results show that these models have lower adjusted  $R^2$  values and the relative importance of explanatory variables becomes skewed towards one (*building age*) or few variables (*building age*, *building type*, *income level*). A more nuanced picture of the relative importance of

variables is seen in the neighbourhood and district models.

When comparing the adjusted  $R^2$  values of all models, we find very high values for the energy models using administrative units and overall lowest performance for the water models. The high  $R^2$  values for the energy district and neighbourhood models could indicate that these models are somewhat overfitted, which means that the models partly explain noise in the data rather than the actual variance. Future research that applies the same method to another city could be a valuable means for cross-validation these findings. The performance of the water models may be affected by the water consumption data used, which has a greater degree of uncertainty than the energy data because it partially relies on estimates (see 3.2). Another explanation for the lower fit of the water models could be that the aggregation of data in yearly consumption levels for the spatial units used, is less suitable for water than for energy consumption. In the case of water, the resolutions used may still be too coarse and thus mask variations of key variables, considering that previous research found that indoor and outdoor use have different explanatory variables (Chang et al., 2017; Ghavidelfar et al., 2017; House-Peters, Pratt, & Chang, 2010; Rasifaghihi et al., 2020). A higher temporal and spatial resolution may thus be required for better performance of the water models. Higher resolutions can reveal seasonal variation, outdoor water use, and more accurately show spatial variation in explanatory variables – e.g. indicating garden presence per block instead of average garden size at neighbourhood level. The performance of our models thus underlines the importance of using accurate, high resolution data for these types of regression studies and indicate that the optimal resolution of study may differ with the resource of interest.

The importance of using accurate, high resolution data is also seen in the findings for *building size*, *garden type*, and *amount of tree canopy cover*. The data used for these variables may not represent the factor accurately enough. Contrary to our initial hypothesis, *building size* was not found to be a significant explanatory variable for natural gas consumption. However, this finding could also show that there was insufficient differentiation in the data used – houses were only discriminated into those having a living space smaller than 80 m<sup>2</sup> or larger. The factor *garden type* was assessed using '*average m<sup>2</sup> of garden per household*', which does not provide insight into garden characteristics affecting outdoor water use such as impervious surface cover or presence of irrigation-dependent vegetation (Chang et al., 2017). Similarly, the dataset for trees does not provide any information about the size of the tree canopy but merely on the number of trees under management by the municipality. To better assess the factors *garden type* and *amount of tree canopy cover*, remote sensing images may prove a useful data source. Chen et al. (2020a, 2020b), Chang et al. (2017) and Pincetl et al. (2014) show that such images can be used to derive the normalized difference vegetation index (NDVI) and assess the correlation of amount of green space coverage with water and energy demand. The lack of accurate data for the factors *garden type* and *amount of tree canopy cover* is expected to have especially affected the performance of the 100 m-grid energy models, because vegetation is a key element to cause microclimatic variations and related variation in energy demands on this micro-level (Chen et al., 2020a). Future studies should thus ensure that variables properly represent the urban morphological characteristics that result in microclimatic differences.

### 5.4. Mechanisms underlying urban energy and water consumption

Our results show that energy and water consumption have a substantial overlap in terms of key explanatory factors for spatial variation at intra-city level. All but one of the variables that were found to have explanatory value are of significance for two or all three resources. Factors shared by three resources and with relative high importance are *income level*, *household size*, and *building type*, with somewhat lower importance *age* and *garden type* (Table 4). *Building size*, *building age*, *presence of supply network* and *migration history* stand out because they are shared by two resources and are of relatively high importance for at



least one of these. Our collinearity evaluation shows that cross-correlations between these variables are present (Table 3). For example, *income level* is correlated with *building size* ( $r = 0.47$ ) and *migration history* ( $r = -0.65$ ) and *building type* is correlated with *household size* ( $r = -0.65$ ) and the share of people of working age ( $r = -0.56$ ). We hypothesize that these interconnections reflect that consumer and spatial characteristics are generally present in specific combinations and that these “lifestyle clusters” each have characteristic consumption patterns. So, although the factor *lifestyle* was not found significant for any of the three resources, our results actually reflect that lifestyle is an overarching factor, captured by both consumer and spatial characteristics.

The notion of such clusters is also described by others, who identified similar connections between spatial and consumer characteristics as the ones that we found in Amsterdam. Wiedenhofer et al. (2013) and Baiocchi et al. (2015), for example, both stress that energy consumption patterns can be linked to the clusters of urban characteristics present, that are defined by the factors of *population density*, *building type*, *income level*, *household size*, and *age* (e.g. “affluent urban dwellers”). The presence of clusters and the interconnections between factors could also clarify some of the unexpected relationships found, such as the positive correlation between *garden type* and energy consumption. These findings could reflect that a larger average *garden size* is indicative for a lifestyle with a generally high energy consumption, i.e. of high-income households (*income level*), living in larger homes (*building size*) and individual dwellings (*building type*), with a larger average *garden size*. Models such as regression tree analysis or machine learning algorithms like Random Forest could help identify such clusters based on consumption patterns and underlying, interconnected consumer and spatial characteristics. Regardless of the specific clusters, the explanatory variables identified in this study can already provide guidance for decision makers, urban planners and urban designers when formulating urban development plans and circularity strategies. We show the importance of taking future resource demands into consideration when designing development plans, as urban form and the lifestyle of envisioned future residents underlies water and energy consumption patterns. Moreover, this study shows that analyses using detailed spatial units by sampling aggregated values can be useful to inform spatially explicit approaches, policies as well as strategies aimed at certain groups of citizens.

## 6. Conclusions

Our multiple linear regression analyses have revealed the factors underlying residential electricity, natural gas and water consumption in Amsterdam. We found twelve spatial, infrastructural and consumer characteristics that are significant explanatory variables for spatial variability in residential energy and water consumption. Explanatory variables with high importance for electricity, natural gas and water consumption are *income level*, *household size*, and *building type*. Important correlations not described hitherto are those between *migration history* and water consumption, and between *the share of people between 25 and 44 years old* and electricity and water consumption. We conclude that the explanatory variables for resource consumption are sensitive to the spatial resolution of observation. Especially the three districts models reveal the influence of the spatial unit used, as they have some distinct findings in the relative importance of explanatory variables compared to the neighbourhood and grid models. Our neighbourhood and district models perform considerably better than the grid models, partially because the shape of these units correspond to urban form and partially because some factors could not be assessed with the grid models. We thus also conclude that for regression analyses to provide comprehensive and accurate insight into explanatory variables for consumption it is crucial to use accurate and high resolution data of a long-list of potential factors of influence that is aligned with urban divisions. When data is missing, model performance is clearly affected and less variance in consumption levels is explained. Finally, explanatory variables for

consumption are found to be cross-correlated. Insight in these interconnections is needed to understand the complex mechanisms underlying spatial variability in residential water and energy consumption. We recommend identifying in what specific combinations consumer, infrastructural and spatial characteristics are present and whether consumption patterns can be linked to the specific cluster present. Also, more multi-scale, high-resolution case studies are needed to identify optimal shape and size of spatial aggregation units for these sorts of regression analysis and to shed light on site and scale specific mechanisms underlying spatial variability in consumption. Such insights are valuable for urban planners and designers when formulating context-specific interventions for consumption reduction.

## CRedit authorship contribution statement

**Ilse M. Voskamp:** Project administration, Conceptualization (lead), Methodology, Writing - original draft (lead), Writing - review & editing, Supervision, Validation, Visualization, Funding acquisition (supporting). **Megan N. Visscher:** Conceptualization, Methodology, Writing - original draft, Software (lead), Investigation, Formal analysis, Visualization, Data curation. **Corné Vreugdenhil:** Conceptualization, Methodology (lead), Writing - review & editing, Software, Supervision, Validation, Visualization, Data curation. **Ron J.A. van Lammeren:** Writing - review & editing, Validation, Resources. **Nora B. Sutton:** Writing - review & editing, Supervision (supporting), Validation (supporting), Visualization (supporting), Resources, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research is part of the Urban Pulse II research project run under the auspice of and partially funded by the Amsterdam Institute for Advanced Metropolitan Solutions (AMS), the Netherlands. We gratefully acknowledge Waternet for providing water consumption data that were essential to perform this research.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.scs.2021.102977>.

## References

- Agudelo-Vera, C. M., Mels, A. R., Keesman, K. J., & Rijnaarts, H. H. M. (2011). Resource management as a key factor for sustainable urban planning. *Journal of Environmental Management*, 92(10), 2295–2303. <https://doi.org/10.1016/j.jenvman.2011.05.016>
- Baiocchi, G., Creutzig, F., Minx, J., If, T. D., & Pichler, P. (2015). A spatial typology of human settlements and their CO<sub>2</sub> emissions in England. *Global Environmental Change*, 34, 13–21. <https://doi.org/10.1016/j.gloenvcha.2015.06.001>
- Beloin-Saint-Pierre, D., Rugani, B., Lasvaux, S., Mailhac, A., Popovici, E., Sibiude, G., ... Schiopu, N. (2017). A review of urban metabolism studies to identify key methodological choices for future harmonization and implementation. *Journal of Cleaner Production*, 163, S223–S240. <https://doi.org/10.1016/j.jclepro.2016.09.014>
- Bettignies, Y., Meirelles, J., Fernandez, G., Meinherz, F., Hoekman, P., Bouillard, P., & Athanassiadis, A. (2019). The scale-dependent behaviour of cities: A cross-cities multiscale driver analysis of urban energy use. *Sustainability (Switzerland)*, 11(12). <https://doi.org/10.3390/su10023246>
- Bich-Ngoc, N., & Teller, J. (2018). A review of residential water consumption determinants. In O. Gervasi, B. Murgante, S. Misra, E. Stankova, C. M. Torre, A. Maria, ... D. Hutchison (Eds.), *Computational science and its applications – ICCSA 2018* (pp. 685–696). Springer International Publishing. <https://doi.org/10.1007/978-3-319-95174-4>
- Broto, V. C., Allen, A., & Rapoport, E. (2012). Interdisciplinary perspectives on urban metabolism. *Journal of Industrial Ecology*, 16(6), 851–861. <https://doi.org/10.1111/j.1530-9290.2012.00556.x>
- CBS. (2014). *Kerncijfers wijken en buurten*.

- CBS. (2016a). *Kerncijfers wijken en buurten*. Retrieved from <https://opendata.cbs.nl/s/tatline/#/CBS/nl/dataset/83487NED/table?fromstatweb>.
- CBS. (2016b). *Statistische gegevens per vierkant en postcode*.
- Chang, H., Bonnette, M. R., Stoker, P., Crow-Miller, B., & Wentz, E. (2017). Determinants of single family residential water use across scales in four western US cities. *Science of the Total Environment*, 596–597, 451–464. <https://doi.org/10.1016/j.scitotenv.2017.03.164>
- Chen, H., Han, Q., & Vries, B. D.e. (2020a). Modeling the spatial relation between urban morphology, land surface temperature and urban energy demand. *Sustainable Cities and Society*, 60(January), Article 102246. <https://doi.org/10.1016/j.scs.2020.102246>
- Chen, H., Han, Q., & Vries, B. D.e. (2020b). Urban morphology indicator analyzes for urban energy modeling. *Sustainable Cities and Society*, 52(May 2019), Article 101863. <https://doi.org/10.1016/j.scs.2019.101863>
- Ciancio, V., Salata, F., Falasca, S., Curci, G., & Golasi, I. (2020). Energy demands of buildings in the framework of climate change : An investigation across Europe. *Sustainable Cities and Society*, 60(102213). <https://doi.org/10.1016/j.scs.2020.102213>
- Cui, X. (2018). How can cities support sustainability: A bibliometric analysis of urban metabolism. *Ecological Indicators*, 93, 704–717. <https://doi.org/10.1016/j.ecolind.2018.05.056>
- Dijst, M., Worrell, E., Böcker, L., Brunner, P., Davoudi, S., Geertman, S., ... Zeyringer, M. (2018). Exploring urban metabolism—Towards an interdisciplinary perspective. *Resources, Conservation and Recycling*, (May (1))<https://doi.org/10.1016/j.resconrec.2017.09.014>. Elsevier B.V.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 027–046. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Fox, C., McIntosh, B. S., & Jeffrey, P. (2009). Classifying households for water demand forecasting using physical property characteristics. *Land Use Policy*, 26(3), 558–568. <https://doi.org/10.1016/j.landusepol.2008.08.004>
- Ghavidelfar, S., Shamseldin, A. Y., & Melville, B. W. (2017). A multi-scale analysis of single-unit housing water demand through integration of water consumption, land use and demographic data. *Water Resources Management*, 31(7), 2173–2186. <https://doi.org/10.1007/s11269-017-1635-4>
- Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, 17(1), 1–27. <https://doi.org/10.18637/jss.v017.i01>
- Groemping, U., & Lehrkamp, M. (2018). *Package “relaimpo”: Relative importance of regressors in linear models*. Retrieved from (pp. 1–36) <http://arxiv.org/abs/1007.5516>.
- Hall, M. H. P. (2011). A preliminary assessment of socio-ecological metabolism for three neighborhoods within a rust belt urban ecosystem. *Ecological Modelling*, 223(1), 20–31. <https://doi.org/10.1016/j.ecolmodel.2011.08.018>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. <https://doi.org/10.1007/b94608>
- House-Peters, L.A., & Chang, H. (2011). Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, 47(5). <https://doi.org/10.1029/2010WR009624>
- House-Peters, L., Pratt, B., & Chang, H. (2010). Effects of Urban spatial structure, sociodemographics, and climate on residential water consumption in hillsboro, oregon1. *Journal of the American Water Resources Association*, 46(3), 461–472. <https://doi.org/10.1111/j.1752-1688.2009.00415.x>
- Jayarathna, L., Rajapaksa, D., Managi, S., Gifford, R., Athukorala, W., Torgler, B., ... Wilson, C. (2017). A GIS based spatial decision support system for analysing residential water demand : A case study in Australia. *Sustainable Cities and Society*, 32, 67–77. <https://doi.org/10.1016/j.scs.2017.03.012>
- Kennedy, C. A., Stewart, I., Facchini, A., Cerosimo, I., Mele, R., Chen, B., ... Sahin, A. D. (2015). Energy and material flows of megacities. *Proceedings of the National Academy of Sciences of the United States of America*, 112(19), 5985–5990. <https://doi.org/10.1073/pnas.1504315112>
- Kennedy, C., Pincetl, S., & Bunje, P. (2011). The study of urban metabolism and its applications to urban planning and design. *Environmental Pollution*, (August)<https://doi.org/10.1016/j.envpol.2010.10.022>
- Lenzen, M., Dey, C., & Foran, B. (2004). Energy requirements of Sydney households. *Ecological Economics*, 49(3), 375–399. <https://doi.org/10.1016/j.ecolecon.2004.01.019>
- Li, H., & Kwan, M.-P. (2018). Advancing analytical methods for urban metabolism studies. *Resources, Conservation and Recycling*, 132(July 2017), 239–245. <https://doi.org/10.1016/j.resconrec.2017.07.005>
- McPhearson, T., Haase, D., Kabisch, N., & Gren, Å. (2016). Advancing understanding of the complex nature of urban systems. *Ecological Indicators*, 70, 566–573. <https://doi.org/10.1016/j.ecolind.2016.03.054>
- Meng, Q., Xiong, C., Mourshed, M., Wu, M., Ren, X., Wang, W., ... Song, H. (2020). Change-point multivariable quantile regression to explore effect of weather variables on building energy consumption and estimate base temperature range. *Sustainable Cities and Society*, 53(101900). <https://doi.org/10.1016/j.scs.2019.101900>
- Nelson, J. K., & Brewer, C. A. (2017). Evaluating data stability in aggregation structures across spatial scales: Revisiting the modifiable areal unit problem. *Cartography and Geographic Information Science*, 44(1), 35–50. <https://doi.org/10.1080/15230406.2015.1093431>
- Openshaw, S. (1984). *The modifiable areal unit problem: Concepts and techniques in modern geography* (38th ed.). Norwich, UK: GeoBooks.
- Pavelescu, F. (2014). An extensive study on the disturbances generated by collinearity in a linear regression model with three explanatory variables. *Romanian Journal of Economics*, 38(1(47)), 89–106.
- Pickett, S. T. A., Cadenasso, M. L., Rosi-Marshall, E. J., Belt, K. T., Groffman, P. M., Grove, J. M., ... Warren, P. S. (2017). Dynamic heterogeneity: A framework to promote ecological integration and hypothesis generation in urban systems. *Urban Ecosystems*, 20(1). <https://doi.org/10.1007/s11252-016-0574-9>
- Pincetl, S., Bunje, P., & Holmes, T. (2012). An expanded urban metabolism method: Toward a systems approach for assessing urban energy processes and causes. *Landscape and Urban Planning*. <https://doi.org/10.1016/j.landurbplan.2012.06.006>
- Pincetl, S., Chester, M., Circella, G., Fraser, A., Mini, C., Murphy, S., ... Sivaraman, D. (2014). Enabling future sustainability transitions: An urban metabolism approach to Los Angeles. *Journal of Industrial Ecology*, 18(6), 871–882. <https://doi.org/10.1111/jiec.12144>
- Pistoni, R., & Bonin, S. (2017). Urban metabolism planning and designing approaches between quantitative analysis and urban landscape. *City, Territory and Architecture*, 4(1). <https://doi.org/10.1186/s40410-017-0076-y>
- Poças Ribeiro, A., Harmsen, R., Rosales Carreón, J., & Worrell, E. (2019). What influences consumption? Consumers and beyond: Purposes, contexts, agents and history. *Journal of Cleaner Production*, (February (1))<https://doi.org/10.1016/j.jclepro.2018.10.103>. Elsevier Ltd.
- Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 52(52), 145. <https://doi.org/10.2307/621706>
- Porse, E., Derenski, J., Gustafson, H., Elizabeth, Z., & Pincetl, S. (2016). Structural, geographic, and social factors in urban building energy use: Analysis of aggregated account-level consumption data in a megacity. *Energy Policy*, 96, 179–192. <https://doi.org/10.1016/j.enpol.2016.06.002>
- Pulido Barrera, P., Rosales Carreón, J., & de Boer, H. J. (2018). A multi-level framework for metabolism in urban energy systems from an ecological perspective. *Resources, Conservation and Recycling*, 132, 230–238. <https://doi.org/10.1016/j.resconrec.2017.05.005>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Retrieved from Vienna: R Foundation for Statistical Computing <https://www.r-project.org/>.
- Rasifaghini, N., Li, S. S., & Haghighat, F. (2020). Forecast of urban water consumption under the impact of climate change. *Sustainable Cities and Society*, 52(101848).
- Rosales Carreón, J., & Worrell, E. (2018). Urban energy systems within the transition to sustainable development. A research agenda for urban metabolism. *Resources, Conservation and Recycling*, 132(August 2017), 258–266. <https://doi.org/10.1016/j.resconrec.2017.08.004>
- Stoker, P., & Rothfeder, R. (2014). Drivers of urban water use. *Sustainable Cities and Society*, 12, 1–8. <https://doi.org/10.1016/j.scs.2014.03.002>
- Thomson, G., & Newman, P. (2018). Urban fabrics and urban metabolism – From sustainable to regenerative cities. *Resources, Conservation, and Recycling*, 132, 218–229. <https://doi.org/10.1016/j.resconrec.2017.01.010>
- UN. (2019). *World urbanization prospects: The 2018 revision*. New York.
- Villarin, M. C. (2019). Methodology based on fine spatial scale and preliminary clustering to improve multivariate linear regression analysis of domestic water consumption. *Applied Geography*, 103(March 2018), 22–39. <https://doi.org/10.1016/j.apgeog.2018.12.005>
- Voskamp, I. M., Spiller, M., Stremke, S., Bregt, A. K., Vreugdenhil, L. C., & Rijnaarts, H. H. M. (2018). Space-time information analysis for resource-conscious urban planning and design: A stakeholder based identification of urban metabolism data gaps. *Resources, Conservation and Recycling*, 128, 516–525. <https://doi.org/10.1016/j.resconrec.2016.08.026>
- Voskamp, I. M., Sutton, N. B., Stremke, S., & Rijnaarts, H. H. M. (2020). A systematic review of factors influencing spatiotemporal variability in urban water and energy consumption. *Journal of Cleaner Production*, 256, Article 120310. <https://doi.org/10.1016/j.jclepro.2020.120310>
- Weisz, H., & Steinberger, J. K. (2010). Reducing energy and material flows in cities. *Current Opinion in Environmental Sustainability*, 2(3), 185–192. <https://doi.org/10.1016/j.cosust.2010.05.010>
- Wiedenhofer, D., Lenzen, M., & Steinberger, J. K. (2013). Energy requirements of consumption: Urban form, climatic and socio-economic factors, rebounds and their policy implications. *Energy Policy*, 63, 696–707. <https://doi.org/10.1016/j.enpol.2013.07.035>
- Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research and Evaluation*, 18(9), 1–14.
- You, Y., & Kim, S. (2018). Revealing the mechanism of urban morphology affecting residential energy efficiency in Seoul, Korea. *Sustainable Cities and Society*, 43(June), 176–190. <https://doi.org/10.1016/j.scs.2018.08.019>
- Zhang, Y., Yang, Z., & Yu, X. (2015). Urban metabolism: A review of current knowledge and directions for future study. *Environmental Science & Technology*, 49(19), 11247–11263. <https://doi.org/10.1021/acs.est.5b03060>