# Quantifying households' carbon footprint in cities using socioeconomic attributes: A case study for The Hague (Netherlands)

Patel, R.G.; Marvuglia, Antonino; Baustert, Paul; Huang, Yilin; Shivakumar, Abhishek; Nikolic, I.; Verma, T.

**DOI**
10.1016/j.scs.2022.104087

**Publication date**
2022

**Document Version**
Final published version

**Published in**
Sustainable Cities and Society

Sustainable Cities and Society 86 (2022) 104087

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

# Quantifying households' carbon footprint in cities using socioeconomic attributes: A case study for The Hague (Netherlands)

Ruchik Patel [a,b], Antonino Marvuglia [b,*], Paul Baustert [b], Yilin Huang [a], Abhishek Shivakumar [c], Igor Nikolic [a], Trivik Verma [a]

[a] Faculty of Technology, Policy and Management, Delft University of Technology, Delft 2628BX, The Netherlands
[b] Luxembourg Institute of Science and Technology (LIST), 5, avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg
[c] Department of Economic and Social Affairs, United Nations, NY, USA

ARTICLE INFO

*Keywords:*
Consumption-driven emissions
Process-based LCA
Cities
Urban policies
Random forest
Demographic clustering

ABSTRACT

Cities consume almost 80 percent of world's energy and account for 60 percent of all the emissions of carbon dioxide and significant amounts of other greenhouse gases (GHG). The ongoing rapid urbanization will further increase GHG emissions of cities. The quantification of the environmental impact generated in cities is an important step to curb the impact. In fact, quantifying the consumption activities taking place inside a city, if differentiated by socioeconomic and demographic groups, can provide important insights for sustainable-consumption policies. However, the lack of high-resolution data related to these activities makes it difficult to quantify urban GHG emissions (as well as other impacts). This paper presents a methodology that can quantify the carbon footprint of households in cities using consumption data from a national or European level, where the resource consumption is linked to socioeconomic attributes of a population. The methodology is applied to analyzing the environmental impact by household resource consumption in the city of The Hague in the Netherlands. The key insights reveal potential intervention areas regarding resource consumption categories and demographic groups that can be targeted to reduce GHG emissions due to consumption-driven activities in the city.

## 1. Introduction

Cities account for almost 80% of world's total energy consumption and 60% of global carbon dioxide emissions (Lanau, Herbert, & Liu, 2021). The demand for resources in cities has caused an ever increasing impact on the environment (Feng, 2015). Around two thirds of the global anthropogenic greenhouse gas (GHG) emissions can be traced back to household consumption patterns (Feng, 2015; Ivanova et al., 2020). This is the reason why the discourse about sustainable lifestyles currently attracts significant attention (Vita et al., 2019). Human consumption activities are largely driven by individual preferences, which can be associated with the households' socioeconomic (Vinholes et al., 2012) and geographic conditions, as well as local culture and context (Huang & Warnier, 2019; Porse, Derenski, Gustafson, Elizabeth, & Pincetl, 2016); the preferences are also shaped by a range of policies at the national and local levels (O'Rourke & Lollo, 2015). Addressing everyday consumption habits to generate virtuous changes requires governments to support households in transitioning towards more sustainable lifestyles (Sonigo, Bain, Kong, Fedrigo, et al., 2012).

Around 55% of global population lives in cities. This is expected to rise to 68% in 2050. Given the high population rate in urban areas, mere *information-provision* policies at the national scale are burdensome (Tsuda, Hara, & Uwasu, 2013) and ineffective at addressing variable lifestyle-induced consumption habits (Ölander & Thøgersen, 2014). Enacting sustainable-consumption policies at national scales falls short of recognizing the role of differentiating responsibility in addressing human consumption patterns among actors (e.g. citizens, institutions, corporations) and acknowledging the complex and integrated systems of behaviour, culture and governance that shape consumption patterns (O'Rourke & Lollo, 2015). Even from a utilitarian economic standpoint, the gains realized through nudging peoples and households towards more sustainable lifestyles will be better achieved through tailored and integrated policy initiatives (Carfagna et al., 2014) that are aware of individual contexts, and regional and local differences in business, culture and society (Muraca, 2012).

Such a complex challenge requires tools for supporting sustainable transitions that can evaluate context-specific environmental impacts of household consumption habits. Because of their close proximity to local

* Corresponding author.
*E-mail address:* antonino.marvuglia@list.lu (A. Marvuglia).

https://doi.org/10.1016/j.scs.2022.104087
Received 13 November 2021; Received in revised form 21 July 2022; Accepted 22 July 2022
Available online 27 July 2022

**Nomenclature**

| | |
|---|---|
| $BoP$ | Basket of Products |
| $CBS$ | Central Bureau of Statistics |
| $COICOP$ | Classification of Individual Consumption According to Purpose |
| $DESA$ | Department of Economic and Social Affairs |
| $DNFCS$ | Dutch National Food Consumption Survey |
| $DSE$ | Demographic and socioeconomic |
| $EC-JRC$ | European Commission - Joint Research Centre |
| $EF$ | Environmental footprint |
| $EU$ | European Union |
| $GHG$ | Greenhouse Gases |
| $GIS$ | Geographic information system |
| $GWP$ | Global warming potential |
| $IPCC$ | Intergovernmental panel on climate change |
| km | kilometers |
| $LCA$ | Life cycle assessment |
| $LCI$ | Life cycle inventory |
| $MAPE$ | Mean absolute percentage error |
| mm | millimeters |
| $NPISHS$ | Non-profit institutions serving households |
| $OLS$ | Ordinary least squares |
| $PB$ | Planetary boundaries |
| $RF$ | Random forest |
| $RIVM$ | Rijksinstituut voor Volksgezondheid en Milieu |
| $RP$ | Representative product |
| $UK$ | United Kingdom |
| $UN$ | United nations |

citizens, local city governments (compared to national governments) can more easily influence household resource consumption (Gouldson et al., 2016) by spurring more responsible environmental behaviours (Darmawan, Mangundjaya, & Herdiansyah, 2018; Hoornweg, Sugar, & Lorena Trejos Gómez, 2011) among all actors through swift actions (Rosenzweig, Solecki, Hammer, & Mehrotra, 2010). This is the main reason that this study takes households as units of analysis that can inform local policy-making.

The quantification of the environmental impacts generated by the consumption of the citizens of the entire European Union (EU), has been already addressed by the European Commission - Joint Research Centre (EC-JRC), which has developed the Consumption Footprint indicator (EC-JRC, 2022; Sala & Castellani, 2019). This indicator aims at assessing EU consumption with a process-based life cycle assessment (LCA) approach, looking at the entire supply chain of about 150 representative products (RPs). The RPs were selected for five areas of consumption (namely, housing, mobility, food, household appliances and household goods). For each RP, the consumption per capita was calculated based on consumption statistics that are available at European level and the environmental impact was assessed based on a life cycle inventory (LCI) database. However, when the spatial target goes down to the city (or even the district) level, the availability of data with this spatial resolution to quantify the consumption of households represents very often a problem.

One way to solve this problem is to use average consumption data on a national (or even European, in some cases) per-capita basis and multiply it by the local population (which is normally available, at least at the urban level), as done by Genta, Sanyé-Mengual, Sala, and Lombardi (2022) for the cases where city or regional data were lacking, such as for the quantification of extra-urban trips or the purchase of household goods. However, this approach is evidently a lumped one, and is not able to catch local differences. The approach for data disaggregation used in Genta et al. (2022) is based on proportional scaling applied to the data retrieved from the combination of different statistical surveys and available reports. The authors (Genta et al., 2022) applied a further step that consisted in the calculation of a single weighed score using the Environmental Footprint (EF) 3.0 method (EC-JRC, 2018; Fazio et al., 2018) and its normalization factors and further assessed the consumption of the city against the EF-based Planetary Boundaries (PB). Cárdenas-Mamani, Kahhat, and Vázquez-Rowe (2022) uses geographic information systems (GIS) data to estimate household energy use in the building sector at district level for the city of Lima in Peru.

From a practical point of view, the process of quantifying the environmental impact of household consumption behaviours in a city is complex and often relies almost entirely on proprietary data. To quantify the GHG emissions of household resource consumption, scholars and decision-makers use two basic approaches as follows Broekhoff, Erickson, and Piggot (2019).

First, translating the spending behaviour of households on goods and services into GHG emissions using an environmentally extended input–output database, such as Exiobase. For example, Hasegawa, Kagawa, and Tsukui (2015), created a multi-region input–output table of 47 prefectures in Japan to analyze the per capita carbon footprint at the prefecture levels in Japan. Similarly, Steen-Olsen, Wood, and Hertwich (2016) used the Norwegian consumer expenditure survey and combined it with a multi-region input–output table to calculate the carbon footprint of Norwegian households.

Second, quantifying through consumer expenditure surveys or *ad hoc* models the physical amount of goods or services consumed, and translating those into GHG emissions through a process-based LCI database, as done by Hillman and Anu (2010) for eight U.S. cities and Genta et al. (2022) for the city of Turin in Italy. In the study conducted by Dorr, François, Poulhès, and Wurtz (2022) this approach is applied with a combination of databases and models to realize a multi-sector life cycle environmental assessment of the city of Montreuil in France. The assessment focussed on the three largest carbon-emitting sectors in cities: food, residential buildings and daily mobility.

With the first approach, expenditure data is often aggregated at the national level through household surveys (Guan, Hubacek, Weber, Peters, & Reiner, 2008; Menyah & Wolde-Rufael, 2010; Peters et al., 2012). As a consequence, environmental policies cannot take into account the variability among household behaviours, unless micro-data for consumer spending is available for each household in the defined geographic region. For example, Froemelt and Wiedmann (2020) used the full dataset of Australian household expenditure survey comprising of 400 attributes of more than 100,000 Australian households to investigate Carbon footprints of households of Sydney and Melbourne. Similarly, Froemelt, Dürrenmatt, and Hellweg (2018) used the Swiss household budget survey which contains detailed data on the consumption behaviour of 9734 Swiss households to assess environmental impacts of Swiss households. Most city-level GHG emission inventories that use the second approach rely on empirical data related to consumption activities (Dhakal, 2009; Kennedy, Ramaswami, Carney, & Dhakal, 2011) that has been directly collected by local city governments. While the second approach allows for more specific evaluation of consumption impacts due to the reference flow corresponding to physical items instead of monetary spending, rarely do cities collect data on all the resources or goods consumed by their residents (Broekhoff et al., 2019).

Although estimating the environmental impacts of consumption behaviour in cities is imperative, the data related to *all* the activities taking place in a city that generate GHG emissions is often not available at the same spatial or temporal resolution. Most emissions data are curated at the national or city level via surveys (Lin, Yu, Bai, Feng, & Wang, 2013), census statistics (Broekhoff et al., 2019), or input–output tables (Corsatea et al., 2019). In addition, scaling-down projections from national levels to small spatial boundaries (e.g. a neighbourhood

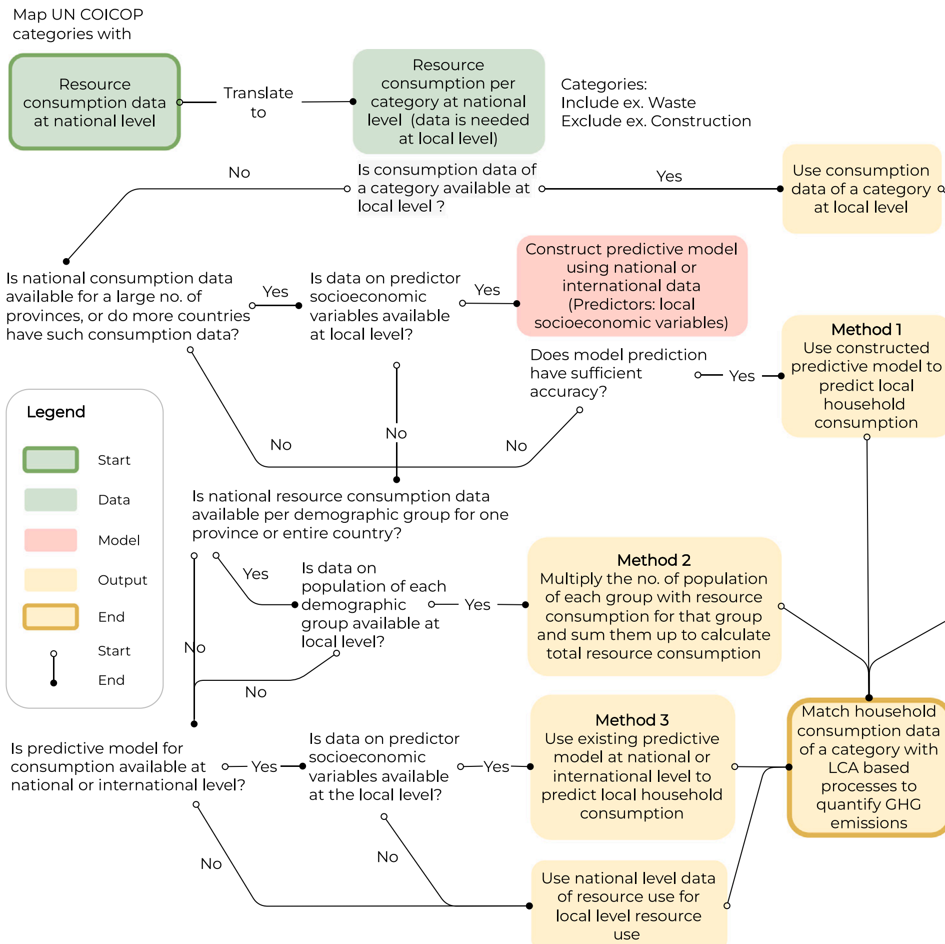


**Fig. 1.** A methodology to quantify household resource consumption and carbon footprint in cities.

or city) often evades the contextual use of resource consumption at smaller spatial scales (Balouktsi, 2020) hiding the contributions at the intersections of multiple sectors, institutional support structures and socioeconomic structure of the households. Thus, the lack of accurate data on resource use and population's consumption of goods and services at a city level poses a major challenge to thoroughly quantify the environmental impact of a city.

This article tackles this data challenge with the aim of quantifying the consumption-driven carbon footprint of a city and providing a structured decision tree to help modellers who approach the same issue. By combining data and models available both locally and at national and continental scales, a methodology is presented to assess the GHG emissions at the local city level. The analysis is put forth with an example of The Hague in the Netherlands, with a level of detail that goes down to the administrative neighbourhood level. The calculation of consumption-driven GHG emissions for households in a city, as applied in this paper, can be estimated in a three-step procedure. First, the major resource-consumption activities of households that result in GHG emissions are identified. The consumption categories can be applied to other cities in the world. Second, using regression models, national consumption data is disaggregated to calculate the per-capita (and per-household) consumption of all the resource consumption categories that have been identified within the administrative boundaries of a city. Third, by using a LCI database (Finnveden et al., 2009) and appropriate impact assessment methods, the GHG emissions of resources consumed are quantified, expressed in terms of their global warming potential (GWP).

The analysis illustrates how a city can identify most energy-intensive activities by the type of households, neighbourhoods and communities. The approach enables investigating the contributions of household-driven emissions across different sectors and by the type of households, so that local administrators can design bespoke policies to proportionately address environmental impacts of consumption behaviours in highly urbanized cities (e.g., by stimulating the use of public transport and discouraging car use). This methodology can be easily extended to other major Dutch cities, since the same models and data are available in the Netherlands and similar data is available for other western countries.

## 2. Methodology

In general, the methodology classifies household resource consumption in a city into categories and quantifies the resulting GHG emissions per category. Fig. 1 provides an overview of the decision pathways of which methods can be used to synthesize the household consumption data.

The consumption classification is based on the international reference COICOP (Classification of Individual Consumption According to Purpose) (United Nations, 2018) with adaptations (see Section 2.1). COICOP was first introduced in 1999 by the United Nations (UN) Statistical Commission and last revised in 2018 by the UN Department of Economic and Social Affairs (DESA). It provides 13 categories of goods and services used in household expenditure statistics (United Nations, 2018).

For each consumption category, relevant data sources need to be identified and transformed into the desired (local household consumption) data representation (hereafter referred to as data synthesis) using existing theories and models found in literature (see Section 2.2).

If the household consumption data is available for a consumption category, it is directly used to quantify the ensuing emissions. This data, however, is often unavailable. In such cases, the consumption data at the national level, if available, is used and mapped to the local household consumption. Some major consumption categories do have national data available in some countries collected by governmental organizations. In the Netherlands, e.g., the *Central Bureau of Statistics* (CBS) collects national data related to mobility, waste, electricity consumption, etc. If the national level consumption data is neither available, international data could be an option. But such data shall be used with caution since the local context can differ greatly.

This study proposes three methods, in order of preference, to synthesize household consumption data at the city level: (1) constructing predictive models, (2) ratio-based normalization, and (3) using existing theoretical models or empirical predictive models for data synthesis. They can be applied according to the availability of national consumption data and socioeconomic data of the populations. The details are explained in Section 2.2.

The synthesized household consumption data per category are then translated into GHG emissions. This study used *ecoinvent* (version 3.6), a well-known LCI database (Wernet et al., 2016). The impact assessment method used to estimate the corresponding GWP was the IPCC 2013 GWP 100a, which expresses the GHG emissions in kilograms of $CO_2$ equivalent over a time horizon of 100 years (Hirashi et al., 2014).

### 2.1. A classification of household consumption

Because the COICOP categorization is intended for the analysis of living standards, this study used *six consumption categories* adapted from COICOP to suit the research need for *resource consumption* and *environmental impact* that are directly linked to households. The adaptation also considered the household budget surveys that are normally conducted in all European countries to determine consumer price indices (Sala et al., 2019).

The six categories of household resource consumption are: (I.) Food, (II.) Water, (III.) Energy, (IV.) Mobility, (V.) Basket of Products (BoP),[1] and (VI.) Municipal Solid Waste. The rationale of the classification is as follows.

- The divisions in COICOP related to housing and services are excluded from the categories in the current study since it is assumed that no (or only marginal) *direct* resource consumption of these expenditures occurs on the *household* side, but also because data were not available in the Netherlands for these items. Examples are rentals for housing, maintenance and repairing of the houses, health care, education, restaurants and accommodation, information and communications services, insurance and financial services.
- The divisions of *Food and non-alcoholic beverages*, and *Alcoholic beverages, tobacco and narcotics* in COICOP are combined into category (I.) Food, as the Dutch National Institute for Public Health and the Environment (RIVM) conducts regular surveys that combines the consumption of food, alcoholic beverages and non-alcoholic beverages. The consumption of tobacco and narcotics is not included in the RIVM survey.[2]
- The division of *Water, electricity, gas and other fuels* in COICOP is separated into two categories of (II.) Water and (III.) Energy, because the GHG emissions due to water consumption are considerably lower than those of electricity and gas. Their separation eases comparison across categories. For category (III.) Energy, only electricity and gas consumption were included as they are the major household energy sources in the Netherlands.
- The division of *Transport* in COICOP is mapped to category (IV.) Mobility (which means people's mobility and not transport of goods). However, the purchase of vehicles was excluded since it is again assumed that no direct resource consumption thereof occurs on the household side. Thus, the Mobility category in the current analysis includes the use of personal cars and public transport. The latter is mainly comprised of trains, trams and buses in the Netherlands.
- The other divisions of products in COICOP are all categorized into (V.) BoP in the current analysis. This includes, among others, clothing and footwear, household equipment, routine household maintenance, and personal care products.
- Household waste is not in COICOP but in the current analysis it is included as category (VI.) Municipal Solid Waste, since waste management creates significant GHG emissions and households have direct influence on the amount of waste they generate.
- The categories of individual consumption expenditure of non-profit institutions serving households (NPISHS) and general government in COICOP were excluded in the current analysis because the GHG emissions resulting from these activities occur outside the households and are beyond the direct influence of the households.

There are many other collective (and business) activities that can be attributed to the demands from households (e.g., the use of postal services, education, insurances, energy consumed in stores and other public areas). These activities were excluded from the current analysis because the factors that influence the choices of consumption at a collective or business level are different from those at an individual household level. Consequently, different policy instruments are required to promote sustainable consumption (and production) at a collective (and business) level. This study focuses on the consumption at an individual household level upon which households have direct influence.

### 2.2. Data collection and synthesis

To quantify the environmental impact of resource consumption in each category, it is key to obtain the consumption data per category at a household level. Consumption data at such a resolution, however, is uncommon.

Relevant surveys and statistics, if available, are often reported at a provincial or national level (Broekhoff et al., 2019). In response to this issue, an approach was adopted to synthesize household consumption data per category. This is achieved in two ways: (1) *data refinement*, i.e., adding details to lower resolution consumption data; and (2) *data generation*, i.e., generating data according to theoretical or empirical models.

The two ways of consumption data synthesis are both based on the demographic and socioeconomic (DSE) data of citizens (i.e., consumers) at different geographical levels. Consumer DSE characteristics, such as age, gender, income and education level, are often reported to have direct linkages to consumption behaviours. Kamakura and Mazzon (2013) firstly reviews several techniques used in social sciences to identify socioeconomic classes and classify individuals in social strata. Then the authors apply a monotonically-constrained latent class model for socioeconomic stratification across 21 consumption categories, which is robust to missing data. The study shows that socioeconomic stratification explains differences in consumption priorities that cannot be explained only with the differences in household composition. The results from their simulations suggest that between the two possible ways how socioeconomic status may affect consumption (budget effect and shifts in consumption priorities), differences in consumption priorities across strata have the strongest effect on consumption. Lévay, Vanhille, Goedemé, and Verbist (2021) explores

[1] This includes the rest of consumption items such as clothes and household products.

[2] https://statline.rivm.nl/#/RIVM/nl/dataset/50038NED/table?ts=1635194556786.

the relationship between the carbon footprint of consumption and socioeconomic characteristics of Belgian households. Using ordinary least square (OLS) regression analysis the authors found out that income, household size, age, education and dwelling size are significantly and positively associated with household GHG emissions, while unemployment, living in an apartment (rather than in a house), and the tenant status (as opposed to being owner of the dwelling) are negatively correlated. Although the study does not explore causality relationships, but only correlations, a dominance analysis determined that income and household size resulted as the two most important determinants of household consumption-related emissions. Finally, using data from a representative house budget survey conducted in the UK, and applying multivariate analysis, Büchs and Schnepf (2013) studies the association of household socioeconomic descriptors like income, household size, education, gender, work status and territorial location (rural vs urban) with four emission areas: home energy, indirect emissions, transport and total $CO_2$ emissions. One of their main findings is that these associations vary considerably across emission domains.

The approach adopted in the present study assumes that consumption at a provincial, national or higher level, either reflected by empirical data or theoretical models, can be used as a basis to synthesize data at a lower (sub) regional level, adjusted by consumer DSE differences among these levels (Horta & Keirstead, 2017). This is the *first principle* in the current study to synthesize household consumption data (deemed as dependent variables) from DSE data (deemed as independent variables). The data synthesis is performed with households as social units. Some of the possible approaches for data refinement at a lower geographical level starting from data and information known at a higher geographical level and/or in similar spatial contexts, are well described in Horta and Keirstead (2017). In the following, the three main methods applied in this study are briefly recalled. In Method (1) and (2), sub-regional DSE data are used to refine regional consumption data. Method (3) uses quantitative models from previous studies to generate household consumption data.

*Method 1.* When the DSE data (on independent variables) and resource consumption data (on dependent variables) are available in many regions at a higher (and also same) geographical level, a *predictive model*, e.g., a regression model (with external calibration), can be constructed using data of those higher geographical levels. This method has been used by Stevens, Gaughan, Linard, and Tatem (2015) to disaggregate census level data in Kenya, Cambodia and Vietnam and predict population density in finer grids of 100 m by using random forest (RF) models with input variables such as lights at night, temperature, precipitation, distance to roads, etc. For example, if $X_u$ is a vector of independent variables at the upper level (for instance income, education and age) and $E_u$ is a vector of the corresponding energy consumption at the upper level, a regression model can be described by the equation:

$$E_u = \beta X_u + \varepsilon \tag{1}$$

where $\varepsilon$ is an error term assumed to be normally distributed, $\epsilon \sim N(0, \sigma^2)$ and $\beta$ is a vector of model coefficients determined by a least squares fit. The lower level energy consumption $\hat{E}_l$ can then be predicted by:

$$\hat{E}_l = \beta X_l \tag{2}$$

where $X_l$ are the same independent variables, but measured at the lower level.

Given the first principle used in the current study, the regression model can therefore be applied to synthesize data (i.e., to predict the dependent variables) at lower level regions, provided that the DSE data at lower level regions are available.

An important requirement for this method is that the data of DSE and the corresponding resource consumption are available for many areas, e.g., different municipalities in a province, various provinces in a country, or numerous countries in the EU. This means sufficient data points must exist at a higher geographical level to be able to meaningfully fit and test a regression model so that the model can be used to generate consumption data in different areas at a lower geographical level. When the fitted model shows a good level of accuracy for given error metrics, the model is used to predict the resource consumption in lower level regions. In case the regression model is not accurate enough or the data needed are not available, then Method (2) shall be explored.

*Method 2.* When Method 1 is not applicable, the *ratio-based normalization* method (Horta & Keirstead, 2017) may be applied to synthesize consumption data at a lower geographical level. This requires DSE data to be available both at higher and lower geographical levels, as well as the resource consumption data at a higher geographical level. In this case, the consumption data does not need to be available for many regions (as in Method 1). To recall the example given earlier, if $x_u$ and $x_l$ are the values of a DSE metric of interest respectively at the upper and lower levels (e.g., the population expressed in number of inhabitants), the lower level energy consumption $\hat{E}_l$ can be estimated by:

$$\hat{E}_l = \frac{x_l}{x_u} E_u \tag{3}$$

This method can be performed at a chosen unit level of a DSE metric. In Horta and Keirstead (2017), four unit levels were mentioned: per unit area, per-capita, per-household, and per median income. The overall resource consumption is then the sum of the consumption of each unit. Note that, compared to Method (1), this method is less robust since its accuracy decreases as the population of a unit level decreases.

*Method 3.* When Method 1 and Method 2 are not applicable (e.g., due to the lack of consumption data), theoretical and/or empirical models in literature can also be used to generate resource consumption data. The models, if they exist, need to be calibrated to fit the specific context of a study. This approach is used to generate (II.) Water and (III.) Energy consumption data (see Sections 3.2 and 3.3), since the two consumption categories are well studied and modelled with relation to DSE data. The models used were calibrated for the situation in the Netherlands (by the original studies). A less desirable approach is to use a model developed in a context that is different from the intended study, e.g., a consumption model for a country that is not comparable to the country of interest, since this introduces a higher level of uncertainty in the validity of results.

## 3. Case study of the Hague

The approach to synthesize sub-regional consumption data based on regional data is applied to the city of The Hague (*Den Haag* in Dutch), the third largest city in the Netherlands. With a population of over half a million, The Hague comprises 111 neighbourhoods (*buurts* in Dutch) (Fig. 2). The aim of the case study is to synthesize average household consumption data per each of the six categories (Section 2.1) in each neighbourhood, with which the GHG emissions can be estimated thereof. The DSE data at three levels was used: The Hague municipal level, the Dutch national level, and the EU level. At the municipal level, the *Den Haag Cijfers* database[3] provides DSE data with regard to variables such as gender ratio, age, income, education and employment rate per neighbourhood. At the Dutch national level, the CBS provides municipal solid waste data of 491 municipalities as well as mobility data of different regions of the Netherlands grouped by gender and age group of travellers.[4] Furthermore, the National Institute for Public Health and the Environment (*Rijksinstituut voor Volksgezondheid en Milieu* RIVM in Dutch) provides food consumption data in the Netherlands grouped by age, gender and educational level of citizens. At the EU level, the *Eurostat* database provides DSE data of

[3] https://denhaag.incijfers.nl.

[4] https://opendata.cbs.nl/statline/#/CBS/en/navigatieScherm/thema.

**Table 1**
Summary of data sources and methods for consumption data synthesis per category in The Hague neighbourhoods.

| Category | Data sources[a] and methods |
|---|---|
| Food | Census data in 12 Dutch provinces from the National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu, RIVM) (Method 2) |
| Water | Data is generated from a regression model with DSE data as predictors (Method 3) |
| Energy | Data from Den Haag Cijfers about average household consumption of electricity and gas per neighbourhood. The percentage of green energy was estimated (Method 3) |
| Mobility | Survey data about mobility patterns of the population in 12 Dutch provinces from the Dutch Central Bureau of Statistics (CBS) (Method 2) |
| Basket of Products (BoP) | Data from Eurostat about BoP used for all European Union countries over a period of 10 years. A RF model is built to synthesize the BoP data for each neighbourhood of The Hague (Method 1) |
| Municipal solid waste | Data from CBS about solid waste per-household in all 491 municipalities in the Netherlands. A RF model is built to synthesize data about municipal solid waste generated in each neighbourhood of The Hague (Method 1) |

[a]All demographic and socioeconomic (DSE) data of residents in The Hague were obtained from Den Haag Cijfers.


**Fig. 2.** The Hague and its 111 neighbourhoods.

28 European countries (still including UK).[5] These data sources as well as the data synthesis methods for the six categories of consumption are summarized in Table 1.

Fig. 3 provides an overview of the methodology linking the data, resource consumption categories and models to synthesize the household resource consumption data in each neighbourhood of The Hague. On the top of the figure it is indicated the geospatial resolution of the uppermost level for which data is available. In each case the country, province, municipality of neighbourhood of interest for the model was selected. The methods are explained in the rest of this section.

### 3.1. Food consumption

The Dutch National Food Consumption Survey (DNFCS) is conducted every four years to collect data about individual daily consumption of 133 food items in 18 food categories. The DNFCS data is classified by gender, age and education, with which the food consumption per item is quantified by the national average. Thus, the gender, age and education level of The Hague neighbourhoods were used (as indicators) to estimate the food consumption per-capita in each neighbourhood as shown in Fig. 4.

The method of demographic clustering (Rajagopal, 2011) is used for a direct mapping of food consumption. In demographic clustering, population in a smaller geographical unit is divided into clusters based on their socioeconomic attributes. The data on the variable that is to be predicted is available for the demographic clusters at a lower geospatial resolution or a higher aggregation level. The value of the variable at higher geospatial resolution is then the population weighted mean of the values of that variable for individual clusters.

**Table 2**
Indicators and Categories in DNFCS ($n_c = 2 \cdot 2 \cdot 3$)

| Indicators | Categories | $x_i$ |
|---|---|---|
| Gender | Male<br>Female | $x_1 = 2$ |
| Age | 0 to 18<br>18 to 79 | $x_2 = 2$ |
| Education | Primary<br>Secondary<br>Tertiary | $x_3 = 3$ |

In the current case, the number of clusters $n$ is defined by:

$$n = \prod_{i=1}^{3} x_i \tag{4}$$

where $x_{1,2,3}$ are the number of categories for gender, age and education (see Table 2).

As shown in Table 2, the DNFCS data is grouped by 2 genders, 2 age groups and 3 educational levels. Thus, a total of 12 possible combinations (i.e., 12 demographic clusters) can be created. For each cluster, it is assumed that the amount of consumption $c_j$ of food item $j \in (1, \dots, 133)$ is determined by a function $\digamma$:

$$c_j = \digamma(v_1, v_2, v_3) \tag{5}$$

Here $c_j$ is measured in terms of kg/(capita·day) or l/(capita·day) depending on whether $c_j$ is a food or beverage item. The function $\digamma$ was estimated by performing a user-defined grouping of the DNFCS data according to the values of the three demographic variables gender ($v_1$), age ($v_2$) and education ($v_3$). It is also assumed that $\digamma$ remains the same at the national and local level. In practice, if $P$ is the total population of the area at the lower geographical level (e.g. the neighbourhood), and $y_{a_i}$, $y_{b_j}$ and $y_{n_k}$ are respectively the percentages of population in this area that has a value $a_i$ of the variable $v_1$, $b_j$ of the variable $v_2$ and $n_k$ of the variable $v_3$, then the population in the particular cluster for which $v_1 = a_i$, $v_2 = b_j$ and $v_3 = n_k$ is given by:

$$p_{a_i b_j \dots nk} = P \cdot y_{a_i} \cdot y_{b_j} \cdot \dots \cdot y_{n_k} \tag{6}$$

Hence the individual food consumption $c_{j,m}$ of item $j$ in a neighbourhood $m \in (1, \dots, 111)$ in The Hague can be obtained by:

$$c_{j,m} = \frac{\sum_{n_c} p_{a_i b_j \dots nk} \cdot z_{a_i b_j \dots nk}}{P} \tag{7}$$

where $z_{a_i b_j \dots nk}$ is the variable value for that cluster on a lower geospatial resolution.

[5] https://ec.europa.eu/eurostat.

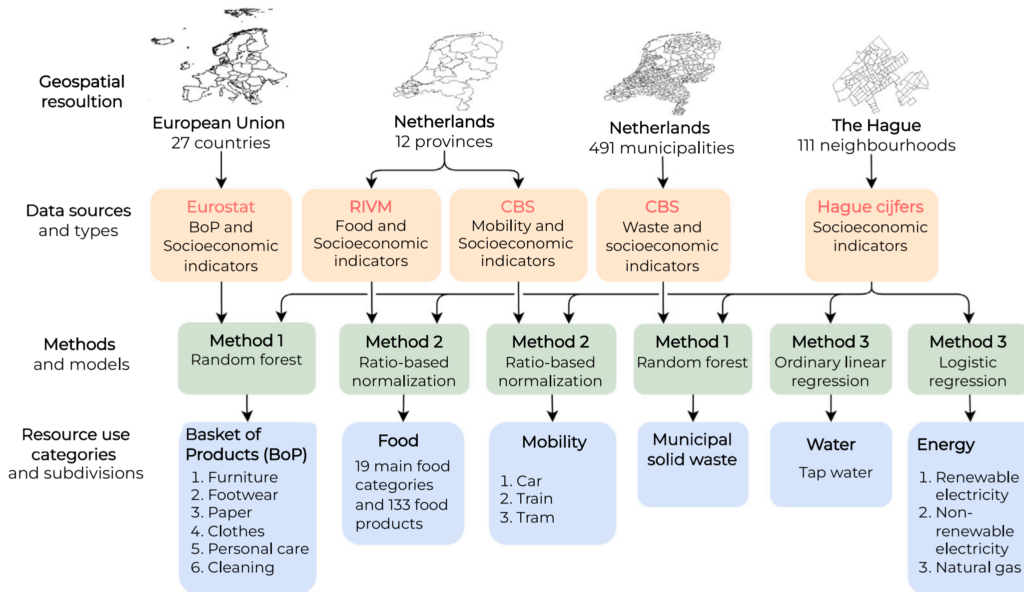


**Fig. 3.** Overview of household consumption data synthesis in the neighbourhoods of The Hague.

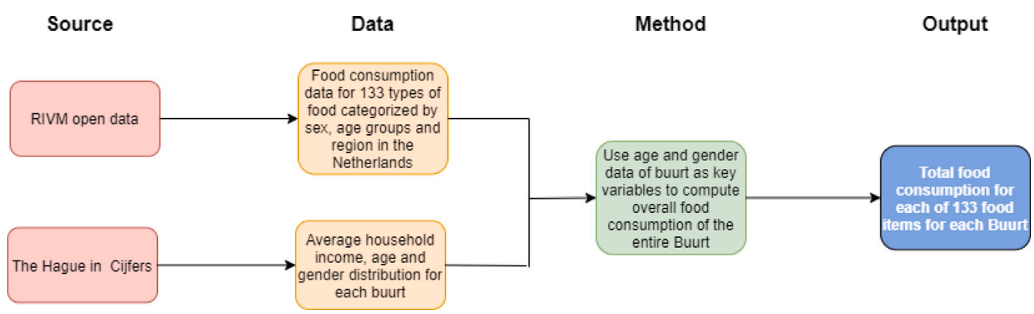


**Fig. 4.** Food consumption data synthesis for different neighbourhoods of The Hague.

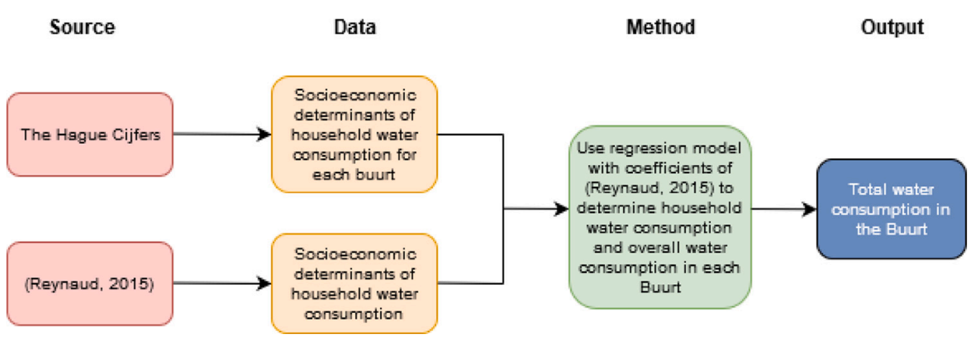


**Fig. 5.** Water consumption data synthesis for different neighbourhoods of The Hague.

### 3.2. Water consumption

The total water consumption in the Netherlands is available from CBS (Dutch Central Bureau of Statistics (CBS), 2018). Knowing only the total national consumption and number of inhabitants (population) both at national and neighbourhood (i.e., neighbourhood) levels, ratio-based normalization (Method 2) could be applied to synthesize the water consumption data at the neighbourhood level, the "metric of interest" being in this case the population. However, since a detailed study (Reynaud, 2015) of water consumption was conducted for Dutch households, based on that study, Method 3 was applied to synthesize water consumption data.

According to Reynaud (2015), the main determinants for water consumption of Dutch households are the cost of water, income, geographical area and level of rainfall. The data about the residential area of households and their income are available in Den Hague Cijfers data set, whereas the price of water is fixed by the water authority. Thus, the regression model in Reynaud (2015) (Eq. (8)) is applied to compute water consumption per-household per year for each neighbourhood of The Hague as shown in Fig. 5:

$$\ln(y) = \alpha \ln(p) + \beta \ln(I) + \gamma \ln(H) + \mu \ln(E) + C \tag{8}$$

where $y$ is the amount of water consumption measured in $m^3$/(capita·year); $p$ the unit water price (including water delivery and sewage treatment) measured in €/$m^3$; $I$ the representative household

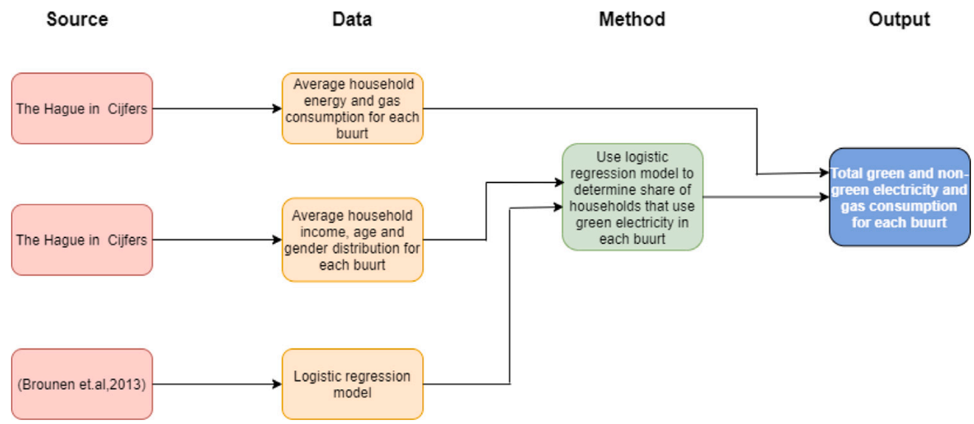


**Fig. 6.** Energy consumption data synthesis for different neighbourhoods of The Hague.

**Table 3**
Parameters of logistic regression model.

| Indicators | Coef. | Value |
|---|---|---|
| *Constant* | $\beta_0$ | 0.059 |
| Male respondent (yes = 1) | $\beta_1$ | −0.032 |
| Age of respondent 60–70 years | $\beta_2$ | −0.5794 |
| Age of respondent >70 years | $\beta_3$ | −0.1892 |
| Tertiary education (yes = 1) | $\beta_4$ | 0.1628 |
| Annual net income per-capita (€) | $\beta_5$ | $1.813 \times 10^{-6}$ |

income measured in €/year; $H$ the area measured in $m^2$; $E$ the evapotranspiration (which takes into account the average rainfall) measured in $\mathrm{mm}/year$; $C$ a constant term; $\alpha$, $\beta$, $\gamma$ and $\mu$ are the elasticity of each term respectively. The elasticity and constant terms were estimated in Reynaud (2015) as follows: $\alpha = -0.275$; $\beta = 0.201$; $\gamma = 0.013$; $\mu = -0.023$ and $C = 2.001$.

### 3.3. Energy consumption

The current household energy consumption in the Netherlands mainly consists of electricity and natural gas. The Hague municipality publishes data about average household consumption of electricity and gas in each neighbourhood.

The GHG emissions differ for renewable or non-renewable sources. However, the data about the source of electricity (renewable or non-renewable) is not available. To determine the percentage of households that opt for green electricity, Method 3 was applied using the model from Brounen, Kok, and Quigley (2013) which studied 1721 Dutch households' energy consumption. The study estimated the probability of "green choice" using a logistic model based on gender, age, education and income level, given by:

$$p = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}} \tag{9}$$

where $p$ is the probability of adopting green energy, $\beta_0$ is a constant, $x_i$ are predictor variables, and $\beta_i$ are the coefficients as shown in Table 3.

As shown in Fig. 6, The Hague cijfers database has the data on the electricity and gas consumed per household for each neighbourhood. Furthermore, data on income, age, gender and education level is also available in The Hague cijfers. Therefore, using these variables, the amount of green and non-green electricity consumed per neighbourhood can be computed.

### 3.4. Mobility

The environmental impacts of mobility are divided into three major categories: (1) private car use, (2) train use and (3) local public transport (bus and tram) use. The CBS collects mobility pattern data (i.e., the distance travelled per person per day per mode of transport) for each of the 12 provinces of the Netherlands. The data set is organized by DSE characteristics of gender and 8 age groups.

Similar to the data synthesis for food consumption, the clusters of people with difference gender and age groups were created for each neighbourhood.

The distance travelled $d_{a,b}$ measured in $\mathrm{km}/(\mathrm{capita{\cdot}day})$ in each neighbourhood $a$ in The Hague for a mode of transport $b$ is given by:

$$d_{a,b} = \sum_i \sum_j p_{a,i} p_{a,j} d_{ij,b} \tag{10}$$

where $i$ is gender; $j$ age group; $p_{a,i}$ and $p_{a,j}$ are respectively the percentage of population in neighbourhood with gender $i$ and age group $j$; $d_{ij,b}$ is the distance travelled per-capita in the province of South Holland (where The Hague is located) by a person of gender $i$ and age group $j$ using mode $b$.

Fig. 7 outlines the methodology to compute mobility data for each neighbourhood of The Hague. The Hague cijfers data set has the data on age and gender distribution in each neighbourhood of The Hague. Thus, the population belonging to each gender and age cluster can be calculated for each neighbourhood. Finally, the number of people that belong to a gender and age cluster in a neighbourhood can be multiplied by the distance travelled per person per mode in that cluster and then aggregated to obtain the overall distance travelled by private and public transport modes for each neighbourhood.

### 3.5. Basket of Products (BoP)

A BoP refers to a set of products that are representative of the consumption by an average European citizen. The European Commission - Joint Research Centre (EC- JRC) focussed on the BoP indicators project since 2010 (Sala & Castellani, 2019). The aim of the study was to comprehensively assess different areas of consumption following a life cycle-based bottom-up approach at the European scale, in order to build a baseline set of indicators to monitor consumption patterns over time and assess the effectiveness of policy scenarios targeting the reduction of the impacts stemming from consumption. Baldassarri, Allacker, Reale, Castellani, and Sala (2017) identified six categories of BoP products: clothing, footwear, house cleaning products, furniture, personal care and paper products, which are used in the current study. The studies described in Büchs and Schnepf (2013) and Lévay et al. (2021) show that BoP is influenced by gender, age and income level. Furthermore, it is also impacted by employment rate (Ganong & Noel, 2019).

Therefore, a RF model and an OLS model (Method 1) with these predictor variables as input and the amount of items per BoP products category as outputs was fitted and the results between RF and OLS were compared based on accuracy. As shown in Fig. 8, the data on

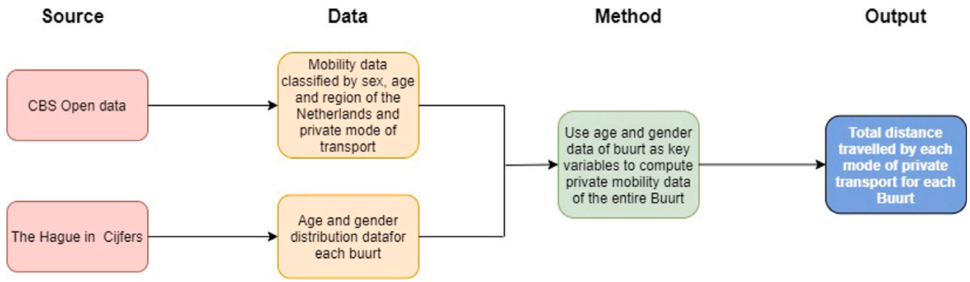


**Fig. 7.** Mobility data synthesis for different neighbourhoods of The Hague.

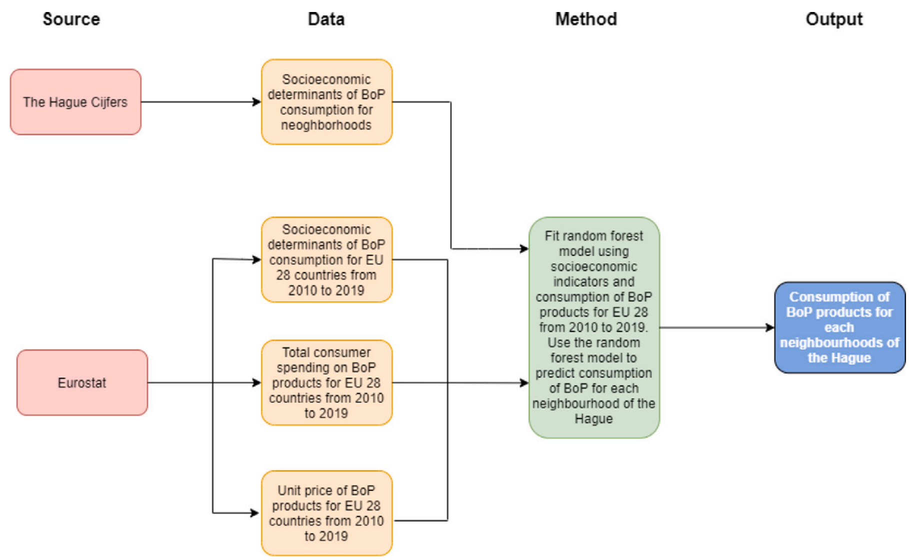


**Fig. 8.** BoP data synthesis for different neighbourhoods of The Hague.

**Table 4**
Validation results of Mean Absolute Percentage Error (MAPE) and R-squared 10-fold cross validation (R 10-f): Random Forest (RF) versus Ordinary Least Square (OLS) models for data synthesis of BoP.

| BoP Category | MAPE RF | MAPE OLS | R 10-f RF | R 10-f OLS |
|---|---|---|---|---|
| Clothing | 9.09 | 20.33 | 0.97 | 0.85 |
| Footwear | 9.51 | 22.19 | 0.91 | 0.54 |
| Cleaning | 13.36 | 50.36 | 0.92 | 0.40 |
| Furniture | 11.00 | 24.31 | 0.94 | 0.75 |
| Pers. care | 11.58 | 35.57 | 0.88 | 0.21 |
| Paper | 10.29 | 24.23 | 0.92 | 0.58 |

the predictor variables and the amount of items per BoP category was obtained from the Eurostats data set (EU, 2020) which contains consumer spending on BoP of 2010–2019 in 28 European countries. Once the RF and OLS models are fitted using the data from Eurostats, the data on input predictor variables obtained from The Hague cijfers dataset is used to predict BoP use for each neighbourhood of The Hague. Two methods were used to validate the models: the MAPE and 10-fold cross validation. In the first case, the data set was randomly split into a training (75% of the data) and a testing (25% of the data) set. The MAPEs were, however, different during testing with different data splits. To overcome the unstable performance, the *k*-fold (with $k = 10$) cross-validation with R-squared error (Hastie, Tibshirani, & Friedman, 2009, p. 241) was used.

Table 4 shows the validation results of the RF models and the OLS regression models for each of the six BoP categories. The RF models show better and more stable performance than OLS models both in MAPE and R-squared values. RF is in general less prone to over-fitting compared to other regression techniques when the ratio between the total number of input variables and the number of noisy variables among them is high and if the training data set has missing values (Deng, Fannon, & Eckelman, 2018; Hastie et al., 2009, p. 596). RF models have been used in different regression and classification tasks in the domain of sustainability and in connection with LCA studies. For example, in Froemelt et al. (2018) RF regression was applied for missing data imputation on utility costs and for modelling electricity demand in a study that aimed at assessing the environmental impacts of household consumption behaviour though the determination of lifestyle archetypes. In Froemelt, Buffat, and Hellweg (2019) the lifestyle archetypes were derived directly using a RF classifier. In Marvuglia, Leuenberger, Kanevski, and Benetto (2015) RF regression has been used to select most informative variables in the prediction of characterization factors for eco-toxicology and human toxicology of chemical compounds starting from molecular-based properties. The percentage increase in mean squared error (MSE) obtained by each model when shuffling values of one predictor has been used as an indicator of the importance of that predictor variable. A similar problem was addressed by Song (2019), but using a wide set of molecular structural descriptors and aiming at the prediction of the chemical properties of the substances. In the same line of research, Hou, Jolliet, Zhu, and Xu (2020) uses a RF model to estimate ecotoxicity hazardous concentrations 50% ($HC_{50}$) that can then be used to calculate characterization factors for chemicals. In the current study RFs are then used to synthesize the amount of BoP products consumed in each neighbourhood in The Hague.

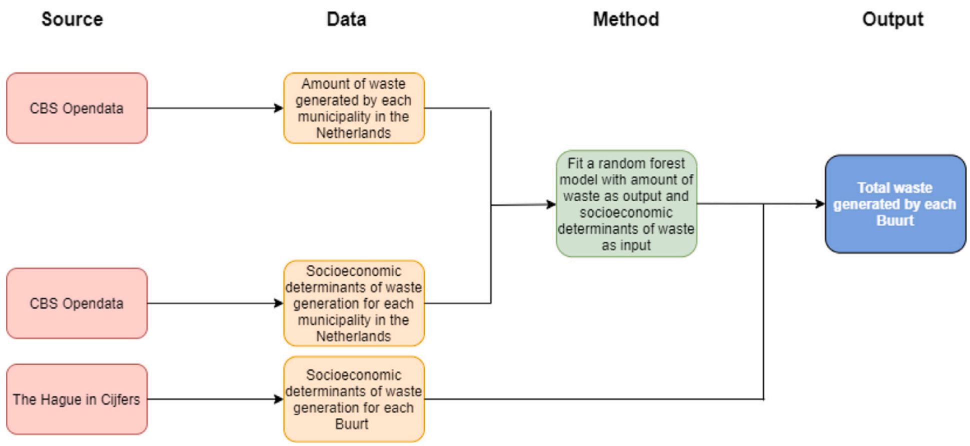


**Fig. 9.** Municipal solid waste data synthesis for different neighbourhoods of The Hague.

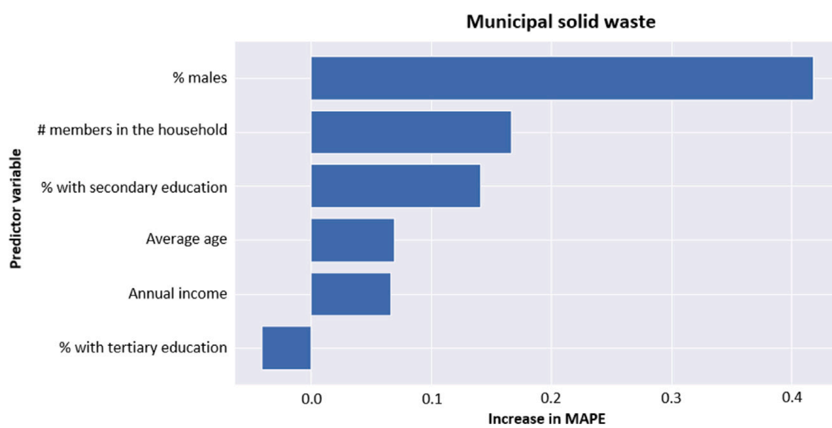


**Fig. 10.** Percentage increase in MAPE obtained with the RF model to predict the municipal solid waste, when a predictor variable is removed.

### 3.6. *Municipal solid waste*

Fig. 9 shows the methodology used for the computation of average municipal solid waste generated in each neighbourhood of the Hague. The CBS provides data on the average household waste generated in each municipality of the Netherlands. Linderhof, Kooreman, Allers, and Wiersma (2001) conducted an empirical study to determine factors that affect the amount of solid waste generated by Dutch households. The predictors include gender, age, income, education level and household size. With the CBS data set containing these predictors for all the 491 municipalities in the Netherlands, the same method as explained in Section 3.5 was used and a RF model with the average household waste as output was fitted. From Den Haag Cijfers, data about the five predictor variables for each neighbourhood of the city was obtained: gender (percentage male); average age; education (percentage of inhabitants with secondary and tertiary education); average income, and household size (number of members in the household). These variables were then used to predict the average municipal solid waste generated in each neighbourhood.

Fig. 10 shows the increase in MAPE in the model used to predict the quantity of municipal solid waste when one of the predictor variables is removed from the model. These values indicate the importance of the variable as predictor of the municipal solid waste. Analogous graphs for the RF models created for all the items of the BoP are shown in Appendix A. In this case one can notice that the percentage of male individuals in the population is slightly more informative than the other variables in predicting the quantity of municipal solid waste produced. However, the difference in MAPE is too low (a mere 0.22%) to establish a dominant position in the ranking between the first variable (percentage of males) and the following ones (number of members in the households, and the others to follow). The situation is clearly different for the other RF models (see Appendix A), especially for the ones used to predict the consumption of footwear, clothing and furniture items, where the annual income is clearly the most informative predictor.

### 3.7. *LCA flows: Consumption activities*

After the resource consumption in the six categories is synthesized for each neighbourhood of The Hague, the synthesized data is matched with the corresponding processes in the *ecoinvent* database (Wernet et al., 2016) to quantify the GHG emissions.

The GHG emissions for the neighbourhoods can be analyzed, e.g., on a per-household or per-capita basis, to inform policy interventions. Many national and international organizations, as well as local governments, have started to center their environmental policies around households and to try to steer a reduction of the environmental impacts of households as a whole (Girod, Stucki, & Woerter, 2017; Soderholm, 2015). Policy interventions are often formulated to target households instead of individuals since services such as energy, water and waste treatment are collectively availed by households as units.

## 4. Results

The methods and models discussed in Section 3 for the six resource consumption categories are applied to each neighbourhood ($N$ = 111) in The Hague to calculate the consumption per-capita in 2018. A summary of the DSE indicators used is shown in Table 5. The data is publicly available in Den Haag Cijfers. The results are discussed in the following paragraphs of this section.

**Table 5**
Descriptive statistics of demographics and socioeconomic indicators of the neighbourhoods in The Hague.

| Indicator | Min | Max | Mean | Std dev |
|---|---|---|---|---|
| Percentage male | 41.48 | 80.1 | 49.54 | 3.87 |
| Percentage female | 19.83 | 58.31 | 50.37 | 3.9 |
| Percentage population: 0 to 15 years | 2.51 | 32.12 | 16.53 | 5.15 |
| Percentage population: 15 to 25 years | 3.52 | 30.96 | 11.72 | 4.29 |
| Percentage population: 25 to 45 years | 10 | 50.83 | 28.37 | 8.73 |
| Percentage population: 45 to 65 years | 16.21 | 40 | 26.3 | 4.61 |
| Percentage population: 65 or older | 0.83 | 54.28 | 17.16 | 10.83 |
| Percentage population: Primary education | 8 | 69 | 30.46 | 14.39 |
| Percentage population: Secondary education | 13 | 46 | 34.99 | 6.41 |
| Percentage population: Tertiary education | 6 | 74 | 34.38 | 17.52 |
| Activity rate (in percentage) | 36.1 | 74.5 | 56.34 | 8.56 |
| Income per-household (€/year) | 17 200 | 89 900 | 36 475.7 | 16 218.12 |

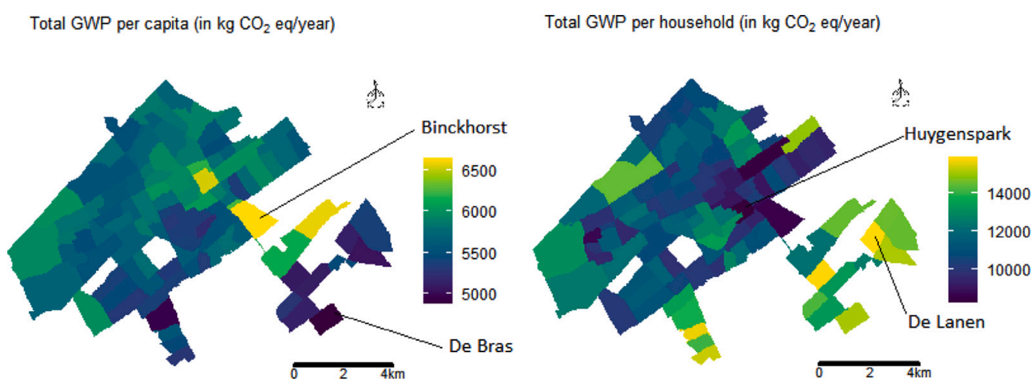


**Fig. 11.** Total GWP for the neighbourhoods of The Hague in 2018: (a) per-capita, (b) per-household.

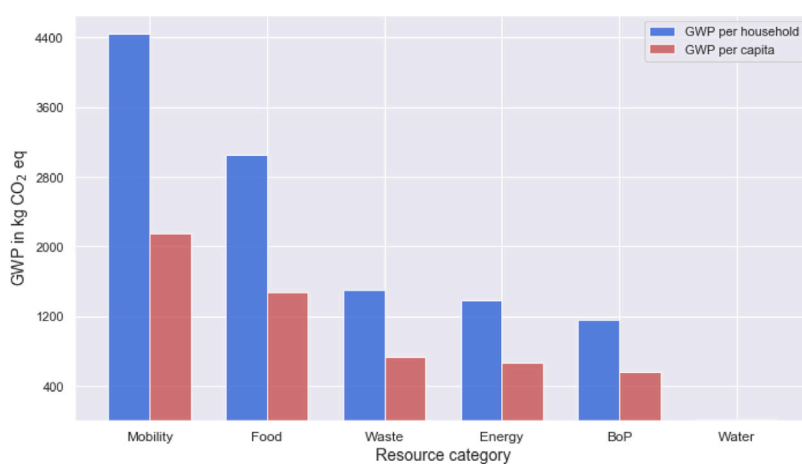


**Fig. 12.** GWP of the six resource categories in The Hague in 2018.

### *4.1. GWP of different neighbourhoods of the hague*

Fig. 11 shows the total GWP (a) per-capita and (b) per-household for the neighbourhoods of The Hague in 2018 by aggregating all resource consumption categories. The average GWP per-capita in the neighbourhoods range from 4.9 to 6.5 tonnes of $CO_2-equivalent$, which is lower than the Dutch national GWP per-capita (9.4 tonnes of $CO_2$ in 2018), because the current calculation excludes services such as post and public utility, leisure activities like dining out at a restaurant, and the emissions in the construction sector.

For the GWP per-household of the neighbourhoods, the trend is the opposite to that of GWP per-capita: the neighbourhoods with the highest GWP per-capita have some of the lowest GWP per-household and vice versa. This suggests that the neighbourhoods with high GWP per-capita have smaller household sizes (i.e., the number of family members).

When the household size increases, its total resource use increases, but the per-capita consumption of shareable goods decreases (as explained in Section 4.4), thus resulting in a possible decrease of the total GWP per-capita. For example, it can be expected that a two-person household would not use twice as much energy or generate twice as much waste compared to a single-person household.

### *4.2. GWP of different resource consumption categories*

Fig. 12 shows the GWP per-capita and per-household for different resource categories for The Hague in 2018. Mobility is the major contributing factor to GWP accounting for almost 45% of net GWP. On average, residents of The Hague travel by car for 30 km/day accounting for high mobility-related GWP. Thus, policies focused on stimulating the use of public transport could considerably reduce the negative environmental impacts of the city. The next major resource use which

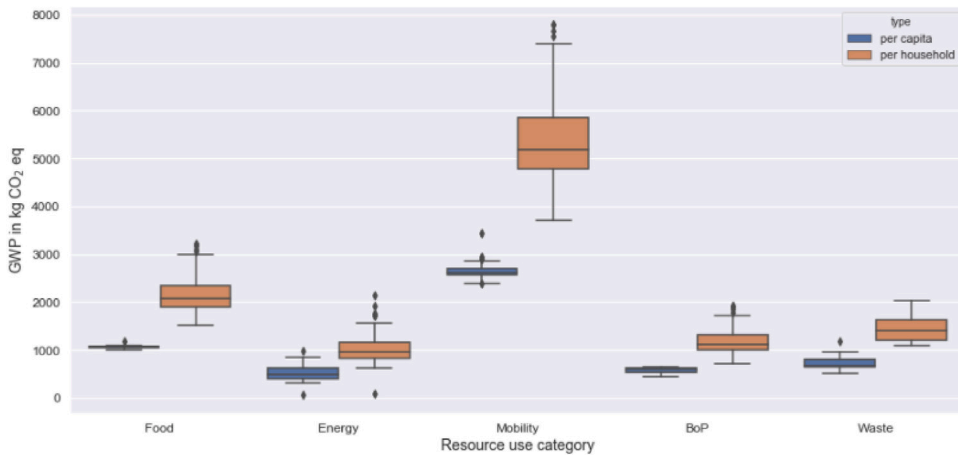


**Fig. 13.** GWP per-capita and per-household for 5 major resource use categories for the neighbourhoods of The Hague in 2018.

has a considerable negative environmental impact is food consumption, which accounts for nearly 20% of total GWP, followed by waste and BoP use, at 15% and 10% respectively. While the food habits of people cannot be controlled, consumption of organic food products (which in the vast majority of the cases have lower carbon emissions compared to traditional ones[6]) can be promoted.

For energy use, the majority of GWP is due to the use of electricity generated from non-renewable sources like coal, oil and natural gas. In The Hague, households are given a choice to select the source of their electricity and they can choose green (renewable) or non-green (non-renewable) sources of electricity. Therefore, promoting green or renewable sources of electricity has a great potential to reduce GWP due to energy use.

Finally, the use of water contributes by less than 1% to total carbon emissions. Thus, the contribution to GWP due to water use in the city is negligible compared to the use of other resources.

### *4.3. GWP per-capita and GWP per-household*

Fig. 13 shows the distribution of GWP per-capita and GWP per-household for the neighbourhoods of The Hague in the five major resource use categories (water is excluded due to its extremely small contribution to GWP, as highlighted above). The distribution of GWP per-capita exhibits much less variance compared to the GWP per-household across all resource use categories. The effect of variance in household size combined with variance in GWP per-capita leads to a higher variance in GWP per-household in every resource use category.

Mobility has the most inequitable distribution of GWP per-household (i.e. the highest variance) even though its distribution of GWP per-capita is equitable (i.e. it is characterized by low variance). This can be interpreted by considering that neighbourhoods with a high GWP per-capita in general also have large household size and, as a result, the variance in GWP per-capita is large. In practical terms, the large differences in GWP between smaller and larger households can be interpreted by the fact that large households have a high GWP per-capita probably due to the combined effect of more members in the household and a higher car ownership and use, whereas smaller households (which are not concerned by this effect) have lower GWP per-capita.

Contrary to that, the variances for GWP per-capita and GWP per-household in the energy use category are comparable. This means that neighbourhoods with smaller GWP per-capita due to energy use have large household sizes and neighbourhoods with larger GWP per-capita have smaller household size. The effects of high GWP per-capita and small household sizes, and low GWP per-capita and large household sizes, balance out, thus resulting in a lower variance in GWP per-household for all neighbourhoods.

As a conclusion of the patterns described above, in a policy perspective it would seem reasonable addressing towards smaller households interventions targeting energy efficiency and towards larger households interventions targeting more sustainable mobility.

### *4.4. Cluster-based analysis of environmental impacts*

The environmental impacts arising from the activities of the residents of a city are correlated to a large extent with their socioeconomic conditions, as shown by Cárdenas-Mamani et al. (2022) and Tang, Wang, Lee, and Yang (2022) for household energy use, and Li et al. (2022) for the entire set of city-level $CO_2$ emissions. In order to further understand this relationship for the residents of The Hague, different neighbourhoods of the city are clustered based on their socioeconomic characteristics and the environmental impacts of neighbourhoods' clusters are analyzed. The break down of environmental impacts into resource use categories, along with information on the environmental impact of socioeconomic groups in different resource use categories, differentiated per neighbourhood, helps policymakers to target their policies to specific socioeconomic groups (Froemelt et al., 2018) and enable them to customize policies and sustainability messages that can encourage sustainable behaviour among the residents of a specific neighbourhood of a city. In the current study, the variables used to cluster the neighbourhoods are : (1) Percentage of population younger than 40 years of age; (2) Percentage of population with tertiary education; (3) Activity rate (1-unemployment rate); (4) Percentage of non-Dutch population; (5) Percentage of single person households; (6) Percentage of rental houses; (7) Number of cars per household; (8) Annual income per household. The indicators were chosen based on the assumption that they are the factors that affect the consumption choices of residents. Several clustering algorithms exist in the literature, the most common of which being *K means* (Jain, 2010). In the case study described in this paper, seven clustering methods available in the Python Scikit-learn library (Pedregosa et al., 2011) were applied to the neighbourhoods of The Hague. The optimal clustering method was then chosen by firstly comparing the performances of the different clustering algorithms across similar performance metrics (more information in Appendix B) and then looking at the practical implications of the obtained neighbourhood clusters through different methods in terms of policy intervention and decision making for the municipality of The Hague. Based on the analysis shown in Appendix B, it was found that *K*

[6] the contribution of organic food to other impact categories than GWP does not show a clear trend with respect to traditional, i.e. non-organic, food, because it depends on several factors. See for example Clark and Tilman (2017).

**Table 6**
Average of the DSE metrics of the neighbourhoods in the 5 clusters identified.

| DSE indicator | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Age (0–40) years (%) | 67.5 | 42.2 | 49.69 | 58.46 | 34.8 |
| Tertiary education (%) | 32.75 | 34.36 | 60.7 | 13 | 39.66 |
| Employment rate (%) | 66.6 | 58.38 | 57.2 | 47.7 | 55.1 |
| Expat population (%) | 11.4 | 19.44 | 29.5 | 13.1 | 23.1 |
| One person households (%) | 23.22 | 47.2 | 45.8 | 48.4 | 62.7 |
| Rental houses (%) | 34.9 | 41.4 | 35.2 | 74.8 | 71.1 |
| Number of cars per-household | 1.24 | 0.77 | 0.99 | 0.57 | 0.46 |
| Income (€/year) | 27 962 | 27 081 | 48 355 | 16 266 | 26 393 |

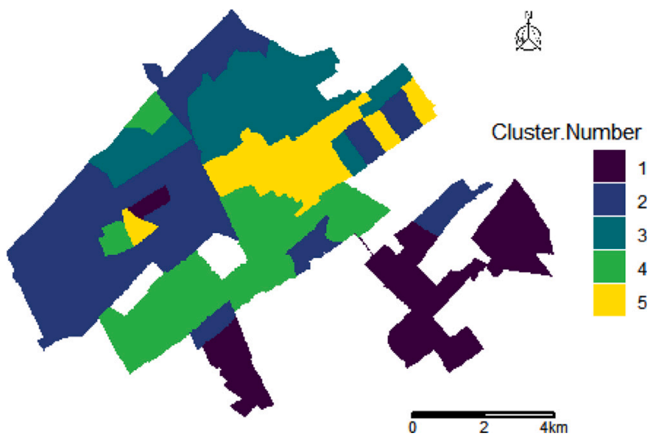


**Fig. 14.** Subdivision of neighbourhoods of The Hague into 5 clusters.

*means* clustering algorithm with 5 clusters is the best method to cluster the neighbourhoods of The Hague. The authors further deliberated on the naming of the clusters and the clusters were named based on the average socioeconomic attributes of the households in the respective clusters.

Fig. 14 shows the neighbourhoods of The Hague and their allocated clusters and Table 6 shows the centroid values of the DSE indicators for the 5 neighbourhoods clusters along with the defining characteristics for each cluster. The clusters are defined as follows: Cluster 1 includes young, highly active middle class households using many cars; Cluster 2 includes relatively older, moderately educated, middle class households with average car use; Cluster 3 includes middle aged, rich, highly educated households with high home and car ownership; Cluster 4 groups low income, low educated households with high degree of unemployment; Cluster 5 includes older, middle class single person households with very low car use. The neighbourhoods with moderately higher standards of living represented by cluster 2 and 3 are along the northern and western parts of the city, whereas the neighbourhoods with relatively lower standards of living are present along the eastern and southern parts of the city.

Fig. 15 shows, for each of the 5 clusters and for each resource use category, the ratio between the GWP per-household of that cluster and the average GWP of The Hague in the same resource use category. Cluster 1 has the highest GWP in terms of food, mobility, BoP use as well as the total GWP, which are respectively 35%, 30%, 45% and 28% compared to the average GWP per-household of The Hague. This can primarily be attributed to the fact that households belonging to Cluster 1 are composed by more members than households belonging to other clusters. In Cluster 1, only 23% of the households are single-person households. As a result, the food consumption and BoP use is expected to be higher compared to households belonging to other clusters.

Resources like food and BoP often show a strong positive correlation with the number of members in a household, as well as with the average income of the household, whereas energy use (Kavousian, Rajagopal, & Fischer, 2013) and waste generation (Noufal, Yuanyuan, Maalla, & Adipah, 2020; Suthar & Singh, 2015) generally exhibit a weaker correlation with the household's size. In other words, when the household's size increases, energy use and waste generation increase at a slower rate than food consumption. This happens because food (which is defined as a private good, as opposed to public goods like energy) cannot be easily shared among household members,[7] while energy and other public goods can, thus causing larger households to have higher aggregate consumption, but lower per-capita consumption for energy (economy of scale effect), but not for food (Wu et al., 2021).

As a result, also the total GWP due to food and BoP consumption of multiple-member households increases more rapidly with the household's size than the GWP due to the energy use. Another feature of Cluster 1 is a very high amount of car ownership along with high activity (employment) rate. Consequently, the car use is expected to be high and the corresponding GWP due to mobility is the highest. The combination of more multiple-people households and high degree of car ownership results in Cluster 1 having the highest GWP per-household.

Cluster 2, which is characterized by moderate values of most of the features, has almost the same GWP per-household as the average of The Hague in all the categories.

Neighbourhoods in Cluster 3, which experience the highest standards of living due to high income, education level and degree of home and car ownership, would be expected to have high GWP per-household. However, Cluster 3 contains many single-person households and its car ownership level is still significantly lower than in Cluster 1. As a result, GWP due to energy use is the highest for Cluster 3 (around 42% more than the average of The Hague) but GWP due to mobility is quite low compared to Cluster 1.

Neighbourhoods in Cluster 4 have the lowest income, education level and activity rate and they have overall high GWP per-household in waste and mobility category. The high GWP related to waste could be attributed to a low activity rate, which results in more people staying at home instead of working, thus generating more waste. It is also possible that features like presence of young residents, low home ownership rate and low education level, whose impact on GWP is not observable directly, results in such a behaviour.

Finally, Cluster 5 is characterized by the highest percentage of single-person households and the lowest car ownership. These factors lead to the lowest GWP overall, as well as the lowest GWP due to mobility. The low GWP due to mobility in neighbourhoods with low car ownership (and vice versa) shown through the analysis also validates the approach for mining mobility data since it is inherently expected that neighbourhoods with lower car ownership would have a low GWP due to mobility.

## 5. Conclusion

One major challenge when quantifying the environmental impact of a city is the lack of data at the local scale. Few local city governments collect such data. Data related to resource consumption such as food,

[7] although this statement cannot be generalized, as shown in Jacobson, Mavrikiou, and Minas (2010).

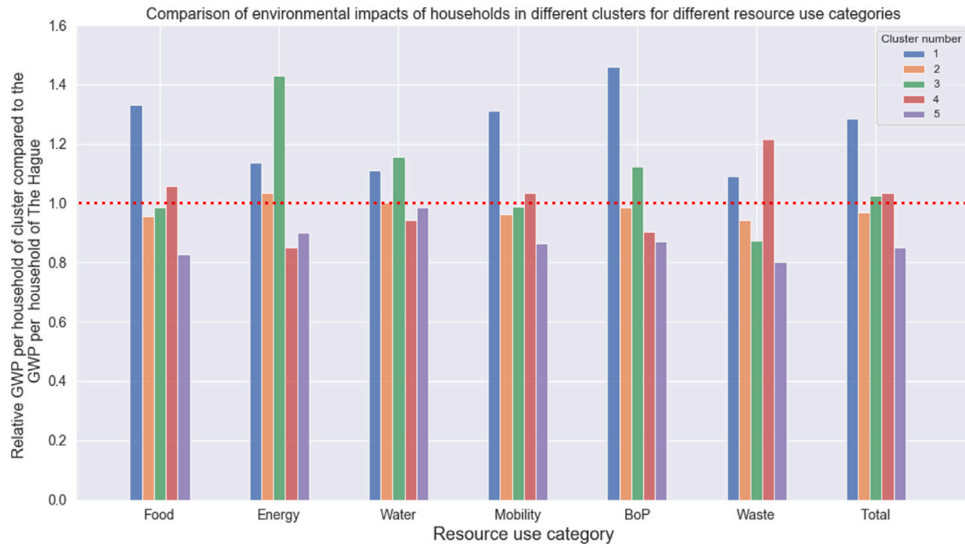


**Fig. 15.** Relative average GWP per-household for different clusters compared to the average GWP per-household of The Hague (dashed line) for different resource use categories in 2018.

mobility, water and energy use, are often only available at the national level, while data on socioeconomic indicators such as income, education and employment rate, can sometimes be available at the local (or neighbourhood) level in countries (such as the Netherlands) that have well-organized data collection processes. At the local level, the lack of resource consumption data represents a hurdle for the quantification of the environmental impacts of households. However, the availability of data on socioeconomic characteristics at the local household level presents an opportunity to model the resource consumption of households.

This paper proposes a methodology that can quantify the carbon footprint of households in cities using consumption data from a national or higher geographical level. The local neighbourhood resource consumption can be synthesized by linking the socioeconomic attributes of local populations to those of national (or higher level) populations. The methodology classifies neighbourhood resource consumption per category, qualifies local resource consumption using predictive models, and translates the consumption to GHG emissions in terms of GWP at local level per category. Analyzing the GWP at a neighbourhood level enables the identification of "hot-spots" where consumption activities and socioeconomic groups have high environmental impacts. This way, policy makers can design environmental policies to nudge targeted local consumers towards more environmentally friendly behaviour.

The methodology was applied to the case study of The Hague in the Netherlands to quantify the GWP of the city's 111 neighbourhoods. As a first step, publicly available datasets of resource consumption (along with the socioeconomic attributes of populations) whose geospatial resolution range from the countries in the EU to the neighbourhoods of The Hague were collected. The data were processed into six household consumption categories. Second, predictive models were constructed from the data of national (or the EU) resolutions. These models were applied to the neighbourhoods of The Hague to derive the consumption patterns per category. When the required data sets were not available, the predictive models in literature were applied instead. Finally, the synthesized local resource consumption data was translated into GHG emissions by matching resource consumption with processes in the *ecoinvent* 3.6 LCI database. The resource consumption categories in this study are derived from the UN's classification of individual consumption according to purpose (COICOP) categories by selecting those that produce direct environmental impacts by households. While the categories can be similar for some households in some countries, the further division into actual materials and energy flows as well as the data collection processes can be different in other countries. Thus, the constructed models for The Hague households are specific to the local context and should not be directly applied to households in other countries without adaptation.

Furthermore, in countries with data collection mechanisms similar to the Netherlands, the population socioeconomic data is collected locally for different city neighbourhoods. When such data is not available, the proposed predictive models are not able to compute GHG emissions at the neighbourhood level.

Lastly, regarding the background LCA data, this study used the *ecoinvent* 3.6 database to quantify the environmental impacts of resource consumption. This database did not contain flows relating to all the 133 food items. Therefore, proxies were used in the cases where the exact item was missing. For example, cheese was used as a proxy for butter and milk as a proxy for yoghurt. These assumptions are coherent with the assumptions made in the EC-JRC study (Sala, De Laurentiis, Barbero Vignola, Marelli, & Sanyé Mengual, 2022) and the GHG values used for these products are in the ranges suggested by other literature studies (Djekic, Miocinovic, Tomasevic, Smigic, & Tomic, 2014; Notarnicola, Tassielli, Renzulli, Castellani, & Sala, 2022). For other food items like eggs (Temme et al., 2015), or tofu and peanut butter (Hamerschlag, 2011) other literature sources have been used.

Despite the highligthed limitations, the methodology and case study presented in this paper provide valuable instruments and implementation examples for detailed quantification of city GHG emissions. The data disaggregation methods and steps described are useful to identify the most relevant method that can synthesize resource consumption data. In addition, many of the model elements can be applicable to other European cities, in particular Dutch cities, since they have similar context and data availability. For example, the RF models of BoP and waste generation (using socioeconomic attributes of European countries) can be applicable to other European cities. The demographic clustering of food consumption and mobility patterns can be applied to other major Dutch cities such as Amsterdam or Rotterdam.

The result of the case study showed that, in The Hague, mobility, household solid waste and energy use were the top three emission sources among the six categories that were accounted for in our study. These three categories were responsible for 70% of the total consumption emissions of the city.

On average, households with low income and education, and high unemployment, generate considerably more waste compared to other households. This confirms the findings by other studies (Noufal et al., 2020; Suthar & Singh, 2015). A preliminary analysis conducted in this study showed a reduction of 30% GHG emissions in the mobility sector

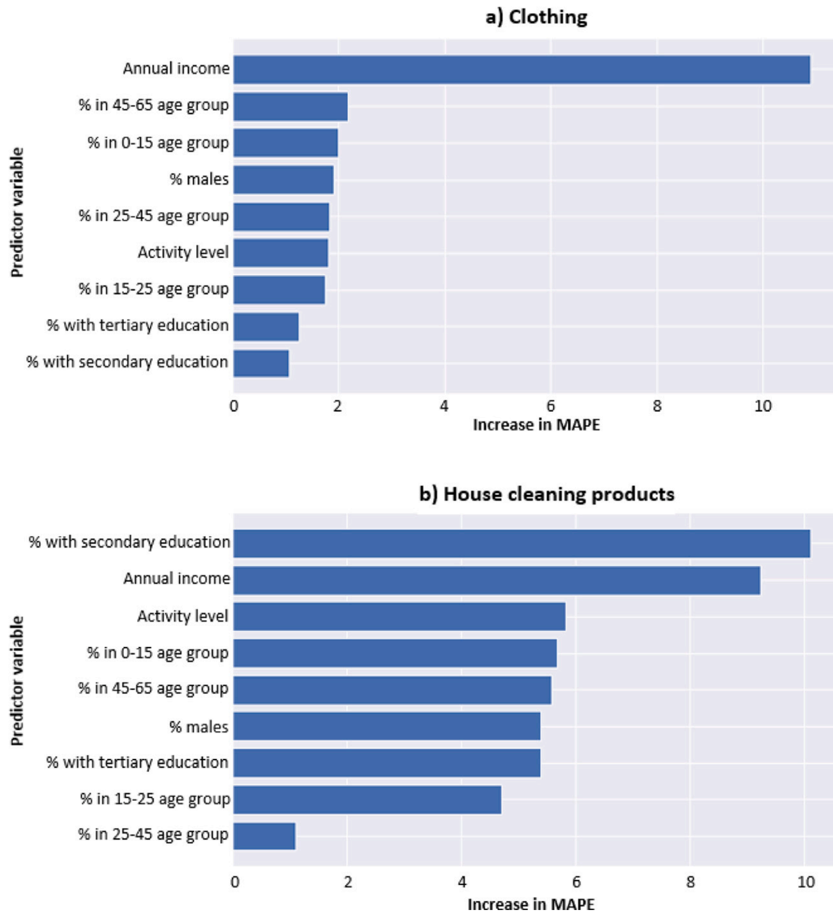


**Fig. A.1.** MAPE increase for the RF models used to predict the consumption of clothing (a) and house cleaning products (b).

in The Hague by an increase in car parking fees and subsidizing public transport. Such insights into the environmental impact per consumption category and per neighbourhood are informative to municipal policy-makers, local governments and community managers for more targeted and effective environmental policies and implementation.

A shown in the introduction section, since the topic of consumption-driven impacts estimation at city level is very timely, other studies similar to the one described in this paper have already been conducted (Dorr et al., 2022; Genta et al., 2022; Hillman & Anu, 2010) using similar approaches. However, every study shows the need to adapt and extend the analysis methodology according to specific data and geographical contexts. This paper can then serve as a general guideline for future studies.

## 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.

## Data availability

The data sources are public and the sources have been cited in the manuscript

## Acknowledgements

At the time when the analyses described in this article were carried out, Ruchik Patel was a MSc. student of Delft University of Technology (TU Delft) conducting an internship at the Luxembourg Institute of Science and Technology (LIST) and Paul Baustert was affiliated with LIST.

## Formatting of funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Appendix A

Figs. A.1 to A.3 show the MAPE increase for the RF models used to predict the consumption of different BoP items when a predictor variable is removed. As can be seen, the increase in MAPE is positive for most of the predictors for different BoP items, which implies that the predictor variable used for the prediction of BoP consumption improve the accuracy of the models.

## Appendix B

Two of the metrics commonly used to assess the performance of clustering algorithms are *silhouette score* (Rousseeuw, 1987) and *Davies–Bouldin index* (Davies & Bouldin, 1979). To define the silhouette score, let us take any object ***i*** in the data set, and denote by $A$ the cluster to which it has been assigned. When cluster $A$ contains other objects apart from ***i***, then one can compute the average distance $a(i)$ of ***i*** from all the objects within $A$. Let us now consider any cluster $C \neq A$ and compute the mean $d(i, C)$ of the distances from ***i*** to all the objects in $C$. Let us select the smallest of those numbers and denote it by $b(i)$. The cluster with this smallest mean distance is termed the "neighbouring cluster" of ***i*** because it is the next best fit cluster for point ***i***. The silhouette width for data point ***i*** in cluster $A$ is defined by Eq. (B.1):

$$\begin{cases} s(i) = \frac{b(i)-a(i)}{max\{(a(i),b(i)\}} & if\ |C_A| > 1 \\ s(i) = 0 & if\ |C_A| = 1 \end{cases} \tag{B.1}$$

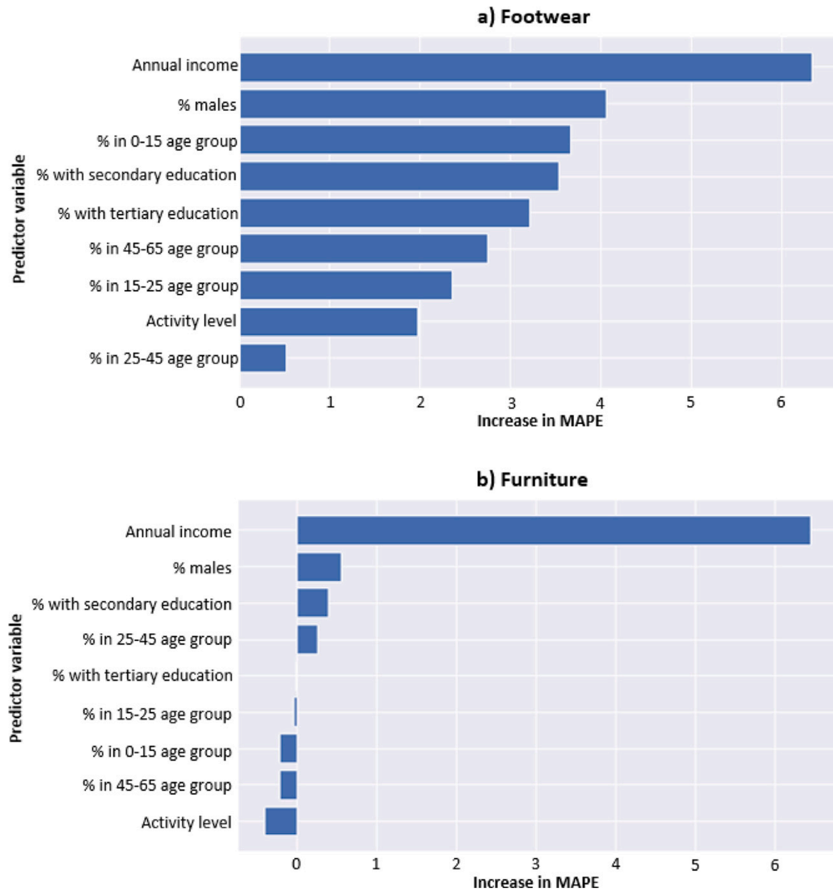


**Fig. A.2.** MAPE increase for the RF models used to predict the consumption of footwear (a) and furniture products (b).

where $|C_A|$ is the cardinality of $A$, i.e. the number of elements in $A$.

By definition, is $-1 \leqslant s(i) \leq +1$. For $s(i)$ to be close to 1, it must be $a(i) \ll b(i)$. Since $a(i)$ is a measure of distance (in other words, dissimilarity) between $i$ and all the elements in its own cluster, a small value of $a(i)$ means that $i$ is well matched to its cluster. For the same reason, a large value of $b(i)$ implies that $i$ is badly matched to its neighbouring cluster. Thus, an $s(i)$ close to 1 means that the data is well clustered. If $s(i)$ is close to −1, then it means that it would have been more appropriate if $i$ had been assigned to its neighbouring cluster. An $s(i)$ near zero means that the object $i$ is on the border of two clusters. The silhouette $S_A$ of cluster $A$ is defined as the average of the silhouette widths of all the objects contained in cluster $A$. Given a certain clustering that divides the data set in $K$ clusters, the mean of the silhouette of the clusters over all the $K$ clusters is the global *silhouette score* of the clustering:

$$S = \frac{1}{K}\sum_{A=1}^{K} S_A \tag{B.2}$$

The Davies–Bouldin index is a measure of similarity of each cluster with its most similar cluster. Similarity is the ratio between the intra-cluster distances and the inter-cluster distances (defined as the distance between the clusters' centroids). A lower value of Davies–Bouldin index is then preferable (Beccali, Cellura, Lo Brano, & Marvuglia, 2004). Fig. B.1 shows the silhouette and Davies–Bouldin scores for different clustering algorithms applied in this study to the neighbourhoods of The Hague. The optimal number of clusters for each algorithm was chosen such that the number of clusters for each method maximized the silhouette score for that method. As mentioned earlier, the higher the silhouette score and lower the Davies–Bouldin index, the better the clustering.

There is often a trade-off between higher silhouette score and lower Davies–Bouldin index value. In the case study described in this paper, seven different clustering algorithms were run on the data set using the Python Scikit-learn package (Pedregosa et al., 2011). The *OPTICS* clustering algorithm had a very low silhouette score, whereas the *mini batch K means* algorithm had a very high Davies–Bouldin score. Thus, also these two algorithms were ruled out from further analysis. Of the remaining five algorithms, *K means* had a slightly higher silhouette score compared to the other algorithms and *mean shift* had the lowest Davies–Bouldin index value. However, the *mean shift* algorithm divided the neighbourhoods into only two clusters, of which one contained only two neighbourhoods (Vlietzoom east and Vlietzoom west) and the other contained the rest of the neighbourhoods. Therefore, even though *mean shift* had a relatively high silhouette score and the lowest Davies–Bouldin index value, the partitioning obtained using this algorithm could not support any practical inference in terms of policy interventions and decision making. Therefore also *mean shift* was ruled out. The silhouette score and Davies–Bouldin index values obtained with the remaining algorithms were very similar. However, all the algorithms except *K means* identified one dominant cluster accounting for almost half of the neighbourhoods of the city (despite their differences in socioeconomic characteristics) and one cluster containing a very small number of neighbourhoods. This partitioning would not allow to properly take into account the socioeconomic differences among the different districts; therefore the *K means* clustering algorithm was chosen as the optimal algorithm for further analysis.

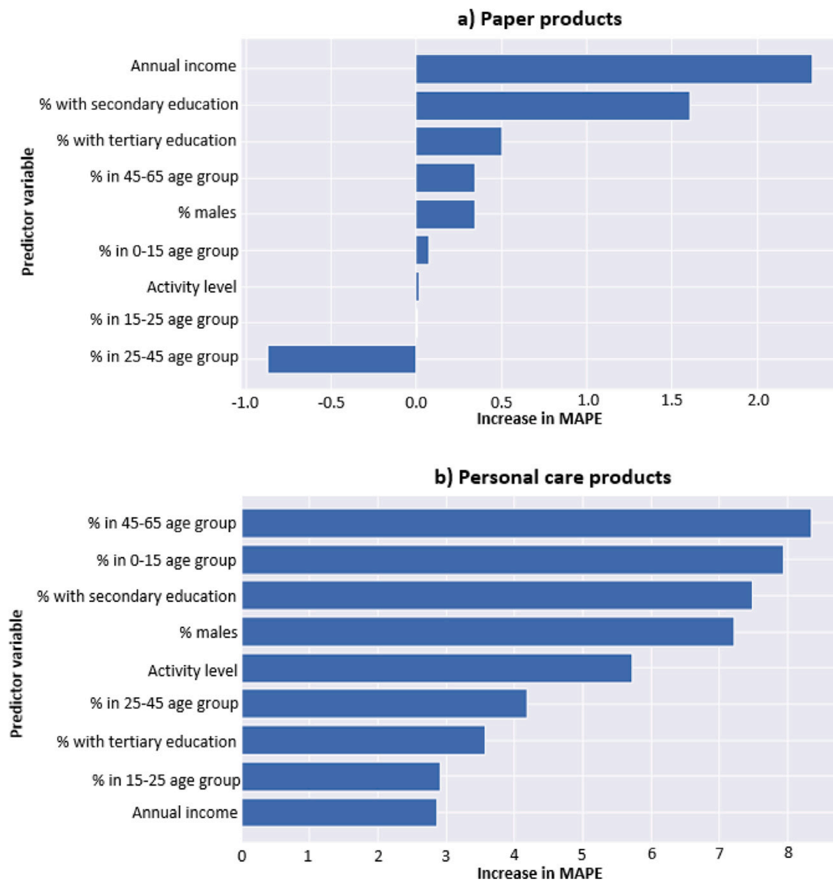


**Fig. A.3.** MAPE increase for the RF models used to predict the consumption of paper (a) and personal care products (b).

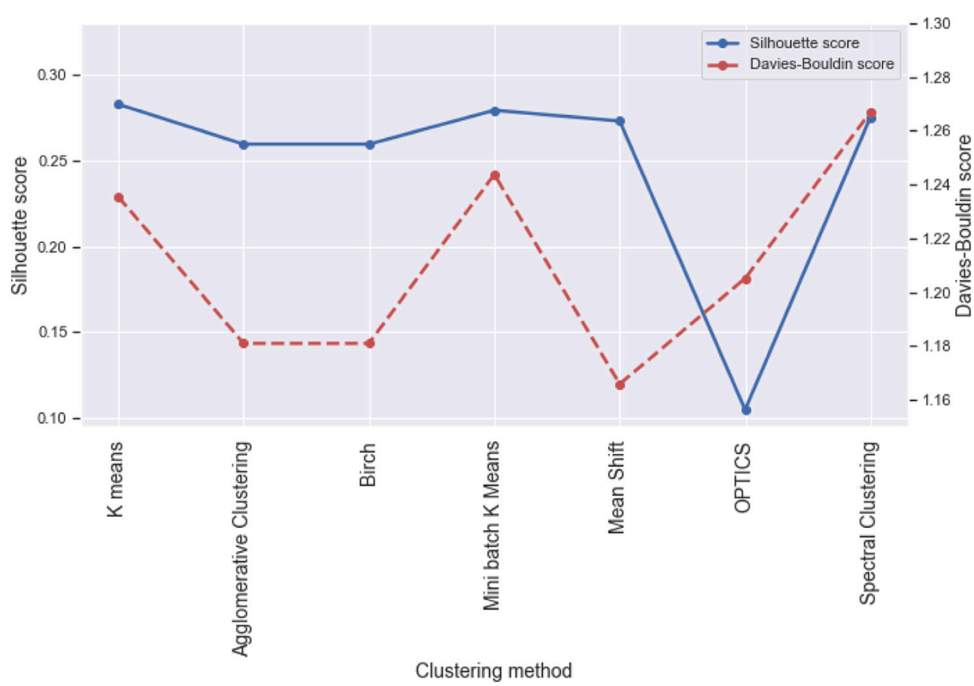


**Fig. B.1.** Silhouette and Davies–Bouldin scores for different clustering methods applied to the neighbourhoods of The Hague.

## References


Baldassarri, C., Allacker, K., Reale, F., Castellani, V., & Sala, S. (2017). *Consumer footprint: Basket of products indicator on housing*: *KJ-NA-28765-EN-C (print),KJ-NA-28765-EN-N (online)*, Luxembourg (Luxembourg): Publications Office of the European Union, http://dx.doi.org/10.2760/05316.

Balouktsi, M. (2020). Carbon metrics for cities: production and consumption implications for policies. *Buildings and Cities*, *1*(1), 233–259. http://dx.doi.org/10.5334/bc.33.

Beccali, M., Cellura, M., Lo Brano, V., & Marvuglia, A. (2004). Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management*, *45*, 2879–2900. http://dx.doi.org/10.1016/j.enconman.2004.01.006.

Broekhoff, D., Erickson, P., & Piggot, G. (2019). *Estimating consumption-based greenhouse gas emissions at the city scale. A guide for local governments*: *Technical report February*, Stockholm Environment Institute (SEI), URL: https://www.sei.org/wp-content/uploads/2019/03/estimating-consumption-based-greenhouse-gas-emissions.pdf.

Brounen, D., Kok, N., & Quigley, J. M. (2013). Energy literacy, awareness, and conservation behavior of residential households. *Energy Economics*, *38*, 42–50. http://dx.doi.org/10.1016/j.eneco.2013.02.008.

Büchs, M., & Schnepf, S. V. (2013). Who emits most? Associations between socio-economic factors and UK households' home energy, transport, indirect and total $CO_2$ emissions. *Ecological Economics*, *90*, 114–123. http://dx.doi.org/10.1016/j.ecolecon.2013.03.007.

Cárdenas-Mamani, U., Kahhat, R., & Vázquez-Rowe, I. (2022). District-level analysis for household-related energy consumption and greenhouse gas emissions: a case study in Lima, Peru. *Sustainable Cities and Society*, *77*, Article 103572. http://dx.doi.org/10.1016/j.scs.2021.103572.

Carfagna, L. B., Dubois, E. A., Fitzmaurice, C., Ouimette, M. Y., Schor, J. B., Willis, M., et al. (2014). An emerging eco-habitus: the reconfiguration of high cultural capital practices among ethical consumers. *Journal of Consumer Culture*, *14*(2), 158–178.

Clark, M., & Tilman, D. (2017). Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. *Environmental Research Letters*, *12*, Article 064016. http://dx.doi.org/10.1088/1748-9326/aa6cd5.

Corsatea, T., Lindner, S., Arto, I., Román, M., Rueda-Cantuche, J., Velázquez Afonso, A., et al. (2019). World input-output database environmental accounts. update 2000-2016. *Publications office of the European union*. Article JRC116234. http://dx.doi.org/10.2760/024036, URL: https://ec.europa.eu/jrc.

Darmawan, D., Mangundjaya, W. L., & Herdiansyah, H. (2018). In searching the appropriate elements: study on the community's environmental friendly behavior in daily activities to achieve an urban sustainability. In *E3S web of conferences, Vol. 74* (p. 08003). http://dx.doi.org/10.1051/e3sconf/20187408003.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*, 224–227. http://dx.doi.org/10.1109/TPAMI.1979.4766909.

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: a comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy and Buildings*, *163*, 34–43. http://dx.doi.org/10.1016/j.enbuild.2017.12.031.

Dhakal, S. (2009). Urban energy use and carbon emissions from cities in China and policy implications. *Energy Policy*, *37*(11), 4208–4219.

Djekic, I., Miocinovic, J., Tomasevic, I., Smigic, N., & Tomic, N. (2014). Environmental life-cycle assessment of various dairy products. *Journal of Cleaner Production*, *68*, 64–72. http://dx.doi.org/10.1016/j.jclepro.2013.12.054.

Dorr, E., François, C., Poulhès, A., & Wurtz, A. (2022). A life cycle assessment method to support cities in their climate change mitigation strategies. *Sustainable Cities and Society*, Article 104052. http://dx.doi.org/10.1016/j.scs.2022.104052.

Dutch Central Bureau of Statistics (CBS) (2018). Household and farm water usage surged in 2018. https://www.cbs.nl/en-gb/news/2020/12/household-and-farm-water-usage-surged-in-2018. [Accessed on: 23.10.2021].

EC-JRC (2018). *Development of a weighting approach for the environmental footprint*: *Technical report*, European Commission - Joint Research Center, URL: https://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml.

EC-JRC (2022). *Consumption footprint: assessing the environmental impacts of EU consumption. Science for policy brief*: *Technical report*, European Commission - Joint Research Center.

EU (2020). Database - eurostat. URL: https://ec.europa.eu/eurostat/data/database.

Fazio, S., Castellani, V., Sala, S., Schau, E., Secchi, M., Zampori, M., et al. (2018). *Development of a weighting approach for the environmental footprint. New models and differences with ILCD*: *Technical report*, European Commission - Joint Research Center, URL: https://eplca.jrc.ec.europa.eu/permalink/supporting_Information_final.pdf.

Feng, J. (2015). From urbanization to urban decay: the problems of modernization, urbanization and industrialization - the case of detroit. In *Global modernization review: new discoveries and theories revisited* (pp. 91–99). World Scientific Publishing Co., http://dx.doi.org/10.1142/9789814616072_0010.

Finnveden, G., Hauschild, M. Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., et al. (2009). Recent developments in life cycle assessment. *Journal of Environmental Management*, *91*(1), 1–21. http://dx.doi.org/10.1016/j.jenvman.2009.06.018, URL: https://www.sciencedirect.com/science/article/pii/S0301479709002345.

Froemelt, A., Buffat, R., & Hellweg, S. (2019). Machine learning based modeling of households: A regionalized bottom-up approach to investigate consumption-induced environmental impacts. *Journal of Industrial Ecology*, *24*(3), 639–652.

Froemelt, A., Dürrenmatt, D. J., & Hellweg, S. (2018). Using data mining to assess environmental impacts of household consumption behaviors. *Environmental Science and Technology*, *52*(15), 8467–8478. http://dx.doi.org/10.1021/acs.est.8b01452.

Froemelt, A., & Wiedmann, T. (2020). A two-stage clustering approach to investigate lifestyle carbon footprints in two australian cities. *Environmental Research Letters*, *15*(10), Article 104096. http://dx.doi.org/10.1088/1748-9326/abb502.

Ganong, P., & Noel, P. (2019). Consumer spending during unemployment: positive and normative implications. *American Economic Review*, *109*(7), 2383–2424. http://dx.doi.org/10.1257/aer.20170537.

Genta, C., Sanyé-Mengual, E., Sala, S., & Lombardi, P. (2022). The consumption footprint as possible indicator for environmental impact evaluation at city level. The case study of Turin (Italy). *Sustainable Cities and Society*, *79*, Article 103679. http://dx.doi.org/10.1016/j.scs.2022.103679.

Girod, B., Stucki, T., & Woerter, M. (2017). How do policies for efficient energy use in the household sector induce energy-efficiency innovation? an evaluation of european countries. *Energy Policy*, *103*(November 2016), 223–237. http://dx.doi.org/10.1016/j.enpol.2016.12.054.

Gouldson, A., Colenbrander, S., Sudmant, A., Papargyropoulou, E., Kerr, N., McAnulla, F., et al. (2016). Cities and climate change mitigation: economic opportunities and governance challenges in Asia. *Cities*, *54*, 11–19. http://dx.doi.org/10.1016/j.cities.2015.10.010.

Guan, D., Hubacek, K., Weber, C. L., Peters, G. P., & Reiner, D. M. (2008). The drivers of Chinese $CO_2$ emissions from 1980 to 2030. *Global Environmental Change*, *18*(4), 626–634. http://dx.doi.org/10.1016/j.gloenvcha.2008.08.001.

Hamerschlag, K. (2011). *Meat eaters guide to climate change + health*: *Technical report*, Washington, D.C.: Environmental Working Group, URL: https://static.ewg.org/reports/2011/meateaters/pdf/report_ewg_meat_eaters_guide_to_health_and_climate_2011.pdf.

Hasegawa, R., Kagawa, S., & Tsukui, M. (2015). Carbon footprint analysis through constructing a multi-region input–output table: a case study of Japan. *Journal of Economic Structures*, *4*(1), 5. http://dx.doi.org/10.1186/s40008-015-0015-6.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). (p. 745). Springer, http://dx.doi.org/10.1007/b94608.

Hillman, T., & Anu, R. (2010). Greenhouse gas emission footprints and energy use benchmarks for eight U.S. cities. *Environmental Science and Technology*, *44*(6), 1902–1910.

Hirashi, T., Krug, T., Tanabe, K., Srivastava, N., Jamsranjav, B., Fukuda, M., et al. (2014). *IPCC 2014, 2013 revised supplementary methods and good practice guidance arising from the kyoto protocol*: *Technical report*, (p. 268). Switzerland: IPCC.

Hoornweg, D., Sugar, L., & Lorena Trejos Gómez, C. (2011). Cities and greenhouse gas emissions: moving forward. *Environment & Urbanization*, *23*(1), 207–227. http://dx.doi.org/10.1177/0956247810392270, URL: www.sagepublications.com.

Horta, I. M., & Keirstead, J. (2017). Downscaling aggregate urban metabolism accounts to local districts. *Journal of Industrial Ecology*, *21*(2), 294–306, URL: https://ideas.repec.org/a/bla/inecol/v21y2017i2p294-306.html.

Hou, P., Jolliet, O., Zhu, J., & Xu, M. (2020). Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environment International*, *135*, Article 105393. http://dx.doi.org/10.1016/j.envint.2019.105393.

Huang, Y., & Warnier, M. (2019). Bridging the attitude-behaviour gap in household energy consumption. In *Proceedings of IEEE PES innovative smart grid technologies Europe (ISGT-Europe)*. IEEE.

Ivanova, D., Barrett, J., Wiedenhofer, D., Macura, B., Callaghan, M., & Creutzig, F. (2020). Quantifying the potential for climate change mitigation of consumption options. *Environmental Research Letters*, *15*(9), Article 093001. http://dx.doi.org/10.1088/1748-9326/ab8589.

Jacobson, D., Mavrikiou, P. M., & Minas, C. (2010). Household size, income and expenditure on food: the case of cyprus. *The Journal of Socio-Economics*, *39*(2), 319–328. http://dx.doi.org/10.1016/j.socec.2009.12.009, Financial Crisis: Economic and Psychological Perspectives.

Jain, A. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.

Kamakura, W. A., & Mazzon, J. A. (2013). Socioeconomic status and consumption in an emerging economy. *International Journal of Research in Marketing*, *30*, 4–18. http://dx.doi.org/10.1016/j.ijresmar.2011.12.001.

Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, *55*, 184–194. http://dx.doi.org/10.1016/j.energy.2013.03.086.

Kennedy, C. A., Ramaswami, A., Carney, S., & Dhakal, S. (2011). Greenhouse gas emission baselines for global cities and metropolitan regions. In *Cities and climate change* (pp. 15–54). The World Bank, http://dx.doi.org/10.1596/9780821384930_CH02, URL: http://elibrary.worldbank.org/doi/abs/10.1596/9780821384930{_}CH02.

Lanau, M., Herbert, L., & Liu, G. (2021). Extending urban stocks and flows analysis to urban greenhouse gas emission accounting: a case of odense, denmark. *Journal of Industrial Ecology*, *25*(4), 961–978. http://dx.doi.org/10.1111/jiec.13110.

Lévay, P. Z., Vanhille, J., Goedemé, T., & Verbist, G. (2021). The association between the carbon footprint and the socio-economic characteristics of Belgian households. *Ecological Economics*, *186*, Article 107065. http://dx.doi.org/10.1016/j.ecolecon.2021.107065.

Li, Z., Wang, F., Kang, T., Wang, C., Chen, X., Miao, Z., et al. (2022). Exploring differentiated impacts of socioeconomic factors and urban forms on city-level $CO_2$ emissions in China: spatial heterogeneity and varying importance levels. *Sustainable Cities and Society*, *84*, Article 104028. http://dx.doi.org/10.1016/j.scs.2022.104028.

Lin, T., Yu, Y., Bai, X., Feng, L., & Wang, J. (2013). Greenhouse gas emissions accounting of urban residential consumption: a household survey based approach. *PLoS One*, *8*(2), e5564. http://dx.doi.org/10.1371/journal.pone.0055642.

Linderhof, V., Kooreman, P., Allers, M., & Wiersma, D. (2001). Weight-based pricing in the collection of household waste: the oostzaan case. *Resource and Energy Economic*, (23), 359–371.

Marvuglia, A., Leuenberger, M., Kanevski, M., & Benetto, E. (2015). Random forest for toxicity of chemical emissions: features selection and uncertainty quantification. *Journal of Environmental Accounting and Management*, *3*(3), 229–241.

Menyah, K., & Wolde-Rufael, Y. (2010). CO2 emissions, nuclear energy, renewable energy and economic growth in the US. *Energy Policy*, *38*(6), 2911–2915. http://dx.doi.org/10.1016/j.enpol.2010.01.024.

Muraca, B. (2012). Towards a fair degrowth-society: justice and the right to a 'good life' beyond growth. *Futures*, *44*(6), 535–545.

Notarnicola, B., Tassielli, G., Renzulli, P. A., Castellani, V., & Sala, S. (2022). Environmental impacts of food consumption in Europe. *Journal of Cleaner Production*, *140*, 753–765. http://dx.doi.org/10.1016/j.jclepro.2016.06.080.

Noufal, M., Yuanyuan, L., Maalla, Z., & Adipah, S. (2020). Determinants of household solid waste generation and composition in homs City, Syria. *Journal of Environmental and Public Health*, Article 7460356. http://dx.doi.org/10.1155/2020/7460356.

Ölander, F., & Thøgersen, J. (2014). Informing versus nudging in environmental policy. *Journal of Consumer Policy*, *37*(3), 341–356.

O'Rourke, D., & Lollo, N. (2015). Transforming consumption: from decoupling, to behavior change, to system changes for sustainable consumption. *Annual Review of Environment and Resources*, *40*, 233–259.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830, URL: http://jmlr.org/papers/v12/pedregosa11a.html.

Peters, G. P., Marland, G., Le Quéré, C., Boden, T., Canadell, J. G., & Raupach, M. R. (2012). Rapid growth in $CO_2$ emissions after the 2008-2009 global financial crisis. *Nature Climate Change*, *2*(1), 2–4. http://dx.doi.org/10.1038/nclimate1332.

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. http://dx.doi.org/10.1016/j.enpol.2016.06.002.

Rajagopal, S. (2011). Customer data clustering using data mining technique. *International Journal of Database Management Systems*, (4), 1–11, URL: https://arxiv.org/ftp/arxiv/papers/1112/1112.2663.pdf.

Reynaud, A. (2015). *Modelling household water demand in Europe insights from a cross-country econometric analysis of EU-28 countries*: *Technical report EUR 27310 EN*, (p. 249). European Union, Joint Research Centre, Institute for Environment and Sustainability, http://dx.doi.org/10.2788/95638.

Rosenzweig, C., Solecki, W., Hammer, S. A., & Mehrotra, S. (2010). Cities lead the way in climate-change action. *Nature*, *467*(7318), 909–911. http://dx.doi.org/10.1038/467909a.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Sustainable Cities and Society*, *20*, 53–65. http://dx.doi.org/10.1016/0377-0427(87)90125-7.

Sala, S., Beylot, A., Corrado, S., Crenna, E., Sanyé-Mengual, E., & Secchi, M. (2019). *Indicators and assessment of the environmental impact of EU consumption. Consumption and consumer footprints for assessing and monitoring EU policies with life cycle assessment*: *Technical report*, (p. 55). Ispra, Italy: JRC, http://dx.doi.org/10.2760/403263.

Sala, S., & Castellani, V. (2019). The consumer footprint: monitoring sustainable development goal 12 with process-based life cycle assessment. *Journal of Cleaner Production*, *240*, Article 118050. http://dx.doi.org/10.1016/j.jclepro.2019.118050.

Sala, S., De Laurentiis, V., Barbero Vignola, G., Marelli, L., & Sanyé Mengual, E. (2022). *The consumer footprint calculator. Estimating the environmental impacts of the consumption of EU citizens and their lifestyle*: *Technical report*, Ispra (VA), Italy: European Commission - Joint Research Center, http://dx.doi.org/10.2760/597000, JRC129382.

Soderholm, P. (2015). Environmental policy and household behaviour. sustainability and everyday life. (p. 272). CRC Press, Taylor & Francis.

Song, R. (2019). *Machine learning for addressing data deficiencies in life cycle assessment* (PhD dissertation), Santa Barbara: University of California, URL: https://escholarship.org/uc/item/2vc7t19w.

Sonigo, P., Bain, J., Kong, M. A., Fedrigo, D., et al. (2012). *Policies to encourage sustainable consumption*: *Technical report*, European Commission.

Steen-Olsen, K., Wood, R., & Hertwich, E. G. (2016). The carbon footprint of norwegian household consumption 1999–2012. *Journal of Industrial Ecology*, *20*(3), 582–592. http://dx.doi.org/10.1111/jiec.12405.

Stevens, F. R., Gaughan, A., Linard, C., & Tatem, A. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, *10*, 1932–6203. http://dx.doi.org/10.1371/journal.pone.0107042.

Suthar, S., & Singh, P. (2015). Household solid waste generation and composition in different family size and socio-economic groups: A case study. *Sustainable Cities and Society*, *14*(1), 56–63. http://dx.doi.org/10.1016/j.scs.2014.07.004.

Tang, W., Wang, H., Lee, X.-L., & Yang, H.-T. (2022). Machine learning approach to uncovering residential energy consumption patterns based on socioeconomic and smart meter data. *Energy*, *240*, Article 122500. http://dx.doi.org/10.1016/j.energy.2021.122500.

Temme, E. H., Toxopeus, I. B., Kramer, G. F., Brosens, M. C., Drijvers, J. M., Tyszler, M., et al. (2015). Greenhouse gas emission of diets in the netherlands and associations with food, energy and macronutrient intakes. *Public Health Nutrition*, *18*(13), 2433–2445. http://dx.doi.org/10.1017/S1368980014002821.

Tsuda, K., Hara, K., & Uwasu, M. (2013). Prospects and challenges for disseminating life cycle thinking towards environmental conscious behaviors in daily lives. *Sustainability*, *5*(1), 123–135.

United Nations (2018). *Classification of individual consumption according to purpose (COICOP) 2018*: *Technical report 99*, United Nations, Department of Economic and Social Affairs, Statistics division.

Vinholes, D. B., Melo, I. M., MacHado, C. A., De Castro Chaves, H., Fuchs, F. D., & Fuchs, S. C. (2012). The association between socioeconomic characteristics and consumption of food items among Brazilian industry workers. *The Scientific World Journal*, *2012*, http://dx.doi.org/10.1100/2012/808245.

Vita, G., Lundström, J. R., Hertwich, E. G., Quist, J., Ivanova, D., Stadler, K., et al. (2019). The environmental impact of green consumption and sufficiency lifestyles scenarios in europe: connecting local sustainability visions to global consequences. *Ecological Economics*, *164*, Article 106322. http://dx.doi.org/10.1016/j.ecolecon.2019.05.002.

Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E., & Weidema, B. (2016). The ecoinvent database version 3 (part i): overview and methodology. *International Journal of Life Cycle Assessment*, *21*(9), 1218–1230. http://dx.doi.org/10.1007/s11367-016-1087-8.

Wu, W., Kanamori, Y., Zhang, R., Zhou, Q., Takahashi, K., & Masui, T. (2021). Implications of declining household economies of scale on electricity consumption and sustainability in China. *Ecological Economics*, *184*, Article 106981. http://dx.doi.org/10.1016/j.ecolecon.2021.106981.