

# Statistical Data-Driven Regression Method for Urban Electricity Demand Modelling

Nina Voulis, Martijn Warnier, and Frances M.T. Brazier

Faculty of Technology, Policy and Management, Delft University of Technology  
{n.voulis,m.e.warnier,f.m.brazier}@tudelft.nl

## Abstract

As the focus of the energy transition within cities worldwide moves towards local communities and neighbourhoods, the need for insights in the dynamics of local electricity demand increases. Detailed local electricity demand information is, however, often not available. This paper proposes a statistical data-driven method to model local electricity demand for mixed urban areas, using a combination of other openly available datasets. Such datasets however are mutually incompatible without further conversion. The proposed method overcomes this problem. Linear regression is used to combine these different datasets, whereby the regression coefficients have the meaning of *scaling factors* for different types of electricity consumers (households, offices, shops, etc.). The method is calibrated and validated using respectively a training and a test dataset of Dutch municipalities, yielding R-squared values for most consumer types between 61% and 98%. The application of the method for local electricity demand modelling is illustrated for three Dutch municipalities with different consumer compositions.

**Keywords**— urban demand, linear regression, energy transition, renewables, data-driven model

## 1 Introduction

Understanding the dynamics of local electricity demand is important to support the transition to renewable, distributed generation in urban areas [1]. For this purpose, a number of knowledge gaps needs to be resolved, including the current lack of electricity demand data and modelling tools for the local level [2, 3]. Existing load forecasting approaches typically cover large areas (*e.g.*, the entire territory serviced by a single utility company) and are therefore too coarse to provide insights in spatial demand variations at the local scale [4].

New urban and building energy models are currently under development, however their application to real areas is restrained by the limited amount of detailed openly available energy demand data [5, 6, 7, 8]. While residential energy demand data are available to some extent, service sector and industrial data are often not (openly) available at all. As a result, service and industrial sectors are not included in most modelling studies, although real urban areas are a mix of residential, service and industrial consumers. The few studies which include mixed electricity consumers, are based on proprietary data (*e.g.*, [9]).

This paper proposes a statistical data-driven method that can be used to estimate *local* urban demand profiles for *mixed* residential and services areas<sup>1</sup>, for which currently no demand profiles are available. The method relies on a combination of energy and non-energy related data which are generally publicly available, and is implemented for the Netherlands.

---

<sup>1</sup>Industrial consumers are left out of scope as they require a case-by-case instead of a statistical approach due to their large size.

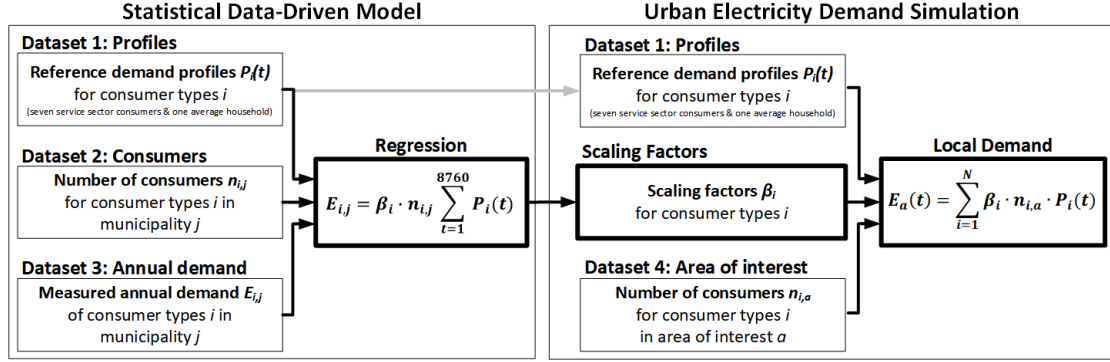


Figure 1: Flow chart for the statistical data-driven model. The model combines three data sources and generates scaling factors  $\beta_i$ . These scaling factors can be used to simulate electricity demand profiles of any urban area of choice for which the local number of consumers of different types is available.

## 2 Methods

The proposed method combines three datasets, each providing details in a different dimension. The first dataset contains *reference demand profiles* for different consumer types, which provide detailed *temporal* information. The second dataset contains *local composition* data for different consumer types, providing detailed *spatial* information. The third dataset contains *aggregated local annual electricity demand data*, which is used to link the two other datasets.

### 2.1 Datasets

The data and sources used in the case study are outlined next. Numeric values for the reference year 2014 are used.

- **Dataset 1: Reference demand profiles.** Both residential and service sector consumers are modelled. For the residential sector, the reference Dutch household demand profile is used [10]. For the service sector, adapted profiles from the United States Department of Energy (U.S. DOE) commercial building reference models are used [11], as the U.S. DOE provides one of the most complete, openly available datasets of energy demand in service sector buildings. The U.S. DOE profiles can be adapted for other regions, in this case for the Netherlands, as described in [12].
- **Dataset 2: Local consumer composition.** The data format of the local consumer composition is the number of residential and different service sector consumers in the area of interest. For the Netherlands, data on municipality-level consumer composition are available from Statistics Netherlands [13].
- **Dataset 3: Annual local electricity demand.** Annual electricity demand data for Dutch municipalities are used. For each municipality, the data for different types of service sector and residential consumers are available from the Dutch Ministry of Infrastructure and Water Management [14].

## 2.2 Statistical Data-Driven Model

The different datasets that are available come from different sources and therefore refer to different “bases”. This is in particular the case for the datasets used in this paper. The reference demand profiles (Dataset 1) pertain to reference buildings (data expressed in kWh/h per building) [10, 11, 12]. The consumer composition data (Dataset 2) pertain to households, businesses and services (collectively termed “consumers”) registered with the Dutch government (data expressed in number of administrative entities in a given municipality) [13]. The annual local electricity demand (Dataset 3) pertains to electricity demand by economic subsectors (*e.g.*, healthcare) in the Netherlands (data expressed in MWh/year per subsector) [14].

This paper proposes a data-driven method which (1) overcomes such base differences, and (2) enables modelling of electricity demand in areas for which only limited data are available. The method relies on linear regression, whereby the *regression coefficients* for each electricity consumer have the meaning of *scaling factors*, thus making the combination of available datasets possible, despite their base differences. The method is shown in Fig. 1 (left block). The regression is given by:

$$E_{i,j} = \beta_i \cdot n_{i,j} * \sum_{t=1}^{8760} P_i(t) \quad (1)$$

where,

$E_{i,j}$  : Annual electricity demand of consumer type  $i$  in area  $j$   
(Dataset 3)

$\beta_i$  : Scaling factor for consumer type  $i$

$n_{i,j}$  : Number of consumers of type  $i$  in area  $j$  (Dataset 2)

$P_i(t)$  : Hourly reference demand profile for consumer type  $i$   
(summed over 8760 hours for one year) (Dataset 1)

## 2.3 Calibration and Validation

The linear regression model is calibrated and validated using data for 383 municipalities in the Netherlands (Datasets 2 and 3). Municipalities are randomly assigned to a training and a test dataset, such that the training dataset contains 268 (70%), and the test dataset 115 (30%) municipalities. The coefficients  $\beta_i$  are estimated based on the training dataset, their predictive power is validated using the test dataset. The R-squared is used as fitness metric.

## 2.4 Urban Electricity Demand Simulation

Once the scaling factors are calculated and validated, they can be used to model electricity demand in any area of interest, if data with the same base as used to calculate the scaling factors is available for that area. In this paper, the base is the number of households, businesses and services (termed “consumers”) registered with the government. Such data is generally available for many areas from local governments or open databases such as [15]. Using scaling factors and

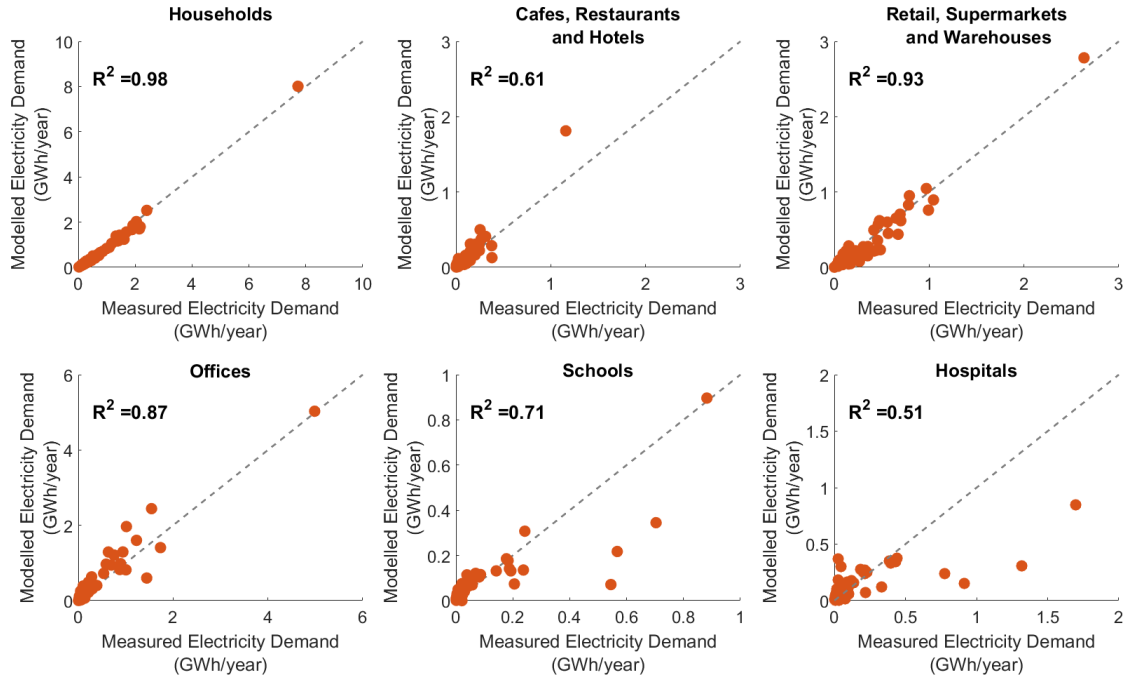


Figure 2: Scatter plots for six consumer types (or categories) for 115 Dutch municipalities (test dataset). Each plot shows how well the modelled annual electricity demand matches the measured electricity demand.  $R^2$  is used as fitness metric, and reported for each consumer type (or category).

the local number of consumers, the local electricity demand for area  $a$  is given by the following equation. (This step is also shown in Fig. 1 in the right block.)

$$E_a(t) = \sum_{i=1}^N \beta_i \cdot n_{i,a} \cdot P_i(t) \quad (2)$$

where,

$E_a(t)$  : Energy demand profile for area  $a$  with resolution  $t$

$N$  : Total number of different consumer types

$n_{i,a}$  : Number of consumers of type  $i$  in area  $a$

In this paper, three municipalities from the test dataset are used as areas of interest to illustrate the application of the proposed method. For these municipalities local demand profiles and local relative annual consumer demand are modelled.

### 3 Results and Discussion

Results of the linear regression method yielding scaling factors are shown and discussed first, followed by results of electricity demand simulation for three municipalities which are used to illustrate the application of the proposed method.

Table 1: Linear regression scaling coefficients and R<sup>2</sup>-values.

Consumer Type	Scaling Factor	R <sup>2</sup> Training Set	R <sup>2</sup> Test Set
Households	0.74	0.97	0.98
Cafes & Restaurants	0.25	0.81	0.61
Hotels	0.002		
Retail & Supermarkets	0.06	0.87	0.93
Warehouses	0.05		
Offices	0.10	0.91	0.87
Schools	0.12	0.76	0.71
Hospitals	0.01	0.43	0.51
Total	(-)	0.97	0.99

### 3.1 Statistical Data-Driven Model

The scaling factors  $\beta_i$  for different consumer types obtained using the proposed linear regression method are summarised in Table 1, alongside with R<sup>2</sup>-values based on municipality-level training and test datasets. Figure 2 shows the scatter plots for each consumer type for the test dataset. (Note that due to differences in consumer classification in different datasets, some consumer types are grouped into consumer “categories”, *e.g.*, consumer types “Cafes and Restaurants”, and “Hotels” are grouped into “Cafes, Restaurants and Hotels”.)

The scaling factors in Table 1 are all less than 1 due to base differences between the datasets used. The dataset used for the number of consumers in each area has administrative registrations as base (Dataset 2), while the reference demand profiles have reference buildings as base (Dataset 1). A single building can house multiple administrative units, for instance, a single office building can house 10 different companies, yielding a scaling factor of 0.1. Furthermore, a relatively large reference building can be used to simulate electricity use in smaller administrative units. For instance, the electricity consumer type “Hotels”<sup>2</sup> includes small lodging rooms and bed-and-breakfasts, yielding a very low scaling factor. Similarly, for instance, “Schools” include all administrative entities which provide education (schools, universities, but also small education and training centres), thus also yielding a relatively small scaling factor.

R<sup>2</sup>-values (Table 1 and Fig. 2) represent the share of variability explained by the regression model. Most obtained R<sup>2</sup>-values vary between 61% and 98%, with lower-end values for broader and more diverse categories (*e.g.*, cafes, restaurants and hotels). Hospitals have a relatively low R<sup>2</sup>-value of 51% (for the test set). Hospital electricity demand is known to be challenging to model [16].

Obtained R<sup>2</sup>-values are compared to values in literature. Fonseca and Schlueter report building-level electricity model errors of 4% to 66%, and area-level electricity model errors of 1% to 19% [9]. Mastrucci *et al.* used linear regression to downscale electricity demand from postcode-level to building level, obtaining an R<sup>2</sup>-value of 81.7% [17]. The method proposed in this paper thus yields a regression model with a predicted variability comparable to, or higher than, that of other energy models in literature.

<sup>2</sup>The names of the consumer types refer to the reference buildings that are used to model the electricity demand of the given consumer type [11].

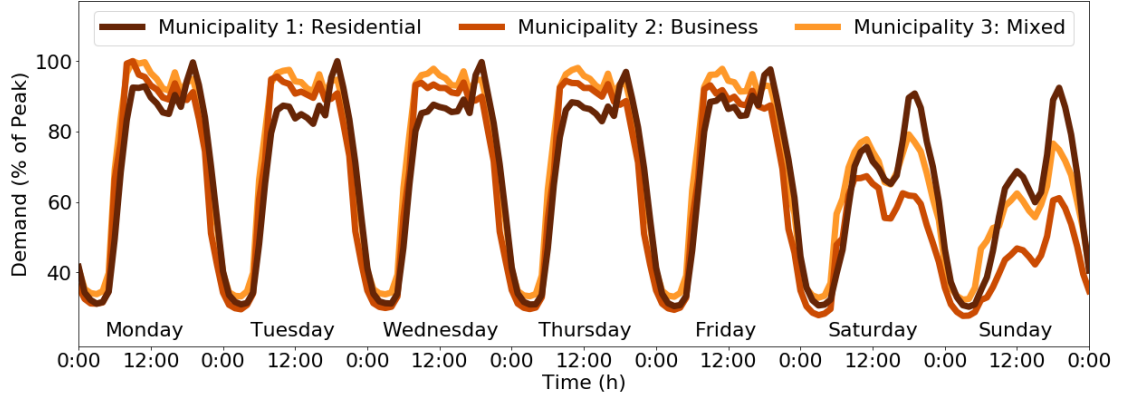


Figure 3: Modelled electricity demand profiles for three municipalities from the test dataset for one week (February 24<sup>th</sup> until March 2<sup>nd</sup>, 2014).

### 3.2 Urban Electricity Demand Simulation

Detailed local electricity demand of mixed areas can be modelled using the scaling factors calculated and validated in the statistical data-driven model. Three Dutch municipalities from the test dataset are used as illustration. Municipality 1 (Doesburg) is predominantly residential, Municipality 2 (Nieuwegein) has a large number of offices, and Municipality 3 (Texel) is mixed, with a relatively high number of cafes, restaurants and hotels.

Table 2 shows the unscaled number of consumers of different types in the three municipalities. This is an example of Dataset 4 in Fig. 1. Note for instance the high number of the “Hotel” type consumers in Municipality 3. This municipality is a popular holiday area on one of the Wadden Sea islands, and has a high number of small-scale holiday rentals. The high number of “Hotels” refers to these small-scale holiday rentals, and therefore is scaled (by a factor 0.002, see Table 1) to represent the real consumer mix and scale. The scaled number of consumers for the three municipalities is shown in Table 3. Note that after scaling, the number of some consumer types is less than 1, in particular “Hotels” and “Hospitals”. This means that the electricity demand of respectively lodging and healthcare in the given municipality can be modelled by a fraction of the electricity demand of the reference building available in [11] and adapted for the Netherlands as described in [12].

Local electricity demand profiles can be modelled using the scaled consumer number data from Table 3 and hourly reference building electricity demand profiles from [11] (adapted for the Netherlands as described in [12]), implementing the approach shown in Fig. 1. The resulting profiles are shown in Fig. 3 for one week (February 24<sup>th</sup> until March 2<sup>nd</sup>, 2014). Note that Municipality 1 as a whole has an electricity demand profile which resembles an average household profile, with an evening demand peak. Municipalities 2 and 3 have relatively flat demand profiles during the day on weekdays, and a relatively low, double-peaked profile during the weekend. The double peak corresponds to lunch and dinner times, with increased demand due to cooking and eating activities both in the households and in cafes, restaurants and hotels.

### 3.3 Model Application for Local Energy Transition

Improved insights in local electricity demand profiles can help integrate distributed renewable resources in local power systems at the distribution level. The synergies and differences in timing between electricity demand on the one hand and solar and wind generation on the other hand,

Table 2: Unscaled number of consumers of different types in three reference municipalities from the test data set.

<b>Consumer Type</b>	<b>Municipality 1: Residential</b>	<b>Municipality 2: Business</b>	<b>Municipality 3: Mixed</b>
Households	5167	27507	6405
Cafes & Restaurants	17	104	197
Hotels	0	3	3972
Retail & Supermarkets	14	445	222
Warehouses	272	1189	562
Offices	14	712	99
Schools	1	85	15
Hospitals	3	43	26

Table 3: Scaled number of consumers of different types in three municipalities from the test data set.

<b>Consumer Type</b>	<b>Municipality 1: Residential</b>	<b>Municipality 2: Business</b>	<b>Municipality 3: Mixed</b>
Households	3800	20231	4711
Cafes & Restaurants	4	26	49
Hotels	0	0.006	8
Retail & Supermarkets	1	28	14
Warehouses	13	55	26
Offices	1	70	10
Schools	0.1	10	2
Hospitals	0.03	0.5	0.3

determine which measures (*e.g.*, storage, interconnection, demand response) are necessary to support local distributed renewable resource integration. For instance, the demand profiles of Municipalities 2 and 3 have a plateau during the day, and thus have a better correspondence with day-peaking solar generation than Municipality 1, for which demand peaks during the evening.

Improved understanding of the breakdown of electricity demand across consumer types can help design targeted energy efficiency and demand response programmes. Fig. 4 shows the relative annual electricity demand of different consumer types in the three modelled municipalities. In Municipality 1, 67% of the modelled annual electricity demand is used by households. In Municipalities 2 and 3 a much smaller share, respectively 34% and 27%, of the modelled annual electricity demand is used by households. In Municipality 2, offices use the most electricity (45%). In Municipality 3, cafes, restaurants and hotels are the largest consumer group (27%). These results indicate that in the three municipalities different consumer groups should be targeted for local demand response and energy efficiency campaigns to achieve the highest impact.

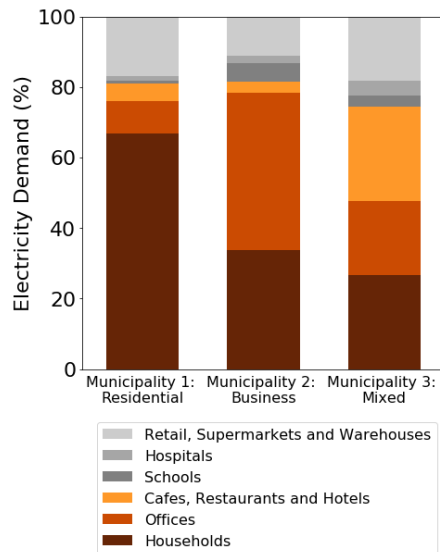


Figure 4: Modelled relative annual electricity demand by consumer type in three municipalities from the test dataset.

## 4 Conclusion

This paper proposes and implements a statistical data-driven method that can be used to model *local* electricity demand in *mixed* residential and service sector areas for which currently no detailed spatio-temporal demand data are available. The proposed method is based on linear regression, whereby the *regression coefficients* have the meaning of *scaling factors*, thus making the combination of openly available datasets possible, despite their base differences.

The linear regression model is calibrated and validated using data of Dutch municipalities. The results yield  $R^2$ -values for scaling factors between 61% to 98% for all consumer types except for hospitals, showing the validity of the method for most consumers types.

Next, the use of scaling factors for modelling of local electricity demand is illustrated for three municipalities with different consumer compositions. The resulting detailed local electricity demand profiles and relative local electricity demand breakdown by consumer type provide insights which can help and guide local distributed renewable resource integration and energy efficiency initiatives. This application shows how the proposed approach can be used as a tool by researchers and local governments for local energy transition projects.

## References

- [1] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi, “Spatial distribution of urban building energy consumption by end use,” *Energy & Buildings*, vol. 45, pp. 141–151, 2012.
- [2] M. R. Pasimemi, I. Petrosillo, R. Aretano, T. Semeraro, A. De Marco, N. Zaccarelli, and G. Zurlini, “Scales , strategies and actions for effective energy planning: A review,” *Energy Policy*, vol. 65, pp. 165–174, 2014.



- [3] A. Mirakyan and R. D. Guio, “Integrated energy planning in cities and territories: A review of methods and tools,” *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 289–297, 2013.
- [4] S. Heiple and D. J. Sailor, “Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles,” *Energy and Buildings*, vol. 40, pp. 1426–1436, 2008.
- [5] J. Allegrini, K. Orehoung, G. Mavromatidis, F. Ruesch, V. Dorer, and R. Evins, “A review of modelling approaches and tools for the simulation of district-scale energy systems,” *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1391–1404, 2015.
- [6] S. Cajot, M. Peter, J.-M. Bahu, A. Koch, and F. Maréchal, “Energy planning in the urban context: challenges and perspectives,” *Energy Procedia*, vol. 78, pp. 3366–3371, 2015.
- [7] H.-X. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 3586–3592, 2012.
- [8] S. Pfenninger, J. DeCarolis, L. Hirth, S. Quoilin, I. Staffel, and I. Sta, “The importance of open data and software: Is energy research lagging behind?” *Energy Policy*, vol. 101, pp. 211–215, 2017.
- [9] J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Applied Energy*, vol. 142, pp. 247–265, 2015.
- [10] NEDU, “Verbruiksprofielen,” <http://nedu.nl/portfolio/verbruiksprofielen>. Last accessed online 22/12/2017.
- [11] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg, M. Yazdanian, J. Huang, and D. Crawley, “U.S. Department of Energy commercial reference building models of the national building stock,” National Renewable Energy Laboratory, Tech. Rep., 2011. [Online]. Available: <http://digitalscholarship.unlv.edu/renew{%}7B{-}{%}7Dpubs/44>
- [12] N. Voulis, M. Warnier, and F. M. T. Brazier, “Impact of service sector loads on renewable resource integration,” *Applied Energy*, vol. 205, pp. 1311–1326, 2017.
- [13] Centraal Bureau voor de Statistiek, “Kerncijfers wijken en buurten,” <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLnl&PA=83220NED&LA=nl>. Last accessed online 22/12/2017.
- [14] Rijkswaterstaat, “Klimaatmonitor,” [klimaatmonitor.databank.nl](http://klimaatmonitor.databank.nl). Last accessed online 22/12/2017.
- [15] OpenStreetMap Foundation, “OpenStreetMap,” <http://www.openstreetmap.org>. Last accessed online 22/12/2017.
- [16] Planbureau voor de Leefomgeving, “Vesta ruimtelijk energiemodel voor de gebouwde omgeving - Data en methoden,” Tech. Rep., 2012.
- [17] A. Mastrucci, O. Baume, F. Stazi, and U. Leopold, “Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam,” *Energy and Buildings*, vol. 75, pp. 358–367, 2014.