Population Estimation Using a 3D City Model: A Multi-Scale Country-Wide Study in the Netherlands

Filip Biljecki*, Ken Arroyo Ohori, Hugo Ledoux, Ravi Peters, Jantien Stoter

3D Geoinformation, Delft University of Technology, Delft, The Netherlands

* f.biljecki@tudelft.nl

Abstract

The remote estimation of a region's population has for decades been a key application of geographic information science in demography. Most studies have used 2D data (maps, satellite imagery) to estimate population avoiding field surveys and questionnaires. As the availability of semantic 3D city models is constantly increasing, we investigate to what extent they can be used for the same purpose. Based on the assumption that housing space is a proxy for the number of its residents, we use two methods to estimate the population with 3D city models in two directions: (1) disaggregation (areal interpolation) to estimate the population of small administrative entities (e.g. neighbourhoods) from that of larger ones (e.g. municipalities); and (2) a statistical modelling approach to estimate the population of large entities from a sample composed of their smaller ones (e.g. one acquired by a government register). Starting from a complete Dutch census dataset at the neighbourhood level and a 3D model of all 9.9 million buildings in the Netherlands, we compare the population estimates obtained by both methods with the actual population as reported in the census, and use it to evaluate the quality that can be achieved by estimations at different administrative levels. We also analyse how the volume-based estimation enabled by 3D city models fares in comparison to 2D methods using building footprints and floor areas, as well as how it is affected by different levels of semantic detail in a 3D city model. We conclude that 3D city models are useful for estimations of large areas (e.g. for a country), and that the 3D approach has clear advantages over the 2D approach.

Introduction

Geographic information science (GIS) and demography have long been closely related, and GIS techniques are ubiquitous in mapping, analysing, and filling gaps in demographic data. In particular, geostatistical techniques are often used to estimate a region's population in the absence of reliable or complete census data [1, 2].

2D GIS datasets (e.g. satellite imagery and maps) have been used extensively in the past 50 years for this purpose, as several of them have been found to be reasonable proxies for

population [2–21]. For instance, Bakillah et al. [22] and Doll et al. [23] estimate the population based on the concentration of surrounding points of interest (e.g. restaurants); Anderson et al. [24] and Sutton [25] use night-time imagery following the hypothesis that city lights indicate the magnitude of the urban extent, which in turn indicates the population. Pozzi and Small [26] infer the population density from a vegetation cover map, based on the idea that less vegetation means more people; Xie [27] finds the relation between the density of road network and population; Steiger et al. [28] analyse georeferenced Twitter data to locate clusters indicating home- and work-related social activities that can serve as a proxy to estimate the residential and workplace population census data; and Lwin et al. [29] do a similar work using geolocated mobile phone usage data.

Among all these methods, many successful approaches rely on 2D datasets (maps) containing *building footprints* (e.g. derived from cadastral records or satellite imagery). The simplest approaches rely on the total number of buildings in a region or the total area of building footprints in it [30–32]. These methods perform reasonably well in homogeneous areas, but they exhibit significant errors in areas where buildings have a great variation in the number of storeys.

With the advancement of remote sensing technologies, such as lidar and aerial photogrammetry [33–37], it is now possible to automatically and remotely measure the height of a building, which can be used to obtain a volumetric representation of a building (3D city model) that is useful for population estimates. In fact, several researchers have indicated that the volume of buildings and the floorspace provide a strong cue for its population [5, 15, 22, 30, 38–46]. For example, Lu et al. [39] use multiple regression models to perform a study in Denver, Colorado, based on both footprint areas and building volumes. Lwin and Murayama [44] and Alahmadi et al. [46, 47] estimate the number of floors from an elevation dataset, and multiply it with the footprint area to get the approximate internal area of the apartments. Their results indicate that the volume-based approach gives more accurate results than the area of the footprints due to heterogeneous building morphologies.

However, despite the frequent indication that volume-based methods can improve on the estimates of area-based methods, there has been no large-scale study that conclusively proves that this is true. Existing studies have several gaps: they usually focus on single metropolitan areas, which can be relatively homogeneous; they seldom compare the accuracy of different approaches within the same region; they derive a building's volume based on a raster dataset, which limits its accuracy; they do not consider how this approach scales between larger and smaller areas; and they do not consider how the level of detail of the used volumetric representation affects the accuracy of the result.

The goal of this paper is to bridge these gaps. We investigate *to what extent 3D city models can be used to estimate the population of a region* by performing a multi-scale country-wide study in the Netherlands. As the Dutch government provides both highly accurate census and building data, we consider that the Netherlands serves as an excellent case study, both for the experiments and the validation of the methods.

We therefore evaluate the use of 3D city models in population estimation in two directions: (1) disaggregation (areal interpolation) to estimate the population of small administrative entities (e.g. neighbourhoods) from that of larger ones (e.g. municipalities); and (2) a statistical modelling approach to estimate the population of large entities from a sample composed of their smaller ones (e.g. one acquired by a government register). We compare the population estimates obtained by both methods with the actual population as reported in the census, and use it to evaluate the quality that can be achieved by estimations at different administrative levels. We also analyse how the volume-based estimation enabled by 3D city models fares in comparison to 2D methods using building footprints and floor areas, as well as how it is affected by

different levels of semantic detail (information on building use) in a 3D city model. We conclude that 3D city models are useful for large scale estimations (e.g. for a country), and that the 3D approach has clear advantages over the 2D approach.

Materials and Methods

Census data

The Netherlands is decomposed hierarchically into 12 provinces (not covered further in this paper), 393 municipalities, 2816 districts, and 12237 neighbourhoods. The population of each of the entities is known thanks to the open dataset of Statistics Netherlands—CBS (*Centraal Bureau voor de Statistiek*). As shown in Fig 1, the dataset consists of sets of polygons representing statistical units—the population within each polygon is stored as an attribute. We use this dataset to validate our results, and its subset to train one of the methods. The properties of

Census neighbourhoods

Density [residents / sq. km]

	0 - 31
	31 - 151
	151 - 1089
	1089 - 3074
	3074 - 5596
	5596 - 28599

Buildings

Residential use [%]

	0 - 20
	20 - 40
	40 - 60
	60 - 80
	80 - 100

Data © Kadaster / Centraal Bureau voor de Statistiek, 2015

Fig 1. Datasets used in this research: census neighbourhoods with building footprints. (Left side:) The Netherlands divided into more than 12 thousand neighbourhoods; and (right side:) two zoomed-in urban areas, where building footprints are visible along with the information on their use (residential share). Note that the maps on the right side show large variations in population density despite neighbourhoods being similarly urbanised. The less populated areas have many non-residential buildings, e.g. industrial and university buildings, showing that information on their use is crucial, and it significantly impacts the quality of the population estimation. The population density classes are divided into quantiles.

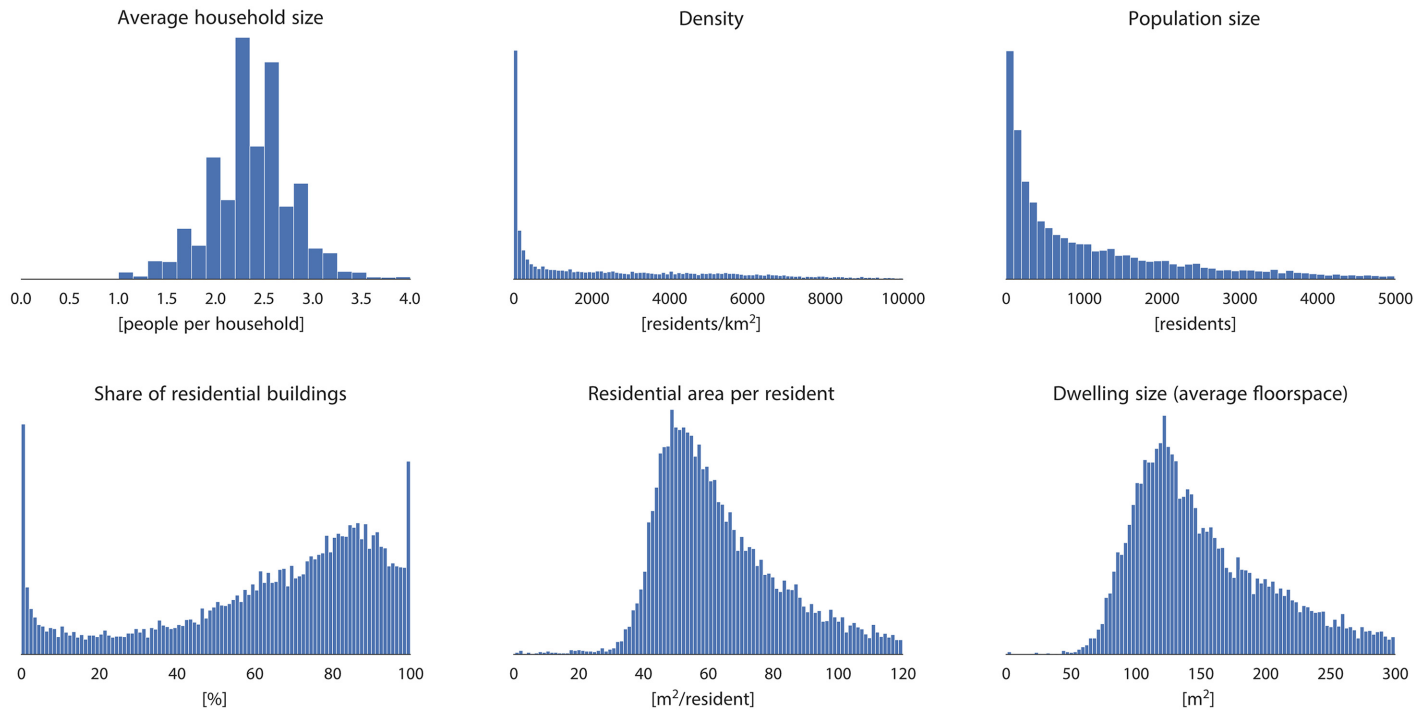doi:10.1371/journal.pone.0156808.g001

Fig 2. Census neighbourhoods statistics. The plots expose substantial housing differences among the neighbourhoods across the country. Derived from data (c) Kadaster / Centraal Bureau voor de Statistiek, 2015.

statistical units across the country vary (see Fig 2), covering widely heterogeneous household sizes, population densities, and dwelling sizes, among others.

3D city model of the Netherlands

3D city models are digital representations of the urban environment, focusing on buildings [48–50]. They are used for many different purposes [50], e.g. the prediction of noise pollution [51]. Their key advantage over 2D maps is that they provide volumetric data, which is beneficial for applications that take advantage of the height or volume of buildings, such as energy demand estimations [52, 53] and visibility analyses [54, 55]. Population estimation is clearly such a case, as high-rise residential buildings are very likely to contain more inhabitants per unit area than low-rise buildings.

3D city models can be created with many different techniques, e.g. from airborne laser scanning, and considerable work has been devoted to their automatic generation [56–58]. In this study, we generate a country-wide 3D city model by combining two open datasets from the Netherlands government: (i) building data from the national register of addresses and buildings (BAG—*Basisregistraties Adressen en Gebouwen*, which is collected and maintained by each municipality, and disseminated as country-wide dataset through the national portal of *Kadaster*, the national mapping agency of the Netherlands; and processed by the NLExtract project)—containing the base geometry, building use, and floorspace information (see Fig 1); and (ii) elevation data—the Height Model of the Netherlands (AHN—*Actueel Hoogtebestand Nederland*), which contains 639 billion elevation points covering the whole country (see S3 Fig for an illustration). The 3D model creation is done using a process called extrusion, where the building footprint is lifted to a certain height to obtain a simple volumetric model [59, 60], yielding so-called block models of buildings (LOD1 according to the CityGML standard [61, 62]). For this
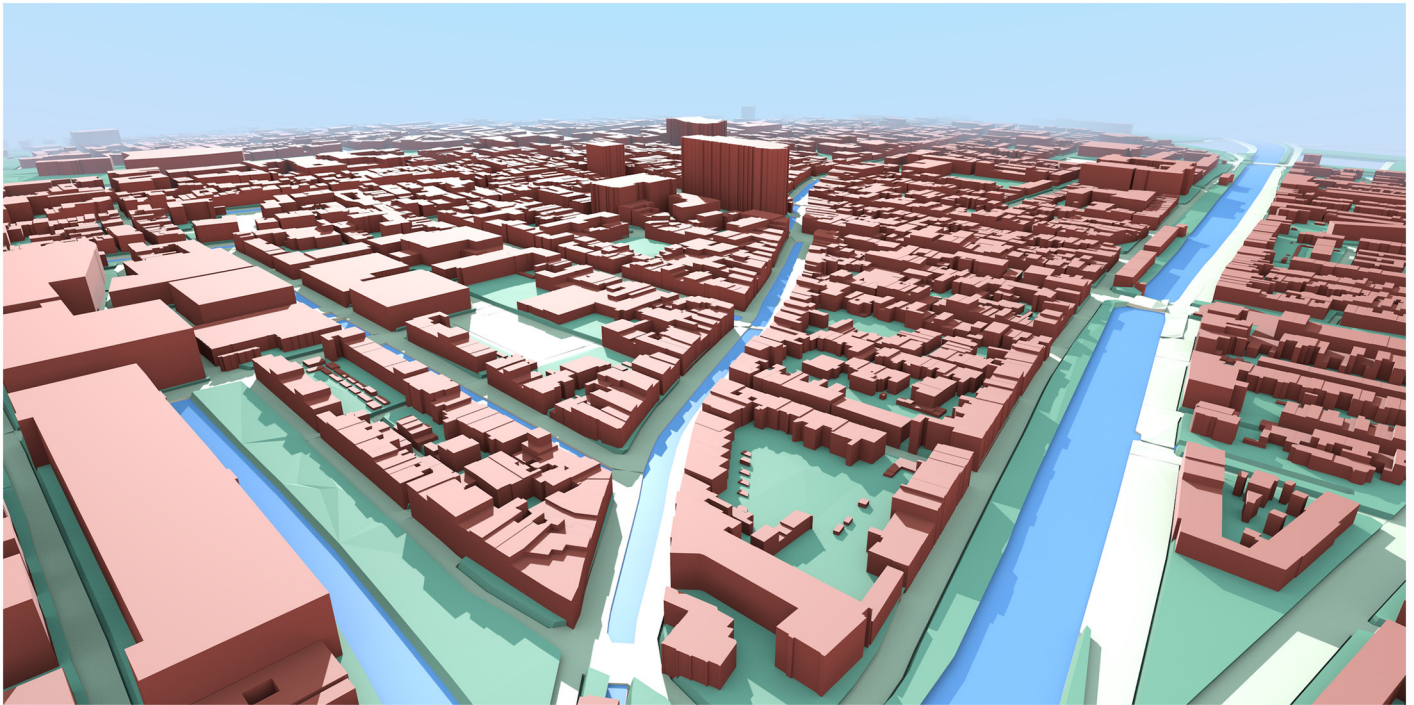
Fig 3. Example of the 3D city model. This example shows a part of the city of Delft, constructed from open data of the Government of the Netherlands ((c) Kadaster and (c) Actueel Hoogtebestand Nederland; see S3 Fig for the illustration of the elevation data).

purpose we have used the software 3dfier, developed by our group and released under an open-source licence (https://github.com/tudelft3d/3dfier). The software analyses all elevation points whose projection is within the footprint of a building, and determines the elevation at the building base and a single value for the height for the building. The height of the building is set to the median of all elevation points, which is considered optimal for building volume estimations [63]. A visual representation of these building block models is given in Fig 3.

3D city models come in different levels of detail (LODs) and with heterogeneous quality [62, 64, 65], both in terms of geometry and in terms of semantic information (e.g. a building's use) [66]. Thus, in order to test how different LODs of a 3D city model affect the population estimations, we construct 9 different LODs using various combinations of different levels of detail in a building's geometric and semantic information (Fig 4). In this way, we can directly compare the quality of the estimations given by the area-based (footprint and floor area) and volume-based approaches.

We consider three geometric LODs: (LOD0) 2D building footprints (the traditional area-based approach without height measurements); (LOD0+) building floorspace (area-based approach in which the vertical extent of the building is available); and (LOD1) volumetric 3D block models (from which the volume of a building can be calculated). For LOD0+, we rely on accurate indoor measurements from the Dutch cadastre, which is a dataset that is rarely available elsewhere. However, it should also be noted there is recent work focused on its automatic reconstruction [67, 68].

The general hypothesis used in this paper, and in related work, is simple: the larger the building, the more people reside in it; and the larger the living capacity of a district, the more populous it is. However, we argue that other building properties should be taken into account as well. The occupancy of a building also depends on its type, e.g. a cathedral, indoor arena, or
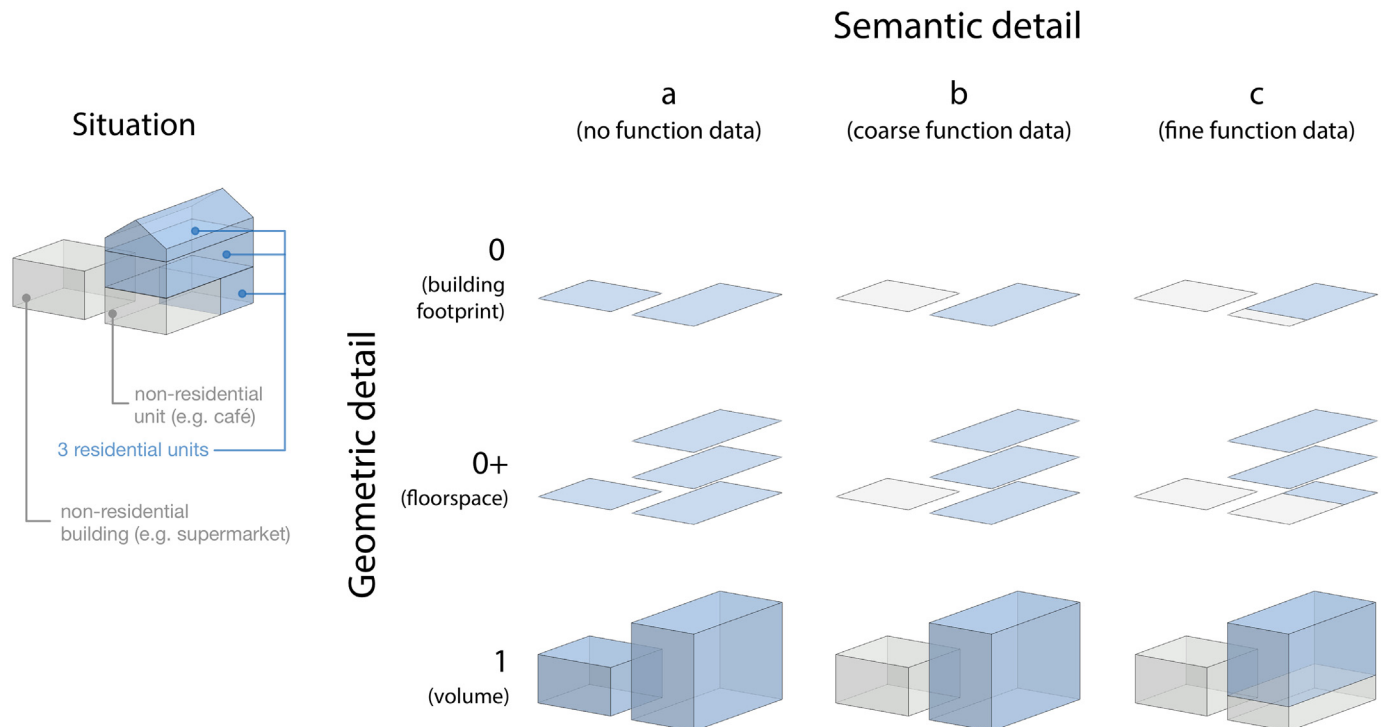
Fig 4. Multi-LOD data used for the experiments. Different granularities, which reflect the different grades of data available in practice. The blue space indicates residential space (proxy for population) as considered for each LOD, which differs depending on the geometry and semantics, and ultimately affects the performance of the methods. In our work we benchmark the performance of each grade of the data for the purpose of estimating the population.

a factory can be very large but at the same time they house zero inhabitants. Therefore, only residential buildings must be taken into account. This is further complicated by mixed-use buildings, which are composed of non-residential and of residential units, e.g. a three-storey building, where the ground floor is occupied by non-residential space (e.g. a restaurant and a shop), and the remaining two floors by residential units (fairly common in the Netherlands). However, such information is not always available, hence we pay special attention to the semantic aspect of data. Therefore, for the semantic part, we distinguish three levels of detail: (a) no data about the function of the building, and hence all buildings are treated equally; (b) a building is either residential or non-residential [42, 69]; and (c) fractional building use, where the share of the residential use within a building is known.

The possible combinations of the three geometric LODs and the three semantic LODs result in the 9 LODs used in this study, e.g. LOD1^{b} denotes a block model with the singular information on the building use.

Existing methods for population estimation

The estimation of population with GIS data and techniques has been extensively reviewed by numerous authors [22, 41, 70–72]. Generally two groups of methods are recognised [70], both of which are used in this paper (Fig 5):

1. Disaggregation (areal interpolation): this is a top-down approach where the population of a larger administrative unit or zone (e.g. region, municipality, census district) is distributed across smaller units (e.g. neighbourhood), usually by weighting it according to different
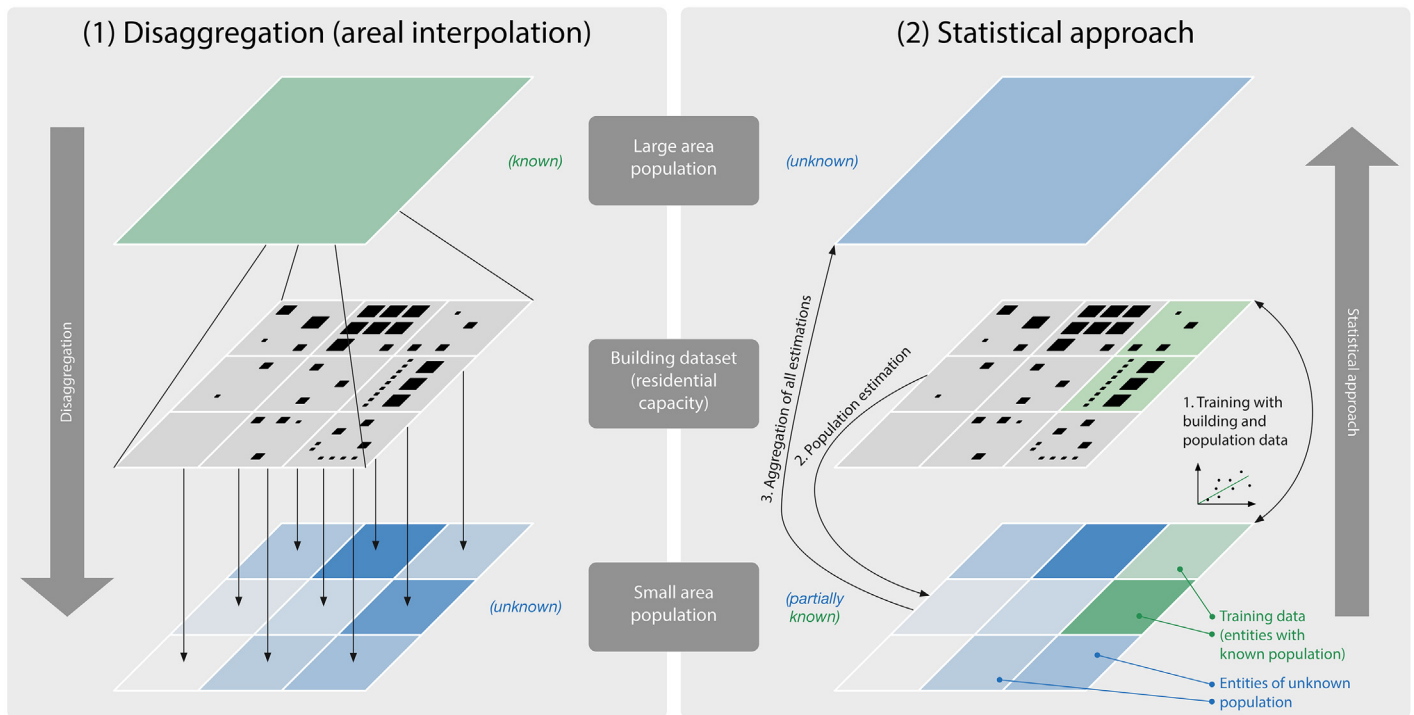
Fig 5. The two population estimation methods used. In this study we employ both methods, and for the residential capacity we use three different indicators in parallel: building footprint area, floorspace area, and building volume. Our work determines the usability of each of the type of geographic information for this purpose.

doi:10.1371/journal.pone.0156808.g005

factors which hint at the population [73–78]. This approach is typically used when the population of a large entity is known (e.g. a city), but the one of its composing entities is not known (e.g. its neighbourhoods).

The disaggregation can be done by simply distributing the population among administrative sub-zones, but it also can be aided by dasymetric mapping to shape smaller surfaces in such a way that variation within each surface is minimised [79]. This is especially useful when the smaller units are political subdivision of the larger (parent) unit often found in choropleth maps (e.g. interpolation from a province to the containing municipalities), because such regions may contain variations in the population density. Dasymetric mapping therefore results in (sub-)units that are more homogeneous [80–83].

2. Statistical modelling approach: first the relationships between population and socio-economic and morphological variables associated with the population density are inferred, e.g. land use [9, 84], proximity to transportation network [71], and distance from the central business district [85]. The deduced relationships are then applied to estimate the population count of unknown areas. In this approach multiple linear regression is most commonly used. The advantage of this bottom-up approach is that a sampling census has to be carried out for only a small area. It is useful in the scenario when only the population of a subset (e.g. a city) of a large area (e.g. a province) is known.

Our proposed method using 3D city models

For our population estimation study, we test three indicators to determine the disaggregation weights and the statistical relationships: (i) area of the 2D building footprints (in m^2), (ii) area

of the building floorspace (in m^2), and (iii) building volume (in m^3). Each of these is tested at three levels of semantic detail, resulting in the 9 aforementioned LODs of the input datasets.

In order to diminish residential and socio-economic variations across a large area, but also to test the performance of different estimation scenarios, we use multiple scales of estimations, as shown in Fig 6. In the disaggregation approach 6 scales are analysed:

D1 Disaggregation from the country level to its 12237 neighbourhoods.

D2 Disaggregation from each of the 393 municipalities to their 12237 neighbourhoods.

D3 Disaggregation from each of the 2816 districts to their 12237 neighbourhoods.

D4 Disaggregation from the country level to its 2816 districts.

D5 Disaggregation from each of the 393 municipalities to their 2816 districts.

D6 Disaggregation from the country level to its 393 municipalities.

On the statistical side we use a random subset of 10% of each statistical level to determine with ordinary least squares the relationships between building space and population, and apply them for three different experiments:

S1 Estimation of the population of the test neighbourhoods (i.e. the remaining 90%).

S2 Estimation of the population of the test districts (i.e. the remaining 90%).

S3 Estimation of the population of the test municipalities (i.e. the remaining 90%).

Furthermore, in the statistical approaches (S1, S2, and S3) we also estimate the population of the Netherlands. This means that we test the suitability of carrying out the census for 10% of the country (training dataset), and estimating the population of the rest of a country (test dataset).

In each of the 9 approaches we carry out separate experiments with the data in the 9 different LODs. This results in a total of 108 experiments.

As in related work [44, 86], we ignore very small buildings (footprint smaller than 20 m^2) such as sheds, garages, etc. which are unlikely to be inhabited (visible in Fig 1 as tiny white footprints in the overly residential areas).

Results and Discussion

Performance and observations

We perform the experiments, and compare them to the actual values, as observed in the governmental census dataset (CBS). We use percentage error because we are dealing with different scales of data (e.g. an error of 1000 residents is not of the same magnitude on the neighbourhood or city level). Furthermore, because of large errors in some statistical units (explained later), instead of the usual mean absolute error and root-mean-square error we use the median absolute error. As in related work [71, 87], we observe that estimations in areas with small populations is prone to a high relative error (see S1 Fig), hence medians are a good option here. The results of all experiments are given in Table 1. Because of many different models and types of data, we focus on the most important results only, however, the elaborated observations are similar with the rest of the models. It should be noticed that both the disaggregation and statistical approach exhibit congruent behaviour in most cases.

The results exhibit a large degree of variation between the accuracy depending on the approach, level of detail of the data, and the scale of the estimations. The smallest error of the volume-based disaggregation approach is in D5/LOD1[b] (the disaggregation from
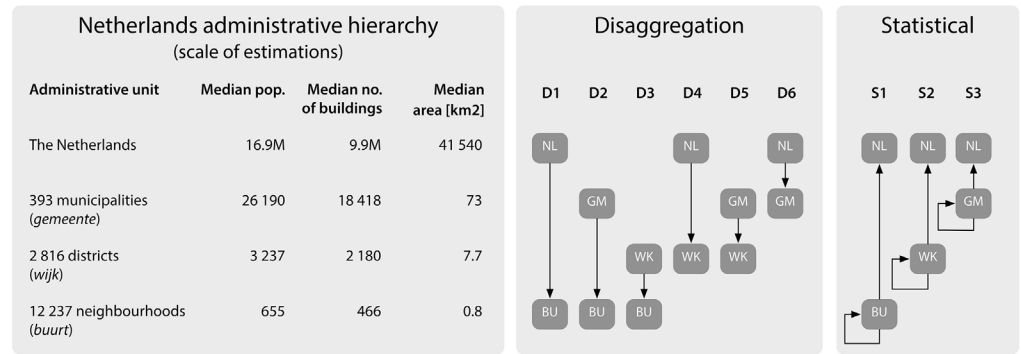
Fig 6. The Dutch statistical hierarchy, and our hybrid multi-scale approach. The hybrid approach refers to both the disaggregation and statistical approach, while multiple scales refer to the level of the statistical units. Statistics of the units obtained from data (c) Kadaster / Centraal Bureau voor de Statistiek, 2015. The provinces are not shown because they have not been considered in our work, and the data refer to the situation in 2015.

municipalities to districts) and it equals 11.8%. The smallest error in the statistical approach was observed in S3 (estimation of the population of cities), resulting in an error of 9.3%. We observe and conclude the following:

- 3D city models and the volume-based approach provide a substantial advantage over traditional 2D maps and the area-based approach because they capture the vertical extent of the building. However, the estimations carried out with 3D models are still less accurate than when using floorspace information. We think that volume does not add value over floorspace

Table 1. Median absolute percentage errors in the population estimates resulting from our experiments.

	a	b	c	a	b	c	a	b	c
				(1) Disaggregation					
	D1 (n = 12237)			**D2 (n = 12237)**			**D3 (n = 12237)**		
0	61.9	41.9	42.4	53.9	25.5	25.7	42.7	17.7	17.7
0+	39.8	20.8	20.8	37.2	16.2	16.4	29.1	12.0	12.0
1	56.4	25.5	25.8	53.0	20.8	20.7	42.4	15.6	15.3
	D4 (n = 3237)			**D5 (n = 3237)**			**D6 (n = 393)**		
0	56.5	37.7	38.2	34.3	15.5	15.5	32.0	25.3	25.5
0+	25.8	16.9	16.5	21.3	9.3	9.2	13.2	11.5	11.4
1	43.5	20.0	20.5	32.0	11.8	11.9	22.1	13.2	13.2
				(2) Statistical approach (local units)					
	S1 (n = 12237)			**S2 (n = 3237)**			**S3 (n = 393)**		
0	85.4	42.0	42.2	56.9	53.1	52.8	74.0	38.7	38.8
0+	35.4	18.3	18.5	41.5	28.9	28.5	20.6	12.6	12.2
1	66.8	24.3	24.8	49.8	26.1	28.6	28.9	9.5	9.3
				(2) Statistical approach (country level)					
	S1 (n = 1)			**S2 (n = 1)**			**S3 (n = 1)**		
0	0.6	1.4	1.4	2.7	5.6	5.7	21.5	1.3	1.7
0+	9.3	2.0	2.2	2.7	0.6	0.5	7.9	1.9	1.9
1	4.1	1.2	1.3	3.1	1.9	1.9	11.7	2.0	1.8

The order of errors in each 3×3 matrix is expressed in the same order as the LODs in Fig 4.

because two flats of the same floorspace but of different volumes (e.g. ceiling height of 2.5 m vs 4 m) generally do not host a different number of residents, unlike what the method would predict. It should be noted, however, that floorspace information is difficult to acquire automatically and it is generally not available.

- In most cases, semantic information on the use of buildings provides a substantial improvement in the estimations over data without such information. This helps to exclude non-residential units, which can significantly skew the estimations. Such behaviour is visible as outliers in the scatter plots in Fig 7 (other observations will be discussed in the continuation). Population estimation without information on the building function is practically unusable in most cases, especially in industrial neighbourhoods (in our experiments we have seen overestimations of more than 5000%). In fact, the results show that in this use case, semantic information is typically more important than the geometric detail (e.g. cf. error of 41.9% in D1/LOD0[b]—semantically enriched 2D footprints vs error of 56.4% in D1/LOD1[a]—plain 3D buildings).

- While semantic data is crucial, it appears that there is inconsistent added value of the detailed (fractional) semantic information versus only binary information. It seems that the difference between binary and fractional semantic information becomes negligible at the neighbourhood level. In fact, in some estimations (e.g. D4/LOD1) the estimations with fractional semantic information (D4/LOD1[c]) are slightly less accurate than when using binary semantic information (D4/LOD1[b]).
  In the floorspace data (LOD0+) there is generally a small improvement of using fractional semantic information rather than binary. A possible reason is that the volume-based estimations are more sensitive to errors in the input dataset.
  For the purposes described in this paper, it does not seem worthwhile to collect detailed building usage, as the binary information suffices. Because such information may be automatically derived from the building morphology, aerial imagery, land use maps, etc. [88–94], this insight is beneficial for estimations that need to be carried out on a large extent where cadastral data is not available.

- Different scales of estimations show different performance and different suitability for the different methods. In the disaggregation, the method works best in hierarchically close units: compare D3 (districts to neighbourhoods) with D1 (country to neighbourhoods). This is because such relations exhibit less difference in housing variations. Furthermore, it seems that disaggregating data to units higher in the hierarchy is more accurate than to units of a finer scale, because larger units such as districts and municipalities capture larger residential differences than small neighbourhoods, i.e. the variation among smaller units is greater than that among larger units. For example, two municipalities may have equal population but within municipalities the population differences among districts may be relatively large (e.g. rural vs urban zones). On the other hand differences among neighbourhoods in a district may be small.

- The statistical approach is of comparable accuracy to the disaggregation because it is also based on coefficients uniform for the whole country, which hide massive disparities among different neighbourhoods and provinces, and it is therefore equally affected by the differences in living standards and residential choices.
  However, for the largest extent (country), the statistical approach is impressively accurate: in the S1/LOD1[b] experiment (statistical approach applied on neighbourhoods with the semantic volume-based LOD1 block model) the population of the Netherlands based on a subset of 10% neighbourhoods has been estimated to 17 100 292, just a 1.2% overestimation from the
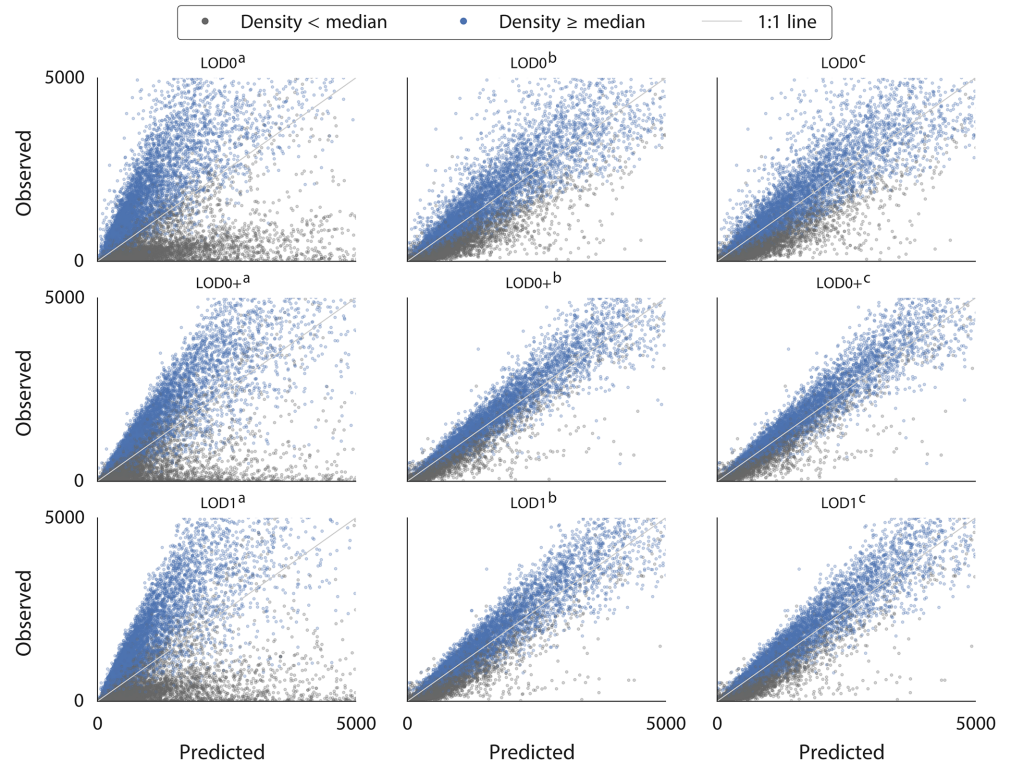
Fig 7. Observed (actual data from the government census) vs predicted scatter plots of the 9 input datasets in the D1 method. The performance of the models depends on the population density of the target area. The lower density refers to areas with the population density lower than the median of all neighbourhoods, and the higher those areas which are denser than the median, indicating urbanised areas. Notice the outliers in the estimations ([a]) that do not take advantage of the semantics—those represent highly industrialised areas without inhabitants or with sparse population. Furthermore, in the experiments carried out with fine-grade data most of the outliers are caused by input data (e.g. mislabelled residential use of a non-residential building) and by districts in which housing standards highly deviate from the average. Observed data (c) Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2015.

true figure. The floorspace-based (LOD0+[b]) data fares even better with a deviation of 0.5% in S2. This finding gives confidence in the use of 3D city models for estimating the population of large areas such as countries, especially in developing countries since the data required for such estimations can be derived automatically and remotely from airborne sensors [95]. However, it should be noted that the model S1/LOD0[a] (building footprints without information on the building use) performed best with an error of 0.6%. It is hard to explain the reason why in this particular model lesser data gave better results, because all errors (induced by different LODs, uncertainty in the input data, different residential choices, etc.) are aggregated in a single number that cannot be decomposed.

- We have noticed that the models tend to overestimate the population in rural areas, and underestimate it in urban areas (see the coloured points in Fig 7). This finding is similar to the observations in related work [38]. The differences are caused by the varying utilisation of living space, which differ between less and more densely populated entities. We use this finding in the succeeding sections for additional insights and we take advantage of it to improve the statistical approach (models S1, S2, and S3).

Sources of error

After analysing the errors we observe different causes of errors. The residential differences (e.g. residential space per resident) is the principal cause of the residuals (the errors very strongly correlate with the average space per resident; $r > 0.99$). There is a variable level of occupancy and variable utilisation of space within each building, i.e. living space per inhabitant considerably varies based on social, economical, and other factors. Some households live in large houses, while others in small studios and dormitories, rendering significant differences in the residential density [43], and presenting a problem for population estimation with remotely sensed data [46]. Furthermore, these differences are also caused by non-residential space within residential units, such as storage rooms, utility rooms, common rooms, gyms, garages, etc., which increase the building size and considered dwelling space, but due to the shortcomings of the data cannot be accounted as non-residential space. It is usually not possible to assume that these characteristics are equally distributed in each entity, as they are not constant among different neighbourhoods and also on larger extent such as among municipalities [68, 96]. This fact is also visible in Fig 2. Therefore it is important to consider different environments when calibrating the method, and accept imperfections as one model cannot fit all situations within a large area such as a province or country.

We had expected that these differences would cancel out within the statistical entities (since one typically contains hundreds of houses, see Fig 1), however, the difference between units, including larger ones such as cities, is still gross. One would assume that a city contains a fair diversity of different configurations, but it turns out that each city has a unique setting which cannot be applied to another one.

Furthermore, another variation of the dwelling density is caused by vacant residential buildings (e.g. empty houses for sale, vacation homes). In our method we can only assume that the vacancy rate is homogeneous in our area of study, consistently with other researchers (e.g. when estimating the energy demand [97]), however, that assumption might deviate from the reality.

When using the data without information on building use (i.e. LODx[a]) many large errors were found in industrial neighbourhoods with huge building volumes, highlighting the importance of using semantics. When using the semantically enriched buildings, the results improved substantially. However, errors in the input data on building use have also caused errors in the estimation of the population. For instance, we have noted that in an industrial neighbourhood a large factory was mislabelled as a residential building, so the population has been pointlessly disaggregated in an empty building, inducing a substantial error. The input datasets that we used were very accurate [98, 99], but occasional small errors induced gross errors in the estimations. Furthermore, it is worthwhile to mention that there were peculiar cases which also caused discrepancies, such as a small neighbourhood with a prison as its sole building. Its inmates are counted as residents in the CBS dataset, but the prison building is not classified as residential in the cadastral dataset, hence the estimate of the neighbourhood exhibited a large error—the population was predicted to be 0, while in reality it is 75.

The 3D geometric aspect (calculated volume) may induce errors to the estimations as well. It has been suggested [64] that geometric errors in 3D city models (e.g. inconsistencies caused by vegetation in the elevation dataset) may substantially influence spatial analyses, especially the computation of the volume [100].

The related work in analysing error propagation in population estimation is limited to 2D [101–103]. For future work it would be interesting to investigate the influence of errors in the input data when using volume-based approaches.

Analysis of errors and enhancement of the statistical approach

We have analysed the errors with demographic and other indicators for each statistical unit in order to understand them better and to potentially improve our methods.

We have analysed the income of each neighbourhood and did not find a correlation with errors. We had presumed that low-income neighbourhoods might have less space per resident, as income may be related to living standard and may drive residential purchasing choices. However, that is not the case, because there are cheap but large country-side properties, and expensive small flats in cities such as Amsterdam, invalidating our assumption.

In the previous section we have noted the particular behaviour of errors with respect to the different population densities of estimated areas. There is a clear difference between more and less urbanised areas caused by the different utilisation of dwelling space (see Fig 8). It is clear that the data on the population density could be used to improve the estimations, but as such it is not available prior to the estimation of population (otherwise we would not need to conduct the estimations).

However, we have realised that there is another indicator that it is associated with the population density, and which is available prior to the estimations: the average building height in a neighbourhood is associated with the population density (see S2 Fig), and consequently to the living space. Therefore, for each neighbourhood we have calculated the average building height (easily available since we have 3D city models), and we have incorporated it in our multiple linear regression model (which now contains two variables: the total building space in the statistical unit, and the average height of buildings in the unit). We have not applied this enhancement to the LOD0 approach in which vertical measurements are not available.

The statistical experiments show that there is an improvement to the models: a reduction of errors by a few percent on average has been observed in the models S1, S2, and S3. Note that the results presented in the previous section are of those with the enhanced models, and that the disaggregation method was not enhanced because of its inherently different approach in which there is no training data.

While we believe that the presented prediction models *might* be further augmented to improve the estimations with additional variables and 2D GIS data such as land use, in this paper we have used only 3D models to determine how accurate the predictions can be if relied solely on them. Adding such additional variables is avoided because of a contradictory situation: if such data is available, it is likely that accurate census data is also available, rendering such estimations unnecessary.

Conclusions and outlook

In this study we have used a 3D city model to estimate the population of 12.2 thousand neighbourhoods, 2816 districts, and 393 municipalities in the Netherlands, and of the Netherlands itself. Our results indicate that in certain circumstances 3D city models can give a good approximation of the population, and that, in most cases, 3D city models add value over traditionally used 2D datasets, but also that they are not accurate enough to replace accurate census techniques employed by governments. Furthermore, there were certain instances when 2D data (even without the information on building use, e.g. S1/LOD0[a]) performed better than 3D data, which is beneficial because such data is simpler to acquire. The main reason why this method is useful is because it does not require expensive and time consuming field surveys and other means of collecting population counts as the data can be acquired automatically and remotely, and it can be carried out more frequently, in contrast to official censuses (usually conducted every decade).
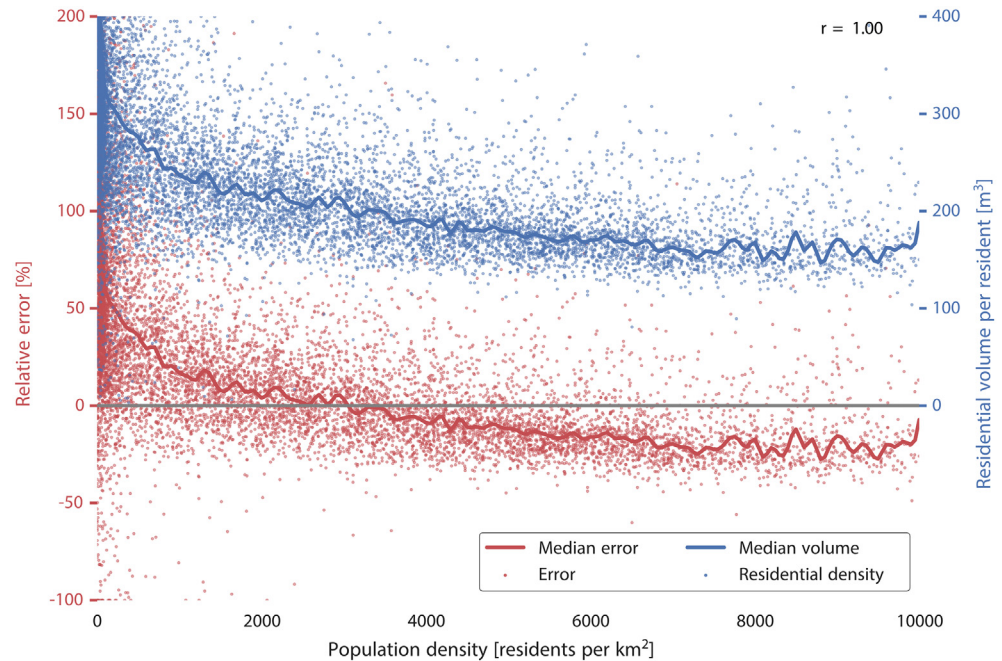
Fig 8. The relations between the errors, population density, and living space per statistical neighbourhood. The errors in the model are from the experiment D1/LOD1^c. Data (c) Kadaster / Centraal Bureau voor de Statistiek, 2015.

One of the strengths of our work over previous studies is that we carried out a country-wide analysis, in which differences between neighbourhoods are more emphasised. Our study is multi-LOD (both area-based and volume-based approaches have been evaluated, along with multiple grades of semantic information), multi-scale (for assessing the suitability of mapping statistical units of different sizes), and multi-method (both the weighted disaggregation and statistical approaches have been employed).

Remote estimation of population with GIS could be applied in areas where census information is not available or it is not reliable, and serves two purposes: (1) as a potential solution to estimate the population count of large areas where a census is not available, or as an intercensal estimate; and (2) for refining the population on a finer scale (e.g. disaggregation of an accurate census of a city among its neighbourhoods).

Our approach is easily applicable in other countries. Governments have started to publicly release building footprints and other GIS data [89], and where data is available many 3D city models have been generated [104–107]. Alternatively, 3D city models may be generated from volunteered geoinformation [108], ensuring the applicability of our method elsewhere. While in this study for the building use we used datasets from the cadastre, it is worth noting that such data can also be derived manually from aerial images, and automatically from the building morphology and other characteristics, or from volunteered geoinformation [88–94]. Such an approach provides an enhancement over previous research, since in related work coarse datasets have traditionally been used, e.g. Kressler and Steinnocher [5] and Silván-Cárdenas et al. [41] distinguish residential buildings from non-residential ones with a zoning map.

Concerning the first application, estimating the population count of large areas where a census is not available, in the 21st century there are still many places around the world where the census has not been carried out in decades, and such remote sensing methods can help to bridge the gap [109, 110]. For instance, Myanmar did not have a reliable census until two years

ago, and in the meantime the authorities were dealing with information which turned out to be significantly erroneous [111, 112], something unthinkable in developed countries nowadays. Obviously, low income countries cannot boast about 3D city models, however, with the development of remote sensing technologies, and surge of volunteered geoinformation and their quality [113, 114], the generation of 3D city models is becoming increasingly simpler and cheaper [107, 108, 115]. Therefore we expect that in the near future country-wide 3D city models will not be a luxury exclusive to developed countries.

With respect to the second purpose, the derived data on the number of residents on a finer scale is beneficial for a multitude of applications [116], such as disaster management (e.g. in flooded areas) [117, 118], analysing accessibility [119], public health [1, 120], crime mapping [121], environmental risk [80, 122], infrastructure planning and transportation sustainability [123], epidemiology [124], territorial classification [125], assessing exposure to noise [51, 126, 127], optimising network coverage (e.g. television) to cover more people [128, 129], for finding areas for landing of stratospheric balloons [129], marketing strategies [44], estimating the quantity of waste [130], estimating energy consumption [131], and in urban simulations [132].

We have also discovered that this method can also be used to detect potential errors in authoritative census and building data (e.g. we have detected erroneous semantic information for some commercial buildings by analysing the large errors in population estimates). Furthermore, we envisage that such method could be used for detecting false residencies (e.g. a large number of people registered in a particular neighbourhood for tax-related reasons, triggering an alert by the population that exceeds the housing capacity in that area).

The results indicate that the estimations are hampered by socio-economic disparities between neighbourhoods, and that population estimation is more reliable when focused on statistical units with a closer proximity. However, this limitation does not seem to affect the estimation of the national population, in which case our method has particularly excelled.

For future work it would be worthwhile to advance the sampling method of the training data in the statistical approach to investigate whether that leads to more accurate estimates. For instance, stratified sampling [133] could be employed instead of the simple random sampling which is used now. Such sampling method could stratify entities based on different characteristics obtainable from 3D city models, such as predominant building types in a neighbourhood, and apply different statistical models to each stratum.

Supporting Information

S1 Fig. Less populated districts exhibit large relative errors, promoting the use of medians. In relative terms, the estimation is more accurate when carried out in more populous areas. These are the results from the experiments S1/LOD1^c. The two histograms show the data divided in two bins (the left one of the statistical units with the population smaller than the median value of all units (710 residents), and the one on the right the units with the population higher than the median). Not to be confused with Fig 8 which shows the relation of errors to the population density (however, notice that in this case as well the methods tend to underestimate the population in more populated areas).
(TIF)

S2 Fig. Association of the population density and vertical extent of the neighbourhood. While the population density is not available for adjusting our models, we have taken advantage of the vertical extent which hints at the population density, and in turn helps in adjusting the prediction between urban and rural areas.
(TIF)

S3 Fig. Elevation dataset (AHN) used to generate the 3D city model. The point cloud was obtained with airborne laser scanning, and the colours represent the elevation. The spatial extent and angle of view correspond to the one shown in Fig 3. The accuracy of the points is within a few centimetres [98]. The whole dataset contains 639B points [134]. Data (c) Actueel Hoogtebestand Nederland.
(TIF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: FB. Performed the experiments: FB. Analyzed the data: FB. Contributed reagents/materials/analysis tools: HL RP. Wrote the paper: FB KAO HL RP JS.

References

1. Anderson W, Guikema S, Zaitchik B, Pan W. Methods for Estimating Population Density in Data-Limited Areas: Evaluating Regression and Tree-Based Models in Peru. PLOS ONE. 2014 Jul; 9(7): e100037. doi: 10.1371/journal.pone.0100037 PMID: 24992657

2. Hillson R, Alejandre JD, Jacobsen KH, Ansumana R, Bockarie AS, Bangura U, et al. Methods for Determining the Uncertainty of Population Estimates Derived from Satellite Imagery and Limited Survey Data: A Case Study of Bo City, Sierra Leone. PLOS ONE. 2014 Nov; 9(11):e112241. doi: 10.1371/journal.pone.0112241 PMID: 25398101

3. Welch R. Monitoring urban population and energy utilization patterns from satellite Data. Remote sensing of Environment. 1980 Feb; 9(1):1–9. doi: 10.1016/0034-4257(80)90043-7

4. Lu D, Weng Q, Li G. Residential population estimation using a remote sensing derived impervious surface approach. International Journal of Remote Sensing. 2006 Aug; 27(16):3553–3570. doi: 10.1080/01431160600617202

5. Kressler F, Steinnocher K. Object-oriented analysis of image and LiDAR data and its potential for a dasymetric mapping application. In: On segment based image fusion. Springer Berlin Heidelberg; 2008. p. 611–624.

6. Lo CP. Automated population and dwelling unit estimation from high-resolution satellite images: a GIS approach. International Journal of Remote Sensing. 1995 Jan; 16(1):17–34. doi: 10.1080/01431169508954369

7. Lo CP. Population Estimation Using Geographically Weighted Regression. GIScience & Remote Sensing. 2013 May; 45(2):131–148. doi: 10.2747/1548-1603.45.2.131

8. Tobler WR. Satellite confirmation of settlement size coefficients. Area. 1969; 1(3):30–34.

9. Kraus SP, Senger LW, Ryerson JM. Estimating population from photographically determined residential land use types. Remote sensing of Environment. 1974 Jan; 3(1):35–42. doi: 10.1016/0034-4257(74)90036-4

10. Lo CP, Welch R. Chinese Urban Population Estimates. Annals of the Association of American Geographers. 1977 Jun; 67(2):246–253. doi: 10.1111/j.1467-8306.1977.tb01137.x

11. Al-garni AM. Mathematical predictive models for population estimation in urban areas using space products and GIS technology. Mathematical and Computer Modelling. 1995 Jul; 22(1):95–107. doi: 10.1016/0895-7177(95)00104-A

12. Wu C, Murray AT. Population Estimation Using Landsat Enhanced Thematic Mapper Imagery. Geographical Analysis. 2007 Jan; 39(1):26–43. doi: 10.1111/j.1538-4632.2006.00694.x

13. Zhan FB, Tapia Silva FO, Santillana M. Estimating small-area population growth using geographic-knowledge-guided cellular automata. International Journal of Remote Sensing. 2010 Nov; 31 (21):5689–5707. doi: 10.1080/01431161.2010.496802

14. Langford M. An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data. Geographical Analysis. 2013 Jul; 45(3):324–344. doi: 10.1111/gean.12012

15. Xie Z. A Framework for Interpolating the Population Surface at the Residential-Housing-Unit Level. GIScience & Remote Sensing. 2013 May; 43(3):233–251. doi: 10.2747/1548-1603.43.3.233

16. Deng C, Wu C, Wang L. Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. International Journal of Remote Sensing. 2010 Nov; 31 (21):5673–5688. doi: 10.1080/01431161.2010.496806

17. Wu C, Murray AT. A cokriging method for estimating population density in urban areas. Computers, Environment and Urban Systems. 2005 Sep; 29(5):558–579. doi: 10.1016/j.compenvurbsys.2005.01. 006

18. Bagan H, Yamagata Y. Landsat analysis of urban growth: How Tokyo became the world's largest megacity during the last 40 years. Remote sensing of Environment. 2012 Dec; 127:210–222. doi: 10. 1016/j.rse.2012.09.011

19. Yuan Y, Smith RM, Limp WF. Remodeling census population with spatial information from LandSat TM imagery. Computers, Environment and Urban Systems. 1997 May; 21(3–4):245–258. doi: 10. 1016/S0198-9715(97)01003-X

20. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE. 2015 Feb; 10(2): e0107042. doi: 10.1371/journal.pone.0107042 PMID: 25689585

21. Gaughan AE, Stevens FR, Linard C, Jia P, Tatem AJ. High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. PLOS ONE. 2013 Feb; 8(2):e55882. doi: 10.1371/journal.pone. 0055882 PMID: 23418469

22. Bakillah M, Liang S, Mobasheri A, Jokar Arsanjani J, Zipf A. Fine-resolution population mapping using OpenStreetMap points-of-interest. International Journal of Geographical Information Science. 2014 Sep; 28(9):1940–1963. doi: 10.1080/13658816.2014.909045

23. Doll CNH, Muller JP, Morley JG. Mapping regional economic activity from night-time light satellite imagery. Ecological Economics. 2006 Apr; 57(1):75–92. doi: 10.1016/j.ecolecon.2005.03.007

24. Anderson SJ, Tuttle BT, Powell RL, Sutton PC. Characterizing relationships between population density and nighttime imagery for Denver, Colorado: issues of scale and representation. International Journal of Remote Sensing. 2010 Nov; 31(21):5733–5746. doi: 10.1080/01431161.2010.496798

25. Sutton P. Modeling population density with night-time satellite imagery and GIS. Computers, Environment and Urban Systems. 1997 May; 21(3–4):227–244. doi: 10.1016/S0198-9715(97)01005-3

26. Pozzi F, Small C. Analysis of urban land cover and population density in the United States. Photogrammetric Engineering and Remote Sensing. 2005 Jun; 71(6):719–726. doi: 10.14358/PERS.71.6. 719

27. Xie Y. The overlaid network algorithms for areal interpolation problem. Computers, Environment and Urban Systems. 1995 Jul; 19(4):287–306. doi: 10.1016/0198-9715(95)00028-3

28. Steiger E, Westerholt R, Resch B, Zipf A. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. Computers, Environment and Urban Systems. 2015 Nov; 54:255–265. doi: 10.1016/j.compenvurbsys.2015.09.007

29. Lwin KK, Sugiura K, Zettsu K. Space–time multiple regression model for grid-based population estimation in urban areas. International Journal of Geographical Information Science. 2016; 30(8):1579–1593. doi: 10.1080/13658816.2016.1143099

30. Wu Ss, Wang L, Qiu X. Incorporating GIS Building Data and Census Housing Statistics for Sub-Block-Level Population Estimation. The Professional Geographer. 2008 Jan; 60(1):121–135. doi: 10.1080/ 00330120701724251

31. Harvey JT. Estimating census district populations from satellite imagery: Some approaches and limitations. International Journal of Remote Sensing. 2010 Nov; 23(10):2071–2095. doi: 10.1080/ 01431160110075901

32. Lwin KK, Murayama Y. Estimation of Building Population from LIDAR Derived Digital Volume Model. In: Spatial Analysis and Modeling in Geographical Transformation Process. Dordrecht: Springer Netherlands; 2011. p. 87–98.

33. Suveg I, Vosselman G. Reconstruction of 3D building models from aerial images and maps. ISPRS Journal of Photogrammetry and Remote Sensing. 2004 Jan; 58(3–4):202–224. doi: 10.1016/j. isprsjprs.2003.09.006

34. Musialski P, Wonka P, Aliaga DG, Wimmer M, van Gool L, Purgathofer W. A Survey of Urban Reconstruction. Computer Graphics Forum. 2013 May; 32(6):146–177. doi: 10.1111/cgf.12077

35. Truong-Hong L, Laefer DF. Quantitative evaluation strategies for urban 3D model generation from remote sensing data. Computers and Graphics. 2015 Jun; 49:82–91. doi: 10.1016/j.cag.2015.03.001

36. Serna A, Marcotegui B. Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning. ISPRS Journal of Photogrammetry and Remote Sensing. 2014 Jul; 93:243–255. doi: 10.1016/j.isprsjprs.2014.03.015

37. Rottensteiner F, Sohn G, Gerke M, Wegner JD, Breitkopf U, Jung J. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing. 2014 Jul; 93:256–271.

38. Dong P, Ramesh S, Nepali A. Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. International Journal of Remote Sensing. 2010 Nov; 31(21):5571–5586. doi: 10.1080/01431161.2010.496804

39. Lu Z, Im J, Quackenbush L. A Volumetric Approach to Population Estimation Using Lidar Remote Sensing. Photogrammetric Engineering and Remote Sensing. 2011 Nov; 77(11):1145–1156. doi: 10.14358/PERS.77.11.1145

40. Lu Z, Im J, Quackenbush L, Halligan K. Population estimation based on multi-sensor data fusion. International Journal of Remote Sensing. 2010 Nov; 31(21):5587–5604. doi: 10.1080/01431161.2010.496801

41. Silván-Cárdenas JL, Wang L, Rogerson P, Wu C, Feng T, Kamphaus BD. Assessing fine-spatial-resolution remote sensing for small-area population estimation. International Journal of Remote Sensing. 2010 Nov; 31(21):5605–5634. doi: 10.1080/01431161.2010.496800

42. Ural S, Hussain E, Shan J. Building population mapping with aerial imagery and GIS data. International Journal of Applied Earth Observation and Geoinformation. 2011 Dec; 13(6):841–852. doi: 10.1016/j.jag.2011.06.004

43. Sridharan H, Qiu F. A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes. Geographical Analysis. 2013 Jul; 45(3):238–258. doi: 10.1111/gean.12010

44. Lwin K, Murayama Y. A GIS Approach to Estimation of Building Population for Micro-spatial Analysis. Transactions in GIS. 2009 Aug; 13(4):401–414. doi: 10.1111/j.1467-9671.2009.01171.x

45. Qiu F, Sridharan H, Chun Y. Spatial Autoregressive Model for Population Estimation at the Census Block Level Using LIDAR-derived Building Volume Information. Cartography and Geographic Information Science. 2010 Jan; 37(3):239–257. doi: 10.1559/152304010792194949

46. Alahmadi M, Atkinson P, Martin D. Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. Computers, Environment and Urban Systems. 2013 Sep; 41:167–176. doi: 10.1016/j.compenvurbsys.2013.06.002

47. Alahmadi M, Atkinson PM, Martin D. A Comparison of Small-Area Population Estimation Techniques Using Built-Area and Height Data, Riyadh, Saudi Arabia. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; 9(5):1959–1969. doi: 10.1109/JSTARS.2014.2374175

48. Kolbe TH. Representing and exchanging 3D city models with CityGML. In: Zlatanova S, Lee J, editors. 3D Geo-Information Sciences. Springer Berlin Heidelberg; 2009. p. 15–31.

49. Billen R, Cutting-Decelle AF, Marina O, de Almeida JP, M C, Falquet G, et al. 3D City Models and urban information: Current issues and perspectives. In: 3D City Models and urban information: Current issues and perspectives—European COST Action TU0801. Les Ulis, France: EDP Sciences; 2014. p. I–118.

50. Biljecki F, Stoter J, Ledoux H, Zlatanova S, Çöltekin A. Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information. 2015 Dec; 4(4):2842–2889. doi: 10.3390/ijgi4042842

51. Stoter J, de Kluijver H, Kurakula V. 3D noise mapping in urban areas. International Journal of Geographical Information Science. 2008; 22(8):907–924. doi: 10.1080/13658810701739039

52. Kaden R, Kolbe TH. Simulation-Based Total Energy Demand Estimation of Buildings using Semantic 3D City Models. International Journal of 3-D Information Modeling. 2014; 3(2):35–53. doi: 10.4018/ij3dim.2014040103

53. Bahu JM, Koch A, Kremers E, Murshed SM. Towards a 3D Spatial Urban Energy Modelling Approach. International Journal of 3-D Information Modeling. 2015 Jul; 3(3):1–16. doi: 10.4018/ij3dim.2014070101

54. Fisher-Gewirtzman D, Shashkov A, Doytsher Y. Voxel based volumetric visibility analysis of urban environments. Survey Review. 2013 Nov; 45(333):451–461. doi: 10.1179/1752270613Y.0000000059

55. Bartie P, Reitsma F, Kingham S, Mills S. Advancing visibility modelling algorithms for urban environments. Computers, Environment and Urban Systems. 2010 Nov; 34(6):518–531. doi: 10.1016/j.compenvurbsys.2010.06.002

56. Xiong B, Jancosek M, Oude Elberink S, Vosselman G. Flexible building primitives for 3D building modeling. ISPRS Journal of Photogrammetry and Remote Sensing. 2015 Mar; 101:275–290. doi: 10.1016/j.isprsjprs.2015.01.002

57. Haala N, Kada M. An update on automatic 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing. 2010 Nov; 65(6):570–580. doi: 10.1016/j.isprsjprs.2010.09.006

58. Sirmacek B, Taubenbock H, Reinartz P, Ehlers M. Performance Evaluation for 3-D City Model Generation of Six Different DSMs From Air- and Spaceborne Sensors. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2012; 5(1):59–70. doi: 10.1109/JSTARS.2011.2178399

59. Arroyo Ohori K, Ledoux H, Stoter J. A dimension-independent extrusion algorithm using generalised maps. International Journal of Geographical Information Science. 2015 Mar; 29(7):1166–1186. doi: 10.1080/13658816.2015.1010535

60. Ledoux H, Meijers M. Topologically consistent 3D city models obtained by extrusion. International Journal of Geographical Information Science. 2011 Apr; 25(4):557–574. doi: 10.1080/13658811003623277

61. Gröger G, Plümer L. CityGML—Interoperable semantic 3D city models. ISPRS Journal of Photogrammetry and Remote Sensing. 2012 Jul; 71:12–33. doi: 10.1016/j.isprsjprs.2012.04.004

62. Biljecki F, Ledoux H, Stoter J, Zhao J. Formalisation of the level of detail in 3D city modelling. Computers, Environment and Urban Systems. 2014 Nov; 48:1–15. doi: 10.1016/j.compenvurbsys.2014.05.004

63. Biljecki F, Ledoux H, Stoter J, Vosselman G. The variants of an LOD of a 3D building model and their influence on spatial analyses. ISPRS Journal of Photogrammetry and Remote Sensing. 2016; 116:42–54. doi: 10.1016/j.isprsjprs.2016.03.003

64. Biljecki F, Heuvelink GBM, Ledoux H, Stoter J. Propagation of positional error in 3D GIS: estimation of the solar irradiation of building roofs. International Journal of Geographical Information Science. 2015 Dec; 29(12):2269–2294. doi: 10.1080/13658816.2015.1073292

65. Arroyo Ohori K, Ledoux H, Biljecki F, Stoter J. Modeling a 3D City Model and Its Levels of Detail as a True 4D Model. ISPRS International Journal of Geo-Information. 2015 Sep; 4(3):1055–1075. doi: 10.3390/ijgi4031055

66. Stadler A, Kolbe TH. Spatio-semantic coherence in the integration of 3D city models. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2007 Jun; XXXVI-2/C43:8.

67. Boeters R, Arroyo Ohori K, Biljecki F, Zlatanova S. Automatically enhancing CityGML LOD2 models with a corresponding indoor geometry. International Journal of Geographical Information Science. 2015 Dec; 29(12):2248–2268. doi: 10.1080/13658816.2015.1072201

68. Shiravi S, Zhong M, Beykaei SA, Hunt JD, Abraham JE. An assessment of the utility of LiDAR data in extracting base-year floorspace and a comparison with the census-based approach. Environment and Planning B: Planning and Design. 2015; 42(4):708–729.

69. Xie Y, Weng A, Weng Q. Population Estimation of Urban Residential Communities Using Remotely Sensed Morphologic Data. IEEE Geoscience and Remote Sensing Letters. 2015 May; 12(5):1111–1115. doi: 10.1109/LGRS.2014.2385597

70. Wu Ss, Qiu X, Wang L. Population Estimation Methods in GIS and Remote Sensing: A Review. GIScience & Remote Sensing. 2005 Mar; 42(1):80–96. doi: 10.2747/1548-1603.42.1.80

71. Brinegar SJ, Popick SJ. A Comparative Analysis of Small Area Population Estimation Methods. Cartography and Geographic Information Science. 2013 Mar; 37(4):273–284. doi: 10.1559/152304010793454327

72. Mennis J. Dasymetric Mapping for Estimating Population in Small Areas. Geography Compass. 2009 Mar; 3(2):727–745. doi: 10.1111/j.1749-8198.2009.00220.x

73. Goodchild MF, Lam NSN. Areal Interpolation—a Variant of the Traditional Spatial Problem. Geo-Processing. 1980; 1(3):297–312.

74. Flowerdew R, Green M, Kehris E. Using areal interpolation methods in geographic information systems. Papers in Regional Science. 1991 Jul; 70(3):303–315.

75. Liu XH, Kyriakidis P C, Goodchild MF. Population-density estimation using regression and area-to-point residual kriging. International Journal of Geographical Information Science. 2008 Mar; 22(4):431–447. doi: 10.1080/13658810701492225

76. Langford M. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. Computers, Environment and Urban Systems. 2006 Mar; 30(2):161–180. doi: 10.1016/j.compenvurbsys.2004.07.001

77. Zandbergen PA, Ignizio DA. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. Cartography and Geographic Information Science. 2013 Mar; 37(3):199–214. doi: 10.1559/152304010792194985

78. Mennis J. Generating Surface Models of Population Using Dasymetric Mapping?. The Professional Geographer. 2003; 55(1):31–42.

79. Wright JK. A Method of Mapping Densities of Population: With Cape Cod as an Example. Geographical Review. 1936 Jan; 26(1):103. doi: 10.2307/209467

80. Maantay JA, Maroko AR, Herrmann C. Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). Cartography and Geographic Information Science. 2007 Jan; 34(2):77–102. doi: 10.1559/152304007781002190

81. Eicher CL, Brewer CA. Dasymetric mapping and areal interpolation: Implementation and evaluation. Cartography and Geographic Information Science. 2001; 28(2):125–138. doi: 10.1559/152304001782173727

82. Mennis J, Hultgren T. Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. Cartography and Geographic Information Science. 2006 Jan; 33(3):179–194. doi: 10.1559/152304006779077309

83. Holt JB, Lo CP, Hodler TW. Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. Cartography and Geographic Information Science. 2004 Jan; 31(2):103–121. doi: 10.1559/1523040041649407

84. Alahmadi M, Atkinson P, Martin D. Fine spatial resolution residential land-use data for small-area population mapping: a case study in Riyadh, Saudi Arabia. International Journal of Remote Sensing. 2015 Sep; 36(17):4315–4331. doi: 10.1080/01431161.2015.1079666

85. Liu X, Clarke K. Estimation of Residential Population Using High Resolution Satellite Imagery. In: Proceedings of the 3rd Symposium in Remote Sensing of Urban Areas. Istanbul, Turkey; 2002. p. 153–160.

86. Greger K. Spatio-Temporal Building Population Estimation for Highly Urbanized Areas Using GIS. Transactions in GIS. 2015 Feb; 19(1):129–150. doi: 10.1111/tgis.12086

87. Zoraghein H, Leyk S, Ruther M, Buttenfield BP. Exploiting temporal information in parcel data to refine small area population estimates. Computers, Environment and Urban Systems. 2016 Jul; 58:19–28. doi: 10.1016/j.compenvurbsys.2016.03.004

88. Henn A, Römer C, Gröger G, Plümer L. Automatic classification of building types in 3D city models. Geoinformatica. 2012 Apr; 16(2):281–306. doi: 10.1007/s10707-011-0131-x

89. Hecht R, Meinel G, Buchroithner M. Automatic identification of building types based on topographic databases—a comparison of different data sources. International Journal of Cartography. 2015 Aug; 1(1):18–31. doi: 10.1080/23729333.2015.1055644

90. Kunze C, Hecht R. Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population. Computers, Environment and Urban Systems. 2015 Sep; 53:4–18. doi: 10.1016/j.compenvurbsys.2015.04.002

91. Neidhart H, Sester M. Identifying building types and building clusters using 3-D laser scanning and GIS-data. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2004; XXXV/B4:715–720.

92. Belgiu M, Tomljenovic I, Lampoltshammer T, Blaschke T, Höfle B. Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data. Remote Sensing. 2014 Feb; 6(2):1347–1366. doi: 10.3390/rs6021347

93. Hermosilla T, Ruiz LA, Recio JA, Cambra-López M. Assessing contextual descriptive features for plot-based classification of urban areas. Landscape and Urban Planning. 2012 May; 106(1):124–137. doi: 10.1016/j.landurbplan.2012.02.008

94. Hermosilla T, Ruiz LA, Recio JA, Balsa-Barreiro J. Land-use mapping of Valencia city area from aerial images and LiDAR data. In: GEOProcessing 2012: The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services. Valencia, Spain; 2012. p. 232–237.

95. Xiong B, Oude Elberink S, Vosselman G. A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. ISPRS Journal of Photogrammetry and Remote Sensing. 2014 Jul; 93:227–242. doi: 10.1016/j.isprsjprs.2014.01.007

96. Swanson DA, Hough GC Jr. An Evaluation of Persons per Household (PPH) Estimates Generated by the American Community Survey: A Demographic Perspective. Population Research and Policy Review. 2012; 31(2):235–266. doi: 10.1007/s11113-012-9227-8

97. Nouvel R, Mastrucci A, Leopold U, Baume O, Coors V, Eicker U. Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. Energy and Buildings. 2015; 107:204–212. doi: 10.1016/j.enbuild.2015.08.021

98. van der Sande C, Soudarissanane S, Khoshelham K. Assessment of Relative Accuracy of AHN-2 Laser Scanning Data Using Planar Features. Sensors. 2010 Sep; 10(9):8198–8214. doi: 10.3390/s100908198 PMID: 22163650

99. VROM. Geometrie en de BAG (Verdiepingsdocument). Dutch Ministry of Housing, Spatial Planning and the Environment; 2008. Available from: http://www.kadaster.nl/web/artikel/download/BAG-verdiepingsdocument-geometrie-1.htm.

100. Biljecki F, Ledoux H, Stoter J. Error propagation in the computation of volumes in 3D city models with the Monte Carlo method. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci. 2014 Oct; II-2:31–39. doi: 10.5194/isprsannals-II-2-31-2014

101. Fisher PF, Langford M. Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Inter-polation by Dasymetric Mapping. The Professional Geographer. 1996 Aug; 48(3):299–309. doi: 10.1111/j.0033-0124.1996.00299.x

102. Fisher PF, Langford M. Modelling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation. Environment and Planning A. 1995 Feb; 27(2):211–224. doi: 10.1068/a270211

103. Sadahiro Y. Accuracy of areal interpolation: A comparison of alternative methods. Journal of Geo-graphical Systems. 1999 Dec; 1(4):323–346. doi: 10.1007/s101090050017

104. Kolbe TH, Burger B, Cantzler B. CityGML goes to Broadway. In: Photogrammetric Week'15. Stutt-gart, Germany; 2015. p. 343–356.

105. Aringer K, Roschlaub R. Bavarian 3D Building Model and Update Concept Based on LiDAR, Image Matching and Cadastre Information. In: Innovations in 3D Geo-Information Sciences. Springer Inter-national Publishing; 2014. p. 143–157.

106. Stoter J, Roensdorf C, Home R, Capstick D, Streilein A, Kellenberger T, et al. 3D Modelling with National Coverage: Bridging the Gap Between Research and Practice. In: Advances in 3D Geo-Infor-mation Sciences. Cham: Springer International Publishing; 2015. p. 207–225.

107. Zhu L, Lehtomäki M, Hyyppä J, Puttonen E, Krooks A, Hyyppä H. Automated 3D Scene Reconstruc-tion from Open Geospatial Data Sources: Airborne Laser Scanning and a 2D Topographic Database. Remote Sensing. 2015 May; 7(6):6710–6740. doi: 10.3390/rs70606710

108. Goetz M. Towards generating highly detailed 3D CityGML models from OpenStreetMap. International Journal of Geographical Information Science. 2013 May; 27(5):845–865. doi: 10.1080/13658816.2012.721552

109. Tatem AJ, Noor AM, von Hagen C, Di Gregorio A, Hay SI. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. PLOS ONE. 2007 Dec; 2(12):e1298. doi: 10.1371/journal.pone.0001298 PMID: 18074022

110. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. PLOS ONE. 2012 Feb; 7(2):e31743. doi: 10.1371/journal.pone.0031743 PMID: 22363717

111. Spoorenberg T. Provisional results of the 2014 census of Myanmar: The surprise that wasn't. Asian Population Studies. 2014 Oct; 11(1):4–6. doi: 10.1080/17441730.2014.972084

112. Spoorenberg T. Myanmar's first census in more than 30 years: A radical revision of the official popula-tion count. Population & Societies. 2015 Nov; 527:1–4.

113. Hecht R, Kunze C, Hahmann S. Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. ISPRS International Journal of Geo-Information. 2013 Dec; 2(4):1066–1091. doi: 10.3390/ijgi2041066

114. van Winden K, Biljecki F, van der Spek S. Automatic Update of Road Attributes by Mining GPS Tracks. Transactions in GIS. 2016;p. n/a–n/a. doi: 10.1111/tgis.12186

115. Stoter J, Ledoux H, Zlatanova S, Biljecki F. Towards sustainable and clean 3D Geoinformation. In: Kolbe TH, Bill R, Donaubauer A, editors. Geoinformationssysteme 2016: Beiträge zur 3. Münchner GI-Runde. Munich, Germany; 2016. p. 100–113.

116. Chen K. An approach to linking remotely sensed data and areal census data. International Journal of Remote Sensing. 2002 Jan; 23(1):37–48. doi: 10.1080/01431160010014297

117. Schneiderbauer S, Ehrlich D. Population Density Estimations for Disaster Management: Case Study Rural Zimbabwe. In: Geo-information for Disaster Management. Springer Berlin Heidelberg; 2005. p. 901–921.

118. Akbar M, Aliabadi S, Patel R, Watts M. A fully automated and integrated multi-scale forecasting scheme for emergency preparedness. Environmental Modelling & Software. 2013 Jan; 39:24–38.

119. Langford M, Higgs G. Measuring Potential Access to Primary Healthcare Services: The Influence of Alternative Spatial Representations of Population. The Professional Geographer. 2006 Aug; 58(3):294–306. doi: 10.1111/j.1467-9272.2006.00569.x

120. Hay SI, Noor AM, Nelson A, Tatem AJ. The accuracy of human population maps for public health application. Tropical Medicine and International Health. 2005 Oct; 10(10):1073–1086. doi: 10.1111/j. 1365-3156.2005.01487.x PMID: 16185243

121. Poulsen E, Kennedy LW. Using Dasymetric Mapping for Spatially Aggregated Crime Data. Journal of Quantitative Criminology. 2004 Sep; 20(3):243–262. doi: 10.1023/B:JOQC.0000037733.74321.14

122. Lin J, Cromley RG. Evaluating geo-located Twitter data as a control layer for areal interpolation of population. Applied Geography. 2015 Mar; 58:41–47. doi: 10.1016/j.apgeog.2015.01.006

123. Meinel G, Hecht R, Herold H. Analyzing building stock using topographic maps and GIS. Building Research & Information. 2009 Nov; 37(5–6):468–482. doi: 10.1080/09613210903159833

124. Vine MF, Degnan D, Hanchette C. Geographic information systems: their use in environmental epidemiologic research. Environmental Health Perspectives. 1997 Jun; 105(6):598–605. doi: 10.1289/ehp. 97105598 PMID: 9288494

125. Wandl A, Nadin V, Zonneveld W, Rooij R. Beyond urban–rural classifications: Characterising and mapping territories-in-between across Europe. Landscape and Urban Planning. 2014 Oct; 130:50–63. doi: 10.1016/j.landurbplan.2014.06.010

126. de Kluijver H, Stoter J. Noise mapping and GIS: optimising quality and efficiency of noise effect studies. Computers, Environment and Urban Systems. 2003; 27(1):85–102. doi: 10.1016/S0198-9715(01) 00038-2

127. Ögren M, Barregard L. Road Traffic Noise Exposure in Gothenburg 1975–2010. PLOS ONE. 2016 May; 11(5):e0155328. doi: 10.1371/journal.pone.0155328 PMID: 27171440

128. Tutschku K. Demand-based radio network planning of cellular mobile communication systems. In: IEEE INFOCOM'98 Conference on Computer Communications Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. San Francisco, CA, United States: IEEE; 1998. p. 1054–1061.

129. INSPIRE Thematic Working Group Buildings. D2.8.III.2 INSPIRE Data Specification on Buildings—Technical Guidelines; 2013.

130. Kohler N, Hassler U. The building stock as a research object. Building Research & Information. 2002 Jul; 30(4):226–236. doi: 10.1080/09613210110102238

131. Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M. A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment. 2010 Jul; 45(7):1683–1697. doi: 10.1016/j.buildenv.2010.01.021

132. Hargreaves AJ. Representing the dwelling stock as 3D generic tiles estimated from average residential density. Computers, Environment and Urban Systems. 2015 Nov; 54:280–300. doi: 10.1016/j. compenvurbsys.2015.08.001

133. Levy PS, Lemeshow S. Stratification and Stratified Random Sampling. In: Sampling of Populations: Methods and Applications. 4th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008. p. 121–142.

134. van Oosterom P, Martinez-Rubi O, Ivanova M, Horhammer M, Geringer D, Ravada S, et al. Massive point cloud data management: Design, implementation and execution of a point cloud benchmark. Computers and Graphics. 2015 Jun; 49(C):92–125. doi: 10.1016/j.cag.2015.01.007