European Journal of Geography

The publication of the European Journal of Geography (EJG) is based on the European Association of Geographers' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual, or empirical way the quality of research, learning, teaching, and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.

Received: 02/06/2023

Revised: 10/09/2023

Accepted: 25/10/2023

Published: 25/10/2023

DOI: 10.48088/ejg.m.bat.14.4.022.036

ISSN: 1792-1341

Research Article

PoD: A Web Tool for Population Downscaling Using Areal Interpolation and Volunteered Geographic Information

Marios Batsaris[1,*], Sofia Zafeirelli[1], Michail Vaitis[1] & Dimitris Kavroudakis[1]

[1] Department of Geography, University of the Aegean, 81100 Mytilene, Greece

[*] Correspondence: m.batsaris@aegean.gr

Abstract: Population data are commonly sourced from censuses, and to meet confidentiality requirements, they are spatially aggregated into standardized enumeration units. However, the need often arises to transform such datasets into user-defined spatial scales, a process known as areal interpolation. Areal interpolation is the process of data transformation across spatial zones and is particularly suitable for aggregated data such as census data. While numerous areal interpolation methods exist, a lack of implementation tools have been witnessed. In this article, we introduce PoD, a web-based solution that encompasses four downscaling schemes. To illustrate the utility of the proposed tool, we conducted a case study using actual data from the city of Mytilini, Greece. We compared the results obtained through PoD with existing R-based implementations, in addition to evaluating their performance using a reference dataset. The outcomes of this evaluation affirm the effectivenes of the proposed PoD tool over alternative implementations.

Keywords: population downscaling, areal interpolation, web tool, GIS

Highlights:

- Areal interpolation is broadly used to facilitate the conversion of population data across spatial zones.
- A notable aspect concerning the areal interpolation of population data pertains to the identified lack of implementation tools.
- The proposed tool demonstrates higher performance compared to existing alternatives.

1. Introduction

Population data play important role in addressing a wide range of geographical issues (Halder, 2018; Papanikolaou & Mitsi, 2020; Paraskevopoulos et al., 2019; Photis & Sirigos, 2015). Census is broadly used as the main source of population data. To meet statistical confidentiality and privacy requirements, census data are spatially aggregated into enumeration spatial units such as census tracts and/or communes. However, there is often a compelling need to transform these datasets into user-defined spatial scales that deviate from the standardized census units. The process of transforming data between different spatial zones is also known as areal interpolation.

Areal interpolation refers to the process of transforming data from a set of spatial units (source) to another (target) (Mennis, 2003) and is suitable for aggregated data (such as census data) (Lam, 1983). Areal interpolation methods either use ancillary information or entirely depend on source and target data (Wu et al., 2005). Perhaps, the simplest form of areal interpolation is the Areal Weighted Interpolation (AWI) (Fisher & Langford, 1995; Kim & Yao, 2010). AWI may be improved by involving additional information to aid the interpolation process (dasymetric mapping). Dasymetric Mapping in the context of Areal Interpolation have dominated the population downscaling related literature as they significantly improve accuracy (Langford, 2006). Ancillary information may be in any form of a categorical layer map such as Land Use Land Cover (LULC) data (Younes et al., 2023), usually acquired by remotely sensed images (Gervasoni et al., 2019; Karunarathne & Lee, 2019; Liu & Martinez, 2019; Wu et al., 2005).

Mobile phone (Bergroth et al., 2022; Kubíček et al., 2019; Peng et al., 2020) and social media geospatial data (Bao et al., 2023; Lin & Cromley, 2015; Park et al., 2020) have been used as alternative sources of ancillary information for population downscaling. While mobile phone data are not always available for public use, recent studies incorporate new forms of spatial data that emerged from the widespread use of the internet and the evolution of web technologies which have enabled users to generate geographic content online. Michael Goodchild (2007) noticed this trend and defined it as Volunteered Geographic Information (VGI). Open Street Map (OSM) is a typical example of VGI with a considerably large community of volunteers that generate, update, and verify spatial content (Bertolotto et al., 2020). OSM data have been extensively incorporated for population downscaling using areal interpolation schemes (Bakillah et al., 2014; Comber & Zeng, 2019; Gervasoni et al., 2019; Guo et al., 2017).

Downscaling of population data using areal interpolation methods undoubtedly has attracted significant attention during the last few decades. However, advances in areal interpolation have primarily focused on developing methods using modern and sophisticated techniques to improve accuracy rather than developing standardized methods accessible to the Geographic Information Systems (GIS) community (Eicher & Brewer, 2001; Langford, 2007; Mennis, 2009). To the best of our knowledge, only a few implementation tools exist in both open-source and off-the-shelf GIS-enabled software.

Sleeter and Gould (2008) introduce a Dasymetric-Mapping Extension (DME) compatible with ESRI ArcGIS versions 9.1 and 9.2. The extension streamlines dasymetric mapping and automates the areal interpolation process through a dialog box. Developed and modified by Michael Gould of the USGS (United States Geological Survey) using a combination of the methods described in Holloway et al. (1997) and Mennis (2003). DME requires two input datasets, a population layer in polygon shapefile format, regardless of geographical level, and an ancillary layer, land use, or a land cover derived ArcGRID raster, with four population density classes.

Intelligent Dasymetric Mapping (IDM) Toolbox developed by Torrin Hultgren on behalf of the United States Environmental Protection Agency (US EPA - https://www.epa.gov/enviroatlas/dasymetric-toolbox) for ArcGIS versions 10.2 and 10.3. IDM was developed using the Arcpy library and arcpy.da and follows the areal interpolation and dasymetric mapping methods proposed in (Mennis & Hultgren, 2006). Qiu et al. (Qiu et al., 2012) created Areal Interpolation Extension for ArcGIS using VB.NET (Visual Basic.NET) and ArcObjects. There are four methods available namely: a) area-weighting, b) pyhnophylactic, c) binary dasymetric, and 3-class regression dasymetric. Source and target shapefiles (vector) required for all methods and ancillary information for binary and 3-class dasymetric methods. Ancillary data may be either in raster or vector type, such as land use, land cover, and road network.

Areal interpolation alternatives also exist within the R programming ecosystem (R Core Team, 2015) through the Comprehensive R Archive Network (CRAN) repository. Pebesma (Pebesma, 2018) developed sf an R package for spatial vector data support. Among many others, simple areal-weighted interpolation functionality is provided in sf. The sf method is an ancillary information-independent method suitable for completely overlapping source and target data. Prener and Revord introduced areal (Prener & Revord, 2019) which extends the sf package's functionality by providing an additional formula suitable for data without complete overlap. Finally, a recent addition named populR (Batsaris & Kavroudakis, 2021) extends areal functionality by introducing a new method that uses a volume-weighted approach.

The use of the existing tools encounters several limitations. One of the main limitations of using off-the-shelf software is that it requires special software and a commercial license to be able to work. ArcGIS extensions are also developed to be used with specific versions of the software and therefore, may not be supported in newer/older software releases. In addition, ancillary data are required for dasymetric interpolation. Moreover, implementations in R, are publicly and freely available for population downscaling but require special software such as R programming language kit, an IDE (Integrated Development Environment), and programming skills.

In this article a web tool named PoD is introduced as an alternative implementation in the context of areal interpolation of population data. PoD has the capacity to operate effectively with minimal data input requirements (source and target data). The contribution of this article focuses on three main pillars. First, to overcome the limitations of using existing implementations (both off-the-shelf and open source), PoD emerged as an addition to the domain of areal interpolation of population data. Second, a comparative analysis of existing R implmentations is carried and finally, as well as it provides four interpolation methods of which one is a novel approach by harnessing VGI from OpenStreetMap (OSM).

The remainder of this paper is structured as follows. First, we provide an extended description of the PoD web tool including architecture, workflow, VGI, and the available methods. Then, numerical experiments and accuracy assessment based on actual data from the city of Mytilini, Lesvos are presented and finally, the conclusions section summarizes the study's key findings and insights while also suggesting directions for further improvements and research.

2. Materials and Methods

2.1 PoD Web Tool

PoD is a web tool focusing on population downscaling using areal interpolation methods and VGI, through a ready-to-work and easy-to-use UI with point-and-click features and minimum user inputs. PoD functionality adopts the areal interpolation schemes of the latest version (v0.2.1) of populR package (Batsaris, 2021) and is composed of a Data Upload module and four downscaling method modules namely: 1) AWI (Areal Weighted Interpolation), 2) VWI (Volume Weighted Interpolation), 3) BDI (Binary Dasymetric Interpolation) and 4) FDI (Float Dasymetric Interpolation).

2.1.1. Architecture

PoD was developed using the R programming language (R Core Team, 2015) extended by shiny (RStudio, 2013), a package for building interactive web apps. As an app developed with shiny, PoD has a common client-server architecture as shown in Figure 1. It consists of a user interface script (UI) that controls the layout and appearance, and a server script responsible for the logic and reactivity of the app. R session is used to facilitate the interaction between these elements. User inputs, trigger updates in the reactive expressions, which are subsequently translated into outputs that update the UI accordingly.

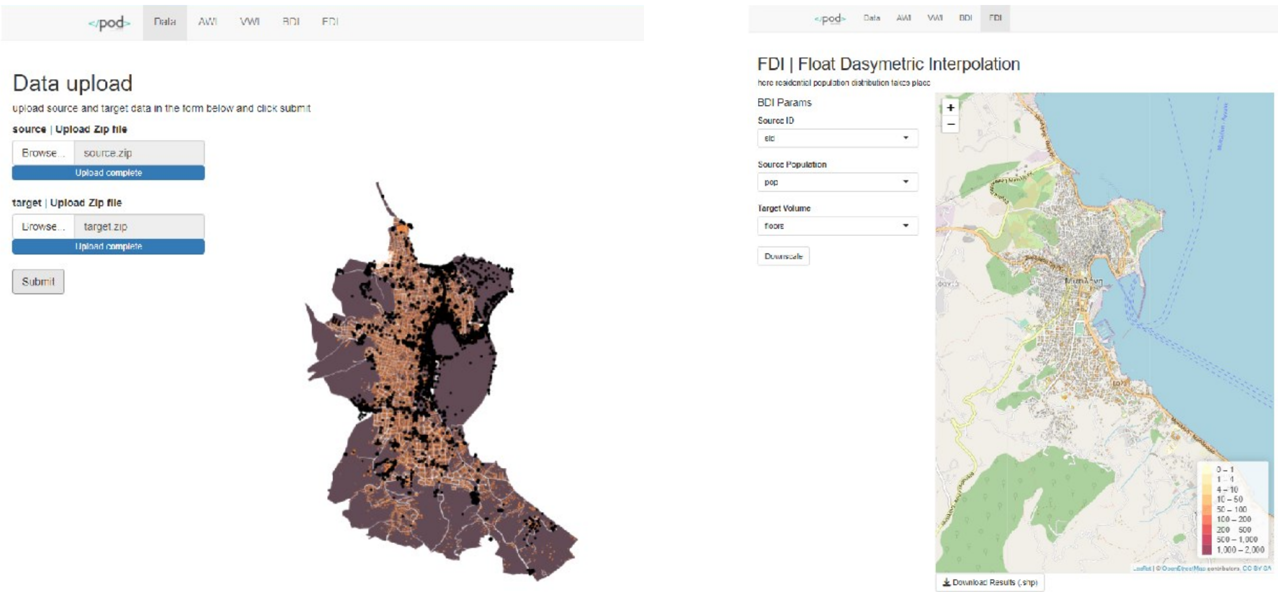
Figure 1. PoD web tool architecture schema

2.1.2. Workflow

Figure 2 presents screens of the PoD web tool. Data upload is a prerequisite. The user uploads source and target geojson files via the data upload module. PoD, then, extracts VGI from the Open Street Map © (OSM) collaborative database right after data upload. PoD plots user inputs

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

and OSM data (Figure 2a). After data upload, users may proceed with population downscaling modules. In each module, users must specify parameters such as source identification number, source population values, and target volume information for VWI, BDI (optional), and FDI. The downscaling results are visualized in interactive leaflet maps (Cheng et al., 2019) and may be downloaded in geojson format for further analysis (Figure 2b).



(a)

(b)

Figure 2. PoD web tool main screens. a) data upload module where user inputs and extracted OSM data are also visualized, b) fdi module in which the user selects the desired parameters from the uploaded data calculates the results, and visualizes them in an interactive leaflet map. The results may be downloaded in ESRI Shapefile format for further analysis.

2.2 VGI

With the growth of user-generated spatial content on platforms such as OSM, VGI has grown dramatically. However, this excessive amount of VGI comes with several notable limitations. First, the quality and accuracy of VGI can vary significantly as it relies on a diverse group of users. Inaccuracies and inconsistencies are common, as contributors may lack geographic expertise or use different mapping tools and techniques. Second, it is important to acknowledge that VGI frequently manifests geographic and thematic biases, in which specific regions or subjects receive excessive attention than others e.g. densely populated urban areas often receive more contributions compared to remote or less populated regions. Lastly, concerns about data currency and relevance emerge because VGI may not always capture recent changes in the real world.

PoD obtains point and polygon data from the Open Street Map © (OSM) collaborative spatial database under the Open Data Commons Open Database License (OdbL) of the OSM Foundation through the osmdata (Padgham et al., 2017) extension right after data upload. The osmdata package downloads OSM map features through queries from the Overpass API (https://www.overpass-api.de/). Each query consists of a bounding box of the study area, the OSM map features to be extracted, and, a data conversion command from XML to simple features (sf) (Pebesma, 2018).

After extracting VGI, PoD then converts the polygon data into points using the center of gravity (centroid) and combines them into a new set of point data. The new set of point data is then used to count points over target polygons to be used as ancillary information for BDI and FDI modules.

2.3 Methods

2.3.1. Areal Weighted Interpolation (AWI)

AWI has an easy and straightforward implementation using simple polygon overlay operations and simple calculations in a GIS (Comber & Zeng, 2019; Qiu et al., 2012). Given a set of city block polygons (s - source) accompanied by population values (v) and an incongruent yet superimposed set of buildings (t - target), AWI proportionately interpolates city block population values based on areal weights calculated by the area of intersection between source and target features. Algebraically may be expressed by equations 1 and 2 (Goodchild & Siu-Ngan Lam, 1980).

$$w_{ts} = \frac{a_{ts}}{\sum_{t \in s} a_{ts}} \tag{1}$$

$$p_{ts} = w_{ts} \times v_s \tag{2}$$

Where: wts: areal weight of t belongs to s
ats: area of t belongs to s
pts: estimated population of t belongs to s
vs: population of s

The main advantage of AWI is that preserves the initial source population values when summarized, however, it assumes homogeneity within the source zone features (Goodchild & Siu-Ngan Lam, 1980; Lam, 1983).

2.3.2. Volume Weighted Interpolation (VWI)

Given the number of floors (n) as additional target information (volume), VWI interpolates population values using volume weights measured by the area of intersection between source and target features multiplied by the number of floors (Lwin & Murayama, 2009) as shown in equations 3 and 4.

$$w_{ts} = \frac{a_{ts} \times n_{ts}}{\sum_{t \in s}(a_{ts} \times n_{ts})} \tag{3}$$

$$p_{ts} = w_{ts} \times v_s \tag{4}$$

Where: wts: volume weight of t belongs to s
ats: area of t belongs to s
nts: number of floors of t belongs to s
pts: estimated population of t belongs to s
vs: population of s

2.3.1. Binary Dasymetric Interpolation (BDI)

BDI is a basic type of dasymetric mapping method (Zandbergen & Ignizio, 2010). Suppose an additional dataset (u), of point data, is retrieved by the OSM spatial database. This dataset is used as a binary mask to highlight populated and unpopulated buildings upon intersection. Unpopulated features are excluded, while areal weights are calculated in populated features (Langford, 2013). Then, BDI may be expressed by equations 5, 6, and 7.

$$w_{ts} = \frac{a_{ts} \times u_{ts}}{\sum_{t \in s}(a_{ts} \times u_{ts})} \tag{5}$$

$$p_{ts} = w_{ts} \times v_s \tag{6}$$

$$u \in (0,1) \begin{matrix} 0 & unpopulated \\ 1 & populated \end{matrix} \tag{7}$$

Alternatively, if volume information is available, then equation 5 may be modified as follows in equation 8.

$$w_{ts} = \frac{a_{ts} \times u_{ts} \times n_{ts}}{\sum_{t \in s}(a_{ts} \times u_{ts} \times n_{ts})} \tag{8}$$

Where: wts: volume weight of t belongs to s
ats: area of t belongs to s
uts: binary value of t belongs to s
nts: number of floors of t belongs to s
pts: estimated population of t belongs to s
vs: population of s

2.3.1. Float Dasymetric Interpolation (FDI)

Consider another dataset (u), of OSM point data. FDI uses this dataset to measure the residential occupancy rate of target features upon intersection. This is a two-step procedure. The first step is to count OSM points over target polygons (c) and divide them by the number of floors (non-residential occupancy rate). Then, the results of the first step are subtracted by 1 to calculate the residential occupancy rate. The residential occupancy rate ranges between 0 and 1, where 0 stands for buildings with non-residential use, 0.1 to 0.9 for both residential and non-residential (mixed) buildings, and 1 for buildings completely used as residencies. FDI is expressed by equations 9 to 11.

$$u_{ts} = 1 - \frac{c_{ts}}{n_{ts}} \tag{9}$$

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

$$w_{ts} = \frac{a_{ts} \times u_{ts} \times n_{ts}}{\sum_{t \in s}(a_{ts} \times u_{ts} \times n_{ts})} \tag{10}$$

$$p_{ts} = w_{ts} \times v_s \tag{11}$$

Where: uts: residential occupancy of t belong to s
cts: count of VGI points over t belongs to s
wts: volume weight of t belongs to s
ats: area of t belongs to s
nts: number of floors of t belongs to s
pts: estimated population of t belongs to s
vs: population of s

3. Numerical Experiments

3.1. Study Area

Numerical experiments were carried out using the city of Mytilini as the case study. Mytilini is a medium-sized Greek city located in the southeast part of Lesvos Island (Figure 3) and is the capital of both Lesvos Prefecture and the North Aegean Region. Mytilini is an important transportation hub as it is the main port as well as the main airport of the Island of Lesvos.

Figure 3. Study area and its location in different scales.

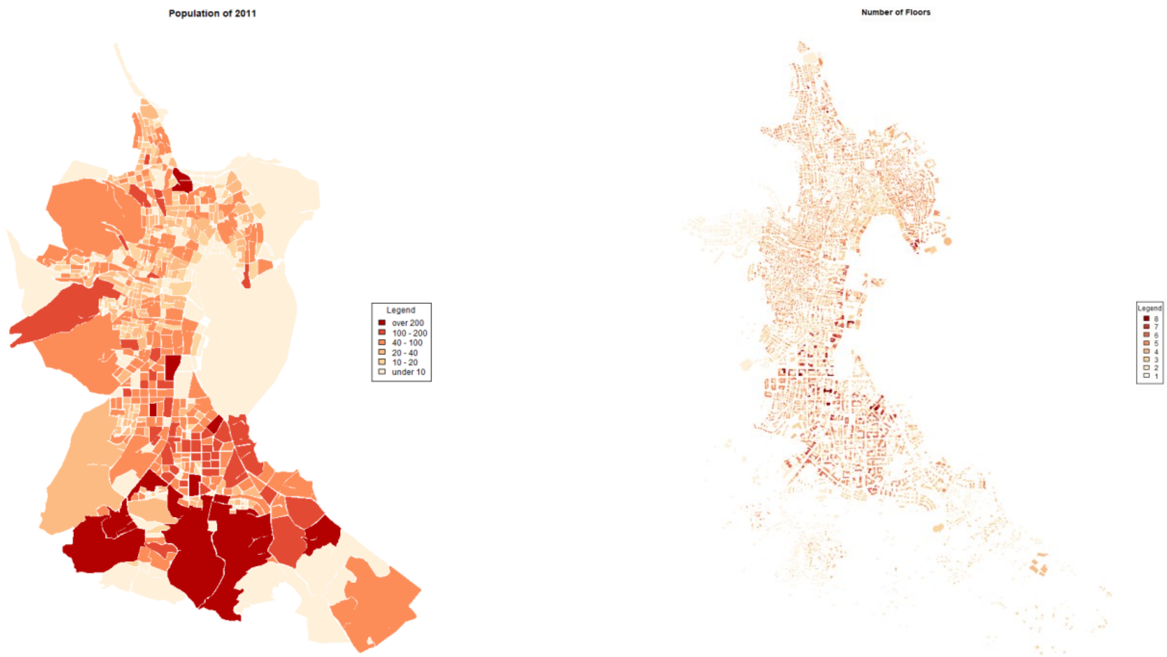
3.2. Data

The data used in this study were provided for academic use by the Hellenic Statistical Authority (ELSTAT) and referred to 2011 (Hellenic Statistical Authority, 2014) and 2001 Population and Housing Census of Greece (Hellenic Statistical Authority, 2009). ELSTAT provided 2011 (de facto) population data aggregated in the city block level of granularity in spreadsheet format. Spatial datasets were also provided. City blocks and building units were retrieved in vector spatial format (ESRI Shapefile - .shp) and correspond to the 2011 Census and 2001 Census accordingly. The study area counts 27,871 residents and is composed of 673 city block polygons (source data) and 11,744 building units (target data). Source and target data are visually presented in Figure 4.

Figure 4. Source (brown) and target (orange) data provided by ELSTAT for the city of Mytilini.

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

Spreadsheet data were joined to the source attribute table using the common city block identification field. As a result, 558 out of 673 city blocks are populated with 25,699 residents because of data inconsistencies, and the spatial distribution of the population is depicted in Figure 5a. Most of the population is concentrated in the southern part of the city, while the northern part of the city is somewhat sparsely populated. In addition, the target dataset consists of more than 11,500 building polygons ranging from 1 to 8 floors. The vast majority of the buildings have up to 3 floors as shown in Figure 6. High-rise buildings (more than 3 floors) appear in the southern part of the city as shown in Figure 5b.



(a) (b)

Figure 5. a) Spatial distribution of the population in the city block level of detail, and b) Building unit number of floors distribution.

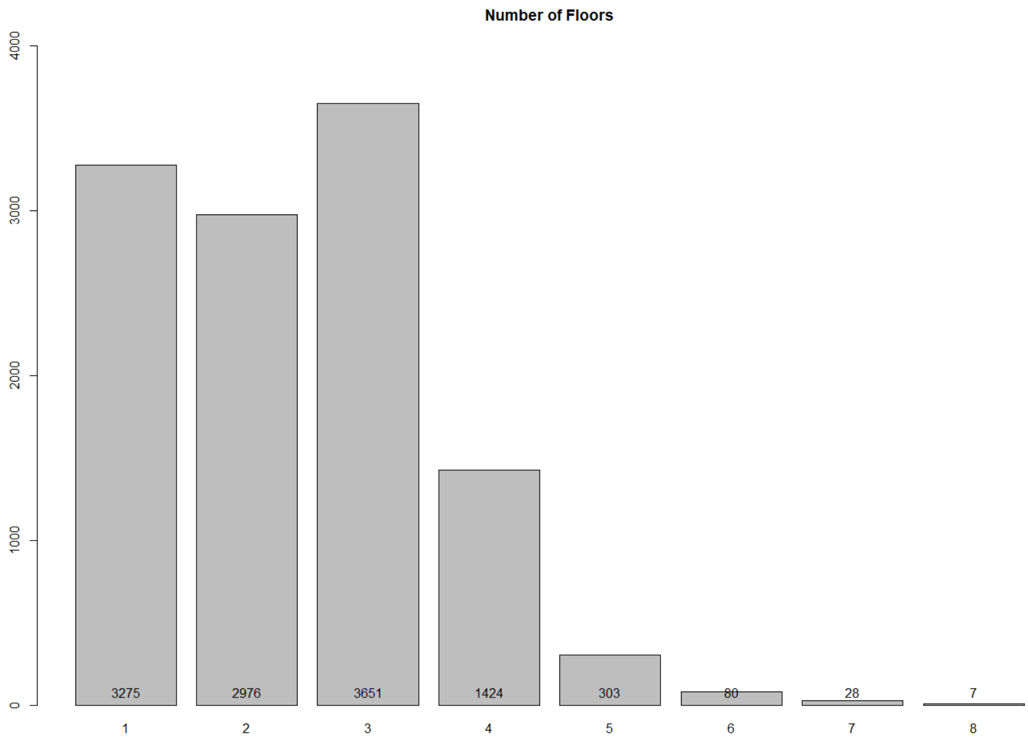


Figure 6. Number of floors for the buildings of the study area.

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

Additional information required for BDI and FDI methods is extracted by the OSM open spatial database upon data upload. OSM provides a large number of map features (OSM Contributors, 2023), however, PoD only uses a subset of point and polygon geometries that are likely to be related to the population's activities. More than 5,000 point and polygon geometries were extracted by the OSM database from the following map features: 1) amenity, 2) clothes, 3) healthcare, 4) leisure, 5) military, 6) office 7) religion, 8) shop, 9) social_facility and 10) tourism. Next, polygons were converted into points using the centroid and combined into a new point dataset. Finally, PoD count points over polygons using overlay operations. As a result, 1.340 OSM points were found over 865 buildings as shown in Figure 7.

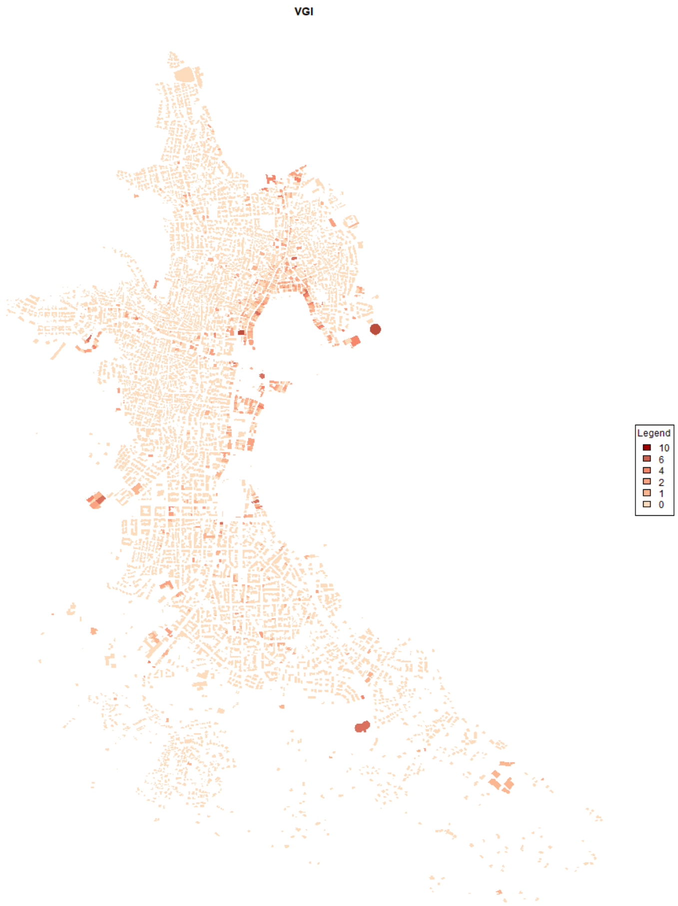


Figure 7. VGI distribution in the buildings of Mytilini.

3.3. Results and Evaluation

In this sub-section, PoD web tool and open-source (areal and sf) implementations are demonstrated and evaluated using a reference dataset from our previous work (Batsaris et al., 2019).

3.3.1. Results

Open-source alternatives include areal-weighted interpolation functionality. The sf package areal interpolation functionality proportionately interpolates source population values using areal weights measured using the total area of source polygons (total weights). The areal package provides two areal-weighted interpolation formulas. The first is similar to sf and the second calculates areal weights by using the sum of the remaining areas after the intersection (sum weights).

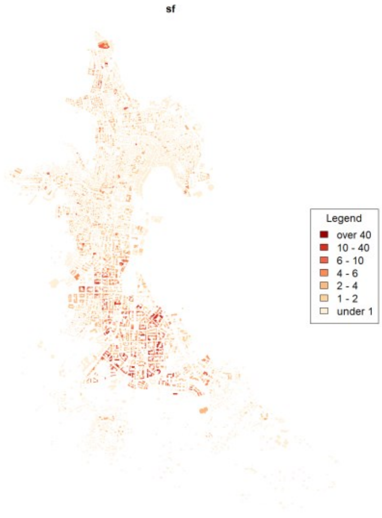
For the reader's convenience method names are shortened as follows.

- sf: sf implementation using total weights (sf)
- awt: areal implementation using total weights (areal)
- aws: areal implementation using sum weights (areal)
- awi: Areal Weighted Interpolation (PoD)
- vwi: Volume Weighted Interpolation (PoD)
- bdia: Binary Daymetric Interpolation using Areal Weights (PoD)
- bdiv: Binary Daymetric Interpolation using Volume Weights (PoD)
- fdi: Float Dasymetric Interpolation (PoD)

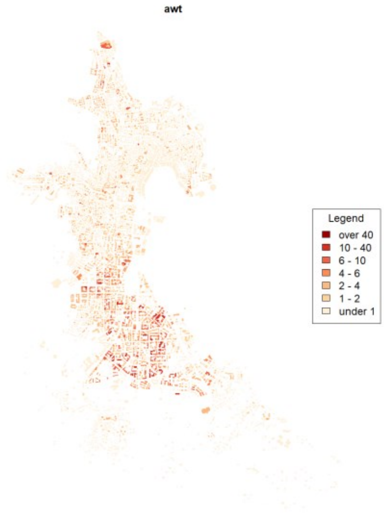
sf, awt, aws, awi and vwi entirely depend on source and target data while bdia, bdiv and fdi use VGI as ancillary information. vwi, bdiv and fdi exploit building volume information and bdia, bdiv and fdi exploit VGI extracted by the OSM spatial database. The implementation results are visually presented in Figure 8. The patterns of population distribution initially appear to be quite similar. Most of the population is concentrated in the southern part of the city while the northern part is somewhat sparsely populated. Slight differences appear when the results are closely

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

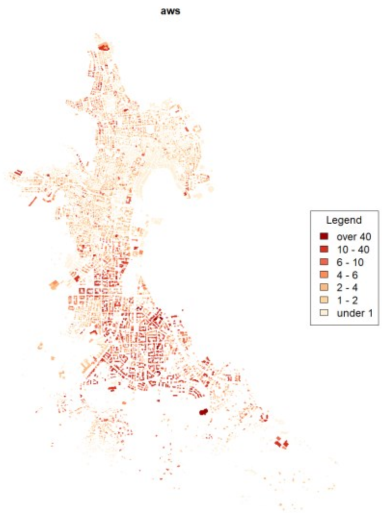
examined. These differences are the result of binary and float constraints as well as the impact of the number of floors during the interpolation process.



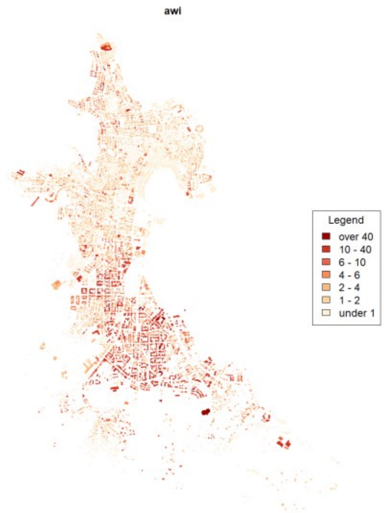
(a)



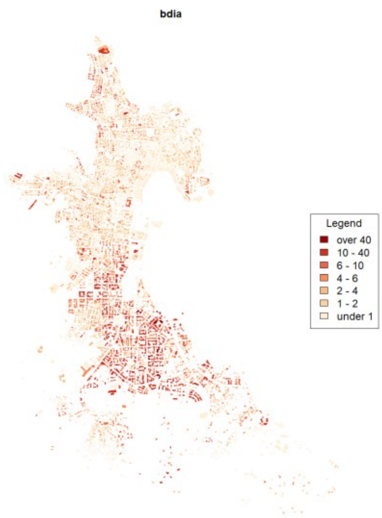
(b)



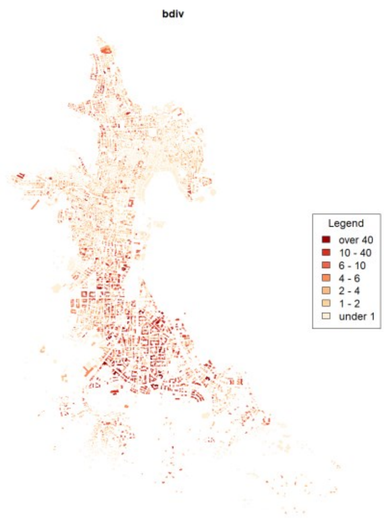
(c)



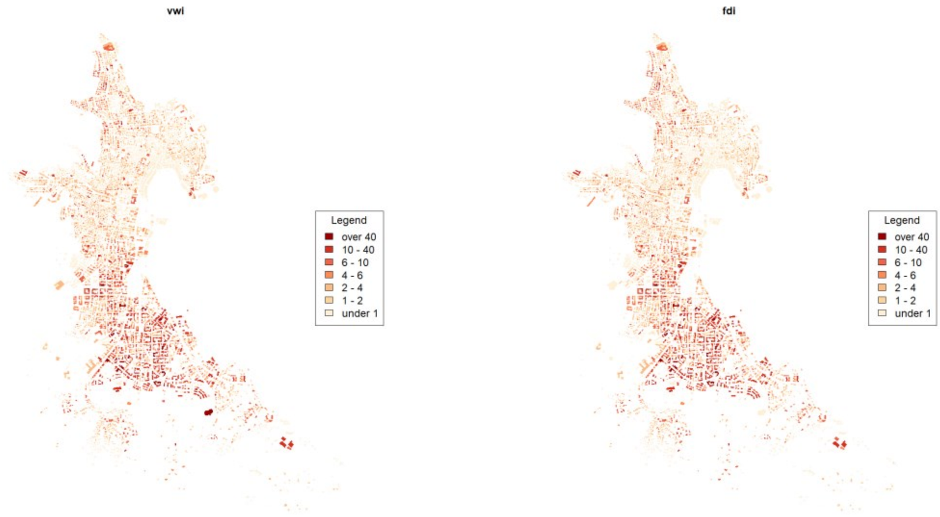
(d)



(e)



(f)



(g) (h)

Figure 8. Results obtained by the implementation of the PoD web tool and open-source alternatives. a) sf, b) awt, c) aws, d) awi, e) bdia, f) bdiv, g) vwi, h) fdi.

As mentioned in the previous section, sf and awt are suitable for completely overlapping data and thus, they significantly underestimate the total population while the rest of the methods preserve the total population of the study area as shown in Table 1. The estimated results were also summarized for each city block and the relationship with the initial source population values was further explored. Scatterplots presented in Figure 9 provide evidence that apart from sf and awt, all methods preserve the initial city block population values when summarized.

Table 1. Comparison of the total estimated population to the total population of the study area.

Method	Total Estimated Population	Total Population
sf	10.271,78	25.699
awt	10.271,78	25.699
aws	25.699	25.699
awi	25.699	25.699
bdia	25.699	25.699
bdiv	25.699	25.699
vwi	25.699	25.699
fdi	25.699	25.699

(a) (b)

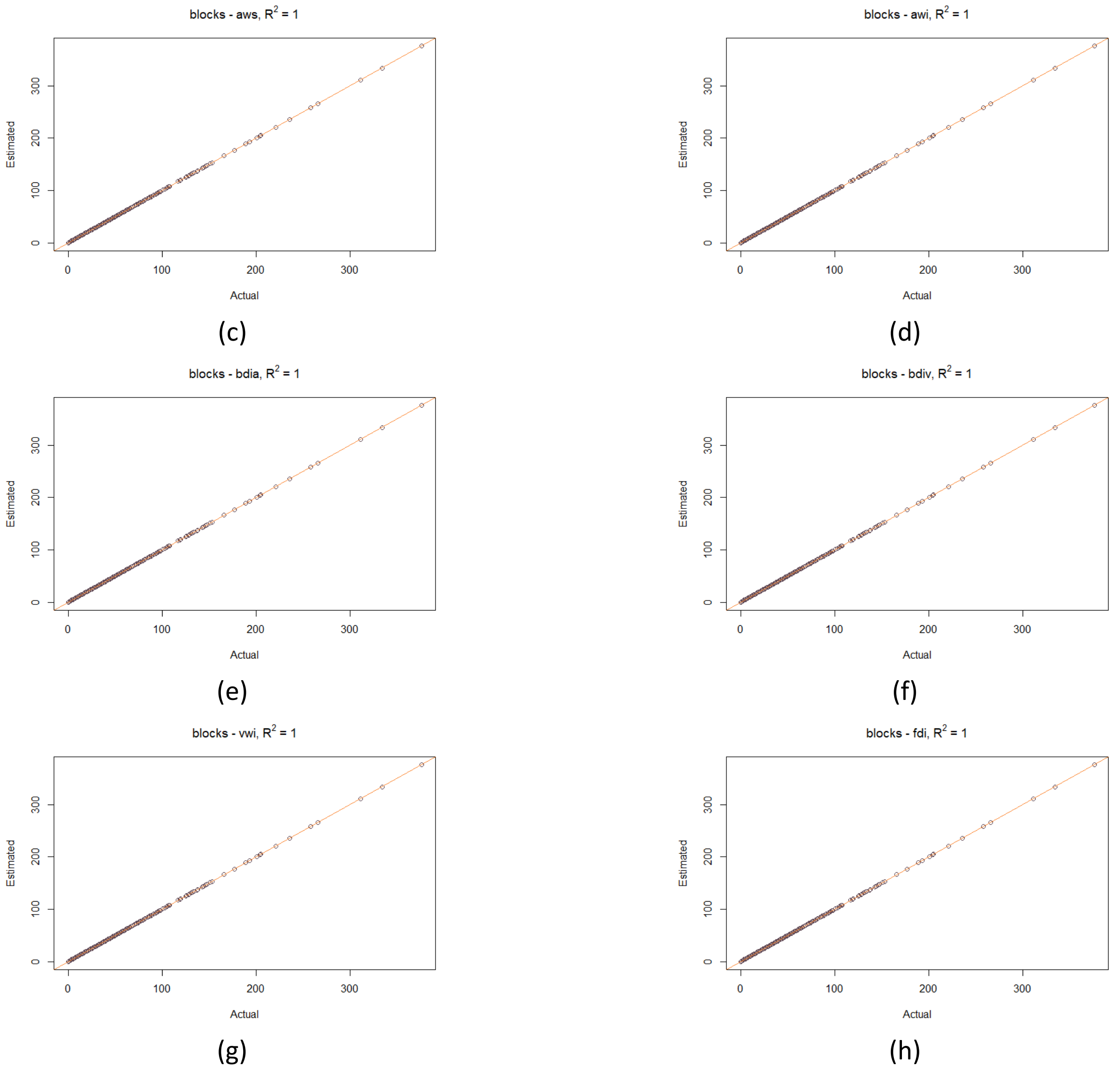


Figure 9. Scatterplots of summarized target estimates in each city block (estimated – y-axis) and initial source population values (actual – x-axis). a) sf, b) awt, c) aws, d) awi, e) bdia, f) bdiv, g) vwi, and h) fdi.

Table 2 presents the estimated values for ten randomly selected buildings of the study area and indicates that sf and awt provide the same results. The same results are also obtained by aws and awi methods, while the rest provide somehow different estimated values.

Table 1. Comparison of the results of the PoD web tool with open-source implementations for 10 randomly selected buildings.

sf	awt	aws	awi	bdia	bdiv	vwi	fdi
0.058	0.058	1.840	1.840	1.840	1.372	1.372	1.372
1.134	1.134	2.405	2.405	2.641	1.038	0.900	0.931
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.421	0.421	0.925	0.925	0.925	1.095	1.095	1.095
0.152	0.152	0.340	0.340	0.514	0.699	0.458	0.560
2.110	2.110	4.588	4.588	4.588	1.546	1.546	1.546
1.146	1.146	2.172	2.172	2.172	0.855	0.855	0.855
0.612	0.612	1.558	1.558	1.581	1.633	1.591	1.600
0.182	0.182	1.888	1.888	1.888	2.862	2.862	2.862
0.443	0.443	0.946	0.946	0.946	0.775	0.775	0.775

3.3.2. Evaluation

Since no official fine-scale population data exist, results obtained by the PoD web tool and open-source tools were further evaluated using a population dataset from our earlier research (Batsaris et al., 2019). This dataset (reference dataset) is based on an extended field survey carried out back in 2010 by the Carto-GI Lab (https://cartogi-lab.aegean.gr/), highlighting populated and unpopulated floors for the buildings of Mytilini, and were validated through random in-situ visits in 2019. Then, the VWI approach is employed for populated floors to obtain population estimates at the building level of detail. The population distribution of the reference dataset (rf) is presented in Figure 10.

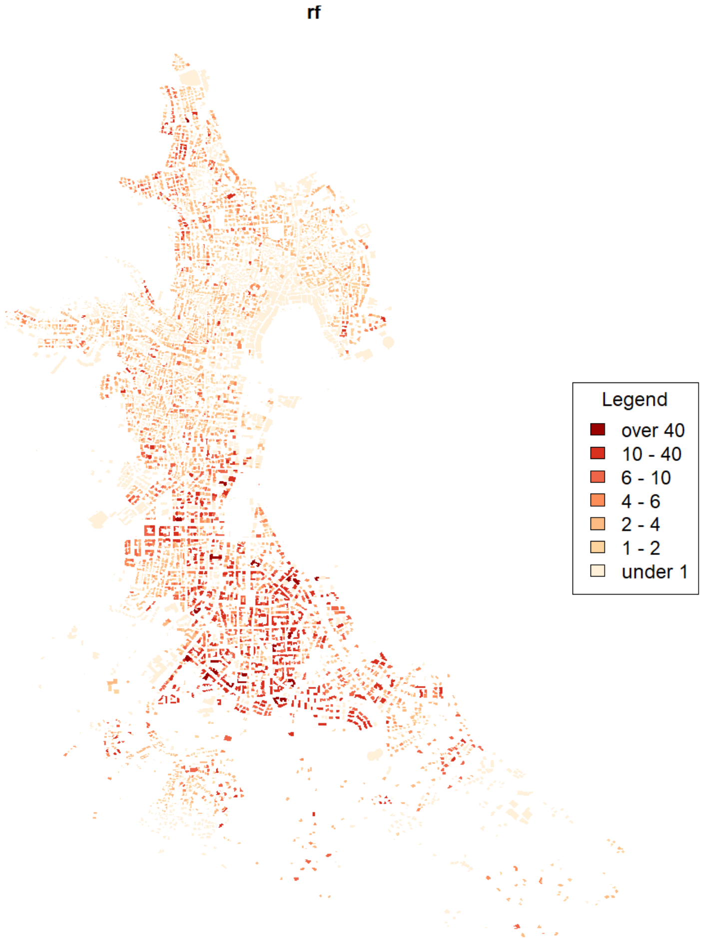
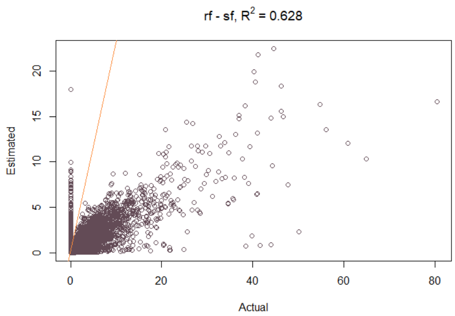
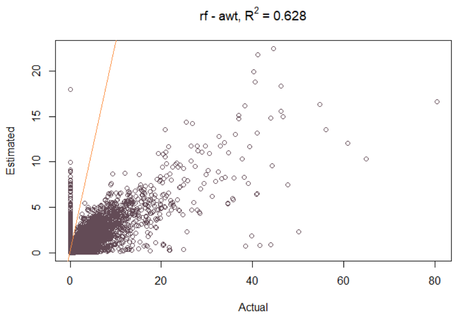


Figure 10. Reference dataset population distribution.

Linear Regression Analysis was carried out to investigate the statistical relationship between rf data and the obtained results (Figure 11). Linear regression R^2 was used to quantify the number of actual (rf) values explained by the results retrieved by the PoD web tool and open-source alternatives



(a)



(b)

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

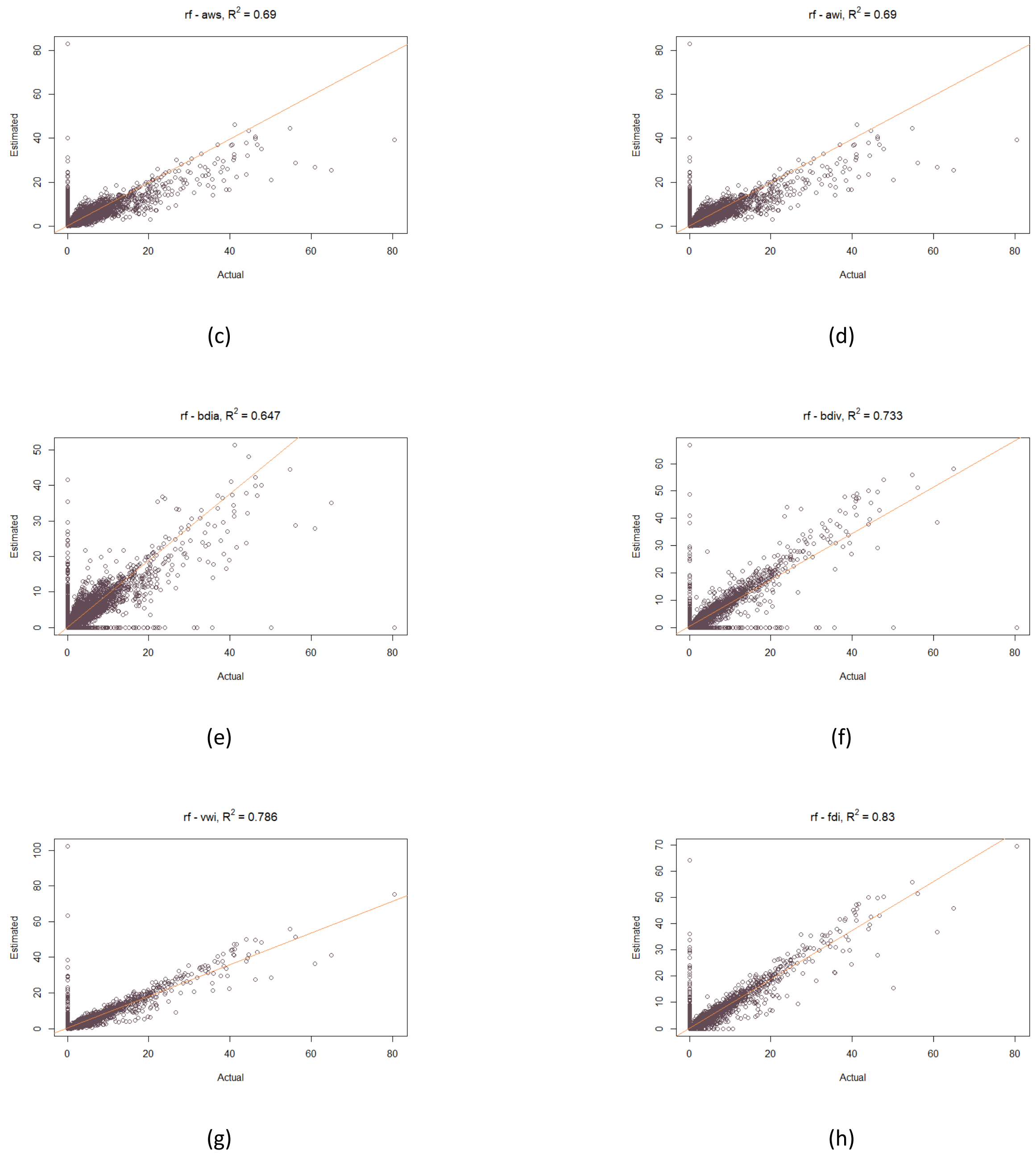


Figure 11. Scatterplots to investigate the relationship between rf and the implemented methods. a) sf, b) awt, c) aws, d) awi, e) bdia, f) bdiv, g) vwi and h) fdi.

There are multiple points identified out of the general trend especially in methods using areal weights (sf, awt, aws, awi and bdia). Outlier values in the y-axis represent buildings that were excluded from the rf dataset due to their status (abandoned/unused), whereas the values assigned by the demonstrated implementations are incorrect. Extreme values can also be found on the x-axis, most notably in bdia and bdiv results, but less so in the fdi results. These values apply to buildings that have been labeled as unoccupied by the binary constraints or the residential occupancy rate, but not in the rf dataset.

Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were also calculated and along with R2 are shown in Table 2. As expected, sf and awt provide the biggest error values and smaller R2 while the best RMSE, MAE, and R^2 values are achieved by fdi and vwi methods.

Using minimal input and several automated areal interpolation schemes, the PoD tool produced reasonable results. One of the most significant limitations is the availability of VGI in the study area. The availability of OSM data is generally good. However, availability can vary depending on many factors, including location and level of community involvement. In many cases, OSM data is incomplete or out of date, whereas in others it is detailed and accurate. The availability of OSM data may have a significant impact on the outcomes of population downscaling. Less OSM data yields results similar to awi and vwi, whereas more OSM data yields very accurate results. For example, in the case of fdi, less OSM data produces results that are quite similar to vwi, while more OSM points may produce results that are similar to rf.

https://doi.org/10.48088/ejg.m.bat.14.4.022.036

Table 1. MAE, RMSE, and R2 values.

method	MAE	RMSE	R^2
sf	1.49	3.35	0.628
awt	1.49	3.35	0.628
aws	0.85	2.27	0.690
awi	0.85	2.27	0.690
bdia	0.88	2.43	0.647
bdiv	0.55	2.19	0.733
vwi	0.50	1.93	0.786
fdi	0.48	1.70	0.830

4. Conclusions

Population downscaling using areal interpolation methods has been extensively used in a wide range of geospatial applications. However, evidence shows that only a few implementation tools exist in off-the-shelf and open-source GIS-enabled tools. In this study, PoD is proposed as a promising addition to the existing population downscaling implementation tools. PoD is a web application that automates the downscaling process through a simple UI with minimum user input and is accessible from modern internet browsers via https://mbatsaris.shinyapps.io/pod_tool/. PoD introduces four areal interpolation schemes namely: Areal weighted interpolation (awi), Volume Weighted Interpolation (vwi), Binary Dasymetric Interpolation (bdi) and Float Dasymetric Interpolation (fdi). Fdi is a novel approach that utilizes VGI from OSM, offering a whole new perspective on population downscaling.

The results of the PoD web tool were further compared to alternative implementations of the areal and sf packages in R programming environment. In detail, vwi and fdi approaches provide the best results by achieving the smallest error measurements and highest R^2 values. As a consequence, the number of floors positively affects the interpolation process. Furthermore, VGI plays an important role in achieving better results, especially when using the fdi approach. Nonetheless, the availability and reliability of VGI are sometimes questionable as it heavily depends on the involvement of the community which may vary in different locations. However, PoD produces reasonable results in comparison to alternative implementations in R.

Conclusively, this work makes the particular contributions listed below:

- Provides a freely accessible, ready-to-work, easy-to-use, and standardized solution for areal interpolation of population data with minimum user input and point-and-click features suitable either for experts or novice users
- Removes the complex stage of ancillary information collection and pre-processing by automated extraction of VGI from the OSM database upon data upload

Yet, there is one major limitation:

- The outcomes of the PoD web tool strongly depend on the availability and reliability of OSM data which can vary between different locations

To improve the results of the proposed web tool additional sources of ancillary information such as social media or other sources of open data may be also incorporated along with VGI. Furthermore, the PoD web tool was demonstrated using an example using residential population data. However, temporal variations of urban population distribution may be helpful in many applications (Pajares et al., 2021) and therefore, future improvements of the PoD tool should focus on this direction.

Conflicts of Interest: The author declares no conflict of interest.

References

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

Bao, W., Gong, A., Zhang, T., Zhao, Y., Li, B., & Chen, S. (2023). Mapping Population Distribution with High Spatiotemporal Resolution in Beijing Using Baidu Heat Map Data. *Remote Sensing*, *15*(2), 1–22. https://doi.org/10.3390/rs15020458

Batsaris, M. (2021). *populR: Population Down-Scaling in R*. CRAN. https://cran.r-project.org/web/packages/populR/

Batsaris, M., & Kavroudakis, D. (2021). populR: an R Package for Population Downscaling. *The R Journal*, *14*(December), 223–234. https://doi.org/https://doi.org/10.32614/RJ-2023-007

Batsaris, M., Kavroudakis, D., Soulakellis, N. A., & Kontos, T. (2019). Location-Allocation Modeling for Emergency Evacuation Planning in a Smart City Context. *International Journal of Applied Geospatial Research*, *10*(4), 28–43. https://doi.org/10.4018/ijagr.2019100103

Bergroth, C., Järv, O., Tenkanen, H., Manninen, M., & Toivonen, T. (2022). A 24-hour population distribution dataset based on mobile phone data from Helsinki Metropolitan Area, Finland. *Scientific Data*, *9*(1), 1–19. https://doi.org/10.1038/s41597-021-01113-4

Bertolotto, M., McArdle, G., & Schoen-Phelan, B. (2020). Volunteered and crowdsourced geographic information: The openstreetmap project. *Journal of Spatial Information Science*, *20*(20), 65–70. https://doi.org/10.5311/JOSIS.2020.20.659

Calka, B., Nowak Da Costa, J., & Bielecka, E. (2017). Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk*, *8*(2), 1440–1455. https://doi.org/10.1080/19475705.2017.1345792

Cheng, J., Karambelkar, B., Xie, Y., Wickham, H., Russell, K., Schloerke, B., Agafonkin, V., Copeland, B., Dietrich, J., Besquet, B., AS, N., Voogdt, L., Montague, D., AB, K., Kajic, R., Bostock, M., Contributors, jQuery F. and, Contributors, L., CloudMade, … RStudioTeam. (2019). *leaflet*. CRAN. https://cran.r-project.org/web/packages/leaflet/leaflet.pdf
Comber, A., & Zeng, W. (2019). Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. *Geography Compass*, *13*(10), 1–23. https://doi.org/10.1111/gec3.12465
Eicher, C. L., & Brewer, C. A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, *28*(2), 125–138. https://doi.org/10.1559/152304001782173727
Fisher, P. F., & Langford, M. (1995). Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment & Planning A*, *27*(2), 211–224. https://doi.org/10.1068/a270211
Freire, S., & Aubrecht, C. (2012). Integrating population dynamics into mapping human exposure to seismic hazard. *Natural Hazards and Earth System Science*, *12*(11), 3533–3543. https://doi.org/10.5194/nhess-12-3533-2012
Gervasoni, L., Fenet, S., Perrier, R., & Sturm, P. (2019). Convolutional neural networks for disaggregated population mapping using open data. *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, 594–603. https://doi.org/10.1109/DSAA.2018.00076
Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y
Goodchild, M. F., & Siu-Ngan Lam, N. (1980). Areal interpolation: a variant of the traditional spatial problem. *Geo-Processing*, *1*(3), 297–312.
Guo, H., Cao, K., & Wang, P. (2017). Population estimation in Singapore based on remote sensing and open data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2W7), 1181–1187. https://doi.org/10.5194/isprs-archives-XLII-2-W7-1181-2017
Halder, J. C. (2018). Population change and land use dynamics: A case study of Paschim Medinipur District, West Bengal, India. *European Journal of Geography*, *9*(3), 23–44.
Hellenic Statistical Authority. (2009). *2001 Population and Housing Census of Greece*.
Hellenic Statistical Authority. (2014). *2011 Population and Housing Census of Greece* (Issue April). https://www.statistics.gr/en/2011-census-pop-hous
Holloway, S. R., Schumacher, J., & Redmond, R. L. (1997). People and place: dasymetric mapping using Arc/Info. *Cartographic Design Using ArcView and Arc/Info*, 1–11.
Karunarathne, A., & Lee, G. (2019). Estimating hilly areas population using a dasymetric mapping approach: A case of Sri Lanka's highest mountain range. *ISPRS International Journal of Geo-Information*, *8*(4). https://doi.org/10.3390/ijgi8040166
Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: Integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, *31*(21), 5657–5671. https://doi.org/10.1080/01431161.2010.496805
Kubíček, P., Konečný, M., Stachoň, Z., Shen, J., Herman, L., Řezník, T., Staněk, K., Štampach, R., & Leitgeb, Š. (2019). Population distribution modelling at fine spatio-temporal scale based on mobile phone data. *International Journal of Digital Earth*, *12*(11), 1319–1340. https://doi.org/10.1080/17538947.2018.1548654
Lam, N. S. N. (1983). Spatial interpolation methods: A review. *The American Cartographer*, *10*(2), 129–150. https://doi.org/10.1559/152304083783914958
Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, *30*(2), 161–180. https://doi.org/10.1016/j.compenvurbsys.2004.07.001
Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, *31*(1), 19–32. https://doi.org/10.1016/j.compenvurbsys.2005.07.005
Langford, M. (2013). An evaluation of small area population estimation techniques using open access ancillary data. *Geographical Analysis*, *45*(3), 324–344. https://doi.org/10.1111/gean.12012
Lin, J., & Cromley, R. G. (2015). Evaluating geo-located Twitter data as a control layer for areal interpolation of population. *Applied Geography*, *58*, 41–47. https://doi.org/10.1016/j.apgeog.2015.01.006
Liu, X., & Martinez, A. (2019). Areal Interpolation Using Parcel and Census Data in Highly Developed Urban Environments. *ISPRS International Journal of Geo-Information*, *8*(7), 302. https://doi.org/10.3390/ijgi8070302
Lwin, K. K., & Murayama, Y. (2009). A GIS approach to estimation of building population for micro-spatial analysis. *Transactions in GIS*, *13*(4), 401–414. https://doi.org/10.1111/j.1467-9671.2009.01171.x
Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping*. *The Professional Geographer*, *55*(1), 31–42. https://doi.org/10.1111/0033-0124.10042
Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass*, *3*(2), 727–745. https://doi.org/10.1111/j.1749-8198.2009.00220.x
Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, *33*(3), 179–194. https://doi.org/10.1559/152304006779077309
Openshaw, S. (1984). The modifiable areal unit problem. In *Concepts and Techniques in Modern Geography* (CATMOG 38). Geo Books.
OSM Contributors. (2023). *OSM Map Features*. https://wiki.openstreetmap.org/wiki/Map_features
Padgham, M., Lovelace, R., Salmon, M., & Rudis, B. (2017). Osmdata. *The Journal of Open Source Software*, *2*(14), 305. https://doi.org/10.21105/joss.00305
Pajares, E., Nieto, R. M., Meng, L., & Wulfhorst, G. (2021). Population disaggregation on the building level based on outdated census data. *ISPRS International Journal of Geo-Information*, *10*(10). https://doi.org/10.3390/ijgi10100662
Papanikolaou, P. V., & Mitsi, T. K. (2020). Analysis of population dynamics of the regional unit of Chania using remote sensing and census data. *European Journal of Geography*, *11*(4), 110–125. https://doi.org/10.48088/EJG.P.PAP.11.4.110.125
Paraskevopoulos, Y., Bardosa, A., & Photis, Y. N. (2019). Eploring the Impact of Network Configuration and Transport Accessibility on Population Dynamics. The Case of Naxos Island, Greece. *European Journal of Geography*, *10*(4), 177–194.

Park, J., Zhang, H., Han, S. Y., Nara, A., & Tsou, M. H. (2020). Estimating Hourly Population Distribution Patterns at High Spatiotemporal Resolution in Urban Areas Using Geo-Tagged Tweets and Dasymetric Mapping. *11th International Conference on Geographic Information Science (GIScience 2021)*, *177*(10), 1–16. https://doi.org/10.4230/LIPIcs.GIScience.2021.I.10

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *R Journal*, *10*(1), 439–446. https://doi.org/10.32614/rj-2018-009

Peng, Z., Wang, R., Liu, L., & Wu, H. (2020). Fine-scale dasymetric population mapping with mobile phone and building use data based on grid voronoi method. *ISPRS International Journal of Geo-Information*, *9*(6). https://doi.org/10.3390/ijgi9060344

Photis, Y. N., & Sirigos, S. A. (2015). Scenario-Based Location of Ambulances for Spatiotemporal Clusters of Events and Stohastic Vehicle Availability. A Decision Support Systems Approach. *European Journal of Geography,* 6(4), 59-75. https://eurogeojournal.eu/index.php/egj/article/view/405/295

Prener, C., & Revord, C. (2019). areal: An R package for areal weighted interpolation. *Journal of Open Source Software*, *4*(37), 1221. https://doi.org/10.21105/joss.01221

Qiu, F., Zhang, C., & Zhou, Y. (2012). The development of an areal interpolation ArcGIS extension and a comparative study. *GIScience and Remote Sensing*, *49*(5), 644–663. https://doi.org/10.2747/1548-1603.49.5.644

R Core Team. (2015). *R: a Language and Environment for Statistical Computing.* http://www.r-project.org/

RStudio, I. (2013). *Easy web applications in R*. RStudio Inc. https://shiny.rstudio.com/

Sleeter, R., & Gould, M. (2008). Geographic information system software to remodel population data using dasymetric mapping methods. *US Geological Survey, Techniques and Methods*, *11-C2*, 1–15. http://pubs.usgs.gov/tm/tm11c2/

Tenerelli, P., Gallego, J. F., & Ehrlich, D. (2015). Population density modelling in support of disaster risk assessment. *International Journal of Disaster Risk Reduction*, *13*, 334–341. https://doi.org/10.1016/j.ijdrr.2015.07.015

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

Younes, A., Ahmad, A., Hanjagi, A. D., & Nair, A. M. (2023). Understanding Dynamics of Land Use & Land Cover Change Using GIS & Change Detection Techniques in Tartous , Syria. *European Journal of Geography*, *14*(3), 20–41. https://doi.org/10.48088/ejg.a.you.14.3.020.041

Zandbergen, P. A., & Ignizio, D. A. (2010). Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartography and Geographic Information Science*, *37*(3), 199–214. https://doi.org/10.1559/152304010792194985

European Journal of Geography