Effects of demand estimates on the evaluation and optimality of service centre locations

Matej Cebecauer a, Konštantín Rosina b and Ľuboš Buzna a,c

aDepartment of Mathematical Methods and Operations Research, University of Žilina, Žilina, Slovakia; bSlovak Academy of Sciences, Institute of Geography, Bratislava, Slovakia; cERA chair for Intelligent Transport Systems, University of Žilina, Žilina, Slovakia

ABSTRACT

Public service systems, such as emergency health care, police or fire brigades, are critical for day-to-day functioning of the society. To design and operate these systems efficiently much data needs to be collected and properly utilised. Here, we use the OpenStreetMap (OSM) data to model the demand points (DPs), which approximate the geographical location of customers, and the road network, which is used to access or distribute services. We consider all inhabitants as customers, and therefore to estimate the demand, we use the available population grids. People are changing their location in the course of the day and thus the demand for services is changing accordingly. In this paper, we investigate how the used demand estimate affects the optimal design of a public service system. We calculate and compare efficient designs corresponding to two demand models, a nighttime demand model when the majority of inhabitants rest at home and the demand model derived from the 24-hour average of the population density. We propose a simple measure to quantify the differences between population grids and we estimate how the size of differences affects the optimal structure of a public service system. Our analyses reveal that the efficiency of the service system is not only dependent on the placement strategy, but an inappropriate demand model has significant effects when designing a system as well as when evaluating its efficiency.

ARTICLE HISTORY

Received 22 September 2014
Accepted 20 September 2015

KEYWORDS

Public service systems; population grids; spatial planning and policy

1. Introduction

Public service systems provide a broad range of critical services to the general public. Real-world examples of such systems are networks of hospitals, schools, ambulances, fire or police stations. In this paper, we evaluate the impact of spatiotemporal variations in the demand for services on the efficient spatial design of the public service systems. Models of the demand are relevant either when designing a service system from scratch or when evaluating the properties of an existing system. The service system design problem consists of finding a suitable set of service centre locations from where the services could be efficiently distributed to customers. When
evaluating the properties of an existing service system, spatial positions of service centres are known, and the goal is to evaluate how well the spatial design corresponds to the demand generated by customers.

Basic data requirements are shared across various types of public service systems. They consist of a suitable set of demand points (DPs) representing geographical positions of customers, the road network infrastructure used to distribute services or access the service centres, and the population density data of suitable resolution to estimate the demand for services. These requirements are identical for the design of a new system and also for the evaluation of an existing system.

Combining the OpenStreetMap (OSM) data with two types of population grids, we build a data model that is later used to compute the optimal design of a hypothetical public service system. To the best of our knowledge, this is the first attempt to quantify the effects of the spatio-temporal variations in the demand on the efficient placement of service centres. Using a large set of efficient service centre locations, we vary the spatio-temporal model of demand and we quantify the effects on the estimated system efficiency. We compare the size of both types of effects. Our results show that the type of the population grid used to model the demand for services should be carefully selected. The misplacement of the population grid might have a significant impact on the efficiency of the resulting service system, a remark of caution to system designers.

The paper is organised as follows: section 2 provides a structured review of the literature. In section 3, we describe the data, preparation procedures and their justifications. Results of numerical experiments are reported in section 4. To conclude, we summarise our main findings in section 5.

2. Background

The number of existing location problems that can be used to find the optimal spatial design of a service system is overwhelming (Daskin 1995, Drezner 1995, Eiselt and Marianov 2011). The p-median problem is one of the most frequently studied and used location problems (Hakimi 1965, Calvo and Marks 1973, Berlin et al. 1976, Janáček et al. 2012). The goal is to locate exactly p service centres in a way that the sum of weighted travel times from all customers to service centres is minimised. The problem is NP-hard (Kariv and Hakimi 1979). For a comprehensive overview of applications and solving methods, see references Marianov and Serra (2002, 2011). Exact solving methods are summarised by Reese (2006) and heuristic methods by Mladenović et al. (2007).

To describe the p-median problem, we adopt the well-known integer formulation proposed by ReVelle and Swain (1970). We consider all n DPs as possible candidate locations. The travel time on the fastest path between DPs i and j is denoted as t_{ij}. We associate with each DP a weight w_i that represents the number of customers assigned to the DP i. Decisions are described by the set of binary variables:

\[
x_{ij} = \begin{cases} 
1, & \text{if demand point } i \text{ is assigned to service centre } j \\
0, & \text{otherwise,}
\end{cases}
\]
\[
y_j = \begin{cases} 
1, & \text{if a service centre at the candidate location } j \text{ is open}, \\
0, & \text{otherwise}. 
\end{cases}
\]

The p-median problem can be formulated as follows:

\[
\begin{align*}
\text{Minimise} & \quad f = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} t_{ij} x_{ij} \\
\text{subject to} & \quad \sum_{j=1}^{n} x_{ij} = 1 \quad \text{for all } i = 1, 2, \ldots, n \\
& \quad x_{ij} \leq y_j \quad \text{for all } i, j = 1, 2, \ldots, n \\
& \quad \sum_{j=1}^{n} y_j = p \\
& \quad x_{ij}, y_j \in \{0, 1\} \quad \text{for all } i, j = 1, 2, \ldots, n.
\end{align*}
\]

The objective function (1) minimises the sum of weighted travel times from all customers to service centres. Constraints (2) ensure that each customer is allocated to exactly one service centre. Constraints (3) allow one to allocate customers only to located service centres, and constraint (4) ensures that exactly \( p \) service centres are located.

### 2.1. Data sets

Service systems typically cover large geographical areas, and the system design or system quality evaluation are data-intensive activities. Customers are modelled by the set of DPs representing their spatial distribution (Francis et al. 2009). Typically, this set can be derived from the specification of the real-world problem or we can approximate DPs from the knowledge of the land use/land cover (LULC). Travel distances or travel times can be deduced from the road network. Considering the different classes of roads helps one in determining the average travel speed more precisely. Typically, when modelling service systems in urban or rural areas the number of DPs is extremely vast, since each private residence might be a DP. Often it is impossible, and also unnecessary, to include all DPs to the model. The solution is to use an aggregation. There is a strong stream of literature studying aggregation methods and the corresponding errors (for a comprehensive overview see and references therein Francis et al. 2009). Various sources of aggregation errors are discussed by Hillsman and Rhoda (1978), Erkut and Bozkaya (1999) and Hodgson et al. (1997). Many service systems are supposed to serve the general public, i.e. all citizens are considered to be customers and the demand for services is assumed to be proportional to the population. Here, we also consider this assumption. However, it should be noted that in cases when only a specific part of the population is served by the system (e.g. children of a certain age), it might be misleading to derive the demand from the general population data. The size of the surrogation error caused by the wrong estimation of the
population could be roughly five times as high as the aggregation error (Hodgson and Hewko 2003).

In this paper, we consider volunteered geographic information (VGI) as a source of data that can be used to extract streets and road infrastructure. Moreover, it provides other useful information such as positions of buildings, residential areas, industrial areas and commercial areas, which can be used to define DPs. VGI is created by volunteers, who produce data through Web 2.0 applications and combine it with the publicly available data (Goodchild 2007). VGI has been criticised due to potential data-quality issues (Flanagin and Metzger 2008, Jackson et al. 2013), and researchers are looking for mechanisms of ensuring the data quality (Bishr and Mantelas 2008) and protecting against deliberate data damages (Neis et al. 2012) or on mitigating the effect of completeness errors in VGI data by machine learning approaches (Hagenauer and Helbich 2012).

One of the most successful examples of VGI is the OSM project. OSM provides the opportunity to download spatial data without any costs or fees and enables the use of data for individual projects (Zielstra and Zipf 2010). OSM data has already been used in various applications, such as automatic derivation of three-dimensional CityGML models (Goetz and Zipf 2012), mapping of land-use patterns (Jokar Arsanjani et al. 2013), designing of a recommendation system providing tourists with the most popular landmarks as well as the best travel routings between the landmarks (Sun et al. 2013) and population mapping at the building level (Bakillah et al. 2014).

To estimate the demand for services, the population density data are needed. The best way to produce a gridded map of a population density is to count the number of people in each cell of the grid (bottom-up approach). Nevertheless, not all of the European countries are releasing bottom-up population grids, and only a few of them provide the resolution finer than 1 km. Often, the distribution of grids is limited by confidentiality rules (Gallego 2010). An alternative is to reallocate data from available census zones (communes, wards) into regular grid cells using disaggregation methods (top-down approach) (Zandbergen 2011). Spatial disaggregation is a variant of the long-studied areal interpolation problem of estimating values for a set of target zones based on values recorded for another set of incongruent source zones (Tobler 1979, Goodchild and Lam 1980). Simple methods estimate the target values based on assumptions such as homogeneous spatial distribution (in areal weighting) or distance decay (Bracken and Martin 1989) without the use of additional spatial information. Intelligent (or dasymetric) methods employ ancillary spatial data related to the disaggregated variable to improve the estimates. Martin et al. (2000) note that the accuracy of the resulting map is influenced more by the quality and appropriateness of the used ancillary data than by the particular disaggregation algorithm. The most frequently used ancillary data in population disaggregation are LULC categorical maps, typically derived from the classification of satellite/aerial imagery (Langford and Unwin 1994, Fisher and Langford 1996, Eicher and Brewer 2001, Mennis 2003, Holt et al. 2004, Gallego 2010). Other types of ancillary data have been proposed such as road networks (Xie 1995, Reibel and Bufalino 2005), satellite images (Harvey 2002, Li and Weng 2005) and cadastral data (Tapp 2010). Multiple types of ancillary data sets can be combined to increase the accuracy (Dobson et al. 2000, Bhaduri et al. 2002, Batista e Silva et al. 2013).

Some disaggregated population models are based on source data from censuses, recording people at their permanent residences (approximately a night-time distribution
of the population). The actual distribution of population changes dynamically following a complex movement pattern. The most common type of movement is daily commuting to workplaces and schools. Models of daytime population reflect the distribution of people in commuting destinations. In some locations extreme differences between the two distributions occur (e.g. for city business districts, industrial zones or shopping malls). Another type of spatio-temporal conceptualisation of population density is ambient population (Dobson et al. 2000, Sutton et al. 2003), modelling an average occurrence of people over a certain period of time (e.g. one day or a year). In such a model certain non-zero densities should be attributed to spaces such as highways, parks or agricultural land. Ambient models may be preferable, if no specific time of the day is characteristic for the studied problem.

3. Methods

In this paper, we compare public service systems designed based on residential and ambient demand models. The demand models are derived from a residential population grid produced by Batista e Silva et al. (2013) (RP hereinafter) and LandScan ambient population grid (Dobson et al. 2000, Bhaduri et al. 2002) (AP hereinafter). Choice of LandScan was straightforward as it is the only one well-recognised ambient population model covering Slovakia. The AP grid is available on the resolution of 30 seconds of arc (approximately 1 km), defined in the geographic coordinate system. We employed the 2012 version. The source population data for this data set comes from a national census that took place in 2011. As candidate data sets for residential population grid, we considered three population grids: Global Rural-Urban Mapping Project (Balk et al. 2006), residential grid that is provided by the Austrian Institute of Technology (Steinnocher et al. 2011) and residential population grid RP by Batista e Silva et al. (2013). Residential grid RP was chosen with respect to the best coincidence in spatial resolution, coordinate system and reference year with the AP grid. The population grid by Batista e Silva et al. (2013) is based on ETRS89-LAEA, the cell size is only 100 m. Thus, it can be projected to the LandScan cells relatively reliably. The disaggregation was based on 2006 commune population counts and CORINE Land Cover maps enhanced by Soil Sealing, Urban Atlas, Tele Atlas and other data were used as ancillary data.

In the next subsection, we analyse the RP and AP grids by comparing the differences between them. Daily travel times are relatively short for the majority of the population (Kölbl and Helbing 2003, Bazzani et al. 2010). Thus, by reducing the resolution of population grids, the differences between AP and RP grids should diminish. Population grids are coming from very diverse sources. The RP grid was produced in Europe and made public by Batista e Silva et al. (2013). LandScan 2012 was made in the United States and we purchased it directly from East View publisher. The aim of these analyses is to use this expected property of population grids to validate their quality and compatibility.

3.1. Comparison of residential and ambient population grids

We evaluate the differences between RP and AP grids for the geographical area of the Slovak Republic considering various scales of the spatial resolution. Grids cannot be compared directly. Therefore, to minimise the bias arising from different spatial
resolutions of both grids, we projected RP cells onto AP cells by recalculating the population proportionally to the area of cell intersections and we resampled the RP grid to the spatial resolution of 1 km. For each grid cell $k = 1, 2, \ldots, s$ in the population grid $l \in \{AP, RP\}$, where $s$ is the number of grid cells, we denote as $c^l_k$ the population of the cell $k$. To evaluate which population grid assigns a larger population to individual cells, we compute the difference:

$$\delta_k = c^\text{RP}_k - c^\text{AP}_k.$$  

(6)

Values $\delta_k$ are shown in Figure 1(a–b) for two selected areas: Bratislava (the capital of the Slovak Republic) and Košice (the second largest Slovak city). The general pattern revealed by this comparison confirms that AP allocates larger population to industrial and commercial areas and smaller population to residential areas. Values $\delta_k$ form an image of absolute differences between population grids and thus they do not allow spotting relative differences, which might be small in terms of absolute values. Therefore, we define the relative difference as

$$\phi_k = \frac{\delta_k}{\frac{1}{2} (c^\text{RP}_k + c^\text{AP}_k)}.$$  

(7)

It is not obvious which value, $c^\text{RP}_k$ or $c^\text{AP}_k$, would be more suitable as a normalisation factor and thus we normalised $\delta_k$ by the average value. Figure 1(e) shows the $\phi_k$ values for the entire area of the Slovak Republic. The AP grid attributes some population to non-urban areas, where the RP is zero (blue areas in Figure 1(e)). Although population density attributed by the AP grid to non-urban areas is low, these areas are geographically large. Thus, they could significantly influence the travel times from service centres to customers.

To check whether RP and AP grids are still comparable, when reducing the spatial resolution, we changed the spatial resolution to 2 km, 4 km, 8 km and 16 km. For the area of the Slovak Republic, we evaluate the similarity of population grids by calculating the Pearson product-moment correlation coefficient $r$ between population values attributed to individual grid cells. We obtained the following sequence of values: $r_{1\text{km} \times 1\text{km}} = 0.721$, $r_{2\text{km} \times 2\text{km}} = 0.853$, $r_{4\text{km} \times 4\text{km}} = 0.919$, $r_{8\text{km} \times 8\text{km}} = 0.949$ and $r_{16\text{km} \times 16\text{km}} = 0.978$. Figure 1(f–g) shows the scatter plots for the spatial resolutions of $1 \text{ km} \times 1 \text{ km}$ and $16 \text{ km} \times 16 \text{ km}$, respectively (for all scatter plots see Figure S2 of the supplementary information file). Total difference between population grids clearly decreases (correlation increases) when decreasing the resolution. This is in agreement with our expectations and it validates, at least partially, the correctness of the used population grids.

In summary, we find large absolute differences between population grids in urban areas and small differences in non-urban areas, which however constitute large geographical areas. The cross-scale analysis reveals that if the resolution is high, differences between population grids are non-negligible. When lowering the resolution, differences are quickly getting small. Thus, these analyses indicate that differences between population grids can have an impact on the efficient spatial design of service systems, when the area served from one service centre is small.
3.2. Data model

In this subsection, we briefly describe the procedure used to prepare the input data for the service system design problem. This procedure consists of four steps that are illustrated in Figure 2. Steps have to be executed in this order: data preparation,
Figure 2. Workflow of input data processing for the service system design problem from OSM data and population grids.

generation of DPs, association of weights with the DPs and computation of road network travel times between all pairs of DPs.

3.2.1. **Data preparation**

In the first step, data are extracted from the OSM database. Geographical features that are suitable to define the position of DPs depend on the application. Here, our goal is to capture the position of inhabitants independent of the time of the day. Therefore, we selected five basic OSM layers that allow estimating the positions of inhabitants when they are at home, at work and also when they are travelling. Thus, to model DPs, we consider data layers describing positions of buildings, roads, residential, industrial and commercial areas. The layer of roads is later used to calculate the travel times between DPs.
3.2.2. Generation of DPs

DPs are generated in two steps. In the first step, we create a spatial grid, which consists of uniform square cells. The size of cells has a twofold effect. On the one hand, it affects the precision in assigning population to DPs. The unified resolution of population grids is $1 \times 1$ km. On the other hand, resolution of cells affects the precision in estimating the travel times between DPs. Our primary goal is to evaluate the effect of using different types of population grids. Hence, to ensure that the precision in measuring travel times is higher than in the demand estimate, we use the resolution that is one order of magnitude higher than the resolution of available population grids. Thus, to generate DPs, we decided for the size of square cells of $100 \times 100$ metres. We intersect each cell with all OSM data layers described in the previous subsection. We search for intersections between cells of the spatial grid and points, polylines and polygons that constitute the selected OSM layers. In the second step, a DP is situated as a centroid of each cell with a non-empty intersection with an OSM data layer.

3.2.3. Association of DPs with weights

In the previous step, we obtained a set of DPs that are described by their coordinates. Next, we assign a weight representing number of customers to each DP. To make both population grids comparable, we scaled the resolution of the residential population grid from $100$ m $\times$ $100$ m down to the resolution of $1$ km $\times$ $1$ km. First, we calculate the spatial geographical areas associated with DPs using Voronoi diagrams. Second, we assign weights to DPs by intersecting Voronoi polygons with the population grid. The population assigned to a DP is proportional to the population and to the area of the population grid cells intersecting the Voronoi polygon.

3.2.4. Computation of road network travel times

In the last step, all DPs are connected to the closest road segment. To minimise the length of the connection between the DP and the road network, when it is necessary, we split the closest polyline describing the road segment by adding an intermediate node. From all road segments and DPs, we compiled the directed graph $G(V, E, t)$, with node-set $V$, edge-set $E$ and edge-travel time function $t: E \to \mathbb{R}_+^+$. Travel time is estimated from road classes considering the speed limits$^1$: 50 km/h for roads classified as residential, service and unclassified; 90 km/h for the primary, secondary and tertiary roads; and 130 km/h for motorways and trunk roads. We measure the travel time between DPs in seconds. The quality and completeness of the OSM road network are discussed in section 1 of the supplementary information file. For the study area of the Slovak Republic, we obtained the graph with 1 966 092 nodes (including 666 824 DPs) and 1 884 537 edges.

4. Results

To find the optimal location of service centres for the p-median problem, we use the state-of-the-art algorithm ZEBRA (García et al. 2011). GIS software offers tools for locating-allocating service centres; however, they are typically based on heuristic algorithms providing suboptimal solutions. We decided to use exact algorithm ZEBRA, to rule out
from evaluations the errors that could be caused by suboptimal solutions. To be able to solve as large instances as possible, we compiled 64-bit version of the ZEBRA algorithm, and we run computations on the high-performance cluster, while setting the memory limit to 90 GB and the time limit to seven days. Our goal is to gain a better understanding of the relation between the level of differences in the demand resulting from using AP and RP population grids and the size of the impact on the optimal location of service centres. In order to do so, we need a simple measure, which can be used to quantify the differences between the demand models. For this purpose, we adapt the measure relative total absolute error, defined in the reference Batista e Silva et al. (2013). For each DP \( t = 1, 2, \ldots, n \), where \( n \) is the number of DPs, we denote as \( g_l^t \) the population projected from the population grid \( l \in \{AP, RP\} \) to the DP \( t \). Similarly as in Equation (7), we normalise the total absolute deviation by the average population attributed to individual DPs:

\[
\varphi = \frac{\sum_{k=1}^{n} |g_k^{RP} - g_k^{AP}|}{\frac{1}{2} \sum_{k=1}^{n} (g_k^{RP} + g_k^{AP})}.
\] (8)

We start by computing values of \( \varphi \) for all 79 districts of the Slovak Republic. The choropleth map and histogram of \( \varphi \) values are shown in Figure 3. The minimum value \( \varphi = 0.416 \) was found for the district of Partizánske and the maximum value \( \varphi = 1.118 \) was obtained for the district of Košice II. The histogram of \( \varphi \) values shows that the large majority of values does not exceed the value 0.8. To study the dependence of the relative errors on the parameter \( \varphi \), we selected a sample of 10 districts (see Table 1), to cover the entire range of \( \varphi \) values uniformly. The sample covers 9.64% of the population of Slovakia.

Nowadays, 113 professional fire stations, 273 emergency ambulance stations, 405 police stations (239 out of them were established by the state government and 166 were established by the self-governed municipalities) and 1500 post offices operate in the

![Figure 3. Choropleth map and histogram of \( \varphi \) values for all 79 administrative districts of the Slovak Republic. Values \( \varphi \) are colour-coded according to the colour key that is shown below the histogram.](image)
Slovak Republic. The entire population that is served by these systems consists of 5,415,949 inhabitants. Taking into account the real number of service centres, we calculated the average number of citizens served from one centre. For the areas that constitute our benchmarks, we estimated the corresponding number of service centres using the information about the population. Based on these numbers, we roughly estimated the realistic range for \( p \) values as 1,...,40.

Let us denote the optimal location vector obtained by solving the \( p \)-median problem (1)–(5) when using the weights \( w^k \), derived from the population grid \( k \in \{AP,RP\} \), as \((y^k_1,y^k_2,\ldots,y^k_n)\). Value of the objective function (1) that is associated with the solution \((y^k_1,y^k_2,\ldots,y^k_n)\) for \( k \in \{AP,RP\} \) and with the weights \( w^l \), derived from the population grid \( l \in \{AP,RP\} \), can be calculated as

\[
f^{k,l} = \sum_{j=1}^{n} \min\{(t_{ij}w^l_{ij}) : i \in \{1,\ldots,n\}; y^k_i = 1\}. \tag{9}\]

Thus, the objective function value \( f^{k,l} \) represents the total sum of minimum travel times between the optimal positions of service centres and the positions of individual citizens. Locations of service centres are optimised with respect to the geographical positions of individual citizens derived from the population grid \( k \). Located service centres are evaluated with respect to the positions of individual citizens derived from the population grid \( l \).

To evaluate the error caused by the interchange of population grids, we adopted two standard error measures used in the location analysis (Francis et al. (2009)). The absolute error is the difference between the objective function values corresponding to two distinct situations. In the first situation, population grids \( k \) and \( l \) are different. Grid \( k \) is used in the case when we search for the optimal location of service centres. Grid \( l \) is

| District       | \( \phi \) | Number of DPs | Size \( [km^2] \) | Population (AP/RP) |
|----------------|-----------|---------------|-----------------|-------------------|
| Partizánske   | 0.416     | 4,873         | 301             | 48,165 / 47,553   |
| Hlohovec      | 0.502     | 5,525         | 267.2           | 46,836 / 45,519   |
| Ružomberok    | 0.589     | 5,791         | 646.8           | 59,915 / 59,225   |
| Ilava         | 0.690     | 5,449         | 359             | 63,031 / 61,168   |
| Revúca        | 0.778     | 6,397         | 730.2           | 42,283 / 40,761   |
| Snina         | 0.820     | 5,510         | 804.7           | 39,609 / 39,157   |
| Myjava         | 0.901     | 5,638         | 327             | 28,574 / 27,884   |
| Detva         | 0.950     | 7,966         | 449.2           | 35,346 / 33,568   |
| Košice IV     | 0.966     | 2,791         | 60.9            | 80,461 / 58,180   |
| Košice II     | 1.118     | 3,537         | 80.5            | 77,989 / 78,793   |

Table 1. Basic information about geographical areas selected based on the value \( \phi \).
used in the case when we calculate the value of objective function corresponding to the optimal solution. In the second situation, we use in both cases the same population grids. Thus, the absolute error is defined as

$$
\Delta_{m}^{k,l} = f^{k,l} - f^{m,m},
$$

where $k, l, m \in \{AP, RP\}$ and $k \neq l$. Objective function values are strongly dependent on the value $p$. While increasing $p$, the objective function value decreases. Therefore, values of the absolute error are not comparable across different $p$ values. For this reason, we evaluate the relative error (Francis et al. (2009), Erkut and Bozkaya (1999), Hodgson and Hewko (2003)):

$$
\Phi_{m}^{k,l} = \frac{\Delta_{m}^{k,l}}{f^{m,m}},
$$

where $k, l, m \in \{AP, RP\}$ and $k \neq l$. Values $\Delta_{m}^{k,l}$ and $\Phi_{m}^{k,l}$ close to zero indicate that population grids can be interchanged without any significant errors.

In the analyses, we calculate the absolute and relative location errors $\Delta_{m}^{k,l}$ and $\Phi_{m}^{k,l}$, where $m = l$ and the absolute and relative evaluation errors $\Delta_{k}^{k,l}$ and $\Phi_{k}^{k,l}$, where $m = k$ for $k, l \in \{AP, RP\}$ and $k \neq l$. The location errors $\Delta_{k}^{k,l}$ and $\Phi_{k}^{k,l}$ quantify the difference between values of the objective function, for two optimally located sets of service centres – first computed on the population grid $k$ and second computed on the population grid $l$ – while the objective function values are calculated using the population grid $l$. Hence, they measure the effect of interchanging the population grids when determining the optimal positions of service centres. The evaluation errors $\Delta_{k}^{k,l}$ and $\Phi_{k}^{k,l}$ represent the difference between two values of the objective function that correspond to one location of service centres. But when calculating the objective function values, we use two different sets of weights derived from the population grids $k$ and $l$, respectively. Thus, evaluation errors quantify the effect of interchanging population grids, when evaluating the objective function value for a given location of centres.

Values $\Delta_{i}^{k,l}$, $\Delta_{i}^{k,l}$, $\Phi_{i}^{k,l}$ and $\Phi_{i}^{k,l}$ obtained for selected geographical areas and selected $p$ values 3, 5, and 10 are shown in Figures 4 and 5. For the complete results covering the realistic range of $p$ values, please refer to Figures S4 and S5 of the supplementary information. We note that we found very similar results, when instead of travel times we use the shortest distances in the objective function. As the travel times are more relevant when dealing with the service systems, we present only them.

For small values of $\varphi$, where both grids are similar, location errors are small. As the value $\varphi$ increases, the location error grows as well. In extreme cases, such as the district of Košice II where $\varphi = 1.118$, we observe exceedingly large location and evaluation errors. The area of the district Košice II includes a large steel-producing industrial area in the south and the densely inhabited areas on the north. Thus, many people travel from the north to the south in the morning and back in the evening. During the night, the industrial area is almost empty. In this case, the population patterns are significantly different when interchanging the population grids. Consequently, different population patterns result in considerable differences in optimal service centre locations (for illustration, see Figure S6 of the supplementary information). On the contrary, for small $\varphi$ values, for example in the district of Partizánske where $\varphi = 0.416$, the population
di
ferences are only modest. Here, we also find the optimal location of service centres forming similar spatial patterns (for illustrations, see Figures S7 and S9 of the supplementary information).

When the value of $\varphi$ is equal to or larger than 0.82, we systematically find significantly larger objective function values when we use the AP grid to evaluate the travel times than what we find for the RP grid. That is because the AP population has a tendency to be more outstretched in space. It has direct consequences for errors. Location errors grow when $\varphi$ is increasing. Relative evaluation errors $\Phi_k^{ij}$ (similarly also absolute evaluation errors $\Delta_k^{ij}$) behave differently. When computing the evaluation error, one layout of the located service centres is evaluated using two different population grids. If $\varphi$ is large enough, for AP we find significantly larger objective function values than for RP, independent of the located service centres. This explains why evaluation errors $\Delta_{RP,AP}$ and $\Phi_{RP,AP}$ are larger than the corresponding location errors and why evaluation errors $\Delta_{AP,RP}$ and $\Phi_{AP,RP}$ are non-monotonic. Thus, the difference between population layouts

![Figure 4](image-url)

Figure 4. (a),(b) The absolute location error $\Delta_k^{ij}$ for selected values of $p = 3, 5, 10$. (c),(d) The absolute evaluation error $\Delta_k^{ij}$ for selected values of $p = 3, 5, 10$. Districts are ordered from left to right according to the ascending $\varphi$ values. More details about the districts are listed in Table 1.
tends to have a stronger impact when evaluating the objective value than when using different optimal layouts of service centres.

Hillsman and Rhoda (1978) warned that location error caused by uncertainty in input data larger than 2% (for large service areas) and 8% (for small service areas) could be large enough to potentially affect the methodological and substantive interpretations of the results of research on spatial systems. When evaluating the difference in the objective function caused by the exchange of population grids, we found that the errors often exceed these values. Moreover, in some cases, when the difference between population grids measured by \( \varphi \) is large, the errors are significantly larger. These findings provide clear evidence that the proper choice of the population grid plays an important role when designing the optimal structure of public service systems. Moreover, the quantity \( \varphi \) seems to be a relevant indicator, which can be able to estimate the role of differences between population grids on the location error, when searching for an optimal location of service centres. Our results cannot be fully generalised to other areas than the Slovak Republic, but we assume that the effects of differences

Figure 5. (a),(b) The relative location error \( \Phi^{k,l}_{RP} \) for selected values of \( p = 3, 5, 10 \). (c),(d) The relative evaluation error \( \Phi^{k,l}_{AP} \) for selected values of \( p = 3, 5, 10 \). Districts are ordered from left to right according to the ascending \( \varphi \) values. More details about the districts are listed in Table 1.
between AP and RP grids, measured by the parameter \( \phi \), could lead to similar errors also elsewhere.

For some applications of service systems, we consider the warnings introduced by Hillsman and Rhoda (1978) as overstated. How relevant is a given value of location or evaluation errors cannot be easily generalised. Each service system involves many uncertainties, which arise from hardly predictable situations. For example, the travel times that we used as a measure to validate the efficiency of a service system may in some areas vary due to traffic conditions by more than 20% (Jenelius and Koutsopoulos (2013)). Thus, in such a case, location error smaller than 20% may not be of high importance. Thus, it is up to the designer to estimate the present level of uncertainties considering the local conditions and the type of the service system.

In Figure 6, we extracted from results, the maximal, the average and the minimal \( \Phi_{k}^{i,j} \) and \( \Phi_{k}^{i,j} \) values to roughly estimate the potential errors that could be caused by the interchange of two population grids as a function of the parameter \( \phi \). Figure 6 can be used in the following way. If the designer knows all other important uncertainties in input data and if, for instance, the largest is the uncertainty in travel times that was estimated to 20\%, than if she/he obtains for this geographical area \( \phi = 0.416 \) or smaller, from Figure 6 one can estimate the range of the relative location and evaluation errors. In this case, the corresponding maximal relative location error \( \Phi_{k}^{i,j} \) for \( \phi = 0.416 \) is below 10\%. Because in this case, the estimated relative location error is significantly lower than the involved uncertainty in other

---

**Figure 6.** The maximal, minimal and average relative location and evaluation errors as a function of the parameter \( \phi \). Maximum, minimum and average values are for a given \( \phi \) (i.e. for a given district) calculated from the results of all experiments, where \( p = 1, \ldots, 40 \). Error bars reflect the standard deviation around the average.
types of input data, the designer could conclude that it is not necessary to take into account the effects of the spatio-temporal demand distribution in the design of this particular service system.

5. Conclusions

Designing the optimal structure of a public service system is a complex task, where several factors can lead to hardly predictable outcomes. We built a detailed model from the publicly available data and to estimate the demand, we combined it with ambient and residential population grids. We computed a large number of optimal designs of a hypothetical public service system and from the numerical experiments we derived the following main conclusions:

- Use of the RP grid has a tendency to lower the value of the objective function systematically and thus the travel times from customers to the closest service centres can be easily underestimated. This result is important, because due to the better availability of residential population data, there is a tendency to use residential data in the location analyses independent of the application (Burkey et al. (2012), Janáček et al. (2012), Nordbeck et al. (2013)).
- Our analyses provide evidence that the public service system designers should be more careful when evaluating the efficiency of existing systems. Errors that are associated with the evaluation of already-located facilities have a tendency to exceed the errors that emerge when searching for efficient location of facilities.
- The average and the maximum location and evaluation errors grow in a non-linear way as the measure of distance between population grids increases. This dependence can be used to assess the size of errors associated with the demand model that can be compared with other input parameters and help designers to decide which errors to account for.

It is important to note that although our study is based on the best available population data, these data describe the reality only to some limited extent. Our results also provide evidence that more efforts, which should result in bottom-up population data for the Slovak Republic, are needed. Currently, freely available high-quality population data for the Slovak Republic, but also some other European countries, do not entirely match the needs for the public service system design. Open data initiatives and recent advances in obtaining population densities from online social networks (Tarasov et al. (2013)) or mobile phone data (Walsh and Pozdnukhov (2011)) are thus in this sense very promising.

Note

1. http://ec.europa.eu/transport/road_safety/going_abroad/slovakia/speed_limits_en.htm
Acknowledgement

This product was made utilising the LandScan (2012) High-Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy. The United States Government has certain rights in this Data Set. Neither UT-Battelle, LLC nor the United States Department of Energy, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the data set.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the research grants VEGA [1/0339/13] ‘Advanced microscopic modelling and complex data sources for designing spatially large public service systems’, VEGA [1/0275/13] ‘Production, verification and application of population and settlement spatial models based on European land monitoring services’, [APVV-0760-11] ‘Designing Fair Service Systems on Transportation Networks’ and FP 7 project ERAdiate [621386] ‘Enhancing Research and innovation dimensions of the University of Zilina in Intelligent Transport Systems’. We thank Dirk Helbing from ETH Zurich for granting access to the Brutus high-performance cluster. We also thank J. Janáček, L. Jánošíková and M. Koháni for thorough reading of the manuscript and useful suggestions. We thank three anonymous reviewers for very useful and constructive review reports that greatly helped improve the manuscript.

ORCID

Matej Cebecauer http://orcid.org/0000-0002-8499-0843
Konštantín Rosina http://orcid.org/0000-0002-4696-1320
Ľuboš Buzna http://orcid.org/0000-0002-5410-6762

References

Bakillah, M., et al., 2014. Fine-resolution population mapping using OpenStreetMap points-of-interest. International Journal of Geographical Information Science, 28 (9), 1–24.
Balk, D.L., et al., 2006. Determining global population distribution: methods, applications and data. Advances in Parasitology, 62, 119–156.
Batista e Silva, F., Gallego, J., and Lavalle, C., 2013. A high-resolution population grid map for Europe. Journal of Maps, 9 (1), 16–28. doi:10.1080/17445647.2013.764830
Bazzani, A., et al., 2010. Statistical laws in urban mobility from microscopic GPS data in the area of Florence. Journal of Statistical Mechanics: Theory and Experiment, 2010 (05), P05001. doi:10.1088/1742-5468/2010/05/P05001
Berlin, G.N., ReVelle, C.S., and Elzinga, D.J., 1976. Determining ambulance - hospital locations for on-scene and hospital services. Environment and Planning A, 8 (5), 553–561. doi:10.1068/a080553
Bhaduri, B., et al., 2002. LandScan: locating people is what matters. Geoinformatics, 5 (2), 34–37.
Bishr, M. and Mantelas, L., 2008. A trust and reputation model for filtering and classifying knowledge about urban growth. GeoJournal, 72 (3–4), 229–237. doi:10.1007/s10708-008-9182-4
Bracken, I. and Martin, D., 1989. The generation of spatial population distributions from census centroid data. Environment and Planning A, 21 (4), 537–543. doi:10.1068/a210537
Burkey, M., Bhadury, J., and Eiselt, H., 2012. A location-based comparison of health care services in four U.S. states with efficiency and equity. *Socio-Economic Planning Sciences*, 46 (2), 157–163. Modeling Public Sector Facility Location Problems, doi:10.1016/j.seps.2012.01.002

Calvo, A.B. and Marks, D.H., 1973. Location of health care facilities: an analytical approach. *Socio-Economic Planning Sciences*, 7 (5), 407–422. doi:10.1016/0038-0121(73)90039-6

CORINE Land Cover, 2006. [online]. Available from: http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-3 [Accessed 10 October 2015].

Daskin, M.S., 1995. *Network and discrete location: models, algorithms and applications*. Hoboken, NJ: John Wiley & Sons.

Dobson, J., *et al.*, 2000. LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66 (7), 849–857.

Drezner, Z., 1995. *Facility location: A survey of applications and methods*. New York: Springer Verlag.

Eiselt, H.A. and Marianov, V., 2011. *Foundations of location analysis*. International Series in Operations Research and Management Science. New York: Springer, Science + Business.

Eischer, C.L. and Brewer, C.A., 2001. Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science*, 28 (2), 125–138. doi:10.1559/152304001782173727

Eiselt, H.A. and Marianov, V., 2011. *Foundations of location analysis*. International Series in Operations Research and Management Science. New York: Springer, Science + Business.

Gallego, F., 2010. A population density grid of the European Union. *Population and Environment*, 31 (6), 460–473. doi:10.1007/s11111-010-0108-y

García, S., Labbé, M., and Marin, A., 2011. Solving large p-median problems with a radius formulation. *INFORMS Journal on Computing*, 23 (4), 546–556. doi:10.1287/ijoc.1100.0418

Goetz, M. and Zipf, A., 2012. Towards defining a framework for the automatic derivation of 3D CityGML models from volunteered geographic information. *International Journal of 3-D Information Modeling*, 1 (2), 1–16. doi:10.4018/IJ3DIM

Hagenauer, J. and Helbich, M., 2012. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science*, 26 (6), 963–982. doi:10.1080/13658816.2011.619501

Hakimi, S.L., 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, 13 (3), 462–475. doi:10.1287/opre.13.3.462

Harvey, J.T., 2002. Estimating census district populations from satellite imagery: some approaches and limitations. *International Journal of Remote Sensing*, 23 (10), 2071–2095. doi:10.1080/01431160110075901

Hillsman, E. and Rhoda, R., 1978. Errors in measuring distances from populations to service centers. *The Annals of Regional Science*, 12 (3), 74–88. doi:10.1007/BF01286124

Hodgson, M. and Hewko, J., 2003. Aggregation and surrogation error in the p-median model. *Annals of Operations Research*, 123 (1/4), 53–66. doi:10.1023/A:102615027615

Hodgson, M.J., Shmulevitz, F., and Körkel, M., 1997. Aggregation error effects on the discrete-space p-median model: the case of Edmonton, Canada. *The Canadian Geographer/Le Géographe canadien*, 41 (4), 415–428. doi:10.1111/cag.1997.41.issue-4
Holt, J.B., Lo, C.P., and Hodler, T.W., 2004. Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science*, 31 (2), 103–121. doi:10.1559/1523040041649407

Jackson, S.P., et al., 2013. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2 (2), 507–530. doi:10.3390/ijgi2020507

Janáček, J., Jánošíková, L., and Buzna, L., 2012. Optimized design of large-scale social welfare supporting systems on complex networks. In: M.T. Thai and P.M. Pardalos, eds. *Handbook of optimization in complex networks.*, Vol. 57 of springer optimization and its applications. US: Springer, 337–361.

Jenelius, E. and Koutsopoulos, H.N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, 53 (0), 64–81. doi:10.1016/j.trb.2013.03.008

Jokar Arsanjani, J., et al., 2013. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, 27 (12), 2264–2278. doi:10.1080/13658816.2013.800871

Kariv, O. and Hakimi, S.L., 1979. An algorithmic approach to network location problems. II: the P-Medians. *SIAM Journal on Applied Mathematics*, 37 (3), 539–560.

Kölbl, R. and Helbing, D., 2003. Energy laws in human travel behaviour. *New Journal of Physics*, 5 (1), 48. doi:10.1088/1367-2630/5/1/348

Langford, M. and Unwin, D.J., 1994. Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31 (1), 21–26. doi:10.1179/caj.1994.31.1.21

Li, G. and Weng, Q., 2005. Using landsat ETM + imagery to measure population density in Indianapolis, Indiana, USA. *Photogrammetric Engineering & Remote Sensing*, 71 (8), 947–958. doi:10.14358/PERS.71.8.947

Marianov, V. and Serra, D., 2002. Location problem in the public sector. In: Z. Drezner and H.W. Hamacher, eds. *Facility location: applications and theory*. Berlin: Springer, 460.

Marianov, V. and Serra, D., 2011. Median problems in networks. In: H.A. Eiselt and V. Marianov, eds. *Foundations of location analysis.*, Vol. 155 of international series in operations research & management science. US: Springer, 39–59.

Martin, D., Tate, N., and Langford, M., 2000. Refining population surface models: experiments with Northern Ireland census data. *Transactions in GIS*, 4, 343–360. doi:10.1111/1467-9671.00060

Mennis, J., 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55 (August 2002), 125–138.

Miladenoivić, N., et al., 2007. The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research*, 179 (3), 927–939. doi:10.1016/j.ejor.2005.05.034

Neis, P., Goetz, M., and Zipf, A., 2012. Towards automatic vandalism detection in openstreetmap. *ISPRS International Journal of Geo-Information*, 1 (3), 315–332. doi:10.3390/ijgi1030315

Nordbeck, O.E., et al., 2013. Access to emergency hospotals: an operative case study for testing gridded population data. In: *Proceeding of the European Forum for Geostatistics* [online], Sofia Conference 2013, 52. Available from: http://www.efgs.info/geostat/1B/frontpage/geostat1b-access-to-hospitals [Accessed 10 October 2015].

Reese, J., 2006. Solution methods for the p-median problem: an annotated bibliography. *Networks*, 48 (3), 125–142. doi:10.1002 (ISSN)1097-0037

Reibel, M. and Bufalino, M.E., 2005. Street weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37 (1), 127–139. doi:10.1068/a36202

ReVelle, C.S. and Swain, R.W., 1970. Central facilities location. *Geographical Analysis*, 2 (1), 30–42. doi:10.1111/j.1538-4632.1970.tb00142.x

Steinnocher, K., Kstl, M., and Weichselbaum, J., 2011. Grid-based population and land take trend indicators - new approaches introduced by the geoland2 core information service for spatial planning. In: *Brussels NTTS conference* [online]. Available from: http://www.cros-portal.eu/sites/default/files/S6P4.pdf [Accessed 10 October 2015].
Sun, Y., et al., 2013. Road-based travel recommendation using geo-tagged images. Computers, Environment and Urban Systems, Available online. doi:10.1016/j.compenvurbsys.2013.07.006
Sutton, P.C., Elvidge, C., and Obremski, T., 2003. Building and evaluating models to estimate ambient population density. Photogrammetric Engineering & Remote Sensing, 69 (5), 545–553. doi:10.14358/PERS.69.5.545
Tapp, A.F., 2010. Areal interpolation and dasymetric mapping methods using local ancillary data sources. Cartography and Geographic Information Science, 37 (3), 215–228. doi:10.1559/152304010792194976
Tarasov, A., Kling, F., and Pozdnukhov, A., 2013. Prediction of user location using the radiation model and social check-ins. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago: ACM, 7.
Tobler, W., 1979. Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, 74, 519–530. doi:10.1080/01621459.1979.10481647
Walsh, F. and Pozdnukhov, A., 2011. Spatial structure and dynamics of urban communities. In: First Workshop on Pervasive Urban Applications (PURBA) at Ninth International Conference on Pervasive Computing, San Francisco: NUI maynooth, 8.
Xie, Y., 1995. The overlaid network algorithms for areal interpolation problem. Computers, Environment and Urban Systems, 19 (4), 287–306. doi:10.1016/0198-9715(95)00028-3
Zandbergen, P.A., 2011. Dasymetric mapping using high resolution address point datasets. Transactions in GIS, 15, 5–27. doi:10.1111/tgis.2011.15.issue-s1
Zielstra, D. and Zipf, A., 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany. In: 13th AGILE International Conference on Geographic Information Science 2010 Guimarães, Portugal [online]. Available from: http://www.agile-online.org/Conference_Paper/CDs/agile_2010/ShortPapers_PDF/142_DOC.pdf [Accessed 10 October 2015].