



DALARNA  
UNIVERSITY

Working papers in transport, tourism, information technology and microdata analysis

## Distance measure and the $p$ -median problem in rural areas

---

---

**Författare 1: Kenneth Carling**

**Författare 2: Mengjie Han**

**Författare 3: Johan Håkansson**

**Författare 4: Pascal Rebreyend**

**Editor: Hasan Fleyeh**

**Nr: 2012:07**

Working papers in transport, tourism, information technology and microdata analysis

ISSN: 1650-5581

© Authors

## Distance measure and the $p$ -median problem in rural areas

Authors<sup>♦</sup>: Kenneth Carling, Mengjie Han, Johan Håkansson<sup>\*</sup>, and Pascal Rebreyend

This version: 2012-12-04

**Abstract:** The  $p$ -median model is used to locate  $P$  facilities to serve a geographically distributed population. Conventionally, it is assumed that the population patronize the nearest facility and that the distance between the resident and the facility may be measured by the Euclidean distance. Carling, Han, and Håkansson (2012) compared two network distances with the Euclidean in a rural region with a sparse, heterogeneous network and a non-symmetric distribution of the population. For a coarse network and  $P$  small, they found, in contrast to the literature, the Euclidean distance to be problematic. In this paper we extend their work by use of a refined network and study systematically the case when  $P$  is of varying size (2-100 facilities). We find that the network distance give as good a solution as the travel-time network. The Euclidean distance gives solutions some 2-7 per cent worse than the network distances, and the solutions deteriorate with increasing  $P$ . Our conclusions extend to intra-urban location problems.

**Key words:** dense network, location model, optimal location, simulated annealing, travel time, urban areas

---

<sup>♦</sup> Kenneth Carling is a professor in Statistics, Mengjie Han is a PhD-student in Micro-data analysis, Johan Håkansson is a professor in Human Geography, and Pascal Rebreyend is a professor in Computer Science at the School of Technology and Business Studies, Dalarna university, SE-791 88 Falun, Sweden.

<sup>\*</sup> Corresponding author. E-mail: [jhk@du.se](mailto:jhk@du.se). Phone: +46-23-778573.

## 1. Distance measures in the $p$ -median model

Consider the problem of allocating  $P$  facilities to a population geographically distributed in  $Q$  demand points such that the population's average or total distance to its nearest service center is minimized. Hakimi (1964) considered the task of locating telephone switching centers and showed that, in a network, the optimal solution of the  $p$ -median model existed at the nodes of the network. Thereafter, the  $p$ -median model has come to use in a remarkable variety of location problems (see Hale and Moberg, 2003).

However, there are three, main challenges with applying the  $p$ -median model on a specific location problem. The first is computational due to the combinatorial feature of the problem. Enumeration of all possible locations, in search of the optimal one, is a formidable task even for  $P$  and  $Q$  small. Hence, much research has been devoted to efficient (heuristic) algorithms to solve the  $p$ -median model (see Handler and Mirchandani 1979, Daskin 1995, and Murray and Church 1996 as examples).

The second challenge is the aggregation error arising from the common practice of aggregating demand points. Hillsman and Rhoda (1978) analysed the errors that may arise in measuring the distance between the population to be served and the facilities. One source of error comes from the aggregation of the population in an area to a single point, where the point shall represent the position of all members of the population in the area. Their research spurred an on-going investigation of this error and techniques to reduce the error (see Francis, Lowe, Rayco, and Tamir 2009 and references therein).

The third challenge is to measure the distance between the demand point and the nearest service center. In his seminal paper, Bach (1981) conducted a thorough investigation of how to measure distance. A number of competing alternatives are the Euclidean (shortest distance in the plane), the rectilinear (or Manhattan distance), the network distance (shortest distance along an existing road or public transport network), and shortest travel time (or cost) along an existing network. Remarkably, Bach (1981) found that the correlation was close to one for network and Euclidean distances when he conducted an empirical examination of two densely populated German cities. Hence, his results

indicate that it does not matter whether the network or the Euclidean distance is used as distance measure. After the publication of Bach (1981), there is little research on the choice of distance measure.

Carling, Han, and Håkansson (2012) compared the Euclidean distance with a coarse road network distance, and travel-time in a two-speed network. They compared the outcome of the  $p$ -median model for the three distance measures for a problem where  $P$  was varied from 2 to 8 facilities ( $Q$  was large and the population spatially disaggregated). They concluded that the Euclidean distance was problematic as it led to suboptimal location of facilities and a distorted understanding of the facilities service area. Spatial aggregation was however found to be inconsequential.

Carling et al (2012) was limited in scope with regard to the  $p$ -median model as it studied the choice of distance measure for  $P$  small in a rural setting with a coarse representation of the network. The aim of this paper is to test whether their conclusion for the  $p$ -median model is of more generality. We do this by systematically vary  $P$  from small to medium in size (2-100 facilities). The experiment is conducted on a refined network in Dalecarlia in Sweden with more than 1,500,000 nodes in which the speed limit for a road segment varies between 30 km/h to 110 km/h. Moreover, there are more than 15,000 demand points representing the population with an error of at the most 175 meters.

The paper is organized as follows: Section two presents the empirical setting and the distance measures. Section three gives the computational approach. Section four presents the results. And the fifth section concludes.

## **2. The empirical setting: Geography and Network**

Figure 1 shows the Dalecarlia region in central Sweden, about 300 km northwest of Stockholm. The size of the region is approximately 31,000 km<sup>2</sup>. Figure 1a gives the geographical distribution of the region<sup>1</sup>. As of December 2010, the Dalecarlia population numbers 277,000 residents. About 65 % of the population lives in 30 towns and villages with between 1,000 and 40,000 residents, whereas the

---

<sup>1</sup> The population data used in this study comes from Statistics Sweden, and is from 2002 ([www.scb.se](http://www.scb.se)). The residents are registered at points 250 meters apart in four directions (north, west, south, and east). There are 15,729 points that contain at least one resident in the region.

remaining third of the population resides in small, scattered settlements. The figure shows the distribution of the residents in the region by squares of 1 km by 1 km. It indicates that the population is non-symmetrically distributed, and also sparsely populated with an average of nine residents per square kilometer (the average for Sweden overall is 21).

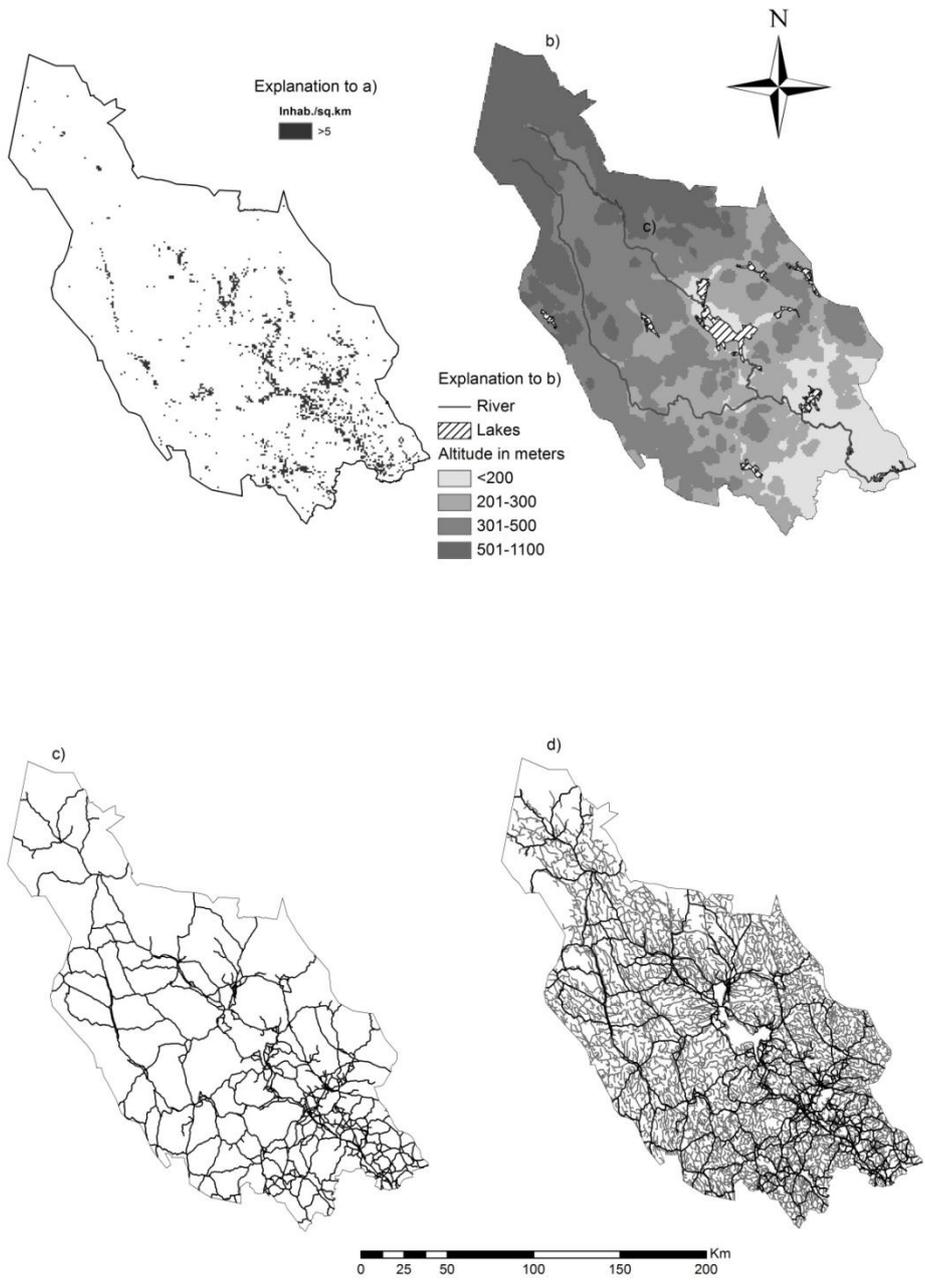


Figure 1: Map of the Dalecarlia region showing (a) one-by-one kilometer cells where the population exceeds 5 inhabitants, (b) landscape, (c) national road system, and (d) national road system with local streets and subsidized private roads.

Figure 1b shows the landscape and gives a perception of the geographical distribution of the

population. The altitude of the region varies substantially; for instance in the western areas, the altitude exceeds 1,000 meters above sea level, whereas the altitude is less than 100 meters in the southeast corner. Altitude variations, the rivers' extensions, and the locations of the lakes provide many natural barriers to where people could settle, and how a road network could be constructed in the region. The majority of residents live in the southeast corner, while the remaining residents are primarily located along the two rivers and around Lake Siljan in the middle of the region.

Figure 1c shows the national road network in the region. The Swedish road system is divided into national roads and local streets that are public, and subsidized and non-subsidized private roads and in Dalecarlia the total length of the road system is 39,452 km.<sup>2</sup> The non-subsidized private roads is the most extensive network amounting to more than 50 per cent of the country's roads and it is primarily built and maintained by companies, and in Dalecarlia for the purpose of transporting timber. The national road system in Dalecarlia totals 5,437 km with roads of varying quality that are, in practice, distinguished by a speed limit.

Table 1: The distribution of speed limits (km/h) in the public road network of Dalecarlia.

Proportion (%)	Speed limit							
	-30	40	50	60	70	80	90	100-
	9	3	31	2	24	19	10	2

Figure 1d adds the local streets and subsidized private roads to the national road network with an additional extension of 14,803 km. This network is very dense compared with the national roads alone. The reason to also depict the subsidized private roads is that they provide an opportunity for the residents to reach the public roads.

The speed limit varies between 30 to 110 km/h in the region's road network. Table 1 gives the proportion of road-kilometers by speed limit for the public road network. The speed limit of 70 km/h is default and the national roads usually have a speed limit of 70 km/h or more. The road network in the towns consists mostly of local streets with low and uniform speed limits (30-50 km/h). Han, Håkansson, and Rebreyend (2012) used the *p*-median model on this road network, and

<sup>2</sup> The road networks are provided by the NVDB (The National Road Data Base). NVDB was formed in 1996 on behalf of the government and now operated by Swedish Transport Agency. NVDB is divided into national roads, local road and streets. The national roads are owned by the national public authorities, and the construction of them funded by a state tax. The local roads or streets are built and owned by private persons or companies or by the municipalities. Data was extracted spring 2011 and represents the network of the winter of 2011. The computer model is built up by about 1.5 million nodes and 1,964,801 road segments.

they noted that it is imperative to include local streets unless  $P$  is small.

### 3. The $p$ -median model and computational aspects

The problem is to allocate  $P$  facilities to the population geographically distributed in  $Q$  demand points such that the population's average or total distance to its nearest facility is minimized. The  $p$ -median objective function<sup>3</sup> is  $\sum_{q \in N} w_q \min_{p \in P} \{d_{qp}\}$ , where  $N$  is the number of nodes,  $q$  and  $p$  indexes the demand and the facility nodes respectively,  $w_q$  the demand at node  $q$ , and  $d_{qp}$  the shortest distance between the nodes  $q$  and  $p$ .<sup>4</sup>

The shortest Euclidean distance,  $d_{qp}^E$  say, is simply the distance in the plane between the nodes  $q$  and  $p$ . To find the shortest network distance and shortest travel-time distance,  $d_{qp}^N$  and  $d_{qp}^T$  say, between the nodes  $q$  and  $p$  is trickier since there may be many possible routes between the nodes in a refined network. We implemented the Dijkstra algorithm (Dijkstra 1959) and retrieve the shortest distance from the center to the residents in each evaluation of the objective function. To obtain the travel-time we assumed that the attained velocity corresponded to the speed limit in the road network.

The  $p$ -median problem is NP-hard (Kariv and Hakimi, 1979). Han et al (2012) discussed and examined exact solutions to the problem as well as heuristic solutions. They advocated the simulated annealing algorithm for the problem at hand and we comply. This randomized algorithm is chosen due to its easiness to implement and the quality of results in case of complex problems. Most important, in our case, the cost of evaluating a solution is high and therefore we prefer an algorithm which keeps the number of evaluated solutions low. This excludes for example algorithms such as Genetic Algorithm and some extended Branch and Bound. Moreover, we may have good starting points obtained from pre-computed trials. Therefore a good candidate is Simulated Annealing (Kirkpatrick, Gelatt, and Vecchi, 1983).

---

<sup>3</sup> Arguments leading to other objective functions can be found elsewhere see e.g. Berman and Krass (1998) and Drezner and Drezner (2007). For instance, a heterogeneous population raises the issue of whether attributes such as the number of residents, average income, educational level, and so on should be considered. To maintain focus, we adhere to the objective function mentioned above.

<sup>4</sup> Facilities are always located at a node in line with the result of Hakimi (1964). Residents are assumed to start the travel at their nearest node, and reaching it by a travel of the Euclidean distance. This assumption is of no importance in this dense road network.

The simulated annealing (SA) is a simple and well described meta-heuristic. Al-khedhairi (2008) gives the general SA heuristic procedures. SA starts with a random initial solution  $s$  and the initial temperature  $T_0$  and the temperature counter  $t = 0$ . The next step is to improve the initial solution. The counter  $n = 0$  is set and the operation is repeated until  $n = L$ . A neighbourhood solution  $s'$  is evaluated by randomly exchanging one facility in the current solution to the one not in the current solution. The difference,  $\Delta$ , of the two values of the objective function is evaluated. We replace  $s$  by  $s'$  if  $\Delta < 0$ , otherwise a random variable  $X \sim U(0,1)$  is generated. If  $X < e^{(\Delta/T)}$ , we still replace  $s$  by  $s'$ . The counter  $n = n + 1$  is set whenever the replacement does not occur. Once  $n$  reaches  $L$ ,  $t = t + 1$  is set and  $T$  is a decreasing function of  $t$ . The procedure is stopped when the stopping condition for  $T$  is reached.

The main drawback of the SA is the algorithms sensitivity to the parameter settings. To overcome the difficulty of setting efficient values for parameters like temperature, an adaptive mechanism is used to detect frozen states and if warranted re-heat the system. In all experiments, the initial temperature was set at 400 and the algorithm stops after 2000 iterations. Each experiment was computed three times with different, random starting points to reduce the risk of local solutions. Among the three trials, we selected the solution with the lowest value of the objective function. The three solutions for each experiment varied slightly, but in an identical manner across the experiments. Hence, for the comparison of distance measure this choice is inconsequential.

Our adaptive scheme to dynamically adjust temperature work as follow: after 10 iteration with no improvement the temperature is increased according to  $newtemp = temp * 3^\beta$ , where  $\beta$  starts at 0.5 and is increased by 0.5 each time the system is reheated. As a result, the SA will never be in a frozen state for long. The temperature is decreased each iteration with a factor of 0.95. The settings above are a result of substantial, preliminary testing on this data and problem. In fact, some of the solutions were compared to those obtained by alternative heuristics.

The number of facilities is varied in the experiments. We consider locating small to medium number of facilities,  $P \in (2,100)$ . The location problem differs as a consequence, not only because  $P$  is varied. Figure 2a shows the solution for  $P = 5$ . The facilities lay far apart in the region and interurban travelling on the national road network is required for a large proportion of the

population. Hence, in this case the rural landscape with its natural barriers and so forth affects the solution indirectly since it has affected the infrastructural setting of national roads and the location of settlements. Consider on the other hand the experiment with  $P = 100$ .

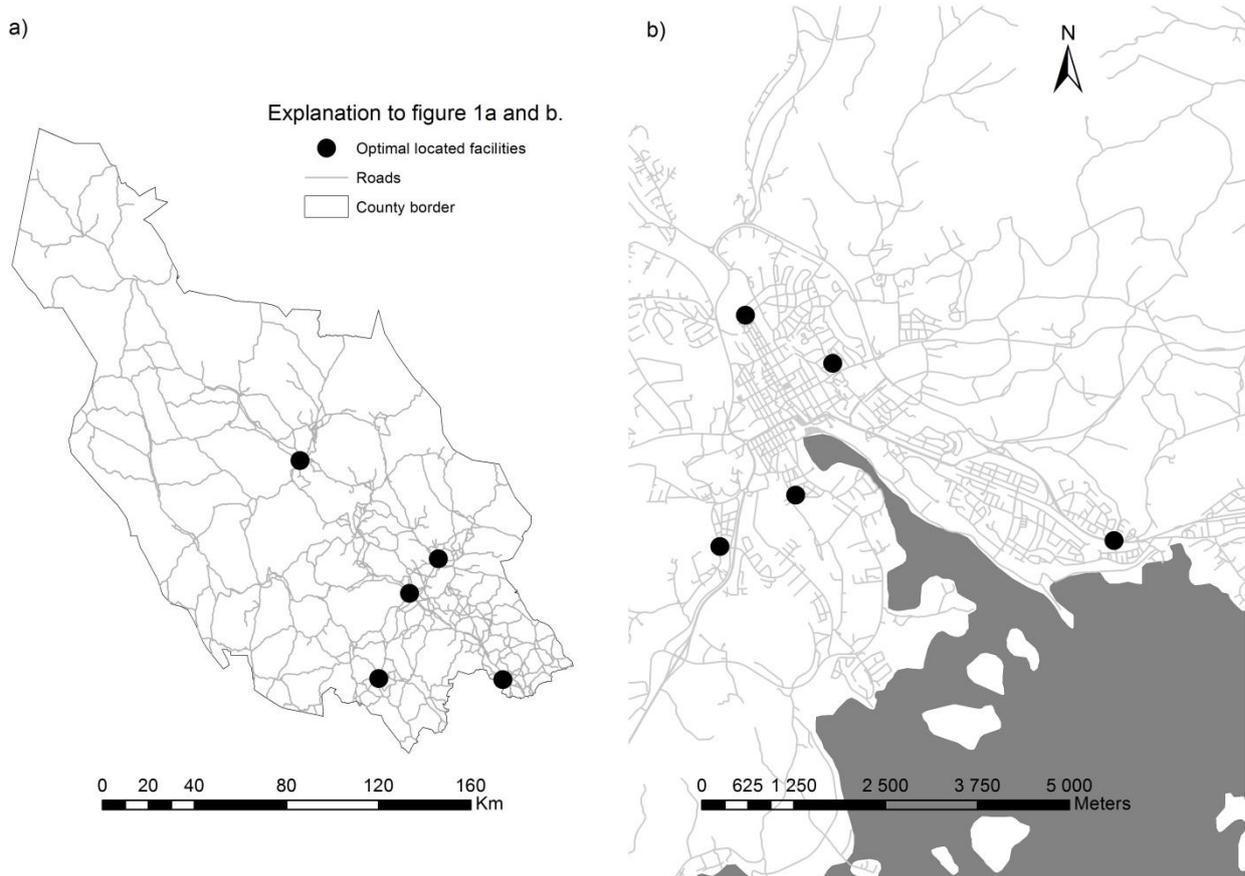


Figure 2: Solution of the  $p$ -median model in Dalecarlia for 5 and 100 facilities. (a) the solution of five facilities and the national road network. (b) the solution of 100 facilities and the road network with both national roads and local streets, focusing on the downtown area of the city of Falun.

Figure 2b shows the solution in the downtown area of the largest city in the region – Falun. There are five facilities located in this area and the population travels to the nearest facility primarily on the local streets in the city. In conclusion, the experiments for which  $P$  are small characterizes a  $p$ -median problem on a rural region with a non-symmetrical distribution of the population and a highly heterogeneous road network. For the experiments with a larger  $P$ , the setting resembles a problem in an urban area. Consequently, the results of the experiments may have some external validity outside this region which is under study.

## 4. Results

In this section, we take, as the benchmark, the solution to the  $p$ -median model when the travel-time is used as distance measure. Table 2 shows the average travel-time in seconds for the residents to their nearest facility in the experiments with  $P$  varying. For  $P = 2$  the average trip is about 25 minutes, a value that decreases to slightly more than 3 minutes for  $P = 100$ . The solutions based on the network distance are virtually identical to those of the travel-time distance as can be seen in Table 2 by comparing the average travel-time for the two measures. To complement the experimental results given in seconds, the travel distance in km on the road network for the residents to the nearest facility is shown on the last row of the table. To sum up, the finding is that the network distance, not accounting for the quality in the road network, produces the same solution to the  $p$ -median model as an elaborated distance measure that accounts for those aspects.

Table 2: The residents' average travel time in seconds to the nearest facility. The travel-time is evaluated for the solutions of the  $p$ -median model for the travel-time and the network measures. Last row gives the average network distance to the nearest facility.

Measure	$P$													
	2	5	8	10	15	20	25	30	35	40	45	50	75	100
<i>Travel-time</i>	1546	973	704	617	505	444	387	348	323	301	290	273	224	198
<i>Network</i>	1540	988	704	618	505	444	387	348	325	301	296	272	224	198
<i>Network (km)</i>	33.7	20.2	13.7	12.1	9.2	7.4	6.6	6.0	5.4	5.1	4.7	4.5	3.6	3.2

Solutions for the  $p$ -median model was also obtained based on the Euclidean measure, and the travel-time between the residents and their nearest facility computed. Generally, these solutions increased the residents' travel-time. Figure 3 shows a relative comparison between the Euclidean solution and the travel-time solution. As an instance, for  $P = 2$ , the average travel-time was found to be 1,630 seconds for the Euclidean solution and 1,546 seconds for the travel-time solution, giving a relative difference of 5.4 per cent. The relative difference was 3.6 per cent on average ranging from 0.0 per cent to 7.0 per cent.

In Figure 3, a regression line is imposed as a function of  $P$ . The significant estimate of the intercept is 2.6 and the estimate of the regression coefficient is 0.03, where the regression coefficient is borderline significant with a  $p$ -value of 0.06. Taken at face-value, the regression coefficient implies a one percentage point worsening of the Euclidean solution for each increment of  $P$  of 30 facilities.

To conclude, the Euclidean measure is potentially problematic since it may provide solutions to the  $p$ -median problem that leads to excessive travel times and distances for the population.

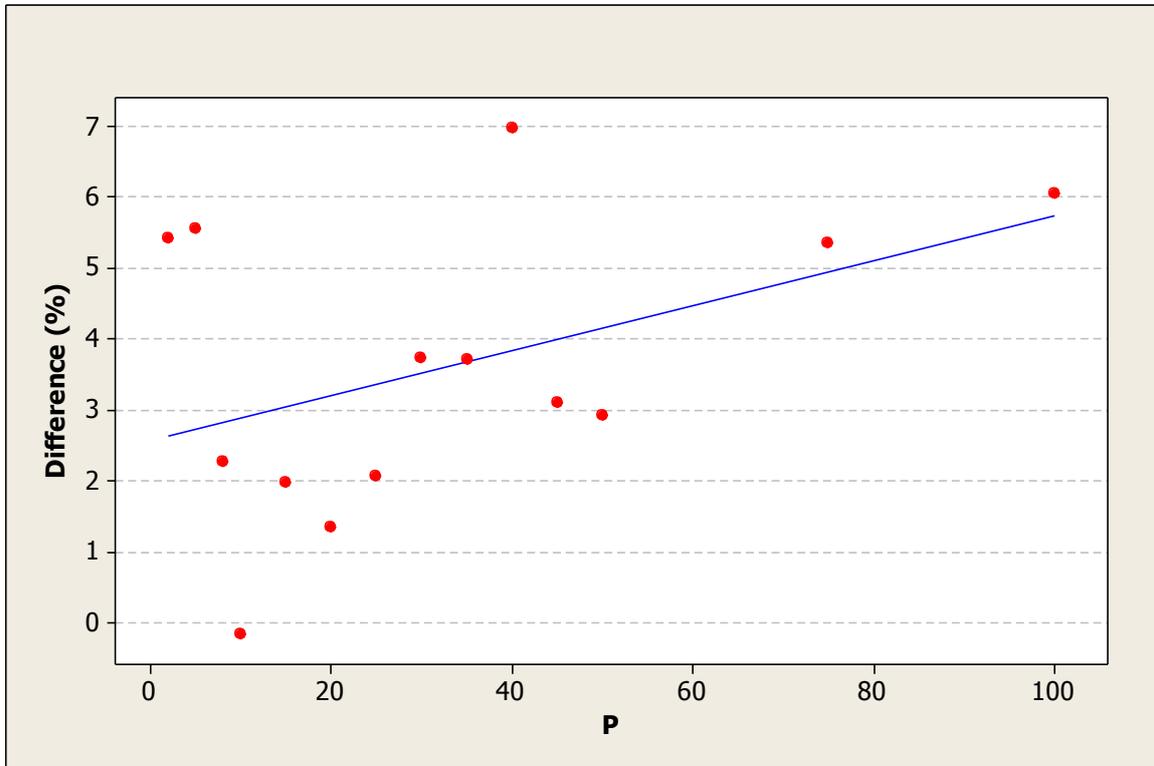


Figure 3: The relative difference between the solutions of the  $p$ -median model based on the Euclidean and the travel-time measure of distance.

## 4. Conclusions

In this study we have examined whether or not the distance measure is of importance when the  $p$ -median model is used to locate facilities. To do this, we have systematically varied  $P$  from small ( $P = 2$ ) to medium size ( $P = 100$ ) in a very dense network with attributed speed limits.

Two main conclusions can be drawn from this investigation. The first is that the Euclidean distance provides solutions to the  $p$ -median model that lead to excessive travel-time for the residents of as much as 7 per cent. The excess seems to increase with the number of facilities to locate.

The second conclusion is that the network distance provided equally good solutions to the  $p$ -median problem as an elaborated network. In spite of the fact that the elaborated network accounted for heterogeneity in the network due to variation in speed limits and the implied variation in road

quality. This finding is startling as the elaborated network showed substantial heterogeneity in terms of speed limits and implied road quality. It should be noted however that the network studied here is very refined and that the findings may not extend to a sparse network.

As a final remark, note that the variation in  $P$  has some implications for interpreting the findings for a rural setting. For  $P$  small, the setting is a problem of locating facilities in inter-urban environment where a large fraction of the population travels between towns to patronize the nearest facility. For the larger values of  $P$ , it is a setting where multiple facilities are located within the towns and the residents travel primarily on local streets within the towns. Hence, we assert that the findings bear some relevance for location problems in urban settings, in addition to rural ones.

## Acknowledgements

Financial support from the Swedish Retail and Wholesale Development Council is gratefully acknowledged.

## References

- Al-khedhairi, A., (2008), Simulated annealing metaheuristic for solving  $p$ -median problem. *International Journal of Contemporary Mathematical Sciences*, 3:28, 1357-1365, 2008.
- Bach, L. (1981). The problem of aggregation and distance for analyses of accessibility and access opportunity in location-allocation models. *Environment & Planning A*, 13, 955–978.
- Berman, O., and Krass, D. (1998). Flow intercepting spatial interaction model: a new approach to optimal location of competitive facilities. *Location Science*, 6, 41–65.
- Carling K., Han M., and Håkansson J, (2012). Does Euclidean distance work well when the  $p$ -median model is applied in rural areas?, *Annals of Operations Research*, Online September 1.
- Daskin, M.S., (1995). *Network and discrete location: models, algorithms, and applications*. New York: Wiley.
- Dijkstra, E.W., (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269–271.
- Drezner, T., and Drezner, Z., (2007). The gravity  $p$ -median model, *European Journal of Operational Research*, 179, 1239-1251.

Francis, R. L., Lowe, T. J., Rayco, M. B., & Tamir, A. (2009). Aggregation error for location models: survey and analysis. *Annals of Operations Research*, 167, 171–208.

Hakimi, S.L., (1964). Optimum locations of switching centers and the absolute centers and medians of a graph, *Operations Research*, 12:3, 450-459.

Hale, T.S., and Moberg, C.R. (2003). Location science research: a review. *Annals of Operations Research*, 32, 21–35.

Han, M., Håkansson, J., and Rebreyend, P., (2012). How does the use of different road networks effect the optimal location of facilities in rural areas?, *Working papers in transport, tourism, information technology and microdata analysis*, ISSN 1650-5581.

Handler, G.Y., and Mirchandani, P.B., (1979). *Location on networks: Theorem and algorithms*, MIT Press, Cambridge, MA.

Kariv, O., and Hakimi, S.L., (1979), An algorithmic approach to network location problems. part 2: The  $p$ -median. *SIAM Journal of Applied Mathematics*, 37, 539-560.

Kirkpatrick, S., Gelatt, C., and Vecchi, M., (1983), Optimization by simulated annealing. *Science*, 220:4598, 671-680.

Murray, A.T., and Church, R.L., (1996). Applying simulated annealing to location-planning models, *Journal of Heuristics*, 2, 31-53.