J Geograph Syst (2000) 2:167–184

Journal of Geographical Systems

Distance predicting functions and applied location-allocation models.

Some simulations based on the l_p norm and the k-median model

Dominique Peeters[1], **Isabelle Thomas**[1,2]

[1] Department of Geography, Center of Operations Research and Econometrics
[2] National Fund for Scientific Research
Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
(e-mail: Isabelle@geog.ucl.ac.be)

Received: 12 October 1998 / Accepted: 17 September 1999

Abstract. Distances between demand points and potential sites for implementing facilities are essential inputs to location-allocation models. Computing actual road distances for a given problem can be quite burdensome since it involves digitalizing a network, while approximating these distances by l_p-norms, using for instance a geographical information system, is much easier. We may then wonder how sensitive the solutions of a location-allocation model are to the choice of a particular metric. In this paper, simulations are performed on a lattice of 225 points using the k-median problem. Systematic changes in p and in the orientation of the orthogonal reference axes are used. Results suggest that the solutions of the k-median are rather insensitive to the specification of the l_p-norm.

Key words: Location, allocation, p-median, distance predicting function

JEL classification: C6, R3, R4

1 Introduction

Facility location analysis deals with the problem of locating new facilities with regard to existing facilities and clients in order to optimize one or several economic criteria. During the last three decades, there has been an outburst of developments in this area; Hansen et al. (1987) or Labbé et al. (1995) give an overview of the progress realized. Besides the theoretical works, one could also witness an urgent call for applications, both in the private and public sectors. The literature is plentiful with case studies for ambulance and fire protection services, schools, health care facilities, post offices, garbage dumps, electrical power stations, warehouses, department stores, ... Now, a successful application requires a satisfactory representation of the geographical and economic environment of the problem: the demands of the clients, the costs related to the implementation and the operations of the facilities in the various sites, the costs related to the movements of people and commodities, the impact of the nuisance generated by the system, etc. As this generally goes along with a huge data collecting step, a major issue in the development of location-allocation software is the design of an interface with a *geographical information system*, which offers more capability to handle georeferenced data and provides the model with its requested inputs. Moreover, such a GIS should also be useful at assessing the outcomes of the model.

In most real-world location-allocation applications, the *environment* is represented by a discrete space where demand and potential supply are located at nodes and distances are measured along links drawn between nodes. Many simplifications are made in order to represent this environment, leading to many well-known spatial data analysis problems: how much does the level of data aggregation affect location-allocation results (e.g. Current and Shilling 1989; Fotheringham et al. 1995; Daskin et al. 1989; Ruhigira 1994; Plastria 1995; Francis et al. 1996; Hodgson et al. 1997; Bowerman et al. 1999; Erkut and Bozkaya 1999; Francis et al. 1999)? How much does the measure of demand influence location-allocation results (e.g. Beguin et al. 1992; Thomas 1993; Owen and Daskin 1998)? Does the shape of the road network affect the location-allocation results (e.g. Peeters and Thomas 1995)? Do boundaries affect location-allocation models (e.g. Hodgson and Oppong 1989)? Data problems are numerous and often still unsolved.

Obviously, one of the most time consuming step in applied facility location analysis is related to the *measurement of distances* on a communication network. Indeed, it involves the digitalization of the network, the evaluation of the velocity along the different links, the computation of shortest paths, the verification of generally very large distance matrices, for this lengthy procedure is much error-prone. Hence, distance predicting functions are used in order to transform co-ordinates differences between two points into an estimate of the travel distance between them (e.g. Brimberg and Love 1993b; Love and Morris 1972, 1979; Love et al. 1995). *Distance predicting functions* allow rapid estimation of unknown actual distances between pairs of points in a geographic region and can be easily implemented in geographical information systems. A commonly used function is the l_p norm (see e.g., Love and Morris 1972, 1979; Brimberg and Love 1992, 1993a,b; Muller 1982). Fitting an l_p norm to distances on a transportation network means finding empirical values of a parameter p as well as an orientation θ of the reference axes (e.g. Brimberg and Love 1992, 1993a; Brimberg et al. 1995). The empirical fits depend upon the studied example (see Sects. 2.2 and 2.3 for more details) and are estimated by regression methods (e.g. Brimberg et al. 1996).

Of course, approximating actual network distances by l_p distances when solving a location-allocation problem is a potential source of difficulties. The optimal solution of the original problem may indeed differ from the solution of the approximated problem. One might then wish to evaluate the gap between these solutions. However, due to the infinite variety of layouts of data points, it seems extremely hard to provide a general answer to this question. A second interrogation is also of interest: **how sensitive is the solution of the approximated location-allocation problem to modifications of the parameters defining the $\mathbf{l_p}$ norm**? The answer should tell whether the degree of accuracy for measuring the parameters p and/or θ really matters for approximating the solution of the original problem. In this paper, we will investigate this question. More particularly, we will test the *robustness* of the solutions of a representative location-allocation model (the k-median) to variations of the parameters characterizing the l_p norm. Finally, a word of caution: it should be clear that our aim is not to discuss the confidence intervals for unknown distances (e.g. Love et al. 1995), nor to find the best estimates for one particular case study (e.g. Berens and Körling 1985).

The remainder of this paper is organized into three sections. Section 2 outlines the formal terms of the location-allocation problem and the distance measurement problem; it describes the design of the experiments: the settlement system (§2.1), the distance predicting function (§2.2 and 2.3), how irregularities are introduced in the grid (§2.4), and the location model chosen (§2.5). In Sect. 3 we detail the effects of p (§3.1) and θ (§3.2) on the location results when $k = 3$, while we summarize the findings for irregular lattices of points (§3.3) and for other values of k (§3.4). The final section gives some concluding comments.

2 Design of the experiment

2.1 The settlement system

Choosing a regular lattice rather than an irregularly shaped layout enables one to isolate the tested problem from many other sources of variation and to control for the spatial layout (e.g. Anselin 1986; Haining 1986; Peeters and Thomas 1995). The chosen hypothetical settlement system can be seen to differ from the shapes of many observed settlements encountered in socio-economic research. However, it enables the researcher to conduct the tests without the additional complexities introduced by empirical lattices, and leads to more clear-cut interpretations of the results.

The spatial layout of the data set is a regular 15-by-15 squared lattice of points ($N = 225$ points). The choice of the shape and size (N) of the lattice is governed by preceding results (e.g. Peeters and Thomas 1995; Arnold et al. 1996): it appears to be the best compromise between feasibility, reasonable computing time and practical results. Each point i is a demand point as well as potential supply site. Each point i is characterized by its (x_i, y_i) co-ordinates, and is a node bound by a link to its eight closest neighbors. The way of measuring the length of the link is defined in Sects. 2.2 and 2.3. Demand is equal for all nodes and normalized to one.

2.2 The distance predicting function

The Minkowsky distance (or l_p-norm) is often used as a surrogate for travel time as it often provides a good fit to network distances (e.g. Love and Morris 1972, 1979; Brimberg and Love 1992, 1993b; Huriot and Perreur 1990). Its functional form is given by:

$$l_p(m,n) = (|d_x|^p + |d_y|^p)^{1/p}$$

where $d_x = |x_m - x_n|$ and $d_y = |y_m - y_n|$.

When $\boldsymbol{p<1}$, which is known as the *hyperrectilinear* case (see e.g. Juel and Love 1985), the triangle inequality postulate is not always satisfied; hence l_p is not a norm in the topological sense. On the l_p-sphere, the circle bends inward (see Fig. 1 which shows the locus of points whose distance with respect to a

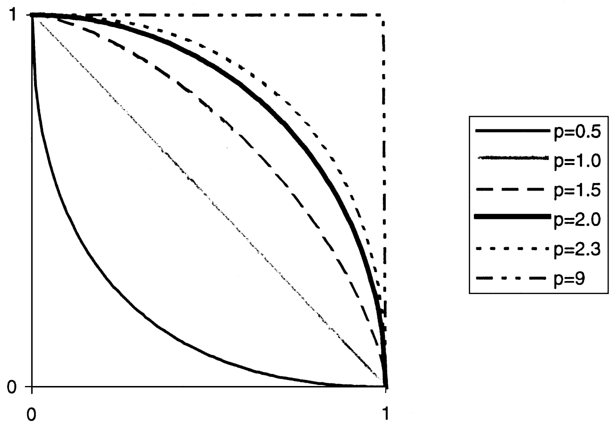

Fig. 1. Mathematical spheres for some critical values of p

central point remains constant). A value of p smaller than 1 may however be not totally unrealistic in applied location studies: Muller (1985, p. 191) gives the example of a motorist taking a detour to avoid traffic as an evidence of a situation where the triangle inequality rule is no longer applicable.

When $\boldsymbol{p=1}$, the distance is called rectangular or *rectilinear* (Manhattan or l_1). The isovectors on the l_p-sphere are diamond-shaped: they take the shape of a square from which the diagonals are parallel to the axes (Fig. 1). For every $\boldsymbol{p>1}$ and finite, the spheres are strictly convex. When $\boldsymbol{p=2}$, distance we have the well-known *Euclidean* distance. For values of $\boldsymbol{p>2}$, the corresponding Minkowsky distances are shorter than the Euclidean distance (Fig. 1) (see e.g. Muller 1985, p. 191 for the relevance of such values of p). The balls in the limit case $\boldsymbol{p=\infty}$ are squares whose sides are parallel to the axes.

Examples of empirical fittings of l_p-distances functions to actual data are found in the literature. They aim at finding the best surrogate for travel time between cities or within cities. Empirical values of p vary from 0.9 to 2.29 (e.g. Love and Morris 1979; Berens and Körling 1985; Brimberg and Love, 1993b). Most often cited measures of p vary between 1.5 and 2.0. Estimating empirically a travel time by an l_p-distance enables one to avoid the time-consuming task of measuring travel time for each link of the network. The aim of this paper is *not* to find the best estimation of travel time, but to test the effects of changes in p on location-allocation results, and therefore to examine the necessity of finding an accurate value of p for distance estimation in location-allocation modeling. 17 values of p are chosen in accordance with the literature about empirical fittings of p as well as with the geometry of space. In our example, p varies from 1.0 to 9: values between 1.0 and 2.3 were selected with a step of 0.1, while outside this range 3, 6, and 9 were chosen. Because of the properties of the l_p-distance, values of p smaller than 1 were not kept in the simulations.

No weight to distance is considered: some authors use a weight factor (sometimes two, one associated with each direction) which accounts for peculiarities in the network: hills, bends, rivers, etc. (e.g. Love and Morris 1972, 1979; Brimberg and Love 1992, 1993a,b). This should make our network more irregular and the results of the test less clear to interpret. Furthermore, we assume the same value of p throughout the entire network; in practical situations, it could be advisable to regionalize the value of the parameter to capture the local characteristics of the transportation network.

In our simulations, 17 measures of p in the l_p-distance are considered.

2.3 The orientation of the reference axes

More accuracy in estimating travel time by an l_p-distance can be obtained by allowing a rotation of the orthogonal reference axes (e.g. Huriot and Perreur 1973; Love and Morris 1973; Brimberg and Love 1992, 1993a,b, Brimberg et al. 1995). θ is the angle measuring the rotation of the orthogonal reference axes X and Y; the original reference axes are those parallel to the edges of the lattice of points. θ is measured counter clockwise; it measures the relative difficulty to travel in any direction: if $p < 2$, it is easier to travel along the directions of the axes, while if $p > 2$, it is easier to follow the direction of the bisectors; when $p = 2$, space is isotropic (e.g. Love and Morris 1979; Brimberg and Love 1993a). In accordance to the demonstrated symmetry about $\Pi/4, \theta$ takes the values 0°, 15°, 30° and 45° in our simulations. These rotations are applied for all p values.

For each of the 17 values of p, 4 values of θ in the l_p-distance predicting function are considered, leading to a total number of 68 simulations.

2.4 Irregular lattices

The choice of a regular lattice of points may be questioned on grounds that it yields too specific results. To assert the robustness of our findings on irregular distributions, we have proceeded in the following way. Each of the 225 points of our regular lattice has been considered as the mean of a bivariate normal distribution characterized by the same standard error σ in the x and y directions. For a given value of σ and each of the 225 distributions, a single point has been created using a pseudo-normal generator. The selected values for σ, which can be viewed as a measure of disorder of the lattice, were 0.00 (regular), 0.25, 0.50, and 1.00. On the four lattices obtained in this way, selected values of p and θ detailed in subsection 3.3 were used to measure the distances.

Experiments were also carried out using a set of 225 points randomly generated in a square. This can be considered as a maximum disordered set.

2.5 The location model

The k-median model is applied (Hakimi 1964). It yields a configuration of k supply sites such that the sum of the distances between the demand points and their respective closest supply site is minimized. The model has been chosen for its high flexibility, the good performance of the algorithms designed to solve it, and its common use in location-allocation applications. Most importantly for our purpose in this paper, distance is one of the key elements in the formulation of the k-median problem. Distance is often used as a surrogate for travel disutility encountered by the customers; thus using a «good» measure of distance is one of the crucial conditions to obtain meaningful solutions. As computing actual distances between points in real spaces is generally time-consuming and prone to errors, approximating those distances by analytical expressions using the coordinates of the points, as geographical information systems allow to do it, seems a priori an interesting idea. Exact solutions of the *k*-median model were computed using a program outlined in Hanjoul and Peeters (1985).

But the optimal configuration of supply sites and the value of the total distance are not the only characteristics of a solution of a *k*-median problem. The number of demand points allocated to a site gives the level of activity of this site, thus its size in terms of human and material resources that should be affected. The shape of the area to be supplied by a site and the corresponding distribution of distances between the demand points and the site are two other important outcomes. The latter can be characterized either by its mean value, its maximum value, its variance, a Lorenz curve, etc. All the measures of spread are to be interpreted in terms of inequity among clients.

In the studied example, the number of supply sites k to be optimally located may theoretically vary between 1 and 225. In this paper, we decided to focus on the results obtained for $k = 3$; according to our experience, it was the value that best isolates the tested effect as it leads to a symmetric solution in the chosen lattice of points. However, to assess the robustness of the findings, experiments with other values of k were also performed. The results are synthesized in Sect. 3.4.

3 Computational experiments

3.1 How much does the choice of p affect the k-median results?

Let us consider a first set of simulations: a regular squared lattice of 225 points, reference axes parallel to the lattice of points ($\theta = 0°$) and p varying from 1 to 9. Let us consider the outputs of the *k*-median for the 17 selected values of p.

3.1.1 Average and maximum distances

Two measures of distance commonly help to evaluate the *k*-median results: (i) the average distance (D_{ave}) between all demand points and their respective closest optimal supply site gives a measure of efficiency of the proposed solution, and (ii) the maximum distance ($D_{\max}$) between a demand point i and the closest facility can be considered as a measure of the equity of the optimal solution. Their variations with p are respectively reported in Figs. 2 and 3.

A *3*-median model was applied for each of the 17 values of p. Spatial differences between the solutions were found because of the isotropy of the studied sample of points: symmetric solutions were suggested by the model. In this particular case, the three optimal sites are located at the vertices of a triangle, sometimes on its base, sometimes on its summit, sometimes rotated from 45° or 90°. Hence, comparability of the optimal locations was difficult for geometric reasons only: for a given value of the objective function, different solutions can be suggested (*alternative optima*). A second set of simula-

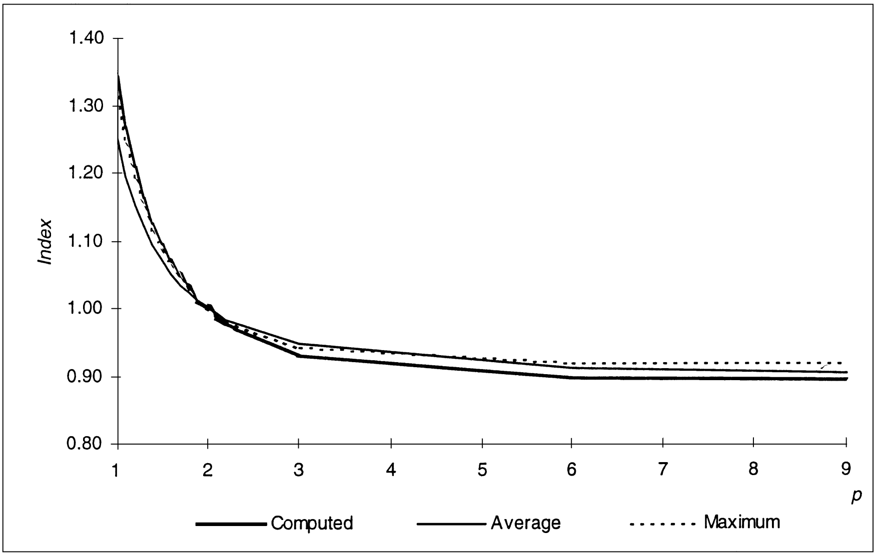

Fig. 2. Relative variation with p of the computed l_p-distance (*computed*) and two outputs of the *3*-median model (the *average* distance and the *maximum* distance (Ratio $= X_p/X_{p=2}$)

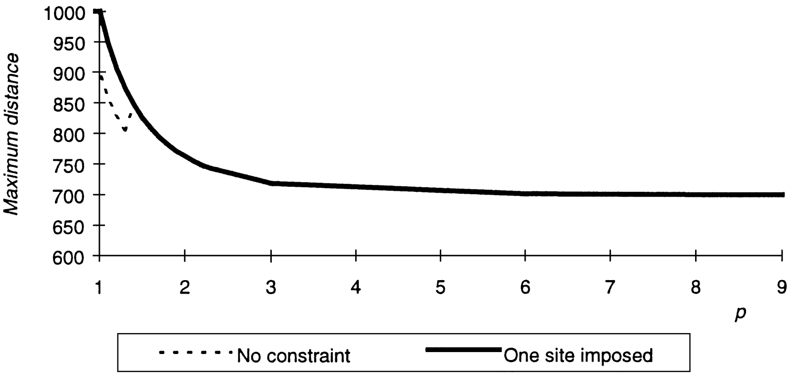

Fig. 3. Computed variation of $D_{\max}$ with p for the *3*-median model applied with no constraint and with one imposed location

tions was therefore performed imposing one location to the model. This enables one to avoid natural symmetries, to better compare the spatial solutions and to interpret the tested effects without confusion. In this studied example, imposing one location has no effect on the average distances and little on the maximum distance for values of $p < 1.5$ (Fig. 3). The following sections will refer to the constrained model only.

Both distances ($D_{\text{ave}}, D_{\max}$) (Figs. 2 and 3) decrease with p, in accordance to the definition of the l_p-distance (see Sect. 2). Figure 2 compares the relative variation with p of the average distance (D_{ave}), the maximum distance ($D_{\max}$) as well as the expected distance (distance simply computed with the l_p

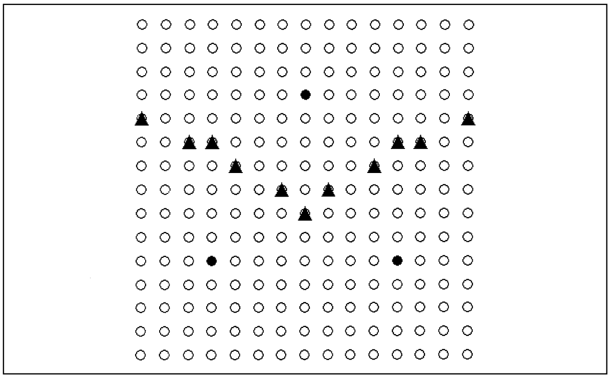

Fig. 4. Optimal locations and changes in allocation with p when $\theta = 0°$. •: optimal location; ○: demand point; △: changes in allocation

formula). Differences are slight and insignificant. The ratio of the maximum distance ($D_{\max}$) to the average distance (D_{ave}) is often considered as a balance between equity and efficiency. In our simulations this ratio varies regularly with p, but the fluctuation is very slight: between 2.15 to 2.26 with a minimum value when $p \in [1.8, 3.0]$. Relative differences in the computed values of D_{ave} and $D_{\max}$ are not insignificant with p, *but* regular.

p influences the absolute value of the average and the maximum distances; the observed variations with p are proportional to the expected one (Fig. 2).

3.1.2 Optimal locations and allocation

For $k = 3$ and forcing one location to be open at site $j = 53$, the same two other *locations* are always retained as optimal, for every p in $[1, 9]$ (Fig. 4): this suggests that for this set of experiments, the optimal locations are *not* sensitive to the choice of the l_p-distance.

Each demand point i is *allocated* to its closest optimal supply site j. Each supply point j is therefore characterized by a zone (a market area) including the allocated demand points and by the size of the facility, which is proportional to the number of points in the supply area. Both criteria (quantity of demand allocated to the site; shape of the zones) are used to evaluate the k-median results. In accordance with the preceding results, the design of the optimal market areas are almost the same whatever p in $[1, 9]$. Figure 4 gives the 3 optimal supply sites j as well as the 11 demand points i whose allocation changes with p. The observed differences are located along the perpendicular bisectors of the triangle formed by the 3 optimal locations and are observed for very large values of p ($p > 3.0$). These locations are explained by the geometry of the studied environment and can easily be simulated by Voronoi diagrams.

Hence, errors made on the use of p when estimating distance by the l_p-distance have little influence on the optimal locations and allocations proposed by the k-median model. No difference is observed for commonly used values of p (see Sect. 2.2).

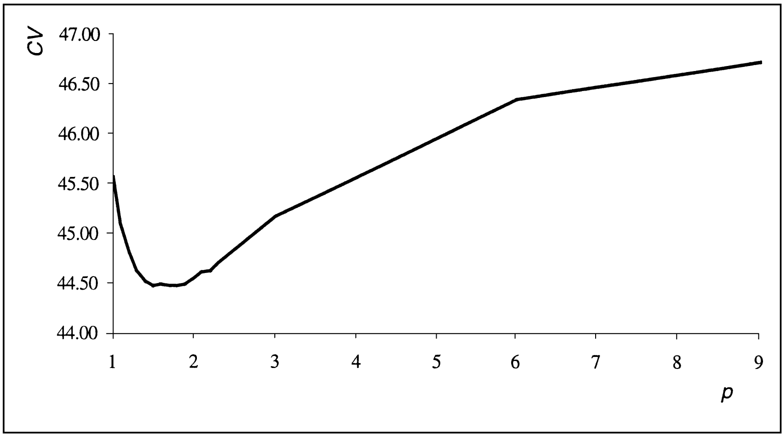

Fig. 5. Variations with p of the CV measured on the distances distribution, when $\theta = 0°$

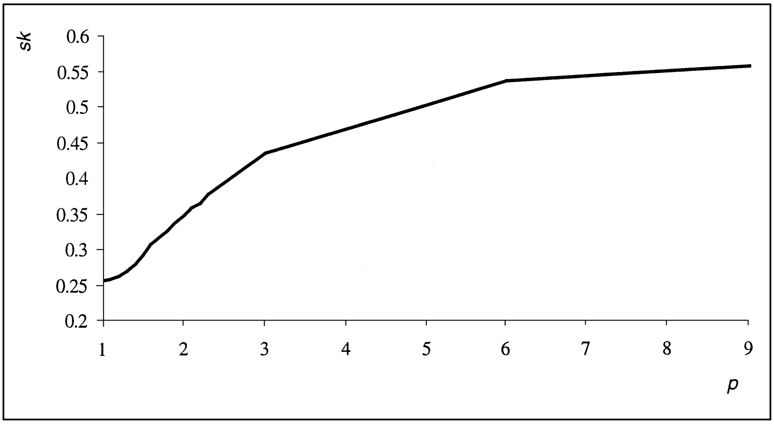

Fig. 6. Variations with p of sk measured on the distribution of distances when $\theta = 0°$

3.1.3 Distribution of distances

Let us make a last verification of the tested hypothesis. The distribution of the distances between each demand point i and its closest optimal supply site j is another way of evaluating the efficiency and the equity of the optimal solutions of location-allocation models. Figures 5 to 7 illustrate the variation of three simple statistics that describe the statistical distributions of a variable: (***1***) the coefficient of variation (CV) which measures the dispersion of a variable independently of the size of the data values; (***2***) the coefficient of skewness (sk) takes the value 0 when the data are perfectly symmetric, it is positive when the mean is larger than the median, and negative when the mean is smaller than the median. sk measures the tendency of a distribution to stretch out in a particular direction; (***3***) the kurtosis (K) measures the peakedness of the distribution.

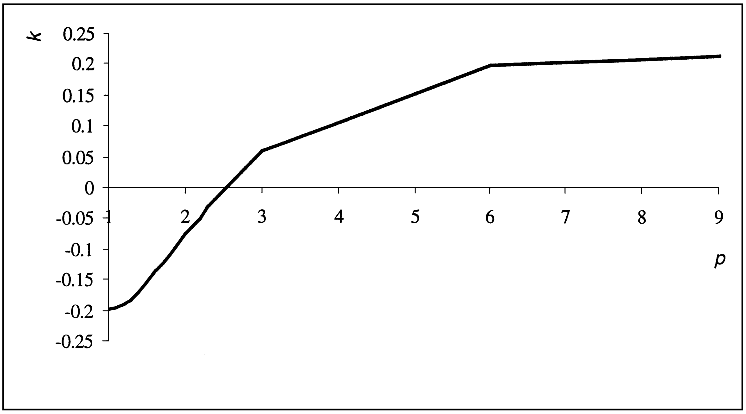

Fig. 7. Variations with p of K measured on the distances distribution when $\theta = 0°$

Whatever the studied indicator, variations with p are very slight and insignificant; they are, however, regular with p: sk and K increase regularly with p, CV reaches minimum values when $p \in [1.5, 2.1]$.

3.1.4 Conclusions

Within the limits of this experiments, it appears that p has little influence on the location-allocation results, especially in the range widely used in practice ($p \in [1.5, 2.0]$). Optimal locations do not vary with p. Optimal allocations change only slightly at the borders of the market areas. The absolute values of the average and maximum distances vary only as theoretically expected by the distance formulation.

3.2 How much does the orientation of the reference axes affect the k-median results?

Let us consider here a regular squared lattice of 225 points and 17 values of p, but this time with 4 different values of θ (see Sect. 2.3).

3.2.1 The effect of θ on the average and maximum distances

The *average distance* D_{ave} was computed for each p and θ values. We observed almost no difference with θ between average distances computed by the *3*-median model: D_{ave} is not significantly affected by the directional changes of the reference axes, whatever p in $[1, 9]$. *Maximum distances* are reported in Fig. 8 for each studied value of p and θ. Differences between maximum distances are small: for a fixed value of p, the maximum observed difference is less than 10%, whatever θ. When p is greater than 1.5 and smaller than 2.3, differences are the smallest: they are less than 3.0%.

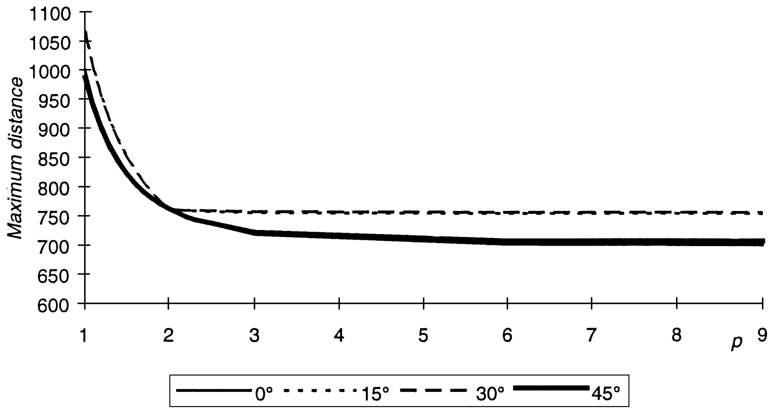

Fig. 8. Variation with p and θ of the maximum distance ($D_{\max}$)

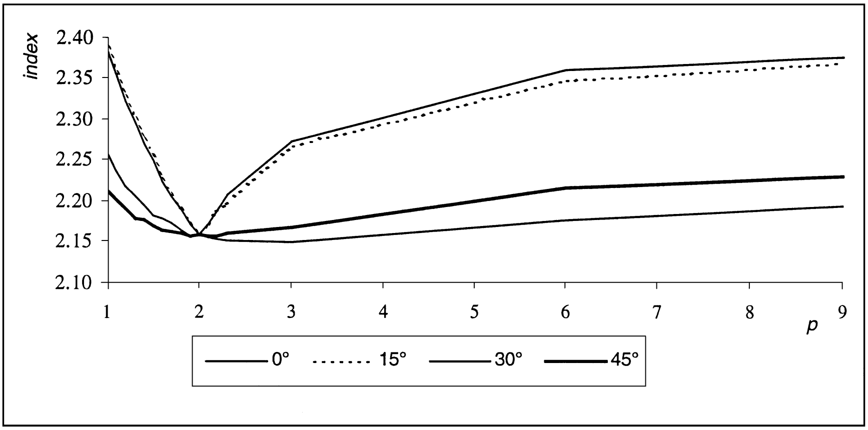

Fig. 9. Variation with p and θ of the ratio between maximum ($D_{\max}$) and average distances (D_{ave})

As a first conclusion, average and maximum distances are slightly influenced by the orientation of the reference axes.

Figure 9 compares the maximum distance to the average distance for each value of p and θ. The computed ratio expresses the balance between equity and efficiency: the highest the index, the more equity is privileged against efficiency. Differences are *not* important: the index varies between 2.1 and 2.4. When $p > 2.0$, $\theta = 0°$ solutions are always the best balance between equity and efficiency. When $p < 2.0$, $\theta = 45°$ is the most balanced solution; this is to be explained by the symmetry demonstrated by Brimberg and Love in 1993b. This figure also illustrates the equivalence of the solutions when Euclidean distance is considered ($p = 2$) and the resemblance between $\theta = 30°$ and $\theta = 15°$ solutions. Differences are small and regular with p.

As a conclusion, the two most often used *3*-median operational outputs are almost not affected by the direction of the reference axes as far as the commonly used values of p in the l_p-distance are concerned ($p \in [1.5, 2.3]$).

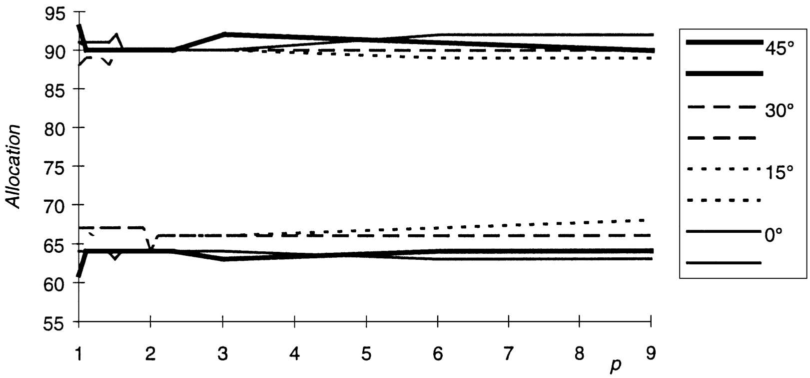

Fig. 10. Variations with p and θ of the minimum and maximum allocation

Errors on the use of p when estimating travel time (see Sect. 2) do almost not affect the *3*-median results as far as D_{ave} and $D_{\max}$ are concerned. Operational location results (choice of j, allocation of i to j) are the same whatever p and θ.

3.2.2 The effect of θ on the optimal location and allocation

One site is fixed open at $j = 53$; the other two *optimal sites* proposed by the *3*-median are almost always the same as those illustrated in Fig. 4. In 1 case out of 68, a shift of one site is observed; it occurs for a value of $p = 1.0$. In all other cases, whatever θ and p, the same supply sites are chosen and hence the solutions are stable.

Figure 10 shows the variation with p and θ of the quantities of *demand allocated* to the 3 optimal supply sites. Small differences between minimum and maximum allocation correspond to a balance between supply sites; large differences should not be chosen by the decision-maker especially when the service is not hierarchically organized. The figure shows that all p and θ values naturally lead to almost the same balanced configuration: p and θ have little influence on the allocation results (difference smaller than 10%). For geometric reasons only, values of p close to 2.0 lead to totally identical values of maximum allocations.

Figure 11 illustrates the demand points i whose allocation varies with p for a fixed value of θ, that is to say the main differences in the shape of the market areas for a given set of reference axes. Two cases are illustrated: $\theta = 15°$ and $\theta = 45°$. In all computed cases, differences are not numerous (17 points out of 225) and are always located on the perpendicular bisectors of the triangle formed by the three optimal locations.

The same maps were designed for fixed values of p and different values of θ. One case is illustrated ($p = 1.1$, Fig. 12): differences in allocation are again located on the perpendicular bisectors of the triangle and are not numerous (13 pairs out of 225). As already mentioned, the differences are located on the perpendicular bisectors and can easily be designed by simple Voronoi diagrams.

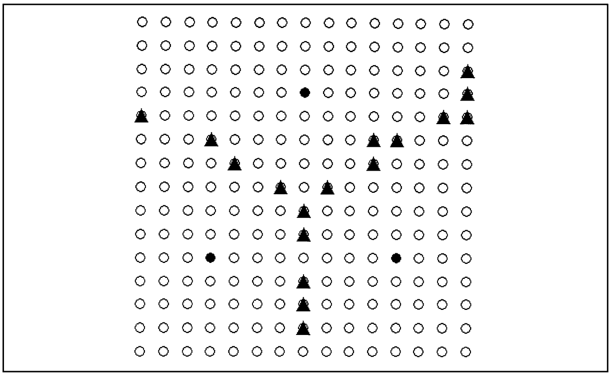

Fig. 11. Optimal locations and allocation changes with p, when $\theta = 15°$ and $\theta = 45°$. •: optimal location; ○: demand point; △: changes in allocation

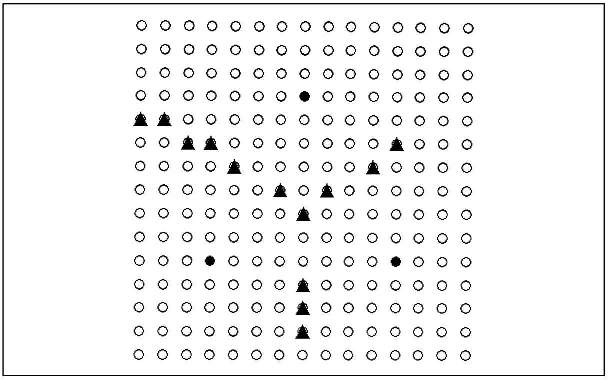

Fig. 12. Optimal locations and allocation changes with θ, when $p = 1.1$. •: optimal location; ○: demand point; △: changes in allocation

3.2.3 The effect of θ on the distributions of distances

As in Sect. 3.1, distribution of distances are summarized by three indicators: CV, sk and K; Figs. 13, 14 and 15 refer to their variations with p and θ. Once again, differences are slight and not significant, *but* regular with p. Euclidean distances are of course not sensitive to θ changes. For geometric reasons only, solutions with $\theta = 0°$ are symmetric to those with $\theta = 45°$.

3.2.4 Evaluation of the relative effects of θ and p

Tables 1 and 2 summarize the amplitude of the effects of the changes in p and θ on four previously measured indicators: the average and maximum distances, the ratio between equity and efficiency, and the coefficient of variation measured on the individual distances between each demand point and its closest supply site.

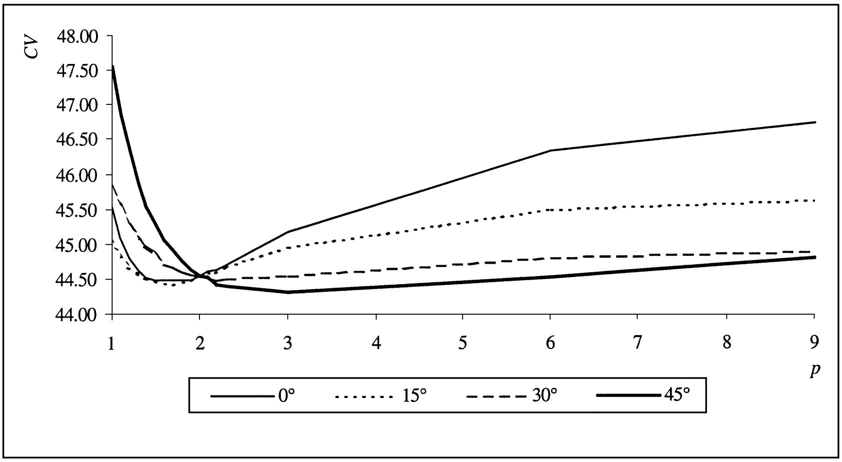

Fig. 13. Variation of the variation coefficient (CV) with p and θ

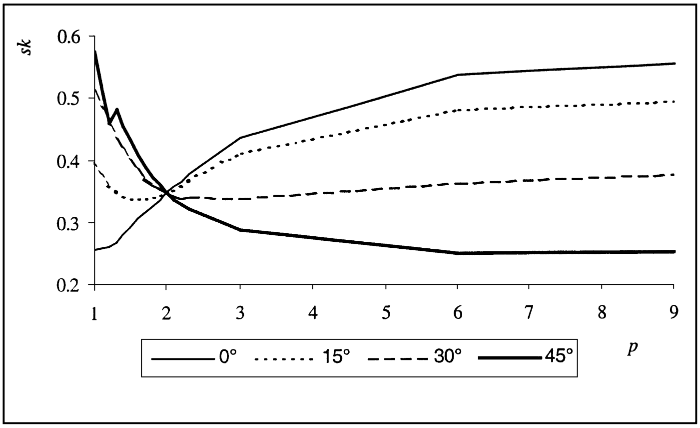

Fig. 14. Variation of the skewness (sk) with p and θ

Table 1 gives the relative changes of the 4 selected indicators with p. For each value of θ, the relative difference of each indicator is computed. Differences on the absolute values of D_{ave} and $D_{\max}$ are important due to the distortion of the space $(p \in [1, 9])$. Computing the differences for the ratio of D_{ave} to $D_{\max}$ lead to differences smaller that 11%. CV is quite stable: CV is a relative value and has a maximum difference of 7.3% between solutions. Relative differences with p between computed indicators are always smaller than 10%.

Table 2 gives the relative changes with θ for 4 fixed values of p. Changes are slight whatever the studied indicator (smaller than 10%).

Whatever p in $[1, 9]$ and θ, differences in location-allocation results are almost always less than 10%, that is to say little in comparison to many other sources of errors that can be introduced when measuring real-world inputs.

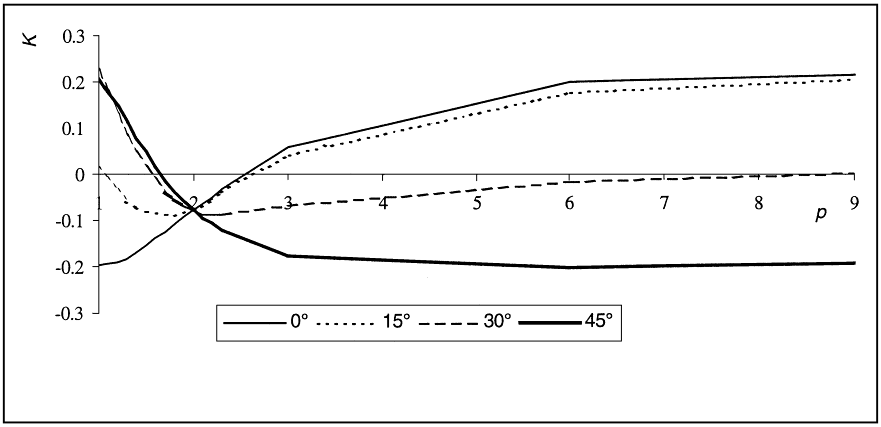

Fig. 15. Variation of the kurtosis (K) with p and θ

Table 1. Maximum differences between the computed D_{ave}, D_{max}, and CV for some fixed values of θ. (index = 100^* $(X_{max} - X_{min})/X_{min}$)

θ	D_{max}	D_{ave}	*Ratio*	*CV*
0°	42.9	38.9	5.0	5.1
15°	41.5	40.6	10.6	2.7
30°	40.9	40.6	10.4	3.0
45°	39.6	40.7	3.5	7.3

Table 2. Maximum differences between observed values for some fixed values of p. (index = $100^*(X_{max} - X_{min})/X_{min}$)

p	D_{max}	D_{ave}	*Ratio*	*CV*
1.0	8.1	0.9	7.8	5.6
1.5	3.6	0.3	3.9	1.9
2.2	1.6	0.0	1.6	0.5
9	8.0	0.6	8.3	4.3

3.3 How far does irregularity in the lattice of points affect the k-median results?

In order to evaluate the robustness of the conclusions discussed in the preceding sections, further simulations were performed on irregular lattices of points. Two sets of simulations were performed.

In the first set of experiments, described in Sect. 2.4, p takes respectively the values of 1.0, 1.5, 2.0 and 3.0 and θ the values of 0°, 30°, 45° and 60° (values of θ have much more sense for irregular lattices than for regular ones).

Table 3. Variation with p, σ and θ of the 3-median solutions (index of the 2 optimal locations)

σ	p	0°	30°	45°	60°
	1.0	154–162	154–162	154–162	154–162
0.00	1.5	154–162	154–162	154–162	154–162
	2.0	154–162	154–162	154–162	154–162
	3.0	154–162	154–162	154–162	154–162
	1.0	153–162	154–162	154–162	154–162
0.25	1.5	154–162	154–162	154–162	154–162
	2.0	154–162	154–162	154–162	154–162
	3.0	154–162	154–162	154–162	154–162
	1.0	153–177	147–170	162–170	138–162
0.50	1.5	153–162	147–170	164–162	138–162
	2.0	153–162	153–162	153–162	153–162
	3.0	147–170	153–162	153–162	153–162
	1.0	91–176	138–162	110–162	138–162
1.00	1.5	138–162	138–162	138–162	138–162
	2.0	138–162	138–162	138–162	138–162
	3.0	138–162	138–162	138–162	138–162

Hence, 16 combinations of p and θ are possible for each value of σ. For a 3-median model (with one location being fixed in $j = 53$), each value of σ leads to a different spatial solution (local optima) but there is no/little change in the optimal locations with p and θ for a given σ (see Table 3).

The second set of simulations (17 values of p, 4 values of θ) were performed on an irregular lattice of 225 points generated randomly. The results lead to conclusions similar to those developed in Sects. 3.1 and 3.2. They are also confirmed by other experiments conducted on random lattices and reported in Peeters and Thomas 1997.

3.4 Do other values of k yield significantly different results?

The same experiments were conducted for other values of k (number of facilities). They lead to the same conclusions, but less clear-cut to present because of the presence of numerous alternate optima. Let us take the example of $k = 9$. For a regular lattice of 225 points ($\sigma = 0.00$), very slight differences are observed if none of the 9 sites is imposed. When forcing 2 locations ($j = 38$ and $j = 113$) to avoid alternate optima, the optimum location patterns are totally identical, whatever p and θ. With other values of σ (0.25, 0.50 or 1.00), differences appear only for values of $p = 1.0$, all other solutions being equal, whatever p and θ.

4 Conclusions

The objective of this paper is to simply test how robust the solutions of a location-allocation model are to changes in the parameters of a distance predicting function and consequently with what accuracy these parameters should be estimated in real-world applications. The effects of systematic changes in p

and/or θ in the l_p norm are systematically analyzed on the k-median results, using regular as well as irregular lattices of points. All simulations lead to the following conclusion: changes in p and θ have *little* effect on the location-allocation results.

Practical implications arising from our simulations are twofold. *First*, small errors on the empirical estimation of distances by the l_p distance predicting function very slightly affect the k-median results. This is particularly true for $p \in [1.5, 2.3]$. This result is of prime importance when considering the amount of work involved in estimating the real-world distances by l_p-distances: it saves data collecting and computing time whenever such applications are considered. *Second*, preceding papers have already shown the significant influence of the shape of the network on the k-median results (Peeters and Thomas 1995), and of the influence of extreme points to locate centers (Papini 1994). Given these two preceding papers and the results suggested in this paper, one can suggest that the co-ordinates of the points, and hence the shape of the transportation network), have more importance on the k-median results than the *way* the distance is measured.

The results presented here should be moderated by some critical appraisals. Theoretical regular layouts such as the squared lattice used in this paper enables to well isolate the tested effect: it isolates the problem and holds constant many sources of variation. It is theoretically very satisfactory, but practically limits the scope of the experiment. Further research will soon be conducted in order to see how much the tested hypothesis is sensitive to the shape of the network or weight variations.

Acknowledgements. The authors would like to thank Pierre Hanjoul, Hubert Beguin and the referees for many helpful comments on earlier drafts of the paper.

References

Anselin L (1986) Some further notes on spatial models and regional science. *Journal of Regional Science* 26:799–802

Arnold P, Peeters D, Thomas I (1996) Networks configuration and p-median results. *Sistemi Urbani* (forthcoming)

Beguin H, Thomas I, Vandenbussche D (1992) Weight variations within a set of demand points, and location-allocation issues: a case study of public libraries. *Environment and Planning A* 24:1769–1779

Berens W, Körling F (1985) Estimating road distances by mathematical functions. *European Journal of Operational Research* 21:54–56

Bowerman R, Calamai P, Hall G (1999) The demand partitioning method for reducing aggregation errors in p-median problems. *Computers and Operations Research* 26:1097–1111

Brimberg J, Dowling P, Love R (1996) Estimating the parameters of the weighted l_p norm by linear regression. *IIE Transactions* 28:363–367

Brimberg J, Love R (1992) A new distance function for modelling travel distances in a transportation network. *Transportation Science* 26:127–137

Brimberg J, Love R (1993a) Directional bias of the l_p norm. *European Journal of Operational Research* 67:287–294

Brimberg J, Love R (1993b) General considerations on the use of the weighted l_p norm as an empirical distance measure. *Transportation Science* 27:341–349

Brimberg J, Love R, Walker J (1995) The effect of axis rotation on distance estimation. *European Journal of Operational Research* 80:357–364

Current J, Shilling D (1987) Elimination of source A and B errors in p-median location problems. *Geographical Analysis* 19:95–110

Daskin M, Haghani A, Khanal M, Malandraki C (1989) Aggregation effects in maximum covering models. *Annals of Operations Research* 18:115–140

Erkut E, Bozkaya B (1999) Analysis of aggregation errors for the *p*-median problem. *Computers and Operations Research* 26:1075–1096

Fotheringham A, Densham P, Curtis A (1995) The zone definition problem in location-allocation modelling. *Geographical Analysis* 27:1 60–77

Francis R, Lowe T, Rayco M (1996) Row-column aggregation for rectilinear distance *p*-median problems. *Transportation Science* 30:160–174

Francis R, Lowe T, Rushton G, Rayco M (1999) A synthesis of aggregation methods for multi-facility location problems: Strategies for containing error. *Geographical Analysis* 31:67–87

Haining R (1986) Spatial Models and Regional Science: a Comment on Anselin's Paper and Research Directions. *Journal of Regional Science* 26:793–798

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

Hanjoul P, Peeters D (1985) A comparison of two dual-based procedures for solving the *p*-median problem. *European Journal of Operations Research* 20:387–396

Hansen P, Labbé M, Peeters D, Thisse J (1987) Facility location analysis. *Fundamentals of Pure and Applied Economics* 22:1–70

Hodgson J, Oppong J (1989) Some efficiency and equity effects of boundaries in location-allocation models. *Geographical Analysis* 21:167–178

Hodgson J, Shmulevitz F, Korkel M (1997) Aggregation error effects on the discrete-space *p*-median model: The case of Edmonton, Canada. *Canadian Geographer* 41:415–428

Huriot J, Perreur J (1973) Modèles de localisation et distance rectilinéaire. *Revue d'Economie Politique*, 83:640–662

Huriot J, Perreur J (1990) Distances, espaces et représentation. Une revue. *Revue d'Economie Régionale et Urbaine* 2:197–237

Juel H, Love R (1985) The facility location problem for hyper-rectilinear distances. *AIIE Transactions* 17:94–98

Labbé M, Peeters D, Thisse J (1995) Location on networks. In: Ball M, Magnanti T, Monma C, Nemhauser G, (eds): *Handbook on operations research and management science: Networks.* North Holland, Amsterdam, pp. 551–624

Love R (1993) Directional bias of the lp-Norm. *European Journal of Operational Research* 67:287–294

Love R, Morris J (1972) Modelling inter-city road distances by mathematical functions. *Operational Research Quarterly* 23:61–71

Love R, Morris J (1979) Mathematical models of road travel distances. *Management Science* 25:130–139

Love R, Walker J, Tiku M (1995) Confidence intervals for $l_{k,p,\theta}$ distances. *Transportation Science* 29:93–100

Muller J (1982) Non-euclidean geographic spaces: Mapping functional distances. *Geographical Analysis* 14:189–203

Owen S, Daskin M (1998) Strategic facility location: a review. *European Journal of Operational Research* 111:423–447

Papini P (1994) How can we locate centers? *Studies in Locational Analysis* 7:143–152

Peeters D, Thomas I (1995) The effect of the spatial structure on the *p*-median results. *Transportation Science* 29:366–373

Peeters D, Thomas I (1997) Distance-l_p et localisations optimales. Simulations sur un semis aléatoire de points. *Cahiers Scientifiques du Transport* 31:55–70

Plastria F (1995) What is the best aggregation point? A case for the gravity centre. *Report BEIF/88, Brussels, V.U.B., Centrum voor Bedrijfsinformatica*, p 14

Ruhigira J (1994) Effets de l'agrégation de la demande sur les configurations optimales de l'offre générées par le modèle de localisation-affectation de le *p*-médiane. Thèse de doctorat en sciences, U.C.L., Louvain-la-Neuve, 27-01-94

Thomas I (1993) Sensibilité du découpage spatial optimal des services d'urgence à la définition de la demande. Quelques réflexions sur un exemple mésogéographique. *L'Espace Géographique* 4:318–332