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 $\theta$ 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 $l_p$ norm? The answer should tell whether the degree of accuracy for measuring the parameters $p$ and/or $\theta$ 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