PUBLIC FACILITY LOCATION: ISSUES AND APPROACHES

Giorgio Leonardi, Editor
International Institute for Applied Systems Analysis, Laxenburg, Austria

RR-82-23
June 1982

Reprinted from *Sistemi Urbani*, volume 3 (1981)

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Reprinted with permission from *Sistemi Urbani* 3:293–470, 1981.

PREFACE

The public provision of urban facilities and services often takes the form of a few central supply points serving a large number of spatially dispersed demand points: for example, hospitals, schools, libraries, and emergency services such as fire and police. A fundamental characteristic of such systems is the spatial separation between suppliers and consumers. No market signals exist to identify efficient and inefficient geographical arrangements, thus the location problem is one that arises in both East and West, in planned and in market economies.

This problem is being studied at IIASA by the Public Facility Location Task (formerly the Normative Location Modeling Task), which started in 1979. The expected results of this Task are a comprehensive state-of-the-art survey of current theories and applications, an established network of international contacts among scholars and institutions in different countries, a framework for comparison, unification, and generalization of existing approaches, as well as the formulation of new problems and approaches in the field of optimal location theory.

The papers collected in this issue were presented at the Task Force Meeting on Public Facility Location, held at IIASA in June 1980. The meeting was an important occasion for scientists with different backgrounds and nationalities to compare and discuss differences and similarities among their approaches to location problems. Unification and reconciliation of existing theories and methods was one of the leading themes of the meeting, and the papers collected here are part of the raw material to be used as a starting point towards this aim. The papers themselves provide a wide spectrum of approaches to both technical and substantive problems, for example, the way space is treated (continuously in Beckmann, in Mayhew, and in Thisse *et al.*, discretely in all the others), the way customers are assigned to facilities (by behavioral models in Ermoliev and Leonardi, in Sheppard, and in Wilson, by normative rules in many others), the way the objective function is defined (ranging from total cost, to total profit, total expected utility for customers, accessibility, minimax distance, maximum covering, to a multi-objective treatment of all of them as in ReVelle *et al.*). There is indeed room for discussion, in order to find both similarities and weaknesses in different approaches.

A general weakness of the current state of the art of location modeling may also be recognized: its general lack of realism relative to the political and institutional issues implied by locational decisions. This criticism, developed by Lea, might be used both as a concluding remark and as a proposal for new challenging research themes to scholars working in the field of location theory.

The papers published in this issue constitute only a part of those presented at the Task Force Meeting. A second set of papers will be published in a forthcoming issue of *Sistemi Urbani*.

ANDREI ROGERS
Chairman
Human Settlements and Services Area

CONTENTS

Sistemi Urbani, 3, 1981, 293-297 Giuda Editori, Napoli, Italia

Public facility location

Introduction

Although facility location problems are common to many fields, these problems are analyzed in such diverse ways that it is often hard to believe they share any common features.

According to urban geographers, regional scientists, and many other social scientists, the geographic distribution of human activities and settlements results from the interplay of complex social, economic, and physical factors. These social scientists have developed the discipline of location analysis to obtain a deeper understanding of such interactions. They usually explain these interactions in terms of the trade-offs that people are forced to make in regard to the spatial separation of needed goods, services, and commodities.

In contrast to this perspective, a vast literature on optimal location models has been produced in the fields of Operations Research (OR) and Management Science (MS). These models often appear under such labels as «plant location problems», «warehouse location problems», and, in a more abstract way, «location-allocation problems»; these names reflect the origins of the models, which have been developed mainly as management tools for private firms. The OR and MS view of the problems is somewhat narrower than the social scientist's perspective. Most of the effort is placed on developing algorithms to solve the resulting mathematical programs (which are usually very complicated).

In order to synthesize these polar perspectives (as well as those that lie between), IIASA held a Task Force Meeting on location problems in June 1980. A selected group of scholars from both East and West discussed the differences and similarities of their own perspectives, in order to identify areas of unsolved problems and to propose new themes for future theoretical and applied research. A short account of the main conclusions of the Task Force Meeting is given below.

The problem areas

Some well-defined problem areas were identified at the meeting, for which the current state of the art seems to provide unsatisfactory answers.

One set of problems is related to the decision-making processes implied by location questions. It has been recognized that at least two types of actors are involved in the process of deciding on a location:

the *customer* and the *decision maker.* Most current models either ignore this distinction or account for it in an oversimplified way.

Another set of problems is related to the *costs* that a locational decision usually implies and to the *constraints* to which it is subjected. Although many effective techniques are available to handle different types of costs and constraints, some unsolved problems still remain. These problems are of a socio-economic rather than a technical nature, since they relate to who provides the funds and to the way the existing structures are accounted for.

The behavior of customers and decision makers

The participants at the Task Force Meeting agreed that there is a definite need for a better understanding (and better models) of the mechanism through which *demand* for services arises and by which customers make choices among different alternatives in *space.* Two contrasting examples may be used to clarify the problem.

In a classic warehouse location problem, a firm must locate a set of warehouses for a homogeneous good, which in turn will be shipped to some demand points. The firm will obviously seek to minimize the total shipping costs plus the costs of establishing the warehouse. It is well known that this cost-minimizing criterion implies that each demand point will be served by the *nearest* warehouse only. It is important to note that no model for customer behavior is required, since the quantity demanded is assumed to be given and the good is *delivered* from the warehouses to the demand points.

In the case of a shopping center location problem, a firm must locate a set of shopping centers where a good (or a variety of goods) can be sold to attracted customers. It is clear that in this case the customers, and not the firm, will decide where to go shopping, and everybody tells us that they will not always go to the nearest shopping center. A behavioral model that assumes that customers will choose only the nearest facility is a poor model for real behavior; shopping behavior is determined by many rational and nonrational factors: differences in taste, imperfect information, trade-offs between distance traveled and quality (or price) of goods, competition with other shopping centers, and so on.

These two examples have been taken from the private sector, but they can be easily generalized to public facility problems. There are many similarities among customer-choice processes relating to shopping centers, high schools, hospitals, libraries, theaters, or even places of work and residence. These similarities suggest the need for a new interdisciplinary modeling effort.

Closely related to the question of customer behavior is the definition of the role of the decision maker. The two problems are intimately tied together even in the simplest cases, as can be seen from the two examples given above. In the warehouse location problem, the same decision maker (the firm) decides both the location of facilities (the warehouses) and the trip pattern (the delivery of goods from the warehouses to the demand points). In the shopping center location problem, the firm decides the location of facilities (the shopping centers) but not the trip pattern, which instead results from customer choices.

Such examples can also be found for the public sector. For instance, in a primary school location problem the same decision maker (a public authority) usually decides both the location of facilities (the schools) and the trip pattern (the assignment of children to schools). This is not true for a post office location problem, where the public authority decides the location of facilities (the post offices), but cannot force the customer to always use a specific facility.

The general issues raised by these examples are the amount of control a decision maker can exert and the relationships between the goals guiding his decisions and the goals guiding those of the customer.

It usually makes a big difference whether the decision maker is maximizing his profits, as in the shopping center example, or maximizing customer welfare, an obligation of every public authority. It also makes a difference whether the location questions are posed in a market economy or in a planned one, since many private problems in the former become public problems in the latter and vice versa.

Costs and constraints

Some questions related to costs and constraints in location problems are well known and lead to discussions of a very technical nature; these will be mentioned but not pursued here.

These questions touch on the introduction of economies of scale in the cost of establishing the facilities and the indivisibility requirements placed on the units to be located. In the mathematical literature, problems of this sort are known as nonconvex and combinatorial optimization problems. The difficulties associated with solving them still constitute a challenge for applied mathematicians.

Two other problems related to costs and constraints deserve more detailed discussion here. One is related to costs - not so much the way cost functions are modeled as where the money to pay the costs comes from. Most location problems are formulated as if there were no direct relationship between the customer using a given kind of facility and the money available to establish and operate the facility. It has been shown, however, by means of some simple examples that charging prices to customers and adding the resulting revenue to the available budget usually improves the overall performance of the system, not only in private, profit-making cases but also in the case of a public authority concerned with customer welfare. If this is the case, why should we think of location problems only as pure «physical planning» problems (i.e., location and size being the only decision variables), rather than allowing pricing policies to be introduced as well? And why shouldn't we also introduce taxation policies? The new type of location problem would then have a list of decision variables made up of the traditional physical ones (size and location of facilities), plus some suitable pricing and taxation rules.

When a stock of facilities already exists, however, the location of new facilities may not be required; instead, pricing and taxation policies may become the main tool for providing equitable access to all customers. Education, health care, and housing are typical examples for urban services where taxation policies, welfare schemes, and public allowances are much more effective than geographical distribution.

The issue of constraints does not so much concern the topology of the set of feasible location patterns, but rather the proper definition of constraints arising from the existing environment in which a location problem usually has to be solved. Indeed, most location problems are formulated for very improbable human settlements where there is demand, but no available facilities. This formulation, artificial as it is, does not constitute a serious limitation in many *developing* countries, where the stock of existing facilities is limited. However, this is not the case in most developed countries. Every kind of facility already exists in most urban areas, so the literal implementation of an «optimal» location pattern, as would follow from the above formulation, would result in a crazy pattern of demolition and reconstruction. Something is therefore missing in the standard formulation: expanding or demolishing the existing stock of facilities is not accounted for in the usual list of decision variables, nor is the implied cost of such actions. Decisions to expand or demolish lead to a dynamic formulation of the location problem, since they cannot be considered on a daily basis without taking into account the future performance of the system. As with pricing and taxation policies, capacity expansion or reduction may be needed even when new locations are not required. When many facilities already exist, decisions to locate new ones may be unreasonable, but the fluctuation of demand over time and space may require adjustments in the size of the existing facilities.

Some conclusions

Pricing, taxation, expansion, and reduction considerations pose a new challenge for location research. They suggest that optimizing location is an unnecessarily restrictive approach to urban management and not necessarily the best one. The goal of improving access to urban services can be reached by using many other tools, and the resulting decision problems require the development of new models and techniques.

Models of customer choice also deserve attention in future research activities. Although the literature on location models deals with this problem unsatisfactorily, much progress has been made in related field, such as transport models and housing-market models. An interdisciplinary effort would therefore greatly improve the state of the theory and applications of customer-choice models.

A third theme for future research underpins the whole discussion, although it has never been stated explicitly. On the one hand, when the locations of some facilities are changed, new traffic flows of people and goods are generated, thus affecting the transport network. On the other hand, a new geographical distribution of facilities causes a new distribution of land values and residential preferences. As well, changes in the transport network and in the location of households lead to changes in facility locations. A true systems approach is therefore required, taking into account interactions among the main subsystems of the urban system, including housing, transportation, and other services.

G. LEONARDI
Leader of the Public Facility Location Task
Human Settlements and Services Area
IIASA

Sistemi Urbani, 3, 1981, 299-317 Giuda Editori, Napoli, Italia

On the location of an obnoxious facility

P. Hansen

Institut d'Economie Scientifique et de Gestion, Lille, France, and Faculté Universitaire Catholique de Mons, Belgique.

D. Peeters

Unité de Géographie Economique, Université Catholique de Louvain,
1348 Louvain-la-Neuve, Belgique.

J.-F. Thisse

SPUR, Unité de Science et de Programmation Urbaines et Régionales,
Faculté des Sciences Appliquées, Bâtiment Vinci, 1 Place du Levant,
1348 Louvain-la-Neuve, Belgique

Received 15 May 1981

Abstract. The problem of locating an obnoxious facility in a continuous and bounded subset of the plane is considered. Localization theorems and resolution methods are proposed for both the minimization of the total nuisance cost and the maximal nuisance cost, when the cost supported by an inhabitant is only assumed to be decreasing and continuous in distance. The locational pattern of nuclear power plants in France is used as an illustration of the properties obtained.

Key words: continuous location, obnoxious facility, total cost minimizing, maximum cost minimizing, nuclear power plants location.

1. Introduction

In June 1980, the Belgian government protested to the French government concerning the establishment of several nuclear power plants along the border between the two countries. A glance at fig. 1 gives the impression that the decision of Electricité de France corresponds to a deliberate choice. The strategy would consist in putting the nuclear power plants on the outskirts of France. (This is especially well illustrated by the locations chosen along the Belgian and German borders and along the Atlantic coast). However, we also observe some interior sitings. Again, fig. 1 suggests an alternative rule. Roughly, the interior plants appear to be set up in regions with low-density population and far from the main towns. (This is very clear for some plants located to the south of Paris). The first purpose of this paper is to provide some simple rationales of those observations by using tools of location theory.

The second purpose is more general. It is adressed to the problem of *locating an obnoxious facility in a continuous space.* An obnoxious facility is a facility necessary to the whole population but generating strong negative externalities on the surrounding population. Apart from the above-mentioned nuclear stations, further examples of interest are given by incinerators, garbage dumps or sewage plants. As spatial externalities decrease with the distance from the source, the planner attempts to place such a facility as far away as possible from population centers, rather than close to those centers as in the classical Weber or Rawls models. The problem of siting an obnoxious facility has been tackled in a non-formal way by Wolpert *et al* (see, e.g. Austin, 1974; and Austin, Smith, Wolpert, 1970). Church and Garfinkel (1978) have placed it within the framework of location theory. They propose to locate a facility of this kind with the aim of maximizing the total weighted distance to the inhabitants. An alternative objective in which the minimal weighted distance between the facility and an inhabitant is to be maximized, has then been considered by Drezner and Wesolowsky (1980) and by Dasarathy and White (1980). In this paper we deal with generalized versions of those models: the facility is established in order to minimize either the *total nuisance cost* or *the*

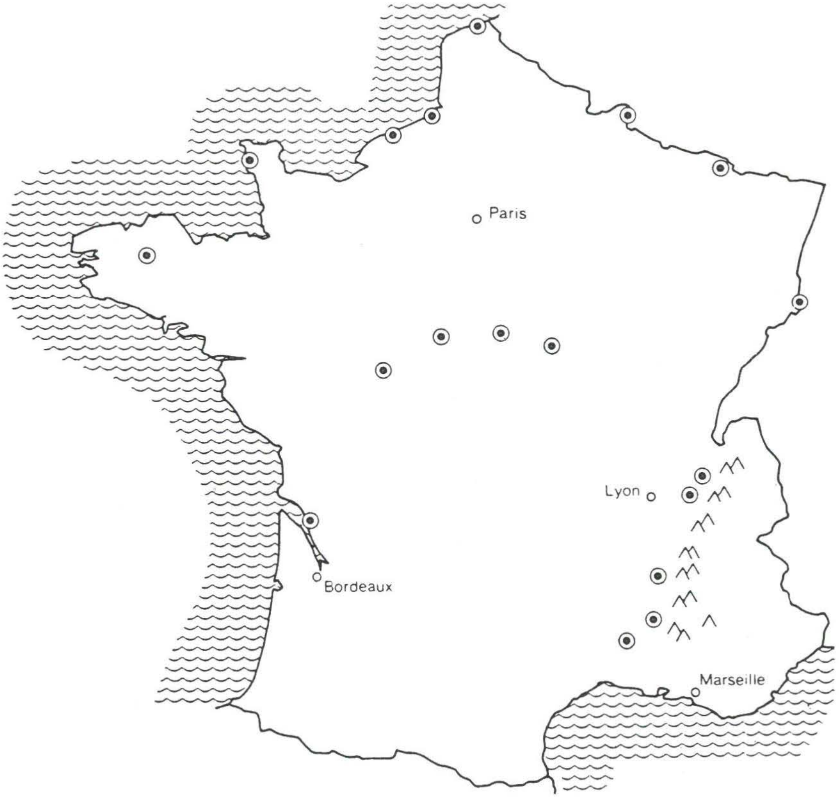


Figure 1

maximal nuisance cost. Here the cost supported by an individual is *solely* assumed to be decreasing and continuous in distance. This is, it seems, the most significant formulation in many practical issues. Indeed, it is well-known that not only the impact, but also the rate of change of the externality decreases when the distance increases (see Papageorgiou, 1978). For instance, the perceived advantage for Belgium from moving nuclear plants ten miles further away from the country is larger when those plants are sited close to the border than when they are far from it. Thus, functions linear in distance often constitute too rough an approximation of the real nuisance costs. Furthermore, in the models under scope, the location can take place anywhere in some continuous and bounded areas of the plane and the distance between the facility and a locality is derived from a norm defined on $\mathbb{R}^2$. Such a formulation appears to be especially relevant for the case where the pollution diffuses throughout the space, rather than along particular lines.

The paper is organized as follows. In Section 2, the models are introduced and localization theorems are derived. These results are used to shed some light on implicit objectives of Electricité de France's locational policy. Resolution methods are then proposed in Section 3. Some remarks complete the paper in Section 4.

2. Models and properties

The *anti-Weber problem* (in short AWP) is defined as follows:

(i) There is one facility to be located and any point $s \in S \subset \mathbb{R}^2$ is a feasible location; S is assumed to be closed and bounded. (Note that the cardinality of S may be finite or infinite).

(ii) There are m localities in the area concerned with the facility; the i-th locality is given by a point d_i of $\mathbb{R}^2$, $i = 1 \ldots m$.

(iii) The distance between the facility located at s and locality d_i is expressed by $\| s - d_i \|$, where $\| . \|$ is a norm defined on $\mathbb{R}^2$. The choice of a particular norm depends on the nature of the nuisance; examples of norms used in location theory are the 1_p-norm, with $p \in [1,2]$ (see Morris, Verdini, 1979) and the weighted one-infinity norm (see Ward, Wendell, 1980).

(iv) The nuisance cost supported by the inhabitants of the i-th locality is given by a decreasing and continuous function of the distance from the facility; it is denoted by $D_i(\| s - d_i \|)$.

(v) The facility must be set up is a point of S where the total nuisance cost defined by

$$E(s) = \sum_{i=1}^{m} D_i (\| s - d_i \|) \tag{1}$$

is minimized.

The *anti-Rawls problem* (in short ARP) is similarly defined by assumption (i)-(iv) and by

(v′) The facility must be located in a point of S where the maximal nuisance cost

$$H(s) = \max_{i=1\ldots m} D_i(\|s - d_i\|) \tag{2}$$

is minimized.

The following remarks are in order. First, we notice that the *only* difference with the Weber and Rawls problems is that the nuisance costs are decreasing in distance while the access costs are increasing (see Hansen, Thisse, 1981; Hansen, Peeters, Richard, Thisse, 1981). Second, without the boundedness assumption on S, both the AWP and the ARP would admit a trivial solution, namely the limit of any sequence of points (s_n) which verifies the condition $\|s_n\| \to \infty$. Clearly, this solution is not practical since it amounts to dumping the refuses abroad. Third, criteria (1) and (2) refer to two different social choice rules: the utilitarian objective and the leximin criterion proposed by Rawls (1971). The first one can be viewed as a measure of the loss of welfare incurred by the overall community and the second one as an equity measure relative to the worse-off locality. Fourth, and last, in the particular case when the nuisance costs are linear in distance, i.e. $D_i(\|s - d_i\|) = a_i - b_i \cdot \|s - d_i\|$ with a_i and b_i positive, the objectives (1) and (2) boil down to the *maxisum criterion* given by

$$\max \sum_{i=1}^{m} b_i \|s - d_i\|, \tag{3}$$

and to the *maximin criterion*

$$\max \min_{i=1\ldots m} b_i \|s - d_i\| \tag{4}$$

when the constants a_i are equal.

The concept of remoteness is used to characterize the solution to the AWP and to the ARP: Given a set X, we say that $s \in S$ is *remote from* X iff $x \in X$ exists such that the straight half-line starting from x and passing through s contains no point of S beyond s. This concept is illustrated in fig. 2 where s_1 and s_2 are remote from X, but not s_3 and s_4. Note that any point of S remote from X is a boundary point of S, but the converse is not true as shown by s_4 in fig. 2.

Denote by C the convex hull of $\{d_1 \ldots d_m\}$. We have:

THEOREM 1. *The set constituted by the points of* $S \cap C$ *and by the points of* $S - C$ *remote from* C *contains at least one solution to the AWP.*

Proof: Let s^* be a solution to the AWP (such a solution always exists by the Weierstrass theorem). If $s^* \in S \cap C$, the theorem is proved. Then, assume $s^* \in S - C$. In this case, there exists a point $\bar{s} \in C$ such that $\|\bar{s} - d_i\| \leq \|s^* - d_i\|$, for $i = 1 \ldots m$ (see Wendell, Hurter, 1973). Let s_1 be given by $\lambda_1 s^* + (1 - \lambda_1)\bar{s}$ with $\lambda_1 = \sup\limits_{\lambda \geq 1} \{\lambda ; \lambda s^* + (1 - \lambda)\bar{s} \in S\}$. Clearly, $\|s_1\| < \infty$ since S is bounded. Furthermore, as s_1 is the limit of a sequence of points of S and S is closed, we have $s_1 \in S$. Furthermore s_1 is remote from C by construction.

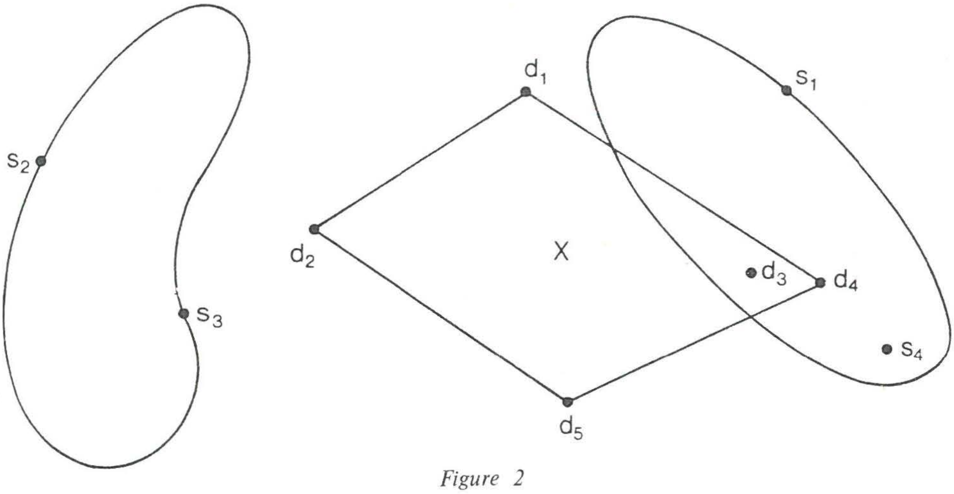


Figure 2

Two cases may arise. In the first one, $\lambda_1 = 1$. Hence, $s^* = s_1$ and the theorem is proved. In the second one, $\lambda_1 > 1$. For any i, we have

$$\begin{aligned} \|s^* - d_i\| &= \| \frac{1}{\lambda_1} s_1 + \frac{\lambda_1 - 1}{\lambda_1} \bar{s} - d_i \| \\ &\leqslant \frac{1}{\lambda_1} \|s_1 - d_i\| + \frac{\lambda_1 - 1}{\lambda_1} \|\bar{s} - d_i\| \end{aligned}$$

since the norm is a convex function. Given that $\|\bar{s} - d_i\| \leqslant \|s^* - d_i\|$, we obtain $\|s^* - d_i\| \leqslant \|s_1 - d_i\|$, for $i = 1 \ldots m$, since $\lambda_1 > 0$. As functions D_i are decreasing, we deduce that $E(s_1) \leqslant E(s^*)$. Consequently, there exists $s_1 \in S$ remote from C which is a solution to the AWP. QED.

Spatially, this theorem means that a solution to the AWP is either a point of the locational polygon C or a point «far» from it. An illustration is contained in fig. 3 where the set of candidate points is constituted by the shaded area and by the heavy lines.

As S is not necessarily convex, we denote by [S] the convex hull of S. The following result then characterizes the solution to (3).

THEOREM 2. *Assume that the nuisance costs are linear in distance. Then, the set of extreme points of* [S] *remote from* C *contains at least one solution to the AWP.*

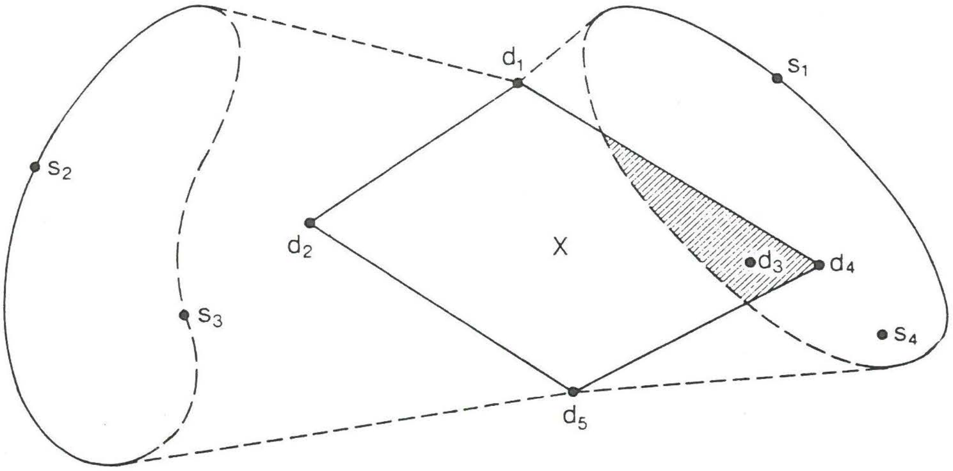


Figure 3

Proof: Function $\sum_{i=1}^{m} b_i \| s - d_i \|$ is convex as the positive weighted sum of norms. Hence, by the theorem of maximization of a convex function (see Roberts, Varberg, 1973, p. 232), it is known that the set of extreme points of [S] contains a maximizer s*, say, of $\sum_{i=1}^{m} b_i \| s - d_i \|$. As all the extreme points of [S] belong to S, s* is a feasible location and, consequently, an optimal solution to (3). If s* is remote from C, the theorem is proved. If not, $\bar{s} \in C$ dominating s* and $s_1 \in S$ exist such that $s_1 = \lambda_1 s^* + (1 - \lambda_1)\bar{s}$ with $\lambda_1 = \sup_{\lambda \geqslant 1} \{ \lambda ; \lambda s^* + (1 - \lambda)\bar{s} \in S \}$; s_1 is remote from C. By the argument of the proof of Theorem 1, $\sum_{i=1}^{m} b_i \| s - d_i \|$ is constant on $[s^*, s_1]$. Accordingly, if s_1 is an extreme point of [S], the theorem is proved. If not, as the objective function is convex, the following two cases may arise: (i) $[s^*, s_1]$ belongs to a contour line *L* of $\sum_{i=1}^{m} b_i \| s - d_i \|$; (ii) $\sum_{i=1}^{m} b_i \| s - d_i \|$ reaches its absolute minimum when

$s \in [s^*, s_1]$. In the first case, as s_1 is not an extreme point of [S], $s' \in [S]$ and $s'' \in [S]$ may be found such that $s_1 \in]s', s''[$. Given that L defines a convex set, point s′, say, is situated outside L so that

$\sum_{i=1}^{m} b_i \| s' - d_i \| > \sum_{i=1}^{m} b_i \| s^* - d_i \|$, a contradiction. In the second case, the objective function must be constant and equal to its minimum on S. This is possible only if S is included in the set of points of $\mathbb{R}^2$ where $\sum_{i=1}^{m} b_i \| s - d_i \|$ is minimum. But then, any extreme point of [S] belonging to C satisfies the desired properties. QED.

In words, Theorem 2 says that all the interior points of S may be disregarded when looking for a solution to the maxisum problem; only some part of the boundary of S are to be considered. The linear model therefore leads to a substantial reduction in the set of candidates, when compared with the general model (1).

A locality $d_j \in S$ may be an optimal solution to the AWP. Yet, in the case when the externality strongly decreases in the neighborhood of the facility, it is expected that d_j is never a minimizer of the total nuisance cost. Indeed, compared with d_j, we observe that locating the facility in the vicinity of d_j leads to a relatively large decrease of D_j and to relatively small variations in the other costs. Hence, provided the distance from d_j is small enough, the gain in D_j should exceed the variation in $\sum_{\substack{i=1 \\ i \neq j}}^{m} D_i$.

This is shown in the next theorem. (Note that it is true for any 1_p-norm but the proof is then more tricky).

THEOREM 3. *Assume that functions* D_i *are continuously differentiable on* $]0, \infty[$ *and that* $\|.\|$ *is the Euclidean norm. If the marginal nuisance cost associated with* d_j *is* $-\infty$ *at zero and if* d_j *is not an isolated point of* S, *then* d_j *is not a local minimizer of the total nuisance cost.*

Proof: As $\|.\|$ is not differentiable at zero, E(s) is not differentiable at d_j and we cannot use the traditional optimization techniques. Rather, we will prove that, provided Θ is small enough, $\frac{d}{d\Theta} E(d_j + \Theta \bar{s})$ is negative for any $\bar{s}$ such that $\|\bar{s}\| = 1$.

It is clear that there exists $\overline{\Theta} > 0$ such that $d_i \notin [d_j, d_j + \Theta \bar{s}]$ for $\Theta \in [0, \overline{\Theta}]$ and for any $i \neq j$. Hence, for $\Theta \in]0, \overline{\Theta}]$, we have

$$\frac{d}{d\Theta} E\ (d_j + \Theta \bar{s}) = \frac{d}{d\Theta} \left\{ \sum_{\substack{i=1 \\ i \neq j}}^{m} D_i (\| d_j + \Theta \bar{s} - d_i \|) + D_j (\Theta) \right\}$$

$$= \sum_{\substack{i=1 \\ i \neq j}}^{m} D_i' \cdot \frac{[(d_j^1 + \Theta \bar{s}^1 - d_i^1) \cdot \bar{s}^1 + (d_j^2 + \Theta \bar{s}^2 - d_i^2) \cdot \bar{s}^2]}{\| d_j + \Theta \bar{s} - d_i \|}$$

$$+ \frac{d\, D_j(\Theta)}{d\Theta} ,$$

where D_i' denotes the derivative of D_i w.r.t. the distance.

Given that the first term of the RHS is continuous on $[0, \overline{\Theta}]$, a constant K_j exists such that $\frac{d}{d\Theta} E(d_j + \Theta \bar{s}) < K_j + \frac{d}{d\Theta} D_j(\Theta)$. As $\lim_{\Theta \to 0} \frac{d}{d\Theta} D_j(\Theta) = -\infty$, we can find $\hat{\Theta} \in]0, \overline{\Theta}]$ such that $K_j < \left| \frac{d\, D_j(\Theta)}{d\Theta} \right|$ for any $\hat{\Theta} \in]0, \hat{\Theta}]$, which means that $\frac{d}{d\Theta} E\ (d_j + \Theta \bar{s}) < 0$ whatever $\Theta \in]0, \hat{\Theta}]$. QED.

Let us come to the ARP. It is easy to verify that the argument developed in the proof of Theorem 1 remains valid for this problem. Accordingly, the ARP admits the same localization theorem than the AWP, at least for the general models. On the other hand, Theorem 2 ceases to be true for the maximin problem. To see it, consider the following counter-example. Given a linear segment whose end points correspond to localities with the same weight b_i, the maximin solution is obviously situated at the middle of the segment, and not at one of its extreme points. This suggests that interior locations are more probable in the ARP than in the AWP. Finally, it can be checked that Theorem 3 is still true for the ARP.

We now return to the problem of the locational pattern of nuclear power plants in France. Theorems 1 and 3 - taken in the context of the AWP or of the ARP - are used for providing a possible explanation. To begin with, we recall that Theorem 1, which deals with the very likely case of non-linear pollution functions, says that both interior locations within the locational polygon and border locations may arise when a «push away» policy is followed. In fig. 1, it is seen that many locations correspond to the latter case, i.e. those near Germany, Belgium or the Atlantic coast. Moreover, locations bordering the Alps, a zone probably not adequate for establishing this kind of facility, can be assimilated, it seems, to the previous ones. Theorem 1 takes also into account interior sitings but does not preclude, however, locations close to large towns. As shown by Theorem 3, such locations will be unprobable provided the impact on the population situated in the vicinity of the plants is large compared to that on more remote populations. Several interior locations depicted in fig. 1 agree with this observation.

We are of course aware that other site selection factors are at work; availability of water for some types of nuclear power plants is a major example; also, distribution expenses are an important part of the investment and operating costs of the electricity sector. However, the result of a locational exercise by Dodu and Maréchal (1980) in which only minimization of investment and distribution costs is considered, yields a pattern of locations very different from the actual one. This suggests that the spatial policy of Electricité de France is strongly influenced by the perceived obnoxiousness of the nuclear plants. (Drastic reductions in nuclear programs induced by ecological protestations in several other European countries corroborate the importance of this factor). Hence, given the decision of the French government to maintain its nuclear program, the adopted «push away» policy would appear as the most satisfactory from the ecological viewpoint, at least as far as France is concerned. The resulting increase in distribution costs can then be viewed as the «implicit» price paid by the French government to meet the environmental preoccupations of the population.

3. Methods

The present section is devoted to algorithms for solving the AWP and the ARP. As many geographical areas can be well approximated by polygons, we assume throughout this section that S is defined by the union of a finite number of convex polygons P_j.

We begin with the AWP and present a branch-and-bound method to deal with model (1), similar to the *Big Square-Small Square* algorithm developed by Hansen and Thisse (1981) for solving the generalized Weber problem. The branching rule consists in partitioning a square Q with sides parallel to the axes into four equal subsquares. The bounding rule exploits the partitioning of $\mathbb{R}^2$ obtained by extending the sides of Q; this partition is formed by the square Q, the four side regions and the four corner regions (see fig. 4). With each point d_i we associate a farthest point $\bar{d}_i$ belonging to Q: $\|\bar{d}_i - d_i\| = \max_{s \in Q} \|s - d_i\|$. It is easy to see that:

(i) if $d_i \in Q$, then $\bar{d}_i$ is the vertex of Q farthest from d_i (see d_1 and $\bar{d}_1$ in fig. 4); (ii) if d_i belongs to a side region, then $\bar{d}_i$ is a vertex of the opposite side of Q farthest from d_i (see d_2 and $\bar{d}_2$ in fig. 4); (iii) if

d_i belongs to a corner region, then $\bar{d}_i$ is the vertex of Q diagonally opposed to d_i (see d_3 and $\bar{d}_3$ in fig. 4). As a direct consequence of the decreasing character of the nuisance costs, we have

$\underline{E} = \sum_{i=1}^{m} D_i(\|\bar{d}_i - d_i\|) \leqslant E(s)$ for all $s \in Q$. Furthermore, as $E(s)$ is continuous, one can always find $\varepsilon > 0$ such that the locations within a square whose sides have length ε are approximately equivalent in value.

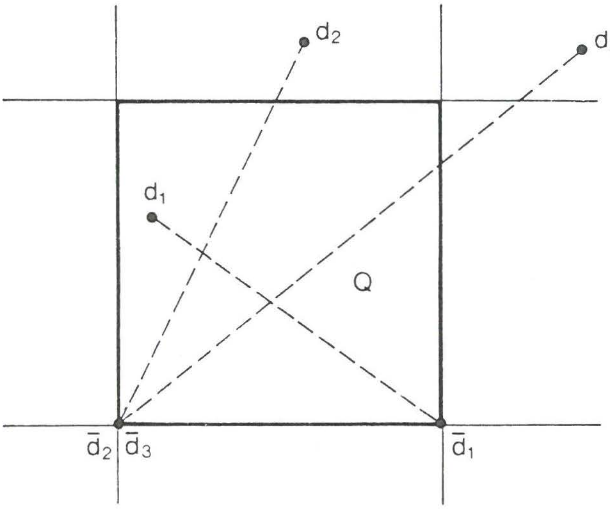


Figure 4

Let us now state the rules of the algorithm.

a. *Initialization.* Let Q_1 be a smallest square containing the set of feasible locations, whose sides are parallel to the axes. Set $I_1 = 1$ if $S = Q_1$ and $I_1 = 0$ otherwise. Let L be the length of a side of Q_1 and ε the tolerance. Compute the value of $E(s)$ for one extreme point of each polygon P_j. Let E_{opt} denote the smallest of the values so obtained and s_{opt} the corresponding feasible location.

b. *New list of squares.* Consider in turn each square Q_h of the current list. Let $Q_k \dots Q_l$ be the four equal subsquares of Q_h.

b.1. If $I_h = 1$, add the four subsquares to the new list and set $I_k = \dots = I_l = 1$.

b.2. If $I_h = 0$, check for each subsquare Q_k if there exists a polygon P_j such that $P_j \cap Q_k = Q_k$ or $P_j \cap Q_k \subset Q_k$. In the former case, add Q_k to the new list and set $I_k = 1$; in the latter one, add Q_k to the new list and set $I_k = 0$. If $P_j \cap Q_k = \varnothing$ for all P_j, delete Q_k.

Replace the current list of squares by the new one.

c. *Improvement of the solution.* Consider in turn each square Q_h of the current list.

c.1. If $I_h = 1$, compute $E(s_h)$ where s_h is the crossing point of the diagonals of Q_h. If $E(s_h) \leqslant E_{opt}$, then set $E_{opt} := E(s_h)$ and $s_{opt} := s_h$.

c.2. If $I_h = 0$, check whether s_h defined as above belongs to S. If yes, proceed as in c.1.

d. *Bounding and deletion of squares.* For each square Q_h in the current list compute $\underline{E}_h$. If $\underline{E}_h \geqslant E_{opt}$, delete Q_h from the current list of squares.

e. *Termination test.* If $L < \varepsilon$, end. Then s_{opt} denotes a near-optimal solution, E_{opt} the corresponding value of the objective and the current list of squares a feasible region containing all the best solutions. Otherwise, let $L := L/2$ and return to step b.

Details on the way to check efficiently the feasibility of points s_h and of squares Q_k are given in Hansen and Thisse (1981), together with a presented of the implementation of an algorithm similar to that one just presented on a computer.

To illustrate, consider the following example. There are three points $d_1 = (12,16)$, $d_2 = (12,0)$ and $d_3 = (0,12)$; the corresponding costs are respectively given by

$$D_1 = 50.\exp\{-0.05\ [(s^1 - d_1^1)^2 + (s^2 - d_1^2)^2]^{1/2}\}\ ,$$

$$D_2 = 55.\exp\{-0.025\ [(s^1 - d_2^1)^2 + (s^2 - d_2^2)^2]^{1/2}\}\ ,$$

$$D_3 = 60.\exp\{-0.05\ [(s^1 - d_3^1)^2 + (s^2 - d_3^2)^2]^{1/2}\}\ .$$

Finally, the polygons are a rectangle P_1 with vertices (0,16), (12,16), (12,10), (0,10) and a triangle P_2 with vertices (8,8), (16,0), (8,0). (See fig. 5).

The initial square has its lower left corner at the origin and the length L of its sides is 16; $s_{opt} = (0,16)$ and $E_{opt} = 109.92$, after the initial step a. The results of the first three iterations are summarized in Table 1 and illustrated in fig. 5. Values of the function E which improve the incumbent E_{opt} are starred and values of the bound $\underline{E}_h$ for which the corresponding square is deleted are underlined. The squares remaining in the current list after three iterations are shaded in fig. 5.

Let us now assume that the costs D_i are linear in distance. Given Theorem 2, the following polynomial procedure can be proposed.

a. *Determination of the convex hull of* S. Determine [S] from the extreme points of all P_j by a standard algorithm for obtaining the convex hull of a finite subset of the plane. Let T denote the set of extreme points of [S].

b. *Finding an optimal solution.* Compute the value of E(s) for each point $s \in T$. Let E_{opt} denote the smallest value so obtained; the corresponding point s_{opt} is an optimal solution.

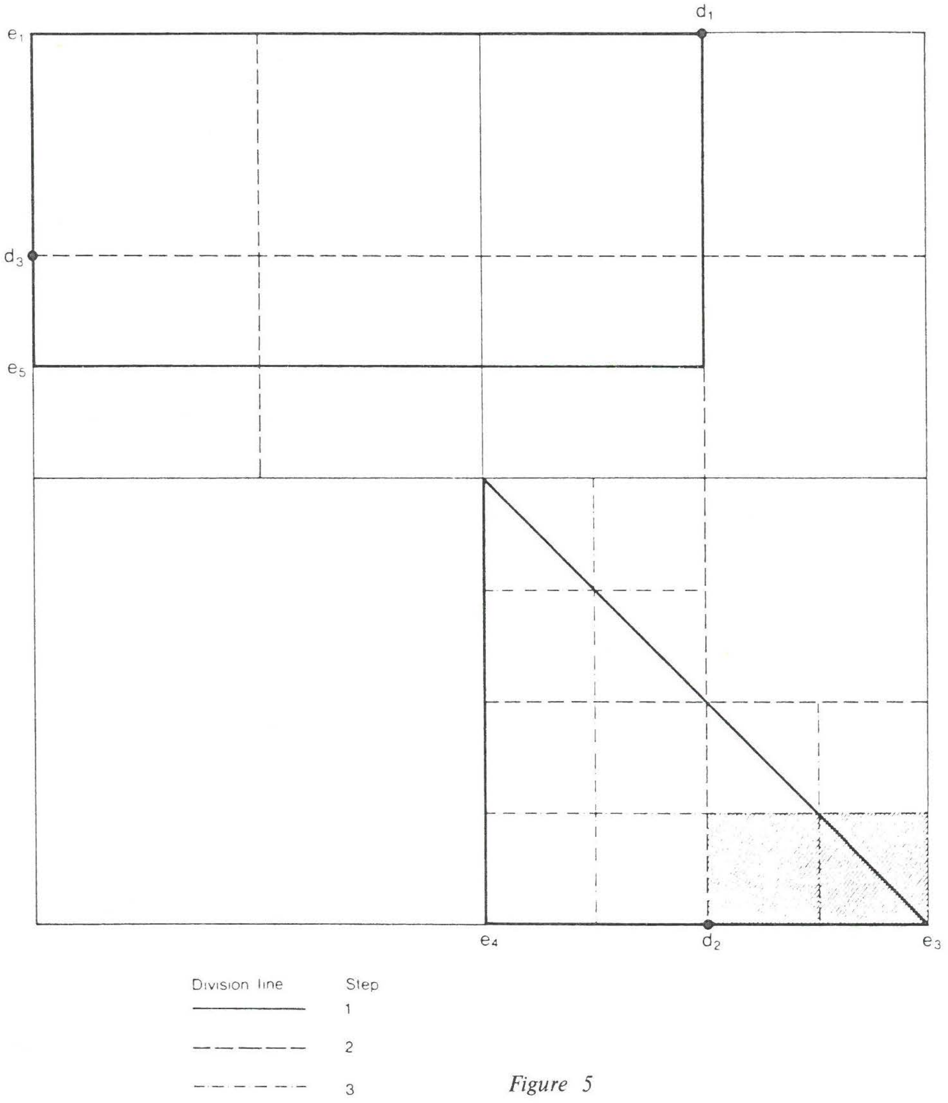


Figure 5

Table 1

Iteration	L	h	s_h	I_h	E	$\underline{E}_h$
0	16	1	(8 , 8)	0	109.92*	78.06
1	8	2	(4 , 12)	0	119.45	96.03
		3	(12 , 12)	0	114.61	94.69
		4	(12 , 4)	0	106.38*	87.97
2	4	5	(2 , 14)	1	117.89	105.14
		6	(6 , 14)	1	117.76	105.50
		7	(6 , 10)	0	117.54	105.11
		8	(2 , 10)	0	118.62	105.51
		9	(10 , 14)	1	118.06	105.98
		10	(10 , 10)	0	115.10	103.94
		11	(10 , 6)	0	110.47	99.72
		12	(14 , 2)	0	101.28*	91.74
		13	(10 , 2)	1	105.48	95.35
3	2	14	(9 , 7)	0	112.44	106.65
		15	(11 , 5)	0	108.46	103.34
		16	(9 , 5)	1	109.74	104.12
		17	(13 , 3)	0	104.09	99.22
		18	(15 , 1)	0	97.76*	93.17
		19	(13 , 1)	1	102.28	97.44
		20	(9 , 3)	1	106.88	101.47
		21	(11 , 3)	1	106.35	101.31
		22	(11 , 1)	1	104.23	99.25
		23	(9 , 1)	1	103.57	98.58

h = index of the square
s_h = center of the square

In step a, the convex hull of n points, where n is the total number of extreme points of all P_j, can be found in order $0(n \lg n)$ with the algorithm of Preparata and Hong (1977); alternatively, Eddy's method (1977) could be used and requires $0(n|T|)$ operations, where $|T|$ denotes the cardinal of T. Step b clearly requires $0(m|T|)$ operations. Hence, the entire procedure's complexity is $0(\max(n \lg n, m|T|))$ or $0((n+m)|T|)$. (Note that the latter cannot exceed $0(\max(n^2, mn))$). The procedure is illustrated by the example given above, but in which

$$D_1 = 50 - 2\,[(s^1 - d_1^1)^2 + (s^2 - d_1^2)^2]^{1/2} \,,$$

$$D_2 = 55 - 2.5\,[(s^1 - d_2^1)^2 + (s^2 - d_2^2)^2]^{1/2} \,, \quad \text{and}$$

$$D_3 = 60 - 3\,[(s^1 - d_3^1)^2 + (s^2 - d_3^2)^2]^{1/2} \,.$$

The comparison of the values of E at the extreme points of [S], i.e. $\{e_1, d_1, e_3, e_4, e_5\}$, in fig. 5 shows that $s_{opt} = e_3 = (16,0)$ and $E_{opt} = 62.02$.

We turn to the ARP and present a very simple method called *Black and White* (*), for solving it. A major advantage of this method is that it can be easily implemented by using a map of the region in which the facility is to be set up and a hand calculator. The only computations to be performed are the evaluations of the cost functions D_i for given locations and the determination of the distances corresponding to given values of the cost functions.

The rules of the algorithm are the following.

a. *Initialization.* Represent on a map the points $d_1 \dots d_m$ and the set S of feasible locations. Shade the part of the map complementary to S. Choose a few feasible points $s \in S$ and compute the corresponding values of H. Let H_{opt} denote the smallest value and s_{opt} the corresponding point.

b. *Elimination of dominated regions.* Compute the radius $R_i = D_i^{-1}(H_{opt})$ for each i. Trace the corresponding iso-cost curves on the map and shade the interior of each of these curves.

c. *Improved solution and test for ending.* Consider all the unshaded regions of the map. If all of them have diameter smaller than a given tolerance, end with s_{opt} being a near-optimal solution and H_{opt} its value.

(*) Indeed, the problem can be interpreted as that of the whisky distillery whose purpose is to locate as far away as possible from the closest temperance league.

Otherwise, select a central point s_h in each unshaded feasible area S_h (or in a few of them if they are numerous). Compute $H(s_h)$ for all points s_h so obtained. Let $H_{opt}: = \min H(s_h)$ and set s_{opt} equal to the corresponding s_h. Then go to step b.

Considering again the data of the example introduced above for Big Square-Small Square, the first three iterations of Black and White are illustrated in fig. 6 and summarized in Table 2.

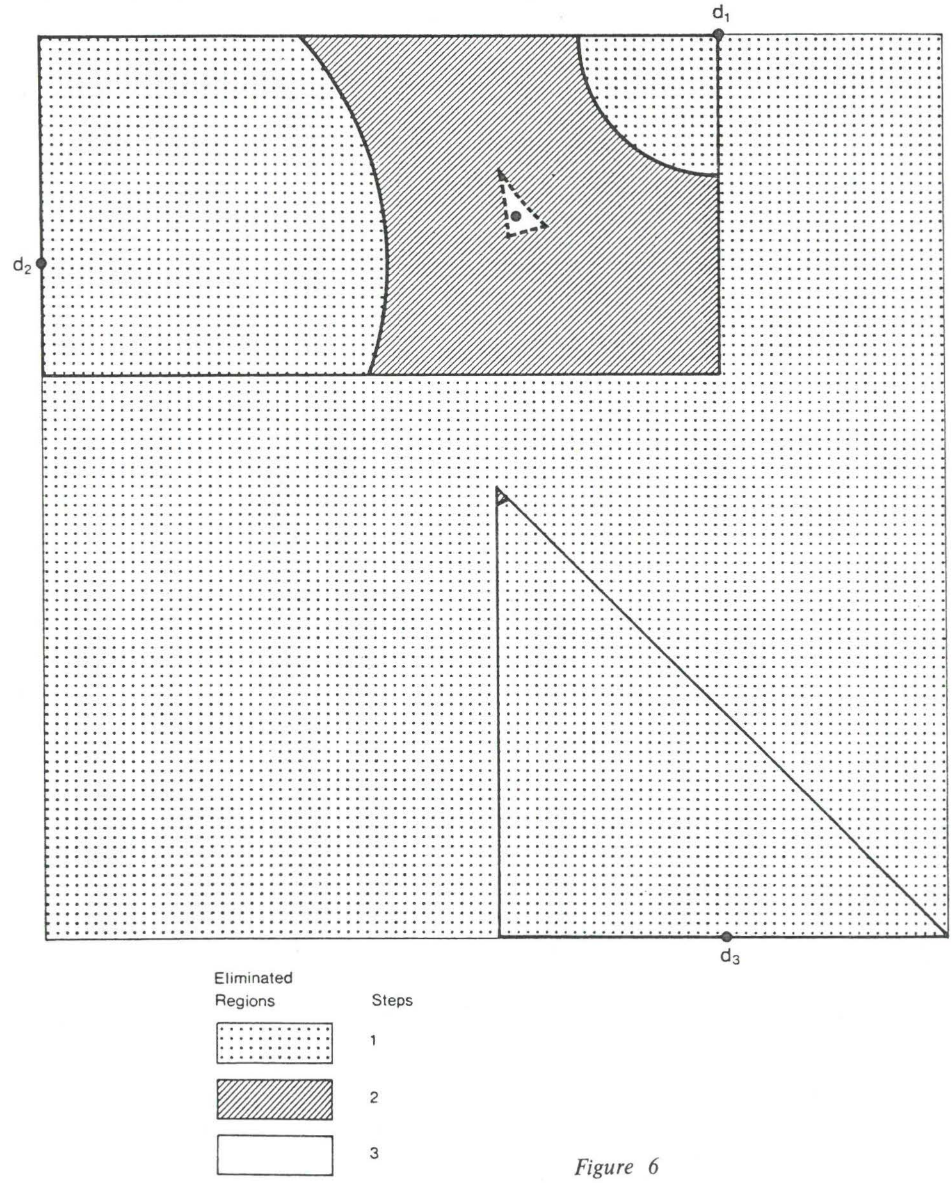


Figure 6

Table 2

Iteration	Point chosen	H	R_1	R_2	R_3
1	(6 , 13)	44.27*			
	(10.7 , 2.7)	51.03			
			2.43	8.68	6.08
2	(8.5 , 13)	39.71*			
	(8 , 8)	43.98			
			4.61	13.03	8.26
3	(8.4 , 12.8)	39.45*			
			4.74	13.30	8.39

The following extensions are possible. The method can take into account non-isotropies in the nuisances. For instance, dominant winds could diffuse pollution further in some directions than in others. Provided that the iso-cost curves associated with the points d_i can be computed or tabulated from observations, the only change needed is to replace the iso-distance curves used in step b by these iso-cost curves. Also, the visual nuisance due to ugly buildings or plants may not be of concern in some places because of the variations in relief. To take this into account, the parts of the area delimitated by an iso-cost curve centered at d_i, which are invisible from d_i should not be shaded.

If desired, the Black and White method can be rendered entirely automatic and used as a black box on a computer. Such an approach has been followed by Drezner and Wesolowsky (1980) who aim to locate a facility at the largest minimum weighted distance from m given points, but not further than a pre-specified distance from any of them. The following two problems then arise in step c: (i) how to determine the feasible regions S_h? (ii) how to find points $s_h \in S_h$ where H is to be evaluated? The first problem can be solved as in Drezner and Wesolowsky. Indeed, given the definition of S, any vertex of a feasible region S_h must be at the intersection of two iso-cost curves or of one iso-cost curve with a side of a polygon P_j or of two sides of such a polygon. As these points are in finite number, they can be all determined. The second problem appears to be more difficult to solve as the regions S_h are usually not convex. One possible solution would consist in selecting a feasible direction from one vertex of S_h and in choosing for s_h the middle of the linear segment defined by the intersection of that direction with S_h.

The above difficulty can be avoided by reversing the procedure, i.e. by choosing values for the objective function and checking if there exist feasible regions for those values, as done by Drezner and Wesolowsky. When following such a tack, the determination of feasible regions associated with a given value of the objective function can of course be interrupted as soon as a feasible location is found; it is needless to determine all the regions S_h. A dichotomous search on H is then adequate and can be initialized on $[\alpha, \beta]$ where α is defined by $\max_{i=1 \dots m} \min_{s \in ET} D_i(\| s - d_i \|)$, ET being the set of extreme points of all polygons, and β by the initial value of H found in step a. This approach guarantees to obtain a solution very close in value to the optimum, even in the (pathological) case where the costs D_i vary largely in a very small region far from d_i, $i = 1 \dots m$. Finally, for those who do not wish to sacrify a map and a bottle of ink, we note that an automatic drawing table could be used (*).

4. Concluding remarks

(i) Methods Big Square-Small Square and Black and White are quite general and applicable to a variety of continuous location problems. Some exemples are discussed in Hansen and Thisse (1981) and in Hansen, Peeters, Richard, Thisse (1981). Also, the maximin problem in three-dimensional space studied by Dasarathy and White (1980) could be treated by an extension of the former method. (The latter one could be used but recognition of feasible regions in $\mathbb{R}^3$ seems difficult).

(ii) A comparison with the problem of locating a desirable facility is of interest. Recall that in the Weber and Rawls problems the total access cost and the maximal access cost are respectively minimized, the access cost associated with d_i being *increasing* in the distance from d_i. First, as far as localization theorems are concerned, Theorem 1 compares with Theorem 1 of Hansen, Peeters and Thisse (1981): in both cases locations in the intersection of the convex hull C of $\{ d_1 \dots d_m \}$ with the set S of feasible locations are considered; but points of S - C *visible* from C - S replace points of S - C *remote* from C. On the other hand, there is no counterpart to Theorem 2 for linear cost functions. Second, a similar analogy holds for resolution methods. In both problems Big Square-Small Square or Black and White applies in the general case, whereas no polynomial algorithm exists for the linear version of the Weber problem (at least when the norm is not rectilinear).

(iii) Theorems 1 and 2 compare to those obtained by Church and Garfinkel (1978) for the location of an obnoxious facility on a network. Thus, Theorem 1 is reminiscent of the main result of those authors, which states that the optimal solution is to be sought in the set of tips and of bottleneck points of the network. Tips correspond to remote points and bottleneck points to points of C $\cap$ S. When the network is a tree, i.e. when the shortest distance displays some convexity, only the set of tips is to be considered as only the set of remote points in Theorem 2.

(iv) Our results suggest that obnoxious facilities will be frequently located at the limit of the area under control of the planner. Considering a larger area does not solve the problem unless very lowly populated zones become available. (Thus American nuclear plants could be located in deserts of the U.S.A.). A more general and more satisfactory approach would consist in introducing a compensation scheme for populations who suffer from the pollution. Locational and non-locational variables are then integrated in procedures which aim to optimize the *global* efficiency of the system (see, e.g. Smets, 1973, d'Aspremont and Gérard-Varet, 1981, for a treatment of such procedures in the case of the transfontier pollution problem).

References

Austin C. M. (1974) The evaluation of urban public facility location: an alternative to benefit-cost analysis, *Geographical Analysis, 6,* 135-145.

Austin C. M., Smith T. E., Wolpert J. (1970) The implementation of controversial facility-complex programs, *Geographical Analysis, 2,* 315-329.

Brady S. D., Rosenthal R. E. (1980) Interactive computer graphical solutions of constrained minimax location problems, *AIIE Transactions, 12,* 241-247.

Church R. L., G^rfinkel R. S. (1978) Locating an obnoxious facility on a network, *Transportation Science, 12,* 107-118.

d'Aspremont C., Gérard-Varet L. A. (1981) Regional externalities and efficient decentralization under incomplete information, in Thisse J.-F., Zoller H. G. (eds.) *Locational Analysis of Public Facilities* (forthcoming).

Dasarathy B., White L. J. (1980) A maximin location problem, *Operations Research, 28,* 1385-1401.

Dodu J.-C., Maréchal P. (1980) Un modèle pour la détermination de la localisation optimale des moyens de production: le modèle Tassili, Electricité de France, *Bulletin de la Direction des Etudes et Recherches, série C, Mathématiques, Informatique, 2,* 5-24.

Drezner Z., Wesolowsky G. O. (1980) A maximin location problem with maximum distance constraints, *AIIE Transactions, 12,* 249-252.

(*) When completing the present paper, we have learned that an approach similar to Black and White has been explored by Brady and Rosenthal (1980).

Eddy W. F. (1977) A new convex hull algorithm for planar sets, *ACM Transactions on Mathematical Software, 3,* 398-403.

Hansen P., Peeters D., Richard D., Thisse J.-F. (1981) The Rawls locational problem revisited, Université Catholique de Louvain, SPUR, R.P. 13.

Hansen P., Peeters D., Thisse J.-F. (1981) Some localization theorems for a constrained Weber problem, *Journal of Regional Science, 21,* 103-115.

Hansen P., Thisse J.-F. (1981) The Weber problem revisited, Université Catholique de Louvain, SPUR, R.P. 10.

Morris J. G., Verdini W. A. (1979) Minisum 1_p distance location problem solved via a perturbated problem and Weiszfeld's algorithm, *Operations Research, 27,* 1180-1188.

Papageorgiou G. Y. (1978) Spatial Externalities, *Annals of the Association of American Geographers, 68,* 465-476 and 477-492.

Preparata F. P., Hong S. J. (1977) Convex hulls of finite sets of points in two and three dimensions, *Communications of the ACM, 20,* 87-93.

Rawls J. (1971) *A theory of justice*, Harvard University Press, Cambridge, Mass.

Roberts A. W., Varberg D. E. (1973) *Convex functions*, Academic Press, New York.

Smets H. (1973) Le principe de compensation réciproque. Un instrument économique pour la solution de certains problèmes de pollution transfontière, O.C.D.E., Direction de l'Environnement.

Ward J. E., Wendell R. E. (1980) A new norm for measuring distance which yields linear location models, *Operations Research, 28,* 836-844.

Wendell R. E., Hurter A. P. (1973) Location theory, dominance, and convexity, *Operations Research, 21,* 314-320.

Riassunto. Il problema di localizzazione di un servizio nocivo viene considerato schematizzando lo spazio mediante una regione continua e limitata del piano. Vengono proposti alcuni teoremi di localizzazione per i casi in cui l'obiettivo sia la minimizzazione vuoi del costo di nocumento totale, vuoi del costo di nocumento massimo. In ambedue i casi, l'unica assunzione circa il costo, che grava su ciascun abitante, è che esso sia continuo e decrescente rispetto alla distanza dalla sorgente nociva. Le proprietà ottenute sono illustrate con un esempio preso dall'assetto localizzativo delle centrali nucleari in Francia.

Résumé. Le problème de la localisation d'un établissement nuisable est examiné. Des théorèmes de localisation et des méthodes de résolution sont proposés pour la minimisation du coût total de nuisance et du coût maximal de nuisance, supposé que les coûts de nuisance supportés par les habitants sont continus et decroissants par rapport à la distance. La politique française de localisation des centrales nucléaires permet d'illustrer la portée des résultats obtenus.

Sistemi Urbani, 3, 1981, 319-343 Guida Editori, Napoli, Italia

Multiple objective facility location

C. ReVelle, J. Cohon, D. Shobrys

The Operations Research Group and Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, Maryland 21218, U.S.A.

Received 15 June 1980

Abstract. Numerous models of location have been created in the past two decades in response to problems which have arisen in both the private and public sector. The models served up to three functions simultaneously: the siting of facilities, the assignment of people or goods to the facilities and the sizing of facilities. The abundance of modelling efforts stems from the multitude of possible ways to conceptualize the movements or flows and assignments which occur in each location problem setting. The large number of efforts also arises because of the mathematical challenges posed by the formulations which include the nefarious zero-one variables. To a very great extent, however, the numerous modelling efforts can be ascribed to different views of the objectives of location problems. Population travel burden, population coverage, number of facilities, transport costs, transport and facilities cost, profits, etc. have all been suggested as objective for location problems. How does one reconcile these often divergent objectives to provide information in a rational manner for decision makers? The ability to tradeoff the levels of achievement of these objectives against one another, depicting at the same time the impact on decisions, is an important need.

Accomplishing such a task raises questions of both a theoretical and practical nature. How to develop the mathematical accounting mechanism which measures and carries the objectives is one such question. How to display the objectives achieved by a given solution is another question. How to compare solutions to facilitate the discarding of inferior location patterns is still a third open question. In this review, we will discuss both our experience in this area and recent results of our research.

Key words: multiple objectives, facility location, trade-off curves.

1. Introduction

A rich literature of location models has developed in the last 15 years. A variety of models has been formulated and applied to facilities ranging from plants and warehouses to libraries, emergency facilities, power plants and nuclear wastes. In this article, we attempt to trace and to categorize these developments by focussing on the objectives used in the models.

The objectives, which often distinguish one location model from another, represent an intriguing element of location analysis where precise statements of objectives are frequently elusive. Multiobjective location analysis, a relatively recent addition to the literature, offers an opportunity for enhancing the utility of location analysis.

We here shall review and evaluate the objectives of two types of location problems:

(1) Those location problems that attempt to characterize the good-oriented decisions, such as are made by corporations and governments. Examples are the location of warehouses, plants and waste disposal sites. It is most common that the number of facilities in these models is determined by the solution to the problems.

(2) Those location problems that attempt to characterize decisions relative to a consuming public. These decisions which are commonly made by governments consider such facilities as schools, hospitals, fire stations, etc. In these models the number of facilities is commonly fixed in advance.

While this dichotomy is not fully accurate (counterexamples will be seen within this paper), it will be useful to us in developing an orderly view of the many problem statements which have evolved. The review will focus on the structuring of objectives for these problems rather than on the development of solution algorithms, although brief mention will be made of the more common solution approaches. We in no sense deny the importance of solution methods; rather we have simply narrowed our purpose in the hope of creating a more coherent picture of the objectives which have been developed for these problems. We will also restrict our attention to problems of location on a network.

2. Location decisions relative to the movement of goods/material

We begin with those location problems which involve decisions about the sites to manufacture, store or dispose of goods and materials. These problems do more than choose locations; they propose shipments from or to the sites, choose capacities for the sites and may even suggest prices for the delivered goods. The earliest location problems of this sort suggested the minimization of the cost of shipment and manufacture or storage. Indeed this theme has been a consistent one for more than two decades, altough variations have occurred. See, for instance, Kuehn, Hamburger (1963), Balinski (1965), Efroymson, Ray (1966), Davis, Ray (1969), Khumawala (1972) and Erlenkotter (1978), among others. All of these investigators structured solution approaches to this problem with only minor variations in the objective.

Their problem can be described with the following set of decision variables and parameters.

Define:

I = the set of demand points ($i = 1, 2, \dots, n$),

J = the set of potential facility sites ($j = 1, 2, \dots, m$),

a_i = the goods demanded by i,

c_{ij} = the cost of shipping one unit of goods from j to i via the least costly route,

f_j = the cost to establish a facility at j,

e_j = the cost per unit of expansion at site j,

x_{ij} = the fraction of the demand at i provided by a facility at site j,

y_j = (0, 1) variable; a value of 1 denotes establishment of a facility at j.

The objective function, using these definitions, is:

$$\text{Minimize} \qquad Z = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}\, a_i\, x_{ij} + \sum_{j=1}^{n} f_j\, y_j + \sum_{j=1}^{m} e_j \sum_{i=1}^{n} a_i\, x_{ij} \tag{1}$$

which, of course, can be condensed to two components, fixed costs plus transport and manufacturing cost.

One significant variation is that suggested by Maranzana (1964) who sought to locate a fixed number of warehouses in such a way as to minimize the cost of shipment. Excluded from his objective was the cost of storage and/or manufacture at the site at which the goods originated. While Maranzana offered no justification for the omission of this cost component, he could have argued that management had dictated the number of warehouses in advance based on other unstated criteria. Alternatively he could have argued that management had given him a budget for facilities within which to work and that the costs of the facilities (the cost to establish and the cost to store) were virtually the same at all sites. Such a structure gives rise to a limit on the number of facilities and eliminates the associated cost component from the objective.

Investigators whose efforts are directed toward the objective of minimum cost rarely offer a justification of this form as the appropriate measure to optimize. Interestingly, another and realistic objective, the maximization of profit (the difference between revenues and costs), provides a theoretical justification for the objective of minimum cost. When the demands are known in advance and the prices at which the goods are sold are also known in advance, one can calculate the total revenues that will be derived from sale of the goods no matter the location decisions. This is a fixed number since all demands are to be fully met, regardless of the net return from sale of goods at a particular point of demand. Given the total value of revenues, we calculate profit as the difference between those revenues and costs, and note that the minimization of costs achieves the maximum profit. This is the theoretical justification of the objective of minimum cost.

More correct from a market economics point of view would be the maximization of profit with the demand set not fully determined rather than the minimization of costs; we discuss the maximum profit objective in the context of location decisions next. There are several assumptions that can be made about the price and the demand for the good at a particular demand point. One assumption is that at each point of demand the price is determined in a competitive market and that the industry in question sells at that price; this is an assumption commonly made by the chemical industry in planning for new capacity.

This is one of the assumptions in the first published paper to raise the objective of maximum profit, Jucker, Carlson (1976); their analyses considered other aspects of the plant location/distribution problem as well. Their statement of the maximum profit problem was in precisely the same notation as we used for the minimum cost problem stated by Balinski [equation (1)], except for the following parameters. They defined:

p_i = the per unit selling price of the product at demand point i, and

$r_{ij} = p_i - e_j - c_{ij}$ = the marginal profit derived from selling an additional unit ot product from plant j at demand point i.

Their objective could then be stated as:

$$\text{Maximize} \qquad a = \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij}\, a_i\, x_{ij} - \sum_{j=1}^{n} f_j\, y_j \qquad (2)$$

The constraint set written by Jucker and Carlson was the same as that for the minimum cost location problem, except that they noted that the constraint which says that each demand point must be supplied, i.e.,

$$\sum_{j=1}^{n} x_{ij} = 1$$

should be replaced by

$$\sum_{j=1}^{n} x_{ij} \leqslant 1$$

in order to allow the option of not supplying a demand point if r_{ij} is negative for the «nearest» open plant. Further, any variable with a negative r_{ij} need not be included in the formulation, a potential computational savings.

It is of interest that the maximum profit objective of Jucker and Carlson is the term known as «producer surplus» in the more general model due to Wagner, Falkson (1975).

The solution methods for the minimum cost problem such as the Branch and Bound algorithm of Efroymson and Ray or relaxed linear linear programming (see Morris, 1978) should carry over with no difficulty to the maximum profit problem.

Jucker and Carlson (1976), Hansen and Thisse (1977) and Erlenkotter (1977) all suggest an additional view of the profit maximization objective, although the Erlenkotter work is more general and integrates several other models of location in the literature. They assume a demand curve exists at i which provides information on the demand generated by a particular price at i. Now the decision is expanded to include the price to charge at i which in turn fixes the quantity to supply to i, as well as locations and flows.

They introduce $p_i(a_i)$ as the demand curve at i where

a_i = quantity demanded, an unknown, and

$p_i(a_i)$ = the price which generates a demand of a_i.

The revenue term of objective then is

$$\sum_{i=1}^{n} p_i(a_i) \cdot a_i$$

and costs of supplyng and producing must be subtracted from this revenue. Further details are omitted but it should be noted that flow quantities rather than 0,1 variables are required to express the objective properly.

Nearly all of the plant location papers consider only corporate objectives. An exception is the model of Cohon *et al* (1980) for the location of electric generating capacity. In this work, the corporate cost objectives are traded against public objectives of environmental quality and perception of safety.

The model developed by Cohon *et al* (1980), for regional energy planning and plant siting policy analyses, uses multiobjective linear programming to estimate the noninferior (efficient) set defined over four objectives: two minimum cost surrogates, water reservoir capacity minimization and the minimization of people residing within 50 miles of a nuclear reactor. These objectives arose from utility and public concerns over power plant site selection.

In selecting power plant sites, utilities were assumed to be most sensitive to those costs which vary appreciably with location. Two such cost categories were identified: costs associated with transmitting power from plants to load centers and costs for shipping coal from mines to plants. Due to the uncertainties of future costs for land acquisitions and other components of transmission line planning, the two cost objectives were treated separately and measured in physical units: megawatt-miles and ton-miles for total transmission and coal shipment, respectively.

Reservoir capacity minimization was used to represent both public and utility desires to avoid new construction of water impoundments, an increasingly controversial activity. The «population proximity» objective was a representation of the public's perception of nuclear plants as sources of danger. The minimization of the objective, measured in megawatts-people, led to the selection of remote sites for nuclear plants.

3. Governmental location decisions relative to a consuming public

Of this class of model, the first formulation in both sequence of development and in terms of its theoretical importance is the p-median problem, so named by Hakimi (1964, 1965). Altough Hakimi was interested in the location of switching centers in a communications network, researchers quickly recognized the applicability of the formulation to the problem of central facilities to which people might come for service.

The possibility of service radiating from the facilities to points of demand was also recognized. The problem with people travelling to facilities may be stated as:

Locate p facilities on a network of demands so that the average travel time of all users is a minimum. Every user is assumed to travel to his nearest facility.

We can assume that the set of eligible sites for facilities is precisely the set of demand points without a loss of generality, a fact asserted and proved by Hakimi (1964, 1965). This problem can be stated as a zero-one programming problem as follows:

$$\text{Minimize} \quad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \, d_{ij} \, x_{ij} \tag{3}$$

$$\text{subject to} \quad \sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, 2, \ldots, n$$

$$x_{ij} - x_{jj} \geqslant 0 \qquad i,j = 1, 2, \ldots, n \qquad i \neq j$$

$$\sum_{j=1}^{n} x_{jj} = p$$

$$x_{ij} = (0,1) \qquad i,j = 1, 2, \ldots, n$$

where a_i = relevant population at demand node i;

d_{ij} = shortest distance, node i to node j;

n = number of nodes;

p = number of facilities; and

x_{ij} = 1 if node i assigns to a facility at j, o otherwise.

Numerous solution procedures have been advanced for this problem statement and a listing through about 1977 is found in ReVelle *et al* (1977). Since that time several additional works have appeared on this subject; these include papers by Narula *et al* (1977), ReVelle *et al* (1979), Boffey (1978) and Galvao (1978). The work of Narula *et al* (1977), in particular seems to offer promise of the ability to handle relatively larger p-median problems than have been solved in the past.

Our interest here, however, is with problem statement not with solution procedure, although we do not deny the importance of the latter. Even so simple a problem as this can be viewed as a problem in two objectives, in this case, the average travel burden and the number of facilities. Trade-off curves which place these two objective in opposition can easily be constructed. There are other ways in which the p-median formulation can be viewed as a multiple objective problem, but we postpone these for a more general discussion which will include other model types. Nonetheless, one additional objective applied to p-median leads to both new insights and new models.

That objective, or consideration, is the maximum time or distance which can separate a user from his nearest facility. That objective was first included in the p-median model by Toregas et *al.* (1971) who showed the form of the trade-off between the average travel distance and the maximum allowable separation. In the graph of fig. 1 a version of which originally appeared in Toregas *et al* (1971) s* is the maximum distance separation above which one observes no effects on the solution to the p-median problem. Reduction in the maximum allowable distance produced tighter and tighter constraints, driving up the average travel burden. Values of maximum distance reduced in increments from s* to s_{min} will gradually increase average distance, until at values below s_{min}, it will be found that no arrangement of p facilities can be found which achieves the desired maximum separation. Interestingly, not all of this trade-off curve is meaningful in the multi-objective sense because some of the points are inferior or dominated. Which points these are will become clearer in a moment.

The addition of a maximum distance constraint produces both a multi-objective view of the p-median and a new problem. The point s_{min} on the graph represents the final contortion of the p facilities in the above problem. No re-arrangement can make the p facilities cover all points of demand within maximum separation values of less than s_{min}. The number of facilities is simply insufficient. What number of facilities is sufficient to insure a particular maximum separation of users and facilities which is less than s_{min}? This leads us to the statement of the location set covering problem as posed by Toregas *et al* (1971).

> *Find the minimum number of facilities and their locations such that each point of demand has a facility within S time units.*

In mathematics, this problem may be stated:

$$\text{Minimize} \quad Z = \sum_{j=1}^{n} x_j \tag{4}$$

$$\text{subject to} \quad \sum_{j \in N_i} x_j \geqslant 1 \qquad i = 1, 2, \ldots, n$$

$$x_j = (0,1) \qquad j = 1, 2, \ldots, n$$

where n = number of nodes;

d_{ij} = shortest distance, node i to node j;

S = maximum allowable distance that may separate node i from its nearest facility; and

$N_i = \{ j \mid d_{ij} \leqslant S \}$.

Again, the assumption here is that demand nodes and facilities sites are co-incident, but the sites and demand nodes could be disjoint or overlap as required by the problem setting. The appropriateness of this problem statement is emphasized by the verbal statement of Huntley (1970) who was searching for criteria for the location of amburlance dispatching points. Huntley posed virtually the same problem, although he made no attempt to solve it; indeed, he was only interested in problem statement. Several methods have been developed to solve this zero-one program (Toregas *et al*, 1971 and Toregas, ReVelle, 1973).

In the location set covering problem, it is logical to examine how the number of facilities and their pattern of deployment are influenced by the maximum value of the separation distance. Such an examination leads to a multiple objective trade-off curve such as the one of fig. 2. Striking properties are associated with the trade-off curve of number of facilities versus maximum distance, properties we can observe from the example curve.

Note that the curve exhibits the expected increase in the number of facilities as the maximum distance is tightened (reduced); the increase occurs, however, in jumps, preceded by flat spots. If we observe the solutions, moving from right to left along these flat spots. such as between S_B and S_A where four facilities are required, reducing the maximum distance has no effect on the number of facilities, although the arrangement of the facilities may be altered to meet the more stringent distance constraint. The solution at S_A could have been found

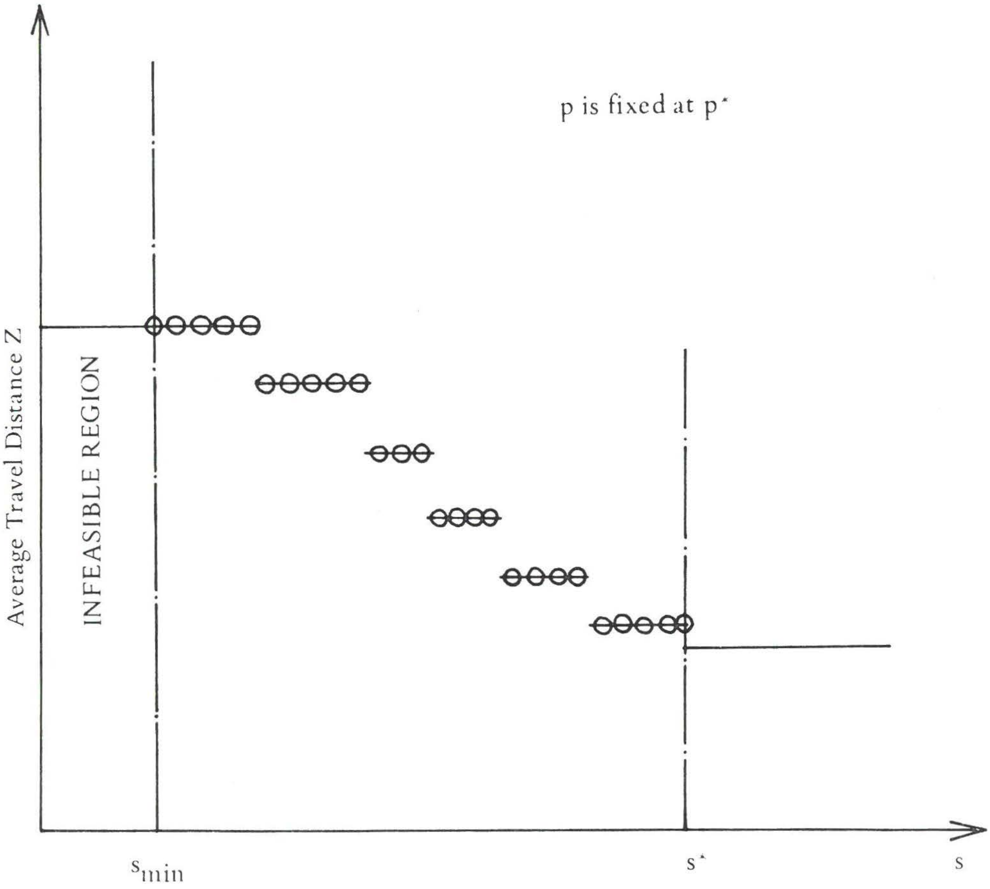


Figure 1 Maximum distance travelled
Average travel distance as a function of the maximal distance travelled

when solving at a requirement of S_B since it is an alternate optimal solution to the problem of minimizing the number of facilities subject to a distance of S_B. Indeed, the arrangement of facilities at S_A is the solution to the problem of minimizing the maximum distance that separates any user from his nearest facility subject to the number of facilities being equal to four. This is a discrete solution space version of the problem which Hakimi (1964, 1965) named the p-center problem.

The striking property in terms of multi-objectives derives directly from the observation that the left most corner point of these flat portions of the curve solve the minimum maximum distance problem. Because number of facilities and maximum distance are the only objectives of concern here, it follows that for a given «flat spot», all solutions to the right of the corner point are dominated by (are inferior to) the left most point. That is, the corner point utilizes the same number of facilities but achieves a better value of the other objective, maximum distance. The non-inferior set then consists of the circled points in the figure. The discovery of dominant or non-inferior solutions in this problem gives us insight into the distance-constrained p-median problem. The same property of non-inferiority is present in the trade-off curve, shown in fig. 1; that is, the left most corner points (circled) dominate all other points on the adjacent flat spot.

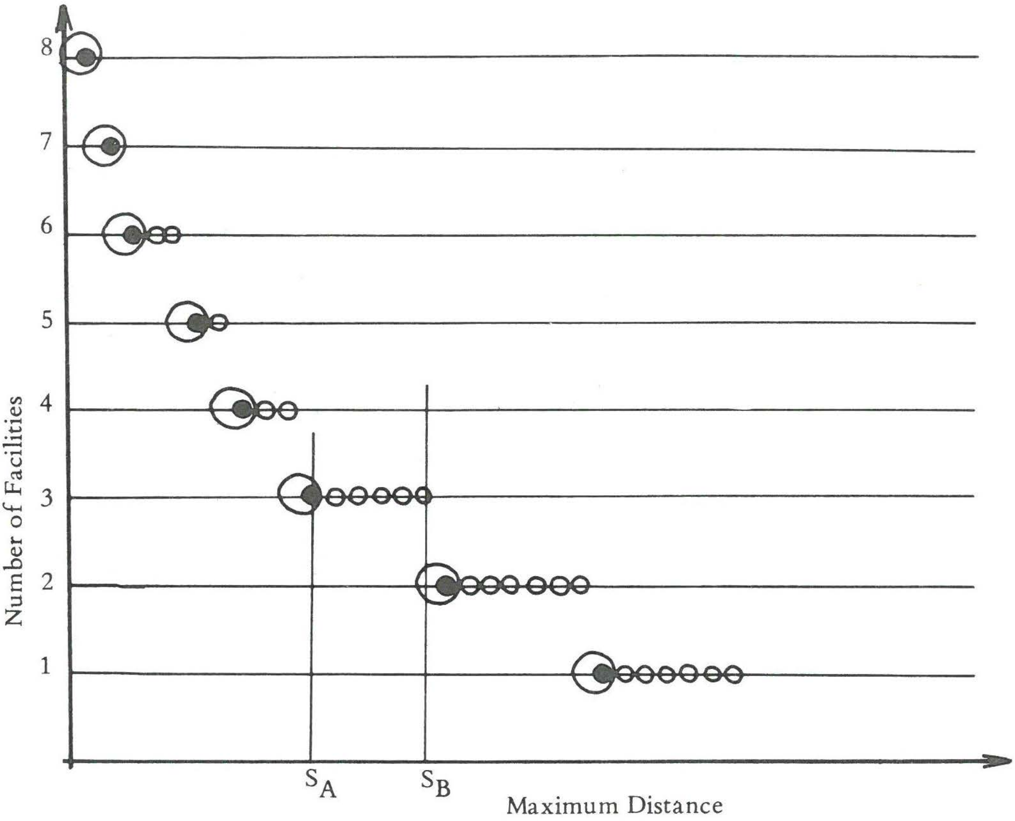


Figure 2 Number of facilities as a function of maximum distance

This trade-off curve from the location set covering problem is an unusual non-inferior set. That the trade-offs are not continuous is a consequence of his being a zero-one problem. Further, all points in the gaps between the corner point solutions are known to be inferior. Usually one cannot eliminate these «gap» points from consideration. Examples in which gap points are possible will be shown later.

The location ser covering problem, over the years since it was introduced, has been an appealing problem statement. The notion of complete coverage of all points of demand, however, was seen as very restrictive. Insistence on complete coverage could lead to the need for more facilities than the budget allowed. If ten facilities are required so that all demand points are covered within 30 minutes and only seven facilities can be afforded, the next logical question is how to deploy the seven for maximal effectiveness. The Maximal Covering Problem, formulated and solved by Church, ReVelle (1974), can be stated as:

Allocate p facilities to positions on the network so that the maximum population will find service within a stated time or distance standard.

In mathematical terms, the problem is:

$$\text{Maximize} \quad Z = \sum_{i=1}^{n} a_i y_i \tag{5}$$

$$\text{subject to} \quad y_i \leqslant \sum_{j \in N_i} x_j \qquad i = 1, 2, \ldots, n$$

$$\sum_{j=1}^{n} x_j = p$$

where the only new definition is

$y_i = 1$ if point i is covered within S, 0 otherwise.

All other terms are as defined earlier. Again, this particular formulation assumes coincidence of demand points and potential facility sites, an assumption easy to relax or alter.

Just as the p-median problem is implicitly a multi-objective problem in travel burden and number of facilities, so too is the Maximal Covering Problem. One would want to examine the trade-off between population covered within a stated distance and the number of facilities allocated to the network. Such a curve is shown in fig. 3. This curve exhibits the expected concave shape in which each additional facility gains less population coverage than did the preceding one. From the graph it can be seen that covering all the demand points requires 12 facilities, a 50 percent increase from the eight facilities needed to cover 90 percent of population. The information for choices by the decision maker is clearly laid out.

It is in the context of the maximal covering location problem that the need for a multiple objective examination of alternatives becomes most apparent. Two examples will be used to illustrate the importance of a multi-objective approach to the maximal covering problem.

First, the population coverage we spoke of by facilities or services implied that the population was always in the same locale, day and night, all seasons, and year after year. Of course, it is not. Work takes people from their dwellings day and night, placing them in shops, offices and factories for substantial portions of their time. Seasons may see population changes during the calendar year due to tourist movements and the movements of migratory labor. Patterns of migration may change the spatial structures of cities and regions through time.

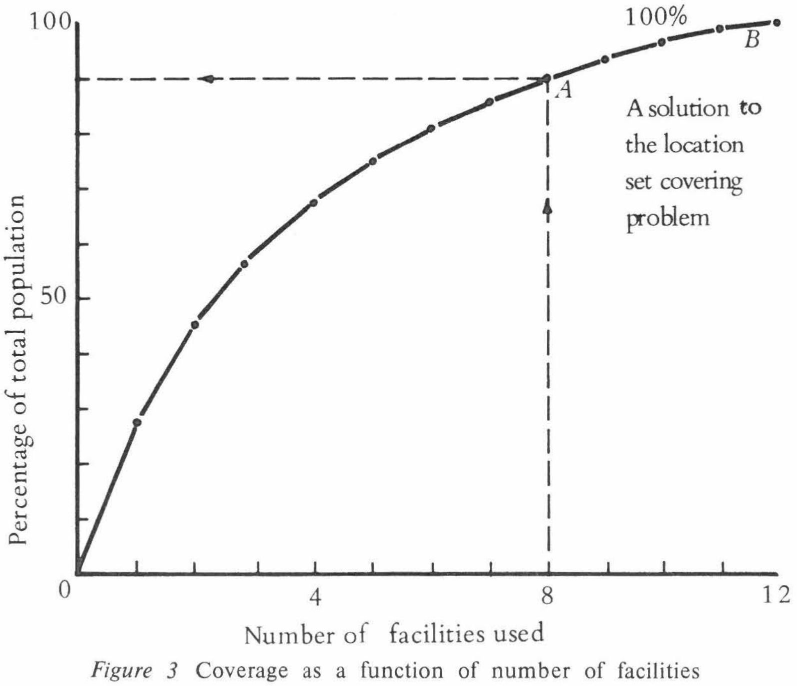
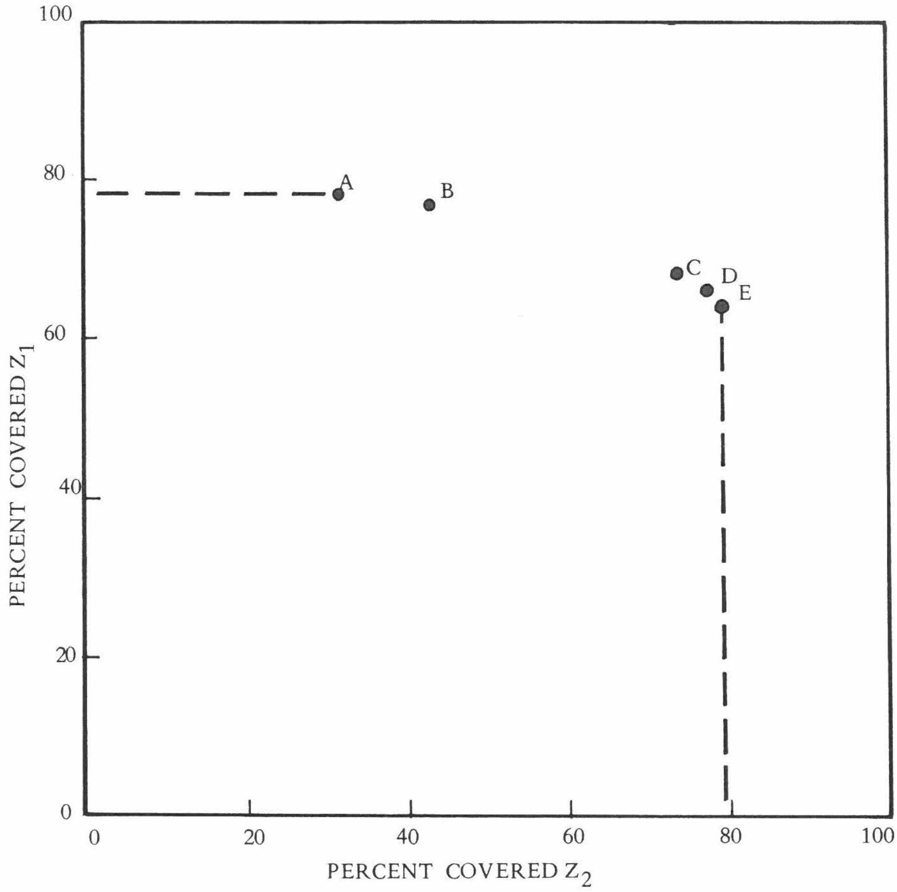


Figure 3 Coverage as a function of number of facilities

Thus in fig. 4, we display a trade-off curve derived from hypothetical data which illustrates how well the facilities that optimize one objective achieve another coverage objective. As one ranges the weights on the two kinds of populations, coverage emphasis gradually shifts from one objective to the other. Two properties of these solutions are worth exploring further.


100
80
60
40
20
0
0
20
40
60
80
100
PERCENT COVERED Z_1
PERCENT COVERED Z_2
A
B
C
D
E

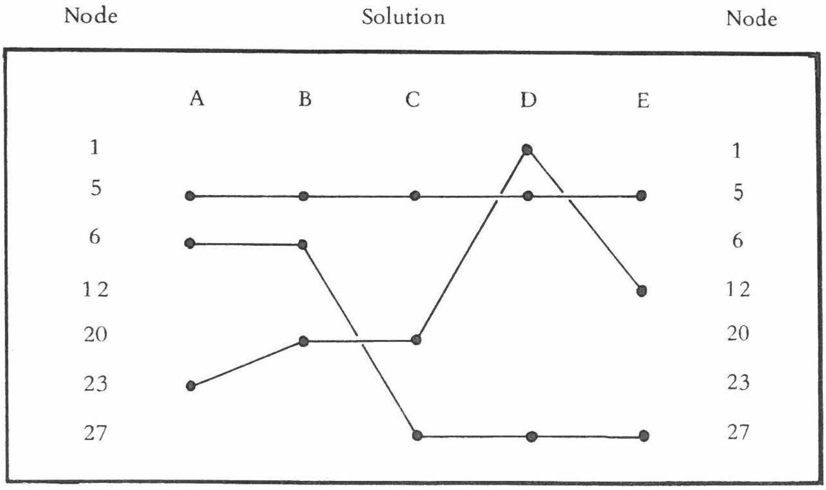

Node
Solution
Node
A
B
C
D
E
1
5
6
12
20
23
27
1
5
6
12
20
23
27

Figure 4

One such property is that the solution points shown are only those on the outer hull of the bicriterion space. Points may exist in the gaps between these hull points which are non-inferior. The weighting method of generating alternatives was used to produce these hull points because the problem is a zero-one programming problem solved by a method which does not admit the addition of constraints. (Constraints on other objectives are likely to produce fractional solutions). The weighting method is unable to produce such gap points as they lie interior to the hull and hence will never be contacted by the outward moving plane of the two-objective function.

Another property of interest pertains to decision space. Note in the graph below the trade-off curve (*) the positioning of facilities and the gradual shift in the recommended positions of the facilities with a change in emphasis on the objectives. It is striking that each successive point on that trade-off curve corresponds to the movement of a single facility from one position to another. More importantly, one facility does not shift position through the entire range of exploration of the two objectives. We have dubbed the set of facilities whose position remains constant through the range of objectives as being in the «core». Such «core» facilities seem to be logical candidates to include in a decision no matter the decision maker's position on the values of the objectives. About facilities in the core, if all objectives have been appropriately enumerated, there can be no disagreement.

Our first example of multiple objectives in the Maximal Covering Problem was of populations which move through time. This raises the possibility of trading present coverage against future coverage, a possibility explored by Schilling (forthcoming) in the context of siting facilities in an environment evolving in time. Not all coverage objectives need have population units though.

The trade-off between covering day and night populations or present and future populations is not confined to the maximal covering problem. Such populations can be considered in the p-median problem as well. The twin objectives in the case of day and night are (1) minimize average travel time of the day populations to their nearest facility and (2) minimize average travel time of the night population to their nearest facility. If this is travel to, let us say emergency rooms, one could consider the different frequencies day and night of accident/ emergency occurrences in each demand zone.

These three models, the-median, the location set covering, and the maximal covering problem are all similar, all related. The location set

(*) This positioning display of objectives and decisions on the same graph first appears in «Displaying Information from Multi-Objective Optimization», Joint National Meeting TIMS/ORSA, Atlanta, November 1977, by D. Schilling, A. McGarity, C. ReVelle and J. Cohon.

covering can be derived conceptually from tightening maximum distance constraints on the p-median; the maximal covering can be derived from the recognition of limited resources and from the expense of complete coverage of all demand points. Furthermore, Church and ReVelle (1976) have shown that the maximal covering model is a special case from a data standpoint of the p-median model. These three models have one other feature in common; they all use some or all of the same basic data: (1) population or demand and its location and (2) shortest distances or time between demand points and facilities.

In a study of the Baltimore Fire Service (see Schilling *et al*, 1979), other objectives were identified for the maximal covering model. In the United States, fire protection location decisions respond to two separate sets of signals. One signal is direct from the body politic; that directive is that location decisions should be based on saving lives or nearness to people. Another signal is from the fire insurance companies who pay off on policies when structures are damaged by fire. The greater the value of the property lost, the greater the payout of the insurance companies; their interest therefore is in the protection of property. The only fire protection location standards in the United States are those developed by the insurance companies. There standards are stated in terms of the nearest equipment to property. High value districts require closer coverage than residential areas. Population densities are absent from the criteria. Local fire departments are graded by the insurance companies on the basis of the degree to which they meet the criteria for nearness to nigh value and to residential districts. High value districts are required to have closer coverage than residential districts.

These two sets of signals, protecting people and protecting property, yield a number of new objectives; the following objectives were developed for the previously mentioned Baltimore Fire Study:

maximize population covered within a distance standard;

maximize the value of property covered within a distance standard;

maximize the area covered within a distance standard.

The degree to which one would wish to cover people and property would be influenced by the risk as exemplified by fire frequency in a given locale. A high fire frequency in a given area would suggest a geater need for coverage than a comparable area of low fire frequency. Thus, in addition to the three objectives above, we formulated three more objectives:

maximize fire frequency covered within a distance standard;

maximize coverage of people at risk (fire frequency times population) within a distance standard;

maximize coverage of property value at risk (fire frequency times property value) within a distance standard.

Alternate configurations of fire suppression equipment were analyzed and compared using these six criteria. The simplest trade-off curves might display population coverage on one axis and property value coverage on the other for a given quantity of equipment available, but clearly proper comparison of alternatives should involve all the dimensions of the decision process.

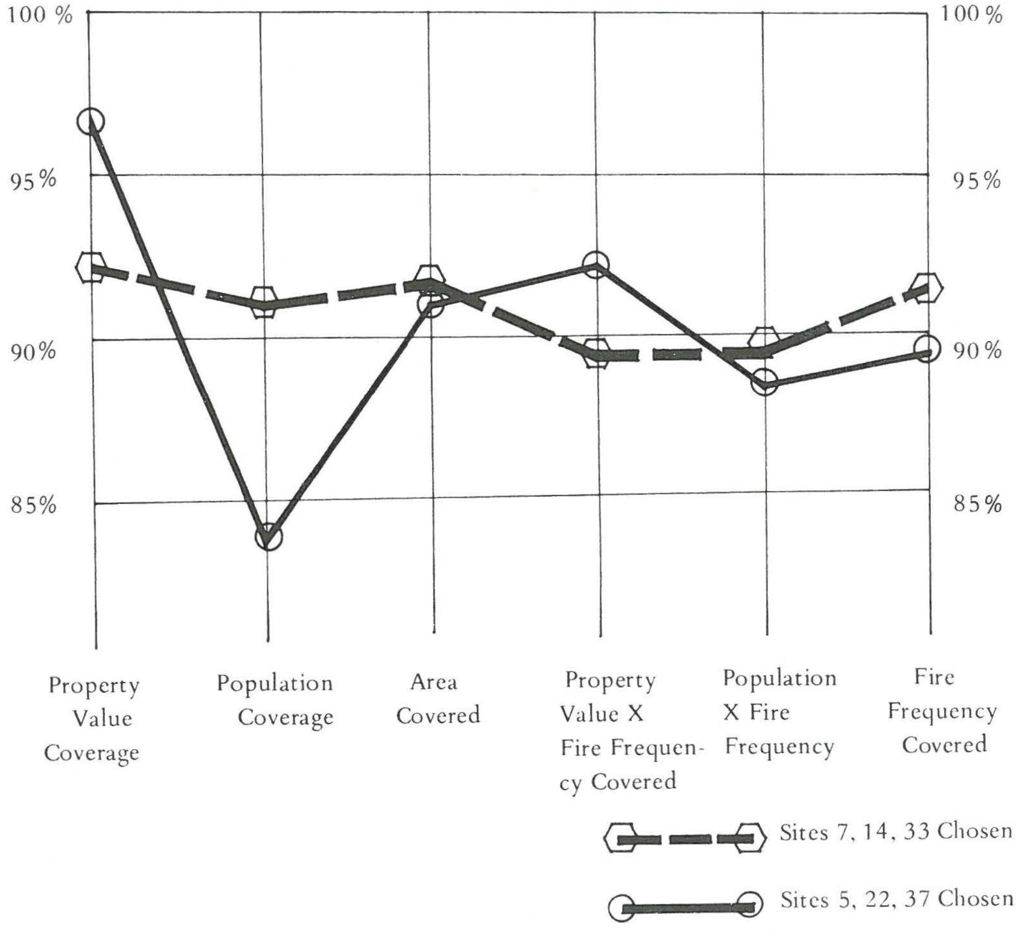


Figure 5 A value path display of two location alternatives

For this reason we developed a method of simultaneously displaying the relative achievement of the six objectives by a given pattern of equipment deployment. The method, while developed to display the objectives achieved by a given set of location decisions, is general in its utility. We call the display technique Value Paths (see Schilling, 1976, and Cohon, 1978) for the lines which trace out the levels of achievement of a particular alternative.

A Value Path display begins with an equal spacing of vertical lines of the same height. Each vertical line represents an objective, and the position of the intersection of the value path with the vertical line indicates the level of achievement for that particular objective. The lines may have physical units or may be in percentage terms, the top of the line representing 100 percent attainment of the maximum possible level of achievement of that objective. A typical value path for a location alternative is shown in fig. 5. Under certain circumstances (some simple, some complex) it is possible to discard a value path because it can be shown that the path is dominated by some other path (by some other set of locations).

Still another objective was identified in the Baltimore Fire Study. While coverage was important to the department, the rule of assignment to closest non-busy company (brigade) can cause some companies to work significantly harder than others. Workload was thus added as a seventh criteria in the study. Workload has been reported on as a criteria by Weaver, Church (1980), and by Siler (1977), both papers addressing the workload of ambulances rather than fire equipment.

Multiple objective optimization can do more than shift facilities around in their positions. In one model we created (Schilling *et al*, 1980), the problem was to allocate two types of equipment for fire protection. In this model a mutli-objective optimization also can suggest the conversion of one type of equipment to another.

In the fire protection system in the United States, rescue services are to be provided by ladder companies (brigades) and the job of extinguishing the fire is borne by the engine companies. Coverage standards are stated both for engine companies and ladder companies. That is, a demand area has engine coverage if an engine company is within a stated distance for such coverage and a demand area has ladder coverage if a ladder company is within a stated distance for this type of coverage. The problem may be stated as:

Determine the trade-off between the property value which can be provided with engine coverage and the population which can achieve ladder coverage for a given number of companies (brigades) which can be allocated to positions on the network as either engine companies or ladder companies.

In mathematics, the problem is stated thusly,

$$\text{Maximize} \quad Z = \sum_{i \in I} a_i^E \, y_i^E \, , \qquad \sum_{i \in I} a_i^L \, y_i^L$$

subject to:

$$y_i^E \leqslant \sum_{j \in E_i} x_j^E \qquad i \in I$$

$$y_i^L \leq \sum_{j \in L_i} x_j^L \qquad i \in I$$

$$\sum_{j \in J} x_j^E + \sum_{j \in J} x_j^L = p$$

where

I = the set of demand areas, indexed by i;

J = the set of sites (indexed by j) at which fire companies may be placed (this may be the current set of fire stations);

a_i^E = the property value of demand area i (assumed to be coverable only by engine companies);

a_i^L = the population of demand area i (assumed to be coverable only by ladder companies);

y_i^E = a (0,1) variable, 1 if demand area i is covered by an engine company, 0, otherwise;

y_i^L = a (0,1) variable, 1 if demand area i is covered by a ladder company, 0, otherwise;

x_j^E = a (0,1) variable, 1 if an engine company is placed at site j;

x_j^L = a (0,1) variable, 1 if a ladder company is placed at site j;

p = the number of companies (or brigades) available;

E_i = sites j eligible to provide engine coverage for demand area i $i = \{ j \mid d_{ji} \leqslant S_E \}$;

L_i = sites j eligible to provide ladder coverage for demand area i $i = \{ j \mid d_{ji} \leqslant S_L \}$;

d_{ji} = shortest distance from site j to demand area i;

S_E = distance standard for engine coverage; and

S_L = distance standard for ladder coverage.

As emphasis is shifted from one objective to another, a graph such as the one shown in fig. 4 is traced out. Now, instead of merely shifting facilities, both the positions and type of equipment are influenced as the objectives are ranged in value. The previous model is a relatively simple view of the decision process. That model might be viewed as applicable when the slate has already been written on and fire stations are immovable in the context of current decisions.

In contrast to this view we conceived of a model in which the slate could be partially erased and redone and one in which the positions for new fire stations were as yet unknown. In the FLEET models, since engine or truck companies could not be put at sites where no station existed, we had to account for this aspect as well (this model is a blend of two models presented in Schilling *et al*, 1979, and Schilling *et al*, 1980. Now the problem may be stated as:

Determine the trade-off between the property value which can be provided with engine coverage and the population which can be provided ladder coverage for a given number of companies (undistinguished by type) which can be allocated to the network only where fire stations exist or are placed by the model.

The problem may be formulated as:

$$\text{Maximize} \quad Z = \sum_{i \in I} a_i^E\, y_i^E \,, \qquad \sum_{i \in I} a_i^L\, y_i^L$$

$$\text{subject to} \qquad y_i^E \leqslant \sum_{j \in E_i} x_j^E \qquad \forall i \in I$$

$$y_i^L \leqslant \sum_{j \in L_i} x_j^L \qquad \forall i \in I$$

$$\sum_{j \in J} x_j^E + \sum_{j \in J} x_j^L = p$$

$$\sum_{j \in J_o} x_j^S = q \,, \qquad \sum_{j \in J_N} x_j^S = r$$

$$x_j^E \leqslant x_j^S \qquad x_j^L \leqslant x_j^S \qquad \forall j \in J$$

where the new terms are

x_j^S = a (0,1) variable, 1 if a station is placed at j;

J_o = the set of sites where stations are currently positioned entering the model;

q = the number of these sites at which facilities will be retained (which ones to retain are unkown);

J_N = the set of new sites which are eligible to be allocated as new stations;

r = the number of these new sites where stations may actually be positioned (which ones to receive stations are unknown).

In this formulation, as emphasis is shifted from one objective to another, a graph of objectives such as that shown in fig. 4 is traced out. In the previous model, positions and the relative numbers of the two kinds of companies are allocated. In this model, the stations to retain are selected; new stations to build are chosen; the companies are allocated to the positions made eligible by the station decisions; and the types of companies are proportioned between engine and ladder companies. All the decisions are open to change as the objectives are ranged in value.

Another category of model has been considered whose basic data has a different character than the models discussed so far. These models use utilization in the objective and note that utilization of facilities is not a characteristic of a demand point but of the intrinsic demand at the point and the friction of space that separates the demand point but from its nearest facility.

The earliest model to focus on the movement of consumers to publicly owned facilities is due to Teitz (1968). While Teitz did not suggest a methodology to solve the problem that he posed, his contribution is of importance because of the problem form which he suggested. Teitz posed the question of the appropriate location design to maximize the utilization of services given a limited budget for the investment in facilities and provision of service. This objective presumes that knowledge of consumer behaviour has been previously obtained. Teitz, suggests, also, that the monetary constraint should include not only the investment costs which other models explicitly or implicitly assume but also the operating costs of the system which are, in turn, a function of the level of utilization. This latter distinction sets the location model of Teitz distinctly apart from many other views of facility systems.

Teitz envisioned a single service distributed from each of N facilities which are of identical scale (size) S. A zero-priced service is dispensed to all who make the journey to a facility. The total number of users per unit of time, Q, (say, users per annum) is given in non-specific functional form in terms of the variables, scale S and number of facilities N; i.e.,

$$Q = Q(S,N).$$

The system cost is composed of operating costs and capital costs.

Let $C_o(Q)$ = annual operating costs as a function of utilization and

$C_c(S,N)$ = annualized capital costs and annual maintenance expenditures as a function of the scale and number of facilities.

The operating costs are in turn a function of S and N through the dependence of utilization on these variables. Hence the operating cost function is

$$C_o(Q) = C_o(Q(S,N)).$$

Total annual system cost is

$$C_T = C_o(Q(S,N)) + C_c(S,N)$$

which must be less than an annual budget figure, B.

The full problem statement is

$$\begin{aligned} &\text{Maximize} && Q(S,N) \\ &\text{subject to} && C_o(Q(S,N)) + C_c(S,N) \leqslant B. \end{aligned}$$

This is the essence of the Teitz model. It was not translated into spatial decisions by Teitz, but other investigators have attempted this step.

The location problem of Holmes, Williams, Brown (1972) is focused ostensibly on the placement of day care centers. In fact, however, the formulation's emphasis is on utilization of facilities. A utilization model is hypothesized for the number who travel from each demand node as function of the distance to the closest facility.

The utilization function for node i is

$$u_{ij} = \begin{cases} a_i - k_i d_{ij} & d_{ij} < S \\ 0 & d_{ij} \geqslant S \end{cases}$$

where u_{ij} = the number of individuals from i who will utilize a facility at j if that is their closest facility;

a_i = the number of individuals from i who will utilize the facility if it is placed at i;

k_i = a coefficient expressing the decline in utilization per unit distance;

d_{ij} = the distance (shortest) from i to j; and

S = a threshold distance beyond which utilization from any demand point falls to zero.

Since u_{ij} is zero when the distance to the nearest facility is S, the value of k_i for each node i is determined by solving $a_i - k_i S = 0$. The value of k_i then is a_i/S. Thus, the utilization function is

$$u_{ij} = \begin{cases} a_i - \dfrac{a_i}{S.} d_{ij} & d_{ij} < S \\ 0 & d_{ij} \geqslant S \ . \end{cases}$$

The maximum utilization model is structured as a p-median type problem:

$$\text{Maximize} \quad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij} x_{ij}$$

$$\text{subject to} \qquad \sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, 2, \ldots, n$$

$$\sum_{j=1}^{n} x_{jj} = p$$

$$x_{jj} - x_{ij} \geqslant 0 \qquad \begin{array}{l} i,j = 1, 2, \ldots, n \\ i \neq j \end{array}$$

$$x_{ij} = 0,1.$$

There are slight differences between the model shown here and that of Holmes *et al* (1972), but the results of application will be the same. In particular, since we defined utilization as 0 beyond S, assignments of a node beyond S can occur at zero cost. Holmes *et al* (1972) allowed no assignment to be made $\left(\sum_{j=1}^{n} x_{ij} \leqslant 1\right)$ if the distance were greater than S because u_{ij} was allowed to be negative beyond S.

If sufficient facilities are available so that all demands will have a facility *within* S, no utilization will fall to zero. ReVelle *et al* (1975) show that this special situation gives rise to equivalence between the maximum utilization and minimum average distance solutions. The objective function then is

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_i - \frac{a_i}{S} d_{ij}\right) x_{ij} \qquad \textit{OR} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \left(1 - \frac{d_{ij}}{S}\right) x_{ij}$$

$$OR \quad \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \left(\frac{S - d_{ij}}{S} \right) x_{ij} \ .$$

It is sufficient to maximize simply

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i (S - d_{ij})\, x_{ij}$$

which may be rewritten as

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i S x_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{n} a_i d_{ij} x_{ij} \ .$$

Since $\sum_{j=1}^{n} x_{ij} = 1$, the first term of the above expression may be stated as

$$S \sum_{i=1}^{n} a_i \sum_{j=1}^{n} x_{ij} = S \sum_{i=1}^{n} a_i$$

which is a constant and hence is non-optimizable. It is sufficient then to:

$$\text{Maximize} \quad Z = - \sum_{i=1}^{n} \sum_{j=1}^{n} a_i d_{ij} x_{ij}$$

or equivalently to:

$$\text{Minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} a_i d_{ij} x_{ij} \ .$$

That is, for this special case of linearly declining utilization and identical threshold distances, maximum utilization is achieved by minimizing average distance. The constraints written above continue to apply.

While this formulation of Holmes *et al* (1972) seeks maximum utilization, scale is not a variable in the model as it is in the Teitz model. The costs of operation which depend on utilization, also are not considered. Thus, although Holmes pursue the same objective as Teitz, their model still stands apart from it.

In 1977, ReVelle and Church showed how the p-median model could be applied to the problem posed by Teitz. That is, they showed how the p-median format could be used incrementally to arrive at both the decisions on location and on the scale of facilities that Teitz suggested were the crux of the problem. Further details of the methodology are omitted because of space in this brief review.

References

Balinski M. (1965) Integer programming: methods, uses and computation, *Management Science, 12,* 3, 253-313.

Boffey B. (1978) On finding p-medians, International Symposium on Locational Decision, Banff, Alberta, Canada.

Church R., ReVelle C. (1974) The maximal covering location problem, *Papers of the Regional Science Association. 32*, 101-118.

Church R., ReVelle C. (1976) Theoretical and computational links between the p-median, location set-covering and the maximal covering location problem, *Geographical Analysis, 8,* 4, 406-414.

Cohon J. (1978) *Multiobjective programming and planning,* Academic Press, New York.

Cohon J., ReVelle C., Current J., Eagles T., Eberhart R., Church R. (1980) Application of a multi-objective facility location model to power plant siting in a Six-State region of the U.S., *Computers and Operations Research, 7,* 1-2, 107-123.

Davis P., Ray T. (1969) A branch-bound algorithm for the capacitated facilities location problem, *Naval Research Logistics Quarterly, 16,* 3, 331-344.

Efroymson M., Ray T. (1966) A branch-bound algorithm for plant location, *Operations Research, 14,* 361-368.

Erlenkotter D. (1977) Facility location with price-sensitive demands: private, public and quasi-public, *Management Science, 24,* 4, 378-386.

Erlenkotter D. (1978) A dual-based procedure for uncapacitated facility location, *Operations Research, 26,* 6,992-1009.

Galvao R. (1978) A dual-bounded algorithm for the p-median problem, International Symposium on Locational Decisions, Banff, Alberta, Canada.

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

Hakimi S. (1965) Optimum locations of switching centers in communication network and some related graph theoretic problems, *Operations Research, 13,* 462-475.

Hansen P., Thisse J-F. (1977) Multiplant location for profit maximization, *Environment and Planning A, 9,* 1, 63-73.

Holmes J., Williams F., Brown L. (1972) Facility location under a maximum travel restriction: an example using day care facilities, *Geographical Analysis, 4,* 3, 258-266.

Huntley H (1970) Emergency health services for the nation, *Public Health Reports,* June.

Jucker J., Carlson R. (1976) The simple plant-location problem under uncertainty, *Operations Research, 24,* 6,1045-1055.

Khumawala B. (1972) An efficient branch-and-bound algorithm for the warehouse location problem, *Management Science, 18,* 12, B-718 - B-733.

Kuehn A., Hamburger M. (1963) A heuristic program for locating warehouses, *Management Science, 9,* 4, 643-666.

Maranzana F. (1964) On the location of supply points to minimize transport costs, *Operations Research Quarterly, 15,* 261-270.

Morris J. (1978) On the extent to which certain fixed charge problems..., *Operational Research, 29,* 71-76.

Narula S., Ogbu U., Samuelsson H. (1977) An algorithm for the p-median problem, *Operations Research, 28,* 4,709-713.

ReVelle C., Bigman D., Schilling D., Cohon J., Church R. (1977) Facility location: a review of context-free and EMS models, *Health Services Research, 12,* 129-145.

ReVelle C., Church R. (1977) A spatial model for the location construct of Teitz, *Papers of the Regional Science Association, 39,* 129-135.

ReVelle C., Church R., Schilling D. (1975) A note on the location model of Holmes, Williams and Brown, *Geographical Analysis 7,* 4, 457-459.

ReVelle C., Rosing K., Rosing-Vogelaar H. (1979) The p-median model and its linear programming relaxation: an approach to large problems, *Operational Research*, *30*, 815-823.

Schilling D. (1976) Multi-objective and temporal considerations in public facility location, Doctoral Dissertation, The Johns Hopkins University, Baltimore, Maryland.

Schilling D., Dynamic location modelling for public sector facilities: a multi-criteria approach, *Decision Sciences,* Forthcoming.

Schilling D., Elzinga D., Cohon J., Church R., ReVelle C. (1979) The TEAM/FLEET models for simultaneous facility and equipment siting, *Transportation Science,*163-175.

Schilling D., ReVelle C., Cohon J., Elzinga D. (1980) Some models for fire protection locational decisions, *European Journal of Operations Research, 5,* 1, 1-7.

Siler K.F. (1977) Level load retrieval time: a new criterion for EMS facility sites, *Health Services Research, 12,* 416-426.

Teitz M. (1968) Toward a theory of urban public facility location, *Papers of the Regional Science Association, 21,* 35-51.

Toregas C., ReVelle C. (1973) Binary logic solutions to a class of location problems, *Geographical Analysis, 5,* 145-155.

Toregas C., Swain R., ReVelle C., Bergman L. (1971) The location of emergency service facilities, *Operations Research, 19,* 6, 1363-1373.

Wagner J., Falkson L. (1975) The optimal nodal location of public facilities with price-sensitive demand, *Geographical Analysis, 7,* 1, 69-82.

Weaver J., Church R. (1980) A multicriteria approach to ambulance location, Proceedings of the IEEE Pittsburgh Symposium.

Riassunto. I numerosi modelli di localizzazione sviluppati nel recente passato, per applicazioni sia pubbliche che private, tentano di risolvere simultaneamente tre problemi: l'ubicazione dei servizi, l'assegnazione ad essi della domanda (flussi di persone o di merci) ed il dimensionamento dei servizi. La gran varietà di modelli esistenti è in parte dovuta ad aspetti tecnici, quali i molti approcci alternativi agli aspetti combinatori del problema o i diversi modelli di comportamento dei flussi di persone e merci. Tuttavia, in modo più sostanziale si può dire che le vere differenze siano dovute a diversi punti di vista circa gli obiettivi che devono guidare le decisioni localizzative.
Gli obiettivi correntemente usati vanno dal costo di trasporto a carico degli utenti, al numero di servizi da installare, ai costi di installazione e manutenzione, ai profitti.
Poiché questi (ed altri) obiettivi sono spesso in conflitto fra loro, è necessario sviluppare tecniche atte a trovare soluzioni che raggiungano un ragionevole compromesso nel grado di raggiungimento di ciascuno di essi.
In questo saggio vengono discusse tecniche ed applicazioni atte a risolvere matematicamente tali problemi multi-obiettivi, a rappresentare graficamente il grado di raggiungimento dei vari obiettivi, a confrontare diverse soluzioni alternative e facilitare l'eliminazione di assetti locatizzativi non convenienti.

Résumé. Les nombreux modèles de localisation développés récemment, pour des applications publiques et privées, tentent de resoudre en même temps trois problèmes: la localisation des services, l'affectation aux services de personnes ou de biens et l'estimation de la dimension des services. La variété des modèles existants est due, d'une part, aux aspects techniques, tels que les nombreuses approches alternatives aux aspects combinatoires du problème, d'autre part aux divers modèles du comportement des flux de personnes et de biens. Dans une large mesure on peut dire que les différences substantielles sont dues aux différents points de vue quant aux objectifs qui doivent guider les décisions de localisation.
Les objectifs couramment utilisés sont le coût de transport supporté par les usagers, le nombre de services qui doivent être installés, les coûts d'installation et d'entretien, les profits, et cetera.
Puisque ces (et autres) objectifs sont souvent en conflit entre eux, il est nécessaire de développer des techniques capables de fournir des solutions qui parviennent à un compromis raisonnable.
Cet essai décrit des techniques et des applications capables de résoudre mathématiquement tels problèmes multi-objectifs, de tracer graphiquement le degré de réalisation de tels objectifs, de comparer différentes solutions alternatives et de faciliter l'élimination des configurations de localisation qui ne sont pas convenables.

Sistemi Urbani, 3, 1981, 345-390 Guida Editori, Napoli, Italia

Public facility location models and the theory of impure public goods

A. C. Lea

Department of Geography, University of Toronto, Toronto, Ontario M5S 1A1, Canada.

Received 15 June 1980

Abstract. A large number of operational location-allocation models exists for optimally locating systems of facilities. Most of these are believed to be suited for use on public facility problems, yet few appear to have ever been used in practice. A fundamental reason for this is that the models are not underpinned by a rigorous theory. Indeed, the literature has not problematized a theory of public facility location in general. There has been a failure to recognize the public/political/institutional nature of the problem. The welfare economic theory of public goods concerns itself with the types of goods and services provided through public facilities. A spatial generalization of this theory to spatially impure public goods can serve as a rigorous foundation of a theory of public facility location. However, the theory of location must be conceived as part of a more general theory of the public space economy and the relationship between location and other key variables, often of higher order, must be explored. These tasks are necessary for the construction of a new breed of relevant operational location models. In exploring this theme, models in the conventional wisdom are criticized, the theory of pure and impure public goods is surveyed and generalized, and some key questions to be addressed in a theory of the public space economy are set out and their implications for location theory and operational models are examined. The paper is largely non-technical and no prototypes of the new models called for have been set out.

Key words: public facility, location-allocation model, public goods, impure goods, the public space economy, political-institutional problems.

1. Introduction

In most western countries an increasing share of the national product is provided by governments. Most of the goods and services are delivered, or made available, through systems of public facilities. The locational configurations of these facilities are clearly important long-run policy considerations. Since theories and models of public systems and policy have received a great deal of attention in the last two decades, it should occasion no surprise to learn that there is a large body of literature on the important problem of public facility location. Despite this literature, there are still no generally accepted theories of public facility location, in contrast to widely accepted theories of private facility location. The vast majority of contributions to this literature sets out mathematical programming formulations for public facility location-allocation models and offers increasingly sophisticated computer algorithms. Although recently there has been a focus on public facility systems, most of the models have been held to be suitable for both private and public facilities. However, it appears that the models have been used only rarely on real public facility problems.

The first thesis of the present paper is that the theoretical underpinnings of these location models are extremely weak. The second thesis is that a spatial generalization of the recently developed theory of public goods can serve as a sound theoretical basis for a normative theory and operational models of public facility location and for a normative theory of the public space economy in general. An important weakness of most existing models is their failure to recognize the importance of the attributes of the goods on services being provided. More fundamentally, the political, economic, legal and institutional contexts of the location problem tend not to be taken into account. Indeed, the existing literature, with few exceptions, fails to problematize theory in general and a theory of social welfare in particular. The largest and most well-developed body of theory to which we can turn for guidance is that of welfare economics, and the liveliest topic of research in this subdiscipline in recent years has been the theory of public goods. Since public facilities should provide public goods, this theory should be of direct relevance to a theory of public facility location. However, as the received theory of public goods is almost entirely aspatial, a first task must be to generalize it. An important outcome of this process is the realization that the problem of location must be firmly situated within a much broader theory - a theory of the public space economy. In exploring the implications of the theory of public goods for theory and models of public facility location, the scope of the paper is already rather broad. Therefore, a detailed examination of technical considerations has been eschewed and no new prototype models have been presented. These will be the subjects of subsequent papers.

Existing public facility location-allocation models are briefly characterized in the second section and these models are selectively criticized in the third section as lacking a rigorous theoretical underpinning. Several highlights of the theory of public goods are reviewed in the fourth section while, in the fifth, the problem of public facility location is placed in context by sketching the outline of a theory of the optimal public space economy which derives in large part from the theory of public goods. In this section a case is made which justifies concern for policies which do not «appear to» affect the location problem. Section 6 overviews a number of important macro level considerations which impinge directly or indirectly on microlevel location-allocation problems. Most of these considerations are not treated as policy variables by «locational» modellers. To be consistent with the theory of the public space economy, public facility location models must have many new attributes. These are discussed in section 7 in a very general way. In section 8 a number of more technical considerations relating to the construction of location models are addressed. No specific models are set out as these will be the subject of several additional papers. Finally the conclusions are summarized in section 9.

2. An overview of existing location-allocation models

For present purposes it is not essential to survey the large and varied literature on public facilities location-allocation models. The reader need only be familiar with the conventional «sense of problem» which can be characterized by a very general overview.

In the simpler models the number of facilities to be located is given exogenously. The problem is to choose a set of locations for the system of facilities that is optimal for serving the public. The locations may be either anywhere in a continuous space, or on a network, or at a predefined set of discrete points. Typically, as in the popular p-median problem, the objective is to minimize aggregate weighted transportation or travel costs, assuming that demand for the good or service in question is unaffected by the size and locational pattern of the facilities (i.e., demand is exogenous) and that customers are served from, or they serve themselves at, only «closest facilities». The solution to the problem provides not only the locations for the given number of facilities but also the «optimal allocation» of customers, or demand points, to facilities, and the size or capacity required at each facility. Extensions of this basic problem include models which deal with stochastic demands as well as multiple time period (dynamic) models. In addition several recent papers have recognized that whereas it may be reasonable to assume closest facility assignments when the good or service is delivered to consumers, this is generally not a reasonable assumption when consumers may decide to which facility or facilities they will travel. Accordingly various spatial interaction theories (e.g., gravity type models) have been invoked in an attempt to capture realistic facility choice and travel processes.

In the more complete location-allocation models the costs of constructing and operating the facilities are explicity included. In these models, the set of decision variables is expanded to include the number of facilities to be included in the system. The models directly address the trade-off between transportation or travel costs, which are a decreasing function of the number of facilities, and construction/production costs which are an increasing function of the number of facilities, because of the «economies of scale» associated with individual facilities. The most common objective of these models - thought to be most appropriate for systems of private facilities - has been the minimization of total system costs. Another variant proposed as being more appropriate for public systems seeks to minimize user (travel) costs while constraining the costs of construction and provision to be within a given budget. Stochastic and dynamic versions of models which include facility-related costs have also been developed. In addition the more realistic spatial-interaction-based allocation processes noted above have been largely developed within models which include facility costs. Before this rather recent development, systems in which consumers travel were treated as isomorphic with systems in which the good or service was delivered. However, most of the models «on the market» still include naive closest assignment rules and inelastic demands and are held to be appropriate for both delivered good and travelled-for good systems, as well as for both private facilities and public facilities.

In the models which include facility-related costs but especially in the models in which the number of facilities is given, some concessions have been made to the fact that the facilities are provided by the public sector. Most attention has been paid to the criterion for optimality and in particular how some concept of equity can be included. Because many p-median solutions may have great variances in the distances consumers are from the nearest facility (a proxy for variance in quality of service) it is common now to include a maximum distance constraint. In the location set covering problem one may minimize the number of facilities required to serve the population so that no consumer (or demand point) is farther than some specified distance from the closest facility. Alternatively the maximal covering location problem seeks to find the location of a fixed number of facilities which maximizes the number of consumers (demand points) covered within a given distance of the closest facility. Proceeding even farther the objective may be some appropriate definition of equity itself. For example, in the m-centre problem, the objective is to minimize the maximum distance between any customer and the closest facility. This minimax objective roughly approximates Rawl's principle of justice. Although it is more common to optimize either efficiency or equity subject to a constraint (or constraints) on the other, it is also possible to jointly optimize some measure of efficiency and some measure of equity.

3. Some major shortcomings of existing models

In this section some problems associated with the conventional wisdom are briefly discussed. The critique stresses problems which seem to be resolvable (conceptually, theoretically, if not computationally) using a public good paradigm. Some of the problems glossed over here will be taken up in subsequent sections in which solutions to them are sought.

It should first be noted that the objectives of most existing models are not recognized as being *proxies* for social welfare, however defined. Although most of the objectives have strong intuitive appeal, the underpinning theory of welfare in general, and the set of assumptions which must be made in particular, are very seldom explored. This is especially surprising given that the seminal and widely cited contributions of Teitz (1968) and ReVelle, Marks and Liebman (1970) had sections explicitly focussing on many of these issues. Whether or not the objective used adequately captures welfare in any public system, and the one being examined in particular, seldom seems to have been rigorously scrutinized. Such scrutiny would seem to require a powerful and comprehensive theory.

Related to the above is the widely accepted central planning paradigm within which these location models are situated. The role of the planner and the model, and the way in which the output of the model is to be used, are not widely discussed (*). Underlying most presentations of operational models are implicit assumptions that not only should the results of the models be imposed on the landscape but that a rational political/planning process will indeed proceed to implement the «optimal solution». This expectation appears incongruous when confronted with the rather simplistic characterization, and especially with the omission of most of the key institutional variables, of the system being modelled. In the short (or even medium) term, in which some budget may be assumed to exist, it is typically neither expedient nor feasible, economically or politically, to restructure totally existing facility systems. Few public systems are constructed *de novo*. Perhaps more locational models based on an incremental view of planning, locally optimizing small interrelated problems, should be developed. Most existing models «seem» to be aimed at «the long run», but for long run problems surely dynamic models are more appropriate and the issues of either optimal or uncertain budgets and demands must be directly addressed. If the models are deemed to be simply aids for public decision-makers, the way in which they «should» be used should be made explicit. These issues are seldom addressed in the existing literature.

If these models are meant to provide «guidance» for better decision making it is perhaps surprising that they make little or no attempt to include the realistic concerns and constraints which preoccupy most decision-makers. Perhaps most significant among these are issues relating to the political support of constituents, the fear of certain forms of opposition, existing or proposed taxation and user charge policies, and the bargaining that characterizes all budgetal processes. The indifference of most public administrators, planners and politicians to existing location-allocation models should not be a surprise. Most fundamentally, the solutions found by location-allocation models do not appear to address the real location problems faced by real decision-makers. To be sure, models have been used and, on occasion, found very helpful by certain enlightened public decision-makers. However, non-users and most of the users have grave reservations about the wisdom, or economic and political expediency, of directly implementing location-allocation model solutions without considerable «hand and eyeball» adjustment. It is taken as axiomatic that, if truly powerful models were available, which address the actual problems faced, they would be much more widely utilized than the current ones.

Some progress has been made, especially recently, in the construction of more elaborate, realistic and theoretically sound models. Mention has been made above to embedded spatial interaction models for consumer travel, to stochastic extensions and to dynamic models. Mention should also be made of models which can cope with more realistic specifications of facility costs and interactions between facilities in addition to those between consumers

(*) One refreshing exception is Liebman (1976).

and facilities. More realistic «multiple criteria» models have been proposed. Finally, considerable progress has been made in devising more efficient and economical search algorithms for dealing with realistically large problems. However, it must be remarked that most of the progress has been fairly narrowly technical. Relatively little attention has been given to constructing a more elaborate normative theory. Clearly one of the principal reasons for this is that the long established models, and the relatively minor adjustments in these which pass for «new models» (together comprising the vast majority of the literature), leave little scope for theoretical elaboration. It is the contention here that truly useful public facilities models will have to be based on a more profound theory than that underlying the transportation problem of linear programming.

One of the first fundamental questions which must be asked is «What kinds of goods and services are provided at, or through, public facilities?» What are the attributes of these goods? It is argued that the answer to this question alone provides a powerful motivation for a different breed of model. This is because almost all conventional models deal with goods and services which are either private goods or siblings and that, at least in western democracies, *the goods and services provided by governments are non-private goods often with very significant «degrees of publicness».* Existing models might not be very far off the mark if goods with strong public dimensions were essentially similar to private goods; (perhaps unfortunately) however, they are markedly different. Existing models may, of course, still be reasonably adequate for the location of public facilities which provide goods and services that are very close to private goods; however there are few instances of these. When it is recognized that public facilities provide various types of pure and impure public goods the theory of public goods should become a significant reference point for a normative theory of public facility location.

Recognition of the public nature of the goods and services provided through public facility systems raises a large number of difficult questions and puzzles to be solved. Many or most of these questions go far beyond those of location (as we will see) and address some of the longstanding thorny issues in political economy, welfare economics and public finance. Clearly only a cursory and selective survey will be possible in what follows. The problem of public facility location must be firmly situated within a whole set of logically higher order questions and then the problem must be recast to be consistent with the new conceptual framework. What seems to be required is a full-blown theory of the optimal (western) public space economy - a theory which has not yet been published. The main questions to be addressed by this theory will be set out below. Before examining the general theory, some of the implications of basing location-allocation models on the theory of public goods will be previewed.

Models based on the theory of public goods must be structured to deal with a variety of public good attributes because this variety impinges on the objective functions, on the constraints, or both. For example, the models must include a consideration of efficient and equitable institutional exclusionary mechanisms. They must be capable of modelling congestion or crowding and determining its optimal extent and incidence. They must include considerations of optimal taxation and user charge policies; ideally these variables should be determined simultaneously with locations, allocations, sizes, etc. (Note how significantly these considerations differ from those of conventional models). The models should consider the relationship between the public and the private sectors. For example, changes in the land rent surface consequent upon public facility location decisions may significantly alter the welfare conclusions that one may reach if these considerations were omitted. Whether the questions being asked relate to the short run or the long run significantly affect the set of decision variables, the way the models are structured and the conclusions reached. These and other issues will be explored further below.

It is important to note that a small body of literature has developed in recent years which could be considered to be spatial welfare economics (*) This literature tends to use the developed body of public goods theory quite liberally. Seldom, however, has the concern been directed specifically at the location of public facility systems at the scale in which the problem is conceptualized in conventional location-allocation models; rather the concern has been the allocation of production to jurisdictions or regions (**) and very frequently the efficiency implications of interjurisdictional spillovers (***). In

(*) Some of the literature relating most closely to public facility location which could be considered spatial or regional welfare economics follows: Tiebout (1956), Williams (1966), Koleda (1971), Vardy (1971, 1973), Bollobas, Stern (1972), Buchanan, Goetz (1972), Stern (1972), Boskin (1973), Lind (1973), Flatters *et al.* (1974), Schuler (1974), Stull (1974), Talley (1974), Fisch (1975, 1976, 1977, 1980), Getz (1975), Richter (1975, 1978, 1979), Sakashita (1975), Sandler (1975), Wheaton (1975), Hamilton (1976), Helpman *et al.* (1976), Kanemoto (1976), Le Roy (1976), Mathur (1976), Urban Systems Group (1976), Greenberg (1977, 1978), Harford (1977, 1979), Helpman, Pines (1977), Henderson (1977a, 1977b, 1979), Morrison (1977), Pestieau (1977, 1980), Richardson (1977), Westhoff (1977), Wright (1977), Coelho, Williams (1978), Honey, Stratham (1978), Miyao (1978), Papageorgiou (1978), Sonstelie, Portney (1978), Thrall, Casetti (1978), Wooders (1978, 1980), Casetti, Thrall (1979), Ellickson (1979), Premus (1979), Rose-Ackerman (1979), Rufolo, (1979), Thrall (1979), Wildasin (1979), Homma, Yamada (1980), Starrett (1980).

(**) Only a very small literature in this tradition has gone beyond the issue of allocating public good capacity to regions to consider the problem of public facility location explicitly. Most of this exceptional literature follows: Tiebout (1961), Teitz (1968), Smolensky *et al.* (1970), Bollobas, Stern (1972), Borukhov (1972), Zeckhauser (1973), Davies (1974), McMillan (1975), Wagner, Falkson (1975), Capozza (1976), Coelho, Wilson (1976), Fisch (1976), Erlenkotter (1977), Schuler, Holahan (1977), Bigman, ReVelle (1978, 1979), Leonardi (1978, 1980), Harford (1979), Lea (1979a, 1979b, 1980). Some other literature addresses the problem of location of public infrastructure in continuous space: see, for example, Schuler (1974), Kanemoto (1976), Fisch (1977), Wright (1977) and Papageorgiou (1978).

(***) The literature which stresses spillover effects and their resolution includes: Wiesbrod (1965), Olson (1969), Koleda (1971), Vardy (1971, 1972, 1973), Buchanan, Goetz (1972), Sandler, Shelton (1972), Flatters *et al.* (1974), Talley (1974), McMillan (1975, 1976), Wheaton (1975), Holtmann *et al.* (1976), Kiesling (1976), Rothenberg (1976), Sandler, Cauley (1976), Greenberg (1978), Sandler (1978 b).

general, this literature derives from welfare and urban economists who have acquired some interest in location, rather than from location theorists who have become interested in the theory of public goods. This literature, then, tends to be at some remove from what I have in mind in proposing a theory of the public space economy.

4. Pure and impure public goods

The theory of public goods derives largely from work in the 1950's by Musgrave (1959) and Samuelson (1954, 1955). Although a great deal of attention has been given to this theory in recent years in the literatures of economics and public finance, relatively little work has been directed to relaxing the pure or polar nature of the goods under scrutiny (*). Public goods, according to conventional useage, are not defined as those goods provided by governments but rather by the attributes of the goods themselves. Three attributes seem to be fairly widely recognized:

1. *Non-Exclusion:* Once the good has been produced or made available to one person it cannot be withheld from anyone wishing to consume it. For pure public goods it is impossible or infeasible to exclude potential users.
2. *Joint Supply:* Once the good has been made available, equal quantities of identical quality services are made available to any number of additional people at no additional costs. (There are zero marginal costs over increments in the number of consumers or amount consumed, but not necessarily over the amount produced).
3. *Non-Rejectability:* Once supplied the good must be fully and equally consumed by all. Self exclusion is not feasible or is not economic. (Neither this nor either of the above implies that the utility derived from equal consumption is equal).

Pure public goods probably do not exist (although national defence and the legal system are often cited). The concept of a (pure) public good was put forward as a polar case; unfortunately most subsequent work on the theory has stuck to this polar - and spatially irrelevant - case. There are also very few pure private goods (exclusive, rejectable goods with no externality effects) although these are the subject of most economic theory. Thus most goods lie somewhere in the three-dimensional continuum defined by these three attributes. Most goods are «non-private goods» or «impure public goods» which are not equally available to, or consumed by, all consumers, have some congestion effects, are excludable at some non-trivial expenditure of resources, or are partially rejectable. One of the most important characteristics of non private goods is that when someone consumes them the amount left available for others is either not diminished at all

(*) The books by Buchanan (1968) and Head (1974) exemplify the preoccupation with pure public goods.

(pure public goods) or is not diminished by the full amount consumed (impure public goods). Thus they have the interesting and important property (which must be captured in models) that the amount consumed may not bear a direct relationship at all to the amount produced (provided or made available); in fact, they should generally be measured in quite different units.

In some sense the theory of public goods is a theory of market failure. The closer a good is to the public good pole, the less likely it is that private firms will provide the good at all. (Thus one does not observe private firms «providing» national defence although they may well «produce» components under contract). If a good is not provided by private firms, but is desiderable, chances are that a reasonably efficient collective provision process (government or the State) could provide them at levels which would cause a net welfare gain to society (a Pareto improvement). This is a most compelling rationale for the existence of government (*). For essentially the same reasons governments can generate welfare gains by mitigating (e.g., internalizing) the externalities that inevitably arise when private market processes produce and distribute non-private goods. The principal reason for non-provision by private firms relates to the non-exclusion attribute: if firms are not able to effectively exclude consumers, they cannot extract a price and will collect no revenues. Private firms may well produce and distribute goods which are fairly close to the private good pole but they will tend to produce them at suboptimal levels and sell them at too high a price. Monopolies and public utilities often purvey such goods. Indivisibility, large fixed costs, increasing returns to scale and «jointness» in consumption (all related concepts) underpin many (quasi) public goods. This is a celebrated problem in public economics and the principal motivation for the marginal cost pricing debate. If firms were forced to charge marginal cost prices for their outputs, in order to be socially efficient in the short run, they would lose money in the long run because marginal costs are necessarily less than average costs.

There is scope under these conditions for public policy to either attempt to adjust private behaviour (regulation, price control, standards, etc.) or to intervene directly in the provision process by undertaking distribution (through public facilities) and perhaps production, or both. The attribute of jointness in consumption alone (even where exclusion and rejection are feasible and economic) poses significant problems. If the marginal cost of serving an additional consumer is effectively zero, the efficient price for short run allocation is zero, and again the

(*) Most of the literature in the new welfare economic tradition is based on the assumption that the principal function of governments, apart from redistributing income, is to provide (impure) public goods and services which are more efficiently provided collectively than privately. A few of the works (largely in the tradition of the public choice approach) which make the case most explicitly and eloquently are: Arrow (1969), Rothenberg (1970), Tullock (1970), Olson (1971), Bird, Hartle (1972), Musgrave, Musgrave (1973), Breton (1974), Head (1974), and Buchanan (1975).

distributing agency is left without revenue. The government must intervene in some way in order to move the economy toward the Paretian frontier. To some extent the normative theory of public goods is a theory about how collective provision institutions (governments) should behave. Note, however, that actual *collective* provision does not guarantee that any welfare improvements will be made; we can only say that a potential improvement exists. Associated with the theory of public goods must be a related theory of optimal collective decision-making institutions. Some properties of such institutions are, however, suggested by the theory of public goods.

Most of the goods which are provided by western governments have non-polar positions on more than one of the three publicness dimensions. Models must be able to deal with these goods and services. Urban parks, for example, are relatively non-exclusive and are clearly joint or non-congestible only until some capacity is reached. Most roads, mosquito abatement schemes, and civil defence (air raid warning) systems are also in this class. However, whereas the parks are quite rejectable, civil defence warning signals are not. For goods provided by or at discrete public facilities (for example, schools, hospitals, libraries, museums, airports, fire fighting services, ambulances) both exclusion and rejection seem to be quite feasible and economic. In these cases jointness in consumption becomes the principal publicness attribute. Further, the positive or negative production and/or consumption externalities associated with these goods should also be taken into account in devising efficient and equitable production and/or distribution systems. The concept of externality is very close to the concept of a public good; indeed a pure public good (bad) can be considered as a polar form of externality in that everyone must consume it equally.

The attributes of pure and impure public goods and services which pose problems for private providers also pose difficult problems for collective institutions in pursuit of efficiency. In the absence of exclusion, rational self-interested individuals will not voluntarily pay for a good if they expect that others will finance the good (the free rider problem). The related preference revelation problem is that individuals will tend to «under-reveal» their true preferences for public goods if they perceive that their tax burden will be associated with their preference statement. However, the theory of public finance states that benefit-related taxes must be used if the first-order conditions for optimal output levels are to be satisfied. The conflict is obvious.

The government «solution» to the public good problem is to directly provide the good (at one level) and to extract coercive taxes to pay for it. Demand is assessed by subsidiary political processes and mechanisms which are unrelated to price. The possibility that government solutions of this nature converge to optimal equilibria is clearly remote. Recently several «planning solutions» to the non-exclusive public goods provision problem have been devised (*). These methods appear to have the desirable attribute of generating optimal public decisions with a fairly low level of information even when participants are aggressively self-interested. It is conceivable that such demand-revealing processes could actually be put into effect for some types of public goods and public decisions.

The principal first order condition for optimality in (pure) public good provision [often called the Samuelson condition after Samuelson (1954)] is that the sum of the marginal rates of substitution over all consumer-citizens be equal to the marginal cost of provision: $\sum_i \mathrm{MRS}^i_{YX} = \mathrm{MC}_{YX}$, where Y is the public good and X is the numeraire private good and i indexes individuals. Thus the payment of marginal benefit prices ($P^i = \mathrm{MRS}^i$), abstracting from the free rider and preference revelation problems, is one solution which would satisfy the condition. This is known as the Lindahl solution which arises in a pure voluntary exchange setting (**). Whereas in private good markets in equilibrium everyone faces a single price and adjusts the quantity consumed, here everyone is forced to consume the same quantity of the public good and, to be in equilibrium, everyone must face a different effective price. Of course other solutions can be devised such that the sum of the prices paid equals the marginal cost but these have different implications for the redistribution of inframarginal surplus.

Conditions for the efficient provision of impure public goods seem to depend on the circumstances and the author (***). There seems to be little agreement on these conditions partly because slightly different goods seem to be involved and because of the poor development of the theory of impure public goods. It is perhaps needless to say that with the early development of the spatial theory of impure goods the conditions will likely be even more diverse. Only an extremely small portion of the extant theory of public goods is explicitly spatial in any sense; much work remains to be done.

(*) The principal demand revealing processes are presented by Clarke (1971, 1972), Drèze, de la Vallée Poussin (1970), Malinvaud (1971, 1972), Milleron (1972), Roberts (1976), Green *et al.* (1976, 1978), Tideman, Tullock (1976), Groves, Ledyard (1977), and Tideman (1977). This last entry is, in fact, a special volume of *Public Choice* devoted to demand revealing schemes. Other methods of assessing demand are represented by the following works: Bohm (1972), Hori (1975), Strauss, Hughes (1976), Bradford, Hildebrandt (1977) and Maital (1979).

(**) See Lindahl (1919, 1958). See Head (1974) for a modern interpretation.

(***) For a fair sampling of the variety of slightly different efficiency conditions for impure public goods, often for slightly different problems, the reader may peruse the following: Buchanan, Stubblebine (1962), Mishan (1969), Evans (1970, 1971), Ng (1971, 1973), Meyer (1971), Oakland (1972), Oates (1972), Ellickson (1973), Kamien *et al.* (1973), Winch (1973), Baumol, Oates (1975), Sandler (1975), Lancaster (1976), Freeman, Haveman (1977), De Serpa (1977, 1978), Muzondo (1978) and Weymark (1979).

Most of the research in the normative public good literature has been directed toward deriving necessary and sufficient conditions for efficient output levels of public goods under a variety of assumptions. In the literature on impure goods the focus has been mainly on various forms of congestion and ways in which it can be treated. It is perhaps surprising that much less work has been done on suggesting institutional frameworks and mechanisms wherein these conditions are likely to be met. In fact, to some considerable extent, the details of the (first best) efficiency conditions themselves depend on the institutional context. This provides an important rationale for devoting considerable resources to the development of a larger theory which includes the (new) theory of public facility location. We turn now to sketch the questions which would have to be addressed by such a larger theory.

5. Needed: a general theory of the public space economy

The theory of public goods provides one broad but rather polar base for the development of a theory of the public space economy (*). The existing body of rather pure theory yields only reluctantly to spatial elaboration and generalization. This is not surprising when one considers that, even without a spatial dimension, the theory is rather complex and esoteric. Nevertheless some preliminary work by the present author indicates that considerable progress can be made in the development of operational planning models related to the location and allocation of public investment. One problem encountered is that much of the public goods literature is cast as positive science (yielding empirically testable propositions) rather than as explicitly normative theory; often in fact it is difficult to separate what is normative from what is positive(**). This may be a salient problem in the new welfare economic and public choice paradigms; however it is also a much more pervasive conceptual problem relating to public sector issues in general.

It has been suggested above that in order to more fully appreciate the context of the locational problem it should be situated in a broad conceptual framework of logically higher order and related problems. The following twelve questions are thought to be among the most central ones to be addressed by a (full-blown) theory of the public

(*) It should be stressed that there are many other conceivable paradigms on which to construct a theory of the public space economy. For example, much work has been done on the anarchistic, Marxian, and other critical theories of the State. Most of this work, however, is aspatial and has been proposed more as a critique of the role and functions of the contemporary western capitalistic state than as a concrete agenda for rationalizing the space economy even under some radically different political institutions.

(**) See the discussions in Buchanan (1968) and Mueller (1979).

space economy (*). As the focus here is on explicitly normative theory, the questions are all phrased «normatively»; positive, or descriptive, versions can readily be derived. A positive theory of the public space economy is also extremely important even if the principal focus is normative theory. Without the former, the latter could easily become utopian and irrelevant.

Key questions to address in a theory of the public space economy

1. What goods and services should be provided by governments?
2. What should be the size(s) (areal extent, population, etc.) and spatial arrangement (shape, nesting, etc.) of political jurisdictions or «provision regions»?
3. What mix of goods should be provided within each jurisdiction or region (or what should be the «mapping» of goods to jurisdictions or levels of jurisdictions)?
4. Under what conditions should governments directly distribute (non-retradable) goods and services (vs. making them available) and what rules should be used in distribution?
5. What levels of output or public provision of various goods and services should there be (in each jurisdiction)?
6. What types and levels of taxation and user charges should there be and how should these be organized?
7. What types and levels of interjurisdictional transfers should there be and how should these be organized?
8. How should the physical provision systems be organized? For example, should some systems be hierarchically structured? Should some goods and services share certain facilities in common, etc.?
9. *Where should public facilities, infrastructure, or resources be located?* (This location question should be distinguished from the allocation questions addressed in numbers 3 and 5 above).
10. How should the public and private sectors be interfaced?
11. What institutions and mechanisms (political, collective, etc.) will achieve the most desirable «solutions» as answers to the above questions?
12. How can we move from current institutions and solutions to those called for in number 11?

(*) These questions clearly pertain most directly to a western liberal capitalistic democracy. Thus they take certain values as given at the outset; these will not be further addressed here. Also, it should be emphasized that these twelve questions are not necessarily the complete set which must be embraced. Rather they should be treated as representatives of the kinds of questions I have in mind.

It is clear that almost all of the questions are interrelated and can not really be posed, much less effectively answered, in isolation. The answer to many of the questions significantly affects the way the location problem is conceptualized and modelled. Some of the more important interactions will be briefly explored below. First, however, it is expedient to address a more fundamental question that surely must be asked: should the location theorist, analyst, or problem solver attempt to embrace a larger set of decision, control, or policy variables than the conventional location, allocation, capacity, etc.? If a case can be made for going beyond these, where should one stop? This is related to the general problem of how to bound or close a theory, model, or system. The answer is certainly not straightforward and depends on one's theory of planning and ultimately on ideology.

The conventional approach in the location-allocation modelling of public facility systems tends to focus only on question 9. There is nothing *intrinsically* wrong with this. What should accompany a focus on question 9 is a *list of assumptions* made about the answers to the other 11 questions. These assumptions should be reasonable (if the policy output is to have any meaning) and the model constructed should be tied directly to the assumptions made. If one examines the vast literature on public facility models, it is extremely rare to see more than one or two assumptions stated which relate to the questions posed above or to similar questions. Further, these are seldom defended as reasonable and often they even seem to be ignored. It is usually difficult even to discover what assumptions seem to be «implicit». Without a discussion of the assumptions made, it is very difficult to fathom the theory which may implicitly underpin the models. The theoretical preambles or asides that are often included in the public facility location literature are typically extremely narrowly based (myopic) and tend to beg more questions than they answer. This observation tends to support my contention that, in fact, there really is no theory of public facility location beyond what might be considered to be primitive, eclectic, heuristic, and *ad hoc* statements or precepts.

The argument could be made that locational modellers are typically hired by decision-makers or agencies of government to solve specific locational problems within contexts or settings which, for all intents and purposes, may be considered given and immutable. Under these conditions, the modeller need not be concerned about (the optimality of) the «environment» of the locational problem. However, the environment should always be taken into account if the model is to be relevant. Even if the environment is not assumed to be optimal, some assumptions must be made about it, and, when these are stated, a theory supporting the model should be set out. If we take the goal of planning to be the seeking of social welfare improvements (however defined) it seems unwise to rigorously optimize a small and rather artificially bounded (by the adjective «locational») problem. Planners and applied social scientists are seldom asked for *one solution* to a problem. Several different solutions, based on different assumptions, may be set out. If a change can be made in a «non-locational» policy which would lead to an improved locational policy, surely this change should be identified, perhaps even advocated.

As an example of this problem consider the provision of fire fighting facilities in the Minneapolis-St. Paul Metropolitan Area. In this area fire services are provided by each local municipality of which there are over 200. Some of these municipalities are as small as a large city block. If a location analyst were asked to recommend an optimal locational pattern for fire fighting facilities in one of these municipalities, one approach would be to use of the standard (e.g., p-median type) models. However, it seems clear that the outcome would be only a local optimum or a second-best solution. The astute analyst would (perhaps in addition) investigate the possibility of contracting services from adjacent municipalities and the possibility of rationalizing the facility systems of all municipalities as one integrated Metropolitan system (*). In the Metropolitan Toronto area fire fighting units are also provided by the constituent municipalities and, although these are much larger than those in Minneapolis-St. Paul, the distribution of facilities illustrates many of the inefficiencies which are not really solved by local relocation decisions. The police services in Metropolitan Toronto however, are provided at the Metropolitan level. This was a response to strong arguments relating to the inefficiency of existing interjurisdictional spillovers and the existence of potential economies of scale. The location of police facilities has been largely rationalized and is now relatively efficient and uncontroversial.

A further case can be made for considering a broader set of issues (such as those set out above). It is entirely possible that a conceptual framework and model addressing only those variables unequivocally associated with location will lead to a proposed solution which demonstrably reduces social welfare or is Pareto inferior to the current system. This conclusion is analogous to the general theory of second best in welfare economics (**). Consider further the Minneapolis-St. Paul fire station problem. A locational model may suggest the addition of a third facility for a small elongated municipality. Although the cost of the new facility may be more than offset by the benefits to the residents of the municipality alone, it might be easy to show that, for the Metropolitan area (or even an area just slightly larger than the municipality), there is a net welfare loss. Further, the construction of a new facility may make it more difficult to rationalize the whole system and its very construction may well mitigate against the adoption of any efficiency-seeking centralization proposals in the future.

(*) For interesting discussions of contracting for services see Ostrom *et al.* (1961), Warren (1964), Friesma (1970), and Ahlbrandt (1973a, 1973b). The possibility of organizing fire services at the Metropolitan level is well within the realm of possibility because a Metropolitan level of government for the provision of other services is already in existence.

(**) This theory is described by Lipsey, Lancaster (1956). Its relevance in a public goods context is discussed by Sandler (1978) and applied by Rosekamp (1980).

6. Macro-level considerations impinging on micro-level location problems

We turn now to examine briefly why answers to the questions to be addressed by a theory of the public space economy are important and impinge on each other and especially on locational problems and models.

The answer to the first question of what goods should be provided by governments, at least within the paradigm of modern welfare theory, is that governments should provide pure and impure public goods and should otherwise attempt to attenuate externalities. This proposition becomes problematical for those goods that could be (or are) provided privately or publicly but with necessary (institutional) levels of inefficiency in each sector, or with different implications for the distribution of income. In these cases rather difficult case-specific cost-benefit analyses would have to be undertaken. The solution to this problem would take us far beyond the scope of this paper (*). If the proposition that governments should be in the business of providing (only) these goods is accepted (as seems to be reasonable after some consideration) then the attributes of these goods ought to be taken into consideration explicitly in all models of public sector systems, including *a fortiori* those of locational systems.

The second question relates to the optimal size and arrangement of jurisdictions. Jurisdictions serve as the basis not only for the provision of public goods and services but also for demand articulation and political decision-making (**). In determining the optimal size and arrangement we must first have a complete list of the goods and services to be provided; this list depends on the outcome of question 1. We must then systematically take into account all of the variables relating to size and arrangement. For exemple, the costs of production, tax collection, and travel or transportation (all as a function of jurisdiction size, population, and/or areal extent) must be determined. However already we observe important dependencies. For

(*) This problem is addressed by Head, Shoup (1969, 1973).

(**) Many authors have approached the question of optimal jurisdictions and optimal jurisdictional systems in many different ways. Often especial consideration is given to fiscal federal systems because they seem to have many desirable properties for resolving spillovers, conflicts, etc. Just a small sampling of this literature follows: Ylvisaker (1959), Stigler (1962), Hirsch (1964), Musgrave (1969, 1971), Ostrom (1969), Tullock (1969), Breton (1970), Rothenberg (1970), Koleda (1971), Young (1971, 1976), Bird, Hartle (1972), Oates (1972, 1977a, 1977b), Evans (1973), Head (1973), Musgrave, Musgrave (1973), Buchanan (1974), McGuire (1974), Cox, Dear (1975), Kiesling (1976), McMillan (1976), Sandler, Shelton (1976), Breton, Scott (1977, 1978), Ellickson (1977), Silver (1977). The work of Tullock (1969) seems to be particularly incisive.

example, if goods and services are optimally contracted out (question 4) then we need not be directly concerned with production costs (and perhaps not with location either). Further the technology to be used and the structural and locational organization of the system must be known for each jurisdiction size. The taxation system must be determined for one to assess its costs. The location related problems (questions 8 and 9) must be solved for each size of jurisdiction in order to assess transportation or travel costs. Discovering how benefits or consumer surplus varies with jurisdiction size seems to be especially problematical.

The costs of political decision-making, and the bureaucracy of implementation, would be expected to vary systematically with jurisdiction size (*) (but assumptions must be made about the type of political system involved). The costs and benefits to individuals, relating to their attempts to articulate demand, tend to be such that net benefits are greatest in small jurisdictions. In larger jurisdictions it often becomes rational to spend little or no resources in articulating preferences. Political externality costs tend to increase with jurisdiction size because, in general, increased size tends to imply an increased variance in preferences so that the provision of a single level of good alienates more people more adversely (in the absence of lump sum benefit taxation) (**). Another important consideration, much neglected by public location models, is the cost of the interjurisdictional spillover effects which will almost always exist. In general «spillunder» effects are not a problem because of the possibility of having a multifacility system in a single jurisdiction. Non-internalized spillovers lead to suboptimal provision levels; the answer to question 7 (types and levels of interjurisdictional transfers) must be known to assess the welfare losses from spillover effects. One would expect that in general, and for obvious reasons, the spillover costs would decrease as jurisdiction size increases.

A rather more simple approach to the problem of optimal jurisdiction size now has a fairly large literature in the «theory of clubs» (***).

(*) Various theories of bureaucracy such as those by Tullock (1965), Downs (1967) and Niskanen (1971) (all of which are broadly similar) would have to be consulted.

(**) This and similar propositions have been proven by Barzel (1969), Oates (1972) and Bish (1971) others and are discussed simply by Bish (1971). Political externalities are those associated with single levels of public output which people with different strenghts of preferences face in common.

(***) The rapidly growing theory of clubs is nicely captured in the following set of papers: Buchanan (1965), Pauly (1967, 1970a, 1970b), McGuire (1972, 1974), Tollison (1972), Oakland (1972), Barr (1972), Musgrave, Musgrave (1973), Ng (1973, 1978), Polinsky (1973), Ellickson (1973), Ng, Tollison (1974), Roberts (1974), Chamberlin (1974), Allen *et al.* (1974), Fisch (1975, 1976), Berglas (1976a, 1976b), Lancaster (1976), Helpman, Hillman (1976, 1977), Hillman (1977), Adams, Royer (1977), Stiglitz (1977), De Serpa (1977, 1978), Topham (1977), Henderson (1979), Boadway (1980), and Brennan, Flowers (1980). The simple diagrammatic exposition of this theory by Allen *et al.* (1974) is an excellent introduction.

Following Buchanan (1965) almost all the club theoretic literature has abstracted from space and from the problem of non-exclusion; thus the theory pertains to exclusive (but partially joint) goods in an aspatial environment. In most articulations of this theory there are two decision variables - optimal public output in a club or jurisdiction and the size of its population. The good being shared by the club members is assumed to be only partially joint so that as the number of users increases, either the cost per user increases or the average benefit per unit of consumption decreases. The optimal club size and output level are determined by the trade-off between the economies achieved with a larger group shaving the large fixed costs, and the increases in congestion associated with a larger consuming group.

This simple model has been very attractive in capturing several of the most important dimensions of the problem of optimal collective consumption. Recent literature has relaxed some of the initial simplifying assumptions in interesting ways which have provided new insights into the nature of the problem. For example, different tastes and preferences for the club good are now accomodated, as are a variety of interesting cost sharing rules or taxation schemes, and forms of congestion. This literature has also provided some interesting and instructive differences of opinion. Berglas (1976a), for example, believes that private firms can provide club goods perfectly efficiently. Boadway (1980) has further examined this proposition. Ng (1973) claims that Buchanan's conditions are not those of an optimal equilibrium because they maximize average rather than total net benefits. Berglas (1976a) holds that Ng's new conditions are inappropriate. Helpman and Hillman (1977) attempt to demonstrate that the problem addressed by Berglas and Buchanan on the one hand, and by Ng on the other, are rather different. (Ng (1978) disagrees). Brennan and Flowers (1980) have recently clarified the key issues of the debate. The debate turns on whether the problem pertains to a single club or a system of clubs, a set of joiners or the whole population. That such a debate could take place amongst eminent welfare economists is illustrative of the subtleties involved in even the simplest of models in this area. If a lesson in not learned from this debate, it will very likely be repeated in the context of public facility location problems.

This has been a very cursory overview of some of the issues involved in resolving the issue of jurisdiction size. Note that these issues would have to be addressed for a whole set of shapes and spatial arrangements of jurisdictions and hierarchical and other nested structures. A hierarchical structuring of jurisdictions, as in a fiscal federal sistem, will certainly influence the shape of many of the cost and benefit curves. For example, it is much more likely that interjurisdictional spillovers would be approximately internalized in a federal system (in which higher order governments may impose rules and policies on lower tiered governments) than in an uncooperative small-n-person bargaining game. Still, however, the problem is not really captured without simultaneous consideration of question 3.

Question 3 pertains to the optimal mapping of goods into jurisdictions or the optimal mix of goods in each (level of) jurisdiction (*). Optimal jurisdictional «scope» is a complex aggregation problem which could be approached conceptually as follows. Assume we have been able to order the goods in increasing size of the optimal single good jurisdictions. Select a set of discrete sized jurisdictions to approximate a continuum. It seems reasonable to constrain the aggregation to similar sized regions. With n goods and m jurisdictions, the number of combinations which would have to be assessed is easily calculated if there are no constraints; for hierarchically structured (nested) systems, assessing the number of combinations alone becomes a problem.

A further complication must also be considered. The various costs and benefits, defined above with respect to jurisdiction size, are clearly affected by the level of aggregation. A few examples will suffice. There may be savings in the production costs when two or more complementary goods are provided by the same government. As the «scope» of a jurisdiction increases, individuals will tend to find their votes and other means of preference revelation have much less meaning and «political externalities» will tend to increase; however at the same time participation costs may decrease as there would tend to be fewer governments with which to interact. Decision-making costs would likely have a U-shape when plotted against scope, holding the number of jurisdictions constant.

It should be clear that these sets of related problems are rather simple to conceptualize separately, are difficult to interface, and are extremely intractable even if one were able to estimate all the necessary cost and benefit curves. To the extent that problems of location tend to presuppose these difficult issues have been satisfactorily resolved, it seems necessary to attack them, and attempt to construct theories of them, or failing this, at least to problematize them.

The answer to question 4 requires conditions under which it would be optimal for a government to contract with some other agency - perhaps some other (level of) government - to provide a good or service. This question has been addressed in a small body of interesting literature but remains to be satisfactorily answered (**). It is sufficient to say here that if there are significant scale economies in production, and if the distribution costs are not significantly affected by having production outside the jurisdiction (television transmission, etc.), then very often a strong case can be made that a small jurisdiction should contract out the good. The possibility is thus open for trade in public goods (***). Further it

(*) This problem has been conceptualized by Breton, Scott (1977, 1978).

(**) For interesting discussions of contracting for services see Ostrom *et al.* (1961), Warren (1964), Friesema (1970) and Ahlbrandt (1973a, 1973b).

(***) Breton (1970), Connolly (1972, 1976), James (1974), Kiesling (1974) and others have discussed trade in (excludable) public goods.

may be argued that if the essential public good problems are associated primarily with distribution (vs. production), contracting with private firms should be considered. Of course, if the service is purchased on contract, the public facility location problem changes dramatically and exists only at a higher level of abstraction.

Question 5 relates to the optimal level of provision of public goods. First, robust efficiency conditions for impure goods must be derived. The conditions must necessarily involve summation over some relevant population and so this relevant population (usually, but not necessarily, the jurisdiction) must be clearly defined. The cost function for production or at least provision must be known, and a whole set of technological (e.g., congestion) and sociological (e.g., reactions to crowding, etc.) variables will have to be measured. The conditions must be capable of embracing the full range of different types and degrees of impure public goods. (It seems likely that different sets of conditions will be required for different types of goods). Perhaps the most difficult problem will be deriving conditions which are directly related to different types of political institutions. For example, Tollison and Willett (1978) have recently proposed a fiscal federal voting system in which individual votes are weighted such that they taper off with «distance» from the issue (facility, policy impact area, etc.). The optimality conditions for such a system will be rather different from those associated with conventional one-man-one-vote systems (*).

Question 6 pertains to optimal taxation and user charge policies. The theory here seems to be relatively well developed (although there is no general agreement) and can be succinctly summarized. Individuals should be assessed charges equal to the marginal costs they impose on the system. If exclusion is not feasible or economic then these charges are not possible. To the extent the marginal cost charges (e.g., congestion tolls) are not sufficient to make up the costs of the system (**) it is optimal to levy lump sum marginal benefit taxes to make up the difference. Actually all that is required is that the marginal tax price be equal to the marginal benefit for each individual for there to be a Lindahl equilibrium; this leaves open the possibility that inframarginal tax prices could be varied to redistribute income (***). There are many difficulties here. First, one must have a demand revelation mechanism

(*) For discussions of equilibrium conditions for traditional majority voting schemes see Slutsky (1977), Westhoff (1977), Flowers (1978), Brueckner (1979a) and references cited in these works.

(**) For interesting analyses of the conditions under which marginal cost congestion prices yield sufficient revenues to support the facilities, see Oakland (1972) and Muzondo (1978).

(***) The general issues are summarized by McGuire, Aaron (1969), Samuelson (1969) and Head (1974).

to be able to correctly assess marginal benefits. (In this area considerable progress has been made recently, as noted above). Second, lump sum taxation is generally held to be non-operational. Third, the issue of distribution of income is extremely problematical. Thus some other form of taxation may in fact be more efficient. This may depend on answers to many of the other questions including, for example, the question of the optimal sizes and scopes of jurisdictions.

Existing public facility locational models seem to side-step a whole set of essential issues in their omission of revenue generation schemes and their incidence. If the goal is to maximize social *net* benefits, then the omission of incidence of the costs requires very strong assumptions about the additivity of utilities and surpluses, assumptions which are probably repugnant to many of the modellers themselves. One can not even address the issue of equilibrium of the consumers of public goods (with its attendant implications for demand articulation, migration, political action, demoralization, social unrest, etc.), without considering tax incidence. It is ironic that considerable pains are often taken to include issues of equity in public facility location models (using average distance travelled by various subgroups, for example) without any discussion whatsoever of who is paying for the facilities (*). Clearly, even without any particular concern for equity, some way must be found for interfacing the problems of optimal taxation and optimal public good levels with the problems of location and allocation. The very definition of optimal location depends fundamentally on the tax institutions assumed and these clearly vary (empirically and optimally) in time and space.

Question 7 concerns optimal interjurisdictional transfers for the internalization of externalities and is intimately related to the previous question of taxation. If jurisdictions must have boundaries (**), then it seems that there will be interjurisdictional spillovers associated with almost all goods which cannot be made completely exclusive (***). The «local public goods», so frequently used in the literature (defined as those public goods which are uniformly «available» up to the edge of jurisdictions and then suddenly vanish) will be extremely rare unless there is exclusion at the boundary. This is especially true of goods for which consumers must travel. Olson (1969) has articulated a «principle of fiscal equivalence» which seems to be a widely accepted policy

(*) See, for example, Morrill and Symons (1977).

(**) Gale (1976) and Gale, Atkinson (1979) have proposed a solution to the general interjurisdictional externality problems inspired by the theory of fuzzy sets. Essentially individuals are allowed to choose different stakes in various issues of interest which are controlled by other jurisdictions (cf. Tollison, Willett, 1978).

(***) Casual observation and many studies indicate that the welfare losses caused by non-internalized externalities are likely to be substantial. For example, see Weisbrod (1965), Holtmann *et al.* (1976), Dear *et al.* (1977), Mehay (1977) and Greene (1977). However, in most cases, prevention of benefit spillovers will be a less efficient strategy than compensation for them.

prescription relating to spillovers. One essential implication of the principle is that jurisdictions should be structured so that those who benefit are taxed. The problem is that most taxes are confined to those within jurisdictions whereas benefits are not. (There may also be «tax exporting» when there are no benefit exports). Thus strict application of the principle is simply not possible in a spatial world.

When benefits fall off with distance to public facilities (for any reason) McMillan's (1975) model and further research by the present author (*) have demonstrated that three policy variables (in addition to location itself) must be simultaneously optimized. These variables are the sizes of the jurisdictions, the levels of provision, and interjurisdictional grants. It has been widely recognized that, if jurisdiction sizes are fixed, interjurisdictional grants are required and these should be «matching grants» (**). However it has certainly not been widely recognized that in the absence of simultaneous variation in the sizes of jurisdictions an optimal matching grants policy does not produce a globally optimal policy of output levels of the public good. The implications of this conclusion are clear and rather illustrative. One should take the full context of the problem into account.

It should also be noted that precisely because spillovers are externalities, and there is likely to be a small number of jurisdictions involved due to distance decay effects, voluntary bargaining is rather unlikely to give rise to an optimal equilibrium (even abstracting from the fixed jurisdiction sizes). For this reason it will be efficient to have a solution which is, to some extent, imposed by a higher level of government. (However, each lower level jurisdiction should have some input into the policy outcome). A fiscal federal hierarchical arrangement of jurisdictions has obvious advantages for resolution of spillover inefficiency problems so that this issue must be included in the problem of the optimal structuring and sizing of jurisdictions.

Interjurisdictional spillovers are externalities which must be internalized if they cannot be ruled out in jurisdiction formation. Whether or not transfers of particular types are proposed and how these are organized impinges on locational decisions. Without compensation for spillovers, a locally optimal strategy would be to never locate beneficial facilities, and to always locate noxious ones, near the boundaries of neighbouring jurisdictions (***). Appropriate compensatory mechanisms essentially allow, or force, one to consider the reverberations of locational decisions throughout the space economy.

(*) See Lea (1978, 1980).

(**) Matching grant policies are discussed by, among others, Breton (1965), Olson (1969), Connolly (1970), Hirsch (1970), Oates (1972), Vardy (1972), Musgrave, Musgrave (1973), Le Grand (1975), Harford (1977), Rittenoure, Pluta (1977), Sheshinski (1977), Jurion (1979a, 1979b) and Mieszkowski, Oakland (1979).

(***) That such locational outcomes are so widely observed seems to be fairly unequivocal evidence of inefficiency and likely inequity. This provides further support for the thesis that the location problem should be addressed in a wider context.

It should be clear that the questions examined, even superficially, to this point are highly correlated with the construction of truly relevant locational models which are consistent with an overall theory of the public space economy. We address questions 8 and 9 relating to micro level locational considerations in the next section and forgo a discussion of the other higher level questions 10, 11, and 12 entirely.

7. Public facility location-allocation models within a theory of the public space economy

The stage has been set to deal with the question of appropriate location-allocation models (questions 8 and 9), at least in a general way, rather succinctly. First the important distinction is made between those goods which are delivered to consumers (delivered goods) and those for which consumers must travel (travelled-for goods). It is thought that this distinction is especially important. Then the attributes of impure public goods in spatial contexts will be discussed. In particular, we will see that it is especially useful to generalize the concept of the jointness of a public good to that of the spatial jointness of an impure good.

Delivered goods include all those goods which are delivered to consumers at their homes, etc., in a conventional sense (sewer lines, mail, fire, ambulance and police services, etc.); extra costs are incurred in the delivery process. They also include other goods which require no particular delivery effort such as mosquito abatement, radio signals, civil defence warning signals, and the like. In addition, they include goods which may not be consumed at all but for which an «option demand» for possible future consumption exists; examples include remote national parks, museums, and subway systems (*). Travelled-for goods comprise all those goods for which consumers must travel, for example, schools, hospitals, libraries, museums, parks, swimming pools. Note that, at least in this case, it is much easier to describe the facilities than the services they provide. Consumption generally takes place at the destination facility and consumers usually, but not necessarily, bear the transportation costs. The distinction between these two kinds of goods is important because of the difference in source and nature of the «impurities» involved and the fact that the differences must be considered in model construction.

Delivered goods *may* be produced/delivered in such a way that they are «equally available» over large territories or large groups. Very often, however, the amount and/or the quality of services received declines with distance (as, for example, a radio signal). Holding locations of facilities and individuals constant, individuals frequently have little or no choice in the level of service received. No private expense need be incurred in consumption of these goods except where exclusion is

(*) That some goods yield benefits without being directly consumed is of considerable importance. See Weisbrod (1964).

feasible and user charges are optimal. For travelled-for goods, it is not possible to have, or conceive of, a good which is equally available in the sense of yielding the same potential benefit even to a small spatially dispersed group (except perhaps when transportation is provided freely and takes little time). There can be no such thing as a «locally public» travelled-for good. Consumers who must pay private costs of public consumption will tend to reject the full amount of the good made available to them. Whereas for delivered goods it is possible to speak of an objective amount of provision «available» at any given location, this is not meaningful for travelled-for goods except in a rarified atmosphere of identical preferences, incomes, etc. Individuals simply evaluate the friction of distance differently and this must be directly modelled. Note, however, that we can still speak of «equal availability» in some sense - in fact, it is pervasive. Abstracting from exclusion, everyone has equal «access» to the facilities. *Once at the facilities* everyone faces the same public good supply. There are two ways in which one can handle less than full consumption. An individual can be assumed to consume a joint good - «the good itself as well as travel to it» - in which case individuals are assumed to discount for distance. Or the travel cost (and/or time) can be entered into the individual's budget constraint as he attemps to maximize utility. Bigman, ReVelle (1978) have constructed a model in which a travelled-for good is treated as a *pure public good*. For a critique of this and other fundamental errors in their model see Lea (1979a, 1979b).

In general the existing theory of impure goods and the related theory of clubs are based on a fairly simple conception of the source of the impurity. Apart from the relatively minor attention paid to differences in institutional excludability, all impurities seem to be attributed to congestion or crowding. It is quite remarkable that almost no attention has been given to the impurities due to the fixity of locations and the frictional effects of space. A spatial theory must embrace both congestion and the several different important effects of space.

The exclusion dimension of impurity need receive no special attention in a spatial theory. It can remain «institutional exclusion» and we can solve for the optimal degree of exclusion as a decision variable in any model (*). There is little doubt that greater distances tend to imply greater *de facto* exclusion but this phenomenon is best treated as a generalization of jointness (**). Normally rejection is deemed relevant only in the case of (impure) public bads. In cases in which consumers bear private cost burdens (the travelled-for case in particular) the concept of

(*) Although a good deal of the public good literature assumes that the degree of exclusiveness is a technical attribute of the good Goldin (1977) has summarized the strong case for considering exclusion as a *decision variable*. Also see the model by Kamien *et al.* (1973).

(**) In a spatial world in which all goods/facilities are necessarily provided in particular spatial patterns, expanding the meaning of exclusion to include reduced benefits due to space would mean that exclusion in always present. This would tend to cause unnecessary confusion.

rejection could certainly be broadened to include the reduction in consumption (and therefore benefits) caused by greater distances or travel costs. However, it is also possible to consider this phenomenon to be part of an expanded concept of spatial jointness.

Crudely a good will be called *spatially joint* if, in addition to being joint (in the conventional sense of non-rival consumption), it does not matter where the additional consumers are located. Additional consumers could have costless access to the same quality of services as current consumers, i.e., the service area of the good could be extended infinitely over space at zero cost. If costs increase, whether to the government or to the individual directly, then we have impure spatial jointness. It is clear that there are no real world instances of *pure* spatially joint public goods but that there are many instances of less than pure ones. Note that whereas the notion of extension of a jurisdiction or benefit area would normally be considered a long run phenomenon we are considering it here in a short run sense. The good «protection from mosquitos» is purely joint within a given sprayed area but this good is certainly not purely spatially joint because either more spray must be used to expand the protected area or else would-be mosquito haters from outside the area must incur greater expense to take advantage of the refuge.

There are many interesting and diverse bases for spatial jointness; only a few will be briefly explored here. It is sufficient to note that the bases tend to be different for delivered and travelled-for goods although the general phenomenon of congestion and crowding, usually stemming from fixed capacities, tends to be an important underpinning of both. For delivered goods it is common for the quality of service simply to physically decay with greater distances independent of use. In addition, there may be economies of scale in both production and delivery. Public agencies often have the possibility of adjusting the amount actually delivered to various locations by varying the delivery system and/or the delivery technology (for example, fire truck or ambulance response strategies may be altered); the service area may be extended without any additional inputs to production *per se*. In the case of travelled-for goods, economies of scale in production may also be extremely important. The spatial impurity here is usually underpinned by the costs of travel incurred by individuals. In this context we must be concerned about rigorously predicting consumers' choice of facility or facilities, number of trips made, and amount consumed per trip. It has been noted above that considerable progress seems to have been made recently in the area of more realistic demand and allocation models (*).

(*) See for example, the work by Coelho, Wilson (1976), Hodgson (1978), Leonardi (1978, 1980), Tapiero (1978) and Sheppard (1980).Also, Erlenkotter (1977) and Hansen, Thisse (1977) have shown how traditional cost-minimizing location-allocation models can be easily modified to cope with elastic demands and maximal net benefits.

It remains to incorporate the phenomenon of congestion into these models. It may be difficult to do this well, even disregarding problem of the tractability of the models (*).

For the purpose of constructing public good theoretic models, it is wise to distinguish between the level of the public good *produced* (available) at the facilities, the amount actually available at any given location, and the amount of the good actually consumed by an individual at any given location. The main decision variables will be location specific production levels, but this variable is clearly not the most important one for consumers. For delivered goods, the amount of good received will be less than or equal to the amount produced (because of decay). If the good is desirable and there are no user charges, then the amount consumed will tend to equal the amount «received». In the case of travelled-for goods, the amount «objectively available» to any individual in space could be considered to be identical with the amount produced. However, because of the privately borne transportation cost, the amount consumed will be less than (or equal to) the amount considered to be available. The pure public good outcome, prescribing consumption of the whole amount of the good available, is thus extremely unlikely.

The discussion above indicates that there will be significant problems of measurement in public goods location problems. In addition to the problem that the units of a good consumed need bear no relationship to the units of the good produced, it must be clearly recognized that it will usually be expedient to measure them in entirely different units. (Swimming pools are measured in square metres of swimming space while the good consumed may be man-hours of recreation). Costs will relate more closely to production units while benefits will relate more closely to consumption units. For meaningful theory and operational models, however, a way must be found of relating these to each other. This measurement (and to some extent conceptual) problem is not intrinsic to the public goods paradigm. Rather, this paradigm points to, and demands a solution to, this problem. Other *ad hoc* «theories» characterizing most of the public facility location literature in no way solve this problem; instead they are sufficiently superficial that the problem is simply ignored.

Public facility location models, to be developed within the theory of the public space economy, must be closely matched with the environment of the problem as has been noted. Many of the variables which should be decision variables for the comprehensive system model will necessarily have to be exogenous to the location problem *per se* in order to have tractable models. (For example, the mix of goods to be provided by jurisdictions will surely have to be given). Nevertheless, these models will have to include a larger number of decision variables than conventional models. Some of these stem directly from the «problematic» of impure goods. For example, it could be argued cogently that even for micro level problems of location (or, *especially*

(*) Leonardi (1980) has made some headway with this problem.

here), one should solve for optimal levels of exclusion, optimal levels of congestion, and optimal taxes and/or user charges. Other variables should be examined which are not particularly attributable to the theory of public goods. For example, I believe that the models should be capable of dealing with hierarchically structured systems of facilities (an analogy to fiscal federalism in some sense) and with the consideration that interactions frequently take place between the facilities themselves (*). In addition, there is another significant dimension to all locational problems which really should be addressed. Rather than locating each different type of facility system piecemeal and independently, considerable effort should be devoted to the development of models which simultaneously deal with different systems. Such models should be capable of dealing with intersystem interactions and also the very real possibility that the sharing of facilities and other infrastructure will be optimal in many contexts. The possibility of shared facility systems seems to be ruled out by most existing approaches. These are just a few examples of improvements which should be made in the location-allocation models so they become more appropriate characterizations of the real problems faced by public decision-makers.

8. Some technical considerations relating to the construction of operational location-allocation models for impure public goods

It would not be possible to present any exemplary or prototype «new» models without significantly extending the length of the paper. The scope of the present section has therefore been restricted to a discussion of a selection of important technical issues which must be addressed in the construction of operational location-allocation models. There have been a few contributions in the literature which proposed particular models that have several of the attributes which should be possessed by the new models I have in mind. Particular reference should be made to the conceptual frameworks and welfare theoretic models of Tiebout (1961), Smolensky *et al.* (1970), Wagner, Falkson (1975), Mc Millan (1975), Capozza (1976), Erlenkotter (1977), and Schuler, Holahan (1977). The works by McMillan, Schuler and Holahan are particularly consistent with the new theory of impure public goods in a spatial context, although these, like the others, are not, in my view, sufficiently general or comprehensive.

One of the more fundamental issues relates to the use of the concept of the «public». Welfare economics and the theory of public goods are based on a concept of «methodological individualism». Social welfare in this paradigm is based on the welfare of each of the constituent members of society. (The problem of aggregating the welfare of individuals is briefly taken up below). One of the key problems to be addressed is precisely how to measure, or proxy, individual welfare in a multivariate world. (This problem has received a good deal of

(*) White (1979) has discussed this latter problem.

attention). However, because of the goal of constructing operational models, it may be expedient to consider that the fundamental unit of analysis is the household (family) or perhaps even some larger group with well defined similar attributes (preferences, incomes, etc.). Although this may significantly reduce the number of variables, any initial aggregation will have to be based on a good theory.

The selection of the appropriate basic units should be based on a thorough study of the locational problem and its context. Such a study is also an absolutely necessary basis for selection of the decision variables. Clearly the decision variables will vary considerably with the type of problem involved, and to some extent with the terms of reference of the study, but the list suggested in the discussion of the foregoing sections may be long indeed. It is not necessary that all relevant instrument or policy variables be included in a single («locational») model, only that they all be rigorously taken into account. This, of course, may require a set of models which interact with each other in clearly specified ways (*). Particular attention must be given to the units in which the variables are measured. For example, the difference between production units and consumption units of the impure goods must be explicitly recognized. It has been suggested above that it would be desirable to have models which are capable of dealing with several (or many) different public good provision systems simultaneously, because of the strong interdependencies which characterize reality. In the short run, however, it is probably expedient to develop separate (sub)models for the delivered good and travelled-for good cases because these tend to involve quite significantly different considerations. It is clear that the process of selection of decision variables, and the process of structuring the actual models in which these are optimized must be informed by a powerful theory. Otherwise the exercise will be non-productive and perhaps counter-productive.

In addition to the «location of impure good capacities» (**) (measured in production units), it is thought that at least the following additional (decision) variables should generally be included (***):

- different types of impure goods;
- different types of production technologies and facility types;

(*) If there are a set of interacting submodels, considerable care must be taken to ensure their consistency. Some of the new methods of multi-level optimization may be used on this type of problem if the submodels are fairly simple. If the models are poorly structured and/or nonlinear, simulation methods will have to be given serious consideration.

(**) The locations may be considered to be in continuous space, on a network, or in a discrete or punctiform space. The latter seems to allow the most operational models.

(***) Of course, some of the variables need not be strict decision variables; this depends on how the model(s) is(are) structured. The point is that they should be included.

- different types of structural organizations for the facility systems;
- different possibilities regarding the sharing of infrastructure;
- different types/modes of delivery technology (delivered goods);
- the distance decay of delivered goods by mode;
- realistic demand generation (delivered and travelled-for goods);
- realistic facility selection submodels (especially for travelled-for goods);
- different forms of institutional exclusionary policies - including various forms of user charges, the passing of laws, the construction of physical barriers, etc.;
- different tax forms which are legal and feasible;
- congestion and crowding;
- the reaction of individuals to prices, taxes, congestion, etc..

Beyond these, two additional general dimensions of the impure good problem should be given consideration at the next level of generation. The first is in the direction of a general equilibrium model and the second takes us into the realm of political economy and public choice theory. In market economies, land values tend to reflect access to public goods. Land rent adjustments must be taken into account in long run models because these impinge significantly on the welfare of individuals (*). In addition, individuals may adjust to changes in the location of facilities in various ways (which also affect land rents) - most notably by migration to other locations and, in fact, other jurisdictions. A good deal of work has been done on the problem (or solution) of interjurisdictional migration following the seminal model of Tiebout (1956) (**). These should be consulted for ideas as to how the process of migration can be modelled along with public facility location.

The second additional dimension of the problem is much more difficult to model and will be noted only briefly. This involves

(*) A number of works in the reference list address the issue of land rent adjustments. These are: Neuberger (1971), Barr (1972), Boskin (1973), Lind (1973), Flatters *et al.* (1974), Schuler (1974), Stull (1974), Getz (1975), Sakashita (1975), Wheaton (1975), Fisch (1976, 1977), Hamilton (1976), Helpman *et al.* (1976), Kanemoto (1976), Le Roy (1976), Greenberg (1977, 1978), Helpman, Pines (1977), Henderson (1977a, 1977b), Morrison (1977), Richardson (1977), Stiglitz (1977), Wright (1977), Courant, Rubinfeld (1978), Miyao (1978), Papageorgiou (1978), Thrall, Casetti (1978), Wooders (1978), Brueckner (1979a, 1979b), Casetti, Thrall (1979), Ellickson (1979), Richter (1979), Rose-Ackerman (1979), Rufolo (1979), Thrall (1979), and Wildasin (1979).

(**) The works included in the reference list which build models including interjurisdictional migration are: Williams (1966), Buchanan, Goetz (1972), Vardy (1973), Flatters *et al.* (1974), McGuire (1974), Schuler (1974), Richter (1975), Sakashita (1975), Weaton (1975), Berglas (1976b), Ellickson (1977, 1979), Fisch (1977), Greenberg (1977, 1978), Greene (1977), Pestieau (1977, 1980), Topham (1977), Wright (1977), Miyao (1978), Sonstelie, Portney (1978), Wooders (1978, 1980), Premus (1979), Rose-Ackerman (1979), Homma, Yamada (1980), and Starrett (1980). Sakashita's work is the only one to attempt to address this problem at the level of specific locations of public facilities. His work is particularly insightful.

optimization with respect to collective choice institutional variables. If the political decision rules relating to the impure goods at issue are not provided for in the constitution, then some attempt should be made to select the best ones. This may involve analysis of such things as voting rules, extent of political «representation», different demand articulation processes, and similar considerations. Although most attention seems to have been given to voting rules, this is only one aspect of the political-institutional setting. The problems here involve selection of the appropriate institutional variables to scrutinize, and measurement of the various costs and benefits associated with changes in them (*). Although the problems associated with the whole political-institutional dimension are thought to be the most intractable, they are also likely to be the ones in which simplifying assumptions are most appropriate (because certain conditions may be truly exogenous). Simple models of simple alternatives may well be adequate.

It should be clear from the above discussion that the appropriate time frame for the model(s) is the long run. The locational system can only be altered in the long run. Also, most of the other considerations discussed above relate to the long run. Models of the short run *allocation* of public goods abound in the literature of welfare economics and public finance. Typically, these models optimize welfare as a function of utilities and lead to neat (simple) first order conditions for efficiency (see the references in footnote (***) at p. 355). To be appropriate for the long run, our model(s) must not only be capable of dealing with the short run (price) rationing and allocation problem but also must be capable of dealing with significant changes in the structure of the system. As has been recognized in the literature cited at the outset of this section as being most consistent with public goods theory, consumer's and producer's surplus seem to be the appropriate concepts for the measurement of net social benefits under the circumstances (**). The problems associated with the use of measures of economic surplus are well known [see Currie *et al.* (1971)]. Also, to be suitable for the

(*) Public choice theorists have attempted to deal-with these problems (of treating traditional political variables as economic ones). Although the literature of the public choice school is now very large, most of the key issues are summarized in Mueller (1979). Some other literature in this style included in the list of references follows: Buchanan (1968, 1974, 1975), Margolis (1968), Arrow (1969), Tullock (1969, 1970), Breton (1970), Rothenberg (1970), Niskanen (1971), Olson (1971), Young (1971, 1976), Tollison (1972), Zeckhauser (1973), Goldin (1977), O'Hare (1977), Rich (1977), Sheshinski (1977), Silver (1977), Slutsky (1977), Flowers (1978), Sonstelie, Portney (1978), Tollison, Willett (1978), and Brueckner (1979a). The works of Flowers (1978) and Tollison, Willett (1978), although not posed as operational models, are particularly instructive. Garrison (1978) also discusses some important institutional considerations associated with public facility systems.

(**) Some of the literature which has been put forward as being appropriate for locational (long run) problems has failed to appreciate the deficiencies of the approach of simply optimizing utilities. See, for example, Talley (1974), Sandler (1975), Bigman, ReVelle (1978), and Harford (1979).

long run, particular attention should be given to the discounting of future benefits and costs. The literature on discount rates for public projects is large and theoretically rigorous, but is somewhat lacking in agreement on key issues.

Because of the complexity of the overall problem that should be addressed by the model, ways should be sought of disaggregating it. One particularly helpful breakdown involves deriving the equilibrium of individuals as one problem, and optimizing social welfare, as a function of individuals' welfares as the other. We address each of these in turn. It should be noted that this disaggregation is only for analytical (and likely computational) convenience.

In the first problem, we should seek to derive the welfare (consumer's surplus) of individuals as functions of the decision variables. One way of doing this would be to start with (empirically based) utility functions and to optimize these subject to budget and institutional constraints. From this process, we may derive demand functions which are functions of the instrument variables (e.g., output levels, distances from facilities, prices, taxes, exclusion rules, decision rules, etc.). With the demand functions (which should be «compensated» Hicksian demand functions), we can then find consumer's surpluses as functions of the decision variables. This general approach allows one to capture the (equilibrium) adjustments of individuals to changes in policies (which are typically ignored in conventional location models). For example, it is possible that many families will install private swimming pools or relocate (or...) if the location of swimming pools is substantially altered, or the accessible facilities become more congested. These reactions must be considered in the locational decision. For this first problem, the relevant costs to include in the budget constraint are those borne by individuals. These include user charges, travel costs, taxes, and a whole range of costs related to political institutions.

The benefits and disbenefits associated with impure goods are typically multidimensional in nature. For example, a new urban park may be used directly by local residents, may provide a pleasant view, may attract outsiders, cut down (or increase) noise in the neighbourhood, attract mosquitos, and have any number of other effects on specific individuals. For this reason, Lancaster's «new theory of consumption» [Lancaster (1971)] seems to be particularly appropriate for public good problems (*). Indeed, an increasing number of public goods models are being cast within this framework (**). One of the advantages

(*) Lancaster's (1971) new demand theory stresses that people have demands for the attributes of goods provided, not necessarily for the goods themselves. For example, congestion may be considered one attribute of a good and, in spatial models, it may be appropriate to consider the location (the distance away) as an attribute.

(**) Some of the public goods models cast in Lancasterian terms are: Oakland (1972), Sandmo (1973, 1975), Lancaster (1976), Rothenberg (1976), Sandler, Cauley (1976), De Serpa (1977, 1978), Hillman (1977), and Muzondo (1978).

of the use of consumer's surplus is that the consumer's surpluses of various goods or attributes of goods, for any individual, can simply be added up to get an aggregate measure of welfare. This is particularly important for a problem in which the benefits tend to be so varied and dissimilar.

Regardless of the theory of demand used, the problem of empirically estimating demand curves for impure goods is difficult. Besides the powerful demand revealing processes (noted in footnote (*) at p. 355), a variety of other interesting possibilities has been developed in recent years which should be looked into (*). There are also now a number of approaches to the estimation of consumers' surpluses associated with facilities in spatial contexts which seem to be fairly good. Neuberger (1971) has discussed many aspects of the use of consumer's surplus in transport and land use plans. Bollobas, Stern (1972) and Stern (1972) have shown, rather elegantly, how the concept of surplus can be used to derive the socially optimal size and structure of market areas. Cesario (1976) and Cesario, Knetsch (1976) present one approach to measuring the benefits from recreational facilities to which individuals must travel for consumption. Finally, Williams (1976), Coelho, Wilson (1976), Coelho, Williams (1978), and Leonardi (1978, 1980) have shown that certain gravity-model-type objectives can be interpreted as consumer surpluses. Thus we have a powerful new methodology for measuring the benefits of travelled-for facilities which duly considers that consumers do not simply patronize closest facilities. However, the realistic consideration of multipurpose trip-making has yet to be adequately modelled in a way that it can be operationalized.

Note should be made of the problem of demand generation effects of altering locational patterns. As Sheppard (1980) has pointed out, it is not sufficient to model tripmaking behaviour or facility patronage as a spatial interaction process; we must, in addition, account rigorously for the fact that total demands are affected by changes in facility patterns. Sheppard's theory indicates how this may be done by taking account simultaneously of the number of trips made and the amount of the service used per visit [see also Berglas (1976b)]. The individual trades off costs of travel, attractiveness of destination, and, in addition, the costs of maintaining an inventory of the good. It is important that any new public facilities models deal with both the number of trips and the amount consumed per trip rather than the aggregate variable «amount consumed».

It remains to construct rigorous models of demand and surplus which incorporate congestion and crowding. This must be a high priority as these are extremely important dimensions of public consumption. It is

(*) See, for example, Bohm (1972), Hori (1975), Strauss, Hughes (1976), Bradford, Hildebrandt (1977) and Maital (1979).

ironic that facilities which tend to be better located tend also to be the most congested. The first difficulty in modelling congestion is the measurement problem (which exists *a fortiori* because congestion is an externality). Various approaches are found in the literature. Congestion can be considered as a subtraction from the «benefits» derived from each unit of consumption, or it can be considered as an additional «cost» incurred with each unit of consumption. Each may be suitable in different contexts. Note, however, that measuring consumption using both the number of trips and the amount consumed per trip should aid in the construction of theoretically rigorous models. Another difficult problem in modelling congestion is the interdependencies involved. An individual's equilibrium with respect to consumption depends on the consumption of other individuals. It is necessary to make certain assumptions about an individual's expectation of congestion. Indeed, this consideration, the uncertainty of certain costs, and the probabilistic nature of facility patronage, strongly suggest that one should use expected utilities, expected demand, and expected surpluses in public facilities models.

After consumers' surpluses have been expressed as a function of the decision variables, the second problem of welfare maximization seems to pose fewer difficulties of detail. In this model, social welfare is expressed as some aggregation of consumers' and producers' surpluses. Essentially the objective should be considered as the average net benefit per capita. The constraints relate to the resources used in the production and delivery of the impure good plus a host of institutional constraints as well as accounting or definitional constraints. The precise nature of the specification of the model would depend on the precise nature of the problem at hand.

It does not seem possible to say much more. However, there is one very significant item which should be addressed within the welfare problem - the issue of equity or the distribution of income.

The new welfare economics within which the theory of public goods has been developed is based on the Pareto norm and no further, stronger, ethical norms. The Pareto norm (that welfare is improved if at least one person's welfare is increased and no one else's welfare is reduced) is ethically fairly weak and uncontentious and seems to command wide assent. We could attempt to discover a range of locational and other policies that satisfies the Pareto norm. Unfortunately, however, there is an infinity of such policies, related to an infinity of income distributions in society. We will have learned a good deal in the process but we will not contribute much to the construction of operational location models.

Il we require specific values for the decision variables (such as specific locations, specific tax-levels etc.) it is simply not possible to side-step the problem of dealing directly with interpersonal utility comparisons. If the social welfare function is taken to be a function of the utilities of all individuals in society we would require a welfare weight for each person in order to have an operational model, yielding a specific policy solution. These welfare weights have to come from somewhere and are very unlikely to be provided directly and unequivocally by political decision-makers (although every attempt should be made to get them in this way). If the objective is to include a direct summation of consumers' surpluses, the value judgement implied is that all individuals welfares should be considered equal.

In recent years considerable attention has been given to the possibility that redistributions of income can be judged (solely) with the Paretian norm which underlies all of the new welfare economics (*). The argument is that people have preferences for the distribution of income to particular other individuals, including themselves [Hochman, Rodgers (1969, 1974), Pauly (1973)] or they have preferences for the overall distribution of income in society [Thurow (1971), Breit (1974)]. Thus particular, or general, redistribution can itself be considered just another (multidimensional, impure) public good for which we tend to expect a suboptimal output precisely because of the public good and externality problems. To the extent that preferences for redistribution do exist they should surely be taken into account in structuring public facility systems. However, the problem of preference or demand revelation becomes particularly acute in this realm.

If it were possible to solve the whole «problem of equity» by recourse to the Pareto norm, then, in essence, the whole issue of values and politics would become relatively insignificant. Unfortunately, although the Pareto norm can be used for a superficial redistribution of income, to claim that it solves the whole problem requires the assumption that the initial income distribution be accepted as optimal (**). For fundamental redistribution, the Paretò norm can never be adequate. Some stronger ethical norm must always be invoked. It is extremely unlikely that any strong norms can be derived which will command wide acceptance (***).

Precisely where the stronger norms which are required should come from is an unanswered question. Buchanan's (1975) theory provides some guidance if one is willing to trace the problem back to the constitutional level of decision-making. For most, however, this is likely to be unsatisfactory. Perhaps value judgements relating to equity can be

(*) The Paretian norm has underlain all of the discussion to this point in the paper. The possibility of Pareto efficient redistribution was first proposed by Hochman, Rodgers (1969, 1974). Other discussions are provided by Thurow (1971, 1973), Mishan (1972), Pauly (1973), Breit (1974), Buchanan (1974), and Kleiman (1978).

(**) This important point is stressed by Mishan (1972).

(***) Rawls (1971) represents one attempt. Recent criticism, however, indicates that his theory is not widely accepted.

derived directly from asking political decision-makers or observing the outcomes of political decisions. However, as both of these are unlikely, perhaps the best way of proceeding is to derive a whole range of «solutions» to the problem at hand, each of which is based on a different income distributional norm. Politicians would then be forced to select a particular solution. No matter how one arrives at the solutions, it is incumbent upon the analyst to point out clearly the implications of each particular solution for the (re)distribution of income (*).

9. Concluding comments

As attempt has been made to point out many of the theoretical deficiencies of most of the existing location-allocation models for public facilities. Most have *ad hoc* theoretical underpinnings. They tend to be extremely simple models of extremely complex public problems. In large measure, the deficiencies stem from a failure to recognize the public/political/institutional nature of the problem of public facility location.

Because almost all of the goods and services provided by, or at, public facilities have (and should have) one or more of the attributes of public goods, the recently developed theory of public goods seems to be a very appropriate foundation on which to build a theory of public facility location. However, a recognition that public facilities provide public goods brings with it a clear realization that the problems of location and allocation do not exist in a vacuum. The problem of location must be situated within a whole set of related problems and the construction of a theory about this whole interrelated set is required. The key questions to be addressed by a theory of the (optimal) public space economy have been articulated, briefly clarified, and related to one another. It has been shown that a number of logically higher level problems must be addressed (and solved) in order for locations and allocations to be meaningfully optimized. For example, optimal public facility locations have little meaning if the goods provided can be more efficiently provided by private facilities (or contracted out), if the system of jurisdictions is decidedly irrational, or if inappropriate user charges and/or taxes are used.

An attempt has been made to describe how the theory of public goods must be altered to deal with impure goods in a spatial context. Unfortunately, only a very small proportion of the received theory of public goods takes space into account. In «spatializing» the theory of (impure) public goods, stress has been placed on showing that the

(*) Savas (1978) has discussed some important aspects to be considered in the reporting of the distributional implications of public policies.

problems of delivered goods and travelled-for goods are significantly different and should be modelled differently and that we can generalize the jointness dimension of impure (public) goods. The dimensions of (institutional) exclusion and rejection, although still important variables to consider, can be treated the same way as they are treated in aspatial theory.

In the final section of the paper were surveyed some rather more technical considerations which should be taken into account in theconstruction of the theory of impure goods and on the general theory of the optimal space economy. It is suggested that after the problem has been defined and the decision variables set out, the larger problem be decomposed into subproblems. The first problem would be the equilibrium of individuals (households) as functions of the key decision variables. Here the individual is allowed to maximize utility subject to budget constraints in an attempt to adjust to changes in the policy variables. Consumer welfare should be measured using consumer's surplus. The problems associated with measuring demands and surpluses were briefly addressed. In the second problem, welfare is maximized as a function of the welfare of individuals. The constraints will be system specific. The principal difficulty with the second model relates to interpersonal welfare comparisons which simply must be made in order to derive specific policies. It is suggested that the problem be solved with a set of different equity norms and the range of solutions set before political decision-makers.

I believe that a spatially generalized theory of public goods and a derivative theory of the optimal public space economy will provide a powerful theory for public facility location and that this theory can be operationalized in the form of rather more broadly conceived location-allocation models than those currently available.

References

Adams R.D. Royer J.S. (1977) Income and price effects in the economic theory of clubs, *Public Finance, 32,* 2, 141-158.

Ahlbrandt R.S. (1973a) *Municipal fire protection services: comparison of alternative organizational forms,* Sage Professional Papers in Administrative and Policy Studies, O3-002, Beverly Hills, CA.

Ahlbrandt R.S. (1973b) Efficiency in the provision of fire services, *Public Choice, 16,* 1-15.

Allen L., Amacher R., Tollison R. (1974) The economic theory of clubs: a geometric exposition, *Public Finance, 29,* 3-4, 386-391.

Arrow K.J. (1969) The organization of economic activity: issues pertinet to the choice of market versus non-market allocation, in U.S. Joint Economic Committee, *The analysis and evaluation of public expenditures: the PPB System,* U.S. Government Printing Office, Washington, D.C..

Barr J.L. (1972) City size, land rent and the supply of public goods, *Regional and Urban Economics, 2,* 67-103.

Barzel Y. (1969) Two propositions on the optimum level of producing public goods, *Public Choice, 6,* 31-37.

Baumol W.J., Oates W.E. (1975) *The theory of environmental policy,* Englewood Cliffs, Prentice-Hall, N.J..

Berglas E. (1976a) On the theory of clubs, *American Economic Review, Papers and Proceedings, 66,* 2, 116-121.

Berglas E. (1976b) Distribution of tastes and skills and the provision of local public goods, *Journal of Public Economics, 6,* 4, 409-423.

Bigman D., ReVelle C. (1978) The theory of welfare considerations in public facility location problem, *Geographical Analysis, 10,* 229-240.

Bigman D., ReVelle C. (1979) An operational approach to welfare considerations in applied public facility location models, *Environment and Planning A, 11,* 1, 83-96.

Bird R.M., Hartle D.G. (1972) The design of governments, in Bird R.M., Head J.G. (eds.) *Modern fiscal issues,* University of Toronto Press, Toronto.

Bish R.L. (1971) *Public economy of metropolitan areas,* Markham Press, Chicago.

Boadway R. (1980) A note on the market provision of club goods, *Journal of Public Economics, 13,* 131-137.

Bohm P. (1972) Estimating demand for public goods: an experiment, *European Economic Review, 3,* 2, 111-130.

Bollobas B., Stern N. (1972) The optimal structure of market areas, *Journal of Economic Theory, 4,* 174-194.

Borukhov E. (1972) Optimal service area for provision and financing of local public goods, *Public Finance, 27,* 3, 267-281.

Boskin M.J. (1973) Local government tax and product competition and the optimal provision of public goods, *Journal of Political Economy, 81,* 1, 203-210.

Bradford D.F., Hildebrandt G.G. (1977) Observable preferences for public goods, *Journal of Public Economics, 8,* 111-132.

Breit W. (1974) Income redistribution and efficiency norms, in Hochman H.M., Peterson G.E. (eds.) *Redistribution through public choice,* Columbia University Press, New York.

Brennan G., Flowers M. (1980) All «Ng» up on clubs?: some notes on the current state of club theory, *Public Finance Quaterly, 8,* 2, 153-169.

Breton A. (1965) A theory of government grants, *Canadian Journal of Economics and Political Science, 31,* 175-187.

Breton A. (1970) Public goods and the stability of federalism, *Kyklos, 23,* 4, 882-901.

Breton A. (1974) *The economic theory of representative government,* Aldine Press, Chicago.

Breton A., Scott A. (1977) The assignment problem in federal structures, in Feldstein, M.S., Inman R.P. (eds.) *The economics of public services,* Macmillan, London.

Breton A., Scott A. (1978) *The economic constitution of federal states,* University of Toronto Press, Toronto.

Brueckner J.K. (1979a) Spatial majority voting equilibria and the provision of public goods, *Journal of Urban Economics, 6,* 3, 338-351.

Brueckner J.K. (1979b) Property values, local public expenditure and economic efficiency, *Journal of Public Economics, 11,* 223-245.

Buchanan J.M. (1965) An economic theory of clubs, *Economica, 32,* 1-14.

Buchanan J.M (1968) The demand and supply of public goods, Rand McNally,Chicago.

Buchanan J.M. (1974) Who should distribute what in a federal system, in Hochman H.M., Peterson G.E. (eds.) *Redistribution through public choice,* Columbia University Press, New York.

Buchanan J.M. (1975) *The limits of liberty: between anarchy and leviathan,* University of Chicago Press, Chicago.

Buchanan J.M., Stubblebine W.C. (1962) Externality, *Economica,* 29, 371-384.

Buchanan J.M., Goetz C.J. (1972) Efficiency limits of fiscal mobility: an assessment of the Tiebout model, *Journal of Public Economics, 1,* 25-43.

Capozza D.R. (1976) *Optimal spacing and pricing in the public sector,* Paper presented at Meeting of North American Conference of Regional Science Association, Toronto.

Casetti B., Thrall G. (1979) Tax schedules in the ideal city: equilibrium versus optimality, *Geographica Polonica, 42,* 33-48.

Cesario F.J. (1976) Demand curves for public facilities, *Annals of Regional Science, 10*, 3, 1-14.

Cesario F.J., Knetsch J.L. (1976) A recreation site demand and benefit estimation model, *Regional Studies, 10*, 97-104.

Chamberlin J.R. (1974) Provision of collective goods as a function of group size, *American Political Science Review, 68,* 707-716.

Clarke E.H. (1971) Multipart pricing of public goods, *Public Choice, 11*, 17-34.

Clarke E.H. (1972) Multipart pricing of public goods: an example, in Mushkin S. (ed.) *Public prices for public products,* The Urban Institute, Washington.

Coelho J.D., Wilson A.G. (1976) The optimum location ad size of shopping centres, *Regional Studies, 10,* 413-421.

Coelho J.D., Williams H.C.W.L. (1978) On the design of land use plans through locational surplus maximization, *Papers of Regional Science Association, 40,* 71-86.

Connolly M. (1970) Public goods, externalities and international relations, *Journal of Political Economy, 78,* 279-290.

Connolly M. (1972) Trade in public goods: a diagrammatic analysis, *Quarterly Journal of Economics, 86,* 61-78.

Connolly M. (1976) Optimal trade in public goods, *Canadian Journal of Economics, 9,* 4, 702-705.

Courant P.N., Rubinfield D.L. (1978) On the measurement of benefits in an urban context: some general equilibrium models, *Journal of Urban Economics, 5,* 3, 346-356.

Cox K.R., Dear M. (1975) Jurisdictional organization and urban welfare, Discussion Paper 47, Department of Geography, Ohio State University, Columbus, Ohio.

Currie J.M., Murphy M.A., Schmitz A. (1971) The concept of economic surplus and its use in economic analysis, *The Economic Journal, 81,* 741-799.

Davies O.A. (1974) Optimal facility location in a one-dimensional spatial market, *Geographical Analysis, 6,* 3, 239-264.

Dear M., Fincher R., Currie L. (1977) Measuring the external effects of public programmes, *Environment and Planning A, 9,* 2, 137-147.

De Serpa A.C. (1977) A theory of discriminatory clubs, *Scottish Journal of Political Economy, 24,* 1, 33-41.

De Serpa A.C. (1978) Congestion, pollution and impure public goods, *Public Finance, 33,* 1-2, 68-83.

Downs A. (1967) *Inside bureaucracy,* Little, Brown, Boston.

Dreze J., de la Vallée Poussin D. (1970) A tâtonnement process for public goods, *Review of Economic Studies, 38,* 133-150.

Ellickson B. (1973) A generalization of the pure theory of public goods, *American Economic Review, 63,* 3, 417-432.

Ellickson B. (1977) The politics and economics of decentralization, *Journal of Urban Economics, 4,* 135-149.

Ellickson B. (1979) Local public goods and the market for neighborhoods, in Segal D. (ed.) *The economics of neighborhood,* Academic Press, New York.

Erlenkotter D. (1977) Facility location with price sensitive demands: private, public and quasi-public, *Management Science, 24,* 4, 378-386.

Evans A.W. (1970) Private good, externality, public good, *Scottish Journal of Political Economy, 17,* 79-89.

Evans A.W. (1971) Definitions and welfare conditions of public goods: a reply, *Scottish Journal of Political Economy, 18,* 203-208.

Evans A.W. (1973) Public goods and metropolitan consolidation, in Proceedings of the 28th Congress of the International Institute of Public Finance, *Issues in Urban Public Finance.*

Fisch O. (1975) Optimal city size. The economic theory of clubs and exclusionary zoning, *Public Choice, 24,* 59-70.

Fisch O. (1976) Optimal city size, land tenure and the economic theory of clubs, *Regional Science and Urban Economics, 6,* 1, 33-44.

Fisch O. (1977) Spatial equilibrium with local public goods: urban land rent, optimal city size, and the Tiebout hypothesis, *Regional Science and Urban Economics, 7,* 3, 197-216.

Fisch O. (1980) Spatial equilibrium with locational interdependencies: the case of environmental spillovers, *Regional Science and Urban Economics, 10*, 201-209.

Flatters F., Henderson V., Mieszkowski P. (1974) Public good, efficiency and regional fiscal equalization, *Journal of Public Economics, 3,* 2, 99-112.

Flowers M.R. (1978) Costs of collective decisions, choice of tax base, and median voter equilibrium, *Public Finance Quarterly, 6,* 3, 305-309.

Freeman A.M., Haveman R.H. (1977) Congestion, quality deterioration and heterogeneous tastes, *Journal of Public Economics, 8,* 2, 225-232.

Friesema H.P. (1970) Interjurisdictional agreements in metropolitan areas, *Administrative Science Quarterly, 15,* 242-252.

Gale S. (1976) A resolution of the regionalization problem and its implications for political geography and social justice, *Geografiska Annaler, 58B,* 1, 1-16.

Gale S., Atkinson M. (1979) Towards an institutionalist perspective on regional science: an approach via the regionalization question, *Papers of the Regional Science Association, 43*, 59-82.

Garrison W.L. (1978) Thinking about public facility systems, in *The National Research Council in 1978,* National Academy of Sciences, Washington, D.C..

Getz M. (1975) A model of the impact of transportation investment on land rents, *Journal of Public Economics, 4,* 57-74.

Goldin K.D. (1977) Equal access vs. selective access: a critique of public goods theory, *Public Choice, 29,* 53-71.

Green J., Kohlberg E., Laffont J.-J. (1976) Partial equilibrium approach to the free-rider problem, *Journal of Public Economics, 6,* 4, 375-394.

Green J., Laffont J.-J. (1978) A sampling approach to the free rider problem, in Sandmo A. (ed.) *Essays in public economics,* Lexington Books, Lexington, Mass..

Greenberg. J. (1977) Existence of an equilibrium with arbitrary tax schemes for financing local public goods, *Journal of Economic Theory, 16,* 137-150.

Greenberg J. (1978) Pure and local public goods: a game-theoretic approach, in Sandmo A. (ed.) *Essays in public economics,* Lexington Books, Lexington, Mass..

Greene K.V. (1977) Spillovers, migration and public school expenditures: the repetition of an experiment, *Public Choice, 29,* 85-93.

Groves T., Ledyard J. (1977) Optimal allocation of public goods: a solution to the «free rider» problem, *Econometrica, 45,* 4, 783-809.

Hamilton B.W. (1976) Capitalization of intrajurisdictional differences in local tax prices, *The American Economics Review, 76,* 743-753.

Hansen P., Thisse J.-F. (1977) Multiplant location for profit maximisation, *Environment and Planning A, 9,* 63-73.

Harford J.D. (1977) Optimizing intergovernmental grants with three levels of government, *Public Finance Quarterly, 5,* 1, 99-116.

Harford J.D. (1979) The spatial aspects of local public goods: a note, *Public Finance Quarterly, 7,* 1, 122-128.

Head J.G. (1973) Public goods and multi-level government, in David W.L. (ed.) *Public finance, planning and economic development,* Macmillan, London.

Head J.G. (1974) *Public goods and public welfare,* Duke University Press, Durham, N.C.

Head J.G., Shoup C.S. (1969) Public goods, private goods and ambiguous goods, *Economic Journal, 89,* 567-572.

Head J.G., Shoup C.S. (1973) Public, private and ambiguous goods reconsidered, *Public Finance, 28,* 3-4, 384-392.

Helpman E., Hillman A.L. (1976) On optimal club size, Working Paper 91, Foerder Institute for Economic Research, Tel Aviv, Israel.

Helpman E., Hillman A.L. (1977) Two remarks on optimal club size, *Economica, 44,* 175, 293-296.

Helpman E., Pines D. (1977) Land and zoning in an urban economy: further results, *American Economic Review, 67, 5,* 982-986.

Helpman E., Pines D., Borukhov E. (1976) The interaction between local government and urban residential location: comment, *American Economic Review, 66,* 5, 961-967.

Henderson J.V. (1977a) Externalities in a spatial context, *Journal of Public Economics, 7,* 89-100.

Henderson J.V. (1977b) *Economic theory and the cities,* Academic Press, New York.

Henderson J.V. (1979) Theories of groups, jurisdictions and city size, in Mieszkowski P., Straszheim M. (eds.) *Current Issues in Urban Economics,* Johns Hopkins University Press, Baltimore.

Hillman A.L. (1977) The theory of clubs: a technological formulation, in Sandmo A. (ed.) *Essays in public economics,* Lexington Books, Lexington, Mass.

Hirsch W.Z. (1964) Local versus areawide urban government services, *National Tax Journal, 17,* 4, 331-339.

Hirsch W.Z. (1970) *The economics of state and local government,* McGraw-Hill, New York.

Hochman H.M., Rodgers J.D. (1969) Pareto optimal redistribution, *American Economic Review, 59,* 542-557.

Hochman H.M., Rodgers J.D. (1974) Redistribution and the Pareto criterion, *American Economic Review, 64,* 4, 752-757.

Hodgson M.J. (1978) Toward more realistic allocation in location-allocation models: an interaction approach, *Environment and Planning, A, 10,* 11, 1273-1286.

Holtmann A.G., Tabasz T., Kruse W. (1976) The demand for local public services: spillovers, and urban decay, the case of libraries, *Public Finance Quarterly, 4,* 1, 97-113.

Homma M., Yamada M. (1980) A dynamic mobility model of individuals between jurisdictions in a system of local governments, *Regional Science and Urban Economics, 10,* 1, 109-121.

Honey R., Stratham J. (1978) Jurisdictional consequences of optimizing public goods, *Annals of Regional Science, 12,* 2, 32-40.

Hori H. (1975) Revealed preference for public goods, *American Economic Review, 65,* 5, 978-991.

James E. (1974) Optimal pollution control and trade in collective goods, *Journal of Public Economics, 3,* 203-216.

Jurion B.J. (1979a) Matching grants and unconditional grants: the case with n goods, *Public Finance, 34,* 2, 234-244.

Jurion B.J. (1979b) Une analyse globale des effets economique de diverses formes de subventions attribuitees par le governement central aux authorités locales, *Revue d'Economie Politique, 89,* 8, 297-313.

Kamien M.I., Schwartz N.L., Roberts D.J. (1973) Exclusion, externalities and public goods, *Journal of Public Economics, 2,* 3, 217-230.

Kanemoto Y (1976) Optimum market and second-best land use patterns in a Von Thünen city with congestion, *Regional Science and Urban Economics, 6,* 1, 23-32.

Kiesling H.J. (1974) Public goods and the possibilities for trade, *Canadian Journal of Economics, 7,* 3, 402-417.

Kiesling H.J. (1976) A model for analyzing the effects of governmental consolidation in the presence of public goods, *Kyklos, 29,* 233-255.

Kleiman E. (1978) Inequality as a public good: unambiguous redistribution and optimality, in Sandmo A. (ed.) *Essays in public economics,* Lexington Books, Lexington, Mass..

Koleda M.S. (1971) A public good model of government consolidation, *Urban Studies, 8, 2,* 103-110.

Lancaster K. (1971) *Consumer demand: a new approach,* Columbia University Press, New York.

Lancaster K. (1976) The pure theory of impure public goods, in Grieson R.E. (ed.) *Public and urban economies,* Lexington Books, Lexington, Mass..

Lea A.C. (1978) *Interjurisdictional spillovers and efficient public good provision,* Paper presented to the Annual Meeting, Association of American Geographers, New Orleans.

Lea A.C. (1979a) Welfare theory, public goods and public facility location, *Geographical Analysis, 11,* 3, 217-239.

Lea A.C. (1979b) Welfare theory, public goods and public facility location: a rejoinder, *Geographical Analysis, 11*, 4, 292-294.

Lea A.C. (1980) Towards a theory of the public space economy, Unpublished Ph. D. Dissertation, University of Toronto.

Le Grand J. (1975) Fiscal equity and central government grants to local authorities, *Economic Journal, 85,* 339, 531-547.

Leonardi G. (1978) Optimum facility location by accessibility maximizing, *Environment and Planning A, 10,* 11, 1287-1305.

Leonardi G. (1980) A multiactivity location model with accessibility and congestion sensitive demand, Working Paper 80-124, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Le Roy S.F. (1976) Urban land rent and the incidence of property taxes, *Journal of Urban Economics, 3,* 2, 167-179.

Liebman J.C. (1976) Some simple minded observations on the role of optimization in public systems decision-marking, *Interfaces, 6,* 4, 102-108.

Lind R.C. (1973) Spatial equilibrium, rents and public program benefits, *Quarterly Journal of Economics, 87,* 2, 188-207.

Lindahl E. (1919, 1958) Just taxation: a positive solution, in Musgrave R.A. Peacock T.A. (eds.) *Classics in the theory of public finance,* Macmillan, London. Originally published 1919 in German.

Lipsey R.G., Lancaster K. (1956) The general theory of second best, *Review of Economic Studies, 24,* 11-32.

Maital S. (1979) Measurement of net benefits from public goods: a new approach using survey data, *Public Finance, 34,* 1, 85-99.

Malinvaud E. (1971) A planning approach to the public goods problem, *The Swedish Journal of Economics, 73,* 1, 96-112.

Malinvaud E. (1972) Prices for individual consumption; quantity indicators for collective consumption, *Review of Economic Studies, 39,* 385-405.

Margolis J. (1968) The demand for urban public services, in Perloff H.S., Wingo L. Jr. (eds.). *Issues in urban economics,* Johns Hopkins Press, Baltimore.

Mathur V.K. (1976) Spatial economic theory of pollution control, *Journal of Environmental Economics and Management, 3,* 1, 16-28.

McGuire M.C. (1972) Private good clubs and public good clubs, *Swedish Journal of Economics, 74,* 84-99.

McGuire M.C. (1974) Group segregation and optimal jurisdictions, *Journal of Political Economy, 82,* 1, 112-132.

McGuire M.C., Aaron H. (1969) Efficiency and equity in the optimal supply of a public good, *Review of Economics and Statistics 51,* 31-39.

McMillan M.L. (1975) Toward more optimal provision of local public goods: internalization of benefits or intergovernmental grants, *Public Finance Quarterly, 3,* 3, 229-260.

McMillan M.L. (1976) Criteria for jurisdictional design: issues in defining scope and structure of river basin authorities, *Journal of Environmental Economics and Management, 3,* 1, 46-68.

Mehay S.L. (1977) Interjurisdictional spillovers of police services, *Southern Economic Journal, 43,* 1352-1359.

Meyer R.A. (1971) Private costs of using public goods, *Southern Economic Journal, 37,* 479-488.

Mieszkowski P., Oakland W.H. (eds.) (1979) *Fiscal federalism and grants-in-aid,* The Urban Institute, Washington, D.C..

Milleron J.-C. (1972) Theory of value with public goods: a survey article, *Journal of Economic Theory, 5,* 419-477.

Mishan E.J. (1969) The relationship between joint products, collective goods and external effects, *Journal of Political Economy, 77,* 329-348.

Mishan E.J. (1972) The futility of Pareto efficient distributions, *American Economic Review, 62,* 3-4, 917-977.

Miyao T. (1978) A probabilistic model of location choice with neighborhood effects, *Journal of Economic Theory, 19,* 2, 347-358.

Morrill R.L., Symons J. (1977) Efficiency and equity aspects of optimal location, *Geographical Analysis, 9,* 3, 215-225.

Morrison C.C. (1977) Public goods and the property tax: a theoretical analysis, *Public Finance Quarterly, 4,* 2, 159-172.

Mueller D.C. (1979) *Public choice*, Cambridge University Press, Cambridge.

Musgrave R.A. (1959) *Theory of public finance: a study in public economy,* McGraw-Hill, New York.

Musgrave R.A. (1969) Theories of fiscal federalism, *Public Finance, 24,* 4, 521-536.

Musgrave R.A. (1971) Economics of fiscal federalism, *Nebraska Journal of Economics and Business, 10,* 4, 3-13.

Musgrave R.A., Musgrave P.B. (1973) *Public finance in theory and practice,* McGraw-Hill, New York.

Muzondo R.T. (1978) Mixed and pure public goods, user charges and welfare, *Public Finance, 33,* 3, 314-330.

Neuberger H.L.I. (1971) User benefit in the evaluation of transport and land use plans, *Journal of Transport Economics Policy, 5,* 52-75.

Ng Y.-K. (1971) Definitions and welfare conditions of public goods, *Scottish Journal of Political Economy, 18,* 199-202.

Ng Y.-K. (1973) The economic theory of clubs: Pareto optimality conditions, *Economica, 40,* 159, 291-298.

Ng Y.-K. (1978) Optimal club size: a reply, *Economica, 45,* 180, 407-410.

Ng Y.-K., Tollison R.D. (1974) A note on consumption sharing and non-exclusion rules, *Economica, 41,* 164, 446-450.

Niskanen W.A. (1971) *Bureaucracy and representative government,* Aldin-Atherton, Chicago.

Oakland W.H. (1972) Congestion, public goods and welfare, *Journal of Public Economics, 1,* 339-357.

Oates W.E. (1972) *Fiscal federalism,* Harcourt, Brace, New York.

Oates W.E. (1977a) *The political economy of federalism,* Lexington Books, Lexington, Mass.

Oates W.E. (1977b) An economist's perspective on fiscal federalism, in Oates W.E. (ed.) *The political economy of fiscal federalism*, Lexington Books, Lexington, Mass.

O'Hare M. (1977) Not on my block you don't: facility siting and the strategic importance of compensation, *Public Policy, 25,* 4, 407-458.

Olson M. (1969) The principle of «fiscal equivalence»: the division of responsabilities among different levels of government, *American Economic Review, 59,* 479-487.

Olson M. (1971) *The logic of collective action,* 2nd edition, Schocken Books, New York.

Ostrom V. (1969) Operational federalism: organization for the provision of public services in the American federal system, *Public Choice, 6,* 1-17.

Ostrom V., Tiebout C.M., Warren R. (1961) The organization of government in metropolitan areas: a theoretical inquiry, *American Political Science Review, 55,* 831-842.

Papageorgiou G.J. (1978) Spatial externalities: I. Theory; II. Applications, *Annals, Association of Americans Geographers, 68,* 4, 465-492.

Pauly M.V. (1967) Clubs, commonality, and the core: an integration of game theory and the theory of public goods, *Economica, 34,* 314-324.

Pauly M.V. (1970a) Optimality; public goods and local governments: a general theoretical analysis, *Journal of Political Economy, 78,* 572-585.

Pauly M.V. (1970b) Cores and clubs, *Public Choice, 9,* 53-65.

Pauly M.V. (1973) Redistribution as a local public goods, *Journal of Public Economics, 2,* 35-58.

Pestieau P. (1977) The optimality limits of the Tiebout model, in Oates W.E. (ed.) *The political economy of fiscal federalism,* Lexington Books, Lexington, Mass.

Pestieau P. (1980) Fiscal mobility and local public goods: a survey of the empirical and theoretical studies of the Tiebout model, Research Paper No. 6, SPUR, Université Catholique de Louvain, Belgium.

Polinsky A.M. (1973) Collective consumption goods and local public finance theory: a suggested analytic framework, in Proceedings of the 28th Congress of the International Institute of Public Finance, *Issues in Urban Public Finance.*

Premus R. (1979) Community selection and equilibrium spatial structure of communities, *Growth and Change, 10,* 3, 25-36.

Rawls J. (1971) *A theory of justice,* Harvard University Press, Cambridge.

ReVelle C.S., Marks D.M., Liebman J.C. (1970) An analysis of private and public sector location models, *Management Science, 16,* 692-707.

Rich R.C. (1977) Equity and institutional design in urban service delivery, *Urban Affairs Quarterly, 12,* 3, 383-410.

Richardson H.W. (1977) *The new urban economics and alternatives,* Pion, London.

Richter D.K. (1975) Existence of general equilibrium of multiregional economies with public goods, *International Economic Review, 16,* 1, 201-221.

Richter D.K. (1978) Existence and composition of a Tiebout general equilibrium, *Econometrica, 46*, 779-805.

Richter D.K. (1979) A computational approach to the study of neighborhood effects in general equilibrium land use models, in Segal D. (ed.), *The Economics of Neighborhoor,* Academic Press, New York.

Rittenoure R.L., Pluta J.E. (1977) Theory of intergovernmental grants and local government, *Growth and Change, 8,* 3, 31-37.

Roberts D.J. (1974) A note on returns to group size and the core with public goods, *Journal of Economic Theory, 9,* 3, 350-356.

Roberts D.J. (1976) The incentives for correct revelation of preferences and the number of consumers, *Journal of Public Economics, 6,* 4, 359-374.

Rose-Ackerman S. (1979) Market models of local government: exit, voting and the land market, *Journal of Urban Economics, 6*, 3, 319-337.

Rosekamp K.W. (1980) Tax prices and the optimal supply of a public good: the Lindahl solution and second best ones, *Public Finance, 35*, 1, 114-119.

Rothenberg J. (1970) Local decentralization and the theory of optimal government, in Margolis J. (ed.) *The analysis of public output,* National Bureau of Economic Research, New York.

Rothenberg J. (1976) Inadvertent distributional impacts in the provision of public services to individuals, in Grieson R.E. (ed.) *Public and urban economics,* Lexington Books, Lexington, Mass.

Rufolo A.M. (1979) Efficient local texation and local public goods, *Journal of Public Economics, 12*, 351-376.

Sakashita N. (1975) *Optimal location of public facilities under influence of the land market,* Paper presented to the European Regional Science Meetings, Budapest, Hungary.

Samuelson P.A. (1954) The pure theory of public expenditures, *Review of Economics and Statistics, 36,* 387-389.

Samuelson P.A. (1955) Diagramming exposition of a theory of public expenditure, *Review of Economics and Statistics, 37,* 350-356.

Samuelson P.A. (1969) Pure theory of public expenditure and taxation, in Margolis J., Guitton H. (eds.) *Public Economics,* Macmillan, London.

Sandler T.M. (1975) Pareto optimality, pure public goods, impure public goods and multi-regional spillovers, *Scottish Journal of Political Economy, 22,* 1, 25-38.

Sandler T.M. (1978a) Public goods and the theory of second best, *Public Finance, 33,* 3, 331-344.

Sandler T.M., Shelton R.B. (1972) Fiscal federalism, spillovers and the export of taxes, *Kyklos, 25*, 4, 736-753.

Sandler T.M., Cauley J. (1976) Multiregional public goods, spillovers, and the new theory of consumption, *Public Finance, 31,* 3, 376-395.

Sandmo A. (1973) Public goods and the technology of consumption, *Review of Economic Studies, 40,* 517-528.

Sandmo A. (1975) Public goods and the technology of consumption: a correction, *Review of Economic Studies, 42,* 167-168.

Savas E.S. (1978) On equity in providing public service, *Management Science, 24,* 8, 800-808.

Schuler R.E. (1974) The interaction between local government and urban residential location, *American Economic Review, 64,* 4, 682-696.

Schuler R.E., Holahan W.L. (1977) Optimal size and spacing of public facilities in metropolitan areas: the maximum covering location problem revisited, *Papers, Regional Science Association, 39,* 137-156.

Sheppard E.S. (1980) Location and the demand for travel, *Geographical Analysis, 12,* 2, 111-128.

Sheshinski E. (1977) The supply of communal goods and revenue sharing, in Feldstein, M.S., Inman R.P. (eds.) *The economics of public services,* Macmillan, London.

Silver M. (1977) Economic theory of the constitutional separation of powers, *Public Choice, 29,* 95-107.

Slutsky S. (1977) A voting model for the allocation of public goods: existence of an equilibrium, *Journal of Economic Theory, 14,* 299-325.

Smolensky E., Burton R., Tideman N. (1970) The efficient provision of a local non-private goods, *Geographical Analysis, 2,* 330-342.

Sonstelie J.C., Portney P.R. (1978) Profit maximizing communities and the theory of local public expenditure, *Journal of Urban Economics, 5,* 2, 263-277.

Starrett D.A. (1980) Measuring externalities and second best distortion in the theory of local public goods, *Econometrica, 48,* 627-642.

Stern N. (1972) The optimal size of market areas, *Journal of Economic Theory, 4,* 154-173.

Stigler G.J. (1962) The tenable range of functions of local government, in Phelps E.S. (ed.) *Private wants and public needs,* W.W. Norton, New York.

Stiglitz J.E. (1977) The theory of local public goods, in Feldstein M.S., Inman R.P. (eds.), *The economics of public services,* Macmillan, London.

Strauss R.P., Hughes G.D. (1976) A new approach to the demand for public goods, *Journal of Public Economics, 6,* 3, 191-204.

Stull W.J. (1974) Land use and zoning in an urban economy, *American Economic Review, 64,* 3, 337-347.

Talley W.K. (1974) Optimality and equity in the provision of public goods by a polycentric political system, *Southern Economic Journal, 41,* 2, 220-227.

Tapiero C.S. (1978) Optimal location - size of a facility on a plane with interaction effects of distance, *European Journal of Operational Research 2,* 2, 107-115.

Teitz M.A. (1968) Toward a theory of urban public facility location, *Papers, Regional Science Association, 21,* 35-51.

Thrall G.I. (1979) Public goods and the derivation of land value assessment schedules with a spatial equilibrium setting, *Geographical Analysis, 11,* 1, 23-35.

Thrall G.I., Casetti E. (1978) Local public goods and spatial equilibrium in an ideal urban centre, *Canadian Geographer, 12,* 4, 319-333.

Thurow L.C. (1971) The income distribution as a pure public good, *Quarterly Journal of Economics, 85,* 2, 327-336.

Thurow L.C. (1973) The income distribution as a pure public good: response, *Quarterly Journal of Economics, 87,* 316-319.

Tideman T.N., Tullock G. (1976) A new and superior process for making social choices, *Journal of Political Economy, 84,* 6, 1145-1159.

Tideman T.N. (ed.) (1977) *Public choice,* Vol. 29, No. 2, Special Supplement.

Tiebout C.M. (1956) A pure theory of local expenditures, *Journal of Political Economy, 64*, 416-424.

Tiebout C.M. (1961) An economic theory of fiscal decentralization, in Margolis J. (ed.): *Public finance: needs, sources and utilization,* National Bureau of Economic Research, New York.

Tollison R.D. (1972) Consumption sharing and non-exclusion rules, *Economica, 39,* 276-291.

Tollison R.D., Willett T.D. (1978) Fiscal federalism: a voting system where spillovers taper off spatially, *Public Finance Quarterly, 6,* 3, 327-342.

Topham N. (1977) Consumer mobility and the distribution of local public goods, *Public Finance, 32,* 2, 254-266.

Tullock G. (1965) *The politics of bureaucracy,* The Public Affairs Press, Washington, D.C..

Tullock G. (1969) Federalism: problems of scale, *Public Choice, 6,* 19-29.

Tullock G. (1970) *Private wants, public means,* Basic Books, New York.

Urban Systems Group (1976) Public goods with consumption indivisibility, *Regional Science and Urban Economics, 6,* 1, 45-50.

Vardy D.A. (1971) Efficiency in the supply of regional public goods that are subject to congestion, Discussion Paper 65, Institute for Economic Research, Queen's University, Kingston.

Vardy D.A. (1972) Intergovernmental transfers and Pareto optimality, *Finanzarchiv, 31,* 1, 68-88.

Vardy D.A. (1973) Population mobility and efficiency in the provision of regional public goods, in Proceedings of the 28th Congress of the International Institute of Public Finance, *Issues in Urban Public Finance.*

Wagner J.L., Falkson L.M. (1975) The optimal nodal location of public facilities with price sensitive demand, *Geographical Analysis, 7,* 1,69-84.

Warren R. (1964) A municipal services market model of metropolitan organizations, *Journal, American Institute of Planners, 30,* 193-204.

Westhoff F. (1977) Existence of equilibria in economies with a local public good, *Journal of Economic Theory, 14,* 1, 84-112.

Weisbrod B.A. (1964) Collective consumption services of individual consumption goods, *Quarterly Journal of Economics, 68,* 471-477.

Weisbrod B.A. (1965) Geographic spillover effects and the allocation of resources to education, in Margolis J. (ed.) *The public economy of urban communities,* Johns Hopkins Press, Baltimore.

Weymark J.A. (1979) Optimality conditions for public and private goods, *Public Finance Quarterly, 7,* 3, 338-351.

Wheaton W.C. (1975) Consumer mobility and community tax bases: the financing or local public goods, *Journal of Public Economics, 4,* 377-384.

White A.N. (1979) Accessibility and public facility location, *Economic Geography, 55,* 1, 18-35.

Wildasin D.E. (1979) Local public goods, property values and local public choice, *Journal of Urban Economics, 6,* 521-534.

Williams A. (1966) The optimal provision of public goods in a system of local governments, *Journal of Political Economy, 74,* 18-33.

Williams H.C.W.L. (1976) Travel demand models, duality relations and user benefit analysis, *Journal of Regional Science, 16,* 147-166.

Winch D.M. (1973) The pure theory of non-pure goods, *Canadian Journal of Economics, 6,* 2, 149-163.

Wooders M.H. (1978) Equilibria, the core and jurisdictional structures in economies with a local public good, *Journal of Economic Theory, 18,* 328-348.

Wooders M.H. (1980) The Tiebout hypothesis: near optimality in local public goods economics, *Econometrica, 48,* 6, 1469-1486.

Wright C. (1977) Financing public goods and residential location, *Urban Studies, 14,* 51-58.

Ylvisaker P. (1959) Some criteria for a «proper» areal division of governmental powers, in Maass A. *Area and power,* The Free Press, New York.

Young D.R. (1971) Institutional change and the delivery of urban public services, *Policy Sciences, 2,* 4, 425-438.

Young D.R. (1976) Consolidation or diversity: choices in the structure of urban governance, *American Economic Review, 66,* 2, 378-385.

Zeckhauser R. (1973) Determining the qualities of a public good - a paradigm on town park location, *Western Economic Journal, 11,* 39-60.

Résumé. Un bon nombre de modèles opérationnels de localisation-allocation existent pour la localisation optimale de systèmes de services. On retient que beaucoup d'entre eux pourraient être appliqués aux problèmes de services publics. Cependant très peu d'entre eux ont jamais été utilisés en pratique. Une raison fondamentale pour ceci est que ces modèles ne sont pas appuyés d'une théorie rigoureuse.
En effet, la littérature n'a pas problematisé une théorie de la localisation de services publics en général. On a manqué de reconnaître la nature publique/politique/institutionnelle du problème. La théorie de bien être economique des bien publics s'occupe de types de services et de biens fournis à travers les services publics.
Une généralisation spatiale de cette théorie appliquée aux biens publics spatiallement impurs peut servir comme fondement rigoureux d'une théorie de localisation de services publics. Néanmonis, la théorie de localisation doit être conçue comme une partie intégrante d'une théorie plus générale de l'economie de l'espace public et le rapport entre localisation et d'autres variables clé, souvent d'une ordre supérieur, doit être examiné.
Les tâches sont nécessaires pour pouvoir construire une nouvelle génération de modèles de localisation qui soient opérationnels et pertinents. En amplifiant ce sujet les modèles du type conventionnel sont critiqués, la théorie de biens publics pur et impurs est examinée et généralisée, et certaines demandes clé à propos de la théorie de l'économie de l'espace public sont présentées est les conséquences pour des modèles opérationnel de la théorie de localisation sont examinées. Cette étude est largement non-technique et ne présente pas de prototype pour les nouveaux modèles dont on par le dans l'article.

Riassunto. Esiste un gran numero di modelli operazionali di localizzazione-allocazione per localizzare in modo ottimale sistemi di servizi. La maggior parte di questi, si ritiene, siano adatti per essere usati nei problemi dei servizi pubblici, ma tuttavia solo pochi sono stati applicati praticamente. Il motivo sta nel fatto che questi modelli non sono sostenuti da una rigorosa teoria. Infatti, la letteratura non ha «problematizzato» una teoria di localizzazione dei servizi pubblici in generale. Non è stata riconosciuta la natura pubblica/politica/istituzionale del problema. La teoria economica del benessere dei beni pubblici si occupa dei tipi di beni e di servizi forniti tramite servizi pubblici. Una generalizzazione spaziale di questa teoria applicata ai beni pubblici, ma spazialmente impuri, può servire come fondamento rigoroso di una teoria di localizzazione dei servizi pubblici. Tuttavia la teoria di localizzazione deve essere concepita come parte di una teoria più generale dell'economia dello spazio pubblico e deve essere esaminato il rapporto tra localizzazione ed altre variabili chiave, spesso di un ordine superiore.
Questi sviluppi sono necessari per la costruzione di una nuova generazione di modelli operazionali di localizzazione. Nello sviluppare questo argomento, si criticano i modelli convenzionali, la teoria di beni pubblici puri ed impuri ed il modello è esaminato e generalizzato e certe domande chiave che vengono poste dovrebbero essere indirizzate alla teoria dell'economia dello spazio pubblico; le loro conseguenze sono esaminate per la teoria di localizzazione dei modelli operazionali.
Questo articolo è largamente non-tecnico e non vengono proposti prototipi dei nuovi modelli qui trattati.

Sistemi Urbani, 3, 1981, 391-401 Guida Editori, Napoli, Italia

Some new sources of instability and oscillation in dynamic models of shopping centres and other urban structures

A. G. Wilson

School of Geography, University of Leeds, Leeds LS2 9JT, England.

Received 15 June 1980

Abstract. May and others have shown that simple non-linear difference equations can exhibit very complicated dynamic behaviour. These results and associated methods are briefly summarised. It is then shown that they offer new insights into the dynamics of shopping centre developments both in respect of these being modelled by difference equations and when they are modelled using differential equations which are then integrated numerically. The methods are applied to two different dynamic shopping models and, in a concluding section, some speculations are presented on the effect of these ideas on more complicated and realistic models.

Key words: dynamic models, instability, oscillations, shopping centres, non-linear systems.

1. Complicated dynamics and simple non-linear difference equations

May (1974, 1975, 1976, May and Oster, 1976) present a number of interesting results relating to first order difference equations of the form

$$X_{t+1} = F(X_t) . \tag{1}$$

The same kinds of results can be obtained for a variety of functions, F, but here we use the main example of his 1976 paper which turns out to be directly applicable to shopping model dynamics. First, however, we comment on the distinctions involved between modelling dynamical systems through difference equations or differential equations. If the main state variables are changing continuously, then differential equations are appropriate; if the events can be considered to be discrete, then difference equations offer the correct formulation. In ecology, if populations have relatively long lives relative to the time periods of analysis, then differential equations represent the correct formulation. If, however, generations do not overlap, but the new populations are still dependent on those of the previous time periods, as with insects, then the model should be formulated in terms of difference equations.

In the shopping centre case, models have been presented in terms of differential equations (Wilson, 1976; Harris and Wilson, 1978) but have been simulated in terms of difference equations (White, 1977, 1978). It turns out that a difference equation formulation may be more appropriate in some circumstances, but we reserve this discussion for later and now concentrate on May's example. Let

$$N_{t+1} = N_t(a - bN_t) \tag{2}$$

be a first order difference equation describing the growth of a population, N. t is a time subscript and a and b are constants. This is one possible difference-equation equivalent of the logistic equation of growth. By the transformation

$$X = bN/a \tag{3}$$

it can be written in the more convenient form

$$X_{t+1} = aX_t(1 - X_t) \tag{4}$$

X can then be considered to vary between 0 and 1, though if X ever exceeds 1, it then diverges to $-\infty$ and the negative numbers are unrealistic in many applications. However, we will ignore this complication here and assume the value of a and the initial condition avoid this. It can always be avoided by using the alternative form of logistic equation

$$X_{t+1} = X_t \exp[r(1 - X_t)] \tag{5}$$

though this is more difficult to handle.

The relation between X_{t+1} and X_t can be plotted as a humped curve, as in fig. 1. It attains a maximum at $X = 1/2$ of $a/4$, and since X must remain less than 1, this implies $a < 4$. We also require $a > 1$, or $X_{t+1} \rightarrow 0$ for large t. Thus, for non-trivial dynamic behaviour

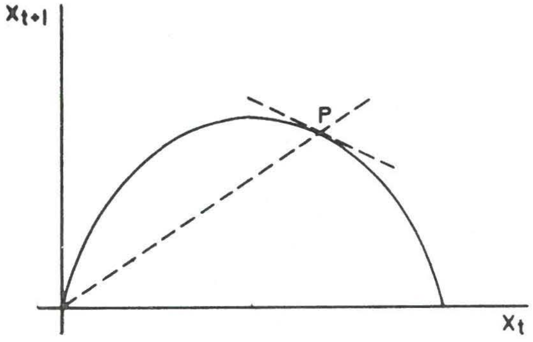


Figure 1

$$1 < a < 4 . \tag{6}$$

The possible equilibrium values of X are found by putting $X_{t+1} = X_t$ in equation (4). This is equivalent to seeking the intersection of the humped curve and the 45° line which is also plotted on fig. 1. Thus P is an equilibrium point.

Let $X = X^*$ be the equilibrium point. For later notational convenience, we also write equation (4) in the form of equation (1) with

$$F(X) = aX(1 - X) . \tag{7}$$

At equilibrium,

$$X_{t+1} = X_t = X^* \tag{8}$$

and so equation (4) gives

$$X^* = aX^*(1 - X^*) \tag{9}$$

which has the non-zero solution (for P)

$$X^* = (a - 1)/a . \tag{10}$$

The slope of the curve at this point is

$$\left.\frac{dF}{dX}\right|_{X = X^*} = a - 2aX^* \tag{11}$$

which is

$$\left.\frac{dF}{dX}\right|_{X = X^*} = 2 - a \tag{12}$$

[substituting from (10)].

Consider points within a small increment ±Δ on either side of X*, as in fig. 2. If the slope of the line is between ±1, and if

$$X_t = X^* + \Delta , \qquad \text{say} \tag{13}$$

then

$$X^* < X_{t+1} < X_t \tag{14}$$

which means that the equilibrium point is stable. Otherwise it is unstable. Since we know the slope to be $2-a$ from (12), for this to lie between ± 1 we must have

$$1 < a < 3 \tag{15}$$

for stability.

We saw earlier that we must have $a > 1$ anyway for non-trivial behaviour, so the interesting condition is $a < 3$. We should note that as a increases, the hump in fig. 1 steepens and it is easy to see that a point will be reached when the modulus of the tangent at the equilibrium point exceeds 1. This occurs when $a = 3$. When $a > 3$, the equilibrium solution X^* becomes unstable. It is then possible to see if there is another kind of equilibrium point two time periods apart; that is, satisfying

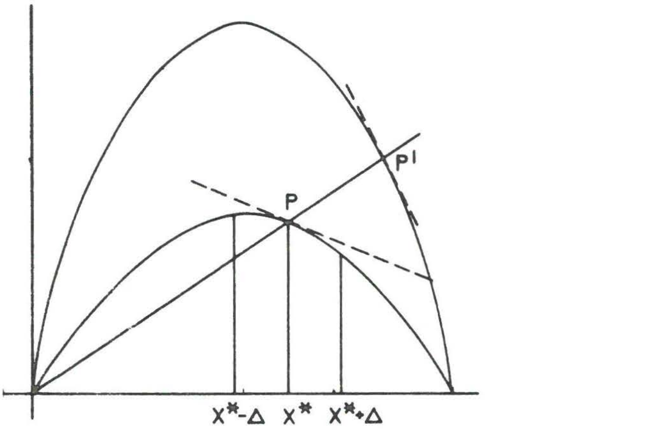


Figure 2

$$X_{t+2} = F[F(X_t)] \, . \tag{16}$$

If X_{t+2} is plotted against X_t, the curve has two humps. Three cases are shown in fig. 3. Case (a) has $a < 3$. There is only one equilibrium point and it is stable. Case (b) shows the 45° line touching the curve. This represents the limiting case, $a = 3$. In case (c), $a > 3$, the tangent at the original equilibrium point now exceeds 45° and is unstable, but there are two new equilibrium points, $X^{(2)*}$ and $X^{(2)**}$. However, there is then another critical value of a which bifurcates into a four-cycle of stable points, then eight and so on, as shown in fig. 4. Beyond a

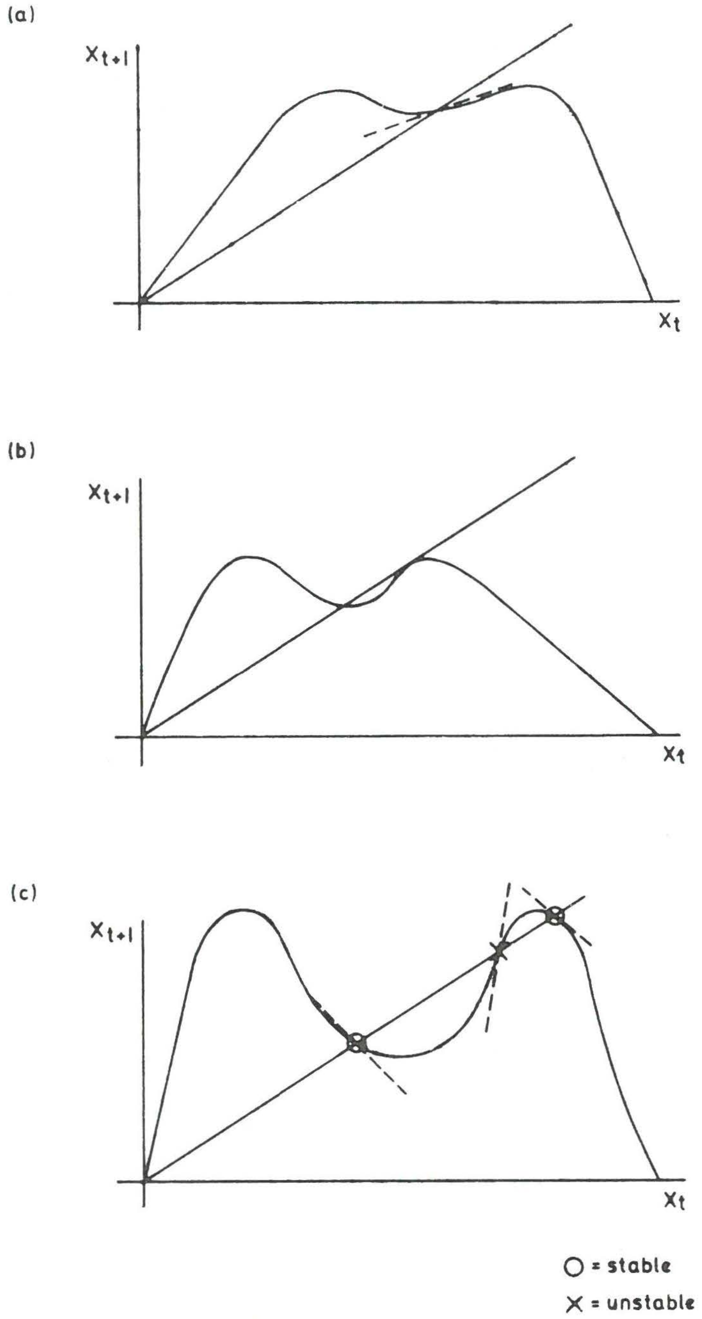


Figure 3

value a_c, the behaviour becomes chaotic. That is, it oscillates without any observable periodic structure.

We now explore how to apply these results, and these methods of analysis, to models of shopping centre dynamics.

2. Model 1: linear growth

Consider now a set of shopping centres across a set of discrete zones. Let the size of the centre in zone j be W_j. Then, if D_j is the revenue potentially attracted to j, a suitable differential equation for the growth of W_j is (Wilson, 1976)

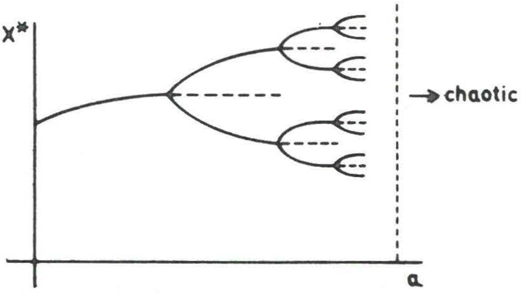


Figure 4

$$\dot{W}_j = \varepsilon(D_j - kW_j) \tag{16}$$

for suitable constants ε and k. The difference equation form which suggest itself is

$$W_{jt+1} - W_{jt} = \varepsilon(D_j - kW_{jt}) \tag{17}$$

(where, without loss of generality, the time period is taken as one, or as a factor merged into ε). This can be written

$$W_{jt+1} = \varepsilon D_j + (1 - \varepsilon k) W_{jt} . \tag{18}$$

Although this is a linear first order equation, and therefore does not have the interesting bifurcation properties of May's examples, we can apply his methods. Equation (18) expresses a linear relationship between W_{jt+1} and W_{jt} and an equilibrium will be at the intersection of this line and the 45° line

$$W_{jt+1} = W_{jt}. \tag{19}$$

Various examples are shown in fig. 5. In relation to stability of equilibrium, the same argument applies as before: the slope of the «curve» is now of course the gradient of the line and if this is between ± 1, then any intersection in stable - the argument of (13) and (14) above still holds. Four cases are distinguished on fig. 5: (a) and (c) are stable equilibrium points; in case (b), there is no equilibrium point with positive W_j; and in case (c), the equilibrium point is unstable. We can collect these results together in terms of the gradient of the line:

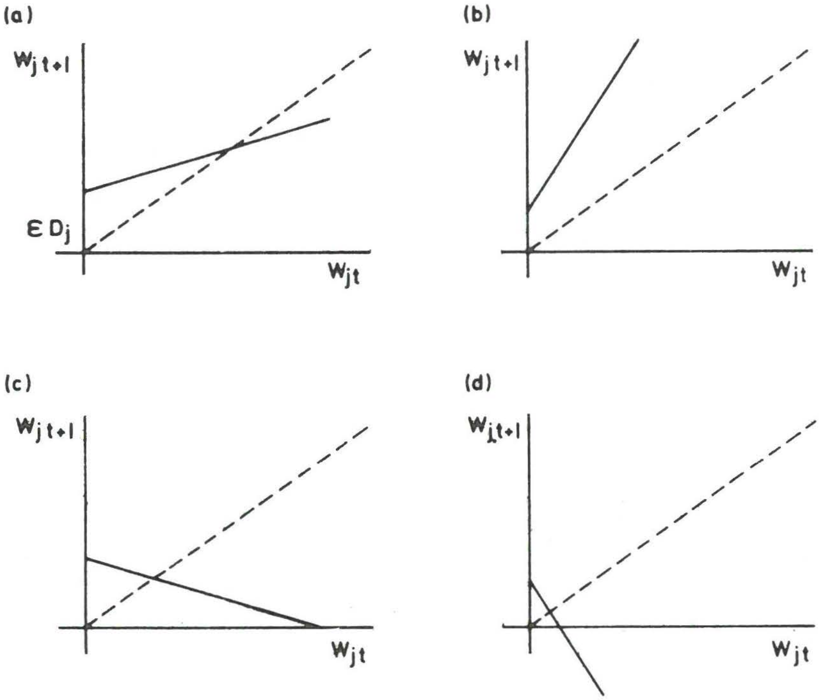


Figure 5

(a) $\quad 0 < 1 - \varepsilon k < 1 \quad$ (20)

(b) $\quad 1 < 1 - \varepsilon k \quad$ (21)

(c) $\quad -1 < 1 - \varepsilon k < 0 \quad$ (22)

(d) $\quad -1 > 1 - \varepsilon k \ . \quad$ (23)

The interesting and new feature about these relationships is that results about stability are related to general conditions involving two of the parameters in the model (18). Case (b) is immediately seen to be geographically nonsensical: it implies $\varepsilon k < 0$ when both of these parameters should be positive. (a) or (c) will hold provided the product εk is sufficiently small. indeed, combining (20), (22) and (23), and assuming $\varepsilon k > 0$, the condition can be restated as

$$\varepsilon k < 2 \tag{24}$$

for stability, and

$$\varepsilon k > 2 \tag{25}$$

for instability. This also gives some clue as to the nature of the instabilities in difference equations. They arise because of the time lags involved in responding to a change. The greater the values of ε or k, the more rapid is the change from period to period and the more difficult it is to get back to equilibrium through feedback.

This analysis has been conducted as though D_j was fixed. In practice, of course, it is not and is given by

$$D_j = \sum_i S_{ij} \tag{26}$$

$$= \sum_i \frac{e_i p_i W_j^\alpha e^{-\beta c_{ij}}}{\sum_k W_k^\alpha e^{-\beta c_{ik}}} \tag{27}$$

since S_{ij}, the flow of revenue from residents of i to shops in j, is given by

$$S_{ij} = \frac{e_i p_i W_j^\alpha e^{-\beta c_{ij}}}{\sum_k W_k^\alpha e^{-\beta c_{ij}}} \tag{28}$$

e_i is per capita expenditure at i, P_i the population of i and c_{ij} the cost of travel from i to j. α and β are constants. In a previous analysis of equilibrium and stability, the focus has been on the stability of the equilibrium value, once it has been achieved (Harris and Wilson, 1978). Equations (16) or (17) show that the equilibrium point is

$$W_j^* = D_j / k\ . \tag{29}$$

It was shown by Harris and Wilson that the stability of equilibrium depends on the values of the parameters like α, β and k. Here, we have seen that if D_j can be assumed constant, there is an additional condition (24). This can perhaps be interpreted as follows: if an equilibrium value of D_j is calculated using Harris and Wilson (1978) methods, say as D_j^{equil}, then the ability to achieve a stable equilibrium in a simulation will require (24) to hold. Since it is a condition on parameters which are not j-dependent, this presumably means there will be difficulties in simulation in any cases where it is not satisfied. White (1979), for example, has reported simulations of this type which have not converged.

3. Model 2: logistic growth

The model given by (16) and (17) implies a steep rate of growth for W_j from a $W_j = 0$ starting point. This can be slowed down at the origin, but still bounded above, by adding a factor W_j. Equation (16) then becomes

$$\dot{W}_j = \varepsilon(D_j - kW_j)W_j \ . \qquad (30)$$

This does not, of course, change the position of the equilibrium point which is still given by (29). We saw in section 1 that there are at least two versions of difference equations which approximate logistic growth and we work with the one given by equations (2) and (4). The obvious modification of equation (30) is to give

$$W_{jt+1} - W_{jt} = (D_j - kW_{jt})W_{jt} \qquad (31)$$

which can be written

$$W_{jt+1} = [(1 + \varepsilon D_j) - \varepsilon k W_{jt}]W_{jt} \ . \qquad (32)$$

This is of the same form as equation (2), and if we write

$$X_j = \frac{\varepsilon k W}{(1 + \varepsilon D_j)} \qquad (33)$$

then the equation takes the canonical form (4) with

$$a = 1 + \varepsilon D_j \ . \qquad (34)$$

We can then immediately apply May's results on stability. Note that while D_j has the «dimension» of money, equation (31) shows ε to have the dimension of $(\text{money})^{-1}$, and so εD_j is a dimensionless constant. The «hump» of the curve in fig. 1 will be steeper for increasing values of either ε or D_j.

A recap of section 1 shows that we require

$$1 < 1 + \varepsilon D_j < 3 \tag{35}$$

for a stable single equilibrium point (using (b)), which is obviously

$$0 < \varepsilon D_j < 2 \ . \tag{36}$$

Clearly εD_j is always positive, but not necessarily less than 2. As it exceeds 2, then there is first a two period cycle, then a four-period one up to a chaotic regime which sets in at $a = 3.8495$, or $\varepsilon D_j = 2.8495$. We should also recall that the system goes into divergent oscillations if $a > 4$, or $\varepsilon D_j > 3$.

As with model 1, D_j has been treated as a constant in this analysis. Again, a suitable first guess at it would be D_j^{equil} as predicted by the Harris and Wilson (1978) procedure. It is also more interesting in this case that the stability condition is j dependent, and that through D_j^{equil} it is dependent on the effects of any changes in other zones. This suggests the possibility of very complicated dynamic behaviour for a whole system which is evolving through the difference equation (32).

The periodic, chaotic or divergent behaviour which results from εD_j exceeding 2 can arise in two ways which would need to be sorted out in particular empirical cases. First, since ε implicitly contains the time step length, it means that if this is too large there will be problems arising from such a (technical) choice. This means that special care will have to be taken if discrete simulation involving the logistic equation are used - as for example in the work of Allen and Sanglier (1979). Secondly, the instabilities arise in a real sense because the implied feedback of the decision maker which is represented in the discrete nature of the difference equation formulation and it becomes a matter of empirical investigation as to whether these exist or not.

4. Concluding comments

May (1976) has shown that very simple difference equations exhibit very complicated dynamic behaviour and he suggested a number of fields where the results were potentially applicable. We have shown in this paper that they appear to have a direct application in geographical dynamics. It is perhaps a coincidence that the correspondence of equation (3) with May's example is so exact, and of course this involves the restrictive condition that D_j should be treated as a constant. What will be even more interesting will be to explore the consequences of these kinds of bifurcation phenomena in more complicated economic models. For example, a retail model might be linked to a residential location model (Wilson, 1980) and this would, through the P_i's in equation (27), have an impact on the D_j's. For particular values of ε, a «jump» in D_j resulting from a P_i change could then lead, say, to new periodic behaviour in W_j. It is also clear that, though the main argument has been cast in terms of shopping centres, the methods and principles are more widely applicable to other urban structures. There is also beginning to be an extension of May's ideas to interacting populations in ecology - see for example Beddington, Free and Lawton (1975) on the investigation of dynamic complexity in prey-predator equations. There is much scope for numerical and empirical experiment and investigation.

References

Allen P. M., Sanglier M. (1979) A dynamic model of growth in a central place system, *Geographical Analysis, 11*, 256-272.

Beddington J. R., Free C. A., Lawton J. H. (1975) Dynamic complexity in predator-prey models formed in difference equations, *Nature, 255*, 58-60.

Harris B., Wilson A. G. (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models, *Environment and Planning A, 10*, 371-388.

May R. M. (1974) Biological populations with non-overlapping generations: stable points, stable cycles and chaos, *Science, 186*, 645-647.

May R. M. (1975) Biological populations obeying difference equations: stable points, stable cycles and chaos, *Journal of Theoretical Biology, 51*, 511-524.

May R. M. (1976) Simple mathematical models with very complicated dynamics, *Nature, 261*, 459-467.

May R. M., Oster G. F. (1976) Bifurcation and dynamic complexity in simple ecological models, *The American Naturalist, 110*, 573-599.

White R. W. (1977) Dynamic central place theory: results of a simulation approach, *Geographical Analysis, 9*, 226-243.

White R. W. (1978) The simulation of central place dynamics: two sector systems and the rank-size distribution, *Geographical Analysis, 10*, 201-208.

Wilson A. G. (1976) Retailers' profits and consumers' welfare in a spatial interaction shopping model, in Masser I. (ed.) *Theory and practice in regional science*, Pion, London, 42-59.

Wilson A. G. (1980) Aspects of catastrophe theory and bifurcation theory in regional science, *Papers, Regional Science Association, 44*, 109-118.

Riassunto. È noto come alcune semplici equazioni non lineari alle differenze finite possano avere una soluzione dal comportamento dinamico assai complesso. In questo saggio viene mostrato come tali fenomeni sorgono naturalmente nel tentativo di spiegare la dinamica dell'assetto spaziale dei centri commerciali. Alcuni modelli discreti e continui (equazioni differenziali) vengono analizzati; infine, possibili generalizzazioni di tale metodo di analisi a sistemi più complessi vengono discusse.

Résumé. On sait que quelques simples équations non linéaires aux différences finies peuvent avoir une solution dont le comportement dynamique est très complexe. Cet essai montre comme tels phénomènes se produisent naturellement en essayant d'expliquer la dynamique de la configuration spatial des centres commerciaux. On analyse quelques modèles discrets et continus (equations différentielles); et on discute ensuite quelques possibles généralisations de telle méthode d'analyse pour des systemes plus complexes.

Sistemi Urbani, 3, 1981, 403-413 Guida Editori, Napoli, Italia

Continuous models of transportation and location

M. J. Beckmann

Department of Economics, Brown University, Arlington Ave., Providence, Rhode Island 02912, USA.

Received 15 June 1980, in revised form 20 September 1980

Abstract. In this paper spatial interaction is modelled in terms of a continuous flow field. Section 1 presents a general framework. Section 2 formulates the allocation problem for a discrete set of given facility locations and a continuous distribution of customers. Section 3 introduces capacity limitations. Section 4 derives a cost function for the service density. Section 5 formulates the optimum facility location problem in terms of this density and solves two simple cases. In conclusion we mention some worth-while problems for further exploration.

Key words: continuous models, transportation and location, facility location.

1. Introduction: the continuous model of transportation

Consider a region in which facilities for a given service (e.g. stores offering a given set of goods) and its customers are widely dispersed. We may represent supply and demand in terms of continuous density distributions in a subset of the Euclidean plane. In general it will not be economically advantageous to limit customers to locally available facilities. Thus the need for moving customers to service locations arises. While it is conceivable that the resulting motion is highly irregular and includes the possibility of some customers from a given location going to every other location for service and of some service locations serving customers from every other location; an efficient arrangement will minimize the amount of such cross hauling. Let us in fact assume that the minimum amount of movement takes place consistent with 2 requirements:

each customer receives a certain level of service regardless

of his/her residential location

no service facilities over-utilized.

We are then dealing with the situation described by the *continuous model of transportation* (Beckmann, 1952; Beckmann, Puu, forthcoming). Amount of service received must here be interpreted as number of trips to a service facility. For location (x_1, x_2) and its customers let the area density of trips originated in this way per unit time be denoted by

$$q(x_1,\ x_2)\ .$$

Let the density of service available be similarly denoted by

$$s(x_1, x_2) .$$

The excess demand

$$q - s$$

gives rise to a net outflow of person trips. Denote the flow of trips by $\Phi(x_1, x_2)$. Thus we assume that there is only one direction of movement at any given point. The strenght of the flow field $|\Phi|$ is the number of trips passing through a unit cross section per unit time in the direction of Φ.

By a well-known argument from fluid or thermodynamics this net outflow is equal to

$$\operatorname{div} \Phi = \frac{\partial \Phi_1}{\partial x_1} + \frac{\partial \Phi_1}{\partial x_2} .$$

We have here a first relationship between the given demand for and supply of service and the flow of customers, their trips,

$$\operatorname{div} \Phi = q - s . \tag{1}$$

In the case of commodity flows one can set up a relationship between the commodity price at various locations and the commodity flow. In the case of a service which is of uniform quality everywhere the same argument may be applied. The situation is more complex if the quality of service is considered to be a function of location. This problem will be addressed below in the case of discrete sources of supply of service. Here we retain the assumption of a uniform quality. If a competitive market has been organized for this service then the local price of the service and the flow of customers must be related. Customers have a motivation to go elsewhere for service if and only if the saving in service price makes up for the costs of transportation.

Let

$$p = p(x_1, x_2)$$

denote the service price. Then

$$- \operatorname{grad} p$$

denotes the direction and amount of price decline per unit distance in the direction where the decline is largest. Compare this to the cost of movement. In the simple model considered here the cost is assumed isotropic (i.e. independent of direction) but it may depend on location. Denote it by

$$k = k(x_1, x_2).$$

In competitive market equilibrium the equality of price advantage and cost of movement assumes the form

$$k = |\text{grad } p|.$$

In fact a stronger statement can be made. The direction of movement must be that in which price falls steepest, the gradient direction. Thus

$$k \cdot \frac{\Phi}{|\Phi|} = - \text{grad } p. \tag{2}$$

In order that an equilibrium exists in a closed region aggregate supply must be at least equal to aggregate demand

$$\iint q \, dx_1 dx_2 \leqq \iint s \, dx_1 dx_2 . \tag{3}$$

On the boundary Γ flow must vanish in the direction of the normal n to the boundary

$$\Phi_n = 0 \quad \text{on } \Gamma. \tag{4}$$

Given (3) the conditions (1), (2) and (4) determine the direction of the flow field $\frac{\Phi}{|\Phi|}$ uniquely. In the case of equality between aggregate supply and aggregate demand, the level of prices p is undetermined. Interpreting p as a potential function we may say that the absolute level of the potential is indeterminate.

The fact that the movement of customers follows a gradient field indicates that there are no closed flowlines, i.e. no cross hauling. The flow lines will be straight, if and only if transportation cost k is uniform at all locations. Generally speaking movement will be from locations of excess demand to locations of excess supply.

A flow field in a simple connected region which is continuous along its boundary must have at least one singularity in its interior. The point singularities consistent with a gradient field are either points of confluence or of effluence. The first is associated with a local minimum of the potential function p, the second with a local maximum of p.

Potential curves are lines of equal p-value and show the customer locations where service costs are equal, or where access to the service is equally good. The potential lines are at right angles to the flow lines.

The equations (1), (2) are in fact the Euler-Lagrange equations solving the following minumum problem

$$\text{Min} \quad \iint_A k|\Phi| \, dx_1 dx_2 \tag{5}$$

subject to (1)
and to the boundary condition (4).

This means that individual optimization described by (1) and (2) also achieves a system's optimum (5). The prices $p(x_1, x_2)$ are identical with the Lagrange multipliers λ associated with the constraint (1) in a Lagrange function for the minimum problem

$$\iint - k|\Phi| + \lambda(\operatorname{div}\Phi - q + s) \; dx_1 dx_2 \, . \tag{6}$$

2. Discrete facilities at fixed locations

Consider now a single service facility at a given location (the center) serving customers whose demand for service $q(x_1, x_2)$ has a given continuous distribution function.

If transportation cost $k(x_1, x_2) = k$ is uniform then the flow lines are the radius vectors emanating from the center and the potential curves are concentric circles. (The same is still true when $k = k(r)$ depends only on the distance r from the center).

Consider next two service locations without capacity limit offering services perceived by all to be of equal quality. Then each customer is served best at the nearest facility and the flow lines will be two sets of radius vectors separated by the normal bisector of the line connecting the two service centers (fig. 1).

Continuing with three and more centers a set of market areas is generated surrounding each service center and bounded by line segments intersecting in triple corners. This is familiar from the location theory of market areas (Beckmann 1968, chapter 3).

The solution can be characterized by a potential function which is zero at each service center and has the same value at any boundary point between adjacent market areas no matter from which direction the point is approached. The last fact follows from the continuity of λ. The first is seen as follows.

Return to the formulation (5) and observe that in the case of unlimited service capacities, the source sink equation (1) has the form

(1a)	$\operatorname{div} \Phi = q$	outside service centers
(1b)	no constraint	at service centers.

The fact that there is no constraint implies that p vanishes at service centers.

The lines of constant p denote locations at which the service is equally costly. The lines give information about the accessibility of service to customers at their various locations. In particular a low level of p means a cheap service and hence good access. Of interest in judging how well the entire region is served are the extremes of p,

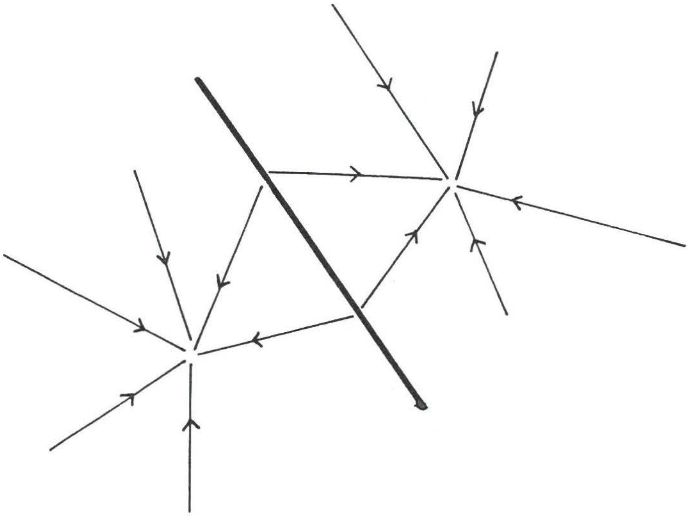


Figure 1

$$p_{\min} \quad \text{and} \quad p_{\max} .$$

Their difference $p_{\max} - p_{\min}$ indicates the range of variation in service cost to customers in the region.

The integral over all service cost is also of interest. Consider

$$\iint k|\Phi|\, dx_1 dx_2 = \iint (\text{grad } p)'\Phi \; dx_1 dx_2 \qquad \text{using (2)}$$

$$= \iint (\text{div}(p\Phi) - p \text{ div}\Phi)\, dx_1 dx_2$$

$$= -\iint p \text{ div}\,\Phi \; dx_1 dx_2 + \int (p\Phi)_n ds$$

by the Gauss integral theorem (Courant, John, 1965)

$$= -\iint p\,(q-s)\; dx_1 dx_2 \;. \qquad \text{using (1), (4)}$$

Thus finally

$$\iint k|\Phi|\; dx_1 dx_2 = \iint p\,(s-q)\; dx_1 dx_2 \;. \tag{6}$$

Observe that $p(x_1, x_2)$ is the negative of the cost of service to customers at location (x_1, x_2).

Equation (6) may also be written as a budget equation

$$\iint q\,(-p)\; dx_1 dx_2 = \iint s\,(-p)\; dx_1 dx_2 + \iint k|\Phi|\; dx_1 dx_2. \tag{7}$$

Total cost of service to customers	=	Total receipts of suppliers	+	Total transportation cost

3. Capacity limitations

Suppose that the two centers have limited capacities not matching the demand by nearest customers. Then the previous solution is invalid and the potential function will in general differ between service centers. Economically speaking the potential function p now contains a scarcity rent for the service facility.

The dividing lines between adjacent market areas are now hyperbolas.

To implement the solution, differential service charges have to be imposed in order to motivate customers to avail themselves of the right facilities so that the overall result is a system's optimum. Failure to set such charges requires either allocations by administrative dictum or in the case of free choice an uneconomical imbalance of supply and demand at the various centers. Equal access to service facilities is an impossibility anyway unless all demands are met locally (see below). Recall that p measures cost of service and will depend on location.

Charging differential fees (or giving differential subsidies) is merely an extension of the fact that different customer locations are differently served, to the service location themselves. If one wishes to combine efficiency in resource allocation which freedom of choice there seems to be no alternative to charging (or paying) differential prices at different facilities.

4. Density distribution of facilities

Turn now to the case of many service facilities. In the spirit of the continuous transportation model let us consider their density not at the facility itself but averaged over the areas served. Thus an activity level z at a point location gives rise to a density

$$s = \frac{z}{A}$$

where A is the area served. What is the cost function in terms of this density?

We assume a fixed cost F of the service and transportation cost proportional to distance and ignore the proportional service cost (which is strictly proportional to the given aggregate demand in the absence of local variation).

If a service intensity s is supplied by one facility for a circular region of radius R then total cost is

$$F + sk \int_{o}^{R} 2\pi r^2 \, dr$$

and cost per area is

$$\frac{F}{\pi R^2} + \frac{2}{3} k \, Rs \; . \tag{8}$$

Minimizing average cost i.e. finding the optimal R

$$R = \sqrt[3]{\frac{3Fs}{\pi k}} \; .$$

Substituting into the cost density function (8)

$$c(s) = \left(\frac{3F}{\pi}\right)^{\frac{1}{3}} k^{\frac{2}{3}} s^{\frac{2}{3}} .$$

The cost function is therefore of the type

$$c(x_1, x_2, s) = a(x_1, x_2)\, s^{\frac{2}{3}} \tag{9}$$

where the constant a depends on local conditions as they affect the fixed cost F of the service and the transportation cost k of customers.

Notice that his cost function is concave implying increasing returns to scale.

5. Towards an analysis of optimum density distribution

Consider now the problem of locating facilities optimally. The objective functions is the sum of production and transportation cost. Since local transportation cost is already included in the cost function

$$a(x_1, x_2)\, s^{\frac{2}{3}}$$

any additional transportation cost must be that of customers entering a local service area from elsewhere. Thus when a low cost service area is adjacent to a high cost area as shown by their respective $a(x_1, x_2)$ function, some flow will take place between these areas measured by a flow vector Φ. The additional transportation cost is therefore $k|\Phi|$.

Consider the following optimization problem

$$\underset{s,\Phi}{\text{Min}} \int\int a(x_1, x_2)\, s^{\frac{2}{3}} + k|\Phi|\; dx_1 \cdot dx_2$$

subject to (1) $\operatorname{div} \Phi = q - s$ in A

(4) $\Phi_n = 0$ on Γ

(10) $s \leq m$.

The last condition on permissible levels of density is needed to avoid singular solutions where a finite amount is produced in an infinitesimal area.

The Lagrangean of this problem is

$$L = \iint_A - a s^{\frac{2}{3}} - k|\Phi| + \lambda\,[s - q - \operatorname{div}\,\Phi] \tag{10}$$
$$+ \mu(m - s)\,dx_1 dx_2\ .$$

A set of necessary - but in general not sufficient - conditions are the Euler equations

$$s \left\{ \begin{smallmatrix} = \\ > \end{smallmatrix} \right\} 0 <=> \frac{2}{3}\, as^{-\frac{1}{3}} \left\{ \begin{smallmatrix} > \\ = \end{smallmatrix} \right\} \lambda - \mu \tag{11}$$

$$k\,\frac{\Phi}{|\Phi|} = -\operatorname{grad}\,\lambda. \tag{2}$$

Notice that the first condition can always be satisfied by setting $s = 0$. Clearly the equation (11) and (2) do not have unique solutions.

When $a(x_1, x_2) \equiv a$ and $q(x_1, x_2) \equiv q$ are uniform then the optimal solution is $s \equiv q$ so that no cross areal flows occur. (Still in the small customers must move themselves to the local facility).

To illustrate the problem in less trivial cases consider a one-dimensional situation. On the interval $(-1,0)$ costs are constant and low $= a_0$, and on the interval $(0,1)$ costs are high and constant $= a_1$.

Let the demand density be uniform $= q$. A good guess of a solution is that some customers will come from the high cost area to be served in the adjacent part of the low cost region where service is supplied at maximum density m (fig. 2). This results in prices

$$\lambda = \lambda_0 + kr \qquad -1 \leq r < 1 \tag{12}$$

The service level in the low cost region is then

$$s = \left[\frac{2}{3}\ \frac{a}{\lambda_0 + k(r - r_0 - \mu)} \right]^3 = m \qquad r_0 \leq r \leq 0 \tag{13}$$

implying

$$\mu = \mu_0 - kr\ . \tag{14}$$

Some export of services will occur from the high density area also into the left hand region. The service density s declines according to the law

$$s = \left[\frac{2}{3} \frac{a_0}{\lambda_0 - k(r - r_0)}\right]^3 \qquad -1 \leq r \leq r_0 \qquad (15)$$

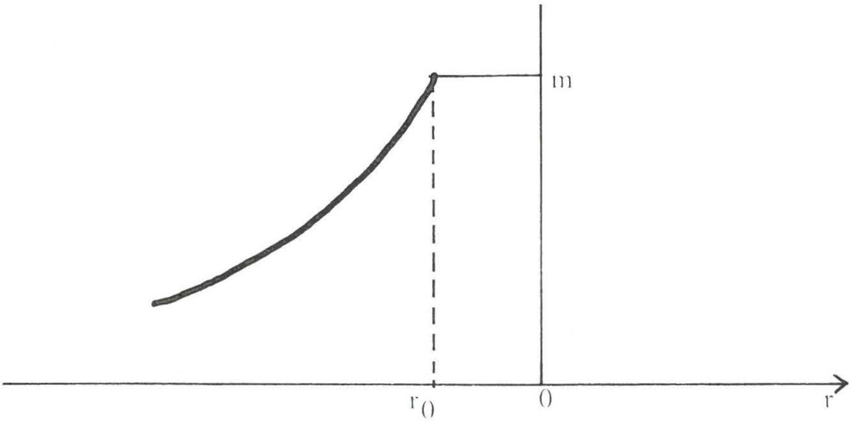


Figure 2

The point r_0 where the density of service begins to decline is determined by equating supply and given demand

$$2q = -r_0 m + \int_{-1}^{r_0} \left(\frac{2}{3} \frac{a_0}{\lambda_0 - kr}\right)^3 dr \qquad (3)$$

where λ_0 is given by

$$m = \frac{2}{3} \left(\frac{a_0}{\lambda_0 - kr_0}\right)^3 . \qquad (16)$$

In the right hand zone service levels are identically zero, provided the region is small enough.

6. Concluding remarks

This paper has studied only one of many scenarios and has merely given a sketch of the situation as seen in the context of continuous flow analysis.

For a comprehensive classification of facility location problems cf. Leonardi (1980). It would be interesting to study the more complex problems of welfare maximization rather than cost minimization which arise when demand is dependent on access, and the even more involved case when quality differences are perceived differently by different customers so that cross hauling will not only happen but will be economical. This raises the even broader issues of social versus private cost in facility location. To see whether and how the continuous flow model can contribute to an understanding of these important questions is a challenging task for the future.

References

Beckmann M. J. (1952) A continuous model of transportation, *Econometrica, 20*, 643-660.

Beckmann M. J. (1968) *Location theory*, Random House, New York.

Beckmann M. J., Puu T. (forthcoming) A continuous transportation model, Research reports, IIASA, Laxenburg.

Courant R., John F. (1965) *Introduction to calculus and analysis*, Interscience Publishers, New York.

Leonardi G. (1980) A unifying framework for public facility location problems, WP 80-79, IIASA, Laxenburg; forthcoming in *Environment And Planning A*.

Riassunto. In questo articolo l'interazione spaziale è modellizzata in termini di un campo continuo di flussi. Nella sezione 1 si presenta un quadro generale di riferimento e nella sezione 2 si formula il problema di allocazione per un insieme discreto di date localizzazioni degli impianti ed una distribuzione continua dei consumatori. Nella sezione 3 si introducono dei vincoli di capacità e nella sezione 4 si costruisce una funzione di costo per la densità dei servizi. Nella sezione 5, infine, il problema della localizzazione ottimale degli impianti è formulata in termini di questa densità e due semplici casi sono studiati. Nella parte conclusiva si presentano alcuni problemi di particolare interesse per ulteriori approfondimenti.

Résumé. Dans cet essai on formalise l'interaction spatial selon un champ continue de flux. La section 1 présente un aperçu général du problème et dans la section 2 on formule le problème d'allocation pour un ensemble discret de localisations des installations et une distribution continue des consommateurs. Dans la section 3 on introduit des contraintes de capacité et dans la section 4 on construit une fonction de coût pour la densité des services. En conclusion, dans la section 5 on formule le problème de localisation optimale des installations selon cette densité et on résout deux cas simples. Dans la partie finale on suggère des voies de recherche particulièrement intéressantes.

Sistemi Urbani, 3, 1981, 415-433 Guida Editori, Napoli, Italia

A theory of health care facility location in cities. Some notes

L. D. Mayhew

Human Settlements and Services Area, IIASA, International Institute for Applied Systems Analysis, Schlossplatz 1, Laxenburg, A-2361, Austria

Received 15 June 1980

Abstract. A theory of health care facility location in cities must address many questions. Some involving narrow allocation problems will be precisely stated; others will depend on ethical or clinical judgement. This paper considers only a small section of this spectrum covering a middle ground between the precise and the imprecise. The objective is not to produce an operational model whose outputs can be put to immediate use. It is to apply a simple and consistent rationale at different points in time to location patterns in one city, London. The arguments are general enough, however, for application elsewhere. The results are interpreted in terms of this rationale, and examples are given for developing more operational models from them. A point of departure from comparable approaches is that space is dealt with in a continuous way. This enables insights not possible using discrete methods that partition space into zones.

Key words: health care, facility location, continuous space, hierarchic structure, transformation.

1. The locational environment

1.1. *Basic characteristics*

The factors controlling the levels of acute health service provision are varied: they include the state of medical technology, the method of finance, the age-sex and income structure of the population, resource availability and, ultimately, societal values. How these factors combine to influence demand is a matter for dispute. At one extreme, demand is argued as being finite, in theory measurable, and so capable of independent assessment and analysis. At the other, it is considered wholly a function of supply, so that whatever is provided by health authorities gets used. Broadly speaking, the first position characterizes a market-based health care system, and the second, a free or planned system. Of course, in between there is a large gradation of systems types. Here, only the two extremes will be discussed.

1.2. *Urbanization*

Urbanization is still proceeding at a rapid rate in many countries. In developed countries this process is more or less complete, and the stock of health care facilities has been in position for many years. A complicating issue in these cases, however, has been the redistribution at lower densities of the population from the city centre to the periphery. This has resulted in a massive increase in the extent of the urban region, affecting greatly the nature and form of the health care facilities provided. The mobility of the population has also changed, creating a new set of travel preferences and locational attractions. For health authorities in these cities, the rapidity and diversity of these changes raises many practical, economic, social, and political issues. Because of the constraints they face, their problem is largely one of managing and investing in an aging stock of facilities (which to close, enlarge, or update) rather than in planning new facilities in new locations. For example, over 50% of the acute hospital stock in London was built and operating by 1900. Thus, the opportunities for large scale changes are small, and most of the adjustments that take place are piecemeal and, taken in isolation, economically very inefficient. The question is whether, from the changes in the sizes and distributions of facilities, any generalizations can be drawn: for example, does locational behavior produce in aggregate a coherent strategy aimed at servicing the urban population, or do the piecemeal changes talked of sum to nothing. The analysis commences with a framework for evaluating this and other problems.

2. The locational framework

2.1. *Districting*

Although many factors contribute to locational decisions, one which is always important is the subject of patient accessibility. How facilities are spaced in an urban region, however, clearly depends on the prevailing costs of travel. If these costs are high, it can be argued that locations will be chosen so that their catchment areas avoid extremes of distance; if they are low, then other factors will operate that give more weight to the size of the service population, so enabling a more cost-effective pattern of services.

An illustration of this point is shown in fig. 1, which shows the complications in cities caused by a variable population density. In the top half, an evenly populated city is partitioned into five equal-spaced sub-regions each served by an imaginary hospital (L_1 to L_5). It is important to notice that the population P serviced by each facility is the same while the dividing line d between each sub-region is co-terminous with the points of maximum travel (MC) and the total distance (TC) of the contained population from each facility. In the bottom half of fig. 1, in which a centre to periphery decline in density is shown, this property vanishes. Holding constant the locations, we note that (i) the population influenced by each centre decreases from left to right, (ii) the total distance of travel also declines because fewer people are travelling; and (iii) the intersections of the divides (d, d′, d′) under each criterion,

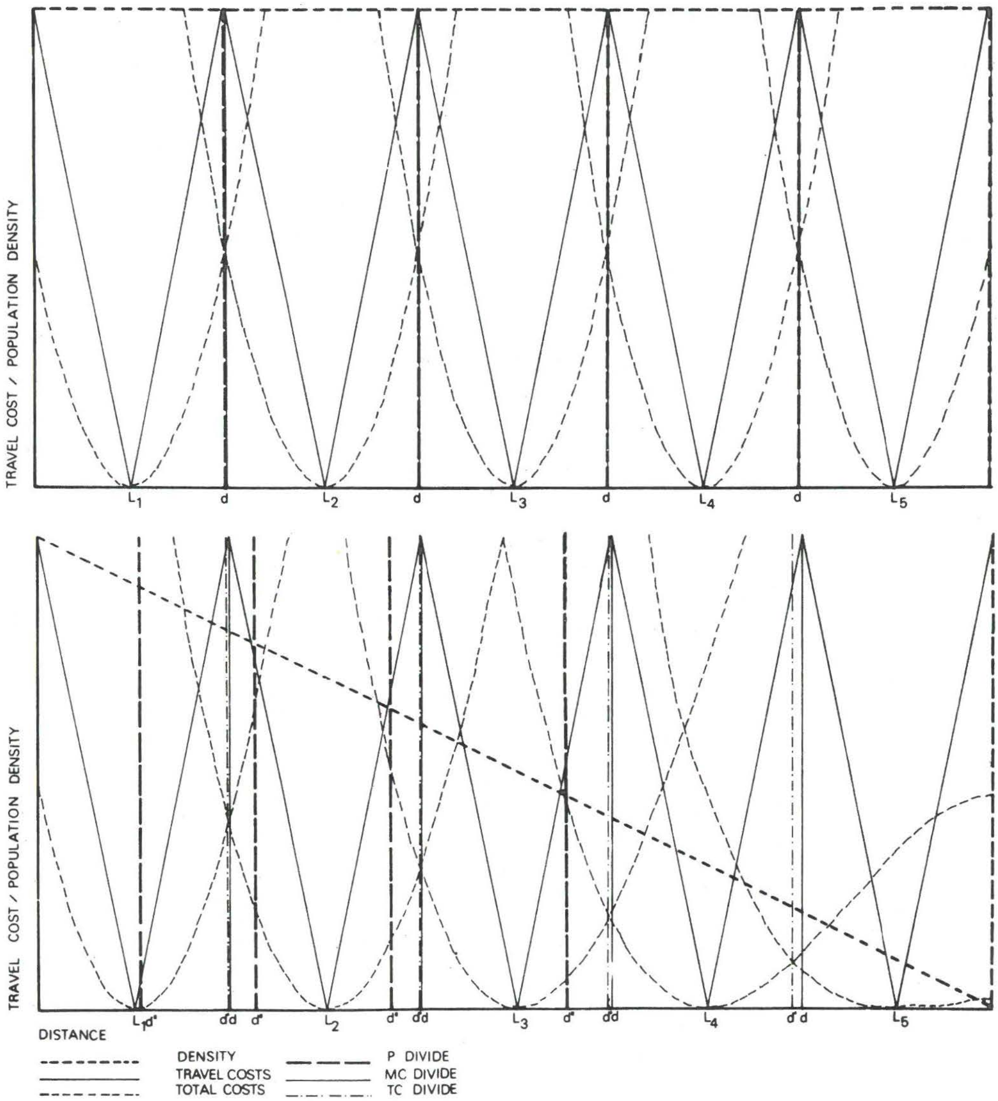


Figure 1 Urban districting: above) a unifamily populated region; below) a non-unifamily populated region. Shows how two regions are partitioned according to the three districting criteria, P, TC and MC

MC, TC, and P, become increasingly dislocated with distance from the centre.

Suppose now that the sizes of the facilities in each sub-region are proportioned in the uneven case to the contained population. Under P – districting (equal population), all facilities – measured, say, in terms of bed capacity – would be the same: under TC (equal total travel distance) and MC they would decline, the latter more rapidly by varying as the population density. Only in a uniformly populated region could the sizes associated with all three be the same.

We can make a case for each type of districting: P-districted facilities could be built to similar specifications to obtain the best economic returns to scale and other advantages of uniformity. The increased distance of travel at low densities, however, could be effective in reducing unit consumption, so introducing a measure of spatial inequality. TC-districted facilities take accessibility into account, but to lesser degree than MC-districted facilities for which the maximum distance in every sub-region is the same. Difficulties will be experienced with of these types of organization, however, in providing an economic mix of services at very low population densities.

2.2. *Hierarchies and transformations*

The exact pattern of densities varies between cities and between times. The recurrence of certain mathematical urban density functions (Clark, 1951), however, provides one common link. Similarities in the organization and functioning of health care facilities provide another (Shigan, Kitsul, 1980). The suggestion is, therefore, that the arrangement of facilities in different cities at the same time, or the same city at different times can be regarded as transformations of one another. The specification of the transformations is the ultimate goal of a dynamic spatial theory. The problem is difficult because it involves combinations of discrete and continuous processes (location decisions and population dynamics) plus uncertainty with regard to consumer behaviour. Nevertheless, certain transformations are easy to produce, and are relevant to the discussion.

In fig. 2 the districting principle is extended to the plane. Embedded in the geometry is a hierarchy of five levels organized on well-known lines (Christaller, 1933; Dietrich, 1977) for supplying services at varying intensity. In the hierarchy, there exists a centrally located facility, which in addition to supplying high order services throughout the region, also subsumes the functions provided by layers lower in the hierarchy. Facilities in lower layers are more numerous, but they attract patients from more limited areas. At the lowest level, a facility serves only the immediate locality, providing only those low order services that are in general demand and that are used most frequently. Finally, some facilities bordering the region share services in an unspecified way with neighbouring regions: their sub-regions are truncated by the urban perimeter.

Fig. 2 has two parts. In (b) a system is shown in which each level serves equal populations. The distortion of districts into curvilinear polygons is inevitable under such a transformation. This is hence P-districting: (a), on the other hand, is based on the MC principle. The grid super-imposed on (a) demonstrates how the system must bend to get from (b) to (a) either in time or between cities.

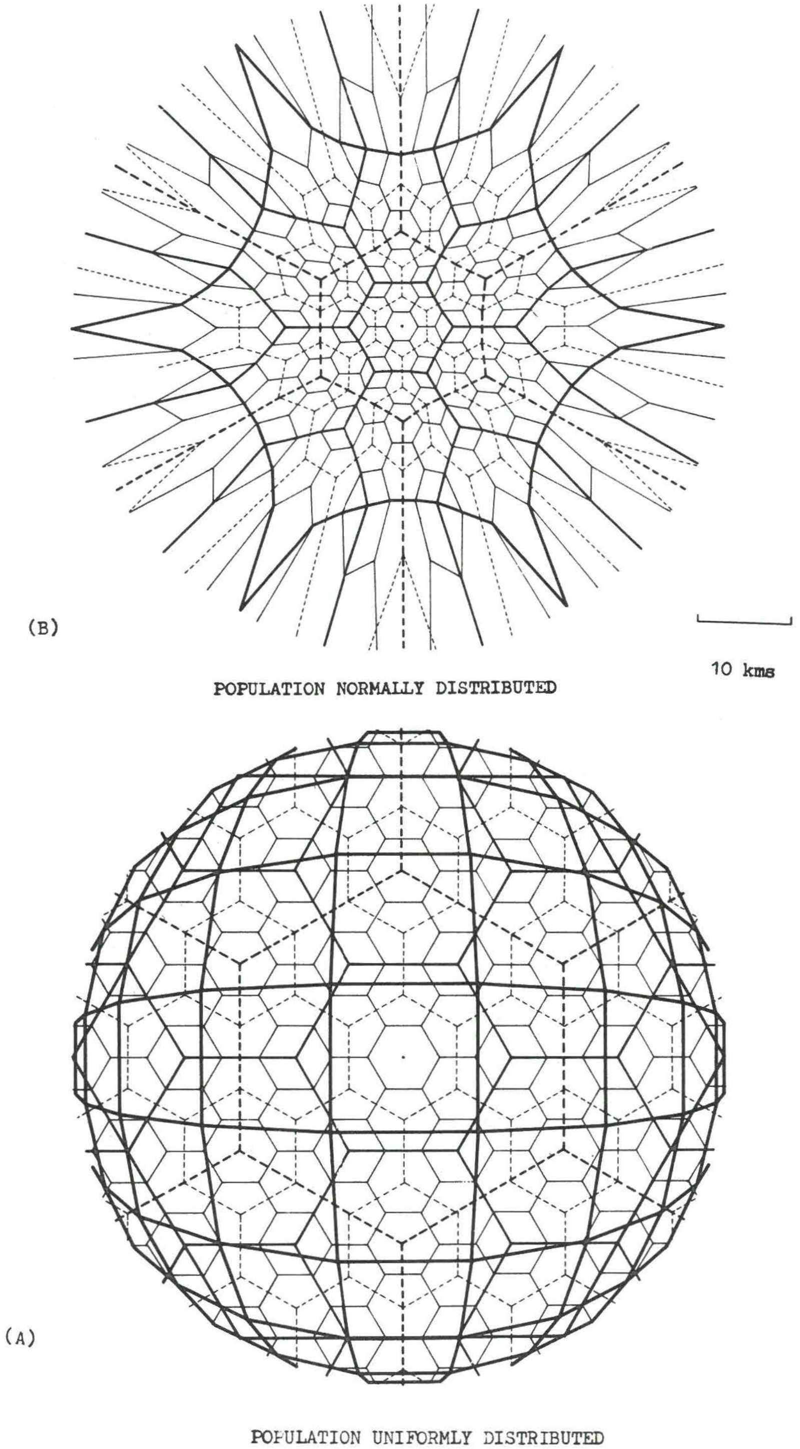

(B)
10 kms
POPULATION NORMALLY DISTRIBUTED
(A)
POPULATION UNIFORMLY DISTRIBUTED

Figure 2 Hierarchies and transformations: a) MC-districting; b) P-districting assuming a normally distributed population

The transforming function in fig. 1 is an equation fitted to the 1971 distribution of population in London. It is

$$D(r) = 92.767 \exp(-0.00354 r^2) \tag{1}$$

where D is the density in persons per hectare and r is the distance in kilometers from the city centre. The procedure for obtaining a P-transform, if cities are radially symmetric, is to set equal the integral of the density function in the region of interest to the integral over an image region, and then solve for r. That is

$$\int\int_{R'} \Phi(\rho)\ \rho\, d\rho\, d\lambda = \int\int_{R} D(r)\ r\, dr\, d\vartheta \tag{2}$$

where R and R′ are the region and image region respectively. For MC-districting the solution is always the identity transformation. Equation 2 can be more generally written, to include non-radially symmetric cases as well. That is

$$\int\int_{R'} \Phi(\rho, \lambda)\, \rho\, |J|\, dr\, d\vartheta = \int\int_{R} D(r, \vartheta)\ r\, dr\, d\vartheta \tag{3}$$

where $|J|$ is the Jacobian determinant

$$\pm J = \frac{\partial \rho}{\partial r}\frac{\partial \lambda}{\partial \vartheta} - \frac{\partial \lambda}{\partial r}\frac{\partial \rho}{\partial \vartheta}\ . \tag{4}$$

For fig. 2b above, let $\Phi(\rho, \lambda)$ = constant and $D(r) = A \exp(-br^2)$. Then, if $\lambda = \vartheta$, J simplifies to $\vartheta\rho/\partial r$, and the required equation is

$$r = \left[-\frac{1}{b} \log(1 - p(\rho))\right]^{\frac{1}{2}} \tag{5}$$

where $p(\rho)$ is the proportion of the population out to ρ in the uniformly populated image region.

As the city develops along a time path, growing in population and area, the existing health care facilities cannot move with it, because they are fixed in position for the duration of their functioning. Services will approximate the theoretical change partly by the development of new facilities, but mostly by the shuffling of resources between existing sites.

3. The impact of time on facility behaviour at particular locations

In this section the impact of the evolving urban system on locational behaviour at specific locations is discussed. Consumer demand is considered in two ways: (a) as finite and supply independent and (b) as supply dependent. These are the extremes in the continuum noted in Section 1.

3.1. *Case (a)*

Let unit demand be a monotonic declining function of accessibility costs represented by the distance ρ from a facility. Specialized services have gently sloping unit demand curves (UDCs); general services have UDCs with steep slopes and high vertical intercepts at zero distance. For a facility in the urban plane located at (r, ϑ) with respect to the city centre (0,0), the demand density in service category *a* at (ρ, λ), a polar co-ordinate pair originating at (r, ϑ), is

$$Q_{a(r\vartheta)} = \Phi_{(r\vartheta)}(\rho, \lambda) \; q_a(\rho) \, . \tag{6}$$

The total demand is thus

$$\underline{Q}_{a(r\vartheta)} = \int_0^{2\pi} \int_0^{q(\rho)=0} \Phi_{(r\vartheta)}(\rho, \lambda) \, q_a(\rho) \, \rho \; d\rho \, d\lambda \tag{7}$$

where

$\Phi_{(r\vartheta)}(\rho, \lambda)$ = population density function originating at (r, ϑ)

$[\Phi_{(r\vartheta)}(\rho, \lambda) = D\,\{[\rho^2 + r^2 - 2\rho r \cos(\pi - \lambda)]^{\frac{1}{2}}, \; \lambda\}$

$q_a(\rho)$ = UDC for service *a* at (r, ϑ).

$\underline{Q}_{a(r\vartheta)}$ = realised demand in service *a* at (r, ϑ).

A simulation was carried out to determine the effects of density change on the demand for services at three locations in the urban plane (0 kms, 5 kms, and 10 kms from the centre). The parameter changes arising in the example shown occur to the density function, not to demand. They are based on population density gradients in London from 1801 to 1971. These indicate a falling central density and an increasing suburban density, a pattern that is common to many cities. Fig. 3 gives the details. On the horizontal axis is time; on the vertical axis is realised demand. The horizontal lines represent hypothetical

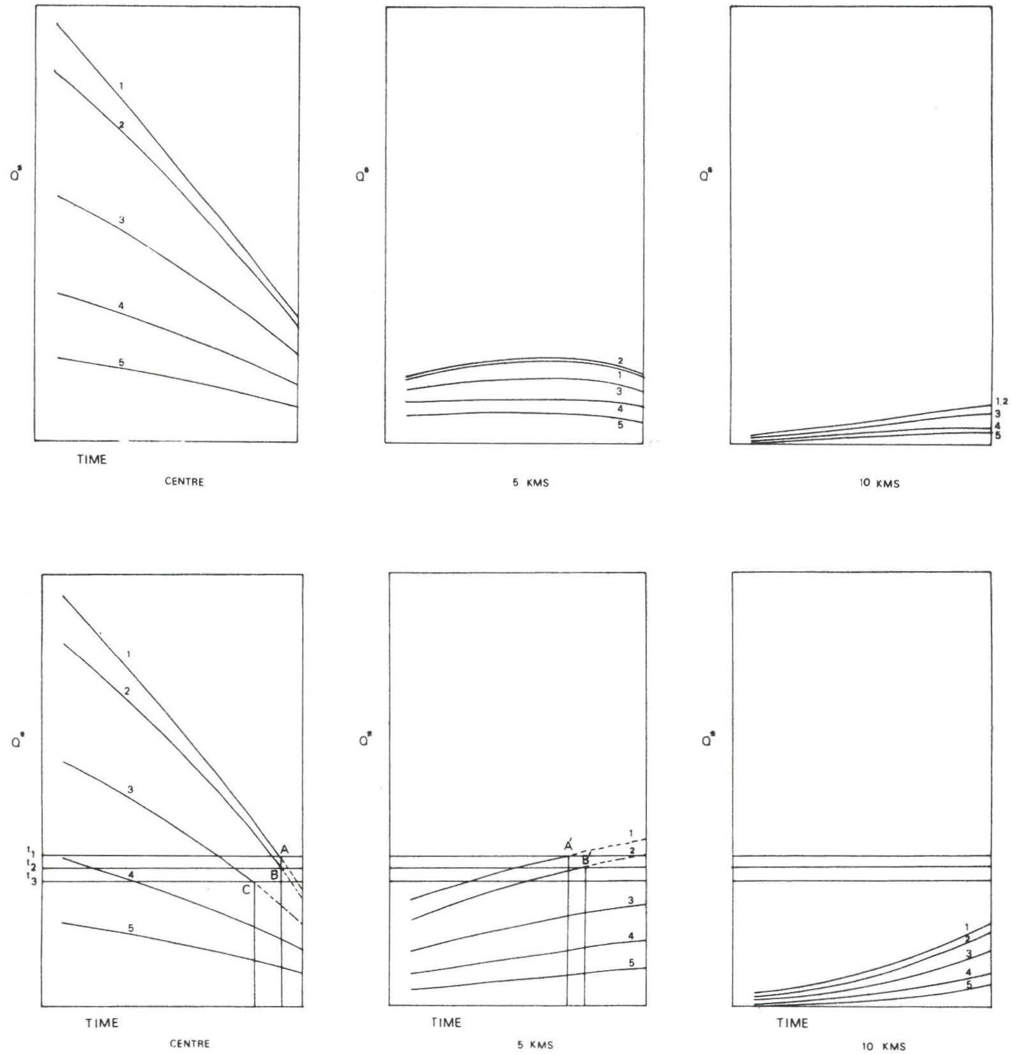


Figure 3 The impact of urban spread on the demand for different health care services at locations spaced at different distances from the city centre [case (a)].

thresholds for each service - levels of demand that must be attained before a service is provided. It is noted,

1. Demand at the centre is never exceeded by demand in the same service category elsewhere. Though this makes this location the most attractive, its attractiveness declines with time.
2. The decline in demand is unevenly split between service categories. Short distance, high-volume services are the most affected.

3. The decline in central services is offset by an increased demand in the suburbs but at different rates depending on the service category.

3.2. *Case (b)*

For the second type of consumer behaviour demand rises to meet the supply available. Despite changes in density, if resources are fixed, the cases treated in each service category are unchanged. The symptoms of spatial disparity are unequal utilization rates in different parts of the city due to differences in population access. Fig. 4, based on 1977 data, shows clearly this effect for parts of the London region. The horizontal axis shows the resources available; the vertical axis, the cases treated. In this instance, the central areas are generating more patients per capita than the city periphery where the supply is less.

A refinement in case (b) is the inclusion of terms to reflect the competition for resources in different locations. This is the approach taken in Mayhew and Taket (1980), which uses a discrete attraction constrained gravity model. For the continuous case in current notation,

$$Q_{(r\vartheta)}\,(\rho\,,\,\lambda) \;=\; R_{(r\vartheta)} \left[\frac{\Phi_{(r\vartheta)}(\rho\,,\,\lambda)\; q(\rho)}{\int_0^{2\pi} \int_0^{q(\rho)=0} \Phi_{(r\vartheta)}\,(\rho\,,.\lambda)\; q(\rho)\; \rho\, d\rho\, d\lambda} \right] \tag{8}$$

where

$$R_{(r\vartheta)} \;=\; \text{resources available at } (r,\vartheta)$$

$$\Phi_{(r\vartheta)}\,(\rho\,,\,\lambda) \;=\; \text{density of population or relative «need» at } (\rho\,,\,\lambda)$$

$$q(\rho) \;=\; \text{a space discount function that decreases with } \rho\,.$$

The denominator in (8) ensures that demand does not exceed supply in each location, r, ϑ. Namely that

$$\underline{Q}_{(r\vartheta)} \;=\; \int_0^{2\pi} \int_0^{q(\rho)=0} Q_{(r\vartheta)}\,(\rho,\,\lambda)\; q(\rho)\; \rho\, d\rho\, d\lambda \;=\; R_{(r\vartheta)}. \tag{9}$$

Facility behaviour under case (a) or case (b) consumer demand types is broadly comparable. Adjustment in case (a) is quicker, since the

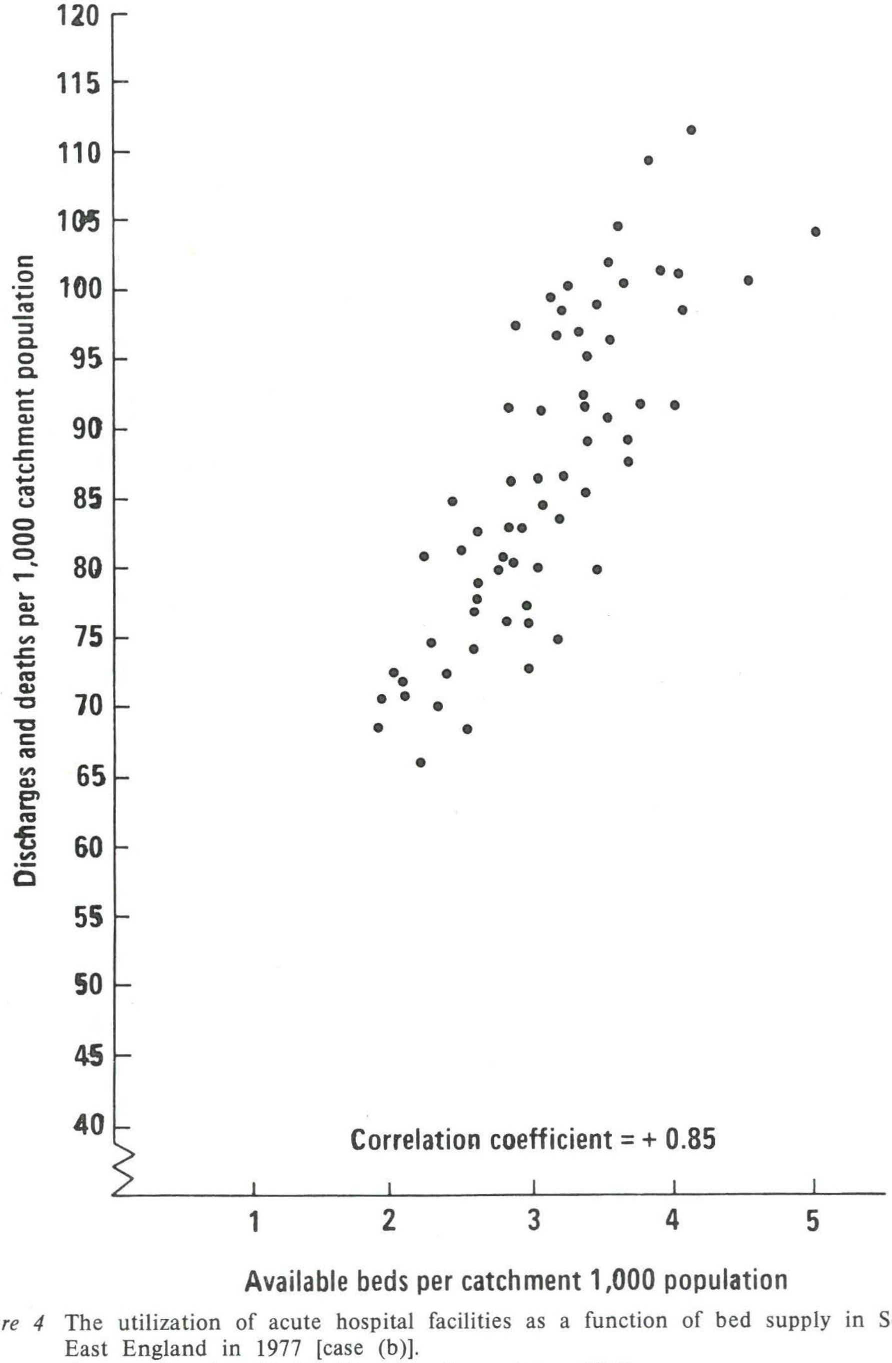


Figure 4 The utilization of acute hospital facilities as a function of bed supply in South-East England in 1977 [case (b)].
(Source: London Health Planning Consortium, 1979).

impact on services is immediately registered in the form of changing demand. Typical of the options available for each hospital are:

1. Facility relocation to a site offering a more even pattern of demand or an equitable pattern of utilization rates. This is a long-term solution.
2. Specialization to concentrate on services with dispersed demand.
3. The creation of entirely new services based on technological developments in medicine.
4. A complete change of function, say to caring for the chronically sick.
5. Closure.

4. Treatment of space

In this section the assumption equating travel costs with distance is relaxed. Generally, travel costs in urban areas are modified by factors determined by the mode or modes of travel, congestion factors and the network geometry. If costs are now represented by time, then a minimum time path between A and B is given by the smallest value of the integral

$$\int_{A}^{B} c(r)\, dr \tag{10}$$

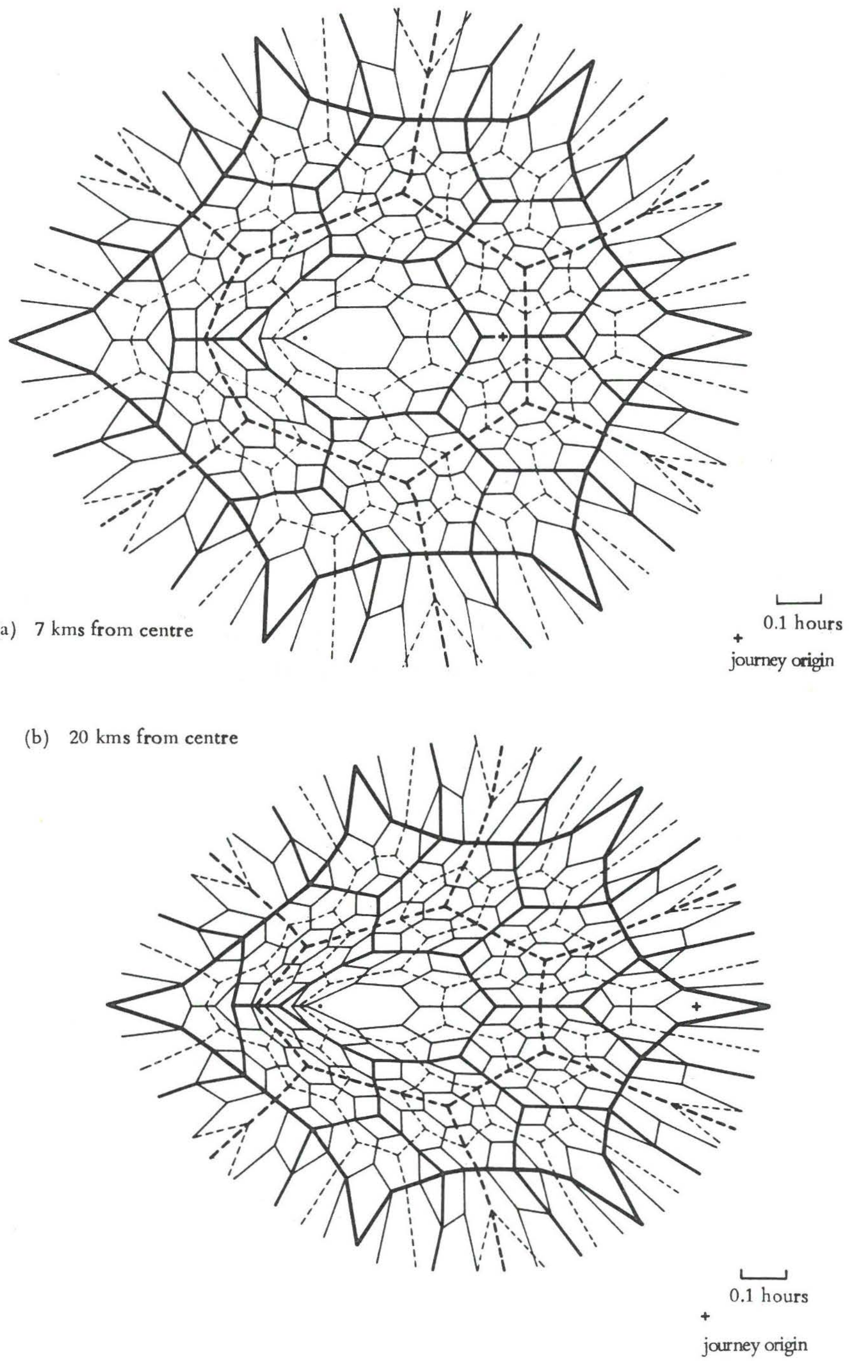
where

c(r) = the cost of travel (time per unit of distance) as a decreasing function of distance from the city centre.

Treating accessibility nonlinearly has many consequences: for example, journey paths may be curved instead of straight as when costs are Euclidean (Angel, Hyman, 1977). Two examples of the effects in health facility location will suffice to show some of the complications that can arise.

The first highlights the problem of creating a map of travel times in which to embed the spatial framework shown in fig. 2. It turns out that one map is needed for each point in the urban plane. Two such maps – for 7 kms and 20 kms from the centre – are shown in fig. 5. On each map, distance is scaled to the travel time. Only the city centre (not shown) gives a symmetric map, because here the shortest paths are all radials.

Consider next the total travel time of the urban population at different distances from the centre based on these maps. This can be written down as an integral over the urban plane as a function of r, the distance from the centre. For Euclidian distance, the point of



(a) 7 kms from centre

(b) 20 kms from centre

Figure 5 Maps of travel times in the urban plane in which distance from the journey origin is equal to travel time. No two origins in the urban area give the same map

minimum aggregate travel is the city centre. For the nonlinear case, the result depends on the cost surface. As instances, the following expression for c(r) was used,

$$c(r) = zr^{-p} \qquad o < p < 1 \tag{11}$$

where p and z are parameters. The total travel time was calculated from a procedure used to solve (10) for different values of the parameters. Fig. 6 shows the results. The conclusions are:

1. The city centre is no longer the most accessible point in the urban plane.
2. The «minimum cost» location is a few kilometers «off-centre».
3. This location is dependent on p but not on z in equation (11).

Transport developments that give undue priority to the suburbs will destabilize minimum cost locations and cause them to migrate out. In the model this is equivalent to an increase in p. Interestingly, it is possible to detect this effect in London. In recent years three large teaching hospitals (layer 1 in the theoretical scheme) have vacated central positions for new locations between 5 kms and 10 kms out.

The second illustration of introducing nonlinear travel costs concerns the problem of marginal districting. The object is to plot a critical isochrone of the travel time about each location. Using distance, this would merely be a circle of constant radius. We know that the journey time between A and B can be expressed as,

$$t = t(r_1\,\vartheta_1\,,\ r_2\,\vartheta_2\,,\ z,\ p) \tag{12}$$

where $r_1\,\vartheta_1$ and $r_2\,\vartheta_2$ are the points A and B above. Re-expressing time as a function of r_2, we obtain

$$r_2 = r_2(r_1\,\vartheta_1\,,\ \vartheta_2\,,\ t\,,\ z\,,\ p). \tag{13}$$

By fixing r_1, ϑ_1, t, z, and p and allowing ϑ_2 to range through 2π radians the desired isochrones can be plotted. This is done in fig. 7 for some accident and emergency centres in London (a sub-set of acute hospital facilities). The critical isochrone is ten minutes, and the values for z and p are 0.33 and 0.75 respectively. As anticipated, because travel costs are highest in the centre, the areas influenced by each location is an increasing function of distance from the centre. Most areas can be approximated by off-centred circles, but near the city centre the areas may distort to cardiods, because of heavy traffic congestion. If this map is true, then gaps in provision can be detected, and steps can be taken to correct them. For example, an obvious extension to this approach is the set coverage problem - selecting the minimum number of facility sites that exactly cover the urban area for a desired critical value of the isochrone (Mayhew, forthcoming).

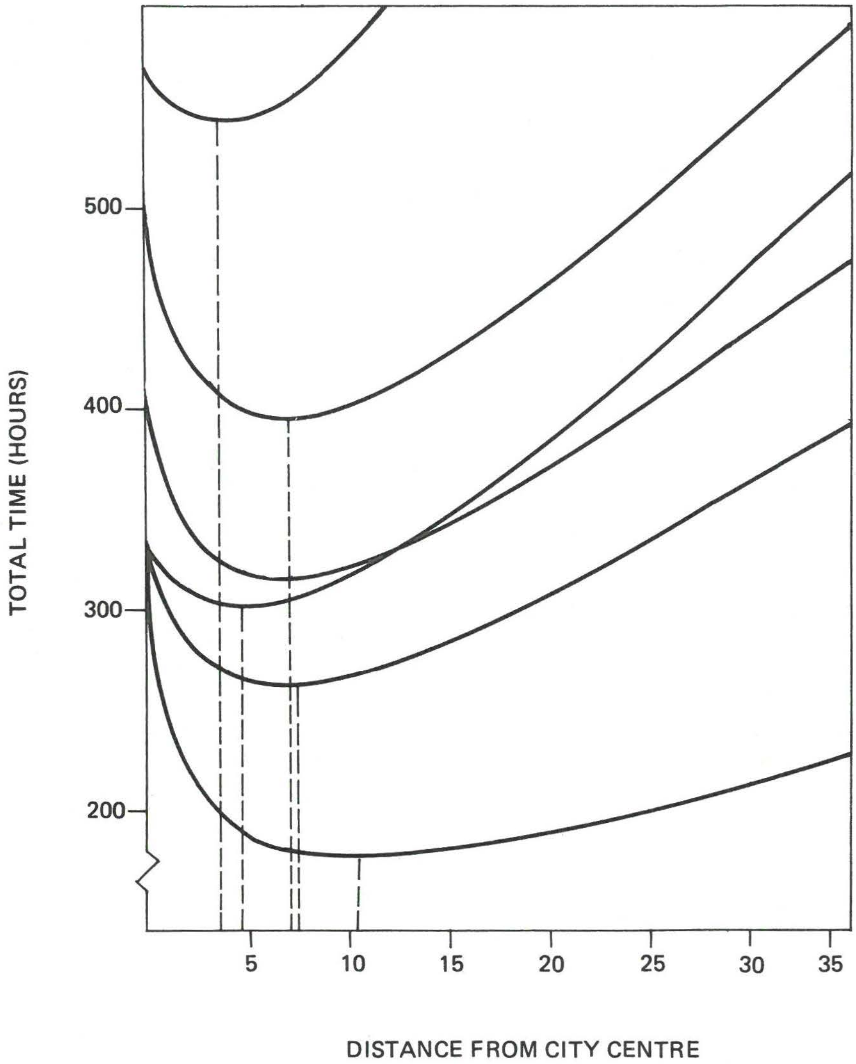


Figure 6 Aggregate travel time as a function of distance from the city centre r for different values of z and p. The points of minimum aggregate travel time are located off-centre for the cost surface [equation (11)]

5. Empirical notes

5.1. *Acute facilities in London*

An important empirical result concerns the districting principle discussed in Section 2. Fig. 8 shows the locations of all acute hospital facilities within the study radius at five points in time. In examining these patterns, it is important for current purposes to know whether changes

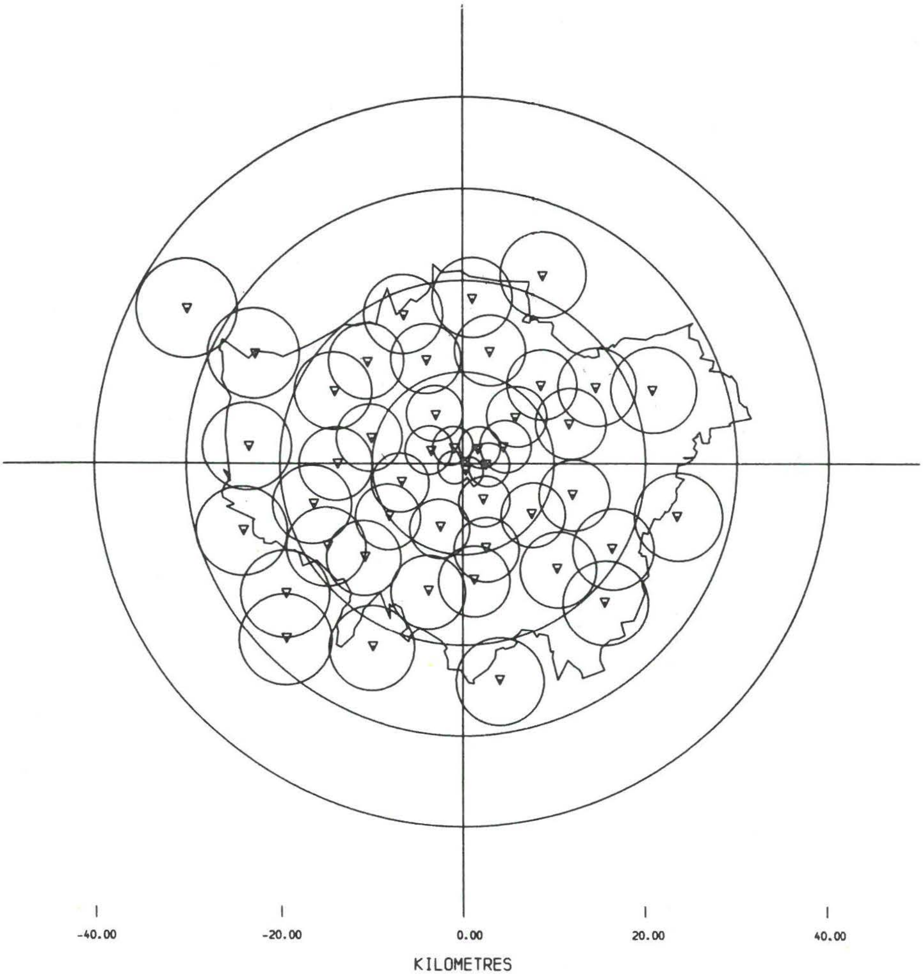


Figure 7 Ten minute isochrones around selected accident and emergency centres in London. Response areas increase with distance from the city centre (z = 0.33; p = 0.75; t = 0.167 hours)

in facility districting can be detected over time. Accordingly, the urban area was partitioned into concentric rings 2.5 kms in width. The numbers of beds and hospitals in each ring were totalled and proportioned, and then multiplied by N^t the total population, to provide the number of person equivalents. Ten linear regressions of the following from were carried out.

$$Y_i^t = B_0^t + B_1^t X_i^t + u_i^t \tag{14}$$

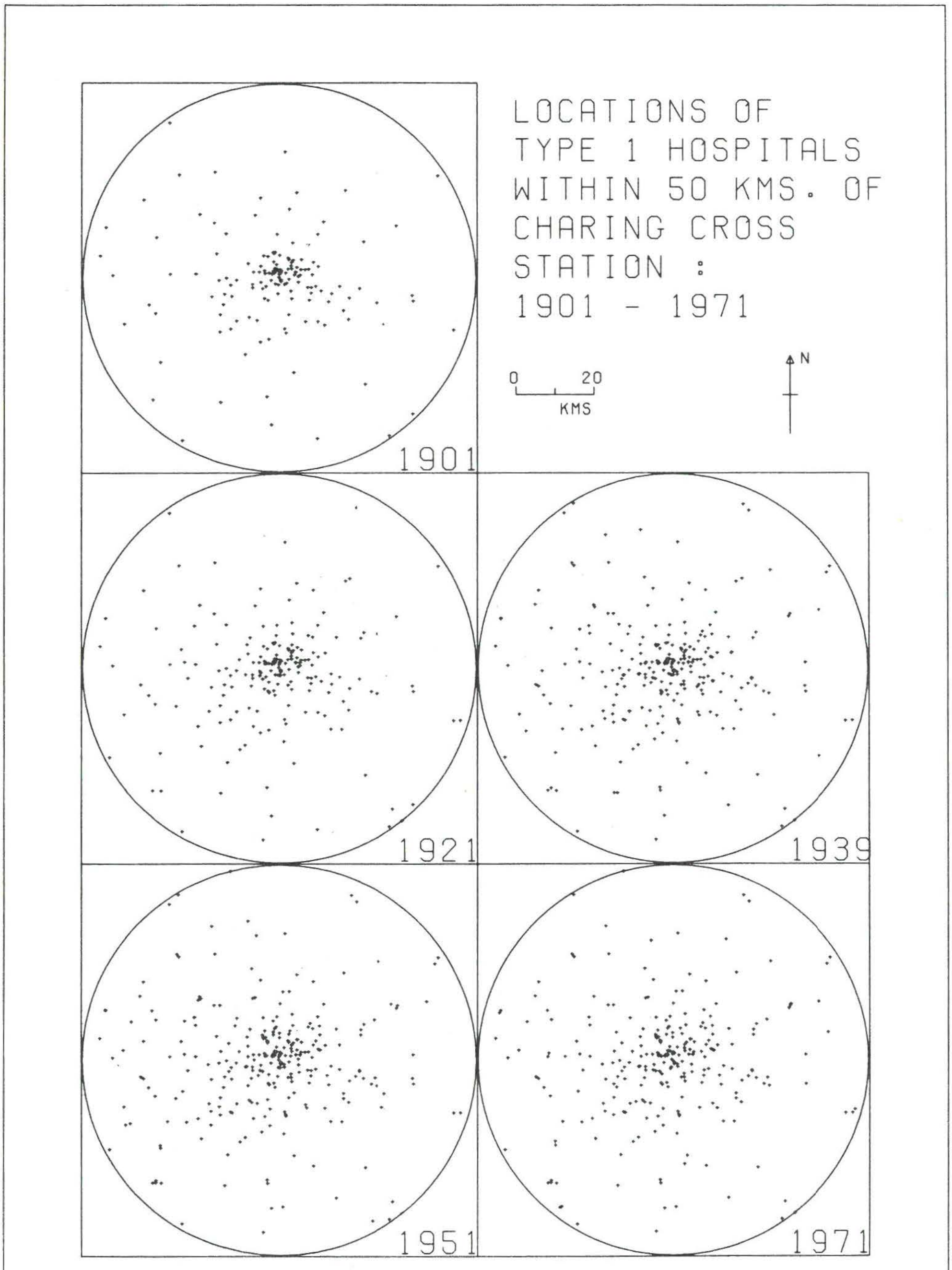


Figure 8 The location of acute facilities in London, 1901-1971 (Source: Mayhew, 1979)

where

Y_i^t = logarithm of the person-equivalent density of either beds or hospitals. It is given by

$$Y_i^t = \left[\log \frac{P_i^t N^t}{A_i} \right]$$

where A_i is the area of the ith ring; P_i^t is the proportion of hospitals or beds in ring i at time t, and N^t is the urban population

X_i^t = the logarithm of the actual population density

B_0^t, B_1^t = the regression coefficients

u_i^t = the ith error term.

The results all gave values of R^2, the coefficient of explanation, of between 0.85 and 0.93 for both beds and hospitals. The theoretical interest is in the slope values (B_1^t). Fig. 9 shows the details. For P-districting, hospitals and beds «map» exactly into the population density, so that the result is a 45°-line passing through the origin. For MC districting, hospital equivalents would plot horizontally.

6. Discussion

From the diagram the bed density in 1901 was lower than the equivalent density of population the while hospital density was higher. Thus facilities were closely packed relative to the population but they were small in size. The hospitals concerned were of the «cottage» type - a form of facility still popular in some rural areas. In 1971, the bed and hospital slopes had converged almost to one, the facilities converting to a form of P-districting. The consequence of this has been much larger suburban facilities and a gradual phasing out of the cottage hospitals mentioned above. The dynamics of this process are also evident if a plot is made of mean facility size (in bed units) on distance from the city centre. In 1901 a peak of approximately 360 beds is observed near the centre, but this declines rapidly to facilities averaging less than 20 beds at distances above 10 kms. By 1901, suburban facilities were enlarged and developed to the extent that two new peaks – one at 15 kms and another at 25 kms – were formed, partly offsetting the former attractiveness of the centre. These hospitals would correspond to layer two in the earlier theoretical structure (fig. 2).

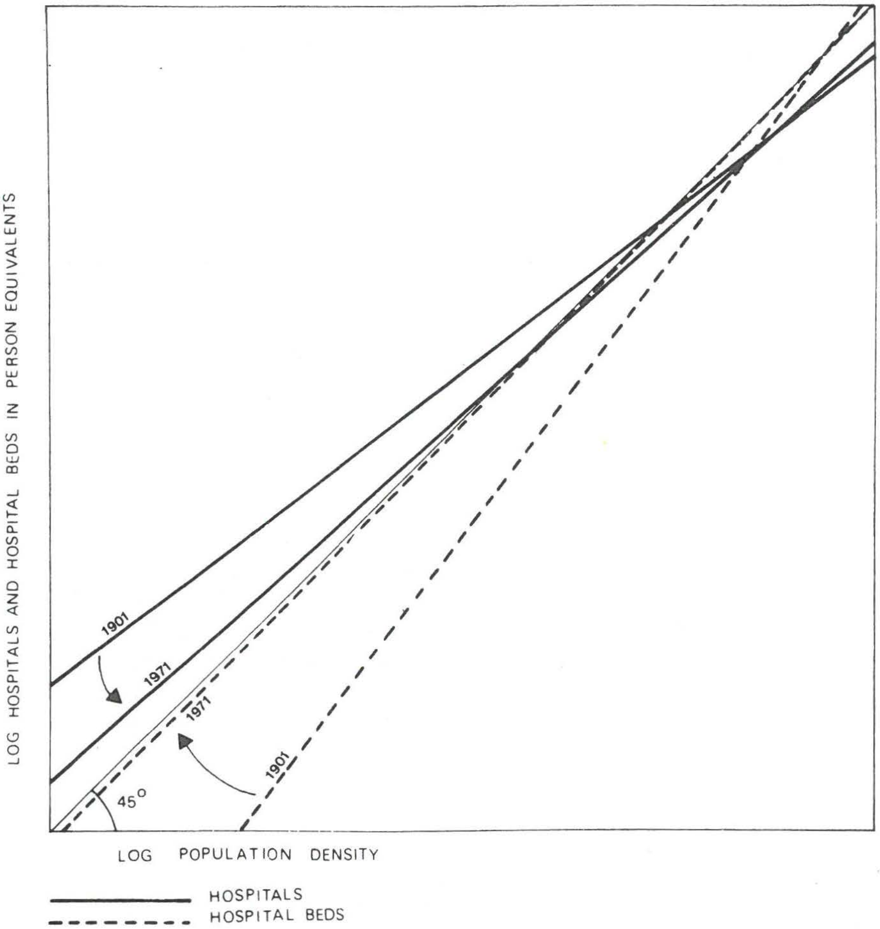


Figure 9 Changes in hospital districting in London for 1901 and 1971 (Source: Mayhew, 1979).

7. Conclusions

This paper has introduced aspects of research into health care facility locations in cities. Although different approaches to health care provision exist in different countries, and although variations occur in the sizes and dynamics of cities, the problems facing providers – government, agencies, charities, etc. – are surprisingly similar, and are of increasing importance. As the financing of health care services becomes ever more expensive, it becomes more urgent at develop rational systems of health care provision, that acknowledge not only past provision but also likely changes in the urban environment.

References

Angel S., Hyman G. (1976) *Urban fields*, Pion, London.

Christaller W. (1933) *Central places in Southern Germany* [translated by Baskin C.W. (1966)], Prentice-Hall, Englewood Cliffs, USA.

Clark C. (1951) Urban population densities, *Journal of the Royal Statistical Society, Series A 114,* 490-496.

Dietrich C. (1977) Macromodels of inpatient and outpatient health care systems in the Federal Republic of Germany, Modelling Health Care Systems, Proceedings of the IIASA Workshop, International Institute for Applied Systems Analysis, Laxenburg, Austria.

London Health Planning Consortium (1979) Acute hospital services in London: A planning profile, HMSO, London.

Mayhew L. D. (1979) The theory and practice of urban hospital location, Ph. D. Thesis, Dept. of Geography, Birkbeck College, University of London, London.

Mayhew L. D. (forthcoming) Automated isochrones and the location of accident and emergency services in cities, *Professional Geographer.*

Mayhew L. D., Taket A. (1980) RAMOS - A model of health care resource allocation in space, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Shigan E., Kitsul P. (1980) Alternative approaches to modelling health care demand and supply, WP-80-80, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Tobler W. R. (1961) Map transformations of geographic space, Unpublished Ph. D. Dissertation, University of Washington.

Riassunto. Una teoria generale della localizzazione dei servizi sanitari dovrebbe in linea di principio essere in grado di dare una risposta a molti problemi, dai più tecnici, quali la localizzazione in senso stretto, a quelli meno trattabili in modo rigoroso, di natura etica o clinica. Questo saggio si propone di sondare solo una piccola parte di tali problemi, che tuttavia sono collocati a metà strada tra i due estremi. Non vengono proposti modelli normativi generali, ma piuttosto viene sviluppato uno schema interpretativo dello sviluppo spazio-temporale di un sistema di servizi sanitari in un'area specifica, la città di Londra. L'approccio sviluppato è tuttavia abbastanza generale da essere applicabile ad altri casi. Una caratteristica che lo distingue da altri approcci analoghi è l'uso dello spazio continuo. Ciò permette approfondimenti che non sarebbero possibili mediante l'usuale discretizzazione dello spazio in zone.

Résumé. Une théorie générale de la locatisation des services sanitaires devrait, en principe, être capable de répondre à beaucoup de questions: des plus techniques tels que la localisation proprement dite, à celles qui, de nature éthique ou clinique, ne peuvent pas être traitées aussi rigoureusement. Cet article aborde seulement une partie des problèmes et se situe à mi-chemin entre les deux extrêmes. On ne propose pas un modèle operationnel dont le résultat peut être immédiatement utilisé, mais plutôt on développe un schéma interprétatif du développement spatial et temporel d'un système de services sanitaires dans une zone particulière, la ville de Londres. L'approche décrite est, cependant, assez générale et peut être appliquée à d'autres cas. En outre, l'utilisation de l'espace continu la distingue des autres approches analytiques. Ceci permet des développements qui ne seraient pas possibles si on se basait sur une articulation de l'espace en zones (espace discret).

Sistemi Urbani, 3, 1981, 435-454 Guida Editori, Napoli, Italia

Public facility location with elastic demand: users' benefits and redistribution issues

E. S. Sheppard

Human Settlements and Services Area, IIASA, International Institute for Applied Systems Analysis, Schlossplatz 1, Laxenburg, A-2361, Austria

Received 15 June 1980, in revised form 20 September 1980

Abstract. Realistic consumer behavior must be incorporated as a datum in locating public facilities, since they serve customer demands as expressed through spatial interaction. General formulae for interaction and accessibility, independent of any particular spatial interaction distribution function, allow incorporation of elastic total demand for facilities. From this viewpoint extensions of «locational surplus» can be considered, involving a tradeoff between achieving analytical correctness versus requiring restrictive behavioral assumptions. Such a user oriented measure provides no more guarantee of equity than other cost related objectives. As a result measures of the redistribution effects of facility location patterns are developed for use as parts of a multi-objective approach to facility planning.

Key words: consumer behavior, spatial interaction, elastic demand, income redistribution, multi-objective planning.

0. Introduction

Public facility location models typically consist of two components, the choice of a set of locations for facilities, and the subsequent allocation or distribution of consumers to facilities. The latter component is essentially used to evaluate the benefits of a particular location pattern and is thus the key to choosing the best combination of facilities. For any given allocation mechanism a mixed integer programming problem exists which must be solved to find the «optimal» locations.

This paper limits consideration to the allocation problem. In particular, situations will be analyzed where facility customers choose the facility that they will visit, and where the benefits of a location pattern will be evaluated from the point of view of maximizing customers' well-being. In order to do this a model is necessary which accurately describes the spatial behavior of customers. The purpose of this paper is to take such a model and use it as the basis for measuring the benefits of location patterns. The paper will be divided into three parts; a description of the interaction theory; a measure of consumers' surplus; and a method of evaluating the redistributional consequences of facility patterns.

1. A model of spatial interaction

The starting point of this paper is that people do not typically visit the closest facility available. Rather, for a number of reasons the customers living at a demand point, i, will visit any one of a number of possible destinations, j. This may be for a number of reasons, which will not be discussed here; varying preferences, different behavioral norms, or different constraints on choice (Sheppard, 1978, 1980b). As a result the «demand» for use of j by customers from i is elastic.

Demand elasticity is of two forms. First, the number of trips made from i, or the demand for travel to the type of public facility to be located, will (among other factors) be inversely related to the accessibility of facility sites to demand points. Places closer to the system of facilities will generate more trips per unit of time than places further away. This captures part of the so-called «hidden demand» problem plaguing public facilities which may be substantiated on empirical and theoretical grounds (London Health Planning Consortium, 1979; Sheppard, 1980a). Define the demand for travel by one individual as g_i. Second, the fraction of trips from i that terminate at a given destination j, h_{ij}, will depend on the attractiveness of j relative to other destinations. Assuming that the attractiveness of one location is independent of that of another location:

$$I_{ij} = O_i \, g_i \, h_{ij} \tag{1}$$

$$h_{ij} = f_{ij} / \sum_{k=1}^{n} f_{ik} \tag{2}$$

where I_{ij} is the number of trips from i to j, O_i is the population of i, f_{ij} is the attractiveness of j as perceived at i, and there are n facility locations. From the point of view of demand theory, the demand for location j, I_{ij}, depends on the «substitution» effect, h_{ij}, of the relative attractiveness of j, and the aggregate effect of the general attractiveness of this public facility (g_i).

Each of these effects has received separate treatment with public facility models. The substitution issue has been most recently dealt with by Beaumont (1980) who specifies f_{ij} as a gravity model; one of many forms suggested for f_{ij} in the literature. Both gravity and intervening opportunities approaches combine spatial separation with the *in situ* attractiveness of j. The demand for travel component has also been given attention (Erlenkotter, 1977; Wagner and Falkson, 1975), but attempts to combine the two are few (Albernathy and Hershey, 1972; Leonardi, 1980b; Sheppard, 1980a; Tapiero, 1980).

In bringing the two components together, a functional form for g_i, as a function of accessibility, φ_i, of i to all destinations, is necessary. Several possibilities can be suggested:

$$g_i = \alpha \varphi_i^{\beta} \tag{3}$$

$$g_i = \frac{1}{\beta} e^{\beta \varphi_i} \tag{4}$$

$$g_i = N_i (1 + e^{\alpha - \beta\varphi_i})^{-1} \tag{5}$$

$$g_i = N_i \varphi_i^{\beta} (\varphi_i^{\beta} + k)^{-1} . \tag{6}$$

Each function has its merits: (3) has a constant elasticity of demand with respect to access of β; (4) has a demand elasticity equal to φ_i; (5) always has a finite maximum (equal to N_i, the «need» at i) as φ_i increases and is S-shaped (logistic); and (6), which has been already examined for public facilities (Leonardi, 1980b), also has a finite maximum, and can be S-shaped for certain values of $\beta^2 > d$. Choice of one of (3) - (6) would in practice depend on theoretical considerations, and on available empirical data.

In (3) - (6) «accessibility» has been used as a surrogate for attractiveness. The reason for this is that accessibility is not solely related to the cost of travel, but also (through f_{ij}) to the *in situ* attractiveness of the destination. The rationale for this is that accessibility is a concept related to users' behavior. Thus a place travelled to more frequently is regarded as more accessible from the origin, as perceived by residents of that origin, than a place which is less traveled to, irrespective of whether the different behavior is due to transport cost, variations in facility size, information availability, or travel constraints (Sheppard, 1979).

From the assumption that the attractiveness of j, f_{ij}, is independent of f_{ik}, and defining φ_{ij} as the accessibility of j from i:

$$\varphi_i = \sum_j \varphi_{ij} . \tag{7}$$

Accessibility φ_{ij} is a ratio scale, comparable across all origins i and facility sites j. The general definition of φ_{ij} suggested here is that φ_{ij} be proportional to the probability that a randomly sampled trip, made in the study area for the purpose of visiting a public facility of the type to be located, is observed to originate at i and terminate at j.

Single purpose trips

If we assume that travel occurs directly to the facility and then home again with no other stops:

$$\begin{aligned} \varphi_{ij} &= f_{ij} \Big/ \left(\sum_i \sum_j f_{ij} \right) \\ &= \alpha \cdot f_{ij} . \end{aligned} \tag{8}$$

Since accessibility is taken as a ratio scale, the constant α may be ignored, and

$$\varphi_i = \sum_j f_{ij} \tag{9}$$

is a definition of accessibility.

Multiple purpose trips

With trips that involve a chain of several individual links, with other stops before (or after) stopping at one of the public facility sites, the definition (9) is too simple. Indeed multiple purpose trips are extremely complex, as the decision to visit any given type of public facility may be made at any time during the chain of actions that comprise the total trip, and can depend on particular other types of stops that the tripmaker decides on, as well as the locations where such stops are made. As a very simple version of this process, let us assume that empirically it is possible to isolate that subset of all trips made which include at least one stop at the type of public facility to be located. Define the large set of M locations which may be visited, for shopping, social, or work-related purposes, during the trip, including all the origin locations i and the public facility locations j. It is now possible in principle to construct an M by M matrix of probabilities, p_{mn}, where p_{mn} is the likelihood that a randomly sampled link from a trip in this subset is observed as occurring between m and n.

This matrix, P, is a transitive matrix, since each row sum is less than one. Now the accessibility of j from i is equal to the probability that a trip from i reaches j by any one of a number of indirect routes (via k, 1, etc.), as well as by the direct link as captured in p_{ij}:

$$\varphi_{ij} = \sum_{b=1}^{\infty} p_{ij}^{(b)} \tag{10}$$

where $p_{ij}^{(b)}$ is the probability of reaching j from i, with b-1 intermediate stops. In matrix form:

$$\varphi = \varphi \cdot \mathbf{i} \tag{11}$$

$$= [(\mathbf{I} - \mathbf{P})^{-1} - \mathbf{I}] \cdot \mathbf{i} \tag{12}$$

where **I** is the M by M identity matrix, **i** is a M by 1 vector of ones, φ is the M by 1 vector of accessibilities φ_i, and φ is the M by M matrix of φ_{ij}'s. φ is in fact a matrix of space potentials (Sheppard, 1979).

It should be emphasised that the probabilities p_{mn} must all be measured in the same units. This is not a problem if p_{mn} is derived empirically as relative trip frequencies. However it is a problem if a theoretical derivation is attempted. Thus if, for example, we treat p_{ij} as measured by h_{ij}, where

$$\begin{aligned} p_{ij} &= h_{ij}/\left(\sum_i \sum_j h_{ij}\right) \\ &= h_{ij}/M \end{aligned} \tag{13}$$

then this would be correct only if the number of trips leaving each origin, i, per unit of time were the same. But of course that removes the question of elastic demand for travel by assumption. The relationships between g_i and f_{ij} in the case of multiple purpose trips are enormously complex. The solution as outlined here is pragmatic rather than theoretically rigorous.

An alternative approach to multiple purpose trips should also be mentioned. This is to capture the externality benefits of multiple purpose trips directly in the specification of f_{ij}. f_{ij} typically includes a measure of the *in situ* attractiveness of the facility at j. If the definition of *in situ* attractiveness is extended to encompass the closeness of j to other complementary opportunities that might be used during a multiple purpose trip, then this may account for some of the multiple trip effects. This is analogous to capturing elastic travel demand by extending the definition of origin generating effects to include accessibility of destinations.

An example

For the case of single purpose trips, which is all that will be treated in the following sections, then the interaction model, from (1), (2) and (9), becomes:

$$I_{ij} = O_i g(\varphi_i) \cdot \varphi_i^{-1} \cdot f_{ij} \ . \tag{14}$$

The function $g(\varphi_i)$ could be derived theoretically from microeconomic considerations. However, given the dubious nature of assuming perfect information and decision making characteristics in real situations $g(\varphi_i)$ would be better estimated empirically, taking due consideration of customer type, and other *in situ* origin specific factors that would influence the demand for use of the facility. Note that:

$$g(\varphi_i) = \left(\sum_j I_{ij}\right) \varphi_i^{-1} O_i^{-1} \ . \tag{15}$$

In particular, if we assume the functional form in equation (3):

$$I_{ij} = \alpha O_i \varphi_i^{\beta-1} \cdot f_{ij} \tag{16}$$

a case which is curiously similar to the origin constrained Alonso «theory of movement» (Alonso, 1978; Ledent, 1980). To call this an origin constrained interaction model would be a loose use of the term, however, since the number of trips is not fixed.

2. Users' benefits

2.1. *The Neuburger approach*

The case of consumers' surplus as a measure of user benefit in the case of transport improvements has been treated by Neuburger (1971) who argues that the change in benefits due to improvements is equal to the perceived net benefits of the new system plus the fall in transport costs incurred under changed user behavior on the new system. However, he restricts himself to the case where *in situ* destination attractivities are held constant while travel costs are changed on some routes. He is thus able to use travel costs as the measure of benefits. Fig. 1 reproduces an example of a demand curve of this type. Benefits perceived by the user are the increase in demand as perceived costs fall from PC_1 to PC_2 (ABDE); perceived costs are the extra perceived travel cost of the new trips (BCHG). Actual benefits to the system are the cost reduction on old trips (IJKL) less the incurred cost of new trips (LMHG). The total benefit is ABDE + IJKL - BCHG - LMHG.

However, in the case where destination attractivities are changed, which by definition occurs in facility location problems, this is inadequate. Consider an addition of attractiveness to all destinations in such a way that h_{ij} is unchanged $\forall i,j$. Then more trips will occur at no greater cost, perceived or actual, per trip. Then $PC_1 = PC_2$, $AC_1 = AC_2$; ABDE = IJKL = 0; and BCHG + LMHG > 0. Thus user benefits are negative, despite the construction of new facilities and the new trips generated by these.

Analytically, Neuburger defines consumer surplus by:

$$\Delta S_1 = \sum_i \sum_j \int_{c_{ij}^1}^{c_{ij}^2} I_{ij}\, d c_{ij}\, P_{ij}(c) \tag{17}$$

in which c_{ij}^2 is the old travel cost, c_{ij}^1 is the new travel cost, and $P_{ij}(c)$ is the path of integration in cost space between the initial and final matrices of transport costs (Beaumont, 1980). If $I_{ij} = O_i a_j e^{-\beta c_{ij}} / \sum_k a_k e^{-\beta c_{ik}}$, where a_j is the *in situ* attractiveness of site j, then

$$\partial I_{ij} / \partial c_{im} = \partial I_{im} / \partial c_{ij} \tag{18}$$

implying that integration is independent of the path. $P_{ij}(c)$ may then be dropped from (17), and

$$\Delta S_1 = \frac{1}{\beta} \sum O_i \log \left[\frac{\sum_j a_j \exp\{-\beta c_{ij}^2\}}{\sum_j a_j \exp\{-\beta c_{ij}^1\}} \right] . \tag{19}$$

However, if the travel costs are unchanged, but facility sizes are increased, (17) is not a valid index, as only changes in travel costs, c_{ij},

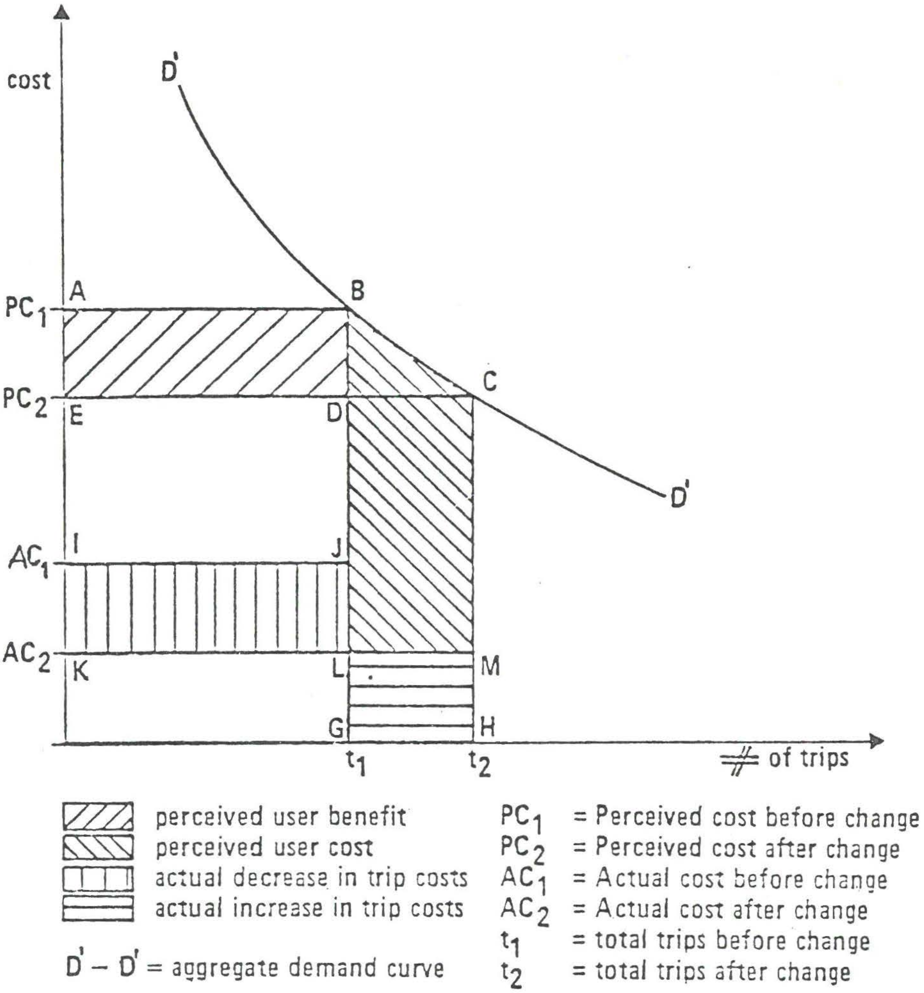


Figure 1 Consumer surplus after Neuburger

are allowed for. If these do not change, (17) must be zero. If changes in the *in situ* attractivenesses a_j are also to be considered, then (17) should be integrable with respect to a_j as well as c_{ij}. But this implies that the integrability conditions (18) should be also expanded to include the effects of a_m and a_j on I_{ij} and I_m. If this is done, then the definition of I_{ij} that will satisfy (18) must also be changed, implying that (19) is no longer correct.

Coelho and Williams (1978) have made one such extension in the following manner. Define:

$$I_{ij} = \alpha \cdot O_i \cdot f_{ij} / \left(\sum_k f_{ik} \right) \tag{20}$$

implying an inelastic demand for travel, α. Then define surplus as:

$$\Delta S_1 = \sum_i \sum_j \int_{F_{ij}^1}^{F_{ij}^2} I_{ij} \, d F_{ij} \, P_{ij}(F) \tag{21}$$

where $F_{ij} = 1n(f_{ij})$. Now it may be shown that:

$$\partial I_{ij} / \partial F_{im} = \partial I_{im} / \partial F_{ij} \qquad \forall \, m,j \, . \tag{22}$$

Thus integration is independent of the path, $P_{ij}(F)$, which may thus be eliminated from (21). Then

$$\Delta S_1 = \sum_i \sum_j \int_{F_{ij}^1}^{F_{ij}^2} \alpha \cdot O_i \, e^{F_{ij}} / \left(\sum_k e^{F_{ik}} \right) d F_{ij} \tag{23}$$

$$= \sum_i \alpha \cdot O_i \, \log \left[\sum_j e^{F_{ij}^2} / \sum_j e^{F_{ij}^1} \right] . \tag{24}$$

If no facilities existed initially, $F_{ij}^1 = 0$ for all i,j and:

$$\Delta S_1 = \sum_i \alpha \cdot O_i \, \log \sum_j e^{F_{ij}^2} \, . \tag{25}$$

This may be generalised to the case of elastic demand with ease. Define:

$$I_{ij} = g(\varphi_i) \cdot O_i \cdot f_{ij} \cdot \varphi_i^{-1} \tag{26}$$

where $\varphi_i = \sum_j f_{ij} = \sum_j e^{F_{ij}}$. (27)

Now it may be shown by differentiation that:

$$\frac{\partial I_{ij}}{\partial F_{im}} = \frac{O_i \cdot e^{F_{ij}} e^{F_{im}}}{\varphi_i} \left[\frac{\partial g(\varphi_i)}{\partial \varphi_i} - \frac{g(\varphi_i)}{\varphi_i} \right] = \frac{\partial I_{im}}{\partial F_{ij}}. \tag{28}$$

This is true for all j,m so the integrability condition is not negated by including elastic demand. Thus even with elastic travel demand the Coelho/Williams surplus measure is given, in the case of building public facilities where none existed previously, by

$$\Delta S_1 = \sum_i \sum_j \int_0^{F_{ij}} I_{ij}\, dF_{ij} \ . \tag{29}$$

In interpreting (21) or (29) as measures of consumers' surplus, some rationale is necessary for integration with respect to F_{ij}. The most reasonable possibility is if F_{ij} can be interpreted as the utility u_{ij} of travelling to j, given that the decision has been made to travel from i. In that case, for both (21) and (29)

$$\Delta S_1 = \sum_i \sum_j \int_{u_{ij}^1}^{u_{ij}^2} I_{ij}\, du_{ij}. \tag{30}$$

This is a justifiable extension of the Marshallian price-based surplus measure poineered by Neuburger, from a micro-economic viewpoint.

It is possible to support this interpretation of surplus, if it is true that actual behavior is the result of individuals acting rationally with a particular form of stochastically distributed utilities. This results if we assume an individual's utility is given by:

$$u_{ij} = E(u_{ij}) + \varepsilon_{ij} \tag{31}$$

and further, that the expectation of u_{ij}, $E(u_{ij}) = F_{ij}$ and that ε_{ij} is distributed according to the Weibull distribution. Then (31) generates (26) as the result of a multinomial logit choice model (Sheppard, 1978). Of these assumptions, the most crucial is the Weibull distribution of individual utilities. No evidence seems to have been forthcoming so far showing that it has empirical validity for spatial choice. Further, an empirical prediction that would be expected from random utility theory, that when populations are appropriately disaggregated then spatial behavior would be more uniform, also does not seem to be true from evidence available thus far. Thus, solid reasons for accepting this interpretation of the surplus function are not as yet forthcoming.

2.2. *An alternative approach*

In the light of this, other approaches to specifying benefits seem worth exploring. One such approach is to treat «demand» or interaction on a link i,j as being positively related to attractiveness, f_{ij}, instead of the usual negative demand curve parameterised by cost. It is essential to distinguish demand changes for each origin-destination pair, as it is quite possible to conceive of facility location patterns decreasing demand in some cases while increasing it in others. The demand on any link i,j is I_{ij}, and defining

$$I_{ij} = A_i f_{ij} \tag{32}$$

where

$$A_i = O_i\, g(\varphi_i)\, \varphi_i^{-1}. \tag{33}$$

Then the effect of constructing a new facility, anywhere in the system, on demand for j from i is:

$$\Delta I_{ij} = A_i^2 f_{ij}^2 - A_i^1 f_{ij}^1 = I_{ij}^2 - I_{ij}^1 \tag{34}$$

where surperscripts 1 and 2 refer to «before» and «after» calculations. The situation is depicted in fig. 2. We now have a positively sloping demand curve, but this simply reflects the fact that the vertical axis measures benefits rather than the costs. The case depicted is where $A_i^2 > A_i^1$. OD_1 represents the origin-specific demand curve linking trips to j with the accessibility of j from i. Point E is the trips attracted by the actual value f_{ij}^1. OD_2 is the same curve after facility construction (including in this case an increase in the attractiveness of j). The line D′EHD′, analogous to D′D′ in fig. 1, is the demand curve representing the change in demand at i for j. This is drawn as a dotted line since it will vary depending on how the new facility locations affect A_i and f_{ij}.

In fact, it seems difficult to conceive of how any consistent shape can be ascribed to the «demand curve» between E and H. Indeed there is not even any good reason to suppose that as I_{ij} changes from I_{ij}^1 to I_{ij}^2 that f_{ij} will smoothly shift between f_{ij}^1 and f_{ij}^2. There are many combinations of f_{ij} and A_i that will generate values of $I_{ij}^1 \leq I_{ij} \leq I_{ij}^2$ with

f_{ij} varying anywhere between zero and infinity. This contrasts with economic demand curves that can be conceived of as being well behaved with respect to price shifts. Thus demand may fluctuate in very different ways as we move from E to H, depending on where in the system new roads or facilities are built.

However, this is perhaps not such a serious issue, as it is really immaterial what happens between E and H. We are simply interested in the difference between the new situation, represented by H on the link i,j, and the old situation shown by E. When trips have been disaggregated by links, presumably each trip maker between i and j will on the average receive a benefit of f^2_{ij} no matter whether he is a new trip-maker or someone who used this or another link previously. Previously, each trip on this link received a benefit of f^1_{ij}. From this we can deduce that the change of benefits, on link i,j, will be:

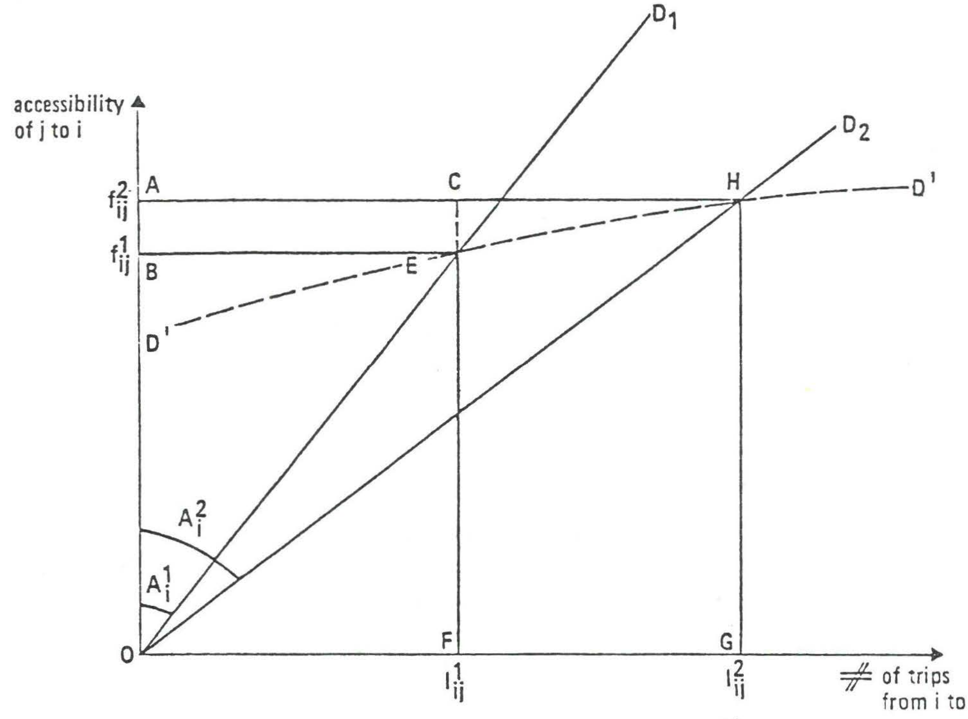


Figure 2 Demand and accessibility on link $\vec{ij}$

$$\Delta S_{ij} = I^2_{ij} f^2_{ij} - I^1_{ij} f^1_{ij}. \tag{35}$$

For facility construction, as opposed to demolition, $f^2_{ij} \geq f^1_{ij}$, and (35) cannot be negative unless $h^2_{ij} < h^1_{ij}$.

Total customers surplus then becomes:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\Delta S_{ij} = \sum_{i}\sum_{j} I_{ij}^{2} f_{ij}^{2} - I_{ij}^{1} f_{ij}^{1}. \tag{36}$$

In the case of constructing a new set of facilities from scratch, $I_{ij} = 0, \forall i,j$, and

$$\Delta S_2 = \sum_{i}\sum_{j} I_{ij}^{2} f_{ij}^{2}. \tag{37}$$

It so happens that this approach can also be given a utility theoretic foundation. If we use Luce's choice theory, which can explicitly allow for non-transitive behavior without ascribing variations in choices solely to varying utilities, then f_{ij} may be interpreted as (Sheppard, 1978):

$$E(u_{ij}) = f_{ij} \tag{38}$$

and

$$\Delta S = \sum_{i}\sum_{j} I_{ij} \cdot E(u_{ij})\,. \tag{39}$$

Thus there are two approaches to specifying user benefits. The first arrives at a benefit function by assuming that spatial choice follows the multinomial logit model, which also happens to meet the necessary integrability conditions. The second approach suggests that when both a_j and distances d_{ij} may be changed, the resulting degree of freedom makes it difficult to conceive of the demand function being well-behaved and integrable. Thus integrability conditions, which are not satisfied by (37), are held to be unnecessary. The resulting benefit function can still be given a utility-theoretic interpretation, if necessary, by relying on an alternative probabilistic choice theory. Since we at present have no basis for choosing between these competing choice theories, it is only possible to conclude that the two surplus approaches are equally tenable, and should be compared empirically.

2.3. *Programming issues*

Application to public facility location would involve choosing the set of facility sites, described by a realisation of a set of binary variables

L, and a set of facility sizes X, to maximise either equation (29) or (37) subject to constraints such as the following:

$$\sum_i I_{ij} \leq x_j \qquad \forall j \tag{40}$$

$$\sum_j (c_j x_j + b_j) \leq B \tag{41}$$

$$x_j \leq 0 \tag{42}$$

where x_j is the capacity of facility j. Constraint (40) ensures facility capacities are not exceeded; and (41) represents a cost function with fixed cost b_j, linear variable cost c_j, and total budget B.

In practical situations I_{ij} would be estimated prior to solving a problem of this sort. It is after all true that interactions, while responding to location patterns, are not control variables but rather represent individuals' behavior to which location patterns conform. Thus we would have a precisely given functional form for I_{ij}, where f_{ij} and $g(\varphi_i)$ have been both specified and calibrated.

Two comments can be made on the programming question raised here. First, objective (25), with inelastic demand for travel, can be reformulated as a mathematically equivalent problem of maximising «entropy» (Leonardi, 1980a; Coelho and Williams, 1978). See Appendix A. It should be realised that this is just a restatement of the first surplus objective. Second, given that I_{ij} is calibrated *a priori* it may well be possible to collapse constraints giving the definition of I_{ij} into the objective function removing an inherently non-linear constraint. Coelho and Williams (1978) employ both of these strategies with equation (32) and are able to show that the dual of a similar problem, with no binary variables, is relatively simple and sometimes unconstrained. Even with binary variables this may greatly simplify the search for a globally optimal location pattern for both (29) and (37) (Erlenkotter and Leonardi, 1980).

Finally, note both approaches assume greater benefits imply that more trips will be made, and fewer benefits lead to fewer trips. Particularly when considering the latter it is clear that this implies people are able to make a choice, and that they are able to reduce their trips if accessibility falls. In many cases of need, and also of ingrained habits, this would seem unrealistic, suggesting that this approach has more relevance for expansion than contraction questions.

2.4. *Equity questions*

There is no reason *a priori* to expect that maximising users benefits need imply achievement of an equitable distribution of facilities. There are two issues here. First, in maximising benefits, there will be a bias towards choosing those locations where a greatest increase in benefits can be achieved at least cost. These locations will be where the base level of demand is highest and increases most rapidly with greater accessibility, typical of populous and high income locations, and also where accessibility to the population as a whole is highest, The lower the budget relative to demand levels, the greater this bias is likely to be. This reinforces any bias due to the implicit assumption in these surplus measures that the marginal utility of income is constant.

Further, the question of whether facilities should be located in a «progressive» manner to redistribute real benefits in favor of the less well-off cannot be broached. There is, of course, a principle frequently evoked that if we concentrate on maximising benefits *a priori* then there is more to be redistributed *ex post.* However, this assumes that distribution and efficiency can be separated. Given that competitive systems seek efficiency only because of the expectation that individuals will reap the benefits of their own actions, the possibility of redistributing all the benefits of efficiency would itself reduce the competitive drive that is supposed to achieve efficiency. Thus maximising both efficiency and redistribution seems socially infeasible under competitive conditions. For these, and other reasons outlined below, it seems essential that some measure of equity or effective redistribution should be considered *a priori* as a possible goal for facility location, not just as an *ex post* «political» decision. This may then be logically extended to consider progressive equity solutions.

3. Redistribution and equity goals

It is axiomatic that any public facility location policy must have an effect on the distribution of benefits, or real income. Those living nearer desirable facilities will have their real income increased, while those further from desirable facilities or closer to undesirable facilities will have their real incomes decreased. Since government is supposedly mandated to redistribute income as well as provide public services (among a number of activities), it is curious that no attempt has been made to evaluate the redistributive consequences of facility locations. Welfare economic objectives are not sufficient to handle this problem as they also ignore fundamental redistributive consequences of «welfare-maximizing» objectives (Dobb, 1969; Lea, 1979). Since facility location solutions may conflict with redistribution objectives it seems obvious that the effects of the former on the latter be measured even if income redistribution is not desirable as a sole goal of public facility policies. This section will develop a simple measure of real income redistribution in the case of single purpose trips.

3.1. *A redistributional measure with elastic demand*

Consider consumers distributed in space such that the *per capita* income of those at location i is given by the deterministic variable Y_i. The demand for the public facility at j by this population is:

$$I_{ij} = O_i \, g(\varphi_i) \, f_{ij} \, \varphi_i^{-1}. \tag{43}$$

The expected cost of travel per trip, per time period is then given by:

$$\bar{c}_i = \sum_j h_{ij} \cdot c_{ij} \tag{44}$$

$$= \varphi_i^{-1} \cdot \sum_j f_{ij} \, c_{ij}. \tag{45}$$

Now the income accruing to individuals at i due to consumption of the public good, measured in real terms (i.e. in terms of units of the good) is given by the number of trips made $g(\varphi_i)$, multiplied by the amount bought per trip. We assume that the amount bought per trip is constant, given by α. Although this last assumption is not true for rational utility maximising consumers (Sheppard, 1980a), even the analytical solutions for this case break down for «free» goods consumed at the facility site, which characterise a number of public facilities. So this assumption is not an unreasonable starting place. Thus we take real income $= \alpha \cdot g(\varphi_i)$.

One measure of redistributive effects on income is:

$$y_i = (1 - \beta_i) Y_i + \gamma^k \cdot \alpha \cdot g(\varphi_i) \tag{46}$$

where β_i is the proportion of Y_i spent on travelling for the public good, γ^k is a multiplier converting real units of the good into monetary units, and y_i is the income net of redistribution effects. Note that:

$$\beta_i Y_i = g(\varphi_i) \cdot \bar{c}_i \, . \tag{47}$$

The multiplier γ^k is problematic to determine, but is essential if we are to develop a common measure of benefits that is translatable to different income groups. Essentially, money is being used as the measure of value to which heterogeneous goods are reduced. Other measures of value are of course also possible.

From (46) and (47):

$$y_i = Y_i + g(\varphi_i) \cdot (\gamma^k \cdot \alpha - \bar{c}_i). \tag{48}$$

In the case of inelastic demand for travel ($g(\varphi_i) = g$) this reduces to:

$$y_i = Y_i - g \cdot \bar{c}_i + k \tag{49}$$

where $k = g \cdot \gamma^k \cdot \alpha$.

It remains to determine g_i and α. α is an empirical constant; the average consumption per visit. g_i has been discussed previously, but now a further complication has been added. With different incomes we would expect g_i to depend on income and accessibility; $g(Y_i, \varphi_i)$. A number of functions are possible; however some reasonable restrictions can be imposed. First, we would expect that the proportion of income spent on obtaining the good, β_i, which itself is a function of travel cost, accessibility and income (35), would decrease as $\bar{c}_i$ increases. Partial differentiation of β_i with respect to $\bar{c}_i$ shows that this implies $\partial g_i / \partial \bar{c}_i < - g_i / \bar{c}_i$, or the elasticity of travel demand with respect to average trip cost is less than minus one. Similarly, if good k is a necessity, this would imply $\partial \beta_i / \partial Y_i < 0$, since low income groups would be constrained to spend a greater part of their money on it that high income groups. This would imply $Y_i \partial g_i / g_i \partial Y_i < -1$. If the elasticity of travel demand with respect to income were greater than minus one, $\partial \beta_i / \partial Y_i$ would be positive, suggesting a luxury public good.

3.2. *Programming issues*

To locate facilities under an egalitarian income redistribution policy:

$$\underset{L,M}{\text{Max}} - \sum_i y_i \ln y_i \tag{50}$$

subject to constraints (48), (45), (40), (41) and (42). Of these, (45) and (48) are definitions that may well be easily embedded into the objective function. The program is one striving for progressive equity, biasing location patterns in favor of the less well-off. Equation (50) is but one of many possible measures of inequality that could be used (Sen, 1973); one which is particularly sensitive to the extremes of the income distribution. Again it seems advisable to experiment with other measures to see how sensitive the optimal solution is to choice of inequality measure. Eventually, choice of the appropriate measure will probably depend on the biases of the investigator.

The formulations developed above still have many shortcomings in describing the benefits to users of any particular location pattern. Questions of negative externalities associated with being too close to certain facilities (such as hospitals and highways) are not discussed. The effects of income variation on travel patterns, and the measurement of Y_i (in real or monetary terms) and its deflation for trips to purchase other goods are not covered. However the more limited aim of broaching the subject has at least been tackled.

4. Users' behavior and the status quo

The are still some fundamental issues, left uncovered in the technical discussions above, relating to the spatial behavioral content and assumptions. It has been assumed that a model of interaction can be calibrated *a priori* and used as an input to the public facility problem. This is not a necessary assumption, since if we have insufficient information to describe detailed behavior we can use information theoretic methods to obtain least biased Bayesian prior estimates of the pairwise interactions that are consistent with the behavioral information which is available (Sheppard, 1976). In this case an entropy maximising sub-problem would replace the interaction equation in the programming model (Coelho and Williams, 1978).

There is, in fact, a sense in which it could be argued that it may be inadvisable to have a very precisely calibrated model of interaction behavior. This is because behavior may change in a fundamental manner after locating facilities; indeed we may wish it to do so, and this cannot be captured by projecting precisely given past behavioral patterns into the future. Essentially the interaction formulae represent a constraint to which location patterns must adjust. We assume past patterns should persist into the future, and indeed we reinforce this persistence by planning in conformity to it. This is the general *status quo* bias in planning (Olsson, 1974).

Essentially, I am arguing that allowing for elastic demand for travel, while capturing an element of hidden demand, may not capture all the changes in behavior. By locating a large park in a poor downtown area, this may not only increase the usage of parks due to their increased availability, but may cause a discrete shift in people's preferences (or culture) to a greater orientation toward park-based activities. Thus, for example, even if the park were closed down again peoples' usage of parks may remain higher than it was before. The reverse can also be true; making certain public services difficult to use by low income groups can effectively remove such services from their set of activities, and induce into their *apparent* preferences an observed underutilisation of these possibilities compared to other groups. Such effects will be more severe for low income groups, since they already face more severe spatial constraints. This undoubtedly accounts in part for why there are social class differences in utilising public services, particularly if one takes into account other features of these services that can make them seem alien to, say, the working class. In fact a progressive equity criterion for facility location can be used to try and counteract this type of social bias. Making services most accessible to the less well-off is one part of a policy that will encourage them to participate fully in those public services which they can benefit from.

A second strongly related issue is the implicit assumption, in taking spatial behavior as a datum, that observed behavior does represent revealed preferences. This is indeed made explicit in the benefit-measures discussed in section 2. As I have argued in detail elsewhere, there are many ways in which spatial constraints can influence observed patterns as strongly as preferences, and this is particularly true for those with lowest incomes. If we ignore this possibility and treat all behavior as representing relatively free choice we introduce a *status quo* bias into our results that can favor the more well-off (Sheppard, 1980b). The question of how issues like these can be introduced into public facility provision is an open one, but one of prime importance for future research.

5. Conclusion

A way of introducing elastic demand into public facility problems has been outlined, and used to discuss possible benefit measures. It should, however, be emphasised that this raises other problems which come to the fore if excessive reliance is placed on observed behavior. One other side issue that perhaps is worth pointing out is that the resulting models are different from least transport cost models, and do not reduce to these as distance friction becomes infinite. However, such a reduction should not be expected. With elastic travel demand, trips should converge to zero, rather than to patterns based on using the nearest facility. It seems that the latter convergence will only occur with a doubly-constrained model.

With regard to income redistribution, it could certainly be argued that past public facility location policies have reinforced income inequalities. Trends in the past location of such facilities with negative externalities as highways, hospitals, and drug treatment centers in poorer downtown areas, with parks and libraries in better parts of town could certainly be cited to support this. If governments really are interested in income distribution, then this should be an explicit dimension of all public policies if it is to be given a coherent treatment. This certainly applies to location problems.

Finally, it should be evident that the objectives proposed in this paper are not necessarily amenable to immediate translation into efficient optimal mathematical programming routines. However, the emphasis throughout has been on rigorous specification of the interactions to give a strong theoretical basis for a facility location problem oriented toward customer behavior. This has been done in the belief that it is better to start with good theory, so that it can be seen what compromises are necessary in the interests of practical results, rather than to start with a practical routine that may be only heuristically related to the forces operating in reality.

References

Abernathy W.J., Hershey J.C. (1972) A spatial-allocation model for regional health-services planning, *Operations Research, 20,* 629-642.

Alonso W. (1978) A theory of movements, in Hansen N. (ed.) *Perspectives on Structure, Change and Public Policy,* Ballinger, Cambridge, Massachusetts, 197-211.

Beaumont J.R. (1980) Spatial interaction models and the location-allocation problem, *Journal of Regional Science, 20,* 37-50.

Coelho J.D., Williams H.C.W.L. (1978) On the design of land use plans through locational surplus maximisation, *Papers of the Regional Science Association, 40,* 71-85.

Dobb M. (1969) *Welfare economics and the economics of socialism,* Cambridge University Press, Cambridge, UK.

Erlenkotter D. (1977) Facility location with price-sensitive demands: public, private and quasi-public, *Management Science, 24,* 378-386.

Erlenkotter D., Leonardi G. (1980) Algorithms for spatial interaction based location-allocation models, Paper presented at the IIASA task force meeting on public facility location, Laxenburg, Austria.

Lea A.C. (1979) Welfare theory, public goods and public facility location, *Geographical Analysis, 11,* 217-239.

Ledent J.(1980) Calibrating Alonso's general theory of movement: the case of interprovincial migration flows in Canada, *Sistemi Urbani, 2,* 327-358.

Leonardi G. (1980a) On the formal equivalence of some simple facility location models, Working Paper 80-21, IIASA, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Leonardi G. (1980b) A unifying framework for public facility location models, Working Paper 80-79, IIASA, International Institute for Applied Systems Analysis, Laxenburg, Austria.

London Health Planning Consortium (1979) *Hospital services in London: a planning profile,* HMSO, Her Majesty's Stationary Office, London.

Neuburger H.L.I. (1971) User benefit in the evaluation of transport and land use plans, *Journal of Transport Economics and Policy, 5,* 52-75.

Olsson G. (1974) Servitude and inequality in spatial planning: ideology and methodology in conflict, *Antipode, 6,* 1, 16-21.

Sen A. (1973) *On economic inequality,* Clarendon Press, Oxford.

Sheppard E.S. (1976) Entropy, theory construction and spatial analysis, *Environment and Planning A, 8,* 741-752.

Sheppard E.S. (1978) Theoretical underpinnings of the gravity hypothesis, *Geographical Analysis, 10,* 386-402.

Sheppard E.S. (1979) Geographic potentials, *Annals of the Association of American Geographers, 69,* 438-447.

Sheppard E.S. (1980a) Location and the demand for travel, *Geographical Analysis, 12,* 111-127.

Sheppard E.S. (1980b) The ideology of spatial choice, *Papers of the Regional Science Association, 45,*197-213.

Tapiero C.S. (1980) A probability model for the effects of distance on the demand for multiple facilities, *Environment and Planning A, 12,* 399-408.

Theil H. (1972) *Statistical decomposition analysis: with applications in the social and administrative sciences,* North Holland Publishing Company, Amsterdam.

Wagner J.L., Falkson L.M. (1975) The optimal nodal location of public facilities with price-sensitive demand, *Geographical Analysis, 7,* 69-83.

Appendix A

$$\sum_i O_i \log \sum_j e^{F_{ij}} = \sum_i O_i \log \varphi_i. \quad (A.1)$$

Now $$O_i = \sum_j I_{ij} \quad (A.2)$$

$$\varphi_i = O_i e^{F_{ij}} (I_{ij})^{-1} \qquad \text{from (20).} \quad (A.3)$$

Substituting (A.2) and (A.3) into (A.1):

$$\Delta S_1 = \sum_i \sum_j I_{ij} \log O_i e^{F_{ij}} (I_{ij})^{-1} \quad (A.4)$$

$$= \sum_i \sum_j I_{ij} \log O_i + \sum_i \sum_j I_{ij} F_{ij} - \sum_i \sum_j I_{ij} \log I_{ij}. \quad (A.5)$$

Now in (A.5) the first term is, by (A.2): $\sum_j O_i \log O_i$ which is an exogenous constant. Therefore it can be eliminated for maximisation purposes leaving a surplus function that is the sum of trip utilities and an entropy term. It is of interest to note that if F_{ij} is defined as expected utility from a multinomial choice model, then the second term of (A.5) has an interpretation equivalent to that for ΔS_2 of equation (37).

Riassunto. È chiara la necessità di introdurre il comportamento degli utenti esplicitamente come dato nei modelli di localizzazione dei servizi publici, poiché essi devono essere coerenti con lo schema di interazioni spaziali che gli utenti manifestano. In questo saggio si analizza come i concetti di interazione, accessibilità e surplus localizzativo possano essere generalizzati in modo da tener conto sia dell'elasticità della domanda totale che di altre assunzioni realistiche circa il comportamentoi degli utenti. Tuttavia, è dimostrato come tale approccio comportamentistico non garantisca in generale soluzioni più eque di quelle basate su più semplici criteri di costo. Viene proposto quindi un approccio multi-obiettivi, che tenga conto anche degli effetti di ridistribuzione del reddito indotti dai diversi assetti localizzativi.

Résumé. Il est évident qu'il est nécessaire d'introduire le comportement des usagers explicitement dans les modèles de localisation des services collectifs, car ils doivent être cohérents avec la configuration des interactions spatiales des usagers. Cet essai montre comme les concepts d'interaction, d'accessibilité et de surplus de localisation peuvent être généralisés de façon à considérer soit l'élasticité de la demande totale soit autres assumptions réalistes du comportament des usagers. Néanmoins, il est prouvé qu'une telle approche du comportement n'assure pas, en général, des solutions plus justes que celles obtenues considérant de simples critères de coût. Une approche multi-objectifs est donc proposée qui rende compte aussi des éffets de rédistribution du revenu induits par les différentes configurations de localisation.

Sistemi Urbani, 3, 1981, 455-470 Guida Editori, Napoli, Italia

Some proposals for stochastic facility location models

Y. Ermoliev

Systems and Decision Sciences Area, IIASA, International Institute for Applied Systems Analysis, Schlossplatz 1, Laxenburg, A-2361, Austria.

G. Leonardi

Human Settlements and Services Area, IIASA, International Institute for Applied Systems Analysis, Schlossplatz 1, Laxenburg, A-2361 Austria.

Received 15 June 1980, in revised form 10 January 1981

Abstract. The static facility location model with a spatial interaction-based allocation rule has been first introduced by Coelho and Wilson (1976). The main interest in introducing a spatial interaction-based allocation rule lies in the more realistic trip patterns that result from its use, which in many cases seem to fit the actual data on customer choice better than the simple nearest-facility allocation rule.

A further step towards more realistic models of customer behavior is the introduction of stochastic features, describing both the amount of total demand for facilities and the trip pattern of the customers. In this paper the usefulness of stochastic programming tools to formulate and solve such problems is explored, and some simple, but easily generalizable applied examples are given. Both numerical techniques and exact analytical methods are outlined, and some issues for further reseaarch are proposed.

Key words: static facility location, spatial interaction, stochastic programming, quasi-gradient methods.

1. Introduction

It is well known that a classical «plant location» model is based on very deterministic assumptions. The main limitation of such models is the customer-choice behavior embedded within, that is, the choice of the nearest facility. The need to introduce more realistic behavioral assumptions has been recognized by many authors, among them Coelho and Wilson (1976), Hodgson (1978), Beaumont (1979), and Leonardi (1978, 1980). In all the above references the sharp distance-minimizing behavior is replaced by a smoother spatial interaction (also known as «gravity») model, thus allowing for possible substitution effects across space. Since spatial interaction models have both theoretical and empirical justifications, their use in location modelling seems a promising one. However, the classical spatial interaction models solve only part of the problem. Although they are rooted on stochastic assumptions (Wilson, 1970; McFadden, 1974; Bertuglia and Leonardi, 1979), only the expected values of the underlying stochastic processes are used. A natural further step to be undertaken is therefore to introduce the stochastic behavior explicitly, thus allowing for both uncertainty in customer choice and uncertainty in the knowledge of demand.

The aim of this paper is to explore some of the problems arising when such stochastic features are introduced, as well as to suggest some numerical tools to solve the resulting problems. Due to the exploratory nature of the paper, the examples are kept as simple as possible. However, it is felt that the suggested approach is by far more general than the applications discussed here, and can be easily extended to more complex formulations without any big change in the required theory and tools.

2. Statement of the problem

In its most general form, the static deterministic facility location problem can be formulated as follows:

$$\max_{S,X,L} B(S) - \sum_{j \in L} f_j(x_j) \tag{1}$$

s. t.

$$\sum_{j \in L} S_{ij} = P_i \ , \qquad i \in M \tag{2}$$

$$\sum_{j \in L} S_{ij} = x_j \ , \qquad j \in L \tag{3}$$

$$X \in \Gamma \tag{4}$$

$$L \subseteq Z \tag{5}$$

where

i labels the demand locations, belonging to a given set M

j labels the facility locations, belonging to a set L, to be chosen among all subsets of a given set Z

$S = (S_{ij})$ is the array of total trips made by customers between each demand-facility location pair in the unit time

$X = (x_j)$ is the array of total service capacity (in terms of customers served per unit of time) to be established in each facility location belonging to L

$P = (P_i)$ is the array of total demand (in terms of customers to be served per unit of time) in each demand location belonging to M

Γ is the set of feasible X, accounting for possible physical and economic constraints to be met by the service capacities

$B(S)$ is a real valued function measuring the total benefit which accrues to the customers from a given trip pattern S

$f_j(x_j)$ are real valued functions measuring the cost of establishing a facility with capacity x_j in each location $j \in Z$.

The objective function (1) is therefore the total net benefit, being the difference between customer benefit and establishing costs. It has to be maximized by suitably choosing the subset of locations L, the facility sizes X, the trip pattern S. This choice is subject to:

a. constraint (2), requiring the total demand to be met;

b. constraint (3), requiring the total capacity to be fully used;

c. constraint (4), requiring the facility sizes to meet the physical and economic constraints;

d. constraint (5), requiring the subset of chosen location to belong to the set of possible locations Z.

The general formulation given above can be specialized in many ways, by introducing special assumptions for the functions $B(\cdot)$ and $f_j(\cdot)$ and for the structure of the set Γ (see Leonardi, 1980, for a review).

The simplest possible form of problem (1) - (5) is obtained by introducing the following assumptions:

a. The benefit function has the form

$$B(S) = -\sum_{ij} S_{ij} \ln S_{ij} - \beta \sum_{ij} C_{ij} S_{ij} \qquad (6)$$

where C_{ij} are the travel costs between each (i, j) pair, and β is a given nonnegative constant. Function (6) has been first introduced by Neuburger (1971) in transport planning evaluation and extended to location analysis by Coelho and Wilson (1976) and Coelho and Williams (1978). In the above references it is shown how this function has a sound economic interpretation, being the consumer surplus measure associated with the trip pattern (S_{ij}). Moreover, it has the useful property of embedding the spatial interaction model with an exponential discount factor, which usually has a good empirical fit on actual data.

b. The cost functions are linear and do not depend on the location

$$f_j(x_j) = ax_j, \qquad \forall j.$$

c. The set Γ is

$$\Gamma = \{X : X \geq 0\}$$

that is, no physical and economic constraints must be met, except for the obvious nonnegativity requirement on the size of the facilities.

After introducing the above assumptions and dropping the constant terms, the redundant variables, and constraints, problem (1)-(5) reduces to the much simpler one

$$\min \sum_{ij} S_{ij} \ln S_{ij} + \beta \sum_{ij} C_{ij} S_{ij} \tag{7}$$

s.t.

$$\sum_{j} S_{ij} = P_i \, . \tag{8}$$

Note that, due to the simple form of the cost functions, constraint (5) is no longer required, since an optimal solution will always have L = Z. The combinatorial features of (1)-(5) have thus disappeared, and the problem has been reduced to the smooth concave programming problem (7)-(8). The closed-form solution to (7)-(8) can be easily found to be

$$S_{ij} = P_i \frac{e^{-\beta C_{ij}}}{\sum_{j} e^{-\beta C_{ij}}} . \tag{9}$$

Equation (9) states that trips from demand locations to facilities are made according to a very simple production-constrained spatial interaction model (Wilson, 1971).

Problem (7)-(8) and equation (9) can be used as a starting point to build some simple stochastic generalizations. The first one is as follows. Let it be assumed that the behavior implied by (9) is deterministic, but the demand array P is not known in advance. This assumption is sensible in many long-term planning applications, where the trip behavior is known but the total demand may fluctuate. For instance, in a high school location problem the way customers will choose facilities from each demand location can be reasonably assumed to be known and deterministic, but the total number of students living in each demand location may change over time in an unpredictable way. However, the size of the schools cannot be changed as fast as demand changes, so the planning authority is possibly faced both with unsatisfied demand and overcrowding and with unused service capacity. The above problem can be stated in mathematical terms as follows. Let

$H_i(Y)$ be the distribution function of the total demand in demand location i; that is, if τ_i is the random variable giving the total demand in i, then $H_i(y) = \Pr\{\tau_i \leq y\}$

α_i^+ be the unit cost to be paid for an overestimate of the demand in i

α_i^- be the unit cost to be paid for an underestimate of the demand in i

x_i be the estimate of total demand in i, given by the decision maker.

Then, if $\tau_i \leq x_i$ an overestimate cost $\alpha_i^+(x_i - \tau_i)$ has to be paid, while if $\tau_i > x_i$ an underestimate cost $\alpha_i^-(\tau_i - x_i)$ has to be paid.

The resulting stochastic programming problem is

$$\min_{S,X} \sum_{ij} S_{ij} \ln S_{ij} + \beta \sum_{ij} C_{ij} S_{ij} + \tag{10}$$
$$+ \sum_i \left[\alpha_i^+ \int_0^{x_i} (x_i - y)\, dH_i(y) + \alpha_i^- \int_{x_i}^{\infty} (y - x_i)\, dH_i(y) \right]$$

s.t.

$$\sum_j S_{ij} = x_i \, . \tag{11}$$

The above generalization has been built on the assumption that the total demand is stochastic, while the trip behavior is deterministic. Let this assumption now be reversed, so that the total demand is deterministic, while the trip behavior is stochastic. This assumption can be easily introduced by suitably reinterpreting equation (9), which can be rewritten as follows:

$$S_{ij} = P_i\, q_{ij} \tag{12}$$

where

$q_{ij} = \dfrac{e^{-\beta C_{ij}}}{\sum_j e^{-\beta C_{ij}}}$ is the probability of choosing the destination j for a customer living in origin i.

The interpretation of the quantities q_{ij} defined above as probabilities is rooted on the theory of probabilistic choice behavior (McFadden, 1973). It has also been shown in Bertuglia and Leonardi (1979) that these quantities can be interpreted as steady-state distribution of a suitably defined Markov process. If the customers are assumed to be mutually independent, then (12) can be interpreted as the expected value of the number of trips between i and j, whose actual values have a multinomial distribution with parameters q_{ij}. Let v_{ij} be the actual (random) number of trips from i to j, and define

$\tau_j = \sum_i v_{ij}$ the total number of customers attracted in j

$H_j(y)$ the distribution function of τ_j; that is, $H_j(y) = \Pr\{\tau_j \leq y\}$.

The distribution functions $H_j(y)$ cannot be easily written in closed form, but random draws of τ_j can be computed using the probabilities q_{ij}. Let also the following costs and decision variables be introduced:

α_j^+ is the unit cost to be paid for an overestimate of the demand attracted in j

α_j^- is the unit cost to be paid for an underestimate of the demand attracted in j

x_j is the size of the facility in j.

Since the planned value x_j will be usually different from the actual demand τ_j, a cost $\alpha_j^+(x_j - \tau_j)$ will have to be paid when $\tau_j \leq x_j$ and a cost $\alpha_j^-(\tau_j - x_j)$ will have to be paid when $\tau_j > x_j$.

The resulting stochastic programming problem is

$$\min_X \sum_j \left[\alpha_j^+ \int_0^{x_j} (x_j - y)\, dH_j(y) + \alpha_j^- \int_{x_j}^{\infty} (y - x_j)\, dH_j(y) \right] . \tag{13}$$

Note that the spatial interaction embedding term has been dropped in the objective function, since the customer behavior is already accounted for by the way the distribution functions $H_j(y)$ are built. If $\alpha_j = \beta_j$, $j = \overline{1,n}$, then it can be shown that the solution to problem (13) is given by the *median* of the random vector $\{\tau_j\}$, which for very large values of P_i, $i = 1, m$, is closely approximated by the expected value, i.e.:

$$x_j^* = \sum_j P_j q_{ij}.$$

Although problems (10) - (11) and (13) look quite different, they belong to the same general form and can be solved with the same methods. A further generalization, allowing for a stochastic behavior of both the total demand and the trip behavior would still lead to the same problem form. The rest of this paper will be mainly concerned with problem (10) - (11) and its generalizations, but it must be kept in mind that the theory and the techniques which will be developed apply to problem (13), as well as to its generalizations.

3. The stochastic quasi-gradient method

In order to develop a computational method to solve problem (10)-(11), let it be further simplified. For given x_i by means of equations (9) the optimal values of the variables S_{ij} can be expressed in terms of the variables x_i:

$$S_{ij} = x_i \frac{e^{-\beta C_{ij}}}{\sum_j e^{-\beta C_{ij}}}. \tag{14}$$

Substitution of (14) in the objective function (10) yields:

$$\min_X F(X) \tag{15}$$

where the function F(X) is defined as:

$$F(X) = \sum_i x_i \ln x_i + \sum_i C_i x_i + \\ + \sum_i \left[\alpha_i^+ \int_0^{x_i} (x_i - y)\, dH_i(y) + \alpha_i^- \int_{x_i}^{\infty} (y - x_i)\, dH_i(y) \right] \tag{16}$$

and the constants C_i are given by

$$C_i = -\ln \sum_j e^{-\beta C_{ij}}. \tag{17}$$

The solution of problems like (15) gives rise to two usually difficult problems. First, although the objective function (16) in convex, it is in general nonsmooth. The possible nonsmoothness arises from the distribution functions $H_i(y)$. First, if they are discrete distributions, then $F(X)$ will not have continuous derivatives. Second, it is often difficult or impossible to compute the exact values of the integrals appearing in (16), unless for very special and well-behaved forms of the distribution functions $H_i(y)$. More often than not, such functions are defined not by a closed-form equation, but rather by means of a rule to generate random draws from them.

Such difficulties can be overcome by using direct stochastic programming methods, such as stochastic quasi-gradient methods (see Ermoliev, 1976, 1978 for a review). These methods are a straightforward generalization of the well-know gradient method of deterministic mathematical programming, can be used for quite arbitrary distributions $H_i(y)$, and require very simple computations. For instance, the stochastic quasi-gradient projection method gives rise to the following rule for generating successive approximations to the optimal solution of problem (15):

$$X^{(N+1)} = \max\{0, X^{(N)} - \rho_N \xi^{(N)}\} \tag{18}$$

for

$$N = 0, 1, \ldots..,$$

where

N is an iteration counter

$X^{(N)}$ is the Nth approximation to the solution vector of (15)

ρ_N is a step size, to be suitably chosen at each iteration

$\xi^{(N)} = \{\xi_i^{(M)}\}$ is a random vector, called the stochastic quasi-gradient of $F(X)$ at the point $X^{(N)}$.

The stochastic quasi-gradient of $F(X)$ at $X^{(N)}$ is defined as

$$\xi_i^{(N)} = \begin{cases} \ln e x_i^{(N)} + C_i + \alpha_i^+ , & \text{if } x_i^{(N)} \leq \tau_i^{(N)} \\ \ln e x_i^{(N)} + C_i - \alpha_i^- , & \text{if } x_i^{(N)} > \tau_i^{(N)} \end{cases} \tag{19}$$

where $\{\tau_i^{(N)}\}$ is a sequence of mutually independent random draws from the distributions $H_i(y)$.

The convergence of the sequence $X^{(N)}$, as computed by (18), to the optimal solution of problem (15) is based on the fact that the random vector $\xi^{(N)}$, as defined in (19), is a stochastic estimate of a *subgradient* of the function F (X). It will be briefly recalled (Rockafellar, 1970) that a subgradient $\hat{F}_X(X)$ of a convex function F (X) is a vector such that the inequality

$$F(Y) - F(X) \geq (\hat{F}_X(X),\ y - x)$$

holds for all y (here the outer brakets on the right-hand side denote the inner product of two vectors). A subgradient of a differentiable function F (X) is equal to the gradient

$$F_X(X) = \left(\frac{\partial F}{\partial x_1}, \ldots.., \frac{\partial F}{\partial x_n}\right).$$

It can be shown that the conditional mathematical expectation of

$$\xi^{(N)} = (\xi_1^{(N)}, \ldots.., \xi_n^{(N)}):$$

$$E\,(\xi^{(N)} | X^{(N)})$$

where E denotes expectation, is a subgradient of the function (16) at $X = X^{(N)}$. To do this one must reformulate the problem as a minimax stochastic programming problem and apply the well-known general results (Ermoliev, 1969, 1976, 1978; Ermoliev and Nurminski, 1980). It is easily seen that

$$\begin{aligned}\alpha_i^+ \int_0^{x_i} (x_i - y)\, dH_i(y) + \alpha_i^- \int_{x_i}^{\infty} (y - x_i)\, dH_i(y) &= \\ = E \max [\alpha_i^+ (x_i - \tau_i),\ \alpha_i^- (\tau_i - x_i)].\end{aligned} \qquad (20)$$

Substitution of (20) into (16) yields:

$$F(X) = \sum_i \{ x_i \ln x_i + C_i x_i + E \max [\alpha_i^+ (x_i - \tau_i),\ \alpha_i^- (\tau_i - x_i)] \}. \qquad (21)$$

The requirements under which the sequence $\{X^{(N)}\}$ converge with probability 1 to the solution of (15) are very weak. For instance, a set of sufficient conditions is

$$\rho_N \geq 0\ , \quad \sum_{N=0}^{\infty} \rho_N = \infty\ , \quad \sum_{N=0}^{\infty} [\rho_N]^2 < \infty\ , \quad |\tau_i| < \text{const.}, \qquad \forall i,$$

and such conditions can always be satisfied in applications.

4. Optimality conditions

The numerical method outlined in Section 3 is quite general and can be used no matter how ill-conditioned the distributions $H_i(y)$ are. If, however, these distributions are well-behaved enough, then one may try to develop the exact optimality conditions for problem (15), and possibly find a set of simple equations for the optimal solution.

The starting point to develop necessary and sufficient optimality conditions for problem (15) is to consider it as a minimax stochastic programming problem (21). The general optimality conditions for a stochastic programming problem have been studied in Wets (1974), Ermoliev (1976), and Ermoliev and Justremski (1979). However, the special structure of problem (21) can be exploited to develop the optimality conditions in a more convenient form. Minimization of (21) is a special case of the following more general problem:

$$\min_X Q(X) \tag{22}$$

where

$$Q(X) = \sum_i \left\{ x_i \ln x_i + C_i x_i + E \max_i \left[\sum_j a_{ij}(W)\, x_j + b_i(W) \right] \right\} \tag{23}$$

and

$$a_{ij}(W)\,,\quad b_j(W)$$

are random parameters.

Let, therefore, the optimality conditions for problems (22) be analyzed. Let $\delta = (\delta_1, \ldots.., \delta_n)$ be a vector with nonnegative components, $Q'_\delta(X)$ the directional derivative along the direction δ. Then at an optimal solution $X = X^*$ it must be

$$\lim_{\Delta \to 0} \frac{Q(X + \Delta\delta) - Q(X)}{\Delta} = Q'_\delta(X) = \sum_{i=1}^{n} (\delta_i \ln ex_i + C_i \delta_i) + f'_\delta(X) \geq 0 \tag{24}$$

where $\Delta > 0$,

$$f(X) = E\Psi(X, W)\,, \qquad \Psi(X, W) = \max_i \left[\sum_{j=1}^{n} a_{ij}(W)\, x_j + b_i(W) \right]$$

From this the following conclusion can be drawn: the components of an optimal solution are positive and (24) is satisfied for any direction

δ. Under suitable hypotheses one can assert something about the equalities:

$$E\Psi'_\delta(X, W) = \int \Psi'_\delta(X, W)\, dH(W)$$

and

$$\lim_{\Delta \to 0} \frac{f(X + \Delta\delta) - f(X)}{\Delta} = \lim_{\Delta \to 0} \int \frac{\Psi(X + \Delta\delta, W) - \Psi(X, W)}{\Delta}\, dH(W)$$

(the integrability of the function $\Psi(X, W)$ as a function of W is automatically assumed).

For instance, it is easy to obtain the estimations

$$\left| \frac{\Psi(X + \Delta\delta, W) - \Psi(X, W)}{\Delta} \right| = \frac{1}{\Delta} \left| \max_i \left[\sum_{j=1}^{n} a_{ij}(W)(x_j + \Delta\delta_j) + b_i(W) \right] - \right.$$

$$\left. - \max_i \left[\sum_{j=1}^{n} a_{ij}(W)x_j + b_i(W) \right] \right| = \frac{1}{\Delta} \left| \sum_{j=1}^{n} \left[a_{i^*_\Delta j}(W)(x_j + \Delta\delta_j) + b_{i^*_\Delta}(W) - \right. \right.$$

$$\left. \left. - a_{i^*j}(W)x_j - b_{i^*}(W) \right] \right| \leq \frac{1}{\Delta} \sum_{j=1}^{n} [\max \{ 0, a_{i^*_\Delta j}(W)(x_j + \Delta\delta_j) +$$

$$+ b_{i^*_\Delta}(W) - a_{i^*j}(W)x_j - b_{i^*}(W) \} + \max(0, a_{i^*j}(W)x_j + b_{i^*}(W) -$$

$$- a_{i^*j}(W)(x_j + \Delta\delta_j) - b_{i^*_\Delta} \}].$$

Since

$$\max \{ 0, a_{i^*_\Delta j}(W)(x_j + \Delta\delta_j) + b_{i^*_\Delta}(W) - a_{i^*j}(W)x_j - b_{i^*}(W) \} \leq$$

$$\leq \max \{ 0, a_{i^*_\Delta j}(W)(x_j + \Delta\delta_j) + b_{i^*_\Delta}(W) - a_{i^*_\Delta j}(W)x_j - b_{i^*_\Delta}(W) \} =$$

$$= \max \{ 0, a_{i^*_\Delta j}(W)\delta_j\Delta \} \leq |a_{i^*_\Delta j}(W)\delta_j|\Delta$$

and

$$\max \{ 0, a_{i^*j}(W)x_j + b_{i^*}(W) - a_{i^*_\Delta j}(W)(x_j + \Delta\delta_j) - b_{i^*_\Delta}(W) \} \leq$$

$$\leq \max \{ 0, a_{i^*j}(W)x_j + b_{i^*}(W) - a_{i^*j}(W)(x_j + \Delta\delta_j) - b_{i^*}(W) \} \leq$$

$$\leq |a_{i^*j}(W)\delta_j|\Delta$$

then

$$\frac{1}{\Delta} \left| \Psi(X + \Delta\delta, W) - \Psi(X, W) \right| \leq \sum_{i,j} \left| a_{ij}(W) \right| \left| \delta_j \right|$$

and from the existence of Ea_{ij} for all i,j and the Lebesque convergence theorem one gets

$$\lim_{\Delta \to 0} \frac{f(X + \Delta\delta) - f(X)}{\Delta} = f'_\delta(X) = \int \Psi'_\delta(X, W)\, dH(W) = E\Psi'_\delta(X, W).$$

As is well known

$$\Psi'_\delta(X, W) = \max_{g \in G(X,W)} (\alpha, \delta)$$

where

$$G(X,W) = Co\{a^k(W), \ k \in K(X, W)\},$$

$$a^k(W) = (a_{k1}(W), \ldots\ldots, a_{kn}(W)),$$

$$K(X, W) = \{k : (a^k(W), X) + b_k(W) = \max_i [(a^i(W), X) + b_i(W)]\}.$$

Here Co denotes the set of all linear combinations of the argument vectors. Taking into account this fact, the condition (24) is replaced by

$$\sum_{j=1}^{n} (\delta_j \ln ex_j + C_j \delta_j) + E \max_{g \in G(X,W)} (g, \delta) \geq 0$$

or

$$\sum_{j=1}^{n} (\delta_j \ln ex_j + C_j \delta_j) + \max_{g(X,W) \in G(X,W)} E(g(X, W), \delta) \geq 0,$$

or

$$\max_{g(X,W) \in G(X,W)} E(\ln eX + C + g(X, W), \delta) \geq 0 \tag{25}$$

where

$$\ln eX = (\ln ex_1, \ldots.., \ln ex_n)$$

$$C = (C_1, \ldots.., C_n).$$

Since the condition (25) is fulfilled for any δ, there exist a $g(X, W) \in G(X, W)$ such that

$$\ln eX + C + E\,g(X, W) = 0. \tag{26}$$

Let us now return to the original problem (15) or (21). For this problem $W = (\tau_1, \ldots.., \tau_n)$,

$$\Psi(X, W) = \sum_{i=1}^{n} \max\{\alpha_i^+ (x_i - \tau_i),\ \alpha_i^- (\tau_i - x_i)\}$$

$$G(X, W) = (G_1(X, W) \times \ldots.. \times G_n(X, W))$$

$$G_i(X, W) = \mathrm{Co}\{\alpha_i^k,\ k \in K(X, W)\}$$

where

$$\alpha_j^1 = \alpha_i^+ , \qquad \alpha_i^2 = -\alpha_i^-$$

$$K(X, W) = \begin{cases} \{1\} & \text{with probability } P\{x_i \geq \tau_i\} \\ \{2\} & \text{with probability } P\{x_i < \tau_i\} \\ \{1{,}2\} & \text{with probability } P\{x_i = \tau_i\}. \end{cases}$$

Then from (26) one can obtain the following optimality conditions for the original problem (15): if a point X is an optimal solution, then and only then do multipliers $0 \leq \gamma_i \leq 1$ exist such that

$$\ln ex_i + C_i + \alpha_i^+ H_i(x_i) - \alpha_i^- [1 - H_i(x_i)] + [\gamma_i \alpha_i^+ - (1-\gamma_i)\alpha_i^-]\, dH_i(x_i) = 0.$$

Notice, that similar conditions are mentioned in Ermoliev and Justremski (1979). In particular, if $dH_i(x_i) = 0$ at an optimal solution, or if the distributions $H_i(X)$ are continuous, then one obtains

$$\ln ex_i + C_i + \alpha_i^+ H_i(x_i) - \alpha_i^- [1 - H_i(x_i)] = 0 , \qquad i = \overline{1, n}$$

or

$$H_i(x_i) = \frac{\ln e x_i + C_i + \alpha_i^-}{\alpha_i^+ + \alpha_i^-} , \qquad i = \overline{1, n}.$$

From these equations and for some kinds of distributions $H_i(X)$ it is possible to obtain a closed form for the optimal solution, or at least to compute a good approximate solution by using simple numerical techniques. In the general case with known distributions $H_i(y)$, the generalized gradient method can be used (see Ermoliev, 1976 and 1978):

$$x_i^{N+1} = x_i^N + \rho_N [\ln e x_i^N + C_i + \alpha_i^+ H_i(x_i^N) - \alpha_i^- (1 - H_i(x_i^N)) +$$

$$+ (\gamma_i^N \alpha_i^+ - (1 - \gamma_i^N) \alpha_i^-) \, d H_i(X^N)] , \qquad i = \overline{1, n}$$

where ρ_N, γ_i^N satisfy the sufficient convergence conditions $\rho_N \geq 0$, $\rho_n \to 0$, $\sum_{N=0}^{\infty} \rho_N < \infty$, $0 \leq \gamma_i^N \leq 1$. The values of ρ_N and γ_i^N can be chosen in order to decrease the objective function value. This problem has some important peculiarities: there is a closed form for the set of subgradients and computing the subgradients is easier than computing the values of the objective function. This gives us the opportunity to construct descent methods of nondifferentiable optimization as well as nondescent ones.

5. Concluding comments and issues for further research

The examples discussed in the foregoing sections have been kept as simple as possible, in order to introduce the proposed methods in the easiest way. When some of the simplifying assumptions are dropped, some new and more realistic models are obtained.

One possible path towards generalization is the introduction of more complex cost functions and constraints. For instance, the assumption on linear homogeneous establishing costs can be generalized to linear nonhomogeneous establishing costs

$$f_j(x) = a x + b \quad \text{if} \quad x > 0 , \qquad f_j(x) = 0 \quad \text{if} \quad x = 0 .$$

Such cost functions introduce a fixed charge b to be paid when a facility is established, independently of its size.

The optimization problem assumes therefore combinatorial features, since in this case the decision of which locations to choose is no longer trivial. On the other hand, this generalization is realistic, since it models the economies of scale often found in real services very well. Research on this kind of problem is ongoing, and some first numerical results have already been produced in Ermoliev *et al* (1981).

Another example of possible further research would be to impose more constraints on the sizes of facilities. Some typical and usually required constraints are the limits placed on both the size of facilities and the total budget, or total capacity to be allocated. For instance, schools have usually a minimum feasible size, below which it is not reasonable to build and sometimes a maximum feasible size as well (e.g., when the available space is limited).

Another generalization is obtained by introducing many types of facilities, to be located at the same time. Using the school example again, one may be concerned with locating high schools for different specialities and trainings. All of the above constraints still hold for each type of school. Moreover, some new constraints due to interactions within different schools may be needed. For instance, total demand for each type of school may not be known in advance, and customers may be allowed to choose both the location and the type of schools. This introduces a competition among different schools. Another obvious competition arises from limited available space in each location.

When all the above generalizations are introduced, the resulting model looks much more complicated than the ones discussed in this paper. However, it still belongs to the class of stochastic programs with linear constraints discussed in Ermoliev (1976) and Wets (1974), for which theoretical results and algorithms are available. Some applications of stochastic programming to such location problems are in progress, and they will be the subject of a forthcoming IIASA Working Paper.

References

Beaumont J.R. (1979) Some issues in the application of mathematical programming in human geography, Working Paper 256, School of Geography, University of Leeds, UK.

Bertuglia C.S., Leonardi G. (1979) Dynamic models for spatial interaction, *Sistemi Urbani, 1,* 2, 3-25.

Coelho J.D., Williams H.C.W.L. (1978) On the design of land use plans through locational surplus maximization, *Papers of the Regional Science Association, 40,* 71-85.

Coelho J.D., Wilson A.G. (1976) The optimum location and size of shopping centres, *Regional Studies, 10,* 413-421.

Ermoliev Y.M. (1969) On the stochastic quasi-gradient method and stochastic quasi-feuer sequences, *Kibernetica, 2.*

Ermoliev Y.M. (1976) *Stochastic programming methods,* Nanka, Moscow.

Ermoliev Y.M. (1978) Methods of nondifferentiable and stochastic optimization and their applications, Working Paper 78-62, IIASA, Laxenburg, Austria.

Ermoliev Y.M., Justremski V. (1979) *Stochastic models and methods in economic planning,* Nanka, Moscow.

Ermoliev Y.M., Leonardi G., Vira J. (1981) The stochastic quasi-gradient method applied to a facility location problem, Working Paper 81-14, IIASA, Laxenburg, Austria.

Ermoliev Y.M., Nurminski E.A. (1980) Stochastic quasi-gradient algorithms for minimax problems, in Demster M. (ed.) *Proceedings of the International Conference on Stochastic Programming,* Academic Press, London.

Hodgson M.J. (1978) Towards more realistic allocation in location-allocation models: an interaction approach, *Environment and Planning A, 10,* 1273-1285.

Leonardi G. (1978) Optimum facility location by accessibility maximizing, *Environment and Planning A, 10,* 1287-1305.

Leonardi G. (1980) A unifying framework for public facility location problems, Working Paper 80-79, IIASA, Laxenburg, Austria.

McFadden D. (1973) Conditional logit analysis of qualitative choice behavior, in Zarembka P. (ed.) *Frontiers in Econometrics,* Academic Press, New York.

McFadden D. (1974) The measurement of urban travel demand, *Journal of Public Economics, 3,* 303-328.

Neuburger H.L.I. (1971) User benefit in the evaluation of transport and land use plans, *Journal of Transport Economics and Policy, 5,* 52-75.

Rockafellar R.T. (1970) *Convex Analysis,* Princeton University Press, Princeton, New Yersey.

Wets R. (1974) Stochastic programs with fixed resources: the equivalent deterministic program, *SIAM Review, 16,* 309-339.

Wilson A.G. (1970) *Entropy in urban and regional modelling,* Pion, London.

Wilson A.G. (1971) A family of spatial interaction models, and associated development, *Environment and Planning A, 3,* 1-32.

Riassunto. Il modello di localizzazione statico con la regola di allocazione basata sull'interazione spaziale è stato proposto per la prima volta da Coelho e Wilson (1976). L'uso di un modello di interazione spaziale per assegnare gli utenti ai servizi produce schemi di spostamenti più realistici e più vicini ai dati sperimentali che non la usuale regola di assegnazione al servizio più vicino.
Un ulteriore passo verso la costruzione di modelli più realistici del comportamento degli utenti è costituito dall'introduzione di aspetti stocastici, inerenti sia l'ammontare totale della domanda di servizi, sia il processo di scelta delle destinazioni da parte degli utenti. Questo saggio conduce una prima analisi della possibilità di usare i metodi della programmazione stocastica per risolvere i problemi localizzativi, e discute alcuni semplici esempi e loro dirette generalizzazioni. Viene delineata la struttura generale degli algoritmi risolutivi e vengono forniti alcuni risultati analitici esatti. Infine, vengono proposti alcuni temi per approfondimenti futuri.

Résumé. Le modèle de localisation statique dont la règle de localisation basée sur l'interaction spatiale a été conçu par Coelho et Wilson (1976). L'utilisation d'un modèle d'interaction spatiale pour assigner les usagers aux services produit des configurations de déplacements plus réalistes et plus proches aux données expérimentales que la règle usuelle d'assignement au service le plus proche. Un ultérieur pas vers la construction de modèles plus réalistes du comportement des usagers est constitué de l'introduction des aspects stocastiques, concernants soit le montant total de la demande pour les services, soit le processus de choix des destinations des usagers. Cet essai analyse quelques unes des possibilités d'utilisation des méthodes de la programmation stocastique pour resoudre les problèmes de localisation et décrit quelques simples exemples et leurs directs généralisations. On décrit ensuite la structure générale des algorithmes résolutifs et on fournit quelques résultats analytiques precis. A la fin, on propose quelques thèmes qui peuvent être ultérieurement développés.

PUBLICATIONS IN THE PUBLIC FACILITY LOCATION SERIES

1. Giorgio Leonardi, On the Formal Equivalence of Some Simple Facility Location Models. WP-80-21.
2. Tony J. Van Roy and Donald Erlenkotter, A Dual-Based Procedure for Dynamic Facility Location. WP-80-31.
3. Donald Erlenkotter, On the Choice of Models for Public Facility Location. WP-80-47.
4. Giorgio Leonardi, A Multiactivity Location Model with Accessibility- and Congestion-Sensitive Demand. WP-80-124.
5. Yuri Ermoliev and Giorgio Leonardi, Some Proposals for Stochastic Facility Location Models. WP-80-176.
6. Giorgio Leonardi and Cristoforo Sergio Bertuglia, Optimal High School Location: First Results for Turin, Italy. WP-81-5.
7. Yuri Ermoliev, Giorgio Leonardi, and Juhani Vira, The Stochastic Quasi-Gradient Method Applied to a Facility Location Problem. WP-81-14.
8. Giorgio Leonardi, The Use of Random-utility Theory in Building Location–Allocation Models. WP-81-28.
9. John Beaumont, Towards a Comprehensive Framework for Location–Allocation Models. CP-81-22.
10. Donald Erlenkotter and Giorgio Leonardi, Facility Location with Spatially Interactive Travel Behavior. WP-81-97.
11. Giorgio Leonardi, A Unifying Framework for Public Facility Location Problems. RR-81-28. Reprinted from *Environment and Planning A* 13:1001–1028, 1085–1108, 1981.